UNIVERSIDAD DE CANTABRIA

DEPARTAMENTO DE ECONOMÍA

TESIS DOCTORAL

MODELOS DE COEFICIENTES VARIABLES CON DATOS DE PANEL: ESTIMACIÓN SEMIPARAMÉTRICA DIRECTA Y SU APLICACIÓN EN LA MODELIZACIÓN DEL COMPORTAMIENTO DE LOS INDIVIDUOS

ALEXANDRA SOBERÓN VÉLEZ

2014

University of Cantabria

Department of Economics

DOCTORAL THESIS

Panel data varying coefficient models: direct semi-parametric estimation and its application in modeling the individuals behavior

Alexandra Soberón Vélez

Santander, 2014

Supervisor: Juan Manuel Rodríguez Poo

Universidad de Cantabria

Departamento de Economía

TESIS DOCTORAL

Modelos de coeficientes variables de datos de panel: estimación directa semi-paramétrica y su aplicación en la modelización del comportamiento de los individuos

Alexandra Soberón Vélez

Santander, 2014

Director: Juan Manuel Rodríguez Poo

A mi familia

Agradecimientos

No sería justo por mi parte iniciar este trabajo sin antes mostrar mi más sincero agradecimiento a todas aquellas personas que, directa o indirectamente, han contribuido de algún modo a la realización de esta tesis doctoral.

En primer lugar, a mi director de tesis, el profesor Dr. D. Juan Manuel Rodríguez Poo, por la confianza que ha depositado en mí y por el afán de superación que ha sabido transmitirme a lo largo de estos años de intenso trabajo. Su excelente labor de dirección, sus continuas palabras de aliento y, muy especialmente, su paciencia han sido determinantes para poder llevar a cabo con éxito esta labor de investigación.

En segundo lugar, a la profesora Dr. Dña Ana Carrera Poncela por abrirme las puertas a este mundo académico tan apasionante, por su paciencia y por el continuo apoyo que, de manera totalmente desinteresada, he recibido de su parte tanto en los buenos como en los malos momentos.

Asimismo, a mis compañeros del Área de Econometría y de Hacienda Pública por los valiosos consejos profesionales, académicos y personales que me han proporcionado, así como al resto de miembros del Departamento de Economía por todo el apoyo prestado. Quisiera agradecer especialmente a mis compañeros y amigos Jesús Collado, Marta Echezarreta, Adolfo Fernández, Valeriano Martínez y Patricia Moreno por su ayuda incondicional, por la tranquilidad que me han aportado en ciertas situaciones y por los muchos momentos divertidos.

Expresar también mi gratitud a los profesores Juan A. Cuesta-Albertos, Miguel Delgado, Wenceslao González-Manteiga, Peter Hall, Oliver Linton y Stefan Sperlich. Sus valiosos comentarios y sugerencias han contribuido a mejorar considerablemente la calidad de esta tesis doctoral.

No puedo olvidar tampoco todos aquellos momentos de mil risas y lágrimas compartidos con mis amigos de toda la vida, en especial Ainhoa, Jana, Omayra y Silvia. Igualmente, agradezco a mis amigos de Potes por todo el estímulo que me han proporcionado en esta ardua carrera y por haber confiado en mí en momentos en los que ni yo misma lo hacía.

Reservo un lugar muy especial para mi familia. Nunca podré expresar con palabras mi agradecimiento por todo lo que han hecho por mí en todos estos años. Su apoyo y amor incondicional son los pilares fundamentales de mi vida. A mis padres, Pili y Manolo, y a mi hermano Fran por su paciencia y comprensión en momentos muy complicados y, muy especialmente, a mi abuela Concha. Su experiencia vital, su buen hacer y su dedicación a la familia la han convertido en un ejemplo a seguir de esfuerzo, integridad y fortaleza.

Gracias a todos vosotros por haber estado siempre ahí. Vuestros consejos, apoyo y palabras de aliento han sido determinantes para que este proyecto haya podido llegar a buen puerto y para que yo me haya convertido en la persona que soy ahora.

Contents

List of Tables

List of Figures

Introducción

En las últimas cinco décadas, la complejidad de los modelos econométricos se ha visto considerablemente enriquecida por la disponibilidad de datos de panel. Dado que como este tipo de datos se caracteriza por la observación de un conjunto de individuos (hogares, consumidores, países,...) a lo largo del tiempo, nos permiten extraer cierta información desconocida sobre las características idosincráticas de los individuos. De este modo, la habilidad teórica de este tipo de datos para aislar el impacto de acciones no observadas de los individuos nos permite realizar inferencia consistente sobre una gran variedad de cuestiones que no es posible con otro tipo de datos, como los de sección cruzada o los de series temporales.

Tradicionalmente, la especificación econométrica de este tipo de modelos se ha centrado en el análisis de la relación existente entre una variable endógena y ciertas variables explicativas, teniendo en cuenta la heterogeneidad individual no observable y basándose en supuestos bastante restrictivos sobre las formas funcionales y las densidades. Sin embargo, como se destaca en Wooldridge (2003), estos supuestos suelen ser bastante poco realistas, y existen situaciones en las cuales el riesgo de cometer errores de especificación es elevado. Si este es el caso, los estimadores estándar basados en condiciones de momento están sesgados y su uso puede invalidar los resultados de inferencia.

En este contexto, las técnicas de regresión no paramétricas se han convertido en un instrumento de gran utilidad a la hora de hacer frente a estos problemas. Los modelos de datos de panel totalmente no paramétricos son muy atractivos dado que no realizan supuesto alguno sobre la especificación del modelo, sino que permiten que sean los propios datos los que dibujen la forma de la función de regresión. Sin embargo, aunque este tipo de estimadores son robustos a la incorrecta especificación de la función de regresión, también están sujetos a la maldición de la dimensionalidad. En otras palabras, a medida que el número de variables explicativas aumenta la tasa de convergencia de estos estimadores se ve dramáticamente reducida. Con el objetivo de lograr una mejora en la tasa de convergencia de estos estimadores, la solución propuesta es la incorporación de un componente totalmente paramétrico en el modelo. Estos son los llamados modelos semi-paramétricos. En este caso, lo que es conocido de investigaciones empíricas previas o de la propia teoría económica se modeliza de manera paramétrica, mientras que lo que es desconocido para el investigador se especifica no paramétricamente.

Sin embargo, estos modelos flexibles son incapaces de capturar ciertas características ocultas en los conjuntos de datos. Esta es la principal razón de por qué muchos estudios empíricos han fomentado la introducción de estructuras más flexibles que permitan la variación de los parámetros desconocidos en función de ciertas variables explicativas. En esta situación, los modelos de coeficientes variables originalmente propuestos en Cleveland et al. (1991) aparecen como solución. En concreto, los modelos de coeficientes variables se ca- racterizan por permitir que ciertos coeficientes de la regresión varíen en función de ciertas variables exógenas propuestas por la teoría económica. De este modo, son capaces de explotar la información contenida en el conjunto de datos.

Señalar que en los últimos 15 años los modelos de coeficientes variables han experimentado un desarrollo sin precedentes, tanto desde el punto de vista metodológico como teórico. Dado que abarcan características tanto de los modelos no paramétricos como de los semi-paramétricos, ofrecen un marco general para solventar buena parte de los problemas de especificación de estos modelos. Con el objetivo de tener un mejor entendimiento sobre las principales ventajas ofrecidas por los modelos de coeficientes variables para el análisis empírico, presentamos una serie de aplicaciones.

En la literatura sobre los rendimientos educativos encontramos un primer ejemplo de la mejora conseguida con este tipo de modelos. Como se destaca en Schultz (2003), los efectos marginales de la educación varían en función de distintos niveles de experiencia laboral. De este modo, la omisión de la forma no lineal de la educación junto con el efecto de interacción entre educación y experiencia laboral nos conduce a resultados infrasuavizados sobre el rendimiento de la educación, como se encuentra en Card (2001). En esta situación, un modelo semi-paramétrico de coeficientes variables de la siguiente forma puede ser más recomendable,

$$Y_i = m(Z_i)X_i + W_i^\top \beta + v_i, \quad i = 1, \cdots, N,$$

donde Y_i representa el salario por hora del individuo i (log), X_i denota los años de educación del individuo como una medida de su experiencia educativa, Z_i es una medida de la experiencia laboral, y W_i es un vector de variables de control que incluye indicadores binarios para ciertas características de los individuos: estado civil, tipo de trabajo realizado (ser o no funcionario) o afiliación a un sindicato, entre otras. De este modo, esta modelización nos permite resolver buena parte de los problemas más habituales de esta literatura. Permite que el impacto de la educación sobre el salario varíe en función del nivel de experiencia laboral y, al mismo tiempo, tiene en cuenta la forma no lineal de la educación.

Asimismo, la literatura sobre macroeconomía y economía internacional proporciona otro ejemplo relevante. Cuando examinamos el papel de la inversión directa extranjera (FDI) sobre el crecimiento económico de los países, autores como Kottaridi and Stengos (2010) resaltan que, dado que el efecto positivo de la FDI sobre el crecimiento económico sólo tiene lugar en aquellos países con un mayor nivel de ingresos, en un modelo de crecimiento el coeficiente asociado al flujo de FDI debe ir variando en función del nivel ingreso inicial de cada país. En esta situación, el siguente modelo semi-paramétrico de coeficientes variables es especialmente atractivo dado que nos permite tener en cuenta tanto el impacto no lineal de la FDI como el efecto de interacción entre FDI y el nivel de ingreso,

$$Y_{it} = \alpha_0 + \alpha_1 D_j + \alpha_2 ln(I^d_{it}/Y) + \alpha_3 ln(n_{it}) + \alpha_4(lnXit)(I^f_{it}/Y) + \alpha_5 h_{it} + \epsilon_{it},$$

para $i = 1, \cdots, N$ y $t = 1, \cdots, T$, donde Y_{it} es la tasa de crecimiento del ingreso per cápita en el país i y el período t, I^d_{it}/Y la tasa de inversión doméstica respecto del PIB, n_{it} la tasa de crecimiento de la población, h_{it} el capital humano, I^f_{it}/Y el ratio FDI/PIB y X_{it} el ingreso per cápita al comienzo de cada período.

Finalmente, encontramos otro ejemplo relevante cuando tratamos de determinar el papel de los recursos naturales en el desarrollo económico de las regiones. Como es bien sabido en la literatura sobre economía del desarrollo, las regiones con abundantes recursos tienden a crecer más lentamente que aquellas con recursos excasos; ver Sachs and Warner

(2001). Sin embargo, dado que dentro de los países ricos en recursos naturales nos encontramos tanto con países ganadores como perdedores en términos de crecimiento, el papel de la calidad de las instituciones puede ser un factor decisivo a la hora de determinar el impacto de dichos recursos. De este modo, con el objetivo de tener en cuenta el hecho de que la variación en los resultados de crecimiento entre las regiones con abundantes recursos naturales depende de cómo las rentas de estos recursos son distribuidas a través del acuerdo institucional existente, en Fan et al. (2010) se propone el siguiente modelo semi-paramétrico de coeficientes variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4(Z_i) X_{4i} + v_i, \quad i = 1, \cdots, N,$$

donde Y_i representa la tasa de crecimiento media del país i, X_{1i}, X_{2i}, X_{3i} y X_{4i} denotan el nivel de ingreso inicial, grado de apertura, inversiones y abundancia de recursos, respectivamente, y Z_i es una medida de la calidad institucional aproximada via el procentaje de exportaciones primarias respecto del PIB en 1970.

En este contexto, el objetivo que persigue esta tesis doctoral tiene una doble vertiente. Por un lado, desarrollar nuevas técnicas de estimación que nos permitan obtener estimadores consistentes para modelos de coeficientes variables de datos de panel en los cuales la heterogeneidad individual no observable está correlacionada con algunas covariables. Por otro lado, resaltar las ganancias ofrecidas por esta nueva metodología para el análisis empírico. Con este objetivo, se presenta un análisis no paramétrico de un modelo estructural sobre la evolución de los ahorros preventivos de los hogares españoles bajo distintas fuentes de incertidumbre. De este modo, esta tesis doctoral se divide en cinco capítulos y la estructura de la discusión es la siguiente.

En el Capítulo 1, se realiza una revisión intensiva de la literatura econométrica sobre modelos de datos de panel semi-paramétricos y totalmente no paramétricos. Primero, los modelos de datos de panel totalmente no paramétricos son analizados tanto con efectos fijos como aleatorios. Posteriormente, se repasan los modelos parcialmente lineales bajo tres especificaciones distintas: efectos fijos y aleatorios, y presencia de covariables endógenas. Concluimos con una revisión sobre los modelos de coeficientes variables de datos de panel. Para cada una de estas áreas, discutimos tanto el modelo básico a estimar como la metodología propuesta. Además, también analizamos las principales propiedades asintóticas de los estimadores propuestos.

En el Capítulo 2, presentamos una nueva técnica de estimación para modelos de datos de panel en los cuales los coeficientes a estimar son funciones suaves de otras variables explicativas establecidas por la teoría económica y donde los efectos individuales están arbitrariamente correlacionados con los regresores del modelo de forma desconocida. Como se puede apreciar, la estimación directa a través de técnicas no paramétricas nos proporciona estimadores inconsistentes de los parámetros (funciones) de interés. Para resolver esta situación, recurrimos a una transformación en primeras diferencias. De este modo, siguiendo la idea original propuesta en Yang (2002) para un contexto totalmente distinto, el estimador propuesto se basa en la regresión local lineal de un modelo en primeras diferencias. Como se demuestra en Lee and Mukherjee (2008), dado que originalmente la ecuación de regresión transformada se localiza alrededor de un valor fijo de la muestra, sin considerar el resto de valores, la aplicación directa de técnicas de regresión local lineal a transformaciones en diferencias de modelos de datos de panel genera un sesgo que no desaparece asintóticamente. Para evitar este problema, proponemos recurrir a una ponderación de kernel de mayor dimensión.

Desafortunadamente, esta técnica nos permite eliminar el sesgo, pero al precio de aumentar el término de la varianza de modo que los estimadores resultantes alcanzan una más lenta tasa de convergencia no paramétrica. Para solventar este problema, proponemos un algoritmo de backfitting de una etapa que nos proporciona estimadores que alcanzan la tasa óptima de convegencia de este tipo de problemas. Además, los estimadores resultantes exhiben la propiedad de eficiencia oráculo. En otras palabras, esta técnica nos proporciona un estimador no paramétrico cuya matriz de varianzas y covarianzas de sus componentes es asintóticamente la misma que si el resto de componentes de la ecuación de regresión transformada fueran conocidos. En este capítulo también obtenemos la distribución asintótica de los dos estimadores. Asimismo, dado que la matriz de anchos de banda juega un papel crucial en la estimación consistente de las funciones desconocidas, también proporcionamos un método para calcular esta matriz de manera empírica. Finalmente, a través de un experimento de Monte Carlo comprobamos el buen comportamiento del estimador propuesto en muestras finitas.

En el Capítulo 3, se presenta un estimador basado en una transformación en desviaciones respecto a la media del modelo de regresión. Al igual que ocurre en el ajuste totalmente paramétrico, la principal ventaja de esta transformación respecto a cualquier otra similar es que nos permite obtener las propiedades asintóticas estándar de los estimadores no paramétricos bajo el supuesto de que los términos de error idiosincráticos son *i.i.d.* De este modo, el estimador de efectos fijos que se presenta en este capítulo se obtiene a través de una aproximación local sobre las T funciones aditivas resultantes de esta transformación en desviaciones respecto de la media, donde T es el número de observaciones por cada individuo. Dado que el uso directo de técnicas de aproximación local estándar en la estimación de estos componentes aditivos omite información tan relevante como la suma de las distancias existentes entre un término fijo y los otros valores de la muestra, genera un problema de sesgo más relevante incluso que el del Capítulo 2. Con el objeto de solventar este problema, proponemos considerar una aproximación local alrededor del vector completo de observaciones temporales de cada individuo.

No obstante, el bien conocido equilibrio entre sesgo y varianza vuelve a aparecer. Aunque el uso de una función de kernel de dimensión T resuelve el problema del sesgo, el término de la varianza se ve incrementado considerablemente de modo que el estimador resultante muestra una tasa de convergencia incluso más lenta que la de la transformación en primeras diferencias. Para resolver esta situación y obtener estimadores que alcancen la tasa óptima de convegencia de este tipo de problemas no paramétricos, recurrimos a un algoritmo de backfitting de una etapa. Analizando la distribución asintótica de este estimador de dos etapas apreciamos que un suavizado adicional, introducido a través del procedimiento de backfitting, no logra reducir el sesgo pero permite disminuir la varianza, tal y como se establece en Fan and Zhang (1999). De este modo, este procedimiento de backfitting permite a los estimadores de las funciones desconocidas de interés alcanzar tasas óptimas de convergencia no paramétricas. Asimismo, mostramos que el estimador de efectos fijos propuesto también es oráculo eficiente. Finalmente, a través de un experimento de Monte Carlo se trata de confirmar los resultados teóricos del estimador de efectos fijos.

En el Capítulo 4, se aborda un estudio comparativo sobre el comportamiento de los estimadores propuestos en los capítulos anteriores. Analizando las propiedades asintóticas de los dos estimadores de regresión local lineal (primeras diferencias y efectos fijos) se aprecia que ambos mantienen el mismo orden de magnitud del término de sesgo, pero muestran límites asintóticos distintos para el término de la varianza. En ambos casos las consecuencias son la obtención de tasas de convergencia no paramétricas subóptimas. Como se demuestra en los Capítulos 2 y 3, para solventar este problema es necesario explotar la estructura aditiva del modelo de regresión en diferencias y algoritmos de backfitting de una etapa son propuestos. Bajo condiciones bastante generales, se obtiene que los dos estimadores de backfitting son asintóticamente equivalentes, por lo que un análisis sobre su comportamiento en tamaños muestrales pequeños es muy interesante.

En un contexto totalmente paramétrico, está demostrado que bajo supuestos de exogeneidad estricta el comportamiento de los estimadores en diferencias depende de la estructura estocástica del término de error idiosincrático; ver Wooldridge (2003). Sin embargo, en el ajuste no paramétrico, además de la estructura estocástica existen otros factores como la dimensionalidad y el tamaño muestral que son de gran interés. En concreto, en un estudio comparativo entre estos estimadores analizamos cómo se comporta su error cuadrático medio en promedio (AMSE) bajo diversos escenarios. Destacar que los resultados de simulación obtenidos del AMSE confirman básicamente los resultados teóricos. Asimismo, encontramos que los estimadores de efectos fijos son bastante sensibles al tamaño del número de observaciones temporales por individuo.

En el Capítulo 5 y con el objetivo de demostrar la gran utilidad de los métodos propuestos para el análisis empírico, se considera la estimación no paramétrica de un modelo estructural sobre los ahorros preventivos de los hogares motivado por el modelo del ciclo vital de Modigliani and Brumberg (1954). A partir de esta especificación, estimamos un modelo en el cual los ahorros de los hogares están relacionados tanto con la incertidumbre sobre gastos médicos imprevistos como con la aversión al riesgo de los hogares.

En las pasadas dos décadas, existe una gran cantidad de estudios empíricos que tratan de mejorar nuestro entendimiento sobre el consumo óptimo de los hogares y su comportamiento bajo distintas fuentes de incertidumbre (seguros por desempleo o programas sanitarios públicos) pero sin llegar a resultados concluyentes; ver Starr-McCluer (1996), Gruber (1997), Egen and Gruber (2001), Gertler and Gruber (2002) o Gourinchas and Parker (2002), entre otros. Sin embargo, la mayor parte de estos estudios sufren de escasez de robustez contra distintos tipos de errores de especificación. En este contexto, el objetivo de este quinto capítulo es contribuir a la literatura sobre los ahorros preventivos extendiendo el modelo semi-paramétrico propuesto en Chou et al. (2004) al análisis de modelos de datos de panel. De este modo, lo que nos proponemos es estimar un modelo de ciclo vital que nos permita determinar el comportamiento de los hogares haciendo frente, de manera simultánea, a los errores de especificación más habituales en este tipo de modelos: (i) heterogeneidad de sección cruzada no observada correlacionada con las covariables del modelo; (ii) ciertas funciones que relacionan variables endógenas y exógenas en la ecuación de Euler son desconocidas y deben ser estimadas; y (iii) existencia de covariables endógenas.

En este sentido, el estimador de las funciones de interés que se propone en este capítulo resuelve el problema de endogeneidad utilizando los valores predichos de las variables endógenas y que han sido generados en la estimación no paramétrica de la ecuación en forma reducida. De este modo, el estimador que se presenta tiene la forma de un estimador simple de mínimos cuadrados de dos etapas localmente ponderado. Además, ciertas técnicas de integración marginal son necesarias para estimar un subconjunto de las funciones de interés. Para determinar el comportamiento de estos estimadores obtenemos sus principales propiedades asintóticas. El capítulo finaliza con un experimento de Monte Carlo que analiza el buen comportamiento de los estimadores propuestos en muestras finitas y con una aplicación empírica sobre los ahorros de los hogares españoles bajo distintas fuentes de incertidumbre.

Finalmente, concluimos esta tesis doctoral resaltando las principales conclusiones que se desprenden de estos cinco capítulos y se apuntan posibles líneas futuras de investigación. Las pruebas de los principales resultados obtenidos a lo largo de este trabajo así como los programas de estimación desarrollados para estos estimadores son relegados al apéndice.

Introduction

In the past five decades, the complexity of econometric models has been greatly enriched by the availability of panel data. Since these data sets are characterized by the observation of a group of individuals (households, consumers, countries, etc.) over time, they allow us to extract some unknown information about the idiosyncratic characteristics of individuals. In this way, the theoretical ability of these data to isolate the impact of unobserved actions of individuals enables us to make consistent inferences on a variety of topics that are not possible with other types of data, such as cross-sectional or time series data.

Traditionally, the econometric specification of such models has focused on the analysis of the relationship between an endogenous variable and some explanatory variables taking into account the unobserved individual heterogeneity. This analysis has been based on rather restrictive assumptions about functional forms and densities. However, as it is noted in Wooldridge (2003), sometimes these assumptions are quite unrealistic, and there are situations where the risk of misspecification is high. If this is the case, standard estimators based on moment conditions are biased and their use can invalidate the inference results.

In this context, nonparametric regression techniques have become a very useful tool to address these problems. Fully nonparametric panel data models are very appealing since they do not make any assumption about the model specification, but they allow data set to draw the shape of the regression function by themselves. However, although such estimators are robust to misspecification of the regression function, they are subject to the curse of dimensionality. In other words, as the number of explanatory variables increases, the rate of convergence of these estimators is dramatically reduced. In order to improve their rate it has been usually proposed to include a fully parametric component. These are the so-called semi-parametric models. In this case, what is known from previous empirical research or economic theory is modeled in a parametric way, whereas what is unknown for the researcher is specified nonparametrically.

However, these flexible models are unable to capture some features hidden in data sets. That is the main reason why many empirical studies have encouraged the introduction of more flexible structures that allow for variation of the unknown parameters according to some explanatory variables. In this situation, varying coefficient models originally proposed in Cleveland et al. (1991) appear as a solution. In particular, they are characterized by allowing for certain regression coefficients to vary with some exogenous variables suggested by economic theory. Therefore, they are able to exploit the information of the data set.

Note that in the past 15 years varying coefficient models have experienced an unparalleled growth, from both a methodological and theoretical point of view. As they encompass both nonparametric and semi-parametric models, they offer a quite general setting to handle many of the specification problems of such models. In what follows, we present some empirical applications of varying coefficient models in order to have a better understanding of the main advantages that they offer for empirical analysis.

In the literature on the returns to education we can find a first example of the improvement achieved with this new type of models. As it is noted in Schultz (2003), marginal returns to education vary with different levels of work experience. Thus, the omission of the nonlinearity of education as well as the interaction impact between education and work experience leads us to undersmoothing outcomes of the education performance, as it is found in Card (2001). In this situation, a semi-parametric varying coefficient model of the following form may be more advisable,

$$Y_i = m(Z_i)X_i + W_i^{\top}\beta + v_i, \quad i = 1, \cdots, N,$$

where Y_i is the hourly wages of individual i (log), X_i denotes the years-of-schooling of the individual as a measurement of his education experience, Z_i is a measure of working experience, and W_i denotes a vector of control variables that includes binary indicators for some features of the individuals such as marital status, government employed, union status and so on. In this way, this modeling overcomes the most common problems in this literature. It allows the impact of education on performance to vary with the level of work experience and to consider, at the same time, the nonlinearity shape of education.

Furthermore, the macroeconomic and international economic literature provides another relevant example. When we examine the role of foreign direct investment (FDI) in the economic growth of the countries, authors like Kottaridi and Stengos (2010) claim that, because the positive effect of the FDI on the economic growth only occurs in those countries with a higher level of income, the coefficient of FDI inflows in a growth model must be varying over the initial income level of each country. In this situation, the following semi-parametric varying coefficient model is specially appealing since it enables us to take into account both the nonlinear impact of FDI as well as the interactive effect between FDI and level of income,

$$Y_{it} = \alpha_0 + \alpha_1 D_j + \alpha_2 ln(I_{it}^d/Y) + \alpha_3 ln(n_{it}) + \alpha_4(lnX_{it})(I_{it}^f/Y) + \alpha_5 h_{it} + \epsilon_{it},$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where Y_{it} is the growth rate of income per capita in country i and period t, I_{it}^d the domestic investment rate to GDP, n_{it} the population growth rate, h_{it} the human capita, I_{it}^f/Y is the ratio of FDI to GDP and X_{it} is income per capita at the beginning of each period.

Finally, we find another relevant example when we try to establish the role of natural resources in the economic development of regions. As it is well-known in the literature on development economics, resource-abundant regions tend to grow slower than those with poor resources; see Sachs and Warner (2001). However, since the former constitute both growth losers and winners, the role of the quality of the institutions can be a decisive issue to identify the impact of the natural resources. Thus, in order to take into account the fact that the variation of growth performance among regions with abundant natural resources depends on how resource rents are distributed via the institutional arrangement, in Fan et al. (2010) it is proposed the following semi-parametric varying coefficient model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4(Z_i)X_{4i} + v_i, \quad i = 1, \cdots, N,$$

where Y_i is the average growth rate of country i, X_{1i}, X_{2i}, X_{3i} and X_{4i} denote initial income level, is a measure of the institutional quality approximated via the share of primary exports in GNP in 1970.

In this context, the objective of this doctoral dissertation is twofold. On one hand, to develop new estimation techniques that allow us to obtain consistent estimators for varying coefficient panel data models where the unobserved individual heterogeneity is correlated with some covariates. On the other hand, to emphasize the gains offered by this new methodology for empirical analysis. To this end, a nonparametric analysis of a structural model on the precautionary savings of the Spanish households to different sources of uncertainty is presented. In this way, this doctoral thesis is divided into five chapters and the structure of the dissertation is as follows.

In Chapter 1, an intensive review of the econometric literature on semi-parametric and fully nonparametric panel data models is performed. First, fully nonparametric panel data models with both random and fixed effects are analyzed. Second, we survey partially linear models under three different specifications: fixed and random effects, and presence of endogenous covariates. We conclude with a review of panel data varying coefficient models. For each of these areas, we discuss both the basic model to estimate and the proposed methodology. We also analyze the main asymptotic properties of the resulting estimators.

In Chapter 2, we present a new estimation technique for panel data models where the coefficients to estimate are smooth functions of other explanatory variables established by economic theory, and where the individual effects are arbitrarily correlated with the regressors of the model in an unknown way. As it can be shown, direct estimation through nonparametric techniques renders inconsistent estimators of the parameters (functions) of interest. In order to circumvent this problem, we use a transformation in first differences. Thus, following the original idea in Yang (2002) for an entirely different context, the proposed estimator is based on a local linear regression of a model in first differences. As it is proved in Lee and Mukherjee (2008), since the transformed regression equation is originally located around a fixed value of the sample, without considering the other values, direct application of local linear regression techniques to differencing transformations of panel data models generates a bias that does not disappear asymptotically. To avoid this problem, we propose to use a higher-dimensional kernel weight.

Unfortunately, this technique allows us to eliminate this bias, but at the price of increasing the variance term so the resulting estimators achieve a slower nonparametric rate of convergence. We propose to overcome this problem via a one-step backfitting algorithm that provides estimators that achieve the optimal rate of convergence of such problems. Also, the resulting estimators exhibit the oracle efficiency property. In other words, this procedure provides a nonparametric estimator whose variance-covariance matrix of all its components is asymptotically the same as if the other components of the transformed regression equation were known. In this chapter, we also obtain the asymptotic distribution of both estimators. Likewise, since the bandwidth matrix plays a crucial role in the consistent estimation of unknown functions, we also provide a method to calculate this matrix empirically. Finally, through a Monte Carlo experiment we test the good behavior of the proposed estimator in finite samples.

In Chapter 3, we present an estimator based on a deviation from the mean transformation of the regression model. As in the fully parametric setting, the main advantage of this transformation compared to other similar is that we obtain the standard asymptotic properties of nonparametric estimators under the assumption that the idiosyncratic error terms are *i.i.d.* In this way, the fixed effects estimator that we present in this chapter is obtained via a local approximation of the T additive functions of the deviation from the mean transformation, where T is the number of observations for each individual. Since the direct use of standard local approximation techniques in the estimation of these additive components omits relevant information as the sum of the distances between a fixed term and the other values of the sample, it provides nonparametric estimators with a bias term even more relevant than the obtained in Chapter 2. Thus, to overcome this problem we propose to consider a local approximation around the entire vector of temporal observations of each individual.

Nevertheless, the well-known trade-off between bias and variance reappears. Although the use of a kernel function of dimension T solves the bias problem, the variance term is considerably enlarged so the resulting estimator shows a slower rate of convergence than the first differences transformation estimator. In order to solve this situation and obtain estimators that achieve the optimal rate of convergence of this type of nonparametric problems, we use a one-step backfitting algorithm. Analyzing the asymptotic distribution of this two-stage estimator we appreciate that additional smoothing, introduced through the backfitting procedure, fails to reduce bias but decreases the variance term, as it is established in Fan and Zhang (1999). Therefore, this backfitting procedure allows to the estimators of the unknown functions of interest to achieve optimal nonparametric rates of convergence. Also, we show that the proposed fixed effects estimator is oracle efficient. Finally, through a Monte Carlo experiment we try to corroborate the theoretical findings obtained for the fixed effects estimator.

In Chapter 4, a comparative study on the behavior of the estimators proposed in the previous chapters is addressed. Analyzing the asymptotic properties of the two local linear regression estimators (first-differences and fixed effects) we observe that both maintain the same order of magnitude of the bias term, but they show different asymptotic limits for the variance term. In both cases the consequences are suboptimal nonparametric rates of convergence. As it is shown in Chapters 2 and 3, to solve this problem it is necessary to exploit the additive structure of the regression model in differences and one-step backfitting algorithms are proposed. Under fairly general conditions we find that the two backfitting estimators are asymptotically equivalent, so the analysis of their behavior in small sample sizes is very interesting.

In a fully parametric context, it is shown that under strict exogeneity assumptions the behavior of differencing estimators depends on the stochastic structure of the idiosyncratic error terms; see Wooldridge (2003). However, in the nonparametric setting, apart from the stochastic structure, there exist other factors such as the dimensionality and sample size that are of great interest. Specifically, in a comparative study between these two estimators we analyze how their average mean squared error (AMSE) behave under fairly different scenarios. Note that the simulation results obtained for the AMSE essentially confirm the theoretical outcomes. Also, we have also found out that fixed effects estimators are rather sensitive to the number of time series observations per individual.

In Chapter 5 and with the aim of showing the feasibility and possible gains of the proposed methods for empirical analysis, we consider the estimation of a structural model of household's precautionary savings motivated by the life-cycle hypothesis model of Modigliani and Brumberg (1954). Starting from this specification, we estimate a model where household's savings are related to both uncertainty about unforeseen medical expenses and household risk aversion.

In the past two decades, there is a plethora of empirical studies that seek to improve our understanding about household's optimal consumption and its behavior under various sources of uncertainty (i.e., unemployment insurances or public health programs) but without reaching conclusive results; see Starr-McCluer (1996), Gruber (1997), Egen and Gruber (2001), Gertler and Gruber (2002) or Gourinchas and Parker (2002), among others. However, most part of these studies suffers from lack of robustness against various types of misspecification problems. In this context, the aim of this fifth chapter is to contribute to the literature on precautionary savings by extending the semi-parametric model proposed in Chou et al. (2004) to the analysis of panel data models. Thus, we propose to estimate a life-cycle hypothesis model that allows us to determine the behavior of households while facing simultaneity the misspecification problems that are more common in such models: (i) unobserved cross-sectional heterogeneity correlated with the explanatory variables of the model; (ii) some functions that relate endogenous and exogenous variables in the Euler equation are unknown and must be estimated; and (iii) existence of some endogenous covariates.

In this sense, the estimator of the functions of interest that we propose in this chapter solves the endogeneity problem using the predicted values of the endogenous variables that are generated in the nonparametric estimation of the reduced form equation. Thus, the estimator presented here has the simple form of a two-step weighted locally constant least-squares estimator. Also, certain marginal integration techniques are necessary to estimate a subset of the functionals of interest. To determine the behavior of these estimators, we analyze their main asymptotic properties. The chapter concludes with a Monte Carlo experiment that analyzes the good performance of the proposed estimators in finite samples and with an empirical application about savings of Spanish households under different sources of uncertainty.

Finally, we conclude this dissertation by highlighting the main results extracted from these five chapters and possible future research. The proofs of the main results obtained along this work and the computational programs developed for these estimators are relegated to the appendix.

Chapter 1

Nonparametric and semi-parametric panel data models: recent developments

1.1 Introduction

When we estimate econometric models, a relevant issue is the type of data that we use in the specification/estimation of the model because good statistical properties of estimation results depend largely on the information provided by the data set. In empirical research, longitudinal or panel data sets are specially attractive since, for each individual in the sample, they contain many observations over time. Therefore, these data sets allow us to access certain unobserved information about individuals' behavior, which cannot be captured with other data type, such as conventional cross-section or time series data.

As it is noted in Hsiao (2003), the main advantages relating to longitudinal data set can be classified into two groups. On one hand, the efficiency of the econometric estimators is improved because this type of data structure usually exhibits large sample sizes. Therefore, the degrees of freedom are increased while, at the same time, the collinearity between explanatory variables is usually reduced. On the other hand, its theoretical ability to isolate the effects of specific actions or treatments allows us to make suitable inferences to analyze a variety of economic questions that would not be possible with other data set. Nevertheless, panel data models also present some disadvantages that are mainly related to the presence of unobserved heterogeneity in the form of individual or temporal effects. As it is well-known, in this panel data context a key issue to provide consistent estimators of the parameters (functions) of interest is the role played by the unobserved effects. Ignoring this information may lead us to biased estimators by the omission of relevant variables. Thus, in many empirical applications it is common to consider the cross-sectional heterogeneity as a random variable distributed independently of the regressors (the so-called random effects). However, sometimes this assumption is too strong, specially when correlation between the cross-sectional heterogeneity and the regressors is allowed. In this case of fixed effects, a random effects estimator is inconsistent and we must resort to specific panel data techniques to deal with this statistical dependence. We refer to Arellano (2003), Baltagi (2013) or Hsiao (2003) for an intensive review about parametric panel data models and Maddala (1987) to obtain good arguments about random versus fixed effects.

Nevertheless, the suitable treatment of these unobserved effects is not enough to guarantee the consistency of the estimators because the estimation of the parameters of interest might also depend on some statistical restrictions imposed on the data generated process as well as on the relative values of the number of individuals N and the number of time periods T. Sometimes these assumptions are too restrictive with respect to functional forms and densities and the risk of misspecification is high.

In this context, nonparametric panel data models are very appealing as they do not make any assumption on the specification of the model and allow data to tailor the shape of the regression function by themselves. However, sometimes this flexibility presents some drawbacks. On one hand, it is unable to incorporate prior information properly to the nonparametric modeling and the resulting estimator of the unknown function tends to have a higher variance term. On the other hand, its worst disadvantage is the so-called curse of dimensionality that practically disables standard nonparametric methods when the dimension of the nonparametric covariates is high. In order to solve these drawbacks, semi-parametric panel data models that embody both general nonparametric specification and fully parametric, have been proposed. In contrast to the slower rate of convergence of the nonparametric estimators, semi-parametric models enable us to obtain $\sqrt{N}$-consistent estimators of the unknown parametric components. For early discussions on semi-parametric panel data models see Ullah and Roy (1998), while we refer to Ai and Li (2008) for a review about partially linear and limited dependent nonparametric and semi-parametric panel data models.

In this chapter, we provide an intensive review of the econometric literature about semi-parametric and fully nonparametric panel data models when N tends to infinity and T is fixed. Note that in Su and Ullah (2011) a similar modeling is analyzed, although in this case we include the most recent results and pay special attention to the treatment of models with fixed effects as well as with endogenous explanatory variables.

Throughout the chapter, we denote individuals as i and time as t. Also, we use a standard data sampling scheme in nonparametric panel data regression analysis such as the strict stationarity assumption of the variables involved in the model, i.e., let $(X_{it}, Z_{it}, Y_{it})_{i=1,\cdots,N;t=1,\cdots,T}$ be a set of independent and identically distributed (*i.i.d.*) $\mathbb{R}^{d+q+1}$-random variables in the subscript i for each fixed t and strictly stationary over t for fixed i. Y_{it} is a scalar response variable and (X_{it}, Z_{it}) are explanatory random variables. Also, $\otimes$ and $\odot$ denote the Kronecker and Hadamard product, respectively. In the following, K is a kernel function that we assume it is bounded. Furthermore, we assume that $\int uu^\top K(u)du = \mu_2(K)I_q$ and $\int K^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars.

The rest of the chapter is organized as follows. In Section 1.2 we analyze the literature on fully nonparametric panel data models. In Section 1.3 we focus on partially linear models. Finally, in Section 1.4 we review panel data varying coefficient models. In all sections we distinguish among random effects, fixed effects and we also consider the case of endogenous covariates.

1.2 Nonparametric panel data models

1.2.1 Random effects

In nonparametric panel data models with random effects, the response variable is generated through the following statistical model

$$Y_{it} = m\left(Z_{it}\right) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.1}$$

where Z_{it} is a $q \times 1$ vector of exogenous variables, v_{it} and μ_i are independent and identically distributed terms with zero mean and homoscedastic variances, $\sigma_v^2 < \infty$ and $\sigma_\mu^2 < \infty$, respectively. We assume μ_i and v_{jt} are uncorrelated with each other for all $i, j = 1, 2, \cdots, N$ and T. Note that the error term follows a one-way error component structure; see Balestra and Nerlove (1966) or Hsiao (2003). Therefore, this means that the disturbance term is formed by both the idiosyncratic error term v_{it} and the unobserved cross-sectional heterogeneity μ_i, so it has the following form

$$\epsilon_{it} = \mu_i + v_{it}. \tag{1.2}$$

In this situation, $E(\mu_i|Z_{it}) = 0$ and $E(v_{it}|Z_{it}) = 0$. The researcher is interested in the estimation of the unknown smooth function $m(z) = E(Y_{it}|Z_{it} = z)$.

Denote by $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{iT})^\top$ a $T \times 1$ vector and $V = E(\epsilon_i \epsilon_i^\top)$ a $T \times T$ matrix that takes the form

$$V = \sigma_v^2 I_T + \sigma_\mu^2 \imath_T \imath_T^\top \tag{1.3}$$

and since the observations are independent along individuals, the variance-covariance matrix of the error term has the standard form

$$\Omega = E(\epsilon\epsilon^\top) = I_N \otimes \left[\sigma_1^2 \imath_T \imath_T^\top / T + \sigma_v^2 (I_T - \imath_T \imath_T^\top / T)\right] = I_N \otimes V, \tag{1.4}$$

where ϵ is a $NT \times 1$ vector that contains the ϵ_i's vectors, $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$ and $V = \sigma_1^2 \imath_T \imath_T^\top / T + \sigma_v^2 \left(I_T - \imath_T \imath_T^\top / T\right)$.

The function of interest in (1.1) and its derivatives were at first estimated through a pooled local linear least-squares procedure. However, the resulting estimator should be inefficient given that, by the presence of μ_i in each time period, the composed error term is serially correlated. Therefore, efficient inference using these pooled estimators should require considering the information contained in the variance-covariance matrix. In order to solve this situation, several have been the developments when $N \to \infty$ and T is fixed. Ullah and Roy (1998), Lin and Carroll (2000) or Su and Ullah (2007), among others, consider the estimation of this nonparametric model via a local polynomial regression approach. Later, and with the aim of improving the efficiency of these estimators, Ruckstuhl et al. (2000) and Henderson and Ullah (2005) propose different strategies to incorporate the information contained in the disturbance terms and to improve the efficiency of the previous nonparametric estimators.

For any $z \in \mathcal{A}$, where $\mathcal{A}$ is a compact subset in a nonempty interior of $I\!R$, the basic idea behind the standard nonparametric estimation of $m(z) = E(Y_{it}|Z_{it} = z)$ is to obtain a smoothed average of the Y_{it} values taking into account the values of Z_{it} contained in a small interval of z such as $Z_{it} - z = O_p(h)$, where h is a bandwidth that tends to zero when $N \to \infty$.

In order to understand later developments, it is useful to begin by the analysis of the univariate least-squares procedure. In this simplest case where $q = 1$, this is equivalent to make a Taylor expansion of the unknown smooth function around z, i.e.,

$$m\left(Z_{it}\right) \approx m(z) + m'(z)\left(Z_{it} - z\right) + \frac{1}{2}m''(z)\left(Z_{it} - z\right)^2 + \cdots + \frac{1}{p!}m^{(p)}(z)\left(Z_{it} - z\right)^p \quad (1.5)$$

and the above exposition suggests that we can estimate $m(z), m'(z), \cdots, m^{(p)}(z)$ by regressing Y_{it} on the terms $(Z_{it} - z)^\lambda$, for $\lambda = 0, 1, \cdots, p$, with kernel weights. Here, the quantities of interest can be estimated by minimizing the following criterion function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - \sum_{\lambda=0}^{p}\beta_\lambda (Z_{it} - z)^\lambda\right)^2 K_h(Z_{it} - z), \quad (1.6)$$

where we denote by $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_p)$ the vector of minimizers of (1.6). This expression suggests $\widehat{\beta}_0 = \widehat{m}_h(z)$, $\widehat{\beta}_1 = \widehat{m}'_h(z), \cdots, \widehat{\beta}_p = \widehat{m}_h^{(p)}(z)$. K is a weight function defined in such way that takes low values when Z_{it} is far away from z and high values when Z_{it} is close to z. Also, for each u it holds

$$\int K(u)du = 1 \qquad \text{and} \qquad K_h(u) = \frac{1}{h}K(u/h). \quad (1.7)$$

Now, if we set $p = 0$ we obtain the standard Nadaraya-Watson regression estimator (see Nadaraya (1964) and Watson (1964)), i.e.,

$$\widehat{m}_{NW}(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it}-z)Y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it}-z)}. \tag{1.8}$$

For the case of a panel data set under random effects models, this estimator was originally proposed in Ullah and Roy (1998). If we set $p = 1$ we obtain the local linear regression estimator; see Ruppert and Wand (1994), Fan and Gijbels (1995b) or Zhan-Qian (1996) for a detailed description of this technique. For applications to panel data models with random effects see Lin and Carroll (2000), Ruckstuhl et al. (2000) and Su and Ullah (2007), among others.

If we extend the criterion function (1.6), for $p = 1$, to the multivariate case, i.e., $Z \in I\!R^q$, we obtain the following criterion function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it}-\beta_0-\beta_1^\top(Z_{it}-z)\right)^2 K_h(Z_{it}-z), \tag{1.9}$$

where now we denote by $\widehat{\beta} = (\widehat{\beta}_0 \quad \widehat{\beta}_1)^\top$ a $(1+q)$-vector that minimizes (1.9). Then, let $D_m(z) = \partial m(z)/\partial z^\top$ be a $q \times 1$ vector of partial derivatives of the function $m(z)$ with respect to the elements of the q-vector z, the above exposition suggests $\widehat{m}(z;h) = \widehat{\beta}_0$ and $\widehat{D}_m(z;h) = \widehat{\beta}_1$ as estimators for $m(z)$ and $D_m(z)$, respectively.

Assuming $Z_z^\top K_z Z_z$ is nonsingular, the solution to (1.9) in matrix form can be written as

$$\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \left(Z_z^\top K_z Z_z\right)^{-1} Z_z^\top K_z Y, \qquad \text{where} \qquad Z_z = \begin{bmatrix} 1 & (Z_{11}-z)^\top \\ \vdots & \vdots \\ 1 & (Z_{NT}-z)^\top \end{bmatrix} \tag{1.10}$$

is a $NT \times (1+q)$ matrix, $K_z = diag\left(K_h(Z_{11}-z), \cdots, K_h(Z_{NT}-z)\right)$ a $NT \times NT$ diagonal matrix and $Y = (Y_{11}, \cdots, Y_{NT})^\top$ a $NT \times 1$ vector.

Then, the local linear least-squares (LLLS) estimator of $m(z)$ is

$$\widehat{m}_{LLLS}(z;h) = e_1^\top\left(Z_z^\top K_z Z_z\right)^{-1} Z_z^\top K_z Y, \tag{1.11}$$

where e_1 is a $(1+q)$ selection matrix having 1 in the first entry and all other entries 0.

Under some smoothness conditions on the regression function, some moment conditions on the errors; i.e., $E|v_{it}|^{2+\delta} < \infty$, and assuming $h \to 0$ as $N \to \infty$ such that $Nh \to \infty$, it is straightforward to extend the results in Ruckstuhl et al. (2000, Theorem 1, pp. 53) to the multivariate case. Thus, as N tends to infinity and T is fixed the conditional bias of these pooled estimators can be written as

$$\text{bias}_{NM}\left(\widehat{m}_{NW}(z;h)\right) = \frac{h^2}{2T}\mu_2(K)\left[tr\left(\mathcal{H}_m(z)\right) + 2D_m(z)D_f(z)f(z)^{-1}\right] \quad (1.12)$$

and

$$\text{bias}_{LLLS}\left(\widehat{m}_{LLLS}(z;h)\right) = \frac{h^2}{2T}\mu_2(K)tr\left(\mathcal{H}_m(z)\right), \quad (1.13)$$

respectively, where $\mathcal{H}_m(z)$ denotes the $q \times q$ Hessian matrix of $m(z)$ and $D_f(z)$ is the $1 \times q$ first-order derivative vector of the density function $f(z)$.

Also, the asymptotic distribution of these estimators is

$$\sqrt{Nh^q}\left(\widehat{m}_{NW}(z;h) - m(z)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\sigma_v^2 + \sigma_\mu^2)R(K)}{Tf(z)}\right) \quad (1.14)$$

and

$$\sqrt{Nh^q}\left(\widehat{m}_{LLLS}(z;h) - m(z)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\sigma_v^2 + \sigma_\mu^2)R(K)}{Tf(z)}\right), \quad (1.15)$$

where, let $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the observed covariates vector, we denote $\sigma_\mu^2 = Var(\mu_i|\mathbb{Z})$ and $\sigma_v^2 = Var(v_{it}|\mathbb{Z})$.

Based on these results, we can realize that although the asymptotic variance of both estimators is equal, the asymptotic bias is not. In particular, the bias of the local linear least-squares estimator only depends on the curvature of $m(\cdot)$ at z in a particular direction, measured through $\mathcal{H}_m(z)$, while the bias term of the Nadaraya-Watson estimator emerges mainly from both the curvature of $m(\cdot)$ and the term $D_m(z)D_f(z)f(z)^{-1}$. Also, as it is well-known, the local linear estimator is the best among all linear smoothers and has better performance near the boundary of the support of the density function; see Fan (1993) for more details. Therefore, the local linear least-squares estimator is usually preferred to the Nadaraya-Watson. The form of the asymptotic variance is also of great interest. As it can be observed, its structure is the same that under a pure *i.i.d.* setting without no correlation. This phenomenon is already pointed out in Ruckstuhl et al. (2000).

With the aim of obtaining estimators asymptotically more efficient than the previous ones, if it is possible, it would be necessary to incorporate the covariance structure of the regression model (1.1) in the form of the estimators. To this end, two different approaches have been proposed in the literature.

On one hand, in Henderson and Ullah (2005) an alternative class of estimators is proposed. The estimators are the result of the minimization of the following weighted criterion function

$$\left(Y - \beta^\top Z_z\right)^\top W_z \left(Y - \beta^\top Z_z\right) \tag{1.16}$$

with respect to β, where $\widehat{\beta}$ and Z_z are defined as in (1.10). Let W_z be a weighting matrix based on the kernel function that contains the information of the error structure, in Henderson and Ullah (2005) the following local linear weighted least-squares (LLWLS) estimator of $m(z)$ is proposed,

$$\widehat{m}_{LLWLS}(z;h) = \widehat{\beta}_0 = e_1^\top \left(Z_z^\top W_z Z_z\right)^{-1} Z_z^\top W_z Y. \tag{1.17}$$

In this way, the first-step of this procedure must be the proposal of a specific form for W_z. In particular, in Lin and Carroll (2000) two types of weighting matrices are used, $W_z = K_z^{1/2}\Omega^{-1}K_z^{1/2}$ and $W_z = \Omega^{-1}K_z$, whereas in Ullah and Roy (1998) it is developed an estimation procedure with $W_z = \Omega^{-1/2}K_z\Omega^{-1/2}$. In addition, and as the reader can see in (1.4), if Ω is a diagonal matrix these different specifications of W_z are the same.

On the other hand, in Ruckstuhl et al. (2000) it is proposed to multiply both sides of (1.1) by the square-root of Ω^{-1} obtaining

$$Y^* = m(Z) + \Omega^{1/2}\epsilon, \tag{1.18}$$

where $Y^* = \Omega^{-1/2}Y + (I - \Omega^{-1/2})m(Z)$ is the transformed variable and $\Omega^{-1/2}\epsilon$ is a term that satisfies the independence condition because it has an identity variance-covariance matrix. Note that $Y^* = (Y_{11}^*, \cdots, Y_{NT}^*)^\top$, $m(Z) = (m(Z_{11}), \cdots, m(Z_{NT}))^\top$ and $\epsilon = (\epsilon_{11}, \cdots, \epsilon_{NT})^\top$ are $NT \times 1$ vectors. Therefore, by minimizing the criterion function related to (1.18) the resulting local linear least-squares estimator is

$$\widehat{m}(z) = e_1^\top \left(Z_z^\top K_z Z_z\right)^{-1} Z_z^\top K_z Y^*, \tag{1.19}$$

where Z_z and K_z are defined as in the expression (1.10).

Note that the estimator defined in either (1.17) or (1.19) is an unfeasible estimator because Ω depends on some unknown terms, i.e., σ_v^2 and σ_μ^2. In this line, and based on the spectral decomposition of Ω, in Henderson and Ullah (2005) it is proposed a local linear feasible weighted least-squares estimator where they suggest to replace the unknown covariance components by their consistent estimators.

Let $\widehat{\epsilon}_{it} = Y_{it} - \widehat{m}(Z_{it})$ be the local linear least-squares residuals related to (1.17), in Henderson and Ullah (2005) it is proposed to estimate the unknown terms of the variance-covariance matrix (1.4) such as

$$\widehat{\sigma}_1^2 = \frac{T}{N}\sum_{i=1}^{N}\widehat{\overline{\epsilon}}_{i\cdot}^2 \qquad \text{and} \qquad \widehat{\sigma}_v^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{\epsilon}_{it} - \widehat{\overline{\epsilon}}_{i\cdot})^2, \tag{1.20}$$

where $\overline{\epsilon}_{i\cdot} = T^{-1}\sum_{t=1}^{T}\epsilon_{it}$.

Plugging these consistent estimators in (1.4) they obtain $\widehat{\Omega}$. If we denote by $\widehat{W}_z$ any of the expression of W_z, where Ω is replaced by $\widehat{\Omega}$, the feasible local linear weighted least-squares (FLLWLS) estimator is

$$\widehat{m}_{FLWLS}(z;h) = e_1^\top \left(Z_z^\top \widehat{W}_z Z_z\right)^{-1} Z_z^\top \widehat{W}_z Y. \tag{1.21}$$

On the other hand, in Ruckstuhl et al. (2000) a two-step procedure is presented. In the first-step, a local linear least-squares estimator of the unknown functions of (1.18) is obtained and the residual term that enables us to obtain $\widehat{\Omega}$ is computed. In the second-step, they use this result to calculate $\widehat{Y}^* = \widehat{\Omega}^{-1/2}Y + (I - \widehat{\Omega}^{-1/2})\widehat{m}(Z)$ and they propose to regress $\widehat{Y}^*$ against Z through a local polynomial regression obtaining the following local linear two-step least-squares (LL2SLS) estimator

$$\widehat{m}_{LL2SLS}(z) = e_1^\top \left(Z_z^\top K_z Z_z\right)^{-1} Z_z^\top K_z \widehat{Y}^*. \tag{1.22}$$

Finally, based on this latter transformation in Martins-Filho and Yao (2009) a two-step procedure is also proposed to provide a local linear estimator in a regression model where the error term has a non-spherical covariance structure and the regressors are dependent and heterogeneously distributed.

In Henderson and Ullah (2005) it is shown that under some standard regularity conditions, for N large and T fixed, the asymptotic bias and variance rate of $\widehat{m}_{LLWLS}(z;h)$ are $O_p(h^2)$ and $O_p\left((NTh^q)^{-1}\right)$, respectively. Furthermore, in a Monte Carlo experiment conducted in Henderson and Ullah (2012) it can be seen that the estimator for Ω proposed in Henderson and Ullah (2005) performs better than the corresponding for the local linear least-squares (LLLS) estimator in Ullah and Roy (1998) or Lin and Carroll (2000). They also point out that the local linear two-step least-squares (LL2SLS) estimator exhibits a better performance in terms of mean squared error than any other type of these estimators.

1.2.2 Fixed effects

As we have just seen, the random effects approach includes μ_i in the random error term under the assumption that there is no correlation between μ_i and Z_{it}. However, if μ_i is allowed to be arbitrarily correlated with Z_{it} in the model, an estimation procedure with a one-way error component clearly provides biased estimators of the parameters (functions) of interest because $E(\mu_i|Z_{it} = z) \neq 0$.

For nonparametric panel data model with fixed effects the main focus is

$$Y_{it} = m\left(Z_{it}\right) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.23}$$

where Z_{it} is a $q \times 1$ vector of explanatory variables, $m(\cdot)$ is an unknown function for the researcher, v_{it} is $i.i.d\ (0, \sigma_v^2)$ and μ_i is $i.i.d\ (0, \sigma_\mu^2)$.

As in the parametric case, different estimation methods are developed to estimate nonparametric panel data models with fixed effects of the form of (1.23); see Hsiao (2003), Wooldridge (2003) or Baltagi (2013), for example. As we appreciate hereinafter, they may be classified into two broad approaches. On one hand, there is a first type of nonparametric estimators that use differencing transformations to remove the unobserved heterogeneity from the structural model. Thus, the unknown function of the transformed model can be estimated consistently through any nonparametric approach. On the other hand, a second type of estimators is based on the spirit of the least-squares dummy variable (LSDV) approach to propose estimators of the parameters of interest, i.e., $m(\cdot)$. In the following, the most recent literature on the nonparametric framework based on both approaches is reviewed. Later, we focus on the corresponding estimators for different specifications of these nonparametric models, i.e., allowing for additive structures of the smooth function or the presence of dynamic nonparametric covariates.

Differencing estimators

In this part, we focus on the proposals in Mundra (2005) and Lee and Mukherjee (2008) to show how standard differencing transformations enable us to propose estimators for the first-derivative function solving the statistical dependence problem between μ_i and Z_{it}. Later, we analyze the iterative nonparametric kernel estimator based on a profile likelihood approach developed in Henderson et al. (2008) to estimate the smooth function.

As in the fully parametric case, there are several transformations that enable us to remove the heterogeneity of unknown form. Among the most popular transformations we find out the so-called first differences and differences from the mean. First differences transformation may be understood as the subtract from time t of (1.23) that of time $t-1$, i.e.,

$$Y_{it} - Y_{i(t-1)} = m(Z_{it}) - m(Z_{i(t-1)}) + v_{it} - v_{i(t-1)}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T \quad (1.24)$$

or that of time 1, i.e.,

$$Y_{it} - Y_{i1} = m(Z_{it}) - m(Z_{i1}) + v_{it} - v_{i1}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \quad (1.25)$$

On its part, differences from the mean implies subtracting from time t the within-group mean, i.e.,

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} m(Z_{is}) + v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is}, \quad (1.26)$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$.

As the reader may appreciate, the right hand side of either (1.24), (1.25) or (1.26) are linear combinations of $m(Z_{it})$ for different periods t. When estimating the unknown function $m(\cdot)$ it is necessary to take into account that we have an additive function for each t whose elements share the same functional form. Therefore, direct nonparametric estimation of this function is not as straightforward as it is noted in Su and Ullah (2011).

Assuming $m(\cdot)$ is smooth enough, in Ullah and Roy (1998) it is proposed to use either a first differences or a mean deviation transformation after linearly approximate the unknown function $m(\cdot)$ around z. They expected that the resulting first-difference and fixed effects estimators of the marginal effects of $m(\cdot)$ (i.e., the partial derivatives of $m(z)$ with respect to z) satisfy the standard properties of the local linear regression approach. However, this statement is not true as it is proved in Lee and Mukherjee (2008) because this technique provides estimators with a bias term that does not disappear even in large samples.

To analyze this problem in detail, we consider the univariate problem ($q = 1$) of the first differences regression model (1.24). Then, approximate $m(\cdot)$ by means of a Taylor expansion implies to obtain

$$\Delta Y_{it} = \Delta Z_{it} m'(z) + \Delta v_{it}(z), \tag{1.27}$$

where $m'(z) = \partial m(z)/\partial z$. Let $m''(z) = \partial^2 m(z)/\partial z^2$ be the Hessian matrix of $m(z)$, the error term of this transformed regression is

$$\Delta v_{it}(z) = \Delta v_{it} + \frac{1}{2} m''(\xi) \left((Z_{it} - z)^2 - (Z_{i(t-1)} - z)^2 \right),$$

for some $\xi \in I\!R$ between Z_{it} and z.

On the contrary, the transformed regression of the mean deviation (i.e., within-group) expression is

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = \left(Z_{it} - \frac{1}{T}\sum_{s=1}^{T} Z_{is} \right) m'(z) + \left(v_{it}(z) - \frac{1}{T}\sum_{s=1}^{T} v_{is}(z) \right), \tag{1.28}$$

where the corresponding error term is of the form

$$v_{it}(z) - \frac{1}{T}\sum_{s=1}^{T} v_{is}(z) = \left(v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is} \right) + \frac{1}{2} m''(\xi) \left((Z_{it} - z)^2 - \frac{1}{T}\sum_{s=1}^{T} (Z_{is} - z)^2 \right).$$

For the transformed regression models (1.27) and (1.28), in Lee and Mukherjee (2008) the following local linear estimators of the first-order derivatives are proposed,

$$\widehat{m}'_D(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it} - z)\Delta Z_{it}\Delta Y_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it} - z)\Delta Z_{it}^2} \tag{1.29}$$

and

$$\widehat{m}'_W(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z)\ddot{Z}_{it}\ddot{Y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z)\ddot{Z}_{it}^2}, \tag{1.30}$$

where we denote $\ddot{Y}_{it} = Y_{it} - (T-1)^{-1}\sum_{s=1,s\neq t}^{T} Y_{is}$, and $\ddot{Z}_{it}$ and $\ddot{v}_{it}$ are defined in a similar way. For the sake of simplicity, in Lee and Mukherjee (2008) the leave-one-out average is used in (1.28) instead of the within-group mean when analyzing the asymptotic properties of these estimators.

Let $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the vector of observed covariates, under some standard smoothing conditions in Lee and Mukherjee (2008, Theorem 1, pp. 5) the following asymptotic properties of these two local linear estimators are shown when $N, T \to \infty$

$$E[\widehat{m}'_D(z;h) - m(z)|\mathbb{Z}] = \frac{m''(z)\mu_3(z)}{2\mu_2(z)} + O_p(h^2) \tag{1.31}$$

and

$$E[\widehat{m}'_W(z;h) - m(z)|\mathbb{Z}] = \frac{m''(z)[(\mu_1(z)\mu_2(z) + \mu_3(z))}{2\left(\mu_1^2(z) + \mu_2(z)\right)} + O_p(h^2), \tag{1.32}$$

where $\mu_j(z) = E(Z_{it} - z)^j < \infty$, for $j = 1, 2, 3$.

By analyzing these results we can highlight that it is proved that these two local linear estimators are inconsistent for any sample size because in both specifications, when $N \to \infty$ and $h \to 0$, the bias that does not go away. In particular, as it can be seen in (1.29) and (1.30), this non-degenerated bias is due to the fact that the transformed regression equations are localized around Z_{it}, without taking into account all other values. Consequently, since the distance between Z_{is} and z cannot be controlled by a fixed bandwidth parameter h, the residual terms of the Taylor approximation do not vanish. Therefore, it is not possible to assume that $\Delta v_{it}(z)$ and Δv_{it} are close enough and we can conclude that the local linear regression approach provides inconsistent estimators by the correlation between the transformed error terms $\Delta v_{it}(z)$ and the transformed regressors ΔZ_{it}. The same can be said for $\ddot{v}_{it}(z)$ and $\ddot{Z}_{it}$.

To our knowledge, there are two type of strategies to overcome this problem. On one hand, in Mundra (2005) a direct procedure is developed based on the use of a higher-dimensional kernel weight. On the other hand, in Lee and Mukherjee (2008) is proposed the estimation of a local within transformation that uses a locally weighted average to remove the fixed effects. In the following, we detail the main peculiarities of both techniques.

As we have stated previously, one way of overcoming this problem of non-negligible asymptotic bias is to use a higher-dimensional kernel weight. As it is suggested in Mundra (2005), the bias associated to (1.29) may be removed by considering a local approximation around the pair $(Z_{it}, Z_{i(t-1)})$ obtaining the following first-difference local linear estimator

$$\widehat{m}'_{FLL}(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it}-z)K_h(Z_{i(t-1)})\Delta Z_{it}\Delta Y_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it}-z)K_h(Z_{i(t-1)}-z)\Delta Z_{it}^2}. \tag{1.33}$$

However, note that despite providing the asymptotic distribution of this estimator, Mundra (2005) does not pay special attention to the behavior of the transformed regression errors, and further research of their asymptotic properties may be necessary.

On the other hand, in Lee and Mukherjee (2008) it is followed a differencing strategy that uses the locally weighted average of Z_{it}, for a given z, to remove the unobserved individual heterogeneity and propose consistent estimators that take into account all the values of the regressors involved in the estimation. They denote

$$\widetilde{Z}_{i\cdot} = \sum_{s=1,s\neq t}^{T} W_{is}(z)Z_{is}, \tag{1.34}$$

where $W_{is}(z)$ is the local weight of the form

$$W_{is}(z) = \frac{K_h(Z_{is}-z)}{\sum_{r=1,r\neq t}^{T} K_h(Z_{ir}-z)}. \tag{1.35}$$

Also, $\widetilde{Y}_{i\cdot}(z)$ and $\widetilde{v}_{i\cdot}(z)$ are defined in a similar way as the locally weighted averages of Y_{it} and $v_{it}(z)$, respectively.

Note that $\sum_{s=1,s\neq t}^{T} W_{is}(z)\mu_i = \mu_i$ because it holds $W_{is}(z) \geq 0$ and $\sum_{s=1,s\neq t}^{T} W_{is}(z) = 1$ for any z. Therefore, if we subtract such local averages from (1.23) and denote $Y_{is}^* = Y_{is} - (T-1)^{-1}\sum_{s=1,s\neq t}^{T} Y_{is}$ and $Z_{is}^* = Z_{is} - (T-1)^{-1}\sum_{s=1,s\neq t}^{T} Z_{is}$, the functions of interest are estimated from the following locally weighted average problem

$$\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T} \left(Y_{is}^* - Z_{is}^*\beta\right)^2 K_h(Z_{is}-z). \tag{1.36}$$

Denote by $\widehat{\beta}$ the minimizer of (1.36), the above exposition suggests as estimator for $m'(\cdot)$,

$$\widehat{m}'_{LWA}(z;h) = \widehat{\beta} = \frac{\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T} K_h(Z_{is}-z)Z_{is}^*Y_{is}^*}{\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T} K_h(Z_{is}-z)Z_{is}^{*2}}. \tag{1.37}$$

Under some standard regularity conditions, in Lee and Mukherjee (2008, Theorem 2, pp. 7) it is shown that this local weighted linear estimator $\widehat{m}'_{LWA}(z;h)$ has the following asymptotic properties when $N, T \to \infty$,

$$E\left[\widehat{m}'_{LWA}(z;h) - m(z)|\mathbb{Z}\right] = \frac{h^2}{2}\left(\frac{m''(z)f'(z)}{f(z)}\right)\left(\frac{\kappa_4 - \kappa_2^2}{\kappa_2}\right) + o_p(h^2) \tag{1.38}$$

and

$$Var\left(\widehat{m}'_{LWA}(z;h) - m(z)|\mathbb{Z}\right) = \frac{1}{NTh^3}\left(\frac{\sigma_v^2}{f(z)}\right)\left(\frac{\varphi_2}{\kappa_2^2}\right) + o_p\left(\frac{1}{NTh^3}\right), \tag{1.39}$$

where $\kappa_j = \int z^j K(z)dz$, for $j = 2, 4$, and $\varphi_2 = \int z^2 K^2(z)dz$.

Looking at these results, we can emphasize that the asymptotic bias and conditional variance is of the same order as the standard results of the local polynomial regression. Furthermore, since $h \to 0$ and they assume $NTh^3 \to \infty$ when $N, T \to \infty$, both conditional bias and variance-covariance matrix are asymptotically negligible.

As the reader may appreciate, these procedures are very appealing as they provide consistent estimators of the local marginal effect of the smooth function in a framework of differencing models. However, they are unable to identify the function $m(\cdot)$. In this context, in Henderson et al. (2008) an iterative nonparametric kernel estimator is developed based on a profile likelihood approach to estimate $m(\cdot)$, when N is large and T is fixed.

Let us consider the following differencing model to estimate

$$Y_{it} - Y_{i1} = m(Z_{it}) - m(Z_{i1}) + v_{it} - v_{i1}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{1.40}$$

so the likelihood function $\mathcal{L}(\cdot)$ suggested for the individual i in Henderson et al. (2008) is

$$\mathcal{L}_i(\cdot) = \varphi(Y_i, m_i) = -\frac{1}{2}\left(\ddot{Y}_i - m_i + m_{i1}\imath\right)^\top \Sigma^{-1}\left(\ddot{Y}_i - m_i + m_{i1}\imath\right), \tag{1.41}$$

where $m_i = m(Z_{it})$, $m_{i1} = m(Z_{i1})$ and $\ddot{Y}_i = (\ddot{Y}_{i2}, \cdots, \ddot{Y}_{iT})$, for $\ddot{Y}_{it} = Y_{it} - Y_{i1}$.

The variance-covariance matrix of $\ddot{v}_{it} = v_{it} - v_{i1}$, Σ, and its inverse, Σ^{-1}, are respectively equal to

$$\Sigma = \sigma_v^2\left(I_{T-1} + \imath_{T-1}\imath_{T-1}^\top\right) \quad ; \quad \Sigma^{-1} = \sigma_v^{-2}\left(I_{T-1} - \imath_{T-1}\imath_{T-1}^\top/T\right), \tag{1.42}$$

where we denote by $\widehat{m}_{[\ell-1]}(z)$ the current estimator.

In Henderson et al. (2008) it is proposed to estimate $\widehat{m}_{[\ell]}(z)$ by computing $\widehat{\alpha}_0(z)$, where $(\widehat{\alpha}_0 \quad \widehat{\alpha}_1)$ solve the following equation

$$\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it}-z)G_{it}(z;h) \\ \times\mathcal{L}_{i,tm}\left(Y_i, \widehat{m}_{[\ell-1]}(Z_{i1}), \cdots, \widehat{\alpha}_0 + ((Z_{it}-z)/h)^\top\widehat{\alpha}_1, \cdots, \widehat{m}_{[\ell-1]}(Z_{iT})\right) = 0, \quad (1.43)$$

where $G_{it}(z;h) = \left[1 \quad ((Z_{it}-z)/h)^\top\right]^\top$. Note that $\widehat{\alpha}_1(z)$ provides the next step derivative estimator of $m(z)$.

As it is well-known in the nonparametric literature, because the model to estimate exhibits an additive structure with two functions that share the same functional form, an initial estimator of $m(\cdot)$ can be obtained by the standard backfitting method, originally proposed in Hastie and Tibshirani (1990). However, it is true that in order to impose the additive structure for estimates the series method can be more suitable. In this way, in Henderson et al. (2008) it is recommended to use an iterative method that only implies a least-squares estimation procedure, turning to a series method to obtain an initial estimator of $m(\cdot)$.

Similarly to other nonparametric estimators developed for differencing models, the iterative estimator proposed in Henderson et al. (2008) has the advantage of removing completely the unobserved individual heterogeneity. However, the rate of convergence of the proposed estimators is relatively slow since the asymptotic variance is $O_p\left((N|H|)^{-1}\right)$.

Profile least-squares estimators

When we want to estimate directly nonparametric fixed effects models such as (1.23) we must control by both observed and unobserved covariates. In this way, we need an estimation procedure that provides consistent estimators for the structural parameters of the explanatory variables, $m(\cdot)$, in the presence of "incidental" parameters, i.e., μ_i. Therefore, following the idea of the least-squares dummy variable approach a profile least-squares method may be proposed, where a dummy variable is used to represent each cross-sectional observation.

In this subsection, we first analyze the profile least-squares estimators proposed in Sun et al. (2009), Su and Ullah (2011), Gao and Li (2013) and Lin et al. (2014) under different identification conditions. Later, we focus on alternative feasible forms for the local linear approach following Li et al. (2013).

If we rewrite the model (1.23) in a matrix form we obtain

$$Y = m(Z) + D_0\mu_0 + v, \tag{1.44}$$

where $Y = (Y_{11}, \cdots, Y_{NT})^\top$, $m(Z) = (m(Z_{11}), \cdots, m(Z_{NT}))^\top$ and $v = (v_{11}, \cdots, v_{NT})^\top$ vectors of $NT \times 1$ dimension. Let $\mu_0 = (\mu_1, \cdots, \mu_N)^\top$ be a vector of $N \times 1$ dimension and $D_0 = (I_N \otimes \imath_T)$ a $NT \times N$ dummy matrix.

For any given value of μ_0 we estimate the unknown $m(z)$, let z be an interior point of the neighborhood of Z, by minimizing the standard optimization problem, i.e.,

$$\left(Y - \imath_{NT} m(z) - D_0\mu_0\right)^\top K_h(z)\left(Y - \imath_{NT} m(z) - D_0\mu_0\right), \tag{1.45}$$

where $K_h(z) = diag\left(K_h(Z_{11} - z), \cdots, K_h(Z_{NT} - z)\right)$ is a $NT \times NT$ diagonal matrix and $K_h(Z_{it} - z) = h^{-1}K\left((Z_{it} - z)/h\right)$.

Taking the first-order condition of the objective function in (1.45) with respect to $m(z)$ and rearranging terms we obtain

$$\widehat{m}(z; h) = \left(\imath_{NT}^\top K_h(z)\imath_{NT}\right)^{-1} \imath_{NT}^\top K_h(z)(Y - D_0\mu_0). \tag{1.46}$$

However, since μ_0 is not directly observable this local constant estimator is unfeasible. In order to solve it, the profile least-squares approach suggests to profile out μ_0 by choosing it such as

$$\widehat{\mu}_0 = \left(D_0^\top K_h(z) D_0\right)^{-1} D_0^\top K_h(z)\left(Y - \imath_{NT} m(z)\right). \tag{1.47}$$

Later, the unobserved individual effects can be removed multiplying the true model by $\left(D_0^\top K_h(z) D_0\right)^{-1} D_0^\top K_h(z)$ so the remaining quantities of interest may be estimated using the following minimization problem

$$\left(Y - \imath_{NT} m(z)\right)^\top W_0(z)\left(Y - \imath_{NT} m(z)\right), \tag{1.48}$$

where the weighting matrix now has the form $W_0 = M_0(z)K_h(z)M_0(z)$, for $M_0(z) = I_{NT} - D_0\left(D_0^\top K_h(z)D_0\right)^{-1} D_0^\top K_h(z)$, in such way that $M_0(z)D_0 = 0$. Then, taking the first-order condition with respect to $m(z)$ the local constant estimator is

$$\widehat{m}_{LCLS}(z;h) = \left(\imath_{NT}^\top W_0(z)\imath_{NT}\right)^{-1} \imath_{NT}^\top W_0(z)Y. \tag{1.49}$$

Since the local weighting matrix $W_0(z)$ is designed to remove any time invariant term in (1.23), i.e., $W_0(z)\imath_{NT} \equiv 0$, $\imath_{NT}^\top W_0(z)\imath_{NT}$ is a non-invertible matrix so this method is not feasible. In order to overcome this situation, it is necessary to use another matrix with removes the unobserved cross-sectional term either complete or asymptotically and that also enables us to select only those values of Z_{it} close to z. In this sense, in Lin et al. (2014) it is proposed to replace D_0 by another matrix D that enables $D^\top K_h(z)D$ to be a nonsingular matrix, following Sun et al. (2009). Therefore, in Lin et al. (2014) it is suggested a new weighting matrix $W_h(z)$ that removes asymptotically the unobserved fixed effects and satisfies

$$\left(\imath_{NT}^\top W_h(z)\imath_{NT}\right)^{-1} \imath_{NT}^\top W_h(z)D_0\mu_0 = N^{-1}\sum_{i=1}^{N}\sum_{i}\mu_i = O_p(N^{-1/2}), \quad \text{as} \quad N \to \infty.$$

Thus, let $\mu = (\mu_2, \cdots, \mu_N)^\top$ be a $(N-1)\times 1$ vector and denote by D a $NT\times(N-1)$ matrix of the form $D = (-\imath_{N-1} I_{N-1})^\top \otimes \imath_T$, in Su and Ullah (2006a), Sun et al. (2009), and Lin et al. (2014) it is proposed to replace $D_0\mu_0$ by $D\widehat{\mu} = (\widehat{\mu}_1, \widetilde{\mu}^\top)^\top$ with $\widehat{\mu}_1 = -\sum_{i=2}^{N}\widehat{\mu}_i$ and

$$\widetilde{\mu} = (\widehat{\mu}_2, \cdots, \widehat{\mu}_N)^\top = \left(D^\top K_h(z)D\right)^{-1} D^\top K_h(z)Y,$$

obtaining the following local constant estimator of $m(z)$

$$\widehat{m}_{LCS}(z;h) = \left(\imath_{NT}^\top K_h(z)\imath_{NT}\right)^{-1} \imath_{NT}^\top K_h(z)W_h(z)Y, \tag{1.50}$$

where $W_h(z) = I_{NT} - D\left(D^\top K_h(z)D\right)^{-1} D^\top K_h(z)$ is a $NT \times NT$ matrix such that $W_h(z)D \equiv 0$.

By extending this problem to the local linear regression technique, we can replace D_0 by D in the corresponding expression of the local linear estimator associated to (1.47) obtaining

$$\widehat{\mu}_{PLLLS} = \left(D^\top K_h(z)D\right)^{-1} D^\top K_h(z)(Y - Z_z\beta), \tag{1.51}$$

where Z_z and β are defined as in (1.10).

Then, if we multiply the model (1.23) by $\left(D^\top K_h(z)D\right)^{-1} D^\top K_h(z)$ the quantities of interest can be estimated by minimizing the following concentrated weighted least-squares regression problem

$$(Y - Z_z\beta)^\top W_h(z) (Y - Z_z\beta), \tag{1.52}$$

where now the weighting matrix has the form $W = M(z)^\top K_h(z)M(z)$, for $M(z) = I_{NT} - D\left(D^\top K_h(z)D\right)^{-1} D^\top K_h(z)$, in such way that $M(z)D \equiv 0$.

Let $\widehat{\beta} = (\widehat{\beta}_0^\top \quad \widehat{\beta}_1^\top)^\top$ be a $(1+q)$ vector of minimizers of (1.52), we can propose the following profile local weighted linear estimator

$$\widehat{\beta}_{PLLLS} = \left(Z_z^\top W_h(z) Z_z\right)^{-1} Z_z^\top W_h(z) Y. \tag{1.53}$$

Furthermore, a consistent estimator of the unobserved fixed effects can be obtained by replacing β by $\widehat{\beta}$ in (1.51) but, to our knowledge, no empirical work has been developed based on it.

Note that although this new weighting matrix provides feasible estimators, it fails to remove the fixed effects asymptotically. In particular, in Lin et al. (2014, Theorem 2.1) we can see that under some standard smoothing conditions the profile estimator of a local constant approximation has a compound bias term when $N \to \infty$ and $T \to \infty$: one of the terms is related to the nonparametric approximation of the smooth function, while the other comes from the existence of unobserved cross-sectional heterogeneity. To overcome it, a standard solution can be the estimation of the nonparametric regression under further strong identification conditions about the unobserved individual heterogeneity. In particular, Mammen et al. (2009) or Su and Ullah (2011) impose $\sum_{i=1}^{N} \mu_i = 0$, while Gao and Li (2013) develop a profile least-squares method under the condition $E(\mu_i) = 0$. As it is proved in Sun et al. (2009) for partially linear models, this stronger identification condition allows us to obtain standard asymptotic properties in the nonparametric framework and, simultaneously, to override the individual effects.

Alternatively, in Li et al. (2013) a profile least-squares procedure it is developed in which is not necessary to pay special attention to the invertibility problem remarked in Lin et al.

(2014) or Gao and Li (2013). Again, assuming $\sum_{i=1}^{N} \mu_i = 0$ a profile local linear least-squares method for the nonparametric components of the regression function is proposed. But, unlike the previous methods, in Li et al. (2013) it is assumed that μ is known and a least-squares procedure for the nonparametric components in β is proposed. In this way, the quantities of interest can be estimated by minimizing the resulting criterion function of a local linear fitting with respect to β obtaining as estimator

$$\widehat{\beta} = \left(Z_z^\top K_h(z) Z_z\right)^{-1} Z_z^\top K_h(z)(Y - D\mu). \tag{1.54}$$

As previously, this estimator is not feasible but we can multiply the true model by $e_1^\top \left(Z_z^\top K_h(z) Z_z\right)^{-1} Z_z^\top K_h(z)$ and choose the μ that minimizes the following criterion function

$$(Y^* - D^*\mu)^\top (Y^* - D^*\mu), \tag{1.55}$$

where we denote by

$$D^* \quad = \quad \left(I_{NT} - e_1^\top \left(Z_z^\top K_h(z) Z_z\right)^{-1} Z_z^\top K_h(z)\right) D,$$

and

$$Y^* \quad = \quad \left(I_{NT} - e_1^\top (Z_z^\top K_h(z) Z_z)^{-1} Z_z^\top K_h(z)\right) Y.$$

In this way, the minimizer of (1.55) is of the form

$$\widehat{\mu}_{PLLLS} = \left(D^{*\top} D^*\right)^{-1} D^{*\top} Y^* \tag{1.56}$$

and replacing μ by $\widehat{\mu}$ in (1.54) the profile local weighted linear least-squares estimator is

$$\widehat{\beta}_{PLLLS} = \left(Z_z^\top K_h(z) Z_z\right)^{-1} Z_z^\top K_h(z)(Y - D\widehat{\mu}). \tag{1.57}$$

Finally, under standard smoothing conditions in Li et al. (2013, Theorem 1, pp. 231) it is shown the following asymptotic normality of (1.57) as $N \to \infty$ and T is fixed,

$$\sqrt{Nh}\left(\widehat{m}_{PLLLS}(z;h) - m(z) - b(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\overline{\sigma}^2 R(K)}{f(z)^2}\right), \tag{1.58}$$

where $m''(z) = \partial^2 m(z)/\partial z^2$ is the Hessian matrix of $m(z)$ and

$$\begin{aligned} b(z) &= \frac{h^2}{2}\mu_2(K)m''(z), \\ \overline{\sigma}^2(z) &= \sum_{t=1}^{T}\sigma_t^2(z)f_t(z) = \sum_{t=1}^{T}Var\left(v_{it}^2|Z_{it}=z\right)f_t(z). \end{aligned}$$

As we detail in later sections, these asymptotic properties are similar to the result obtained in Su and Ullah (2006b) for partially linear panel data models with fixed effects.

Additive models

Sometimes there are situations in which nonparametric panel data models of moderate dimension are not suitable, and another modeling is necessary. For example, when we want to perform an economic analysis of the production functions, nonparametric additive models may be more interesting; see Sperlich et al. (2002) for a cross-sectional analysis of this type of models.

At this point, we may be tempted to extend directly the estimation techniques analyzed previously to nonparametric additive models. However, the panel data framework makes this task much more difficult; see Hjellvik and Tjøstheim (1999) or Wooldridge (2005) for more details. In this situation, in Mammen et al. (2009) it is considered the consistent estimation of nonparametric additive panel data models under different forms of the unobserved heterogeneity. They state the main asymptotic properties of the resulting nonparametric estimators when only unobserved temporal heterogeneity is allowed, but we analyze a panel data regression model where both unobserved temporal and individual effects are present, i.e.,

$$Y_{it} = \sum_{j=1}^{p} m_j(Z_{jit}) + \eta_t + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.59}$$

where $m_j(\cdot)$ are the unknown functions that the research has to estimate, while η_t and μ_i) represent the fixed effects.

As it is standard in nonparametric additive models, a possible solution to the estimation of $m_j(\cdot)$ is the use of the marginal integration or the backfitting technique. Because the marginal integration approach works well for low-dimensional covariates models, in Mammen et al. (2009) it is proposed an alternative backfitting procedure based on the smoothed backfitting approach developed in Mammen et al. (1999). For the correct performance of the local constant smooth backfitting estimators, $(\widehat{m}_1, \cdots, \widehat{m}_p)$, in Mammen et al. (2009) it is proposed to use only those covariates that fall into the interval $[0,1]^p$, ignoring the rest of explanatory variables for estimates.

In this way, the kernel function has to integrate to one over the interval $[0,1]$, i.e.,

$$K_h(u,v) = \begin{cases} \frac{K[h^{-1}(u-v)]}{\int_0^1 K[h^{-1}(\omega-v)]d\omega} & \text{if} \quad u,v \in [0,1] \\ 0 & \text{else,} \end{cases}$$

so we can highlight that when the dimension of the covariates q is large, this criterion function causes the loss of a relatively large portion of the observations. In order to minimize this effect, in Mammen et al. (2009) it is suggested to use an arbitrarily large but fixed compact set, whereas the use of data set that depends on the sample size and converge to the whole space can be asymptotically more suitable.

In this situation, the quantities of interest of (1.59) can be estimated using a smoothed least-squares criterion of the form

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\int\left(Y_{it} - \sum_{j=1}^{q}\widehat{m}_j(z_j) - \widehat{\eta}_t - \widehat{\mu}_i\right)^2 K_{h(1)}(z_1, Z_{1it})\cdots K_{h(q)}(z_q, Z_{qit})dz_1\cdots dz_p, \tag{1.60}$$

under the following constraints

$$\int \widehat{m}_j(z_j)\widehat{f}_j(z_j)dz_j = 0, \tag{1.61}$$

$$\sum_{i=1}^{N} N_i\mu_i = 0. \tag{1.62}$$

Denote by $(\widehat{m}_1, \cdots, \widehat{m}_p)$ the minimizers of the criterion function (1.60), the estimators of the arguments of interest can be written as

$$
\begin{aligned}
\widehat{m}_j(z_j) &= \widetilde{m}_j(z_j) - \sum_{t=1}^{T} \frac{N_t}{NT} \widehat{\eta}_t \frac{\widehat{f}_j^t(z_j)}{\widehat{f}_j(z_j)} - \sum_{i=1}^{N} \frac{N_i}{NT} \widehat{\mu}_i \frac{\widehat{f}_j^i(z_j)}{\widehat{f}_j(z_j)} \\
&- \sum_{\ell \neq j} \int \widehat{m}_\ell(z_\ell) \frac{\widehat{f}_{j\ell}(z_j, z_\ell)}{\widehat{f}_j(z_j)} dz_j, && (1.63) \\
\widehat{\eta}_t &= \widetilde{n}_t - \sum_{j=1}^{q} \int \widehat{m}_j(z_j) \widehat{f}_j^t(z_j) dz_j, \quad t = 1, \cdots, T && (1.64) \\
\widehat{\mu}_i &= \widetilde{\mu}_i - \sum_{j=1}^{q} \int \widehat{m}_j(z_j) \widehat{f}_j^i(z_j) dz_j, \quad i = 1, \cdots, N && (1.65)
\end{aligned}
$$

where $(\widetilde{m}_j, \widetilde{\eta}_j, \widetilde{\mu}_i)$ are the following marginal estimators

$$
\begin{aligned}
\widetilde{m}_j(z_j) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_h(z_j, Z_{jit}) Y_{it} / \widehat{f}_j(z_j), && (1.66) \\
\widetilde{\eta}_t &= \frac{1}{N_t} \sum_{i=1}^{N} \mathbf{1}(Z_{it} \in [0,1]^q) Y_{it}, && (1.67) \\
\widetilde{\mu}_i &= \frac{1}{N_i} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) Y_{it}, && (1.68)
\end{aligned}
$$

and the functions $(\widehat{f}_{jk}, \widehat{f}_j^t, \widehat{f}_j^i)$ are the estimators of the kernel density of the form

$$
\begin{aligned}
\widehat{f}_{jk}(z_j, z_k) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h(j)}(Z_{jit} - z_j) K_{h_{(k)}}(Z_{kit} - z_k), && (1.69) \\
\widehat{f}_j^t(z_j) &= \frac{1}{N_t} \sum_{i=1}^{N} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h(j)}(Z_{jit} - z_j), && (1.70) \\
\widehat{f}_j^i(z_j) &= \frac{1}{N_i} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h(j)}(Z_{jit} - z_j). && (1.71)
\end{aligned}
$$

As the reader can see in (1.63)-(1.65), to obtain $\widehat{m}_j(\cdot)$ is necessary to use an iterative procedure since $\widetilde{m}_j(z_j)$ has to be estimated previously. In particular, what it is proposed in Mammen et al. (2009) is to plug-in the current values of $\widehat{m}_\ell$ $(\ell \neq j)$, $\widehat{\eta}_t$ and $\widehat{\mu}_i$ into the right-hand side of (1.63), and later apply (1.64) and (1.65) for updates of $\widehat{\eta}_t$ and $\widehat{\mu}_i$, respectively. Thereafter, this strategy is again done by using the actual values of $\widehat{m}_j(\cdot)$ on the right-hand side of (1.63).

In this way, plugging (1.64) and (1.65) into the right-hand side of (1.63) the resulting estimator of the local constant backfitting smoothing can be written as

$$\widehat{m}(z) = m^*(z) + \int \widehat{H}(z;y)\widehat{m}(y)\widehat{f}(y)dy, \tag{1.72}$$

where

$$\begin{aligned}
m_j^*(z_j) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(Z_{it}\in[0,1]^q)K_h(z_j,Z_{jit})(Y_{it}-\overline{Y}_{\cdot t}-\overline{Y}_{i\cdot}+\overline{Y}),\\
\overline{Y}_{\cdot t} &= \frac{1}{N_t}\sum_{i=1}^{N}\mathbf{1}(Z_{it}\in[0,1]^q)Y_{it},\\
\overline{Y}_{i\cdot} &= \frac{1}{N_i}\sum_{t=1}^{T}\mathbf{1}(Z_{it}\in[0,1]^q)Y_{it},\\
\overline{Y} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(Z_{it}\in[0,1]^q)Y_{it}.
\end{aligned}$$

Also, $\widehat{H}(z,y)$ is a $q\times q$ matrix whose off-diagonal entries ($\ell\neq j$) are

$$\widehat{H}_{j\ell} = \frac{\widehat{f}_{j\ell}(z_j,y_\ell)}{\widehat{f}_j(z_j)} + \sum_{t=1}^{T}\frac{N^t\widehat{f}_j^t(y_\ell)\widehat{f}_j^t(z_j)}{NT\widehat{f}_j(z_j)} + \sum_{i=1}^{N}\frac{N^i\widehat{f}_\ell^i(y_\ell)\widehat{f}_j^i(z_j)}{NT\widehat{f}_j(z_j)} \tag{1.73}$$

and whose diagonal elements are

$$\widehat{H}_{jj}(z,y) = \sum_{t=1}^{T}\frac{N^t\widehat{f}_\ell^t(y_\ell)\widehat{f}_j^t(z_j)}{NT\widehat{f}_j(z_j)} + \sum_{i=1}^{N}\frac{N^i\widehat{f}_\ell^i(y_\ell)\widehat{f}_j^i(z_j)}{NT\widehat{f}_j(z_j)}. \tag{1.74}$$

By analyzing the asymptotic properties of $\widehat{m}_j(\cdot)$, Mammen et al. (2009) confirm that the smooth backfitting local constant estimator exhibits one of the main annoying features of the standard backfitting estimators, that is, $\widehat{m}_j(\cdot)$ has a complex bias expression because it depends on the form of other regression functions, $m_\ell(\cdot)$ for $\ell\neq j$. Note that this result complicates greatly the statistical inference based on it.

With the aim of overcoming this problem, these authors propose a smoothing process using a local linear approximation of the unknown function rather than a local constant. Let $\widehat{m}_1,\cdots,\widehat{m}_p$, $\widehat{m}'_1,\cdots,\widehat{m}'_p$, $\widehat{\mu}_1,\cdots,\widehat{\mu}_N$ and $\widehat{\eta}_1,\cdots,\widehat{\eta}_T$ be local linear estimators defined as the minimizers of

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\int\left(Y_{it}-\sum_{j=1}^{p}\widehat{m}_j(z_j)-\frac{Z_{jit}-z_j}{h(j)}\widehat{m}'_j(z_j)-\widehat{\eta}_t-\widehat{\mu}_i\right)^2 K_{h(1)}(z_1,Z_{1it})\cdots K_{h(p)}(z_p,Z_{pit}) \\ \times dz_1\cdots dz_p, \tag{1.75}$$

where now, instead of using (1.61) as identification restriction, they propose the following

$$\int \widehat{m}_j(z_j)\widehat{f}_j(z_j)dz_j + \int \widehat{m}_j(z_j)\widehat{f}_j'(z_j)dz_j = 0, \tag{1.76}$$

where $\widehat{f}_j$ is the estimator of the kernel density previously defined and

$$\widehat{f}_j'(z_j) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(Z_{it} \in [0,1]^q)K_{h(j)}(Z_{jit} - z_j)(Z_{it} - z_j).$$

Finally, in Mammen et al. (2009) the main asymptotic properties of the local linear backfitting estimator are obtained. They show that this asymptotic bias is very close to the standard one of the local linear estimator when only temporary fixed effects are included in the model of interest.

1.2.3 Endogeneity (dynamic models)

Until this moment, we have focused on several estimation strategies that allow us to obtain consistent estimators for nonparametric panel data models, but ignoring the situation in which the right-hand side of the regression presents lagged dependent variables.

When our aim is the estimation of a dynamic panel data model as the considered in Su and Lu (2013), the following model is analyzed

$$Y_{it} = m(Y_{i(t-1)}, X_{it}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.77}$$

where X_{it} is a $q \times 1$ vector of explanatory variables, $Y_{i(t-1)}$ a scalar lagged dependent variable, μ_i the cross-sectional heterogeneity and v_{it} the error term.

Using a first difference transformation to remove the fixed effects, they obtain

$$\Delta Y_{it} = m(Y_{i(t-1)}, Z_{it}) - m(Y_{i(t-2)}, Z_{i(t-1)}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{1.78}$$

where we may appreciate some particularities that must be taken into account in order to develop a suitable estimation procedure. On one hand, the error term Δv_{it} has the form of moving average process of order 1 (MA(1)) so, in general, it is correlated with the regressor $Y_{i(t-1)}$. Because of this endogeneity problem, conventional kernel estimation based on marginal integration or backfitting procedures does not provide consistent estimators for the smooth estimation. On the other hand, since both additive components share the same functional form and it is assumed $E(\Delta v_{it}|Y_{i(t-2)}, Z_{i(t-1)}) = 0$, to estimate $m(\cdot,\cdot)$ is necessary to solve a Friedholm integral equation of second type.

In this context, to obtain consistent estimators that address both problems, in Su and Lu (2013) an iterative estimator based on a local polynomial regression is developed. Let us denote by $U_{i(t-2)} = (Y_{i(t-2)} \quad Z_{i(t-1)}^{\top})^{\top}$ and assuming $U_{i(t-2)}$ has a positive density on φ, where φ denotes a compact set on $I\!R^{q+1}$. They obtain the following conditional moment condition by the law of iterated expectations, since Δv_{it} is (conditionally) mean-independent of $U_{i(t-2)}$,

$$E\left[\Delta Y_{it} - m(Y_{i(t-1)}, Z_{it}) + m(Y_{i(t-2)}, Z_{i(t-1)})|U_{i(t-2)}\right] = 0 \tag{1.79}$$

and rearranging terms

$$\begin{aligned} m(u) &= -E\left[\Delta Y_{it}|U_{i(t-2)} = u\right] + E\left[m(U_{i(t-1)})|U_{i(t-2)} = u\right] \\ &= r_{t|t-2}(u) + \int m(\overline{u}) f_{t-1|t-2}(\overline{u}|u) d\overline{u}, \qquad \text{for} \qquad t = 3, \cdots, T, \end{aligned} \tag{1.80}$$

where $r_{t|t-2}(u) = E\left[\Delta Y_{it}|U_{i(t-2)} = u\right]$, $f_{t-1|t-2}(\cdot|\cdot)$ is the conditional density function of $U_{i(t-1)}$ given $U_{i(t-2)}$ and $\overline{u}$ is the mean value of u.

By simplicity, let us denote by $\rho_{t-2} = P(U_{i(t-2)} \in \varphi)$ and $\rho = \sum_{t=3}^{T} \rho_{t-2}$, so if we multiply both sides of (1.80) by ρ_{t-2}/ρ and use the fact that $\sum_{t=3}^{T} \rho_{t-2}/\rho = 1$ we obtain

$$m(u) = r(u) + \int m(\overline{u}) f(\overline{u}|u) d\overline{u}. \tag{1.81}$$

Under certain regularity conditions, Su and Lu (2013) rewrite (1.81) as

$$m = r + \mathbb{A}m, \tag{1.82}$$

where $\mathbb{A}$ is a bounded linear operator defined such as $\mathbb{A}m(u) = \int m(\overline{u}) f(\overline{u}|u) d\overline{u}$.

Therefore, from (1.82) we may intuitively conclude that the estimator of the parameter of interest $m(\cdot)$ can be defined as a solution to the Fredholm integral equation of the second kind in an infinite dimensional Hilbert space. However, since both r and $\mathbb{A}m(u)$

are not directly observable, the resulting estimator of (1.82) is unfeasible and an iterative procedure is needed. In this situation, in Su and Lu (2013) it is proposed a plug-in estimator for $\widehat{m}(\cdot)$ as a solution of the following equation

$$\widehat{m} = \widehat{r} + \widehat{\mathbb{A}}\widehat{m}, \tag{1.83}$$

where $\widehat{r}$ and $\widehat{\mathbb{A}}$ are nonparametric estimators obtained from a local polynomial regression of pth order. In particular, $r(u)$ can be estimated using the following weighted least-squares problem

$$\sum_{i=1}^{N}\sum_{t=3}^{T}\left(-\Delta Y_{it} - \sum_{0\leq|j|\leq p}\beta_j^\top((U_{i(t-2)}-u)/h)^j\right)^2 K_h(U_{i(t-2)}-u)\mathbf{1}(U_{i(t-2)}\in\varphi), \tag{1.84}$$

where $\widehat{\beta} = (\widehat{\beta}_0^\top, \cdots, \widehat{\beta}_q^\top)^\top$ is a $(1+q)$-vector that minimizes (1.84) and $h = (h_0, \cdots, h_q)^\top$ is a bandwidth sequence.

Analogously, $\widehat{\mathbb{A}}m(u)$ is defined as the resulting estimator of $\mathbb{A}m(u)$ when $-\Delta Y_{it}$ is replaced by $m(U_{i(t-1)})$ in the problem to minimize (1.84). However, in order to obtain a feasible estimator of $\mathbb{A}$ we need to obtain an estimator for the function $m(\cdot)$. In this case, in Su and Lu (2013) it is proposed to resort to the method of sieves and, after obtaining $\widehat{m}(u)$, they replace it in the final regression to estimate. See Chen (2007) for an intensive revision of the method of sieves.

Let $h! = \prod_{\ell=0}^{d} h_\ell^2$ and $||h||^2 = \sum_{\ell=0}^{d} h_\ell^2$. Under certain standard smoothness conditions, and some assumptions on the behavior of bandwidths, i.e., as $N\to\infty$, T is fixed and $||h||\to 0$, $Nh!/logN\to\infty$ and $N||h||^4h!\to c\in[0,\infty]$, in Su and Lu (2013, Theorem 2.2, pp. 117) it is also established the asymptotic normality of the plug-in estimator. In particular, for the simplest case ($q=1$) they show

$$\sqrt{NTh!}\left(\widehat{m}(u;h) - m(u) - (I-\mathbb{A})^{-1}B(u)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, v(u)\right), \tag{1.85}$$

where $\mathbb{A}$ is a Hilbert-Schmidt operator such that

$$\begin{aligned} B(u) &= \frac{1}{2}\sum_{\ell=0}^{q} h_\ell^2\mu_2(K)\frac{\partial^2 m(u)}{\partial u_\ell^2}, \\ v(u) &= \frac{\sigma^2(u)}{f(u)}R^{q+1}(K), \\ \sigma^2(u) &= \sum_{t=3}^{T}\frac{\rho_{t-2}}{\rho}\sigma_{t-2}^2(u)f_{t-2}(u), \end{aligned}$$

and they denote $\sigma^2_{t-2}(u) = E(v^2_{it}|U_{i(t-2)} = u) + E(v^2_{i(t-1)}|U_{i(t-2)} = u)$.

According to this results, we can conclude the following: The asymptotic variance of this iterative estimator has a similar structure to that presented by a conventional local polynomial estimator of the unknown functions of nonparametric models such as $\Delta Y_{it} = m(U_{i(t-1)}) - m(U_{i(t-2)}) + \Delta v_{it}$, when $m(U_{i(t-2)})$ is observed; The asymptotic bias shows significant variations with reference to the standard results. Specifically, this iterative estimator present an additional operator, $(I - \mathbb{A})^{-1}$, which reflects the cumulative bias of the iterative process.

1.3 Partially linear models

As we have already realized, nonparametric panel data models are much more robust to specification errors than their parametric counterparts. Unfortunately, the most important price to be paid for the use of nonparametric techniques is the curse of dimensionality. One possible way to overcome this problem is to introduce some parametric components to the econometric model. These are the so-called semi-parametric models. After the seminal contribution in Robinson (1988), partially linear models have been quite frequently used in the econometric literature becoming the most popular semi-parametric modeling approach.

In what follows we analyze the statistical properties of partial linear models estimators under alternative assumptions. Mainly, random effects, fixed effects and endogenous covariates.

1.3.1 Random effects

In partially linear panel data models where the error term follows a one-way error component structure, the response variable follows

$$Y_{it} = X_{it}^\top \beta + m(Z_{it}) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.86}$$

where X_{it} and Z_{it} are vectors of exogenous variables of $d \times 1$ and $q \times 1$ dimension, respectively, β is a $d \times 1$ vector of unknown parameters and $m(\cdot)$ is an unknown smooth function.

We consider the standard panel data framework where N is large and T is finite, and the composed error term ϵ_{it} in the regression (1.86) exhibits a structure of the form

$$\epsilon_{it} = \mu_i + v_{it},$$

where μ_i is $i.i.d.(0, \sigma_\mu^2)$ and v_{it} is $i.i.d.(0, \sigma_v^2)$. Again, as in the fully nonparametric panel data models with random effects, we may define V as a $T \times T$ matrix of the same form as in (1.3). Also, since the observations are independent along individuals, the variance-covariance matrix of the composed error term Ω has the standard form as in (1.4).

Following the proposal in Robinson (1988); in Li and Stengos (1996) and Li and Ullah (1998) it is considered the estimation of a semi-parametric partially linear panel data model as the one established in equation (1.86). In order to estimate consistently the parameters of interest, in the previous references it is proposed to take the conditional expectation with respect to Z in the model of interest. Then, if we subtract it from (1.86), it is obtained

$$Y_{it} - E(Y_{it}|Z_{it}) = (X_{it} - E(X_{it}|Z_{it}))^\top \beta + \epsilon_{it}. \tag{1.87}$$

Let us denote by $\widehat{\beta}$ the ordinary least-squares (OLS) estimator of β from (1.87). In matrix notation it can be written as

$$\widehat{\beta}_{OLS} = \left(\widetilde{X}\widetilde{X}^\top\right)^{-1} \widetilde{X}^\top \widetilde{Y} = \beta + \left(\widetilde{X}\widetilde{X}^\top\right)^{-1} \widetilde{X}^\top \epsilon, \tag{1.88}$$

where $\widetilde{X}$ is a $NT \times d$ matrix whose typical row element is $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$, whereas $\widetilde{Y}$ is a $NT \times 1$ vector whose typical row element is of the form $\widetilde{Y}_{it} = Y_{it} - E(Y_{it}|Z_{it})$.

However, since $E(X_{it}|Z_{it})$ and $E(Y_{it}|Z_{it})$ are some unknown terms, (1.88) is an unfeasible estimator. To overcome this problem and to obtain $\sqrt{N}$-consistent estimators, a possible solution is to resort to nonparametric estimation techniques to provide the estimators of these conditional expectations. Specifically, in all previous references it is proposed to use a kernel estimation method to estimate these unknown conditional expectations. Thus, $E(Y_{it}|Z_{it})$ and the density function $f(Z_{it})$ may be estimated through

$$\widehat{Y}_{it} = \widehat{E}(Y_{it}|Z_{it}) = \frac{1}{NTh^q} \sum_{j=1}^{N} \sum_{s=1}^{T} Y_{js} K_h(Z_{it} - Z_{js}) / \widehat{f}_{it}, \tag{1.89}$$

$$\widehat{f}_{it} = \widehat{f}(Z_{it}) = \frac{1}{NTh^q} \sum_{j=1}^{N} \sum_{s=1}^{T} K_h(Z_{it} - Z_{js}), \tag{1.90}$$

where $K_h(Z_{it} - Z_{js}) = K((Z_{it} - Z_{js})/h)$ is the kernel function and h the bandwidth parameter. Note that $\widehat{X}_{it} = \widehat{E}(X_{it}|Z_{it})$ is defined in a similar way.

By substituting the unknown terms in (1.87) by their nonparametric estimators and subtracting the resulting expression from (1.86) we obtain

$$Y_{it} - \widehat{Y}_{it} = (X_{it} - \widehat{X}_{it})^\top \beta + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{1.91}$$

At this point, we must highlight that the resulting estimators of this previous equation may be subject to the random denominator problem when $\widehat{f}_{it}$ is too small. As a solution to this problem, these authors propose to multiply both sides of (1.91) by the density estimator $\widehat{f}_{it}$ following Powell et al. (1989), obtaining

$$Y_{it}\widehat{f}_{it} - \widehat{Y}_{it}\widehat{f}_{it} = (X_{it}\widehat{f}_{it} - \widehat{X}_{it}\widehat{f}_{it})^\top \beta + \epsilon_{it}\widehat{f}_{it}. \tag{1.92}$$

Thus, the ordinary least-squares (OLS) estimator of β now can be written as

$$\widetilde{\beta}_{OLS} = \left((X - \widehat{X})^\top (X - \widehat{X})\widehat{I}\right)^{-1} \left((X - \widehat{X})\widehat{I}\right)^\top (Y - \widehat{Y}), \tag{1.93}$$

where $(X - \widehat{X})\widehat{I}$ is a matrix of $NT \times d$ dimension with a typical row element $(X_{it} - \widehat{X}_{it})^\top I_{it}$, $I_{it} = I(\widehat{f}_{it} \geq b)$ for $I(\cdot)$ being the usual indicator function and $b = b_n(> 0)$ is a trimming parameter.

Under standard regularity conditions, in Li and Stengos (1996, Theorem 1, pp. 392) and Li and Ullah (1998, Proposition 1, page 151) it is shown the following convergence in distribution of this ordinary least-squares estimator,

$$\sqrt{N}\left(\widetilde{\beta}_{OLS} - \beta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Sigma\Phi^{-1}\right), \tag{1.94}$$

where $\Phi = \frac{1}{T}\sum_{t=1}^{T} E(\widetilde{X}_{1t}\widetilde{X}_{1t}^\top f_{1t}^2)$ and $\Sigma = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E(v_{1t}v_{1s}\widetilde{X}_{1t}\widetilde{X}_{1s}^\top f_{1t}^2 f_{1s}^2)$.

Looking at these properties, it is proved that this procedure allows to obtain a consistent semi-parametric estimator in this context of random effects, where $\widetilde{\beta}_{OLS}$ has the same asymptotic efficiency as the unfeasible estimator $\widehat{\beta}_{OLS}$. However, since this estimator ignores the information contained in the error term more efficient estimators can be obtained.

With the aim of using the structure of the variance-covariance matrix Ω, in Li and Ullah (1998) a feasible generalized least-squares (GLS) semi-parametric estimator for the coefficient of the linear component of (1.86) is proposed. Therefore, to compute $\widehat{\Omega}$ they propose to estimate (1.92) obtaining the residuals, $\widehat{\epsilon}_{it} = (Y_{it} - \widehat{Y}_{it}) - (X_{it} - \widehat{X}_{it})^\top \widetilde{\beta}_{OLS}$, so the unknown parameters of Ω may be estimated as

$$\widehat{\sigma}^2_\epsilon = \widehat{\sigma}^2 - \widehat{\sigma}^2_\mu \qquad \text{and} \qquad \widehat{\sigma}^2_1 = T\widehat{\sigma}^2_\mu + \widehat{\sigma}^2_\epsilon,$$

where $\widehat{\sigma}^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \widehat{\epsilon}^2_{it}\widehat{f}_{it}$ and $\widehat{\sigma}^2_\mu = \frac{1}{NT(T-1)}\sum_{i=1}^N \sum_{t=1}^T \widehat{\epsilon}_{it}\widehat{\epsilon}_{is}\widehat{f}_{it}\widehat{f}_{is}$, whereas $\widehat{f}_{it}$ and $\widehat{f}_{is}$ are defined as in (1.90).

Replacing the unknown terms by these estimators in (1.86), the feasible generalized least-squares semi-parametric (FGLS) estimator of β is

$$\widehat{\beta}_{FGLS} = \left((X - \widehat{X})^\top \widehat{\Omega}^{-1}(X - \widehat{X})\widehat{I}\right)^{-1} \left((X - \widehat{X})\widehat{I}\right)^\top \widehat{\Omega}^{-1}(Y - \widehat{Y}), \tag{1.95}$$

whose convergence in distribution, under some standard regularity conditions, is

$$\sqrt{NT}\left(\widehat{\beta}_{FGLS} - \beta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Sigma^{-1}\right), \tag{1.96}$$

where $\Sigma = T^{-1}E[\widetilde{X}_1^\top \Omega^{-1}\widetilde{X}_1]$, for $\widetilde{X}_1$ being a $T \times q$ dimensional matrix with $\widetilde{X}_{1t} = X_{1t} - E(X_{1t}|Z_{1t})$ as a typical row element, while Ω^{-1} is the inverse matrix of (1.4).

By comparing the asymptotic properties of the ordinary least-squares semi-parametric estimator and the feasible generalized least-squares semi-parametric estimator, in Li and Ullah (1998) it is shown that $\widehat{\beta}_{FGLS}$ is asymptotically more efficient than $\widetilde{\beta}_{OLS}$ when the error term has a one-way error component structure.

As we have just shown, the estimation strategy developed in Robinson (1988) can be easily generalized to several contexts within the framework of partially linear panel data models. However, it is true that the presence of heteroskedastic errors in the model of interest makes more difficult the use of this procedure. In this context, in You et al. (2010) it is considered an alternative method to obtain consistent nonparametric estimators that take into account a one-way error component structure and allow for unequal error variances, i.e., heteroskedasticity.

More precisely, in You et al. (2010) it is considered a one-way error component structure with heteroskedasticity of the following form

$$\epsilon_{it} = \mu_i + \sigma(Z_{it})v_{it}, \qquad i = 1, \cdots, N \qquad ; \qquad t = 1, \cdots, T, \tag{1.97}$$

where $\sigma_v^2 = 1$ is assumed without loss of generality and the variance-covariance matrix of the error term is written as

$$\Omega = E(\epsilon\epsilon^\top) = \sigma_\mu^2 I_N \otimes (\imath_T \imath_T^\top) + diag\left(\sigma^2(Z_{11}), \cdots, \sigma^2(Z_{NT})\right). \tag{1.98}$$

In order to obtain $\sqrt{N}$-consistent and asymptotically efficient estimators of the parameters of interest, in You et al. (2010) it is proposed a two-step estimator. In the first-step, standard estimators for β and $m(\cdot)$ are obtained as in Li and Ullah (1998). However, the difference with respect to their proposal is that, in order to compute both $E(Y_{it}|Z_{it})$ and $E(X_{it}|Z_{it})$ a local linear regression is used. In a second-step, the authors suggest to include information about the error component structure to obtain efficiency gains. In fact, they develop a generalized semi-parametric least-squares estimator that turns out to be unfeasible. Finally, they propose consistent estimators for the variance parameters, i.e., σ_μ^2 and $\sigma^2(z)$, using the residuals from the first-step estimators.

We start by giving the expressions for the estimators of β and $m(\cdot)$ with the local linear regression estimation procedure. In order to do so, let us rewrite the model (1.86) as

$$Y_{it} - X_{it}^\top \beta = m(Z_{it}) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.99}$$

so we can obtain an estimator of the quantities of interest by solving the following local linear regression problem

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left((Y_{it} - X_{it}^\top\beta) - \gamma_0 - \gamma_1(Z_{it} - z)\right)^2 K_h(Z_{it} - z). \tag{1.100}$$

Denote by $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ the minimizers of (1.100), they suggest as estimators of $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}(z;h) = \widehat{\gamma}_0$ and $\widehat{D}_m(z;h) = \widehat{\gamma}_1$, respectively,

$$\widehat{\gamma}_0 = (1,0)\left(D^\top W D\right)^{-1} D^\top W(Y - X\beta) = S(Y - X\beta) \tag{1.101}$$

and

$$\widehat{\gamma}_1 = (0,1)\left(D^\top W D\right)^{-1} D^\top W(Y - X\beta), \tag{1.102}$$

where $S = (1,0)\left(D^\top WD\right)^{-1} D^\top W$, $W = diag\left(K_h(Z_{11}-z),\cdots,K_h(Z_{NT}-z)\right)$ is a $NT\times NT$ matrix, $X = (X_{11},\cdots,X_{NT})^\top$ is a $NT\times d$ matrix and D is a $NT\times(1+q)$ matrix such that

$$D = \begin{bmatrix} 1 & (Z_{11}-z)^\top \\ \vdots & \vdots \\ 1 & (Z_{NT}-z)^\top \end{bmatrix}.$$

However, because β is an unknown parameter we can replace $m(Z_{it})$ by $\widehat{m}(Z_{it};h) = \widehat{\gamma}_0$ in (1.99), so the regression to estimate is of the form

$$\widehat{Y}_{it} = \widehat{X}_{it}^\top\beta + \epsilon_{it}^*, \quad i = 1,\cdots,N \quad ; \quad t = 1,\cdots,T, \tag{1.103}$$

where $(\widehat{Y}_{11},\cdots,\widehat{Y}_{NT})^\top = (I-S)Y$, $(\widehat{X}_{11},\cdots,\widehat{X}_{NT})^\top = (I-S)X$, $(\epsilon_{11}^*,\cdots,\epsilon_{NT}^*)^\top = (I-S)\epsilon + (I-S)m(Z)$, I is a $NT\times NT$ identity matrix and $m(Z) = (m(Z_{11}),\cdots,m(Z_{NT}))^\top$.

We denote by $\widehat{\beta}_{LSS}$ the least-squares semi-parametric estimator of (1.103) of the form

$$\widehat{\beta}_{LSS} = \left(X^\top(I-S)^\top(I-S)X\right)^{-1}\left(X(I-S)\right)^\top(I-S)Y, \tag{1.104}$$

whereas the local linear estimator of $m(\cdot)$ is written as

$$\widehat{m}_{LSS}(z;h) = (1,0)\left(D^\top WD\right)^{-1} D^\top W(Y - X\widehat{\beta}). \tag{1.105}$$

Note that these estimators are consistent but they fail to incorporate information from the structure of the variance-covariance matrix. Thus, in a second-step a feasible weighted least-squares semi-parametric estimator for both components is proposed to incorporate this information. In this way, what it is suggested in You et al. (2010) is to estimate both the variance of the error term and the error structure and, later, use this information to provide the efficient semi-parametric estimator.

Based on $\widehat{\beta}_{LLS}$ and $\widehat{m}_{LLS}(\cdot)$, the estimated residuals are

$$\widehat{\epsilon}_{it} = Y_{it} - X_{it}^\top\widehat{\beta}_{LLS} - \widehat{m}_{LLS}(Z_{it}), \tag{1.106}$$

and because $E(\epsilon_{it}\epsilon_{is}) = \sigma_\mu^2$, when $t\neq s$, and $E(\epsilon_{it}^2) = \sigma_\mu^2 + \sigma^2(Z_{it})$, consistent estimators of σ_μ^2 and $\sigma^2(\cdot)$ can be written as

$$\widehat{\sigma}_\mu^2 = \frac{1}{NT(T-1)}\sum_{i=1}^N\sum_{t=1}^T\sum_{t\neq s}^T\widehat{\epsilon}_{it}\widehat{\epsilon}_{is} \qquad \text{and} \qquad \widehat{\sigma}^2(z) = \sum_{i=1}^N\sum_{t=1}^T\omega_{it}(z)\widehat{\epsilon}_{it} - \widehat{\sigma}_\mu^2, \tag{1.107}$$

respectively, where $\omega_{it}(z)$ are some weighting functions of the local linear estimator that follow

$$\omega_{it}(z) = \frac{(Nh)^{-1}K((Z_{it}-z)/h)\,(A_{k2}(z)-(Z_{it}-z)A_{k1}(z))}{A_{k0}(z)A_{k2}(z)-A_{k1}^2(z)},$$

being $A_{ks}(z) = \frac{1}{Nh}\sum_{i=1}^{N}\sum_{t=1}^{T} K\left(\frac{Z_{it}-z}{h}\right)(Z_{it}-z)^{\lambda}$, for $\lambda = 0,1,2$.

Consequently, the estimator of Ω^{-1} is given by $\widehat{\Omega}^{-1} = \text{blockdiag}(\widehat{\Sigma}_1^{-1},\cdots,\widehat{\Sigma}_N^{-1})$, where

$$\begin{aligned}\widehat{\Sigma}_i^{-1} &= diag\left(\widehat{\sigma}^{-2}(Z_{i1}),\cdots,\widehat{\sigma}^{-2}(Z_{iT})\right)\\ &- \left(\widehat{\sigma}_{\mu}^{-2}+\sum_{t=1}^{T}\widehat{\sigma}^{-2}(Z_{it})\right)^{-1}\left(\widehat{\sigma}^{-2}(Z_{11}),\cdots,\widehat{\sigma}^{-2}(Z_{iT})\right)^{\top}\left(\widehat{\sigma}^{-2}(Z_{11}),\cdots,\widehat{\sigma}^{-2}(Z_{iT})\right).\end{aligned}$$

By replacing Ω by $\widehat{\Omega}$, the feasible weighted least-squares semi-parametric estimator (WSLSE) is

$$\widehat{\beta}_{WLSS} = \left(X^{\top}(I-S)^{\top}\widehat{\Omega}^{-1}(I-S)X\right)^{-1}(X(I-S))^{\top}\widehat{\Omega}^{-1}(I-S)Y, \tag{1.108}$$

whereas the feasible estimator of the nonparametric component is

$$\widehat{m}_{WLSS}(z;h) = (1,0)\left(D^{\top}WD\right)^{-1}D^{\top}W(Y-X\widehat{\beta}_{WLSS}). \tag{1.109}$$

Finally, under certain regularity conditions of these heteroskedastic models and comparing the behavior of both estimators in large samples, in You et al. (2010) it is emphasized that $\widehat{\beta}_{WLLS}$ is asymptotically more efficient than the usual semi-parametric estimator $\widehat{\beta}_{LLS}$ because the error component structure is considered.

1.3.2 Fixed effects

In partially linear models with fixed effects, the response variable is generated through the following statistical model

$$Y_{it} = X_{it}^{\top}\beta + m(Z_{it}) + \mu_i + v_{it}, \quad i = 1,\cdots,N \quad ; \quad t = 1,\cdots,T, \tag{1.110}$$

where X_{it} and Z_{it} are vectors of exogenous variables of $p\times 1$ and $q\times 1$ dimension, respectively, β is a $p\times 1$ vector of unknown parameters, $m(\cdot)$ is an unknown smooth function, μ_i the cross-sectional heterogeneity and v_{it} the idiosyncratic disturbances.

In this subsection, we are interested in the review of the recent literature on consistent estimation of the unknown parameters in (1.110) in the presence of unobserved individual heterogeneity that is correlated with the covariates. As in fully nonparametric panel data models with fixed effects, there are many proposals to consistently estimate this type of models. On one side, the so-called profile least-squares method and, in the other side, the differencing methods. In Su and Ullah (2006b) and Zhang et al. (2011) some alternative profile methods are proposed. Furthermore, recently in Ai et al. (2013) an estimator for the additive version of the semi-parametric regression model (1.110) is developed. This estimator ameliorates the dimensionality problem related to the nonparametric covariates. Within the differencing approach, in Baltagi and Li (2002) and Qian and Wang (2012) it is proposed to remove the fixed effects from the regression and to estimate the unknown smooth function using a series method and a marginal integration technique, respectively.

Profile least-squares estimators

On one hand, under the fixed effects approach and treating the unobserved effects as dummy variables to estimate, in Su and Ullah (2006b) it is followed the idea of the least-squares dummy variable estimator to develop a profile likelihood estimator for both parametric and nonparametric terms of the regression model (1.110).

Let $Y = (Y_{11}, \cdots, Y_{NT})^\top$ and $X = (X_{11}, \cdots, X_{NT})^\top$ be vectors of $NT \times 1$ dimension, $\mu = (\mu_2, \cdots, \mu_N)^\top$ a vector of $(N-1) \times 1$ dimension and $D = (I_N \otimes \imath_T)d$ a $NT \times (N-1)$ dimensional matrix, where $d = (-\imath_{N-1} I_{N-1})^\top$ is a $N \times (N-1)$ matrix, the standard locally weighted linear regression to estimate the quantities of interest in (1.110) can be written in matrix form as

$$\left(Y - D\mu - X^\top\beta - \widetilde{Z}\gamma\right)^\top W_H(z) \left(Y - D\mu - X^\top\beta - \widetilde{Z}\gamma\right), \tag{1.111}$$

where let K be a kernel function of the form $K_H(z) = |H|^{-1}K(H^{-1}z)$, $|H|$ is the determinant of a matrix of bandwidth sequences, i.e., $H = diag(h_1, \cdots, h_q)$, $W_H(z) = diag\left(K_H(Z_{11} - z), \cdots, K_H(Z_{NT} - z)\right)$ is a $NT \times NT$ matrix and $\widetilde{Z}$ is a $NT \times (1 + q)$

matrix of the form

$$\widetilde{Z} = \begin{bmatrix} 1 & \left(H^{-1}(Z_{11}-z)\right)^\top \\ \vdots & \vdots \\ 1 & \left(H^{-1}(Z_{NT}-z)\right)^\top \end{bmatrix}.$$

The above exposition suggests as estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}(z;H) = \widehat{\gamma}_0$ and $\widehat{D}_m(z;H) = \widehat{\gamma}_1$, respectively,

$$\widehat{\gamma}_0 = \widehat{m}(z;H) = s(z)(Y - D\mu - X\beta), \tag{1.112}$$

where $s(z) = e_1^\top S(z)$ for $S(z) = \left(\widetilde{Z}^\top W_H(z)\widetilde{Z}\right)^{-1} \widetilde{Z}^\top W_H(z)$, and $e = (1, 0, \cdots, 0)^\top$ is a $(1+q) \times 1$ selection matrix.

However, since μ and β are unknown parameters, they also need to be estimated. In order to do so, let us denote by $m(Z) = (m(Z_{11}), \cdots, m(Z_{NT}))^\top$ a $NT \times 1$ vector and replace (1.112) in the following optimization problem

$$(Y - D\mu - X\beta - m(Z))^\top (Y - D\mu - X\beta - m(Z)). \tag{1.113}$$

Hence, we can write the minimizers of (1.113) as

$$\widehat{\beta}_{PL} = \left(X^{*\top} M^* X^*\right)^{-1} X^{*\top} M^* Y^*, \tag{1.114}$$

and

$$\widehat{\mu}_{PL} = \left(D^{*\top} D^*\right)^{-1} D^{*\top}(Y^* - X^*\widehat{\beta}), \tag{1.115}$$

where $X^* = (I_{NT} - S)X$, $M^* = I_{NT} - D^*\left(D^{*\top}D^*\right)^{-1} D^*$, $D^* = (I_{NT} - S)D$, and $Y^* = (I_{NT} - S)Y$, let $S_i = (s(Z_{i1}), \cdots, s(Z_{iT}))$ be a $T \times T$ smoothing matrix. Note that since in Su and Ullah (2006a) it is introduced the identification condition $\sum_{i=1}^N \mu_i = 0$, then $\widehat{\mu}_1 = -\sum_{i=2}^N \widehat{\mu}_i$.

In this way, using (1.114) and (1.115) the profile likelihood estimator of the nonparametric term of $m(z)$ may be written as

$$\widehat{m}_{PL}(z;H) = s(z)\left(Y - D\widehat{\mu}_{PL} - X\widehat{\beta}_{PL}\right). \tag{1.116}$$

Under some standard smoothness conditions, some moment conditions on the error, i.e., $E|v_{it}|^{2+\delta} < \infty$, for some $\delta > 0$, and some assumptions on the behavior of the bandwidth, in Su and Ullah (2006b, Theorem 3.1, pp. 78) it is shown that as $N \to \infty$ and T is fixed, $\|H\| \to 0$ and $N|H|^2 \to \infty$ the asymptotic normality of this estimator is

$$\sqrt{N}\left(\widehat{\beta}_{FPL} - \beta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Omega\Phi^{-1}\right), \tag{1.117}$$

where Φ is a positive definite matrix of the form $\Phi = \sum_t E\left(\widetilde{X}_{it}\left(\widetilde{X}_{it} - T^{-1}\sum_{s=1}^T \widetilde{X}_{is}\right)^\top\right)$ and $\Omega = \sum_{s=1}^T \sum_{t=1}^T E\left(\widetilde{X}_{it}\left(\widetilde{X}_{is} - T^{-1}\sum_{\ell=1}^T \widetilde{X}_{i\ell}\right)^\top v_{it}v_{is}\right)$, for $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$.

Furthermore, for the estimator of the nonparametric component in Su and Ullah (2006b, Theorem 3.2, pp. 78) it is shown

$$\sqrt{N|H|}\left(\widehat{\gamma}_{PL}(z;H) - \gamma_{PL} - Q^{-1}\begin{pmatrix} \frac{\overline{f}(z)}{2}tr(\mu_2(K)Hm''(z)H) \\ 0 \end{pmatrix}\right)$$
$$\xrightarrow{d} \quad \mathcal{N}\left(0, Q^{-1}\Gamma Q^{-1}\right), \tag{1.118}$$

where $m''(z)$ is the second-order derivative matrix of $m(\cdot)$ at z,

$$Q = \overline{f}(z)\begin{pmatrix} 1 & 0^\top \\ 0 & \mu_2(K) \end{pmatrix} \qquad \text{and} \qquad \Gamma = \overline{\sigma}^2(z)\begin{pmatrix} \int K(u)^2 du & 0^\top \\ 0 & \int uu^\top K(u)du \end{pmatrix},$$

and we denote $\overline{f}(z) = \sum_{t=1}^T f_t(z)$ and $\overline{\sigma}^2(z) = \sum_{t=1}^T E\left((v_{it} - T^{-1}\sum_{s=1}^T v_{is})|Z_{it} = z\right) f_t(z)$.

However, note that despite the great advantages offered by these new procedures, the dimensionality problem characteristic of the nonparametric models is unsolved. In order to avoid the slower rates of convergence of these nonparametric estimators related to the course of dimensionality, a possible solution is to analyze an additive alternative expression of the regression model (1.110). Thus, the panel data partially linear model with fixed effects to estimate is

$$Y_{it} = X_{it}^\top\beta + m_1(Z_{1it}) + \cdots + m_q(Z_{qit}) + \mu_i + v_{it}, \quad i = 1,\cdots,N \quad ; \quad t = 1,\cdots,T, \tag{1.119}$$

where now $m(\cdot) = (m_1(\cdot),\cdots,m_q(\cdot))$ is a vector of unknown functions to estimate and the remaining components are defined as in (1.110).

In this context, in Ai et al. (2013) it is proposed to combine the polynomial spline series approximation with the profile least-squares procedure to obtain a semi-parametric least-squares dummy variables (SLSDV) estimator for the parametric component, and a series estimator of the nonparametric term. Under very weak conditions, these authors show that the semi-parametric least-squares dummy variables estimator is asymptotically normal and the series estimator achieves the optimal rate of convergence of the nonparametric regression. Later, with the aim of obtaining estimators that exhibit the oracle efficiency property, they develop a two-step local polynomial procedure based on a series method that enables us to impose the additive structure of the $m(\cdot)$ function. Since the nonparametric smoothing spline technique is beyond the scope of this study, we refer to Ai et al. (2013) for a detailed analysis of the proposed procedure and the study of the main asymptotic properties of the resulting estimators.

Differencing estimators

Another way to estimate consistently both the parameters and unknown functions of interest in (1.110) is to take first differences in this model. By doing so we obtain

$$\Delta Y_{it} = \Delta X_{it}^{\top}\beta + m(Z_{it}, Z_{i(t-1)}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \qquad (1.120)$$

where $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$.

In order to estimate directly the parameters of interest, in Li and Stengos (1996) it is proposed to take conditional expectations on all nonparametric covariates with the aim of removing the unknown smooth function. However, as it is noted in Baltagi and Li (2002), in such situations this technique presents some drawbacks. On one hand, taking conditional expectations on $(Z_{it}, Z_{i(t-1)})$ implies having to deal with the curse of dimensionality problem. In other words, in this case you have to regress $\Delta Y_{it} - E(\Delta Y_{it}|Z_{it}, Z_{i(t-1)})$ on $\Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})$ by the kernel method, so this estimator has to be defined on $I\!R^{2q}$ rather than $I\!R^{q}$. On the other hand, although these authors suggest how to estimate $m(Z_{it}, Z_{i(t-1)})$, they ignore the additive structure in (1.120) and they do not provide a nonparametric estimator for $m(Z_{it})$.

If $m(\cdot)$ is an unknown function twice differentiable in the interior of its support $\mathcal{A}$, being $\mathcal{A}$ a compact subset or $I\!R^q$, and $E[m''(z)] = E[\partial^2 m(z)/\partial z^2] < \infty$, the unknown function $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$ belongs to the class of additive functions $\mathcal{M}$ ($m \in \mathcal{M}$). Then, with the aim of taking into account the restriction that both additive functions share the same functional form, we distinguish two different approaches. On one hand, in Baltagi and Li (2002) an estimation method based on the series approach is developed. On the other hand, in Qian and Wang (2012) it is proposed an alternative method based on marginal integration techniques.

In particular, in Baltagi and Li (2002) the function $m(z)$ is approximated through the series $\rho^L(z)$ of $L \times 1$ dimension, where $L = L(N)$, provided the approximation function $\rho^L(z)$ meets the following features

i) $\rho^L(z) \in \mathcal{M}$,

ii) as far as L increases, there is a linear combination of $\rho^L(z)$ that can approximate any $m \in \mathcal{M}$ arbitrarily well in mean square error.

In this way, $\rho^L(z)$ approximates $m(z)$ and $\rho^L(Z_{it}, Z_{i(t-1)}) = \rho^L(Z_{it}) - \rho^L(Z_{i(t-1)})$ approximates $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$, where

$$\rho^L(Z_{it}, Z_{i(t-1)}) = \begin{bmatrix} \rho_1(Z_{it}) - \rho_1(Z_{i(t-1)}) \\ \rho_2(Z_{it}) - \rho_2(Z_{i(t-1)}) \\ \vdots \\ \rho_L(Z_{it}) - \rho_L(Z_{i(t-1)}) \end{bmatrix}. \tag{1.121}$$

For any scalar or vector function $W(z)$, $E_{\mathcal{M}}(W(z))$ is denoted as an element that belongs to $\mathcal{M}$ and it is the closest function to $W(z)$ among all the functions in $\mathcal{M}$; see Baltagi and Li (2002) for further details. We denote by $P = (\rho^L_{11}, \cdots, \rho^L_{NT})$ a $NT \times L$ matrix, where $\rho^L_{it} = \rho^L(Z_{it}, Z_{i(t-1)})$. For the sake of simplicity, let us define $\theta(z) = E(X|Z = z)$ and $m(z) = E_{\mathcal{M}}(\theta(z))$.

Thus, let ΔY be a $NT \times 1$ vector with a typical element ΔY_{it}, and ΔX and Δv vectors of $NT \times 1$ dimension defined in a similar way, the expression (1.120) may be written in matrix form as

$$\Delta Y = \Delta X \beta + M + \Delta v. \tag{1.122}$$

By multiplying both sides of (1.122) by $\overline{P} = P\left(P^\top P\right)^{-1} P^\top$ and subtracting the resulting expression from (1.122), they obtain

$$\Delta Y - \Delta\widetilde{Y} = (\Delta X - \Delta\widetilde{X})\beta + (M - \widetilde{M}) + (\Delta v - \Delta\widetilde{v}), \tag{1.123}$$

where we denote $\Delta\widetilde{Y} = \overline{P}\Delta Y = P\gamma_{\Delta Y}$, for $\gamma_{\Delta Y} = \left(P^\top P\right)^{-1} P^\top \Delta Y$. Note that $\widetilde{M}$, $\Delta\widetilde{X}$, and $\Delta\widetilde{v}$ are defined in a similar way.

Thus, the least-squares estimator for β is defined such that

$$\widehat{\beta} = \left((X-\widetilde{X})^\top(X-\widetilde{X})\right)^{-1}(X-\widetilde{X})^\top(Y-\widetilde{Y}), \tag{1.124}$$

whereas for the smooth function $m(z)$ they propose $\widehat{m}(z) = \rho^L(z)^\top\widehat{\gamma}$ as the nonparametric estimator, where

$$\widehat{\gamma} = \left(P^\top P\right)^{-1} P^\top(Y - X\widehat{\beta}). \tag{1.125}$$

Under standard conditions of a series approach, in Baltagi and Li (2002, Theorem 2.1, pp. 108) it is established the following asymptotic normality of $\widehat{\beta}$ as $N \to \infty$ and T is fixed,

$$\sqrt{N}\left(\widehat{\beta}-\beta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Omega\Phi^{-1}\right), \tag{1.126}$$

where, for $\xi_{it} = X_{it} - W(Z_{it})$ and $W(Z_{it}) = E_{\mathcal{M}}(\theta(Z_{it}))$, Φ is a positive definite matrix of the form $\Phi = T^{-1}\sum_{t=1}^T E(\xi_{it}\xi_{it}^\top)$ and $\Omega = T^{-1}\sum_{t=1}^T E\left(\sigma^2_{\Delta v}(X_{it}, Z_{it})\xi_{it}\xi_{it}^\top\right)$, being $\sigma^2_{\Delta v}(X_{it}, Z_{it}) = E\left(\Delta v_{it}^2 | X_{it} = x, Z_{it} = z\right)$.

Furthermore, in order to show the consistency result of the nonparametric estimate, in Baltagi and Li (2002, Theorem 2.2, pp. 108) it is obtained

i) $\sup_{z\in\mathbb{R}^q} |\widehat{m}(z;H) - m(z)| = O_p(\zeta_0(L))(\sqrt{L}/(\sqrt{L} + L^{-\delta})$, where $\zeta_0(L)$ is a sequence of constant that satisfies $\sup_{z\in\mathbb{R}^q} \|P^L(z)\| < \zeta_0(L)$, let $\delta > 0$ be certain constant.

ii) $N^{-1}(\widehat{m}(z;H) - m(z))^2 = O_p(L/N - L^{-2\delta})$.

iii) $\int(\widehat{m}(z;H) - m(z))^2 dF(z) = O_p(L/N + L^{-2\delta})$, where $F(\cdot)$ is the cumulative distribution function of z.

In this way, it is proved that this new procedure provides estimators of the smooth function at the standard nonparametric rate. For the proofs of these results we refer to the appendix in Baltagi and Li (2002).

On the other hand, in Qian and Wang (2012) a non-iterative method based on the marginal integration technique is proposed to provide an estimator of the unknown function $m(Z_{it})$

that takes into account for the additive structure of $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$. Thus, using first differences in (1.110) to avoid the statistical dependence problem between μ_i and the regressors of the model they obtain

$$\Delta Y_{it} = \Delta X_{it}^\top \beta + m(Z_{it}) - m(Z_{i(t-1)}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \quad (1.127)$$

and taking as benchmark the proposal in Li and Stengos (1996), in Qian and Wang (2012) it is proposed to estimate the lineal component β following a weighted density regression model that enables us to avoid the random denominator problem usual in the estimation of nonparametric kernel regression, i.e.,

$$\begin{aligned} f_{it,i(t-1)} \left(\Delta Y_{it} - E(\Delta Y_{it}|Z_{it}, Z_{i(t-1)})\right) &= f_{it,i(t-1)} \left(\Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})\right)^\top \beta \\ &+ f_{it,i(t-1)} \Delta v_{it}, \qquad (1.128) \end{aligned}$$

where $f_{it,i(t-1)} = f(Z_{it}, Z_{i(t-1)})$ is the joint density function. Note that due to the regression in first differences and because we take conditional expectations over all nonparametric covariates, the resulting estimator has to be defined on $\mathbb{R}^{2q}$ rather than $\mathbb{R}^q$.

Assuming that there exist an instrumental variable vector, $W_{it} \in \mathbb{R}^d$, and replacing the unknown parameters in (1.118) by their consistent estimators, i.e., $\Delta\widehat{Y}_{it}$, $\Delta\widehat{X}_{it}$ and $\widehat{f}_{it,i(t-1)}$, respectively, they obtain

$$\widehat{f}_{it,i(t-1)}(\Delta Y_{it} - \Delta\widehat{Y}_{it}) = \widehat{f}_{it,i(t-1)}(\Delta X_{it} - \Delta\widehat{X}_{it})^\top \beta + \widehat{f}_{it,i(t-1)} \Delta v_{it}, \qquad (1.129)$$

where denoting by $\widehat{f}_{it,i(t-1)} = \widehat{f}(Z_{it}, Z_{i(t-1)})$ and $\Delta\widehat{Y}_{it} = \widehat{E}(\Delta Y_{it}|Z_{it}, Z_{i(t-1)})$, these estimators have the form

$$\widehat{f}_{it,i(t-1)} = \frac{1}{NT} \sum_{js} K_h(Z_{it} - Z_{js}) K_h(Z_{i(t-1)} - Z_{j(s-1)})$$

and

$$\Delta\widehat{Y}_{it} = \frac{1}{NT} \sum_{js} \Delta Y_{js} K_h(Z_{it} - Z_{js}) K_h(Z_{i(t-1)} - Z_{j(s-1)}) / \widehat{f}_{it,i(t-1)},$$

being $K_h(u) = K(u/h)$ the kernel function and h the bandwidth parameter. We define $\Delta\widehat{X}_{it} = \widehat{E}(\Delta X_{it}|Z_{it}, Z_{i(t-1)})$ in a similar way.

In this way, the instrumental variable (IV) estimator proposed by these authors is

$$\widetilde{\beta}_{IV} = \left((\Delta W - \Delta\widehat{W})^\top(\Delta X - \Delta\widehat{X})\right)^{-1}\left((\Delta W - \Delta\widehat{W})\widehat{I}\right)^\top(\Delta Y - \Delta\widehat{Y}), \quad (1.130)$$

where, let I be the indicator function previously defined, $(\Delta W - \Delta\widehat{W})\widehat{I}$ and $(\Delta X - \Delta\widehat{X})\widehat{I}$ are $N(T-1)\times d$ matrices whose typical row element are $(\Delta W_{it} - \Delta\widehat{W}_{it})^\top\widehat{I}_{it}$ and $(\Delta X_{it} - \Delta\widehat{X}_{it})^\top\widehat{I}_{it}$, respectively.

By adapting the assumptions in Li and Stengos (1996); in Qian and Wang (2012) it is imposed that $f(Z_{it}, Z_{i(t-1)})$ is a bound density function and at least first-order partially differentiable, with a remainder term that is Lipschitz-continuous. In this situation, in Qian and Wang (2012, Theorem 1, pp. 485) and under fairly standard nonparametric assumptions it is shown the following asymptotic distribution of the IV estimator for the linear component,

$$\sqrt{N}\left(\widetilde{\beta}_{IV} - \beta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Psi^{-1}\Gamma\Psi^{-1}\right), \quad (1.131)$$

where, for $\Delta\widetilde{W}_{it} = \Delta W_{it} - E(\Delta W_{it}|Z_{it}, Z_{i(t-1)})$ and $\Delta\widetilde{X}_{it} = \Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})$, $\Gamma = T^{-2}\sum_{t=2}^T\sum_{s=2}^T E\left(\Delta v_{1t}\Delta v_{1s}\Delta\widetilde{W}_{1t}\Delta\widetilde{W}_{1s}^\top f_{1t}^2 f_{1s}^2\right)$ and Ψ is a nonsingular matrix of the form $\Psi = T^{-1}\sum_{t=2}^T E\left(\Delta\widetilde{W}_{it}\Delta\widetilde{X}_{it}^\top f_{it}^2\right)$.

Furthermore, in Qian and Wang (2012) it is presented a new estimator of the nonparametric component, $m(z)$. This type of results cannot be found in other related papers such as Li and Stengos (1996) and Baltagi and Li (2002). They are only concerned about the parametric component. Let us denote by $\Delta Y_{it}^* = \Delta Y_{it} - \Delta X_{it}^\top\widehat{\beta}$, so the regression model (1.127) can be written as

$$\Delta Y_{it}^* = m(Z_{it}, Z_{i(t-1)}) + \Delta v_{it}, \qquad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \quad (1.132)$$

However, by the fact that $m(Z_{it}, Z_{i(t-1)})$ is an additive function the estimation task of $m(\cdot)$ is greatly complicated. In this context, in Henderson et al. (2008) it is developed an iterative backfitting procedure to obtain $\widehat{m}(\cdot)$ based on a first-order condition of the maximum likelihood criterion, whereas in Lee and Mukherjee (2008) it is proposed to make a Taylor approximation of the smooth function. Nevertheless, this latter approximation removes $m(\cdot)$ from the differencing model so they can only provide an estimator for the first-order derivative function.

In view of these results, in Qian and Wang (2012) a non-iterative method based on the marginal integration technique is proposed. Specifically, they present a two-step procedure in which they first use conventional multivariate nonparametric techniques, such as the Nadaraya-Watson estimator or the local linear regression. Later, the function $m(\cdot)$, evaluated in z_1, is obtained through the marginal integration of the previous estimator. Therefore, using the local linear regression procedure to estimate $m(Z_{it}, Z_{i(t-1)})$, in Qian and Wang (2012) it is proposed to solve the following locally weighted linear least-squares problem for α,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it}^{*} - \alpha - (Z_{it}-z_1)^{\top}\beta_0 - (Z_{i(t-1)}-z_2)^{\top}\beta_1\right)^2 K_H(Z_{it}-z_1)K_H(Z_{i(t-1)}-z_2), \tag{1.133}$$

where z_1 and z_2 are points in the interior of the support of $f(\cdot)$.

This latter expression suggests that let $\widehat{\alpha}$ be a minimizer of (1.133), the estimator for $m(Z_{it}, Z_{i(t-1)})$ is

$$\widehat{m}(z_1, z_2; H) = \widehat{\alpha} = e_1^{\top}\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}\widetilde{Z}^{\top}W\Delta Y^{*}, \tag{1.134}$$

where W and $\widetilde{Z}$ is a $N(T-1)\times N(T-1)$ and $N(T-1)\times(1+2q)$ matrix, respectively, of the following form

$$W = diag\left(K_H(Z_{12}-z_1)K_H(Z_{11}-z_2), \cdots, K_H(Z_{NT}-z_1)K_H(Z_{N(T-1)}-z_2)\right)$$

and

$$\widetilde{Z} = \begin{bmatrix} 1 & (Z_{12}-z_1)^{\top} & (Z_{11}-z_2)^{\top} \\ \vdots & \vdots & \vdots \\ 1 & (Z_{NT}-z_1)^{\top} & (Z_{N(T-1)}-z_2)^{\top} \end{bmatrix}.$$

If our interest is the estimation of the partial derivatives for the unknown functions, i.e. $D_{m_1}(z) = \partial m(z_1, z_2)/\partial z_1$ and $D_{m_2}(z) = \partial m(z_1, z_2)/\partial z_2$, it is enough to minimize (1.133) for β_0 and β_1. Thus, we could propose as estimators for β_0 and β_1, $vec(\widehat{D}_{m_1}(z_1; H)) = \widehat{\beta}_0$ and $vec(\widehat{D}_{m_2}(z_2; H)) = \widehat{\beta}_1$, respectively. However, since the objective of these authors is to provide an estimator of the unknown function $m(Z_{it})$ they propose to integrate marginally the estimated function $\widehat{m}(z_1, z_2)$, i.e.,

$$\widehat{m}(z_1; H) = \int \widehat{m}(z_1, z_2) q(z_2) dz_2, \tag{1.135}$$

where $q(\cdot)$ is a predetermined density function.

With the aim of avoiding strict usual identification restrictions of the marginal integration technique (i.e., $\int m(z_1)q(z_1)dz_1 = 0$ proposed in Hengartner and Sperlich (2005)) or numerical integration methods such as Simpson's or Trapezoidal rules, in Qian and Wang (2012) an alternative strategy is developed. In particular, they propose to generate *i.i.d.* samples of the $q(\cdot)$ distribution such as Z_k^*, for $k = 1, \cdots, NT$, and compute

$$\widehat{m}_{MC}(z_1; H) = \frac{1}{N(T-1)} \sum_{k=1}^{N} T\widehat{m}(z_1, Z_k^*). \tag{1.136}$$

As Qian and Wang (2012) emphasize, if NT is large enough $\widehat{m}_{MC}(\cdot)$ approximates considerably well to $\widehat{m}(\cdot)$ and we choose $q(\cdot)$ to be the density function of Z_{it}, we can use the sample version of (1.134) rather than (1.135), i.e.,

$$\widehat{m}_S(z_1; H) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{m}(z_1, Z_{it}). \tag{1.137}$$

Under standard conditions of the marginal integration technique, these authors show that the nonparametric estimator (1.137) behaves asymptotically equal to (1.135) when $q(\cdot)$ is the density function of Z_{it}, bounded and twice differentiable and when it satisfies $\int m(z_1)q(z_1) = 0$. Therefore, they obtain

$$N^{2/(4+q)} \left(\widehat{m}_S(z_1; H) - m(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(B(z_1), V(z_1)\right), \tag{1.138}$$

where

$$\begin{aligned} B(z_1) &= \frac{1}{2}\mu_2(K)\left(tr(H\mathcal{H}_m(z_1)) - \int tr(H\mathcal{H}_f(z_2))q(z_2)dz_2\right), \\ V(z_1) &= \frac{\overline{\sigma}^2 R^q(K)}{T|H|^{1/2}} \left(\int \frac{q^2(z_2)}{f_2(z_1, z_2)} q(z_2) dz_2\right), \end{aligned}$$

let $\mathcal{H}_m(z_1)$ be the Hessian matrix of $m(\cdot)$ and $\overline{\sigma}^2 = T^{-1}\sum_{t=2}^{T} \sigma_t^2$.

By analyzing in detail these asymptotic results, in Qian and Wang (2012) it is emphasized that if Z_{it} is *i.i.d.* across t as well as i, the asymptotic variance takes the conventional form $\frac{\overline{\sigma}^2 R^q(K)}{T|H|^{1/2}} f(z_1)^{-1}$ when $q(\cdot) = f(\cdot)$. In addition, when Z_{it} is accurately predictable by $Z_{i(t-1)}$, the conditional density function $f(z_1|z_2)$ is close to zero except in a small neighborhood of z_2, and this method may fail. Finally, note that if $\widehat{m}(z_1, z_2)$ is estimated using the Nadaraya-Watson kernel smoothing the asymptotic variance remains without changes, but the asymptotic bias is different.

1.4 Varying coefficient models

In many scientific areas such as economics, finance, politics and so on, nonparametric and semi-parametric panel data models are unable to detect some features hidden in the data set. In this context, varying coefficient models that allow for some coefficients to change smoothly with some covariates appear as a solution. In what follows, we analyze the estimation results of varying coefficient models under random and fixed effects, and the presence of endogenous explanatory variables.

1.4.1 Random effects

In this case, the response variable is generated through the following statistical model

$$Y_{it} = U_{it}^{\top}\beta + X_{it}^{\top} m(Z_{it}) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.139}$$

where U_{it}, X_{it} and Z_{it} are exogenous variables of $p \times 1$, $d \times 1$ and $q \times 1$ dimension, respectively, β is an unknown parameter and $m(\cdot)$ is a smooth function to estimate. Also, the composed error term ϵ_{it} follows a one-way error component structure so it has the form

$$\epsilon_{it} = \mu_i + v_{it},$$

where μ_i and v_{it} are *i.i.d.* random variables with zero mean and finite variances, σ_μ^2 and σ_v^2, respectively. In addition, we assume μ_i and v_{it} are independent.

As previously, we define V as a $T \times T$ matrix of the same form as in (1.3). Also, since the observations are independent along individuals, the variance-covariance matrix of the composed error term Ω has the standard form as in (1.4), and the inverse term of this variance-covariance matrix may be written as

$$\Omega^{-1} = I_N \otimes \left[(T\sigma_\mu^2 + \sigma_v^2)^{-1} \imath_T \imath_T^\top / T + \sigma_v^{-2} (I_T - \imath_T \imath_T^\top / T) \right]. \tag{1.140}$$

As the reader may see in (1.139), this specification has the particularity that encompasses the features of other statistical models of interest. Specifically, if $X_{it} = 1$ the expression (1.139) becomes into the partially linear model previously analyzed in Section 1.2.1, while if $m(Z_{it})$ is a constant parameter we obtain the conventional linear panel data model.

In order to provide consistent estimators of the unknown terms of (1.139), in Zhou et al. (2010) it is shown that it is possible to follow a similar procedure as the proposed in You et al. (2010) to estimate partially linear panel data models with heteroskedastic errors taking into account the information contained in Ω.

To this end, in Zhou et al. (2010) a profile least-squares approach is used to estimate both the parametric component β and the nonparametric term $m(\cdot)$. Thus, for any given β, they propose to estimate the parameter of interest solving the following weighted local linear least-squares problem

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left((Y_{it} - X_{it}^\top \beta) - X_{it}^\top \gamma_0 - (X_{it} \otimes (Z_{it} - z))^\top \gamma_1 \right)^2 K_h(Z_{it} - z), \tag{1.141}$$

where h is the bandwidth parameter.

Denote by $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ the minimizers of (1.141), they suggest as estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}(z; h) = \widehat{\gamma}_0$ and $vec(\widehat{D}_m(z; h)) = \widehat{\gamma}_1$, respectively,

$$\widehat{\gamma}_0 = (1, 0) \left(D^\top W D \right)^{-1} D^\top W (Y - X\beta) = S(Y - X\beta), \tag{1.142}$$

and

$$\widehat{\gamma}_1 = (0, 1) \left(D^\top W D \right)^{-1} D^\top W (Y - X\beta), \tag{1.143}$$

where $Y = (Y_{11}, \cdots, Y_{NT})^\top$ is a $NT \times 1$ vector, $X = (X_{11}, \cdots, X_{NT})^\top$ and $W = diag\,(K_h(Z_{11} - z), \cdots, K_h(Z_{NT} - z))$ a $NT \times q$ and $NT \times NT$ matrix, respectively, whereas

D is a $NT \times d(1+q)$ matrix of the form

$$D = \begin{bmatrix} X_{11}^\top, & X_{11}^\top \otimes (Z_{11}-z)^\top \\ \vdots & \vdots \\ X_{NT}^\top, & X_{NT}^\top \otimes (Z_{NT}-z)^\top \end{bmatrix}.$$

Because (1.142) and (1.143) are unfeasible estimators, in Zhou et al. (2010) it is proposed to substitute $\widehat{\gamma}_0$ into (1.139) as an estimator for $m(z)$ obtaining

$$\widehat{Y}_{it} = \widehat{X}_{it}^\top \beta + \epsilon_{it}^*, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.144}$$

where ϵ_{it}^* are the residual errors, $\widehat{Y} = (\widehat{Y}_{11}, \cdots, \widehat{Y}_{NT}) = (I-S)Y$, $\widehat{X} = (\widehat{X}_{11}, \cdots, \widehat{X}_{NT}) = (I-S)X$ and $\epsilon^* = (\epsilon_{11}^*, \cdots, \epsilon_{NT}^*)^\top = (I-S)\epsilon$, let $\epsilon = (\epsilon_{11}, \cdots, \epsilon_{NT})^\top$ be a $NT \times 1$ vector and I a $NT \times NT$ identity matrix.

Also, let 0_q be a $q \times 1$ vector of zeros, S is a matrix of the form

$$S = \begin{bmatrix} \left(X_{11}^\top \quad 0_q^\top\right) \left(D_{11}^\top W_{11} D_{11}\right)^{-1} D_{11}^\top W_{11} \\ \vdots \\ \left(X_{NT}^\top \quad 0_q^\top\right) \left(D_{NT}^\top W_{NT} D_{NT}\right)^{-1} D_{NT}^\top W_{NT} \end{bmatrix}.$$

In this way, applying the least-squares method to (1.144) we obtain the following profile least-squares (PLS) estimator for β

$$\widehat{\beta}_{PLS} = \left(\widehat{X}^\top \widehat{X}\right)^{-1} \widehat{X}^\top \widehat{Y}, \tag{1.145}$$

whereas the local linear estimator (PLL) for $m(\cdot)$ is

$$\widehat{\gamma}_{0PLL} = \widehat{m}_{PLL}(z;h) = (1,0)\left(D^\top W D\right)^{-1} D^\top W (Y - X\widehat{\beta}_{PLS}). \tag{1.146}$$

Despite the consistency of $\widehat{\beta}_{PLS}$ and $\widehat{m}_{PLS}(z;h)$, they do not take into account the within-group correlation of the error term, so in Zhou et al. (2010) it is proposed to use the one-way error structure to obtain asymptotically efficient estimators. In order to do so, these authors develop a similar procedure as in You et al. (2010) using the residuals of the previous fitting to estimate the structure of the error. Thus, computing the residuals as

$$\widehat{\epsilon}_{it} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it} - U_{it}^\top \widehat{\beta}_{PLS} - X_{it}^\top \widehat{m}_{PLS}(Z_{it}), \tag{1.147}$$

they propose to estimate $E(\epsilon_{it}\epsilon_{is}) = \sigma_\mu^2$, when $t \neq s$, and $E(\epsilon_{it}^2) = \sigma_\mu^2 + \sigma_v^2$ as

$$\widehat{\sigma}_\mu^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t} \widehat{\epsilon}_{it}\widehat{\epsilon}_{is} \qquad \text{and} \qquad \widehat{\sigma}_v^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 - \widehat{\sigma}_\mu^2.$$

According to (1.139), we can replace the unknown terms in (1.140) by their consistent estimators obtaining the consistent estimator for Ω^{-1}, i.e.,

$$\widehat{\Omega}^{-1} = I_N \otimes \left[(T\widehat{\sigma}_\mu^2 + \widehat{\sigma}_v^2)^{-1} \imath_T \imath_T^\top / T + \widehat{\sigma}_v^{-2} (I_T - \imath_T \imath_T^\top / T) \right] \tag{1.148}$$

and using this information to estimate the unknown parameters in (1.139), in Zhou et al. (2010) the following weighted profile least-squares (WPLS) estimator for the parametric term is proposed

$$\widehat{\beta}_{WPLS} = \left(\widehat{X}_\omega^\top \widehat{\Omega}^{-1} \widehat{X}_\omega \right)^{-1} \widehat{X}_\omega^\top \widehat{\Omega}^{-1} \widehat{Y}_\omega, \tag{1.149}$$

where $\widehat{Y}_\omega = (I_{NT} - S_\omega)Y$ and $\widehat{X}_\omega = (I_{NT} - S_\omega)X$, being

$$S_\omega = \begin{bmatrix} \begin{pmatrix} X_{11}^\top & 0_q^\top \end{pmatrix} \left(D_{11}^\top W_{11} \widehat{\Omega}^{-1} D_{11} \right)^{-1} D_{11}^\top W_{11} \widehat{\Omega}^{-1} \\ \vdots \\ \begin{pmatrix} X_{NT}^\top & 0_q^\top \end{pmatrix} \left(D_{NT}^\top W_{NT} \widehat{\Omega}^{-1} D_{NT} \right)^{-1} D_{NT}^\top W_{NT} \widehat{\Omega}^{-1} \end{bmatrix}.$$

Furthermore, the estimator for the nonparametric term has the form

$$\widehat{m}_{WPLS}(z;h) = (I_{q\times 1}, 0_{q\times q}) \left(D^\top W \widehat{\Omega}^{-1} D \right)^{-1} D^\top W \widehat{\Omega}^{-1} (Y - X\widehat{\beta}_{WPLS}) \tag{1.150}$$

and, under some standard regularity conditions, in Zhou et al. (2010, Theorem 5.1, pp. 254) the following asymptotic distribution is obtained when $N \to \infty$

$$\sqrt{N} \left(\widehat{\beta}_{WPLS} - \beta \right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Sigma^{-1}\right), \tag{1.151}$$

where, $\Sigma = E(U_1^\top V U_1) - E(U_1^\top V X_1) \left(E(X_1^\top V X_1) \right)^{-1} E(X_1^\top V U_1)$, for $U_i = (U_{i1}, \cdots, U_{iT})^\top$ and $X_i = (X_{i1}, \cdots, X_{iT})^\top$, while in Zhou et al. (2010, Theorem 5.2, pp. 255) it is shown that as N tends to infinity

$$\sqrt{NTh} \left(\widehat{m}_{WPLS}(z;h) - m(z) - \frac{h}{2} \frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} m''(z) \right) \quad \xrightarrow{d} \quad \mathcal{N}(0, \Sigma), \tag{1.152}$$

where now we denote $\Sigma = ((cv_0 + 2cc_1v_1 + c_1v_2)) \left(E(X_1^\top V X_1)\right)^{-1}$, $c = \mu_2/(\mu_2 - \mu_1^2)$ and $c_1 = -\mu_1/(\mu_2 - \mu_1^2)$, for $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$ and $v_j = \int_{-\infty}^{\infty} u^j K^2(u)du$, being $f(z)$ the density function.

By comparing the asymptotic results of these estimators, in Zhou et al. (2010) it is highlighted that $\widehat{\beta}_{WPLS}$ and $\widehat{m}_{WPLS}(\cdot)$ are asymptotically more efficient than $\widehat{\beta}_{PLS}$ and $\widehat{m}_{PLS}(\cdot)$, respectively, in the sense that they have a lower asymptotic variance-covariance matrix.

1.4.2 Fixed effects

Now, we extend our analysis to the following varying coefficient panel data models with fixed effects,

$$Y_{it} = X_{it}^\top m(Z_{it}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.153}$$

where X_{it} and Z_{it} are $d \times 1$ and $q \times 1$ vector of exogenous variables, respectively, $m(\cdot)$ is a vector that contains d smooth functions, μ_i the unobserved cross-sectional heterogeneity and v_{it} the idiosyncratic disturbance. Also, it is allowed that μ_i is correlated with X_{it} and/or Z_{it}, with an unknown correlation structure.

As it is usual in this literature, to avoid the statistical dependence problem between μ_i and some/all the regressors we may apply the first differences transformation to (1.153) obtaining

$$Y_{it} - Y_{i(t-1)} = X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)}) + v_{it} - v_{i(t-1)}, \tag{1.154}$$

for $i = 1, \cdots, N$ and $t = 2, \cdots, T$, or using a mean deviation transformation that gives us

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = X_{it}^\top m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} X_{is}^\top m(Z_{is}) + v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is}, \tag{1.155}$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$.

The methodological simplicity of these transformations and their good asymptotic results in standard situations make them very interesting tools when the unknown functions are estimated through conventional methods of nonparametric local smoothing. However, the use of these nonparametric techniques for differencing varying coefficient is much more complex than we might expect at first glance; see Su and Ullah (2011). The reason is that, if we denote $\Delta Y_{it} = Y_{it} - Y_{i(t-1)}$ and $\ddot{Y}_{it} = Y_{it} - T^{-1}\sum_{s=1}^{T} Y_{is}$, for each individual i, $E(\Delta Y_{it}|X_{it}, X_{i(t-1)}, Z_{it}, Z_{i(t-1)})$ and $E(\ddot{Y}_{it}|X_{it}, X_{is}, Z_{it}, Z_{is})$ provide linear combinations of $X_{it}^{\top} m(Z_{it})$ for different periods t. Thus, to obtain an estimator of $m(\cdot)$ is necessary to estimate an additive model whose unknown functions share the same functional form. To our knowledge, there is no nonparametric procedure that allows us to estimate $m(\cdot)$ directly. The only available option is the use of iterative techniques such as the backfitting algorithm or the marginal integration method. Nevertheless, note that this latter case presents some awkward features as its high computational cost. In particular, to obtain an estimator of this type we must compute $O(NT^3|H|^{1/2})$ operations, i.e., we have to perform NT^2 regressions and each one requires $O(NT|H|^{1/2})$ operations.

In this context and with the aim of minimizing the high computational cost and the laborious asymptotic analysis associated to these iterative estimators, in Sun et al. (2009) it is proposed a non-iterative estimator based on a local linear regression approach. As we have stated in previous sections, due to the existence of the fixed effects, the unknown parameters in (1.153) cannot be estimated directly.

Let $Y = (Y_{11}, \cdots, Y_{NT})^{\top}$ and $V = (v_{11}, \cdots, v_{NT})^{\top}$ are $NT \times 1$ vectors, and $B(X, m(Z))$ a $NT \times 1$ vector that stacks all $X_{it}^{\top} m(Z_{it})$, the model (1.153) can be written in matrix form such as

$$Y = B(X, m(Z)) + D\mu + V, \tag{1.156}$$

where $\mu = (\mu_2, \cdots, \mu_N)^{\top}$ is a $(N-1)\times 1$ vector and $D = (-\imath_{N-1} I_{N-1}) \otimes \imath_T$ a $NT \times (N-1)$ matrix.

As it is standard in this kind of literature, the parameters of interest can be estimated by minimizing the following criterion function

$$(Y - B(X, m(Z)) - D\mu)^{\top} K_H(z) (Y - B(X, m(Z)) - D\mu), \tag{1.157}$$

where $K_H(z) = diag(K_H(Z_{11} - z), \cdots, K_H(Z_{NT} - z))$ is a $NT \times NT$ diagonal matrix.

Denote by $\widehat{\mu}$ the minimizer of (1.157), it may be written as

$$\widehat{\mu}(z;H) = \left(D^\top K_H(z) D\right)^{-1} D^\top K_H(z)(Y - B(X, m(Z))), \tag{1.158}$$

but this estimator is unfeasible since it depends on some unknown terms such as $m(Z)$. In order to overcome it, in Sun et al. (2009) it is proposed to combine a least-squared method with a local linear regression approach of (1.156) to remove the fixed effects using kernel-based weights. Therefore, they propose to minimize the following weighted criterion function

$$\left(Y - \widetilde{Z} vec\left(\beta(z)\right)\right)^\top W_H(z) \left(Y - \widetilde{Z} vec\left(\beta(z)\right)\right), \tag{1.159}$$

where the weighting matrix now has the form $W_H(z) = M(z)^\top K_H(z) M(z)$, for $M(z) = I_{NT} - D\left(D^\top K_H(z) D\right)^{-1} D^\top K_H(z)$ in such way that $M(z)D\mu \equiv 0_{NT\times 1}$ for all z, and $\widetilde{Z}$ is a $NT \times d(1+q)$ matrix of the form

$$\widetilde{Z} = \begin{bmatrix} X_{11}^\top & X_{11}^\top \otimes (Z_{11} - z)^\top \\ \vdots & \vdots \\ X_{NT}^\top & X_{NT}^\top \otimes (Z_{NT} - z)^\top \end{bmatrix}.$$

Also, we define $\beta_r = (m_r(z) \quad (Hm_r'(z))^\top)^\top$ as a $(1+q)\times 1$ column vector for $r = 1, \cdots, d$, and $\beta(z) = (\beta_1(z), \cdots, \beta_p(z))^\top$ as a $d \times (1+q)$ parameter matrix. Therefore, the first column of $\beta(z)$ is $m(z)$. For the sake of simplicity, they stack the matrix $\beta(z)$ into a $d(1+q)\times 1$ column vector and denote it by $vec(\widehat{\beta})$.

Then, the above expression (1.159) suggests as nonparametric estimators

$$vec(\widehat{\beta}(z;H)) = \left(\widetilde{Z}^\top W_H(z) \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W_H(z) Y. \tag{1.160}$$

Furthermore, under some standard smoothness conditions on the regression, some moment conditions on the errors; i.e., $E|v_{it}|^{2+\delta}$, and as $N \to \infty$, $\|H\| \to 0$ in such way that $N|H| \to \infty$, in Sun et al. (2009, Theorem 3.2) it is shown the following asymptotic distribution of this estimator,

$$\sqrt{N|H|}\left(\widehat{m}_{PLS}(z;H) - m(z) - \frac{1}{2}\Psi(z)^{-1}\Lambda(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Sigma(z)\right), \tag{1.161}$$

where

$$
\begin{aligned}
\Lambda(z) &= |H|^{-1}\sum_{t=1}^{T} E\left((1-\varpi_{it})\lambda_{it}X_{it}X_{it}^{\top} r_H(\widetilde{Z}_{it}, z)\right),\\
\Sigma(z) &= \sigma_v^2 \lim_{N\to\infty} \Psi(z)^{-1}\Gamma(z)\Psi(z)^{-1},
\end{aligned}
$$

and $\lambda_{it} = K_H(Z_{it} - z)$ and $\varpi = \lambda_{it}/\sum_{t=1}^{T}\lambda_{it} \in (0,1)$. Also, let $r_H(\cdot)$ be the second-order term of the Taylor expansion of $m(\cdot)$, $\Psi(z)$ is a nonsingular matrix of the form $\Psi(z) = |H|^{-1}\sum_{t=1}^{T} E\left((1-\varpi_{it})\lambda_{it}X_{it}X_{it}^{\top}\right)$ and $\Gamma(z) = |H|^{-1}\sum_{t=1}^{T} E\left((1-\varpi_{it})^2\lambda_{it}^2 X_{it}X_{it}^{\top}\right)$.

Thus, it is demonstrated that via a one-step procedure that eliminates the fixed effects through kernel-based weights we can obtain a nonparametric estimate of the unknown function that is asymptotically normally distributed.

1.4.3 Endogeneity

In this subsection, we consider a panel data varying coefficient model of the form

$$
Y_{it} = X_{it}^{\top} m(Z_{it}) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{1.162}
$$

where X_{it} is a $d \times 1$ vector whose first element is $X_{1it} = 1$, $m(\cdot)$ is an unspecific smooth function and Z_{it} is a $q \times 1$ vector. Also, some or all components of X_{it} may be correlated with the disturbance ϵ_{it} and we assume $E(\epsilon_{it}|Z_{it}) = 0$ and $E(\epsilon_{it}|X_{it}) \neq 0$. Note that if X_{it} and Z_{it} are exogenous variables, we would be in the static varying coefficient model analyzed in Zhou et al. (2010) for panel data models or in Das (2005), Cai et al. (2006) and Cai and Xiong (2012) in the cross-sectional setup.

As the reader can observe, when $E(\epsilon_{it}|X_{it}) \neq 0$ we obtain $E(Y|Z_{it}, X_{it}) \neq X_{it}^{\top} m(Z_{it})$ in (1.162), so it is not possible to estimate consistently the coefficient functions by projecting Y_{it} on $X_{it}^{\top} m(Z_{it})$. To overcome this situation, in Cai and Li (2008) it is established that by adapting the proposal in Cai et al. (2006) it is possible to obtain a feasible estimator for $m(\cdot)$ via a two-step nonparametric procedure.

Assuming there is a $q \times 1$ vector of instrumental variables W_{it} whose first component is $W_{1it} = 1$ in such way that $E(\epsilon_{it}|W_{it}) = 0$, the orthogonality condition is $E(\epsilon_{it}|Z_{it}, W_{it}) = 0$.

By multiplying both sides of (1.162) by $\Pi(Z_{it}, W_{it}) = E(X_{it}|Z_{it}, W_{it})$ they obtain

$$E\left(\Pi(Z_{it}, W_{it})Y_{it}|Z_{it} = z\right) = E\left(\Pi(Z_{it}, W_{it})X_{it}^{\top}|Z_{it} = z\right)m(z) \tag{1.163}$$

and by the law of iterated expectations we can write

$$E\left(\Pi(Z_{it}, W_{it})Y_{it}|Z_{it} = z\right) = E\left(\Pi(Z_{it}, W_{it})\Pi(Z_{it}, W_{it})^{\top}|Z_{it} = z\right)m(z). \tag{1.164}$$

Under the assumption that $E\left(\Pi(Z_{it}, W_{it})\Pi(Z_{it}, W_{it})^{\top}|Z_{it} = z\right)$ is positive definite, we obtain

$$m(z) = \left(E\left(\Pi(Z_{it}, W_{it})\Pi(Z_{it}, W_{it})^{\top}|Z_{it} = z\right)\right)^{-1} E\left(\Pi(Z_{it}, W_{it})Y_{it}|Z_{it} = z\right). \tag{1.165}$$

Note that the condition $E\left(\Pi(Z_{it}, W_{it})\Pi(Z_{it}, W_{it})^{\top}|Z_{it} = \cdot\right)$ is positive definite guarantees that $m(\cdot)$ is identified but, since $\Pi(Z_{it}, W_{it})$ is an unknown term in Cai and Li (2008) it is proposed a two-step procedure in order to obtain nonparametric estimators for $m(\cdot)$. In the first-step, they suggest to estimate $E(X_{it}|Z_{it}, W_{it})$ through conventional multivariate nonparametric techniques, and in the second-step, to estimate another conditional mean function of $\widehat{\Pi}(Z_{it}, W_{it})Y_{it}$ conditional on $Z_{it} = z$, where $\widehat{\Pi}(Z_{it}, W_{it})$ is the nonparametric estimator obtained at the first-step.

Nevertheless, note that in a framework where $N \to \infty$ and T is fixed, this two-step procedure greatly complicates the analysis of the main asymptotic properties of this estimator. For this reason, in Cai and Li (2008) it is developed an alternative estimator based on a nonparametric generalized method of moments (NPGMM) that requires only one step.

To avoid the endogeneity problem, in Cai and Li (2008) it is proposed to replace the endogenous covariate X_{it} by an instrumental variable W_{it}. However, it is true that we know the relationship between X_{it} and Z_{it} (i.e., the functional coefficient that we want to estimate) but we do not know how W_{it} relates to Z_{it}. In this case, the usual orthogonality condition to provide nonparametric estimators, i.e., $E(\epsilon_{it}|Z_{it}, W_{it}) = 0$, cannot be used and a generalization of this condition for a vector function $Q(\cdot)$ is needed.

In this way, assuming the model is identified for any $p \times 1$ function vector as $Q(Z_{it}, W_{it})$ it is possible to obtain the conditional moment restrictions from the condition

$$0 = E\left(Q(Z_{it}, W_{it})\epsilon_{it}|Z_{it}, W_{it}\right) = E\left(Q(Z_{it}, W_{it})(Y_{it} - X_{it}^{\top}m(Z_{it}))|Z_{it}, W_{it}\right). \tag{1.166}$$

In this context, in Cai and Li (2008) it is proposed to combine the restrictions contained in (1.166) with the conventional local linear regression approach to provide a nonparametric generalized method of moments for the functional coefficients. Assuming $m(\cdot)$ is a twice continuous differentiable function, we obtain the following locally weighted orthogonality conditions

$$\sum_{i=1}^{N}\sum_{t=1}^{T} Q(Z_{it}, W_{it})(Y_{it} - \widetilde{X}_{it}^{\top}\alpha)K_h(Z_{it} - z) = 0, \tag{1.167}$$

where $\widetilde{X}_{it} = \begin{pmatrix} X_{it}^{\top} & X_{it}^{\top} \otimes (Z_{it} - z)^{\top} \end{pmatrix}^{\top}$ and $\alpha = \begin{pmatrix} \alpha_0^{\top} & \alpha_1^{\top} \end{pmatrix}^{\top}$ are vectors of $d(1+q) \times 1$ dimension. Motivated by this local linear fitting, $Q(Z_{it}, W_{it})$ can be chosen such as $Q(Z_{it}, W_{it}) = \begin{pmatrix} W_{it}^{\top} & W_{it}^{\top} \otimes (Z_{it} - z)^{\top}/h \end{pmatrix}^{\top}$ and to ensure that we obtain an unique solution in (1.167), they also impose that the dimension of this $Q(\cdot)$ function must meet $p \geq d(1+q)$.

Then, we denote by S a $d(1+q) \times p$ matrix of the form

$$S = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Q(Z_{it}, W_{it})\widetilde{Z}_{it}^{\top} K_h(Z_{it} - z) \tag{1.168}$$

and by multiplying (1.167) by S we obtain that it is possible to obtain consistent estimates for the quantities of interest using the following conditional moment restrictions

$$\sum_{i=1}^{N}\sum_{t=1}^{T} S_{it}^{\top} Q(Z_{it}, W_{it})(Y_{it} - \widetilde{X}_{it}^{\top}\alpha)K_h(Z_{it} - z) = 0. \tag{1.169}$$

Let $\widehat{\alpha}$ be the minimizer of (1.169), the solution to this problem in matrix form is

$$\widehat{\alpha}_{NPGMM} = \left(S^{\top} S\right)^{-1} S^{\top}\Upsilon, \tag{1.170}$$

where $\Upsilon = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Q(Z_{it}, W_{it})K_h(Z_{it} - z)Y_{it}$. In this way, $\widehat{\alpha}$ provides the nonparametric generalized method of moments (NPGMM) estimator both for the function $m(z)$ and its first-order derivatives, $D_m(z)$.

Finally, establishing some regularity and the strong mixing conditions, in Cai and Li (2008, Theorem 2, pp. 1331) it is shown the following asymptotic normality result in a context where $N \to \infty$ and T is fixed,

$$\sqrt{NTh^q}\left(H\left(\widehat{\alpha}_{NPGMM} - \alpha\right) - \frac{h^2}{2}\begin{pmatrix} b(z) \\ 0 \end{pmatrix} + o_p(h^2)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, f^{-1}(z)\Sigma\right), \tag{1.171}$$

where $H = diag(I_d, hI_{dq})$ is a diagonal $d(1+q) \times d(1+q)$ matrix, let I be an identity matrix, and

$$\begin{aligned} b(z) &= (tr(\mathcal{H}_m(z)\mu_2(K)))_{d\times 1}, \\ \Sigma &= diag\left(R(K)\Omega_m \quad \Omega_m \otimes (\mu_2^{-1}(K)\mu_2(K^2)\mu_2^{-1}(K))\right), \\ \Omega_m &= \left(\Omega^\top\Omega\right)^{-1}\Omega^\top\Omega_1\Omega\left(\Omega^\top\Omega\right)^{-1}, \end{aligned}$$

for $\Omega = E\left[W_{it}X_{it}^\top|Z_{it} = z\right]$ and $\Omega_1 = E\left[W_{it}\epsilon_{it}|Z_{it} = z\right]$.

Furthermore, in Tran and Tsiomas (2010) a two-step procedure based on the local generalized method of moments (LGMM) is proposed, for which they use the general weighting matrix instead of the identity matrix. In addition, they show that the resulting estimator is asymptotically more efficient than the nonparametric generalized method of moments in Cai and Li (2008). Recently, in Cai et al. (2013) it is considered a new type of dynamic partially linear models with varying coefficient where linearity in some regressors and non-linearity in others is allowed.

Finally, note that given the complexity of providing consistent estimators for the parameters of interest of varying coefficient panel data models with fixed effects and endogenous variables as well as the difficulty of the analysis of the main asymptotic properties, to our knowledge no research has been made to estimate this type of models.

Chapter 2

Direct semi-parametric estimation of fixed effects panel data varying coefficient models

2.1 Introduction

This chapter is concerned with the estimation of varying coefficient panel data models. This type of specification consists of a linear regression model where regression coefficients are assumed to be varying, depending on some exogenous continuous variables proposed by economic theory. For example, in the so-called problem of returns to education, when estimating elasticity of wages to changes in education, it has been pointed out that marginal returns to education might vary with the level of working experience; see Schultz (2003). Therefore, conditionally on a level of education, the elasticity is going to change according with the level of working experience.

Within this context, the issue of potential misspecification of the functional form of the varying coefficients has motivated the use of nonparametric estimation techniques in empirical studies. In most cases, the estimation of the functional form of the coefficients has been performed through standard techniques, such as spline smoothers, series estimators, or local polynomial regression estimators; see Su and Ullah (2011). Although, in most cases, a direct application of the previous techniques renders correct inference results, it is true that not much attention has been paid to the asymptotic behavior of these estimators under non-standard settings. Unfortunately, some of these settings are rather relevant in empirical analysis of panel data models. One clear example is the presence, in the econometric model, of some unobserved explanatory variables that, although not varying along time, can be statistically correlated with other explanatory variables in the model (fixed effects). The presence of a heterogeneity of unknown form that is correlated with some explanatory variables is not an easy problem to solve. In fact, under this type of heterogeneity any estimation technique suffers from the so-called incidental parameters problem; see, e.g., Neyman and Scott (1948).

In order to obtain a consistent estimator of the parameters of interest, one possible solution is to transform the model in order to remove the heterogeneity of unknown form. To be more precise, consider a linear panel data model where the heterogeneity μ_i is arbitrarily correlated with the covariates X_{it} and/or Z_{it}, of dimension d and a, respectively,

$$Y_{it} = X_{it}^{\top} m\left(Z_{it}\right) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{2.1}$$

Furthermore, the function $m(z)$ is unknown and needs to be estimated, and the v_{it} are random errors. It is clear that any attempt to estimate $m(\cdot)$ directly through standard nonparametric estimation techniques will render inconsistent estimators of the underlying curve. The reason for this is that $E\left(\mu_i | X_{it} = x, Z_{it} = z\right) \neq 0$. A standard solution to this problem is to remove μ_i from (2.1) by taking a transformation, and then estimating the unknown curve through the use of a nonparametric smoother. There exist several approaches to remove these effects. The simplest approach is probably to take first differences, i.e.,

$$\Delta Y_{it} = X_{it}^{\top} m\left(Z_{it}\right) - X_{i(t-1)}^{\top} m\left(Z_{i(t-1)}\right) + \Delta v_{it}, \quad i = 2, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{2.2}$$

The direct nonparametric estimation of $m(\cdot)$ has, until now, been considered as rather cumbersome; see Su and Ullah (2011). The reason for this is that, for each i, the conditional expectation $E\left(\Delta Y_{it} \middle| Z_{it}, Z_{i(t-1)}, X_{it}, X_{i(t-1)}\right)$ in (2.2) contains a linear combination of $X_{it}^{\top} m\left(Z_{it}\right)$ for different t. This can be considered as an additive function with the same functional form at a different times.

In some special cases, a consistent estimation of the quantities of interest has been provided in the literature. For the unrestricted model $X_{it}^{\top} m(Z_{it}) \equiv m(X_{it}, Z_{it})$, (2.2) becomes a fully nonparametric additive model

$$\Delta Y_{it} = m(X_{it}, Z_{it}) - m\left(X_{i(t-1)}, Z_{i(t-1)}\right) + \Delta v_{it}, \quad i = 2, \cdots, N \quad ; \quad t = 2, \cdots, T.$$

In this case, in Henderson et al. (2008) it is proposed an iterative procedure based in a profile likelihood approach, whereas in Mammen et al. (2009) it is considered the consistent estimation of nonparametric additive panel data models with both time and individual effects via a smoothed backfitting algorithm. Furthermore, for $X_{it}^{\top} m(Z_{it}) \equiv g(Z_{it}) + \widetilde{X}_{it}^{\top}\beta_0$, where $X_{it} = (1, \widetilde{X}_{it}^{\top})^{\top}$ and $m(Z_{it}) = (g(Z_{it}), \beta_0)^{\top}$ for some real valued $g(\cdot)$ and a vector β_0, the regression function in (2.2) becomes a semi-parametric partially additive model

$$\Delta Y_{it} = \beta_0^{\top} \Delta \widetilde{X}_{it} + g(Z_{it}) - g\left(Z_{i(t-1)}\right) + \Delta v_{it}, \quad i = 2, \cdots, N \quad ; \quad t = 2, \cdots, T.$$

Qian and Wang (2012) consider the marginal integration estimator of the nonparametric additive component resulting from the first differencing step, i.e., $G\left(Z_{it}, Z_{i(t-1)}\right) = g(Z_{it}) - g\left(Z_{i(t-1)}\right)$.

The estimation procedure that we introduce in this chapter mainly generalizes the previous results to a rather general varying coefficient model as the one specified in (2.1) in a framework where $N \to \infty$ but T remains fixed. It is based in applying a local approximation to the additive function $X_{it}^{\top} m(Z_{it}) - X_{i(t-1)}^{\top} m\left(Z_{i(t-1)}\right)$. The same idea was proposed in a completely different context in Yang (2002). Because the estimator is based in local approximation properties, we investigate the behavior of the bias remainder term under fairly general conditions. This term, which is negligible in standard local linear regression techniques (see Fan and Gijbels (1995b)), requires much more attention when dealing with first difference estimators. In fact, as it has been already pointed out in Lee and Mukherjee (2008), the direct application of local linear regression techniques to first differencing transformations in panel data models renders to biased estimators and the bias does not degenerate, even with large samples. Using a higher-dimensional kernel weight, our estimation technique overcomes the problem of non-vanishing bias, although, as expected, the variance term becomes larger. The same phenomenon also appears in Henderson et al. (2008), where their final estimator already shows an even larger variance.

In order to obtain the standard rates of convergence for this type of problems (i.e., to reduce the variance holding the bias constant), we propose to use the developments introduced in Fan and Zhang (1999). Their core idea was that the variance can be reduced by further smoothing, but bias cannot be reduced by any kind of smoothing. We apply these ideas to our problem by using a one-step backfitting algorithm. Because it has the form of an additive model, we also show that our estimator is oracle efficient; that is, the variance-covariance matrix of any of the components of our estimator is the same asymptotically as if we would know the other component. Finally, we also propose a data-driven method to select the bandwidth parameter.

As already pointed out previously, to remove the heterogeneous effects, other transformations are available in the literature. To our knowledge, for model (2.1), an estimation of $m(\cdot)$ has been proposed in Sun et al. (2009), through the use of the so-called least-squares dummy variable approach. They estimate $m(\cdot)$ using the following alternative specification

$$Y_{it} = X_{it}^{\top} m\left(Z_{it}\right) + \sum_{j=1}^{n} \mu_i d_{ij} + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{2.3}$$

where $d_{ij} = 1 \quad \text{if} \quad i = j$ and 0 otherwise. Based on this model, they propose a least-squares method combined with a local linear regression approach that produces a consistent estimator of the unknown smoothing coefficient curves. Compared to our method, their estimator exhibits a larger bias. In fact, their bias presents two terms. The first term results from the local approximation of $m(\cdot)$ and it is also present in our estimator. The second term results from the unknown fixed effects and it is zero only in the case that they add the additional (strong) restriction that $\sum_i \mu_i = 0$. This type of restrictions is also used in Mammen et al. (2009).

The rest of the chapter is organized as follows. In Section 2.2, we set up the model and the estimation procedure. In Section 2.3, we study its asymptotic properties and we propose a transformation procedure which provides an estimator that is oracle efficient and achieves optimal rates of convergence. Section 2.4, we show how to estimate the bandwidth matrix empirically and, in Section 2.5, we present some simulation results. Finally, we conclude in Section 2.6. The proofs of the main results are collected in the Appendix 1.

2.2 Statistical model and estimation procedure

To illustrate our technique, we start with the univariate case and then we present our results for the multivariate case. Then, we consider (2.2) with $d = q = 1$. In this case, for any $z \in \mathcal{A}$, where $\mathcal{A}$ is a compact subsect in a non-empty interior of $I\!R$, we have the following Taylor expansion

$$\begin{aligned} X_{it}m\left(Z_{it}\right) - X_{it-1}m\left(Z_{it-1}\right) &\approx m\left(z\right)\Delta X_{it} + m'\left(z\right)\left(X_{it}\left(Z_{it}-z\right) - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)\right) \\ &+ \frac{1}{2}m''\left(z\right)\left(X_{it}\left(Z_{it}-z\right)^2 - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^2\right) \\ &+ \cdots + \frac{1}{p!}m^{(p)}\left(z\right)\left(X_{it}\left(Z_{it}-z\right)^p - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^p\right) \\ &\equiv \sum_{\lambda=0}^{p}\beta_\lambda\left(X_{it}\left(Z_{it}-z\right)^\lambda - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^\lambda\right). \end{aligned}$$

This suggests that we estimate $m\left(z\right)$, $m'\left(z\right)$, $\cdots$, $m^{(p)}\left(z\right)$ by regressing ΔY_{it} on the terms $X_{it}\left(Z_{it}-z\right)^\lambda - X_{it-1}\left(Z_{it-1}-z\right)^\lambda$, $\lambda = 0, 1, \cdots$, with kernel weights. Then, the quantities of interest can be estimated using a local linear regression estimator,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it} - \alpha\Delta X_{it} - \delta\left[X_{it}\left(Z_{it}-z\right) - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)\right]\right)^2 K_h\left(Z_{it}-z\right)K_h\left(Z_{i(t-1)}-z\right), \tag{2.4}$$

see Fan and Gijbels (1995b), Ruppert and Wand (1994), or Zhan-Qian (1996). Here, K is a bivariate kernel such that $K(u,v) = K(u)K(v)$, where for each u, v,

$$\int K(u)du = 1 \qquad \text{and} \qquad K_h(u) = \frac{1}{h}K\left(u/h\right),$$

and h is a bandwidth.

We denote by $\widehat{\beta}_0$ and $\widehat{\beta}_1$ the minimizers of (2.4). The above exposition suggest as estimators for $\alpha = m(\cdot)$ and $\delta = m'(\cdot)$, $\widehat{m}_h(z) = \widehat{\beta}_0$ and $\widehat{m}'_h(z) = \widehat{\beta}_1$, respectively. In particular, for the case of a local constant approximation ($p = 0$; i.e., Nadaraya-Watson kernel regression estimator), the estimator for $m(z)$ has the following closed form:

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}K_h(Z_{it}-z)K_h(Z_{i(t-1)}-z)\Delta X_{it}\Delta Y_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}K_h(Z_{it}-z)K_h(Z_{i(t-1)}-z)\left(\Delta X_{it}\right)^2}. \tag{2.5}$$

In the local linear regression case ($p = 1$), we have

$$\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \left(\sum_{it} K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z) \widetilde{Z}_{it} \widetilde{Z}_{it}^\top \right)^{-1} \sum_{it} K_h(Z_{it}-z) K_h(Z_{i(t-1)}-z) \widetilde{Z}_{it} \Delta Y_{it}, \tag{2.6}$$

where $\widetilde{Z}_{it}$ is a 2×1 vector such that

$$\widetilde{Z}_{it}^\top = \left(\Delta X_{it} \quad X_{it} \left(Z_{it} - z \right) - X_{i(t-1)} \left(Z_{i(t-1)} - z \right) \right). \tag{2.7}$$

Note that in (2.4) we propose a bivariate kernel that also contains $Z_{i(t-1)}$, instead of considering only Z_{it}. The reason for this is that, if we consider only a kernel around Z_{it}, the transformed regression equation (2.2) would be originally localized around Z_{it} without considering all other values. Consequently, the distance between Z_{is} (for $s \neq t$) and z cannot be controlled by the fixed bandwidth parameter and thus the transformed remainder terms cannot be negligible. The consequence of all this would be a non-degenerated bias in this type of local linear estimator, which is removed by considering a local approximation around the pair $\left(Z_{it}, Z_{i(t-1)} \right)$. In Theorem 2.1, we show that the local linear estimator with the bivariate kernel shows the same rate as standard local linear smothers estimators (i.e., with a bias of order $O(h^2)$). Unfortunately, the well-known trade-off between bias and variance term appears and, although the introduction of this bivariate kernel drops the bias out, it enlarges the variance term, which becomes of order $O(1/NTh^2)$. This is also emphasized in Theorem 2.1. Of course, this affects the achievable rate of convergence for this type of problems, which slows down at the rate $\sqrt{NT}h$.

In order to recover the desirable rate of $\sqrt{NTh}$, we propose a transformation that is basically a one-step backfitting algorithm. Let us denote by ΔY_{it}^b the following expression

$$\Delta Y_{it}^b = \Delta Y_{it} + m(Z_{i(t-1)}) X_{i(t-1)}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{2.8}$$

By substituting (2.2) into (2.8), we obtain

$$\Delta Y_{it}^b = m(Z_{it}) X_{it} + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{2.9}$$

As it can be realized from (2.9), the estimation of $m(\cdot)$ is now a one-dimensional problem, and therefore we can use again a local linear least-squares estimation procedure with univariate kernel weights. However, there is still a problem that needs to be solved. In (2.8), the term $m(Z_{i(t-1)})$ is unknown. So, we replace it by the initial local linear regression estimator, i.e., $\Delta\widetilde{Y}_{it}^{b} = \Delta Y_{it} + \widehat{m}_h\left(Z_{i(t-1)}\right)X_{i(t-1)}$, having the following regression model

$$\Delta\widetilde{Y}_{it}^{b} = m(Z_{it})X_{it} + v_{it}^{b}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{2.10}$$

where

$$v_{it}^{b} = \left(\widehat{m}_h\left(Z_{i(t-1)}\right) - m\left(Z_{i(t-1)}\right)\right)X_{i(t-1)} + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T.$$

By doing so, we can estimate $m(\cdot)$ using the following weighted local linear regression

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta\widetilde{Y}_{it}^{b} - \delta_0 X_{it} - \delta_1 X_{it}\left(Z_{it} - z\right)\right)^2 K_{\widetilde{h}}\left(Z_{it} - z\right). \tag{2.11}$$

Let $\widetilde{\alpha}_0$ and $\widetilde{\alpha}_1$ be the minimizers of (2.11). Then, as before, we propose as estimators for $m(\cdot)$ and $m'(\cdot)$, $\widetilde{m}_{\widetilde{h}}(z) = \widetilde{\gamma}_0$ and $\widetilde{m}'_{\widetilde{h}}(z) = \widetilde{\gamma}_1$, respectively.

Now, once our estimation procedure has been fully explained for the univariate case, we proceed to extend our results for the multivariate case, that is, for $d \neq q \neq 1$ in (2.1). In this case, the quantities of interest can be estimated using a multivariate locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it} - \widetilde{Z}_{it}^{\top}\beta\right)^2 K_H\left(Z_{it} - z\right)K_H\left(Z_{i(t-1)} - z\right), \tag{2.12}$$

where we denote by

$$\widetilde{Z}_{it}^{\top} = \begin{pmatrix}\Delta X_{it}^{\top} & X_{it}^{\top}\otimes\left(Z_{it} - z\right)^{\top} - X_{i(t-1)}^{\top}\otimes\left(Z_{i(t-1)} - z\right)^{\top}\end{pmatrix}$$

a $1 \times d(1+q)$ vector. Now, K is a q-variate kernel such that

$$\int K(u)du = 1 \qquad \text{and} \qquad K_H(u) = \frac{1}{|H|^{1/2}}K\left(H^{-1/2}u\right),$$

where H is a $q \times q$ symmetric positive definite bandwidth matrix.

Finally, we denote by $\widehat{\beta} = (\widehat{\beta}_0^{\top} \quad \widehat{\beta}_1^{\top})^{\top}$ a $d(1+q)$-vector that minimizes (2.12). Again, the above exposition suggests as estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}(z;H) = \widehat{\beta}_0$

and $vec(\widehat{D}_m(z;H)) = \widehat{\beta}_1$, respectively. Here, $D_m(z)$ is a $d \times q$ matrix of partial derivatives of the d-function $m(z)$ with respect to the elements of the $q \times 1$ vector z.

It is easy to verify that the solution to the minimization problem in (2.12) can be written in matrix form as

$$\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \Delta Y, \tag{2.13}$$

where

$$\begin{aligned} W &= diag\left(K_H(Z_{12}-z)K_H(Z_{11}-z), ..., K_H(Z_{NT}-z)K_H(Z_{N(T-1)}-z)\right), \\ \Delta Y &= (\Delta Y_{12}, ..., \Delta Y_{NT})^\top, \end{aligned}$$

and

$$\widetilde{Z} = \begin{bmatrix} \Delta X_{12}^\top & X_{12}^\top \otimes (Z_{12}-z)^\top - X_{11}^\top \otimes (Z_{11}-z)^\top \\ \vdots & \vdots \\ \Delta X_{NT}^\top & X_{NT}^\top \otimes (Z_{NT}-z)^\top - X_{N(T-1)}^\top \otimes (Z_{N(T-1)}-z)^\top \end{bmatrix}.$$

The local weighted linear least-squares estimator of $m(z)$ is then defined as

$$\widehat{m}(z;H) = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \Delta Y, \tag{2.14}$$

where $e_1 = (I_d \vdots 0_{dq \times d})$ is a $d(1+q) \times d$ selection matrix, I_d is a $d \times d$ identity matrix and $0_{dq \times d}$ a $dq \times d$ matrix of zeros. Note that the dimensions of W and $\widetilde{Z}$ are $N(T-1) \times N(T-1)$ and $N(T-1) \times d(1+q)$, respectively.

Finally, there are several reasons to choose local linear least-squares estimators against other candidates. First, the form in (2.14) suggests that this estimator is found by fitting a plane to the data using weighted least-squares. The weights are chosen according to the kernel and the bandwidth matrix H. As has already been discussed in Ruppert and Wand (1994), if a Gaussian kernel with (possibly) compact support is chosen, then the weight given to Z_{it} is the value of the Gaussian density with mean $Z_{it} - z$, which has an ellipsoidal contour of the form $(Z_{it} - z)^\top H^{-1} (Z_{it} - z) = c$, for $c > 0$. Clearly, the further from z that Z_{it} is, the less weight it receives. However, H controls both the size and orientation of the ellipsoids at a given density level and therefore it also controls the amount and direction of the weights. Often, instead of taking a matrix H, we adopt a simpler form $H = \mathrm{diag}(h_1^2, \cdots, h_q^2)$. If we have a diagonal bandwidth matrix, this means that the ellipsoids have their axes in the same direction as the coordinate axes, whereas for a general H matrix they will correspond to the eigenvectors of H. Depending on the shape of $m(\cdot)$, there are situations where having a full bandwidth matrix is advantageous. Another important advantage of local linear least-squares kernel estimators is that the asymptotic bias and variance expressions are particularly appealing and appear to be superior to those of the Nadaraya-Watson or other nonparametric estimators. In particular, in Fan (1993) it is shown that the local linear least-squares estimator has an important asymptotic minimax property. Furthermore, unlike the Nadaraya-Watson or other nonparametric estimators, the bias and variance of (2.14) near the boundary of the density of Z are of the same order of magnitude as in the interior. That is a very interesting property because, in applications, the boundary region might comprise a large proportion of the data.

2.3 Asymptotic properties and the oracle efficient estimator

In this section, we investigate some preliminary asymptotic properties of our estimator. In order to do so, we need the following assumptions.

Assumption 2.1 *Let $(Y_{it}, X_{it}, Z_{it})_{i=1,\cdots,N;t=1,\cdots,T}$ be a set of independent and identically distributed (i.i.d.) $\mathbb{R}^{d+q+1}$-random variables in the subscript i for each fixed t and strictly stationary over t for fixed i. They are a sample following (2.1). Furthermore, let $f_{Z_{1t}}(\cdot)$, $f_{Z_{1t},Z_{1(t-1)}}(\cdot,\cdot)$, $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(\cdot,\cdot,\cdot)$ be the probability density functions of Z_{1t}, $\left(Z_{1t}, Z_{1(t-1)}\right)$ and $\left(Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)}\right)$, respectively. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.*

Assumption 2.2 *The random errors v_{it} are independent and identically distributed, with zero mean and homoscedastic variance, $\sigma_v^2 < \infty$. They are also independent of X_{it} and Z_{it} for all i and t. Furthermore, $E\left|v_{it}\right|^{2+\delta} < \infty$, for some $\delta > 0$.*

Assumption 2.3 *μ_i can be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.*

Assumption 2.4 *Let $\|A\| = \sqrt{tr\left(A^\top A\right)}$, then $E\left[\|X_{it}X_{it}^\top\|^2 \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ is bounded and uniformly continuous in its support. Furthermore, let*

$$\mathcal{X}_{it} = \begin{pmatrix} X_{it}^\top & X_{i(t-1)}^\top \end{pmatrix}^\top \qquad \text{and} \qquad \Delta\mathcal{X}_{it} = \begin{pmatrix} \Delta X_{it}^\top & \Delta X_{i(t-1)}^\top \end{pmatrix}^\top .$$

The matrix functions $E\left[\mathcal{X}_{it}\mathcal{X}_{it}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, $E\left[\Delta\mathcal{X}_{it}\Delta\mathcal{X}_{it}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, $E\left[\mathcal{X}_{it}\mathcal{X}_{it}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-3)} = z_3\right]$, and $E\left[\mathcal{X}_{it}\Delta\mathcal{X}_{it}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-2)} = z_3\right]$ are bounded and uniformly continuous in their support.

Assumption 2.5 *The function $E\left[\Delta X_{it}\Delta X_{it}^\top \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ is positive definite for any interior point of (z_1, z_2) in the support of $f_{Z_{it}, Z_{i(t-1)}}(z_1, z_2)$.*

Assumption 2.6 *The following functions $E\left[|X_{it}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, $E\left[|X_{i(t-1)}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, and $E\left[|\Delta X_{it}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ are bounded and uniformly continuous in any point of their support, for some $\delta > 0$.*

Assumption 2.7 *Let z be an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $m_1(\cdot), m_2(\cdot), \cdots, m_d(\cdot)$ are bounded and uniformly continuous.*

Assumption 2.8 *The Kernel functions K are compactly supported, bounded kernel such that $\int uu^\top K(u)du = \mu_2(K)I$ and $\int K^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars and I is the $q \times q$ identity matrix. In addition, all odd-order moments of K vanish, that is $\int u_1^{\iota_1} \cdots u_q^{\iota_q} K(u)du = 0$, for all nonnegative integers $\iota_1, \cdots, \iota_q$ such that their sum is odd.*

Assumption 2.9 *The bandwidth matrix H is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|H| \to \infty$.*

As can be seen, all assumptions are rather standard in the nonparametric regression analysis of panel data models. Assumption **2.1** establishes standard features about the sample and data-generating process. The individuals are independent and, for a fixed individual, we allow for correlation over time. Also, other possible time-series structures might be considered, such as strong mixing conditions (see Cai and Li (2008)) or non-stationary time-series data (see Cai et al. (2009)). Mixing conditions are usually taken into account to make the covariances of the estimator tend to zero at a faster rate. In our case, this is not necessary because our asymptotic analysis is performed for fixed T. However, we believe that non-stationary processes are beyond the scope of this chapter. Note also that marginal densities are assumed to be bounded from above and below. This assumption can be relaxed at the price of increasing the mathematical complexity of the proofs.

Assumption **2.1** is also standard for first-difference estimators; see Wooldridge (2003) for the fully parametric case. Furthermore, independence between the v_{it} errors and the X_{it} and/or Z_{it} covariates is assumed without loss of generality. We could relax this assumption by assuming some dependence based on second-order moments. For example, heteroskedasticity of unknown form can be allowed and, in fact, under more complex structures in the variance-covariance matrix, a transformation of the estimator proposed in You et al. (2010) can be developed in our setting. This type of assumption also rules out the existence of endogenous explanatory variables and imposes strict exogeneity conditions. If this were the case, then an instrumental variable approach, such as the one proposed in Cai and Li (2008) or Cai and Xiong (2012), would be needed. Assumption **2.3** imposes the so-called fixed effects. Note that this assumption is much weaker than the one introduced in Sun et al. (2009) so that their least-squares dummy variable approach can work. Basically, they impose a smooth relationship between heterogeneity and explanatory variables, and in order to avoid an additional bias term, they need $\sum_i \mu_i = 0$.

Assumptions **2.4** and **2.5** are some smoothness conditions on moment functionals. Assumption **2.6** is the equivalent to a standard rank condition for identification of this type of models. Assumptions **2.7**-**2.9** are standard in local linear regression estimators; see Ruppert and Wand (1994). Finally, all our results hold straightforwardly for the random coefficient setting.

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT})$ and $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the observed covariates sample vectors. Under these assumptions, we now establish some results on the conditional mean and the conditional variance of the local linear least-squares estimator.

Theorem 2.1 *Let Assumptions 2.1-2.9 hold. Then, for T fixed and $N \to \infty$,*

$$\begin{aligned}
&E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) \\
&= \frac{1}{2} e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \left(S_{m_1}(z) - S_{m_2}(z)\right) \\
&= \frac{1}{2} \mathcal{B}_{\Delta X \Delta X}^{-1}(z,z)(\mu_2(K_u)\mathcal{B}_{\Delta XX}(z,z) - \mu_2(K_v)\mathcal{B}_{\Delta XX_{-1}}(z,z)) \\
&\quad \times diag_d(tr(\mathcal{H}_{m_r}(z)H))i_d + o_p(tr(H)),
\end{aligned}$$

where for $r = 1, \cdots, d$, $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the rth component of $m(\cdot)$, while for $\ell = 1, 2$ the ith element of S_{m_ℓ} is

$$\left(X_{i(t+1-\ell)} \otimes \left(Z_{i(t+1-\ell)} - z\right)\right)^\top \mathcal{H}_m(z) \left(Z_{i(t+1-\ell)} - z\right).$$

Furthermore, if $\mu_2(K_u) = \mu_2(K_v)$ the bias term becomes

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}\mu_2(K_u)\, diag_d(tr(\mathcal{H}_{m_r}(z)H))i_d + o_p(tr(H)).$$

The variance is

$$Var\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = \frac{2\sigma_v^2 R(K_u) R(K_v)}{NT|H|} \mathcal{B}_{\Delta X \Delta X}^{-1}(z,z)(1 + o_p(1)),$$

where

$$\begin{aligned}
\mathcal{B}_{\Delta XX}(z,z) &= E\left[\Delta X_{it} X_{it}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z), \\
\mathcal{B}_{\Delta XX_{-1}}(z,z) &= E\left[\Delta X_{it} X_{i(t-1)}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z), \\
\mathcal{B}_{\Delta X \Delta X}(z,z) &= E\left[\Delta X_{it} \Delta X_{it}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z),
\end{aligned}$$

$diag_d(tr(\mathcal{H}_{m_r}(z)H))$ stands for a diagonal matrix of elements $tr(\mathcal{H}_{m_r}(z)H)$, for $r = 1, \cdots, d$, and i_d is a $d \times 1$ unit vector.

To illustrate the asymptotic behavior of our estimator, we give a result for the case when $d = q = 1$ and $H = h^2 I$. In this case, the above result can be written as follows.

Corollary 2.1 *Let Assumptions 2.1-2.8 hold. Then, if $h \to 0$ in such a way that $Nh^2 \to \infty$ as N tends to infinity and T is fixed,*

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}c(z,z)m^{''}(z)\,h^2 + o_p\left(h^2\right),$$

where

$$c(z,z) = \frac{\mu_2(K_u)\,E\left[\Delta X_{it}X_{it}|\,Z_{it}=z, Z_{i(t-1)}=z\right] - \mu_2(K_v)\,E\left[\Delta X_{it}X_{i(t-1)}\right|Z_{it}=z, Z_{i(t-1)}=z\right]}{E\left[\left(X_{it}-X_{i(t-1)}\right)^2|Z_{it}=z, Z_{i(t-1)}=z\right]}.$$

Furthermore, if $\mu_2(K_u) = \mu_2(K_v)$, then the bias term has the following expression

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}m^{''}(z)\,h^2 + o_p\left(h^2\right).$$

The variance is

$$Var\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = \frac{2\sigma_v^2 R(K_u)\,R(K_v)}{NTh^2 f_{Z_{it},Z_{i(t-1)}}(z,z)\,E\left[\left(X_{it}-X_{i(t-1)}\right)^2|Z_{it}=z, Z_{i(t-1)}=z\right]}(1+o_p(1)).$$

Note that, in the standard case, $\mu_2(K_u) = \mu_2(K_v)$ and thus we obtain a good result for the bias. In fact, the resulting asymptotic bias has the same expression as in the standard local linear estimator.

As has already been pointed out in other works, the leading terms in both bias and variance do not depend on the sample, and therefore we can consider such terms as playing the role of the unconditional bias and variance. Furthermore, we believe that the conditions established on H are sufficient to show that the other terms are $o_p(1)$ and therefore it is possible to show the following result for the asymptotic distribution of $\widehat{m}(z;H)$.

Theorem 2.2 *Let Assumptions 2.1-2.9 hold. Then, for T fixed and $N \to \infty$,*

$$\sqrt{NT\,|H|}\left(\widehat{m}(z;H) - m(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b(z), \upsilon(z)\right),$$

where

$$\begin{aligned} b(z) &= \frac{1}{2}\mu_2(K_u)\,diag_d\left(tr\left(\mathcal{H}_{m_r}(z)\,H\sqrt{NT\,|H|}\right)\right)\imath_d, \\ \upsilon(z) &= 2\sigma_v^2 R(K_u)\,R(K_v)\,\mathcal{B}^{-1}_{\Delta X\Delta X}(z,z). \end{aligned}$$

Note that the rate at which our estimator converges is $NT\,|H|$. Under the conditions established in the propositions, our estimator is both consistent and asymptotically normal. However, its rate of convergence is suboptimal because the lower rate of convergence for this type of estimators is $NT\,|H|^{1/2}$. As we have already indicated in Section 2.2, in order to achieve optimality we propose to reduce the dimensionality of the problem by redefining ΔY_{it} as in (2.10). Now for the multivariate case,

$$\Delta\widetilde{Y}_{it}^{b} \;=\; X_{it}^{\top} m\left(Z_{it}\right) + v_{it}^{b}, \qquad i = 1, \cdots, N \qquad ; \qquad t = 2, \cdots, T,$$

where

$$v_{it}^{b} = X_{i(t-1)}^{\top}\left(\widehat{m}(Z_{i(t-1)}; H) - m\left(Z_{i(t-1)}\right)\right) + \Delta v_{it}. \tag{2.15}$$

In expression (2.15), $\widehat{m}(Z_{i(t-1)}; H)$ is the first-step local linear estimator obtained in (2.14). Now, we propose to estimate $m\left(Z_{it}\right)$ using a multivariate locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta\widetilde{Y}_{it}^{b} - \left(X_{it}^{\top}\gamma_0 + X_{it}^{\top}\otimes\left(Z_{it}-z\right)^{\top}\gamma_1\right)\right)^{2} K_{\widetilde{H}}(Z_{it}-z), \tag{2.16}$$

where $\widetilde{H}$ is a $q\times q$ symmetric positive-definite bandwidth matrix.

If we define $\widetilde{Z}_{it}^{b\top} = \left(X_{it}^{\top} \quad X_{it}^{\top}\otimes\left(Z_{it}-z\right)^{\top}\right)$ as a $1\times d(1+q)$ vector, (2.16) can be written as

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it}^{b} - \widetilde{Z}_{it}^{b\top}\beta\right)^{2} K_{\widetilde{H}}(Z_{it}-z), \tag{2.17}$$

where we denote by $\widetilde{\gamma} = (\widetilde{\gamma}_0^{\top} \quad \widetilde{\gamma}_1^{\top})^{\top}$ the $d(1+q)$ vector that minimizes (2.17).

Following the same reasoning as before, we can write

$$\widetilde{m}(z;\widetilde{H}) = \widetilde{\gamma}_0 = e_1^{\top}\left(\widetilde{Z}^{b\top}W^{b}\widetilde{Z}^{b}\right)^{-1}\widetilde{Z}^{b\top}W^{b}\Delta\widetilde{Y}^{b}, \tag{2.18}$$

where $\Delta\widetilde{Y}^{b} = (\Delta\widetilde{Y}_{12}^{b}, \cdots, \Delta\widetilde{Y}_{NT}^{b})^{\top}$, $W^{b} = diag\left(K_{H}(Z_{12}-z), \cdots, K_{H}(Z_{NT}-z)\right)$ and

$$\widetilde{Z}^{b} = \begin{bmatrix} X_{12}^{\top} & X_{12}^{\top}\otimes\left(Z_{12}-z\right)^{\top} \\ \vdots & \vdots \\ X_{NT}^{\top} & X_{NT}^{\top}\otimes\left(Z_{NT}-z\right)^{\top} \end{bmatrix}.$$

In order to show the asymptotic properties of this estimator, we need to assume the following about the bandwidth $\widetilde{H}$ and its relationship with H.

Assumption 2.10 *The bandwidth matrix* $\widetilde{H}$ *is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as* N *tends to infinity in such a way that* $N|\widetilde{H}| \to \infty$.

Assumption 2.11 *The bandwidth matrices* H *and* $\widetilde{H}$ *must fulfill that as* N *tends to infinity,* $N\,|H|\,|\widetilde{H}|/\log(N) \to \infty$ *and* $tr(H)\,/tr(\widetilde{H}) \to 0$.

In general, for the kernel function and conditional moments and densities, we need both the smoothness and boundedness conditions already given in Assumptions **2.1-2.8**. These are required to use uniform convergence results such as the ones established by Masry (1996). It is then possible to show the following result.

Theorem 2.3 *Let Assumptions 2.1-2.8 and 2.10-2.11 hold. Then, for T fixed an* $N \to \infty$,

$$E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}\mu_2\left(K_u\right) diag_d\left(tr(\mathcal{H}_{m_r}\left(z\right)\widetilde{H})\right)\imath_d + o_p(tr(\widetilde{H}))$$

and

$$Var\left(\widetilde{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = \frac{2\sigma_v^2 R\left(K_u\right)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{XX}^{-1}\left(z\right)\left(1+o_p(1)\right),$$

where $diag_d\left(tr(\mathcal{H}_{m_r}\left(z\right)\widetilde{H})\right)$ *stands for a diagonal matrix of elements* $tr(\mathcal{H}_{m_r}\left(z\right)\widetilde{H})$, *for* $r = 1, \cdots, d$, *and* i_d *is a* $d \times 1$ *unit vector.*

Finally, focusing on the relevant terms of bias and variance of Theorems 2.1 and 2.2 and following Ruppert and Wand (1994), it can be highlighted that each entry of $\mathcal{H}_m(z)$ is a measure of the curvature of $m(\cdot)$ at z in a particular direction. Thus, we can intuitively conclude that the bias is increased when there is a higher curvature and more smoothing is well described by this leading bias term. Meanwhile, in terms of the variance, we can conclude that it will be penalized by a higher conditional variance of Y given $Z = z$ and sparser data near z.

2.4 Bandwidth selection

As it is clear from previous sections, the bandwidth matrix H plays a crucial role in the estimation of the unknown quantity $m(\cdot)$. In fact, as we have learned from the asymptotic expressions, when choosing H there exist a trade-off between the bias and the variance of our estimator. Consider the simplest case, $H = h^2 I$. If we choose h very small, then according to Corollary 2.1 the bias of our estimator will be reduced (it is of the order of h^2) but at the price of enlarging the variance (the order of this term is $1/NTh^2$). This trade-off should be solved by choosing a bandwidth matrix H that minimizes the mean square error (MSE), which is the sum of the squared bias and variance. There are many different measures of discrepancy between the estimator $\widehat{m}(\cdot; H)$ and the function $m(\cdot; H)$. A comprehensive discussion of these measures has been given in Härdle (1990, Chapter 5). For the sake of simplicity, and taking into account the data-generating process in (2.1), we propose the following measure of discrepancy,

$$MSE\,(H) = E\left[X^\top\left(\widehat{m}(Z;H) - m(Z)\right)\right]^2.$$

In this MSE, the expectation is taken over $Z_1, \cdots, Z_q; X_1 \cdots, X_d$ and $\widehat{m}(Z;H)$ is the estimator defined in (2.14). Therefore, for our problem, we can define the optimal bandwidth matrix H_{opt} as the solution to the following minimization problem,

$$H_{opt} = \arg\min_H MSE\,(H) = \arg\min_H E\left[X^\top\left(\widehat{m}(Z;H) - m(Z)\right)\right]^2.$$

If $Z_1, \cdots, Z_q$; $X_1, \cdots, X_d$ are random variables that are independent of the observed sample $\mathbb{D} = (X_{11}, Z_{11}, \cdots, X_{NT}, Z_{NT})^T$, but they share the same distribution with (X_{11}, Z_{11}), it is straightforward to show that

$$MSE\,(H) = E\left[b^\top(Z)\,\Omega(Z) b\,(Z) + \mathrm{tr}(\Omega\,(Z)\,V(Z))\right], \tag{2.19}$$

where

$$\begin{aligned} b(Z) &= E\left[\widehat{m}(Z;H)|\mathbb{D}, Z\right] - m(Z), \\ V(Z) &= \mathrm{Var}\left(\widehat{m}(Z;H)|\mathbb{D}, Z\right), \quad \text{and} \\ \Omega(Z) &= E\left[XX^\top|Z\right]. \end{aligned}$$

As can be realized from the expression above, we have now formalized the idea of choosing a bandwidth matrix H that minimizes the MSE (i.e., the sum of the squared bias and variance). Note that the way we have defined the measure of discrepancy determines, in our case, the choice of a global bandwidth. That is, we choose a bandwidth that remains constant with the location point. Of course, another possibility would be to choose a bandwidth that varies locally according to this location point (i.e., $H(z)$). In this case, the local MSE criteria would be

$$MSE(z;H) = E\left[X^\top(\widehat{m}(z;H) - m(z))\right]^2,$$

where now the expectation is taken over X. Müller and Stadtmüller (1987) have discussed the issue of local variable bandwidth for convolution-type regression estimators. Furthermore, Fan and Gijbels (1992) have proposed a variable bandwidth for the estimation of local polynomial regression. In our case, we propose to choose a global bandwidth. The reason is twofold. First, all components in our model have been assumed to have the same degree of smoothness. Second, the use of local bandwidths, except for the case where the curve presents a rather complicated structure, increases the computational burden without much improvement in the final results. This is probably because of the local adaptation property that already exhibits local linear regression smoothers.

Unfortunately, the selection of H_{opt} does not solve all problems in bandwidth selection. In fact, as it can be realized, the MSE depends on some unknown quantities and, therefore our optimal bandwidth matrix cannot be estimated from data. There are several alternative solutions to approximate the unknown quantities in the MSE. One alternative is to replace in (2.19) both bias and variance terms by their respective first-order asymptotic expressions that were obtained in Theorem 2.1. This is the so-called 'plug-in' method; for details, see Ruppert et al. (1995)). Another possibility is, as suggested in Fan and Gijbels (1995a), to replace directly in (2.19) bias and variance by their exact expressions. That is

$$\begin{aligned} E\left[\widehat{m}(Z;H)|\mathbb{D},Z\right] - m(Z) &= \left(E[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}] - m(z)\right)|_{z=Z} && (2.20)\\ \mathrm{Var}\left(\widehat{m}(Z;H)|\mathbb{D},Z\right) &= \mathrm{Var}\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right)|_{z=Z}, && (2.21)\end{aligned}$$

where clearly, according to Theorem 2.1

$$\begin{aligned} E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) &= e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\tau && (2.22)\\ \mathrm{Var}\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) &= e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\mathcal{V}W\widetilde{Z}\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}e_1. && (2.23)\end{aligned}$$

Note that τ is a $N(T-1)$ vector such that, for $i = 1, \cdots, N$, $t = 2, \cdots, T$,

$$\begin{aligned}\tau_{it} &= X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)}) \\ &\quad -(X_{it}^\top D_m(z)(Z_{it} - z) - X_{i(t-1)}^\top D_m(z)(Z_{i(t-1)} - z)),\end{aligned}$$

and $\mathcal{V}$ is a $N(T-1) \times N(T-1)$ matrix that contains the V_{ij}'s matrices

$$V_{ij} = E(\Delta v_i \Delta v_j^\top | \mathbb{X}, \mathbb{Z}) = \begin{cases} 2\sigma_v^2, & \text{for} \quad i = j, \quad t = s, \\ -\sigma_v^2, & \text{for} \quad i = j, \quad |t = s| < 2, \\ 0, & \text{for} \quad i = j, \quad |t = s| \geq 2. \end{cases} \tag{2.24}$$

In order to estimate both bias and variance, we need to calculate τ and $\mathcal{V}$. Note that for τ, developing a fifth-order Taylor expansion of both $m(Z_{it})$ and $m(Z_{i(t-1)})$ around z, a local linear polynomial regression of order five would guarantee that the proposed bandwidth selection procedure will be $\sqrt{N}$-consistent for the local linear fit; see Hall et al. (1991) for details. However, for the sake of simplicity, a local cubic polynomial regression would be close to a $\sqrt{N}$-consistent selection rule and it will lead to a reduction in the computational effort. In this case (for $d = q = 1$), the vector $\hat{\tau}$ will contain the (estimated) expressions for the second and third-order derivatives of the local cubic polynomial regression of the terms ΔY_{it} on to $X_{it}(Z_{it} - z)^\lambda - X_{it-1}(Z_{it-1} - z)^\lambda$, $\lambda = 0, 1, \cdots, 3$.

However, in order to estimate $\mathcal{V}$, note that, because of Assumption **2.2**, the estimation of $\mathcal{V}$ is tantamount to the estimation of σ_v^2. To estimate this last quantity, note that under the homoscedastic assumption, we can consistently estimate this by

$$\widehat{\sigma}_v^2 = \frac{1}{2N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta Y_{it} - \Delta X_{it}^\top \widehat{m}^{-i}(Z_{it}; H) + \Delta X_{i(t-1)}^\top \widehat{m}^{-i}(Z_{i(t-1)}; H) \right)^2. \tag{2.25}$$

Note that both $\hat{\tau}$ and $\hat{\sigma}_v^2$ depend on a bandwidth matrix H that needs to be determined from the data. A suitable pilot bandwidth matrix H^*, which can be used for these computations, can be obtained using the global residual squares criterion (RSC) procedure proposed in Fan and Gijbels (1995a). Furthermore, we denote by $\widehat{m}^{-i}(Z_{it}; H)$ the leave-one-out estimator of $m(Z_{it})$. That is, when estimating $m(Z_{it})$ using (2.14), we use all data except those that belong to the ith subject.

Once we have estimated τ and σ_v^2, we can now provide an estimator for $b\left(H\right)$, $V\left(H\right)$, and $\Omega\left(H\right)$. Mainly,

$$\begin{aligned}
\widehat{b}(Z_{it}) &= E\left[\widehat{m}^{-i}(Z_{it};H)|\mathbb{X},\mathbb{Z}\right] - m(Z_{it}) = e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\widehat{\tau},\\
\widehat{V}(Z_{it}) &= \mathrm{Var}\left[\widehat{m}^{-i}(Z_{it};H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\widehat{\mathcal{V}}W\widetilde{Z}\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}e_1,\\
\widehat{\Omega}(Z_{it}) &= \frac{\sum_{j\neq i,t}X_{jt}X_{jt}^\top K_H(Z_{jt}-Z_{it})K_H(Z_{j(t-1)}-Z_{i(t-1)})}{\sum_{j\neq i,t}K_H(Z_{jt}-Z_{it})K_H(Z_{j(t-1)}-Z_{i(t-1)})}.
\end{aligned}$$

The corresponding estimator of the $MSE(H)$, according with (2.19) will be

$$\widehat{MSE}(H) = \frac{1}{N(T-1)}\sum_{it}\left(\widehat{b}^\top\left(Z_{it}\right)\widehat{\Omega}\left(Z_{it}\right)\widehat{b}\left(Z_{it}\right) + \mathrm{tr}\left(\widehat{\Omega}\left(Z_{it}\right)\widehat{V}\left(Z_{it}\right)\right)\right). \tag{2.26}$$

Then, we define the estimator of H_{opt}, $\widehat{H}_{opt}$, as the solution to the following problem,

$$\widehat{H}_{opt} = \arg\min_H \widehat{MSE}\left(H\right).$$

Although we do not provide theoretical properties of this bandwidth, in Zhang and Lee (2000) they have been studied in detail for the local MSE case, and we believe it is straightforward to analyze them for the global MSE case that we present here. Finally, we propose to use the same procedure to estimate the bandwidth matrix H when estimating the oracle efficient estimator.

2.5 Monte Carlo experiment

In this section, we report some Monte Carlo simulation results to examine whether the proposed estimators perform reasonably well in finite samples when μ_i are fixed effects.

We consider the following varying coefficient nonparametric models,

$$Y_{it} = \mu_i + X_{dit}^\top m(Z_{qit}) + v_{it}, \qquad i=1,\cdots,N; \qquad t=1,\cdots,T; \qquad d,q=1,2$$

where X_{dit} and Z_{qit} are scalars random variables, v_{it} is an i.i.d.N(0,1) random variable. The observations follow a data generating process where $Z_{qit} = w_{qit} + w_{qi(t-1)}$ (w_{qit} an i.i.d. uniformly distributed $[0, \Pi/2]$ random variable) and $X_{dit} = 0.5X_{di(t-1)} + \xi_{it}$ (ξ_{it} is i.i.d.N(0,1)).

We consider three different cases of study,

$$
\begin{aligned}
(q=1\ d=1): \ Y_{it} &= X_{1it}m_1(Z_{1it}) + \mu_{1i} + v_{it},\\
(q=2\ d=1): \ Y_{it} &= X_{1it}m_1(Z_{1it}, Z_{2it}) + \mu_{2i} + v_{it},\\
(q=1\ d=2): \ Y_{it} &= X_{1it}m_1(Z_{1it}) + X_{2it}m_2(Z_{2it}) + \mu_{1i} + v_{it},
\end{aligned}
$$

where the chosen functionals form are $m_1(Z_{1it}) = \sin(Z_{1it} * \Pi)$, $m_2(Z_{1it}) = \exp\left(-Z_{1it}^2\right)$, and $m_1(Z_{1it}, Z_{2it}) = \sin\left(\frac{1}{2}(Z_{1it} + Z_{2it}) * \Pi\right)$. We experiment with two specifications for the fixed effects:

(a) μ_{1i} depends on Z_{1it}, where the dependence is imposed by generating $\mu_{1i} = c_0\overline{Z}_{1i.} + u_i$ for $i = 2, \cdots, N$ and $\overline{Z}_{1i.} = T^{-1}\sum_t Z_{1it}$;

(b) μ_{2i} depends on Z_{1it}, Z_{2it} through the generating process $\mu_{2i} = c_0\overline{Z}_{i.} + u_i$ for $i = 2, \cdots, N$ and $\overline{Z}_{i.} = \frac{1}{2}\left(\overline{Z}_{1i.} + \overline{Z}_{2i.}\right)$.

In both cases u_i is an i.i.d.$N(0,1)$ random variable and $c_0 = 0.5$ controls de correlation between the unobservable individual heterogeneity and some of the regressors of the model.

In the experiment, we use 1000 Monte Carlo replications Q. The number of period T is fixed at three, while the number of cross-sections N is varied to be 50, 100 and 200. In addition, the Gaussian kernel has been used and, as in Henderson et al. (2008), the bandwidth is chosen as $\widehat{H} = \widehat{h}I$, and $\widehat{h} = \widehat{\sigma}_z(N(T-1))^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{qit}\}_{i=1,t=2}^{N,T}$.

We report estimation results for both proposed estimators and we use the MSE as a measure of their estimation accuracy. Thus, denoting the φth replication by the subscript φ,

$$
MSE\left(\widehat{m}_l(z;H)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q} E\left[\left(\sum_{r=1}^{d}(\widehat{m}_{\varphi r}(z;H) - m_{\varphi r}(z))X_{it,\varphi r}\right)^2\right]
$$

which can be approximated by the averaged mean squared error (AMSE)

$$
AMSE\left(\widehat{m}(z;H)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{r=1}^{d}(\widehat{m}_{\varphi r}(z;H) - m_{\varphi r}(z))X_{it,\varphi r}\right)^2.
$$

The simulations results are summarized in Tables 2.1-2.3, respectively.

Table 2.1. First-differences estimators. AMSE for $d = 1$ and $q = 1$

		Local polynomial estimator	Backfitting estimator
	N = 50	1.37439	1.27563
T=3	N = 100	1.28384	1.18554
	N = 150	1.23225	1.17689

Table 2.2. First-differences estimators. AMSE for $d = 1$ and $q = 2$

		Local polynomial estimator	Backfitting estimator
	N = 50	5.76503	1.63356
T=3	N = 100	3.25944	1.18840
	N = 150	3.05025	0.99526

Table 2.3. First-differences estimators. AMSE for $d = 2$ and $q = 1$

		Local polynomial estimator	Backfitting estimator
	N = 50	2.08248	1.72158
T=3	N = 100	1.75573	1.49022
	N = 150	1.54960	1.37529

Furthermore, we carried out a simulation study to analyze the behavior in finite samples of the multivariate locally estimator with kernels weights, $\widehat{m}(z; H)$, and the oracle estimator, $\widetilde{m}(z; \widetilde{H})$, as proposed in Sections 2.2 and 2.3. Looking at Tables 2.1-2.3 we can highlight the following.

On one hand, because the proposed estimators are based on a first-difference transformation, the bias and the variance of both estimators do not depend on the values of the fixed effects, so their estimation accuracy are the same for different values of c_0.

On the other hand, from Tables 2.1-2.3, we can see that both estimators show a good performance. For all T, as N increases the AMSEs of both estimators are lower, as expected.

This is because of the asymptotic properties of the estimators described previously. In addition, these results also allow us to test the hypothesis that the oracle estimator generates an improvement in the rate of convergence. Specifically, for the univariate case (Tables 2.1 and 2.3), we appreciate that the achievement of both estimators are quite similar while, on the contrary, in the multivariate case (Table 2.2), the rate of convergence of the oracle estimator is faster than the multivariate locally estimator, as expected. In addition, as we can see in Table 2.2, the results of the local polynomial estimator reflect the curse of dimensionality property, given that as the dimensionality of Z_{it} increases, the AMSE is greater. Thus, the backfitting estimator has an efficiency gain over the local polynomial estimator, as we suspect.

2.6 Conclusions

In this chapter, we introduce a new technique that estimates the varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in an unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters, and we have shown that it is consistent and asymptotically normal. Furthermore, we have shown that it achieves the optimal rate of convergence for this type of problems and it exhibits the so-called oracle efficiency property. Because the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. The Monte Carlo results indicate the good performance of the estimator in finite samples.

Chapter 3

Nonparametric estimation of fixed effects panel data varying coefficient models

3.1 Introduction

Nonparametric estimation of panel data varying coefficient models under fixed effects has been traditionally undertaken through the use of differencing techniques; see Su and Ullah (2011). The main reason is that, direct estimation of varying coefficients through smoothing techniques rends asymptotically biased estimators. This is due to the correlation that exists between the heterogeneity term and the explanatory variables. The differencing approach removes the heterogeneity effect and therefore, it enables us to estimate the function of interest without bias. However, it turns out that the model in differences appears as an additive function with the same functional form at different times. This is why, the proposals to estimate this type of models are closely related to estimation techniques originally designed for additive models. After taking differences with respect to the first observation in time, in Henderson et al. (2008) it is developed an iterative procedure based on a profile likelihood approach. In Mammen et al. (2009) it is proposed a smooth backfitting algorithm, although the specification of the model is slightly different. Recently, in Su and Lu (2013) the unknown function is estimated as a solution of a second order Friedholm integral equation. In their approach they take first differences and they allow for lagged endogenous regressors as explanatory variables. However, these procedures are not very appealing since they are computationally intensive. In view of these results, in Chapter 2 it is presented a direct estimation strategy that is based on a local linear regression for a first-differences model. Afterwards, it is proposed to combine this strategy with a one-step backfitting algorithm. The resulting two-step estimator achieves an optimal rate of convergence and it is shown to be oracle efficient. Unfortunately, due to the first differences, the asymptotic properties of the estimator depend on some strong assumptions on the error term. For example, as in the fully parametric setting it is assumed that the error term exhibits a random walk structure; see Wooldridge (2003) for more details. This is an important drawback because in many situations it is natural to assume either *i.i.d.* or stationary errors.

In this chapter, we present an estimation procedure that uses the deviation from the mean transformation. The advantage of this transformation against others is that, as in the fully parametric setting, we obtain standard asymptotic properties of the nonparametric estimators under *i.i.d.* assumptions on the idiosyncratic error terms. The estimator is based on applying a local approximation on the T additive functions that result from the deviation to the mean transformation, where T is the number of time observations per individual. Note that all asymptotic properties are obtained as N, the number of individuals tends to infinity and keeping T fixed. The use of standard local approximation techniques in this context rends a non-negligible bias in the estimation of the additive components. This is because these techniques approximate the unknown function around a fixed value without considering the sum of the distances between this fixed term and the other values of the sample. In order to cope with this problem, we have to consider a local approximation around the whole vector of time observations for each individual. Unfortunately, the well-known trade-off between bias and variance term appears and, although the introduction of a kernel function of T dimension drops the bias out, it enlarges the variance. Using the same idea as in Chapter 2, we propose to use a one-step backfitting algorithm. The idea, as already pointed out in Fan and Zhang (1999), is that additional smoothing cannot reduce the bias but it can diminish the variance. Therefore, the additional smoothing that is introduced by the backfitting enables us to achieve optimal nonparametric rates of convergence for the estimators of the unknown functions of interest. Furthermore, we also show that the resulting estimators are oracle efficient.

The rest of the chapter is organized as follows. In Section 3.2 we set up the model and the estimation procedure. In Section 3.3 we generalize the direct local linear estimator to the multivariate case and we study its asymptotic properties. In Section 3.4 we show how to apply the backfitting algorithm and we obtain the asymptotic properties for this two-step estimator. In Section 3.5 we perform a Monte Carlo simulation to analyze the behavior in small sample sizes of both estimators. Finally, Section 3.6 concludes the chapter. The proofs of the main results are collected in the Appendix 2.

3.2 Statistical model and estimation procedure

To illustrate the estimation procedure proposed in this chapter we first focus on the univariate regression model and later we extend the results to the multivariate case. Consider the linear panel data model, where the dimensions of X and Z are respectively $d = 1$ and $q = 1$,

$$Y_{it} = X_{it} m\left(Z_{it}\right) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{3.1}$$

Let $\overline{Y}_{i\cdot} = T^{-1} \sum_{s=1}^{T} Y_{is}$ and $\overline{v}_{i\cdot} = T^{-1} \sum_{s=1}^{T} v_{is}$. Taking differences from the mean in (3.1) we obtain

$$Y_{it} - \overline{Y}_{i\cdot} = X_{it} m\left(Z_{it}\right) - \frac{1}{T} \sum_{s=1}^{T} X_{is} m\left(Z_{is}\right) + v_{it} - \overline{v}_{i\cdot}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{3.2}$$

In this case, for any $z \in \mathcal{A}$, where $\mathcal{A}$ is a compact subset in a non-empty interior of $I\!R$ one has the following Taylor expansion

$$\begin{aligned} X_{it} m\left(Z_{it}\right) - \frac{1}{T} \sum_{s=1}^{T} X_{is} m\left(Z_{is}\right) &\approx \left(X_{it} - \frac{1}{T} \sum_{s=1}^{T} X_{is}\right) m(z) + \left[X_{it}\left(Z_{it} - z\right) - \frac{1}{T} \sum_{s=1}^{T} X_{is}\left(Z_{is} - z\right)\right] m'(z) \\ &+ \frac{1}{2}\left[X_{it}\left(Z_{it} - z\right)^2 - \frac{1}{T} \sum_{s=1}^{T} X_{is}\left(Z_{is} - z\right)^2\right] m''(z) + \cdots + \frac{1}{p!}\left[X_{it}\left(Z_{it} - z\right)^p - \frac{1}{T} \sum_{s=1}^{T} X_{is}\left(Z_{is} - z\right)^p\right] m^{(p)}(z) \\ &\equiv \sum_{\lambda=0}^{p} \beta_\lambda \left[X_{it}\left(Z_{it} - z\right)^\lambda - \frac{1}{T} \sum_{s=1}^{T} X_{is}\left(Z_{is} - z\right)^\lambda\right]. \end{aligned} \tag{3.3}$$

This suggests that we estimate $m(z)$, $m'(z), \cdots, m^{(p)}(z)$ by regressing $Y_{it} - \overline{Y}_{i\cdot}$ on the terms $X_{it}(Z_{it}-z)^{\lambda} - \frac{1}{T}\sum_{s=1}^{T} X_{is}(Z_{is}-z)^{\lambda}$, for $\lambda = 1, \cdots, p$, with kernel weights. Then, the quantities of interest can be estimated using a locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \beta_0\left(X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{is}\right) - \beta_1\left[X_{it}(Z_{it}-z) - \frac{1}{T}\sum_{is} X_{is}(Z_{is}-z)\right]\right)^2 \\ \times K_h(Z_{i1}-z, \cdots, Z_{iT}-z); \tag{3.4}$$

see Fan and Gijbels (1995b), Ruppert and Wand (1994) or Zhan-Qian (1996). h is a bandwidth and K is the product of univariate kernels such as $K(u_1, u_2, \cdots, u_T) = \prod_{\ell=1}^{T} K(u_\ell)$, for u_ℓ being the ℓth component of u. Also, for each one it holds

$$\int K(u)du = 1 \qquad \text{and} \qquad K_h(u) = \frac{1}{h}K(u/h).$$

Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the minimizers of (3.4). The above exposition suggests as estimators for $m(z)$ and $m'(z)$, $\widehat{m}_h(z) = \widehat{\beta}_0$ and $\widehat{m}'_h(z) = \widehat{\beta}_1$, respectively. Furthermore, let us denote by $\ddot{Y}_{it} = Y_{it} - \overline{Y}_{i\cdot}$, $\ddot{X}_{it} = X_{it} - \overline{X}_{i\cdot}$, $\beta = (\beta_0 \quad \beta_1)^{\top}$ and

$$\widetilde{Z}_{it}^{\top} = \left(\ddot{X}_{it} \quad X_{it}(Z_{it}-z) - T^{-1}\sum_{s=1}^{T} X_{is}(Z_{is}-z)\right).$$

Then, the criterion function (3.4) can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \widetilde{Z}_{it}^{\top}\beta\right)^2 \prod_{\ell=1}^{T} K_h(Z_{i\ell}-z), \tag{3.5}$$

and $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have the following expression

$$\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \left(\sum_{it}\prod_{\ell} K_h(Z_{i\ell}-z)\widetilde{Z}_{it}\widetilde{Z}_{it}^{\top}\right)^{-1} \sum_{it}\prod_{\ell} K_h(Z_{i\ell}-z)\widetilde{Z}_{it}\ddot{Y}_{it}. \tag{3.6}$$

Note that in (3.4) or (3.5) it would have been usual to introduce a kernel function around Z_{it}. By doing so, the distance between z and any of the terms of the sample $Z_{i1}, \cdots, Z_{i(t-1)}, Z_{i(t+1)}, \cdots, Z_{iT}$ cannot be controlled by a fixed bandwidth and thus the transformed reminder terms cannot be negligible. The consequence of all that is a non-negligible asymptotic bias. Here, we propose to introduce a multivariate kernel function around the vector of values $Z_{i1}, \cdots, Z_{iT}$. This modified version of a local linear regression, as it will be shown later, solves the problem of the bias but it considerably enlarges the variance. More precisely, under rather standard conditions in the next section we show that, asymptotically, the bias term is of order $O\left(h^2\right)$ but the variance is of order $O\left(1/NTh^T\right)$. As the reader may notice, this bound for the variance is rather large. In order to reduce the variance term but keeping the bias of the same order we propose to add to both terms in (3.2) the average term $\frac{1}{T}\sum_s X_{is} m\left(Z_{is}\right)$ and denote by

$$\ddot{Y}_{it}^{*} = \ddot{Y}_{it} + \frac{1}{T}\sum_{s=1}^{T} X_{is} m\left(Z_{is}\right). \tag{3.7}$$

Therefore, combining (3.2) and (3.7) we obtain

$$\ddot{Y}_{it}^{*} = X_{it} m\left(Z_{it}\right) + \ddot{v}_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{3.8}$$

where $\ddot{v}_{it} = v_{it} - \frac{1}{T}\sum_t v_{it}$. Note that equation (3.8) already shows a low dimensional problem where $m\left(\cdot\right)$ could be estimated by a standard nonparametric regression method. Unfortunately, the functions $m\left(Z_{i1}\right), \cdots, m\left(Z_{iT}\right)$ are not observed and the standard locally weighted least-squares procedure would generate unfeasible estimators. To overcome this situation, we propose to replace in (3.7) the $m\left(Z_{is}\right)$ by their corresponding estimators, $\widehat{m}_h\left(Z_{is}\right)$, in (3.6). Let $\ddot{Y}_{it}^{b} = \ddot{Y}_{it} + T^{-1}\sum_{s=1}^{T} X_{is}\widehat{m}_h\left(Z_{is}\right)$ be, the regression problem becomes

$$\ddot{Y}_{it}^{b} = X_{it} m\left(Z_{it}\right) + \ddot{v}_{it}^{b}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{3.9}$$

where the composed error term is of the form

$$\ddot{v}_{it}^{b} = \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(\widehat{m}_h\left(Z_{is}\right) - m\left(Z_{is}\right)\right) + \ddot{v}_{it}.$$

The quantities of interest can be obtained by minimizing the following criterion function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}^{b} - \gamma_0 X_{it} - \gamma_1 X_{it}\left(Z_{it} - z\right)\right)^2 K_{\widetilde{h}}\left(Z_{it} - z\right), \tag{3.10}$$

where $\widetilde{h}$ is the bandwidth of this stage. We denote by $\widetilde{\gamma}_0$ and $\widetilde{\gamma}_1$ the minimizers of (3.10). As previously, we propose as estimators for $m(\cdot)$ and $m'(\cdot)$, $\widetilde{m}_{\widetilde{h}}(z) = \widetilde{\gamma}_0$ and $\widetilde{m}'_{\widetilde{h}}(z) = \widetilde{\gamma}_1$, respectively,

$$\begin{pmatrix} \widetilde{\gamma}_0 \\ \widetilde{\gamma}_1 \end{pmatrix} = \left(\sum_{it} K_{\widetilde{h}} \left(Z_{it} - z\right) \widetilde{Z}^b_{it} \widetilde{Z}^{b\top}_{it} \right)^{-1} \sum_{it} K_{\widetilde{h}} \left(Z_{it} - z\right) \widetilde{Z}^b_{it} \ddot{Y}^b_{it}, \tag{3.11}$$

where $\widetilde{Z}^{b\top}_{it} = \left(X_{it} \quad X_{it} \left(Z_{it} - z\right)\right)$ is a 2×1-dimensional vector.

3.3 Local linear estimator: asymptotic properties

In this section we extend the above results for the case $(d \neq q \neq 1)$. Furthermore, we give the asymptotic expressions for the bias and the variance and we calculate the asymptotic distribution of the local linear regression estimator. Let us consider the multivariate version of (3.5),

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\ddot{Y}_{it} - \widetilde{Z}^{\top}_{it} \beta \right)^2 \prod_{\ell=1}^{T} K_H \left(Z_{i\ell} - z\right), \tag{3.12}$$

where in this case $\beta = \left(\beta_0^\top \quad \beta_1^\top\right)^\top$ is a $d\left(1+q\right) \times 1$ vector and we denote by $\widetilde{Z}^{\top}_{it}$ a $1 \times d\left(1+q\right)$ dimensional vector of the form

$$\widetilde{Z}^{\top}_{it} = \left(\ddot{X}^{\top}_{it} \quad X^{\top}_{it} \otimes \left(Z_{it} - z\right)^\top - T^{-1} \sum_{s=1}^{T} X^{\top}_{is} \otimes \left(Z_{is} - z\right)^\top \right).$$

Let H be a $q \times q$ symmetric positive definite bandwidth matrix, K is the product of q-variate kernels such that for each u it holds

$$\int K(u)du = 1 \qquad \text{and} \qquad K_H(u) = \frac{1}{|H|^{1/2}} K\left(H^{-1/2} u\right).$$

Let us denote by $\widehat{\beta}$ the minimizer of (3.12) and assuming $\widetilde{Z}^\top W \widetilde{Z}$ is nonsingular, the solution of (3.12) can be written as

$$\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \ddot{Y}, \tag{3.13}$$

where $\ddot{Y} = (\ddot{Y}_{11}, \cdots, \ddot{Y}_{NT})$ is a $NT \times 1$ vector while

$$W = \text{blockdiag}\left(K_H(Z_{i1} - z) \prod_{\ell=2}^{T} K_H(Z_{i\ell} - z), \cdots, K_H(Z_{iT} - z) \prod_{\ell=1}^{T-1} K_H(Z_{i\ell} - z) \right)$$

and

$$\widetilde{Z} = \begin{bmatrix} \ddot{X}_{11}^\top & X_{11}^\top \otimes (Z_{11} - z)^\top - T^{-1} \sum_{s=1}^{T} X_{1s}^\top \otimes (Z_{1s} - z)^\top \\ \vdots & \vdots \\ \ddot{X}_{NT}^\top & X_{NT}^\top \otimes (Z_{NT} - z)^\top - T^{-1} \sum_{s=1}^{T} X_{Ns}^\top \otimes (Z_{Ns} - z)^\top \end{bmatrix}$$

are $NT \times NT$ and $NT \times d\,(1+q)$ dimensional matrix, respectively.

Then, (3.12) and (3.13) suggest as estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}\,(z; H) = \widehat{\beta}_0$ and $vec(\widehat{D}_m\,(z; H)) = \widehat{\beta}_1$, respectively. In particular, the local weighted linear least-squares estimator of $m(z)$ is defined as

$$\widehat{m}\,(z; H) = \widehat{\beta}_0 = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z} \right)^{-1} \widetilde{Z}^\top W \ddot{Y}, \tag{3.14}$$

where $e_1 = (I_d \vdots 0_{dq \times d})$ is a $d\,(1+q) \times d$ selection matrix, I_d is a $d \times d$ identity matrix and $0_{dq \times d}$ a $dq \times d$ matrix of zeros.

Once the estimator in its closed form is defined, let us consider the assumptions required to obtain its asymptotic properties. Consider the data generating process defined in (3.1). Furthermore, we assume

Assumption 3.1 *Let $(Y_{it}, X_{it}, Z_{it})_{i=1,\cdots,N; t=1,\cdots,T}$ be a set of independent and identically distributed $\mathbb{R}^{1+d+q}$-random variables in the subscript i for each fixed t and strictly stationary over t for fixed i.*

Assumption 3.2 *The random errors v_{it} are independent and identically distributed, with zero mean and homoscedastic variance, $\sigma_v^2 < \infty$. They are also independent of X_{it} and Z_{it} for all i and t. In addition, $E\,|v_{it}|^{2+\delta}$, for some $\delta > 0$.*

Assumption 3.3 *The unobserved cross-sectional effect, μ_i, can be arbitrarily correlated with both X_{it} and/or Z_{it} with an unknown correlation structure.*

Assumption **3.1** is standard in panel data analysis. We could consider other settings of time-dependence such as strong mixing conditions (as in Cai and Li (2008)) or nonstationary time series (as in Cai et al. (2009)). However, since in this chapter we investigate the asymptotic properties of the estimators as N tends to infinity and T is fixed it is enough to assume stationarity. Assumption **3.2** is also standard for the conventional transformation in deviation from the mean; see Wooldridge (2003) or Hsiao (2003) for the fully parametric case. It also rules out the presence of lagged endogenous variables. Independence between the idiosyncratic error term and the covariates X_{it} and/or Z_{it} is assumed without loss of generality although it can be relaxed assuming some dependence in higher order moments. Finally, Assumption **3.3** imposes the so-called fixed effects.

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT})$ and $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ the observed covariates sample vectors, we also need to impose the following additional assumptions about moments and densities,

Assumption 3.4 *Let $f_{Z_{1t}}(\cdot)$ be the probability density function of Z_{1t}, for $t = 1, \cdots, T$. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.*

Assumption 3.5 *The function $E\left[\ddot{X}_{it}\ddot{X}_{it}^\top | Z_{i1} = z_1, \cdots, Z_{iT} = z_T\right]$ is definite positive for any interior point of $(z_1, z_2, \cdots, z_T)$ in the support of $f_{Z_{i1},\cdots,Z_{iT}}(z_1, z_2, \cdots, z_T)$.*

Assumption 3.6 *Let $\|A\| = \sqrt{tr\left(A^\top A\right)}$, then $E\left[\|X_{it}X_{it}^\top\|^2 | Z_{i1} = z_1, \cdots, Z_{iT} = z_T\right]$ is bounded and uniformly continuous in its support. Furthermore, the matrix functions $E\left[X_{it}X_{is}^\top | Z_{i1} = z_1, \cdots, Z_{iT} = z_T\right]$ and $E\left[\ddot{X}_{it}X_{is}^\top | Z_{i1} = z_1, \cdots, Z_{iT} = z_T\right]$, for $t = s$ and $t \neq s$, are bounded and uniformly continuous in their support.*

Assumption 3.7 *Let z an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $m_1(\cdot), m_2(\cdot), \cdots, m_d(\cdot)$ are bounded and uniformly continuous.*

Assumption 3.8 *The q-variate Kernel functions K are compactly supported, bounded kernel such that $\int uu^\top K(u)du = \mu_2(K)I$ and $\int K^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars and I is the $q \times q$ identity matrix. In addition, all odd-order moments* *of K vanish, that is $\int u_1^{\imath_1}\cdots u_q^{\imath_q}K(u)du = 0$, for all nonnegative integers $\imath_1,\cdots,\imath_q$ such that their sum is odd.*

Assumption 3.9 *The bandwidth matrix H is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as $N\to\infty$ in such a way that $N|H|\to\infty$.*

Assumption 3.10 *For some $\delta>0$, the functions $E\left[|X_{it}v_{it}|^{2+\delta}\,|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right]$, $E\left[|X_{is}v_{it}|^{2+\delta}|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right]$, and $E\left[|\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right]$ are bounded and uniformly continuous in any point of their support.*

This second set of assumptions is more directly related to nonparametric statistics literature. They are basically smoothness and boundedness conditions. Assumption **3.4** imposes smoothness conditions in the probability density function of Z_{1t}, for $t=1,\cdots,T$. Furthermore, Assumptions **3.5-3.6** are smoothness conditions on moment functionals. Assumptions **3.7-3.9** are standard in the literature of local linear regression where, in particular, Assumption **3.9** contains a standard bandwidth condition for smoothing techniques. Finally, Assumption **3.10** is required to show that the Lyapunov conditions hold for the Central Limit Theorem.

Under these assumptions we obtain the following asymptotic expressions for the conditional bias and conditional variance-covariance matrix of the local weighted linear least-squares estimator,

Theorem 3.1 *Assume conditions 3.1-3.3 and 3.4-3.9 hold, then as $N\to\infty$ and T is fixed we obtain*

$$
\begin{aligned}
&E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]-m(z)\\
&=\frac{1}{2}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}\left(z,\cdots,z\right)\left(\mu_2(K_{u_\tau})\mathcal{B}_{\ddot{X}_tX_t}\left(z,\cdots,z\right)-\frac{1}{T}\sum_{s=1}^{T}\mu_2\left(K_{u_s}\right)\mathcal{B}_{\ddot{X}_tX_s}\left(z,\cdots,z\right)\right)\\
&\quad\times diag_d(tr(\mathcal{H}_{m_r}(z)H))\imath_d+o_p\left(tr(H)\right)
\end{aligned}
$$

and

$$
Var\left(\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right) \;=\; \frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{NT|H|^{T/2}}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}\left(z,\cdots,z\right)\left(1+o_p(1)\right),
$$

where τ is any index between 1 and T,

$$\begin{aligned}
\mathcal{B}_{\ddot{X}_t X_t}(z,\cdots,z) &= E\left[\ddot{X}_{it}X_{it}^{\top}|Z_{i1}=z,\cdots,Z_{iT}=z\right] f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z), \\
\mathcal{B}_{\ddot{X}_t X_s}(z,\cdots,z) &= E\left[\ddot{X}_{it}X_{is}^{\top}|Z_{i1}=z,\cdots,Z_{iT}=z\right] f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z), \\
\mathcal{B}_{\ddot{X}_t \ddot{X}_t}(z,\cdots,z) &= E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{i1}=z,\cdots,Z_{iT}=z\right] f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z),
\end{aligned}$$

$diag_d\left(tr(\mathcal{H}_{m_r}(z)H)\right)$ stands for a diagonal matrix of elements $tr(\mathcal{H}_{m_r}(z)H)$, for $r=1,\cdots,d$, where $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the rth component of $m(\cdot)$. Finally, we denote by $\imath_d$ is a $d\times 1$ unit vector.

The proof of this result is done in the Appendix 2.

This theorem shows that $\widehat{m}(z;H)$ is, conditionally on the sample, a consistent estimator of $m(z)$. Furthermore, as it was already remarked in the previous section, although the bias shows the standard order of magnitude for such problems, the variance shows an asymptotic expression that is larger than the expected in this type of problems. In order to achieve an optimal rate of convergence, the variance term must be of order $1/NT|H|^{1/2}$ whereas our result shows a bound of order $1/NT|H|^{T/2}$. Just to clarify the asymptotic behavior of the estimator we show its properties for the univariate case, $d=q=1$ and $H=h^2I$,

Corollary 3.1 *Assume conditions 3.1-3.9 hold, then if $h\to 0$ in such a way that $Nh^2\to\infty$ as N tends to infinity and T is fixed we obtain*

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2}c(z,z)\,m''(z)h^2+o_p\left(h^2\right),$$

where

$$c(z,z) = \frac{\mu_2(K_{u_\tau})E\left[\ddot{X}_{it}X_{it}|Z_{i1}=z,\cdots,Z_{iT}=z\right]-T^{-1}\sum_{s=1}^{T}\mu_2(K_{u_s})E\left[\ddot{X}_{it}X_{is}|Z_{i1}=z,\cdots,Z_{iT}=z\right]}{E\left[\ddot{X}_{it}^2|Z_{i1}=z,\cdots,Z_{iT}=z\right]}.$$

Furthermore, if $\mu_2(K_{u_1})=\cdots=\mu_2(K_{u_T})=\mu_2(K_{u_\tau})=\mu_2(K)$ then the bias term has the following expression

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2}\mu_2(K)m''(z)h^2+o_p\left(h^2\right),$$

whereas if $R(K_{u_1}) = \cdots = R(K_{u_T}) = R(K)$ *the variance-covariance matrix is*

$$\begin{aligned} & Var\left(\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right) \\ &= \frac{\sigma_v^2 R\left(K\right)^T}{NTh^2 f_{Z_{i1},\cdots,Z_{iT}}\left(z,\cdots,z\right) E\left[\ddot{X}_{it}^2|Z_{i1}=z,\cdots,Z_{iT}=z\right]}\left(1+o_p(1)\right). \end{aligned}$$

As a tool to construct asymptotic confidence bands we give also a result that provides the asymptotic distribution of the estimator.

Theorem 3.2 *Assume conditions 3.1-3.3 and 3.4-3.10 hold, then as* $N \to \infty$ *and* T *is fixed we obtain*

$$\sqrt{NT|H|^{T/2}}\left(\widehat{m}\left(z;H\right) - m\left(z\right)\right) \xrightarrow{d} \mathcal{N}\left(b(z), v(z)\right),$$

where

$$\begin{aligned} b(z) &= \frac{1}{2}\mu_2(K_u) diag_d\left(tr\left(\mathcal{H}_{m_r}(z) H\sqrt{NT|H|^{T/2}}\right)\right)\imath_d, \\ v(z) &= \sigma_v^2 R\left(K\right)^T \mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\cdots,z). \end{aligned}$$

The proof of this result is shown in the Appendix 2.

We can compare the results obtained here with those obtained in Chapter 2 for the first differences case. As expected, for both estimators the bias term presents the same linear dependence in the trace of the bandwidth matrix H. However, the variance term differs from one to the other estimator. In the first differences case, see Theorem 2.1 of Chapter 2, up to a constant, the variance term exhibits a dependence from the bandwidth matrix H of order $1/NT|H|$ whereas in our case it is of order $1/NT|H|^{T/2}$. That is, the ratio between the first differences and the deviances from the mean estimators is of order $|H|^{(T-2)/2}$. For $T = 2$, the estimators show the same rate of convergence. This is clearly expected. For $T > 2$, the first-differences estimator under the conditions established above shows a faster rate of convergence for the variance terms as far as the diagonal elements of the bandwidth matrix H tend to zero. This was also expected because the dimensionality of the kernel used in the local linear regression procedure is different in both cases.

Now we show the asymptotic optimality of the first-step backfitting algorithm that is obtained in Section 3.2.

3.4 The backfitting estimator

In the regression model expressed in differences from the mean, by using of a local linear regression approach with a high dimensional kernel weight we can consistently estimate the function of interest but at the price of achieving a slow rate of convergence. However, as it is noted in Section 3.2, we can solve this problem turning to an alternative procedure that allows us to cancel asymptotically all additive terms of the function of interest expected in the model.

Considering the multivariate version of (3.9) and let

$$\ddot{Y}_{it}^{b} = X_{it}^{\top} m\left(Z_{it}\right) + \ddot{v}_{it}^{b}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{3.15}$$

where

$$\ddot{v}_{it}^{b} = \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \left(\widehat{m}\left(Z_{is}\right) - m\left(Z_{is}\right)\right) + \ddot{v}_{it}.$$

The quantities of interest in (3.15) can be estimated by minimizing the following locally weighted linear regression

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\ddot{Y}_{it} - \widetilde{Z}_{it}^{b\top} \gamma\right)^{2} K_{\widetilde{H}}\left(Z_{it} - z\right), \tag{3.16}$$

where $\widetilde{H}$ is a $q \times q$ symmetric positive definite bandwidth matrix, $\gamma = (\gamma_0^{\top} \quad \gamma_1^{\top})^{\top}$ is a $d\,(1+q) \times 1$ vector and $\widetilde{Z}_{it}^{b\top} = (X_{it}^{\top} \quad X_{it}^{\top} \otimes \left(Z_{it} - z\right)^{\top})$ is a $1 \times (1+q)$ vector.

Furthermore, let $\widetilde{\gamma} = (\widetilde{\gamma}_0^{\top} \quad \widetilde{\gamma}_1^{\top})^{\top}$ be the minimizer of (3.16). As estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, we suggest $\widetilde{m}(z, \widetilde{H}) = \widetilde{\gamma}_0$ and $vec(\widetilde{D}_m(z; \widetilde{H})) = \widetilde{\gamma}_1$, respectively, i.e.,

$$\widetilde{m}(z; \widetilde{H}) = \widetilde{\gamma}_0 = e_1^{\top} \left(\widetilde{Z}^{b\top} W^{b} \widetilde{Z}^{b}\right)^{-1} \widetilde{Z}^{b\top} W^{b} \ddot{Y}^{b}, \tag{3.17}$$

where $\ddot{Y}^{b} = (\ddot{Y}_{11}^{b}, \cdots, \ddot{Y}_{NT}^{b})$ is a $NT \times 1$ vector and W^{b} and $\widetilde{Z}^{b}$ are $NT \times NT$ and $NT \times d\,(1+q)$ dimensional matrix, respectively, of the form

$$W^{b} \quad = \quad diag\left(K_{\widetilde{H}}\left(Z_{11} - z\right), \cdots, K_{\widetilde{H}}\left(Z_{NT} - z\right)\right)$$

and

$$\widetilde{Z}^{b} \quad = \quad \begin{bmatrix} X_{11}^{\top} & X_{11}^{\top} \otimes \left(Z_{11} - z\right)^{\top} \\ \vdots & \vdots \\ X_{NT}^{\top} & X_{NT}^{\top} \otimes \left(Z_{NT} - z\right)^{\top} \end{bmatrix}.$$

We now study the asymptotic behavior of the so-called backfitting estimator. At this stage we need the results shown in Theorem 3.1 to hold uniformly in z. In order to do so, we can rely on the well-known results in Masry (1996). In fact, some of the conditions already enounced in Section 3.3 are sufficient to show the uniform rates for $\widehat{m}(z;H)$. However, we need some additional assumptions that relate the bandwidths of both $\widehat{m}\left(z;H\right)$ and $\widetilde{m}(z;\widetilde{H})$.

Assumption 3.11 *The bandwidth matrix $\widetilde{H}$ is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|\widetilde{H}|\rightarrow\infty$.*

Assumption 3.12 *The bandwidth matrices H and $\widetilde{H}$ must fulfill that, as N tends to infinity, $N\left|H\right|\left|\widetilde{H}\right|/log(N)\rightarrow\infty$, and $tr\left(H\right)/tr(\widetilde{H})\rightarrow 0$.*

Then, under these assumptions we obtain the following asymptotic expressions for the conditional bias and conditional variance-covariance matrix of $\widetilde{m}(z;\widetilde{H})$.

Theorem 3.3 *Assume conditions 3.1-3.3, 3.4-3.8 and 3.11-3.12 hold, then as $N\rightarrow\infty$ and T remains to be fixed we obtain*

$$E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right]-m(z) = \frac{1}{2}\mu_2\left(K_u\right)diag_d\left(tr(\mathcal{H}_{m_r}(z)\widetilde{H})\right)\imath_d+o_p(tr(\widetilde{H}))$$

and

$$Var\left(\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right) = \frac{\sigma_v^2R(K)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}^{-1}_{X_tX_t}(z)\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)\mathcal{B}^{-1}_{X_tX_t}(z)\left(1+o_p(1)\right),$$

where $diag_d\left(tr(\mathcal{H}_{mr}(z)\widetilde{H})\right)$ stands for a diagonal matrix of elements $tr(\mathcal{H}_{m_r}(z)\widetilde{H})$, for $r=1,\cdots,d$ and $\imath_d$ is a $d\times 1$ unit vector.

The proof of this result is done in the Appendix 2.

On one hand, we realize that the bias term is influenced by the amount of smoothing, H, as well as the curvature of $m(z)$ at z in a particular direction, measured through each entry of $\mathcal{H}_m(z)$. In this way, we can guess that this estimator exhibits a higher conditional bias when there is a higher curvature and more smoothing. On the other hand, from the standpoint of the conditional variance we can see that it is a bit different from the corresponding for the standard case. In particular, it will be increased when the smoothing is lower and sparse data near z but now also depends on the time-demeaned covariates $\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)$. Regardless, it is proved that the estimation procedure developed in this chapter provides a nonparametric estimator in which the variance-covariance matrix of all its components is asymptotically the same as if we would know the rest of components of the mean deviation transformed expression, the so-called oracle efficiency property.

3.5 Monte Carlo simulations

In this section, Monte Carlo simulations are carried out in order to verify our theoretical results. Furthermore, we analyze the small sample size behavior of our estimator under the statistical setting analyzed in the previous sections.

As it is well-known, the mean squared error (MSE) is a suitable measure of the estimation accuracy of the proposed estimators. Thus, let us denote by φ as the φth replication and Q as the number of replications,

$$MSE\left(\widehat{m}\left(z;H\right)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q} E\left[\left(\sum_{r=1}^{d}\left(\widehat{m}_{\varphi r}\left(z;H\right) - m_{\varphi r}(z)\right)\Delta X_{it,\varphi r}\right)^2\right],$$

which can be approximated by the averaged mean squared error (AMSE) such as

$$AMSE\left(\widehat{m}\left(z;H\right)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{r=1}^{d}\left(\widehat{m}_{\varphi r}\left(z;H\right) - m_{\varphi r}(z)\right)\Delta X_{it,\varphi r}\right)^2.$$

Observations are generated from the following varying coefficient panel data model of unknown form

$$Y_{it} = X_{dit}^{\top} m\left(Z_{qit}\right) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T; \qquad d, q = 1, 2,$$

where X_{dit} and Z_{qit} are random variables generated such that $X_{dit} = 0.5X_{di(t-1)} + \xi_{it}$ and $Z_{qit} = w_{qit} + w_{qi(t-1)}$, where w_{qit} is an $i.i.d.$ uniformly distributed $[0, \Pi/2]$ random variable and ξ_{it} an $i.i.d.N\left(0,1\right)$. Furthermore, v_{it} is an $i.i.d.N\left(0,1\right)$ random variable and $m\left(\cdot\right)$ is a pre-specified function to be estimated.

With the aim of verifying the theoretical results in Sections 3.3 and 3.4 we consider three different cases

$$\begin{aligned}(1)\quad Y_{it} &= X_{1it}m_1\left(Z_{1it}\right)+\mu_{1i}+v_{it},\\(2)\quad Y_{it} &= X_{1it}m_1\left(Z_{1it},Z_{2it}\right)+\mu_{2i}+v_{it},\\(3)\quad Y_{it} &= X_{1it}m_1\left(Z_{1it}\right)+X_{2it}m_2\left(Z_{2it}\right)+\mu_{1i}+v_{it},\end{aligned}$$

where the chosen functionals form are $m_1\left(Z_{1it}\right)=\sin\left(Z_{1it}*\Pi\right)$, $m_2\left(Z_{2it}\right)=exp\left(-Z_{2it}^2\right)$, and $m_1\left(Z_{1it},Z_{2it}\right)=\sin\left(\frac{1}{2}\left(Z_{1it}+Z_{2it}\right)*\Pi\right)$.

In addition, we experiment with two specifications for the individual heterogeneity

a. μ_{1i} depends on Z_{1it}, where the dependence is imposed by generating $\mu_{1i}=c_0\overline{Z}_{1i.}+u_i$ for $i=2,\cdots,N$ and $\overline{Z}_{1i.}=T^{-1}\sum_{t=1}^{T}Z_{1it}$,

b. μ_{2i} depends on Z_{1it},Z_{2it} through $\mu_{2i}=c_0\overline{Z}_{i.}+u_i$ for $i=2,\cdots,N$ and $\overline{Z}_{i.}=\frac{1}{2}\left(\overline{Z}_{1i.}+\overline{Z}_{2i.}\right)$,

where in both cases u_i is an $i.i.d.N\left(0,1\right)$ random variable and $c_0=0.5$ controls the correlation between the fixed effects and some of the regressors of the model.

In the experiment we use 1000 Monte Carlo replications. The number of time observations T is set up to three, while the number of cross-sections N is either 50, 100 or 200. The Gaussian kernel has been used and the bandwidth is chosen as $\widehat{H}=\widehat{h}I$, and $\widehat{h}=\widehat{\sigma}_z\left(NT\right)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{qit}\}_{i=1,t=1}^{N,T}$.

The simulations results that we report in the following tables are based on the AMSE.

Table 3.1. Fixed effects estimators. AMSE for $d=1$ and $q=1$

		Local polynomial estimator	Backfitting estimator
T=3	N = 50	0.40494	0.16476
	N = 100	0.21256	0.10741
	N = 150	0.14272	0.07520

Table 3.2. Fixed effects estimators. AMSE for $d = 1$ and $q = 2$

		Local polynomial estimator	Backfitting estimator
	N = 50	2.46715	0.76366
T=3	N = 100	2.28708	0.50700
	N = 150	1.82741	0.38136

Table 3.3. Fixed effects estimators. AMSE for $d = 2$ and $q = 1$

		Local polynomial estimator	Backfitting estimator
	N = 50	0.84175	0.34369
T=3	N = 100	0.54485	0.21006
	N = 150	0.33999	0.135574

The results from the simulation show some expected outcomes. Mainly, if we analyze the behavior of the first-step estimator, we realize that the AMSE increases when the dimensionality of Z goes from one to two, as expected. This does not happen if we let the dimension of X become larger. This is the course of dimensionality. Furthermore, the rate of decay, as N tends to infinity, is also much slower when $q = 2$ (Table 3.2) than when $q = 1$, Tables 3.1 and 3.3. However, for this first-step estimator, some unexpected results in small sample sizes occur. According to the theoretical results, the rate of convergence of the first-step estimator should not depend on the dimension of the X covariates. However, if we compare Table 3.1 against Table 3.3 we do not conclude the same. There seem to be an effect in the rate of convergence of the AMSE, although this effect is much weaker than in the case $q > 1$.

With respect to the backfitting estimator, the results fully confirm the theoretical findings. As we can see, in all cases the AMSE of the backfitting algorithm is smaller than the correspondent AMSE for the first-step estimator. Furthermore, the rate at which the backfitting estimator decays is faster than in the first step estimator.

3.6 Conclusions

In this chapter, we consider the estimation of a panel data model where the heterogeneity term is arbitrarily correlated with the covariates and the coefficients are unknown functions of some explanatory variables. The estimator is based in a deviation from the mean transformation of the regression model and then a local linear regression is applied to estimate the unknown varying coefficient functions. It turns out that the standard use of this technique rends a non-negligible asymptotic bias. In order to avoid it, we introduce a high dimensional kernel weight in the estimation procedure. As a consequence, the resulting estimator shows a bias that asymptotically tends to zero at usual nonparametric rates. However, the variance is enlarged, and therefore the estimator shows a very slow rate of convergence. In order to achieve the optimal rate, we propose a one-step backfitting algorithm. The resulting two-step estimator is shown to be asymptotically normal and its rate of convergence is optimal within its class of smoothness functions. Furthermore, the estimator is oracle efficient. Finally, we show some Monte Carlo results that confirm the theoretical findings.

Chapter 4

Differencing techniques in nonparametric panel data varying coefficient models with fixed effects: a Monte Carlo analysis

4.1 Introduction

In the previous chapters some new techniques for the estimation of semi-parametric varying coefficient panel data models with fixed effects have been proposed. These new techniques fall within the class of the so-called differencing estimators and enable us to obtain consistent estimators of the unknown objective functions. Under fairly general conditions, we have previously shown the asymptotic properties of the proposed estimators, first-differences and fixed effects. In a fully parametric context, it is well-known that under strict exogeneity assumptions the behavior of the differencing estimators depends on the stochastic structure of the random error term; see Wooldridge (2003). In order to prove whether this statement holds for nonparametric estimators, in this chapter a comparative analysis on the behavior of these nonparametric estimators in finite samples is performed.

As we have already stated, a panel data varying coefficient model where some regression coefficients are allowed to be varying depending on some exogenous continuous variables is of the form

$$Y_{it} = X_{it}^{\top} m\left(Z_{it}\right) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{4.1}$$

where X_{it} and Z_{it} are $d \times 1$ and $q \times 1$ vector of covariates, respectively, $m(Z)$ is a $d \times 1$ vector of smooth functions to estimate, μ_i the unobserved individual heterogeneity and v_{it} the random disturbances.

In order to avoid the statistical dependence problem between μ_i and X_{it}/Z_{it}, in Chapters 2 and 3 we propose to remove the unobserved cross-sectional heterogeneity through differencing transformations. The most popular are first differences and differences from the mean. In the first case, the model to analyze is

$$Y_{it} - Y_{i(t-1)} = X_{it}^{\top} m\left(Z_{it}\right) - X_{i(t-1)}^{\top} m\left(Z_{i(t-1)}\right) + v_{it} - v_{i(t-1)}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{4.2}$$

whereas the deviation from the mean transformation of the regression model implies

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = X_{it}^{\top} m\left(Z_{it}\right) - \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top} m\left(Z_{is}\right) + v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{4.3}$$

However, as it is emphasized in Su and Ullah (2011), direct application of nonparametric regression techniques to estimate the unknown function of interest in either (4.2) or (4.3) is a cumbersome task since it is necessary to consider $m(\cdot)$ as an additive function whose elements share the same functional form. To overcome this situation, in Chapters 2 and 3 we propose direct strategies that provide consistent estimators of $m(\cdot)$, either in first differences or differences from the mean regression equation. In particular, their core idea is to approximate locally the additive function of either $X_{it}^{\top} m(Z_{it}) - X_{i(t-1)}^{\top} m(Z_{i(t-1)})$ or $X_{it}^{\top} m(Z_{it}) - T^{-1}\sum_{s=1}^{T} X_{is}^{\top} m(Z_{is})$ and its derivatives through a Taylor series expansion.

Nevertheless, although these new estimation strategies enable us to solve the non-negligible asymptotic bias resulting from the direct application of standard nonparametric regression techniques to panel data models in differences, they exhibit the standard dilemma of the nonparametric estimates. In other words, any attempt to hold back the bias is offset by an increase of the variance term and therefore the resulting estimators achieve a suboptimal rate of convergence. Because the proposed first-differences and fixed effects estimator present this feature, in Chapters 2 and 3 we propose to ameliorate the variance term exploding the additive structure of the regression model by combining the previous procedure with a one-step backfitting algorithm. In this way, and as it is emphasized in Fan and Zhang (1999), an additional smoothing can reduce the variance without affecting the asymptotic order of the bias so the resulting estimators achieve the optimal rate of convergence of this type of problems, i.e., $NT|H|^{1/2}$.

Since these two backfitting estimators are asymptotically equivalent, up to some different constants, it is interesting to analyze their behavior in small sample sizes. In a fully parametric context, and under strict exogeneity assumptions, the stochastic structure of v_{it}'s is the determinant of the performance of the estimator. However, in the nonparametric setting, apart from the previous issues other factors such as dimension of q and sizes of T, and more importantly N, are of great interest. In particular, in this chapter we would be interested in learning whether, for different values of q and N, the fixed effects estimator is more efficient than the first-differences estimator when v_{it}'s are serially correlated. Or whether, the opposite holds when the errors follow a random walk. The last results are rather standard in fully parametric settings. However, we would like to know whether this behavior, in terms of the empirical AMSE, is affected by the curse of dimensionality as in nonparametric frameworks. Finally, as the sample size increases one might expect that both estimators become equal in terms of the asymptotic rates of convergence.

The rest of the chapter is organized as follows. In Section 4.2, we review the local linear estimation procedures for both differencing estimators that we propose in the previous chapters. We also analyze their main asymptotic properties. In Section 4.3, we review the one-step backfitting algorithm of both estimators that allow them to achieve asymptotically optimal rates. In Section 4.4, we compare the estimators considered via a Monte Carlo simulation. Finally, we conclude in Section 4.5.

4.2 Local linear estimation procedure

In this section, we compare both local linear regression procedures proposed in Chapters 2 and 3. To illustrate these estimation procedures, we first focus on the univariate regression model and later we extend the results to the multivariate case.

Consider the first differences transformation in (4.2) with $d = q = 1$, for any $z \in \mathcal{A}$, where $\mathcal{A}$ is a compact subsect in a nonempty interior of $I\!R$, one has the following Taylor expansion

$$\begin{aligned} X_{it} m\left(Z_{it}\right) - X_{i(t-1)} m\left(Z_{i(t-1)}\right) &\approx m(z)\Delta X_{it} + m'(z)\left[X_{it}\left(Z_{it}-z\right) - X_{i(t-1)} m\left(Z_{i(t-1)}\right)\right] \\ &+ \frac{1}{2}\left[X_{it}\left(Z_{it}-z\right)^2 - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^2\right] \\ &+ \cdots + \frac{1}{p!} m^{(p)}(z)\left[X_{it}\left(Z_{it}-z\right)^p - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^p\right] \\ &\equiv \sum_{\lambda=0}^{p} \beta_{F_\lambda}\left[X_{it}\left(Z_{it}-z\right)^\lambda - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^\lambda\right]. \quad (4.4) \end{aligned}$$

Similarly, for the deviation from the mean regression in (4.3), one has the following

$$\begin{aligned} & X_{it} m\left(Z_{it}\right) - \frac{1}{T}\sum_{s=1}^{T} X_{is} m\left(Z_{is}\right) \approx \left(X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{is}\right) m(z) + \left[X_{it}\left(Z_{it}-z\right) - \frac{1}{T}\sum_{s=1}^{T} X_{is} m\left(Z_{is}\right)\right] m'(z) \\ + \quad & \frac{1}{2}\left[X_{it}\left(Z_{it}-z\right)^2 - \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)^2\right] m''(z) + \cdots + \frac{1}{p!}\left[X_{it}\left(Z_{it}-z\right)^p - \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)^p\right] m^{(p)}(z) \\ \equiv \quad & \sum_{\lambda=0}^{p} \beta_{W_\lambda}\left[X_{it}\left(Z_{it}-z\right)^\lambda - \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)^\lambda\right]. \qquad (4.5) \end{aligned}$$

Both expressions (4.4) and (4.5) suggest that we estimate $m(z), m'(z), \cdots, m^{(p)}(z)$ by regressing, respectively, ΔY_{it} on the terms $X_{it}\left(Z_{it}-z\right)^\lambda - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)^\lambda$ and $\ddot{Y}_i = Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is}$ on $X_{it}\left(Z_{it}-z\right)^\lambda - \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)^\lambda$, for $\lambda = 1, \cdots, p$, with different kernel weights. Thus, the quantities of interest in both cases can be estimated using locally weighted linear regression; see Fan and Gijbels (1995b).

For (4.4), that means to solve

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta Y_{it} - \beta_{F_0}\Delta X_{it} - \beta_{F_1}\left[X_{it}\left(Z_{it}-z\right) - X_{i(t-1)}\left(Z_{i(t-1)}-z\right)\right]\right)^2 \\ &\qquad \times K_h\left(Z_{it}-z\right) K_h\left(Z_{i(t-1)}-z\right), \qquad (4.6) \end{aligned}$$

whereas for (4.5) we have

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \beta_{w_0}\left(X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{is}\right) - \beta_{w_1}\left[X_{it}\left(Z_{it}-z\right) - \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)\right]\right)^2 \\ &\qquad \times K_h\left(Z_{i1}-z, \cdots, Z_{iT}-z\right), \qquad (4.7) \end{aligned}$$

where h is a bandwidth and K is an univariate kernel such that

$$\int K(u)du = 1 \qquad \text{and} \qquad K_h(u) = \frac{1}{h}K\left(u/h\right).$$

Let us denote by $\widehat{\beta}_{F_0}$ and $\widehat{\beta}_{F_1}$ the minimizers of (4.6), and $\widehat{\beta}_{w_0}$ and $\widehat{\beta}_{w_1}$ the corresponding of (4.7). The above exposition suggests as estimators for $m(\cdot)$ and $m'(\cdot)$, $\widehat{m}_h(z) = \widehat{\beta}_{F_0}$ and $\widehat{m}'_h(z) = \widehat{\beta}_{F_1}$, respectively. Meanwhile, the estimators for the deviation from the mean regression are $\widehat{m}_h(z) = \widehat{\beta}_{w_0}$ and $\widehat{m}'_h(z) = \widehat{\beta}_{w_1}$, respectively.

Note that with the aim of avoiding the non-negligible asymptotic bias, in (4.6) we propose a bivariate kernel that enables us to consider a local approximation around the pair $(Z_{it}, Z_{i(t-1)})$, not only around Z_{it} as it is usual. Consequently, the distance between Z_{is} (for $s \neq t$) and z is considered for estimates so that the transformed remainder terms are negligible. The same can be said for (4.7). Although there, the non-degenerated bias must be removed by considering a local approximation around the $T \times 1$ vector $(Z_{i1}, \cdots, Z_{iT})$. Note that the difference in the local approximation makes a substantial difference in terms of the asymptotic variance in both estimators. In fact, in Theorems 2.2 and 3.2 it is shown that under similar conditions the order of the bias for the univariate case will be the same, $O(h^2)$, but the variance is for $T > 1$ rather different. For the first-differences estimator the variance is of order $O(1/NTh^2)$, whereas for the other estimator is of order $O(1/NTh^T)$.

For $d \neq q \neq 1$, the estimators have the following form. Denote by $\widehat{\beta}_F = (\widehat{\beta}_{F_0}^\top \quad \widehat{\beta}_{F_1}^\top)$ a $d(1+q)$-vector that minimizes (4.6) in the multivariate case, that is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta Y_{it} - \Delta X_{it}^\top \beta_{F_0} - \left[X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes\left(Z_{i(t-1)}-z\right)\right]^\top \beta_{F_1}\right)^2$$
$$\times K_H\left(Z_{it}-z\right)K_H\left(Z_{i(t-1)}-z\right), \tag{4.8}$$

where H is a $q \times q$ symmetric positive definite bandwidth matrix and K is a q-variate kernel.

Let $D(z) = vec(D_m(z))$ be a $dq \times q$ vector and $D_m(z) = \partial m(z)/\partial z$ a $d \times q$ partial derivative matrix of the dth component of $m(z)$ with respect to the elements of the $q \times 1$ vector z. Denote by $\mathcal{H}_m(z) = \partial m(z)/\partial z \partial z^\top$ a $dq \times d$ matrix of the Hessian matrix of the

dth component of $m(z)$. We suggest as estimators for $m(z)$ and $D_m(z)$, $\widehat{m}_F(z;H) = \widehat{\beta}_{F_0}$ and $vec(\widehat{D}_{F_m}(z)(z;H)) = \widehat{\beta}_{F_1}$, respectively. Assuming that $\widetilde{Z}_F^\top W_F \widetilde{Z}_F$ is nonsingular, the minimization problem (4.8) has the following solution in matrix form

$$\begin{pmatrix} \widehat{\beta}_{F_0} \\ \widehat{\beta}_{F_1} \end{pmatrix} = \left(\widetilde{Z}_F^\top W_F \widetilde{Z}_F\right)^{-1} \widetilde{Z}_F^\top W_F \Delta Y, \tag{4.9}$$

where $\Delta Y = (\Delta Y_{12}, \cdots, \Delta Y_{NT})^\top$ is a $N(T-1)$ vector while W_F and $\widetilde{Z}_F$ are $N(T-1) \times N(T-1)$ and $N(T-1) \times d(1+q)$ matrix, respectively, of the form

$$W_F = diag\left(K_H\left(Z_{12}-z\right)K_H\left(Z_{11}-z\right), \cdots, K_H\left(Z_{NT}-z\right)K_H\left(Z_{N(T-1)}-z\right)\right)$$

and

$$\widetilde{Z}_F = \begin{bmatrix} \Delta X_{12}^\top & X_{12}^\top \otimes (Z_{12}-z)^\top - X_{11}^\top \otimes (Z_{11}-z)^\top \\ \vdots & \vdots \\ \Delta X_{NT}^\top & X_{NT}^\top \otimes (Z_{NT}-z)^\top - X_{N(T-1)}^\top \otimes \left(Z_{N(T-1)}-z\right)^\top \end{bmatrix}.$$

Then, the local weighted linear least-squares estimator for $m(z)$ is defined as

$$\widehat{m}_F(z;H) = e_1^\top \left(\widetilde{Z}_F^\top W_F \widetilde{Z}_F\right)^{-1} \widetilde{Z}_F^\top W_F \Delta Y, \tag{4.10}$$

where $e_1 = (I_d \vdots 0_{dq\times d})$ is a $d(1+q)\times d$ selection matrix, I_d is a $d \times d$ identity matrix and $0_{dq\times d}$ a $dq \times d$ matrix of zeros.

We focus now on the estimators of a semi-parametric panel data varying coefficient models in deviation from the mean. Let $\widehat{\beta}_w = (\widehat{\beta}_{w_0}^\top \quad \widehat{\beta}_{w_1}^\top)^\top$ be a $d(1+q)$- vector that minimizes the expression (4.5) in the multivariate case, i.e.,

$$\sum_{i=1}^N \sum_{t=1}^T \left(\ddot{Y}_{it} - \left(X_{it} - \frac{1}{T}\sum_{s=1}^T X_{is}\right)^\top \beta_{w_0} - \left[X_{it}\otimes(Z_{it}-z) - \frac{1}{T}\sum_{s=1}^T X_{is}\otimes(Z_{is}-z)\right]^\top \beta_{w_1}\right)^2$$
$$\times \prod_{\ell=1}^T K_H\left(Z_{i\ell}-z\right), \tag{4.11}$$

where now K is the product of univariate kernels such that $K(u_1,u_2,\cdots,u_T) = \prod_{\ell=1}^T K(u_\ell)$ and u_ℓ is the ℓth component of u. We suggest as estimators for $m(z)$ and $D_m(z)$, $\widehat{m}_w(z;H) = \widehat{\beta}_{w_0}$ and $vec(\widehat{D}_{w_m}(z;H)) = \widehat{\beta}_{w_1}$, respectively.

Assuming $\widetilde{Z}_w^\top W_w \widetilde{Z}_w$ is nonsingular, the matrix form of the solution of the minimization problem (4.11) can be written as

$$\begin{pmatrix} \widehat{\beta}_{w_0} \\ \widehat{\beta}_{w_1} \end{pmatrix} = \left(\widetilde{Z}_w^\top W_w \widetilde{Z}_w\right)^{-1} \widetilde{Z}_w^\top W_w \ddot{Y}, \tag{4.12}$$

where $\ddot{Y} = (\ddot{Y}_{11}, \cdots, \ddot{Y}_{NT})^\top$ is a $NT \times 1$vector while W_w and $\widetilde{Z}_w$ is a $NT \times NT$ and $NT \times d(1+q)$ matrix, respectively, such that

$$W_w = blockdiag\left(K_H\left(Z_{i1} - z\right) \prod_{\ell=2}^{T} K_H\left(Z_{i\ell} - z\right), \cdots, K_H\left(Z_{iT} - z\right) \prod_{\ell=1}^{T-1} K_H\left(Z_{i\ell} - z\right)\right)$$

and

$$\widetilde{Z}_w = \begin{bmatrix} \ddot{X}_{11}^\top & X_{11}^\top \otimes \left(Z_{11} - z\right)^\top - T^{-1} \sum_{s=1}^{T} X_{1s}^\top \otimes \left(Z_{1s} - z\right)^\top \\ \vdots & \vdots \\ \ddot{X}_{NT}^\top & X_{NT}^\top \otimes \left(Z_{NT} - z\right)^\top - T^{-1} \sum_{s=1}^{T} X_{Ns}^\top \otimes \left(Z_{Ns} - z\right)^\top \end{bmatrix}.$$

The local weighted linear least-squares estimator for $m(z)$ of a regression in deviation from the mean is then defined as

$$\widehat{m}_w\left(z; H\right) = e_1^\top \left(\widetilde{Z}_w^\top W_w \widetilde{Z}_w\right)^{-1} \widetilde{Z}_w^\top W_w \ddot{Y}. \tag{4.13}$$

Note that for the sake of simplicity we use the same bandwidth matrix for these two estimators. As it is well-known in the nonparametric literature, the optimal bandwidth matrix H can be obtained using several standard procedures such as, for example, the residual squares criterion proposed in Fan and Gijbels (1995a). Then, for empirical applications we must not forget that although the resulting bandwidths are very close, they are different.

Once obtained the nonparametric estimators for both a first-differences model and a regression in deviation from the mean, the next step is to establish the behavior of the two estimators in large samples. Under some standard assumptions collected in the previous chapters, their asymptotic distributions are derived in the next theorems. Assumption **2.1**, or **3.1**, characterizes the data-generating process for a panel data model. Assumption **2.2**, or **3.2**, is a standard strict exogeneity condition and **2.3**, or **3.3**, imposes the so-called fixed effects. In addition, for conditional moments, densities and kernel functions we need some smoothness and boundedness conditions that are collected in Assumptions **2.4-2.5**, **2.7-2.9**, and **3.4-3.9**. Finally, Assumptions **2.6** and **3.10** are required to show that the Lyapunov condition holds.

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT})$ and $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the observed covariates vectors. We denote by $\mathcal{H}_{m_r}(z)$ the Hessian matrix of the rth component of $m(\cdot)$, for $r = 1, \cdots, d$, whereas $diag_d\left(tr\left(\mathcal{H}_{m_r}(z)H\sqrt{NT|H|}\right)\right)$ stands for a diagonal matrix of elements of $tr\left(\mathcal{H}_{m_r}(z)H\sqrt{NT|H|}\right)$ and $\imath_d$ a $d \times 1$ unit vector. Furthermore, $R(K) = \int K^2(u)du$, and $f_{Z_{it},Z_{i(t-1)}}(z,z)$ is the probability density function of the random variable $\left(Z_{it}, Z_{i(t-1)}\right)$. We denote by $f_{Z_{i1},\cdots,Z_{iT}}(z, \cdots, z)$ the probability density functions of $(Z_{i1}, \cdots, Z_{iT})$ evaluated at point z.

In this context, in Chapter 2 it is shown the following result for the locally weighted least-squares first-differences estimator (4.9),

Theorem 2.2 *Assume conditions 2.1-2.9 hold. Then,*

$$\sqrt{NT|H|}\left(\widehat{m}_F(z;H) - m(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b_F(z), \upsilon_F(z)\right),$$

as N tends to infinity and T is fixed, where

$$\begin{aligned} b_F(z) &= \frac{1}{2}\mu_2(K)\, diag_d\left(\mathrm{tr}\left(\mathcal{H}_{mr}(z)H\sqrt{NT|H|}\right)\right)\imath_d, \\ \upsilon_F(z) &= 2\sigma_v^2 R^2(K)\,\mathcal{B}^{-1}_{\Delta X\Delta X}(z,z) \end{aligned}$$

and

$$\mathcal{B}_{\Delta X\Delta X}(z,z) = E\left[\Delta X_{it}\Delta X_{it}^{\top}|Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it},Z_{i(t-1)}}(z,z).$$

On the other hand, for the locally weighted least-squares fixed effects estimator (4.12), in Chapter 3 we obtain the following asymptotic properties

Theorem 3.2 *Assume conditions 3.1-3.10 hold, then as* $N \to \infty$ *and* T *remains to be fixed we obtain*

$$\sqrt{NT|H|^{T/2}}\left(\widehat{m}_w(z;H) - m(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b_w(z), \upsilon_w(z)\right),$$

where

$$\begin{aligned} b_w(z) &= \frac{1}{2}\mu_2(K) diag_d \left(tr\left(\mathcal{H}_{m_r}(z) H \sqrt{NT|H|^{T/2}} \right) \right) \imath_d, \\ \upsilon_w(z) &= \sigma_v^2 R^T(K)\, \mathcal{B}^{-1}_{\ddot{X}\ddot{X}}(z, \cdots, z) \end{aligned}$$

and

$$\mathcal{B}_{\ddot{X}\ddot{X}}(z, \cdots, z) = E\left[\ddot{X}_{it}\ddot{X}_{it}^\top | Z_{i1} = z, \cdots, Z_{iT} = z\right] f_{Z_{i1},\cdots,Z_{iT}}(z, \cdots, z).$$

As we have already pointed out in Theorems 2.2 and 3.2, the use of a higher dimensional kernel weight enables us to solve the problem of non-negligible asymptotic bias. It provides local linear estimators with a bias term of the same order as the standard results, $O_p(tr(H))$. However, as it is usual in the nonparametric techniques any attempt to reduce the bias is offset by an enlargement of the variance term. Thus, these two estimators are consistent but exhibit a suboptimal rate of convergence. Note that the standard rate of this type of problems is $NT|H|^{1/2}$. The first-differences estimator exhibits a rate of order $NT|H|$ and the estimator based in differences from the mean shows a rate of order $NT|H|^{T/2}$.

4.3 One-step backfitting procedure

In this section we analyze alternative procedures to provide nonparametric estimators that exhibit the optimal rate of convergence of such problems. Firstly, we focus on the first differences transformation. Later, we present the corresponding estimator for a regression in deviation from the mean. We conclude with a comparison between the asymptotic properties of the resulting estimators.

As it is noted in Fan and Zhang (1999), the variance can be reduced by further smoothing but the bias cannot be reduced by any kind of smoothing. Thus, in order to achieve optimality we propose to combine previous estimators with a one-step backfitting algorithm.

In this way, this estimation strategy allows us to exploit the additive structure of the model in order to cancel asymptotically one of the two additive terms of the model.

Let $\widehat{m}_F\left(z;H\right)$ be the first-step local weighted linear least-squares first-differences estimator (4.8) and define the variable ΔY_{it}^b such that

$$\Delta Y_{it}^b = \Delta Y_{it} + X_{it}^\top \widehat{m}\left(Z_{i(t-1)};H\right), \tag{4.14}$$

and replace (4.2) in this previous equation obtaining

$$\Delta Y_{it}^b = X_{it}^\top m\left(Z_{it}\right) + \Delta v_{it}^b, \quad i=1,\cdots,N \quad ; \quad t=2,\cdots,T, \tag{4.15}$$

where the composed error term has the form

$$\Delta v_{it}^b = X_{it}^T\left(\widehat{m}\left(Z_{i(t-1)};H\right) - m\left(Z_{i(t-1)}\right)\right) + \Delta v_{it}.$$

By the same reasoning as before, the quantities of interest of (4.15) can be estimated as a solution for γ_F to the following locally weighted linear regression

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it}^b - X_{it}^T\gamma_{F_0} - X_{it}^\top \otimes \left(Z_{it}-z\right)^\top \gamma_{F_1}\right)^2 K_{\widetilde{H}}\left(Z_{it}-z\right), \tag{4.16}$$

where $\widetilde{H}$ is a $q\times q$ symmetric positive definite bandwidth matrix of this step. We denote by $\widetilde{\gamma}_F = (\widetilde{\gamma}_{F_0}^\top \quad \widetilde{\gamma}_{F_1}^\top)^\top$ a $d(1+q)$-vector that minimizes the expression (4.16).

Assuming $\widetilde{Z}_F^{b\top} W_F^b \widetilde{Z}_F^b$ is a nonsingular matrix, we suggest as estimators for $m(z)$ and $D_m(z)$, $\widetilde{m}_F(z;\widetilde{H}) = \widetilde{\gamma}_{F_0}$ and $vec(\widetilde{D}_{F_m}(z;\widetilde{H})) = \widetilde{\gamma}_{F_1}$, respectively,

$$\widetilde{m}_F(z;\widetilde{H}) = \widetilde{\gamma}_{F_0} = e_1^\top\left(\widetilde{Z}_F^{b\top} W_F^b \widetilde{Z}_F^b\right)^{-1}\widetilde{Z}_F^{b\top} W_F^b \Delta Y^b, \tag{4.17}$$

where $\Delta Y^b = \left(\Delta Y_{12}^b,\cdots,\Delta Y_{NT}^b\right)^\top$, $W_F^b = diag\left(K_{\widetilde{H}}\left(Z_{it}-z\right),\cdots,K_{\widetilde{H}}\left(Z_{NT}-z\right)\right)$ and

$$\widetilde{Z}_b = \begin{bmatrix} X_{12}^\top & X_{12}^\top \otimes \left(Z_{12}-z\right)^\top \\ \vdots & \vdots \\ X_{NT}^\top & X_{NT}^\top \otimes \left(Z_{NT}-z\right)^\top \end{bmatrix}.$$

On the other hand, and following a similar procedure as before, we propose a backfitting estimator for a mean deviation regression such as (4.3). Let $\widehat{m}_w\left(z;H\right)$ be the first-step fixed effects estimator proposed in (4.11), they define $\ddot{Y}_{it}^{b} = \ddot{Y}_{it} - T^{-1}\sum_{s=1}^{T} X_{is}^{\top}\widehat{m}_w(Z_{is}; H)$ and replace $\ddot{Y}_{it}$ by (4.3) obtaining

$$\ddot{Y}_{it}^{b} = X_{it}^{\top} m(Z_{it}) + \ddot{v}_{it}^{b}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{4.18}$$

where the error term is

$$\ddot{v}_{it}^{b} = \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top}\left(\widehat{m}_w(Z_{is}; H) - m(Z_{is})\right) + \ddot{v}_{it}.$$

Denote by $\widetilde{\gamma}_w = (\widetilde{\gamma}_{w_0}^{\top} \quad \widetilde{\gamma}_{w_1}^{\top})^{\top}$ the $d(1+q)$-vector that minimizes the following problem

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}^{b} - X_{it}^{\top}\gamma_{w_0} - X_{it}^{\top}\otimes(Z_{it}-z)^{\top}\gamma_{w_1}\right)^2 K_{\widetilde{H}}(Z_{it}-z), \tag{4.19}$$

we propose as estimator for $m(z)$ and $D_m(z)$, $\widetilde{m}_w(z;\widetilde{H}) = \widetilde{\gamma}_{w_0}$ and $vec(\widetilde{D}_{w_m}(z;\widetilde{H})) = \widetilde{\gamma}_{w_1}$, respectively, of the form

$$\widetilde{m}_w(z;\widetilde{H}) = \widetilde{\gamma}_{w_1} = e_1^{\top}\left(\widetilde{Z}_w^{b\top} W_w^{b} \widetilde{Z}_w^{b}\right)^{-1} \widetilde{Z}_w^{b\top} W_w^{b} \ddot{Y}^{b}, \tag{4.20}$$

where $\ddot{Y}^{b} = (\ddot{Y}_{11}^{b}, \cdots, \ddot{Y}_{NT}^{b})^{\top}$, $W_w^{b} = diag\left(K_{\widetilde{H}}(Z_{11}-z), \cdots, K_{\widetilde{H}}(Z_{NT}-z)\right)$ and

$$\widetilde{Z}_w^{b} = \begin{bmatrix} X_{11}^{\top} & X_{11}^{\top}\otimes(Z_{11}-z)^{\top} \\ \vdots & \vdots \\ X_{NT}^{\top} & X_{NT}^{\top}\otimes(Z_{NT}-z)^{\top} \end{bmatrix}.$$

In order to show the asymptotic efficiency of these two backfitting estimators, we need the sampling scheme conditions established in Assumptions **2.1-2.3**, or **3.1-3.3**, and the smoothness and boundedness conditions already considered in Assumptions **2.4-2.8** and **3.4-3.8**. Furthermore, as they are obtained via a one-step backfitting algorithm we need to ensure that both bias and variance rates of the first-step estimates, $\widehat{m}_F(z;H)$ and $\widehat{m}_w(z;H)$, are uniform. Therefore, following Masry (1996) we impose some assumptions about the bandwidth $\widetilde{H}$ and its relationship with H. This is already considered in Assumptions **2.10-2.11** and **3.11-3.12**.

Let $\mathrm{diag}_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H})\right)$ be the diagonal matrix of elements $\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H})$ and i_d a $d\times 1$ unit vector, in Chapter 2 obtain the following asymptotic expressions for the backfitting first-differences estimator (4.17),

Theorem 2.3 *Assume conditions 2.1-2.8 and 2.10-2.11 holds, then, as* N *tends to infinity and* T *is fixed we obtain*

$$E\left(\widetilde{m}_F(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right)-m(z) \quad = \quad \frac{1}{2}\mu_2\left(K\right)\mathrm{diag}_d\left(\mathrm{tr}(\mathcal{H}_{m_r}\left(z\right)\widetilde{H})\right)\imath_d+o_p(\mathrm{tr}(\widetilde{H}))$$

and

$$Var\left(\widetilde{m}_F(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right) \quad = \quad \frac{2\sigma_v^2 R\left(K\right)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{XX}^{-1}\left(z\right)\left(1+o_p(1)\right),$$

where

$$\mathcal{B}_{XX}\left(z\right)=E\left[X_{it}X_{it}^{\top}|Z_{it}=z\right]f_{Z_{it}}(z).$$

Under similar conditions, in Chapter 3 it is proved the following asymptotic results for the backfitting fixed effects estimator (4.17):

Theorem 3.3 *Assume conditions 3.1-3.3, 3.4-3.9 and 3.11-3.12 hold, then as* $N\to\infty$ *and* T *remains to be fixed we obtain*

$$E\left(\widetilde{m}_w(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right)-m(z) \quad = \quad \frac{1}{2}\mu_2(K)diag_d\left(tr(\mathcal{H}_{m_r}(z)\widetilde{H})\right)\imath_d+o_p(tr(\widetilde{H}))$$

and

$$Var\left(\widetilde{m}_w(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right) \quad = \quad \frac{\sigma_v^2 R(K)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{XX}^{-1}(z)\mathcal{B}_{\ddot{X}\ddot{X}}(z)\mathcal{B}_{XX}(z)^{-1}\left(1+o_p(1)\right),$$

where

$$\begin{aligned}\mathcal{B}_{XX}\left(z\right) &= E\left[X_{it}X_{it}^{\top}|Z_{it}=z\right]f_{Z_{it}}(z),\\ \mathcal{B}_{\ddot{X}\ddot{X}}\left(z\right) &= E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{it}=z\right]f_{Z_{it}}(z).\end{aligned}$$

With the aim of itemizing the asymptotic behavior of these two backfitting estimators, $\widetilde{m}_F(z;\widetilde{H})$ and $\widetilde{m}_w(z;\widetilde{H})$, we analyze in detail the bias and variance-covariance matrix of Theorems 2.3 and 3.3. On one hand, in both cases the conditional bias is very close to the standard one of the local polynomial regression estimates. Thus, as each entry of $\mathcal{H}_{m_r}(z)$ is a measure of the curvature of $m(\cdot)$ at z in a particular direction, we can intuitively conclude that these estimators show a higher conditional bias as far as the unknown function exhibits a higher curvature and more smoothness. On the other hand, regarding to the conditional variance we observe that both estimators achieve the optimal rate of convergence, but they show different constants. Thus, while the first-differences estimator exhibits a variance-covariance matrix which increases when the smoothness becomes lower or the data becomes sparse near z, the conditional variance of the fixed effects estimator is also influenced by the time-demeaned covariates $\mathcal{B}_{\ddot{X}\ddot{X}}(z)$.

In this way, it is shown that direct estimation techniques allow obtaining estimators with different rates of convergence that depend on the type of differencing transformation. Meanwhile, one-step backfitting procedures provide estimators that achieve the optimal rate of convergence for both transformations. In this situation, the rate of convergence should not be used as an efficiency criterion between both backfitting estimators and, in order to analyze efficiency, it is necessary to study their finite sample behavior.

4.4 Monte Carlo experiment

In this section, we conduct an extensive Monte Carlo simulation experiment with the aim of comparing the small sample behavior of both first-differences and fixed effects (time-demeaned) nonparametric estimators introduced in Sections 4.2 and 4.3. In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both estimators is going to depend on the stochastic structure of the v_{it}'s random errors. Therefore, a first idea for our simulation would be to check whether this behavior is also fulfilled in the nonparametric case. In order to do so, we will consider a model like (4.1) with three different types of idiosyncratic errors: random walk structure, *i.i.d.* errors and an AR(1) process with correlation parameter equal to 0.5.

However, in the nonparametric setting, apart from the previous issues other factors such as dimension of q and sizes of T and N, are of great interest. In particular, we are interested in learning whether, for different values of q, N and T, the results obtained for different types of idiosyncratic errors still hold. That is, we want to investigate the average mean square error (AMSE) behavior of our estimators when facing the course of dimensionality problem under different specifications of the error term. Furthermore, as T becomes larger the performance of our estimators is another issue of great interest in this context. In particular, as it appears in Theorems 2.2 and 3.2, the asymptotic bound for the variance term of the first-differences estimator is $1/NT|H|$, whereas the corresponding term for the fixed effects estimator is $1/NT|H|^{T/2}$. Therefore, one might expect a different behavior between both estimators as T increases. On the opposite, for the one-step backfitting estimators proposed in Section 4.3 the performance of both estimators should be affected in the same direction by T because the asymptotic bounds for the variance are now the same, i.e., $1/NT|H|^{1/2}$. Finally, it is also of interest to check whether the performance of the one-step backfitting estimators is, as expected, better than the corresponding for the local linear regression estimators in small sample sizes.

In this context, we first investigate the properties in finite samples of the two local linear regression estimators proposed in Section 4.2 and make a comparison between them. Secondly, a similar comparative analysis is made for the one-step backfitting estimators that we present in Section 4.3.

Observations are generated from the following varying coefficient panel data model

$$Y_{it} = X_{dit}^{\top} m\left(Z_{qit}\right) + \mu_{qi} + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \quad d, q = 1, 2, \tag{4.21}$$

where X_{dit} and Z_{qit} are random variables generated such that $X_{dit} = 0.5\zeta_{dit} + 0.5\xi_{dit}$ (ζ_{1it} and ζ_{2it} are *i.i.d.* $N(0,1)$), $Z_{qit} = \omega_{qit} + \omega_{qi(t-1)}$ (ω_{1it} and ω_{2it} are *i.i.d.* $N(0,1)$) and we consider three different cases of study:

$$\begin{aligned}
(q = 1 \ d = 1): \quad & Y_{it} = X_{1it} m\left(Z_{1it}\right) + \mu_{1i} + v_{it}, \\
(q = 2 \ d = 1): \quad & Y_{it} = X_{1it} m\left(Z_{1it}, Z_{2it}\right) + \mu_{2i} + v_{it}, \\
(q = 1 \ d = 2): \quad & Y_{it} = X_{1it} m\left(Z_{1it}\right) + X_{2it} m\left(Z_{1it}\right) + \mu_{1i} + v_{1it},
\end{aligned}$$

where the chosen functional forms are $m_1\left(Z_{1it}\right) = sin\left(Z_{1it} * \Pi\right)$, $m_2\left(Z_{1it}\right) = \exp\left(-Z_{1it}^2\right)$, $m_1\left(Z_{1it}, Z_{2it}\right) = \sin\left(\frac{1}{2}\left(Z_{1it} + Z_{2it}\right) * \Pi\right)$.

By treating the cross-sectional heterogeneity as the fixed effect, we allow that the individual effects can be correlated with one or more of the covariates. In particular, the dependence between μ_{qi} and Z_{qit} is imposed by generating $\mu_{qi} = c_0 \overline{Z}_{i\cdot} + u_i$ and $Z_{i\cdot} = (T\kappa)^{-1} \sum_{q=1}^{\kappa} \sum_{t=1}^{T} Z_{qit}$, where u_i is an *i.i.d.* $N(0,1)$ random variable and $i = 2, \cdots, N$. The correlation between the fixed effects and some of the explanatory variables of the model is controlled by $c_0 = 0.5$. Also, let ϵ_{it} be an *i.i.d.* $N(0,1)$ and v_{it} a scalar random variable, for each model we work with the following three different specification of the error term:

a) $v_{it} = \epsilon_{it}$;

b) v_{it} follows a random walk, such as $v_{it} = 1 + v_{i(t-1)} + \epsilon_{it}$;

c) v_{it} is generated as stationary AR(1) process of the form $v_{it} = \rho v_{i(t-1)} + \epsilon_{it}$.

In each experiment we use 1000 Monte Carlo replications Q. The number of period T is varied to be 3 and 5, whereas the number of cross-sections N takes the values 50, 100 and 150. For the calculations we use a Gaussian kernel and the bandwidth is chosen as $\widehat{H} = \widehat{h}I$, and $\widehat{h} = \widehat{\sigma}_z(NT)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{qit}\}_{i=1,t=1}^{N,T}$.

In order to state the performance of the first-differences and fixed effects estimator, we use the mean square error (MSE) as a measure of their estimation accuracy. Thus, we denote by the φth replication by the subscript φ,

$$MSE(\widehat{m}(z;\widehat{H})) = \frac{1}{Q}\sum_{\varphi=1}^{Q} E\left[\left(\sum_{r=1}^{d}(\widehat{m}_{\varphi r}(z;\widehat{H}) - m_{\ell r}(z))X_{it,\varphi r}\right)^2\right],$$

which can be approximated by the averaged mean squared error (AMSE),

$$AMSE(\widehat{m}(z;\widehat{H})) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{r=1}^{d}(\widehat{m}_{\varphi r}(z;\widehat{H}) - m_{\varphi r}(z))X_{it,\varphi r}\right)^2.$$

4.4.1 Local linear estimator: simulation results

With the aim of analyzing the finite sample behavior of the two local linear regression estimators, $\widehat{m}_F(z;H)$ and $\widehat{m}_w(z;H)$, proposed in Section 4.2, the following tables summarize the results of the simulation. Specifically, Tables 4.1-4.3 contain the AMSE obtained for each of the three varying coefficient specifications proposed in the simulation when the error term is *i.i.d.* Hence, in Table 4.2 we show the impact of the course of dimensionality $(q = 2, d = 1)$, and Table 4.3 shows the possible impact of a larger dimension in the number of linear explanatory variables $(q = 1, d = 2)$. In each table, we present results for $T = 3, 5$ and $N = 50, 100, 150$. Note that our asymptotic results hold as N becomes larger whereas T is kept fixed. With the same structure, Tables 4.4-4.6 summarize the results when the error term follows a random walk, whereas Tables 4.7-4.9 present the results when the idiosyncratic error term is generated as an AR(1) stationary process.

In Tables 4.1-4.3, according to our developments and standard results in fully parametric settings, the fixed effects estimator should perform better than the first-differences one. A quick look to these tables confirms in general our theoretical findings. Furthermore, as expected, as N increases all AMSEs tend to zero and the rates between them are not equal. However, we point out that, the fixed effects estimator is much sensitive to T than the other estimator. As we can realize in all tables, as T increases the relative performance of the fixed effects estimator becomes worse. In Table 4.1 for example, the relative AMSE defined as $AMSE\left(\widehat{m}_F(z;\widehat{H})\right)/AMSE\left(\widehat{m}_w(z;\widehat{H})\right)$, for $N = 150$, goes from 4.46, for $T = 3$, to 3.33, for $T = 5$. This effect can be explained in terms of the asymptotic bounds of both estimators. Moreover, as it was expected from these bounds, when $q = 2$, in Table 4.2, we realize that the relative performance of the fixed effects estimator is even worse when T increases. Keeping $N = 150$, the relative performance falls from 1.69, for $T = 3$, to 0.99, for $T = 5$. Finally, the results shown in Table 4.3 indicate that the dimension d of the vector of covariates X does not affect the asymptotic behavior of the estimators.

Table 4.1. AMSE for $d = 1$ and $q = 1$ when v_{it} is $N.I.D.(0, 1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.50140	0.26676	1.87959
	N = 100	0.45282	0.16728	2.70696
	N = 150	0.43055	0.09647	4.46304
T=5	N = 50	0.43100	0.23248	1.85392
	N = 100	0.41523	0.16562	2.50712
	N = 150	0.41112	0.12333	3.33349

In Tables 4.4-4.6, we show the simulation results for the random walk case. Within this data-generating process, the first-differences estimator should show its better performance. This is true, if we compare the AMSE of the first-differences estimator in Tables 4.1, 4.2 and 4.3 against their counterparts in Tables 4.4, 4.5 and 4.6. In all cases, the AMSE is smaller when the idiosyncratic errors are generated as a random walk. However, if we compare results in Table 4.2 against Table 4.5, we point out that the course of dimensionality affects

Table 4.2. AMSE for $d = 1$ and $q = 2$ when v_{it} is $N.I.D.(0,1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.87609	0.71333	1.22817
	N = 100	0.74099	0.50743	1.46028
	N = 150	0.67584	0.40031	1.68829
T=5	N = 50	0.58338	0.85925	0.67894
	N = 100	0.53208	0.64803	0.82107
	N = 150	0.49571	0.49681	0.99785

Table 4.3. AMSE for $d = 2$ and $q = 1$ when v_{it} is $N.I.D.(0,1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.67624	0.40464	1.66232
	N = 100	0.55552	0.26768	2.07531
	N = 150	0.50958	0.17777	2.86651
T=5	N = 50	0.52788	0.53855	0.98019
	N = 100	0.48977	0.41464	1.18119
	N = 150	0.47359	0.27039	1.75151

negatively the performance of the first-differences estimator. On the contrary, the fixed effects estimator, as expected from our results, worsens its performance compared against the *i.i.d.* setting. It is also seriously affected by the course of dimensionality, as the first-differences estimator. If we compare the relative performance between the estimators, in Table 4.4, the relative AMSE for $N = 150$ goes from 3.23, for $T = 3$, to 3.38 for $T = 5$. In Table 4.5, this relative AMSE goes from 1.19, for $T = 3$, to 0.95, for $T = 5$. That is, in relative terms, the fixed effect estimator still performs better than the first-differences one, except if T becomes larger. This was already remarked in the previous setting but now this effect is much stronger. As in previous cases, the behavior of the fixed effects estimator is much more sensitive to changes in T than the other estimator. That is why, for large T, in the random walk setting the first-differences estimator shows a smaller

AMSE.

Table 4.4. AMSE for $d = 1$ and $q = 1$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.48321	0.34697	1.39266
	N = 100	0.44879	0.21847	2.05424
	N = 150	0.42673	0.13221	3.22767
T=5	N = 50	0.41596	0.21696	1.91722
	N = 100	0.40375	0.16199	2.49244
	N = 150	0.40429	0.11942	3.38545

Table 4.5. AMSE for $d = 1$ and $q = 2$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.80752	0.94763	0.85215
	N = 100	0.72200	0.70594	1.02275
	N = 150	0.65203	0.54788	1.19009
T=5	N = 50	0.50191	0.91147	0.55066
	N = 100	0.47401	0.61333	0.77285
	N = 150	0.44549	0.46605	95588

In Tables 4.7-4.9 there are shown the AMSEs when the idiosyncratic errors are generated according to an AR(1) stationary structure. First, although in most cases, the fixed effects estimator performs better than the other one, its performance compared against the other settings is worse. This is very clear in the case of $q = 2$, (see Table 4.8). Again, we discover that the fixed effects estimator is much more sensitive to the size of T than the first-differences estimator. Moreover, the results in terms of AMSE of the first-differences estimator are better than the ones obtained in the random walk setting. This is somehow unexpected. As a summary, the results of this estimator are much more stable across different specifications of the error term. The same cannot be said about the fixed effects estimator. In fact, it performs quite well with $q = 1$ and under *i.i.d.* or an AR(1) stationary

Table 4.6. AMSE for $d = 2$ and $q = 1$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.63756	0.59151	1.07785
	N = 100	0.54387	0.43507	1.25007
	N = 150	0.50232	0.27174	1.84880
T=5	N = 50	0.49219	0.52216	0.94260
	N = 100	0.45677	0.38412	1.18913
	N = 150	0.45309	0.02615	17.3266

process, but it shows a much worse performance when the errors are generated following a random walk process. As we have already pointed out before, when T grows, the finite sample performance of these estimators worsens considerably.

Table 4.7. AMSE for $d = 1$ and $q = 1$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.45359	0.17422	2.60354
	N = 100	0.42577	0.11878	3.58452
	N = 150	0.41777	0.07595	5.50059
T=5	N = 50	0.42027	0.55042	0.76354
	N = 100	0.41011	0.44902	0.91334
	N = 150	0.40609	0.49980	1.35453

We finish this section by highlighting that as N increases, that is, asymptotically, the AMSE tends to converge for each estimator under different specifications of the error term. For small values of T, the AMSE of the first-differences estimator tends to dominate in terms of the AMSE of the fixed effects one. This can be also observed by looking at the relative AMSE values.

Table 4.8. AMSE for $d = 1$ and $q = 2$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.67023	0.60541	1.00796
	N = 100	0.60011	0.42938	1.39762
	N = 150	0.55613	0.34268	1.62288
T=5	N = 50	0.56328	2.15283	0.26165
	N = 100	0.50476	1.63113	0.30945
	N = 150	0.47522	1.29458	0.36708

Table 4.9. AMSE for $d = 2$ and $q = 1$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.58076	0.39258	1.47934
	N = 100	0.51289	0.22028	2.32835
	N = 150	0.48115	0.13734	3.50335
T=5	N = 50	0.51009	1.83876	0.27741
	N = 100	0.47487	1.23014	0.38603
	N = 150	0.45814	0.88428	0.51809

4.4.2 Backfitting estimator: simulation results

In this section, we show in Tables 4.10-4.18 the analogue results for the one-step backfitting estimators. In Tables 4.10-4.12 we compute the AMSE of the simulations for the first-differences and the fixed effects estimator under the *i.i.d.* setting. In Tables 4.13-4.15 we show the AMSE for the random walk setting and finally in Tables 4.16-4.18 we show all relevant values for the AR(1) stationary process. The bandwidth in the second-step is taken according to Silverman's rule of thumb.

In all error settings, the performance of the one-step backfitting is better than its cor-

Table 4.10. AMSE for $d = 1$ and $q = 1$ when v_{it} is $N.I.D.(0,1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.51583	0.08973	5.74869
	N = 100	0.46329	0.05449	8.50229
	N = 150	0.44410	0.03447	12.8837
T=5	N = 50	0.44147	0.05277	8.36593
	N = 100	0.42812	0.03485	12.2846
	N = 150	0.42790	0.02314	18.4918

Table 4.11. AMSE for $d = 1$ and $q = 2$ when v_{it} is $N.I.D.(0,1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.52639	0.25755	2.04384
	N = 100	0.40277	0.17059	2.36104
	N = 150	0.34501	0.12348	2.79405
T=5	N = 50	0.64658	0.16027	4.22150
	N = 100	0.53277	0.11141	4.78207
	N = 150	0.46933	0.07968	5.89018

Table 4.12. AMSE for $d = 2$ and $q = 1$ when v_{it} is $N.I.D.(0,1)$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.68699	14619	4.84854
	N = 100	0.56451	0.08886	6.35280
	N = 150	0.51538	0.05336	9.65854
T=5	N = 50	0.52992	0.08711	6.08334
	N = 100	0.49683	0.05554	8.94544
	N = 150	0.48044	0.03471	13.8415

Table 4.13. AMSE for $d = 1$ and $q = 1$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
	N = 50	0.52637	0.11840	4.44569
T=3	N = 100	0.46566	0.07394	6.29781
	N = 150	0.44353	0.04497	9.86279
	N = 50	0.43097	0.05039	8.55269
T=5	N = 100	0.42111	0.03325	12.6649
	N = 150	0.42257	0.02254	18.7475

Table 4.14. AMSE for $d = 1$ and $q = 2$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
	N = 50	0.51851	0.34453	1.50498
T=3	N = 100	0.39859	0.23508	1.68411
	N = 150	0.34249	0.16189	2.11557
	N = 50	0.57119	0.15036	3.79882
T=5	N = 100	0.47637	0.10662	4.46792
	N = 150	0.43928	0.07663	5.73248

Table 4.15. AMSE for $d = 2$ and $q = 1$ when v_{it} follows a random walk

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
	N = 50	0.69001	0.20968	3.29078
T=3	N = 100	0.56407	0.12945	4.35743
	N = 150	0.51588	0.07691	6.70758
	N = 50	0.50400	0.08082	6.23608
T=5	N = 100	0.46947	0.03380	13.8896
	N = 150	0.46781	0.03380	13.8405

Table 4.16. AMSE for $d = 1$ and $q = 1$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.47537	0.07127	6.66999
	N = 100	0.44203	0.04663	9.47952
	N = 150	0.43503	0.03056	14.2353
T=5	N = 50	0.44391	0.11273	3.98572
	N = 100	0.43182	0.07360	5.86712
	N = 150	0.42899	0.04329	9.90968

Table 4.17. AMSE for $d = 1$ and $q = 2$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.44274	0.21529	2.05648
	N = 100	0.36115	0.14795	2.44103
	N = 150	0.32366	0.10671	3.03308
T=5	N = 50	0.67840	0.364000	1.86374
	N = 100	0.52601	0.24939	2.10919
	N = 150	0.46826	0.17092	2.73964

Table 4.18. AMSE for $d = 2$ and $q = 1$ when v_{it} is $AR(1)$ with $\rho = 0.5$

		First-Differences estimator	Fixed Effects estimator	$\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$
T=3	N = 50	0.60729	0.13773	4.40928
	N = 100	0.52407	0.04489	11.6745
	N = 150	0.49471	0.04489	11.0205
T=5	N = 50	0.53384	0.23783	2.24463
	N = 100	0.49343	0.14064	3.50846
	N = 150	0.48116	0.08747	5.50086

respondent local linear regression estimators. However, the improvement is much bigger in the fixed effects case. As we expected also, the course of dimensionality is overridden. This is of course the main reason why we have applied this second stage estimation procedure. Furthermore, the rate at which the AMSE tends to zero for both estimators seems to be faster, according to the predictions of the asymptotic results. However, for small sample sizes, although both estimators do have the same rates of convergence the constants are different. This is considered in simulations under different scenarios of the error terms. Hence, under *i.i.d.* and an AR(1) stationary process the first-step backfitting algorithm of the fixed effect estimator performs better than the first-differences one. This can be realized by analyzing the relative AMSE. On the contrary, under the random walk specification, the performance is better in the opposite sense.

4.5 Conclusions

Recently, some new techniques have been proposed for the estimation of semi-parametric fixed effects varying coefficient panel data models. These new techniques fall within the class of the so-called differencing estimators. In particular, we consider first-differences and fixed effects local linear regression estimators. Analyzing their asymptotic properties it turns out that, keeping the same order of magnitude for the bias term, these estimators exhibit different asymptotic bounds for the variance. In both cases, the consequences are suboptimal nonparametric rates of convergence. In order to solve this problem, exploiting the additive structure of this model, a one-step backfitting algorithm is proposed. Under fairly general conditions, it turns out that the resulting estimators show optimal rates of convergence and exhibit the oracle efficiency property. Since both estimators are asymptotically equivalent, it is of interest to analyze their behavior in small sample sizes. In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both first-differences and fixed effects estimators is going to depend on the stochastic structure of the idiosyncratic random errors. However, in the nonparametric setting, apart from the previous issues other factors such as dimensionality or sample size are of great interest. In particular, we would be interested in learning about their relative average mean squared error (AMSE) under different scenarios. The simulation results basically confirm the theoretical findings for both local linear regression and one-step backfitting estimators. However, we have found out that fixed effects estimators are rather sensitive to the size of number of time observations.

Chapter 5

Precautionary savings over the life cycle: a two-step locally constant least-squares estimator

5.1 Introduction

Household save primarily for two reasons, to finance expenses after retirement (life-cycle motive) and to protect consumption against unexpected shocks (precautionary motive). In this situation, this article focuses on the estimation of a stochastic model of household's precautionary savings motivated by the life-cycle hypothesis model (LCH henceforth) of Modigliani and Brumberg (1954). Following the predictions of human capital theory, incomes rise in the early stages of working life by the effect of experience, while later they are reduced by the action of obsolescence and depreciation. Thus, the LCH model indicates that there is a hump-shaped age-saving profile since individuals tend to save from the middle of his life until retirement, and dissave in younger and older ages to maintain a constant level of utility in all periods. In this context, unforeseen adverse conditions and prudent behavior of the households make precautionary savings in a protection tool, finding a different behavior related to both the age of the individual and their risk aversion.

Uncertainty about the household's out-of-pocket medical expenses is a relevant issue for preventive savings as it is noted in Palumbo (1999). Thus, in the past two decades, there is a plethora of empirical studies that have sought to improve our understanding about optimal household consumption and its behavior under various sources of uncertainty; see Starr-McCluer (1996), Gruber (1997), Egen and Gruber (2001), Gertler and Gruber (2002) or Gourinchas and Parker (2002), among others. Nevertheless, many of these empirical studies have been criticized with regard to its lack of robustness against different types of misspecification.

The different sources of misspecification that they usually ignore can be grouped in three major weaknesses. First, the effect of the individual preferences is a relevant aspect when we try to model the household behavior but, unfortunately, most of these previous models omit this issue that appears in econometric models in the form of unobserved heterogeneity. Second, the decision of how much to spend in health-care products depends on some social and demographic features of the household. Considering this variable as exogenous may lead to inconsistent estimators. Finally, these models are usually based on a log-linearized Euler equation that is captured in the form of a fully parametric model and where precautionary savings are contained in the error term. However, such equations can be misleading because they are rather poor approximations of the marginal utility smoothing and they prevent the use of the age of the people as a key factor in their consumption decisions. Therefore, the resulting estimators may be suffering from an omitted variable bias and, as it is noted in Attanasio et al. (1999), we might be incorrectly assigning a significant portion of the hump-shape consumption over the life-cycle to demographic changes in the household, or if we analyze a model with little heterogeneity in preferences we could be assigning too much of the variation of this shape to precautionary savings.

In this way, it is shown how some restrictions concerning to structural relationships in savings go far beyond what is traditionally assumed by parametric models and we need to turn to more flexible procedures such as nonparametric and semi-parametric models; see Chou et al. (2004), Maynard and Qiu (2009), Gao et al. (2012) or Kuan and Chen (2013). However, although these studies can relax the assumptions related to the specification of the model, they do not face individual heterogeneity correlated with some covariates and endogenous variables are not allowed.

In this situation, the aim of this chapter is to contribute to the literature on precautionary savings, extending a semi-parametric method as the one developed in Chou et al. (2004) to the analysis of panel data models that address the main weaknesses of these latter studies. Specifically, what we propose is to estimate a LCH model that allows us to determine the behavior of the households under the following peculiarities: (i) unobserved individual heterogeneity correlated with some of the covariates; (ii) some of the functions that relate endogenous and explanatory variables are unknown in the Euler equation and need to be estimated; and (iii) health-care spending determined endogenously.

Starting from a nonparametric panel data varying coefficient regression model with fixed effects, we propose to estimate the unknown functions of interest through a very simple estimator based on a differencing technique. Furthermore, to cope with the endogeneity issue we perform the estimation procedure using the predicted values of the (possibly) endogenous variables that are generated from the nonparametric estimation of the instrumental variable (IV henceforth) reduced form equation. We establish that the resulting nonparametric two-stage estimator is consistent and asymptotically normal.

As we have stated in previous chapters, direct nonparametric estimation of differencing panel data models with fixed effects has been considered as rather cumbersome in the literature; see Su and Ullah (2011). To our knowledge, among differencing nonparametric estimators, in Henderson et al. (2008) it is proposed to solve this problem using profile likelihood techniques whereas in Su and Lu (2013) the estimator comes out as the solution of a second order Friedholm integral equation. Unfortunately, all these estimators are asymptotically biased in the presence of endogeneity. On the contrary, some IV methods have been proposed in the context of nonparametric panel data varying coefficient models with random effects. In particular, in Cai and Li (2008) it is proposed to estimate the nonparametric functions using the so-called nonparametric generalized method of moments, when endogeneity is allowed. However, note that this procedure do not control for heterogeneity that is correlated with some explanatory variables, so it renders to asymptotically biased estimators when fixed effects are present.

To our knowledge, this is the first technique that enables us to estimate directly the impact of different types of uncertainty on the behavior of households in a context of stochastic dynamic models of precautionary savings with endogenous variables and fixed effects without resorting to restrictive assumptions about functional forms, such as it is common with the well-known log-linearized version of the Euler equation. Furthermore, in contrast with some papers mentioned above, the asymptotic analysis that we perform in this chapter is based on the standard panel data framework, where the number of observations in time is fixed and the number of individuals grows with sample size.

Also, in order to show the feasibility and possible gains of this new procedure, we give a solution for the optimal consumption decision problem using a data set from the consumer expenditure continuous survey (ECPF henceforth) elaborated by the Spanish bureau of statistics (INE henceforth) for the period 1985(I)-1996(IV). Thus, the empirical application discussed in this chapter allows health-care expenditures to have a different impact on household behavior depending on the age-group they belong to, a topic of great relevance to political considerations for their important implications on the individuals welfare.

The structure of this chapter is as follows. Section 5.2 lays out the econometric model and the estimation procedure. In Section 5.3 we study its asymptotic properties and state how to calculate the confidence intervals. In Section 5.4 we illustrate an application with real data to investigate the finite sample performance of our proposed estimation procedure and present some simulation results. Finally, Section 5.5 concludes the chapter. The proofs of the main results are collected in the Appendix 3.

5.2 Conceptual framework

Along with liquidity constraints and habits in consumer preferences, uncertainty about possible economic hardships and household risk aversion are key determinants of household's consumption/saving decisions; see Friedman (1957). According to Eurostat data, health-care expenditures of U.S. households represent a 16.4% of their total consumption in 1986 and a 17% in 1996, while in Spain these expenses are about a 3.4% of the total in 1986 and 5.9% in 1996. The significant impact of health-care expenditures on household wealth, jointly with their persistent and increasing behavior with the age of the individuals, make them a significant part of this uncertainty and precautionary savings appears as an instrument of protection against potential income downturns or unforeseen out-of-pocket medical expenses in the latter stages of life, see Chou et al. (2003). In this context, the aim of this section is to analyze the relationship between the marginal propensity to consume and the age of the main household breadwinner by incorporating uncertainty in the LCH model.

To this end, we solve the basic problem of the consumer i at the time t in the presence of different types of uncertainty, following for example Blanchard and Fisher (1989) and Deaton (1992), but relaxing some assumptions about functional forms and, at the same time, allowing the presence of some endogenous variables. Thus, the system of equations that we consider is

$$\begin{aligned} Y &= \alpha(Z) + W^{\top} m_1(Z) + U^{\top} m_2(Z) + \mu + v, \\ W &= g(X) + \zeta + \xi, \end{aligned} \tag{5.1}$$

where v and ξ are idiosyncratic error terms that are correlated between them. The individual preferences, denoted by μ and ζ, are unobserved and correlated with some/all covariates of the system while both precautionary savings Y and health-care expenditures W are endogenously determined. As we explain in the following, we assume Y is function of the age of the main household breadwinner Z, health-care expenditures W and some other explanatory variables of the households U, whereas W depends on a set of household demographic features X.

To understand the stochastic life-cycle model that we consider in this chapter, expression (5.1), it is assumed that individuals live T periods and they work during the first ones, $T-1$. In addition, in each work period t they receive a stochastic income I_{it} and incur in an out-of-pocket health-care expenditure M_{it}. If M_{it} were known, households would decide how to spend their income between consumption C_{it} and financial wealth A_{it}, maximizing an additively time-separable utility that has a positive third-order derivative ($U'''(\cdot) > 0$). Note that this is the necessary condition for precautionary saving because, when the marginal utility function is convex, increases in the level of uncertainty about future consumption will cause a reduction in current consumption and an increase in savings; see Blanchard and Fisher (1989) for further details. According to Caballero (1990), we use a negative exponential utility function by assuming that the degree of absolute risk aversion and absolute prudence are both constant and equal to α, so household i maximize the following problem at the moment 0,

$$\max_{C_{it}} E_0 \left[\sum_{i=1}^{N} \sum_{t=0}^{T-1} \left(-\frac{1}{\alpha} \right) exp(-\alpha C_{it}) \right], \tag{5.2}$$

subject to

$$\begin{aligned} A_{i(t+1)} &= A_{it} + I_{it} - M_{it} - C_{it}, \\ M_{it} &= M_{i(t-1)} + \epsilon_{it} \quad ; \quad \epsilon_{it} \sim \mathcal{N}\left(0, \sigma^2\right), \end{aligned} \tag{5.3}$$

where health-care expenditures are modeled as a random walk. For the sake of simplicity, we assume there do not exist liquidity constraints so both the discount and the interest rate are both equal to zero.

If we take first-order conditions in (5.2) and use the fact that $E[exp(C_{it})] = exp[E(C) + \sigma^2/2]$, when C_{it} is normally distributed with mean $E(C)$ and variance σ^2, the expected consumption is

$$C_{i(t+1)} = C_{it} + \frac{\alpha\sigma^2}{2} + \epsilon_{i(t+1)}, \tag{5.4}$$

and combining this result with the inter-temporal budget restriction (5.3) we obtain the optimal level of consumption, i.e.

$$C_{it} = \frac{1}{T-t} A_{it} + (I_{it} - M_{it}) - \frac{\alpha(T-t-1)\sigma^2}{4}, \tag{5.5}$$

where $(I_{it} - M_{it} - C_{it})$ is the precautionary savings and $\alpha(T-t-1)\sigma^2/4$ reflects the uncertainty effect. Note that here we do not enter in the debate about the importance of the retirement versus the bequests reasons. Thus, the measure of retirement that we use implicitly includes both.

Analyzing in detail the optimal consumption expression (5.5), we see that increases in uncertainty about future medical expenses (σ^2) or the degree of absolute prudence (α), together with a broad horizon of life ($T-t-1$), causes that agents behave as buffer-stock agents. Thus, C_{it} falls while $(I_{it} - M_{it} - C_{it})$ increases in order to have resources against potential adversities. On the contrary, adult households with high levels of accumulated wealth A_{it} have a high consumption, because their already reduced temporal horizon of life. On the other hand, focusing on the behavior of the expected consumption (5.4) we can see the uncertainty effect on the slope of the consumption path. Since this expression has a constant positive slope that depends on both α and σ^2, we can state that a higher level of σ^2 or α causes a steeper path consumption. Also, if there were no risk about unforeseen medical expenses ($\sigma^2 = 0$), consumption patterns change completely and household incentives to save seem to be more related to fears of potential income downturns.

As a result, these peculiarities together with the fact that the optimal choice problem of the consumer, (5.2) and (5.3), state that the optimal current consumption depends on the lifetime resources as well as on the expected rate of income growth and health-care resources so it is feasible to assume that the household behavior may vary with the age of the home; see Hubbard et al. (1994), Carroll (1997) or Deaton (1992). Thus, we can conclude that the cautious behavior of the households varies with a non-uniform function of the age of the house while both types of uncertainty, σ^2 and α, are age-dependent parameters. Therefore, the effect of different types of uncertainty on the household behavior can be analyzed through an extension of the proposal in Chou et al. (2004); i.e., the estimation of the equation (5.1) where Y represents the household savings (i.e., income I minus consumption C), Z is the age of the main household breadwinner, W the health-care expenditure and U the financial wealth measured, for example, via his permanent income. Remember that because household consumption decisions depend on a vector of their demographic features, such as the age of the main household breadwinner Z, the number of children, and so on, W is considered as an endogenous variable. Therefore, (5.1) states that household's savings are characterized by the risk aversion of the family, $\alpha(\cdot)$, uncertainty about future health-care expenses, $m_1(\cdot)$, and uncertainty about income downturns, $m_2(\cdot)$.

Note that the additive specification of precautionary savings specified in (5.1), but without considering the presence of endogenous variables, is well-known in the literature on the precautionary behavior of households. Many have been the empirical studies that have tried to examine the relationship between household precautionary savings and different sources of uncertainty, but without achieving conclusive results. On one hand, in Starr-McCluer (1996) and Egen and Gruber (2001) it is found that a reduction in the level of uncertainty, for example through unemployment insurance or public health programs, has a negative impact on savings. On the other hand, in Gruber (1997) and Gertler and Gruber (2002) it is established that these types of programs smooth the individual consumption.

Recently, in Chou et al. (2004) it is shown that public health programs have a negative effect on household savings, whereas in Kuan and Chen (2013) it is emphasized that these kind of programs impact more heavily on those households with higher incomes. On its part, in Gourinchas and Parker (2002) it is confirmed the patterns established by the LCH model and they find that, early in life, U.S. households behave as buffers-stock agents and accumulate wealth to unexpected income downturns, whereas around the age forty these savings are mainly for retirement and legacy. Also, in Cagetti (2003) it is determined the role of precautionary savings to explain the behavior of household wealth. It is found that this effect is particularly relevant at the beginning of household's life, whereas close to retirement savings are more related to the aim of the households of maintaining a constant level of utility in all periods of life.

However, as we state previously, despite the interesting results of these studies they ignore some sources of misspecification problems that can override their conclusions, such as the endogeneity issue of the household's consumption decisions or the unobserved individual heterogeneity. In this way, the method that we propose in this chapter attempts to overcome such problems and enables us to estimate the impact of both types of uncertainty on household's precautionary savings; i.e., household risk aversion and unforeseen health-care expenses.

5.3 Econometric model and estimation procedure

Once the economic model is formulated, we proceed to develop the estimation procedure for an extended version of (5.1). For this, we introduce the following sampling scheme for the data set that is standard in nonparametric panel data regression analysis.

Assumption 5.1 *Let* $(Y_{it}, U_{it}, W_{it}, Z_{it}, X_{it})_{i=1,\cdots,N;t=1,\cdots,T}$ *be a set of independent and identically distributed* $\mathbb{R}^{1+a+(M-1)+q+b}$*-random variables in the subscript* i *for each fixed* t *and strictly stationary over* t *for fixed* i.

Generalizing (5.1) to a multivariate panel data model and taking into account the endogeneity problem, the semi-parametric system to estimate is

$$\begin{aligned} Y_{it} &= \alpha(Z_{it}) + W_{it}^{\top} m_1(Z_{it}) + U_{it}^{\top} m_2(Z_{it}) + \mu_i + v_{it}, \\ W_{it} &= g(X_{it}) + \zeta_i + \xi_{it}, \end{aligned} \qquad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{5.6}$$

where $E(v_{it}|\mathbf{Z}, \mathbf{U}) = 0$, $E(\xi_{it}|\mathbf{X}) = 0$ and $E(v_{it}|\xi_{it}) \neq 0$.

Let $\mathbf{Z} = (Z_{11}, \cdots, Z_{NT})$, $\mathbf{X} = (X_{11}, \cdots, X_{NT})$, $\mathbf{U} = (U_{11}, \cdots, U_{NT})$, $\mathbf{W} = (W_{11}, \cdots, W_{NT})$, $\mathbf{Y} = (Y_{11}, \cdots, Y_{NT})$ be $NT \times 1$ vectors, we assume $E(\mu_i|\mathbf{Z}, \mathbf{W}, \mathbf{U}) \neq 0$ and $E(\zeta_i|\mathbf{X}) \neq 0$. Note that the conditional mean restrictions, $E(v_{it}|\mathbf{Z}, \mathbf{U}) = 0$ and $E(\xi_{it}|\mathbf{X}) = 0$, are standardized versions of the usual orthogonality condition of linear models. Also, Z is a subset of X.

As we state previously, the statistical dependence between μ_i and ζ_i and some/all covariates of the system will rend inconsistent direct estimators of the functions of interest. To avoid this situation, we take the standard solution of removing the fixed effects by the first difference transformation; i.e.,

$$\begin{aligned} \Delta Y_{it} &= \alpha(Z_{it}, Z_{i(t-1)}) + \left(W_{it}^{\top} m_1(Z_{it}) - W_{i(t-1)}^{\top} m_1(Z_{i(t-1)}) \right) \\ &+ \left(U_{it}^{\top} m_2(Z_{it}) - U_{i(t-1)}^{\top} m_2(Z_{i(t-1)}) \right) + \Delta v_{it}, \\ \Delta W_{it} &= g(X_{it}, X_{i(t-1)}) + \Delta \xi_{it}, \end{aligned} \tag{5.7}$$

for $i = 1, \cdots, N$ and $t = 2, \cdots, T$, where $\alpha(\cdot) : I\!R^{2q} \to I\!R$ and $g(\cdot) : I\!R^{2b} \to I\!R$ are additive functions, $\alpha(Z_{it}, Z_{i(t-1)}) = \alpha(Z_{it}) - \alpha(Z_{i(t-1)})$ and $g(X_{it}, X_{i(t-1)}) = g(X_{it}) - g(X_{i(t-1)})$, respectively. The coefficient functions $\alpha(\cdot)$, $m(\cdot)$ and $g(\cdot)$ are unknown for the researcher and need to be estimated.

To illustrate our estimation technique, let us consider the simplest case with $q = (M-1) = a = b = 1$. Following the Taylor expansion in (5.7), for any $z_1 \in \mathcal{A}$, being $\mathcal{A}$ a compact subset in a nonempty interior of $I\!R$,

$$\begin{aligned} \alpha(Z_{it}, Z_{i(t-1)}) &\approx \alpha'(z_1) \Delta Z_{it} + \frac{1}{2} \alpha''(z_1) \left((Z_{it} - z_1)^2 - (Z_{i(t-1)} - z_1)^2 \right) \\ &+ \cdots + \frac{1}{p!} \alpha^{(p)}(z_1) \left((Z_{it} - z_1)^p - (Z_{i(t-1)} - z_1)^p \right), \end{aligned}$$

$$\begin{aligned} W_{it}m_1(Z_{it}) - W_{it-1}m_1(Z_{i(t-1)}) &\approx m_1(z_1)\Delta W_{it} \\ &+ m_1'(z_1)\left(W_{it}(Z_{it}-z_1) - W_{i(t-1)}(Z_{i(t-1)}-z_1)\right) \\ &+ \frac{1}{2}m_1''(z_1)\left(W_{it}(Z_{it}-z_1)^2 - W_{i(t-1)}(Z_{i(t-1)}-z_1)^2\right) \\ &+ \cdots + \frac{1}{p!}m_1^{(p)}(z_1)\left(W_{it}(Z_{it}-z_1)^p - W_{i(t-1)}(Z_{i(t-1)}-z_1)^p\right), \end{aligned}$$

and similarly for $U_{it}m_2(Z_{it}) - U_{i(t-1)}m_2(Z_{i(t-1)})$.

These approximations suggest that we can estimate $\alpha'(z_1)$,$\cdots$,$\alpha^{(p)}(z_1)$, $m_1(z_1)$,$m_1'(z_1)$,$\cdots$, $m_1^{(p)}(z_1)$ and $m_2(z_1)$, $m_2'(z_1)$,$\cdots$,$m_2^{(p)}(z_1)$ by regressing ΔY_{it} on the terms $(Z_{it}-z_1)^{\lambda+1} - (Z_{i(t-1)}-z_1)^{\lambda+1}$, $W_{it}(Z_{it}-z_1)^{\lambda} - W_{i(t-1)}(Z_{i(t-1)}-z_1)^{\lambda}$ and $U_{it}(Z_{it}-z_1)^{\lambda} - U_{i(t-1)}(Z_{i(t-1)}-z_1)^{\lambda}$, for $\lambda = 0, 1, \cdots, p$ with kernel weights. Clearly, $m_1(\cdot)$ and $m_2(\cdot)$ are identified functions but $\alpha(\cdot)$ is not. This is due to the structure of the differencing procedure and this leads us to treat these components separately.

In order to estimate these smooth functions $m(\cdot)$, we propose as estimators $\beta_1 = m_1(z_1)$ and $\beta_2 = m_2(z_1)$ that are obtained minimizing the following locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{1it} - \Delta W_{it}^{\top}\beta_1 - \Delta U_{it}^{\top}\beta_2\right)^2 K_{H_2}\left(Z_{it}-z_1\right)K_{H_2}\left(Z_{i(t-1)}-z_1\right). \tag{5.8}$$

The resulting estimator is the so-called local constant regression estimator, also known as the Nadaraya-Watson estimator; see Nadaraya (1964), Watson (1964) or Fan and Gijbels (1995b). However, note that this new estimator exhibits the peculiarity that the kernel weights relate to both Z_{it} and $Z_{i(t-1)}$. If we considered kernels only around Z_{it} the remainder term in the Taylor's approximation would not be negligible, since the distance between Z_{is} $(s \neq t)$ and z_1 does not vanish asymptotically, and therefore the asymptotic bias would also be non-negligible. This problem was already pointed out in Lee and Mukherjee (2008) and solved in Chapter 2 in another context.

Unfortunately, although the resulting estimator of (5.8) is robust to fixed effects, is still biased due to endogeneity. In order to overcome this situation, the IV equation for (5.7) suggests to obtain a consistent nonparametric estimator of $g(X_{it}, X_{i(t-1)})$ and then replace $E(\Delta W_{it}|X_{it}, X_{i(t-1)}) = g(X_{it}, X_{i(t-1)})$ by $\widehat{g}(X_{it}, X_{i(t-1)})$ in the expression (5.8). Note that

$\widehat{g}(X_{it}, X_{i(t-1)}) = \widehat{g}_{it,i(t-1)}$ can be any standard nonparametric estimator such as the local linear regression estimator.

Once the substitution is done and denote by $\Delta\widehat{W}_{it} = (\widehat{g}_{it,i(t-1)}^{\top} \quad \Delta U_{it}^{\top})^{\top}$ a $d \times 1$ vector, where $d = (M-1)+a$, the estimators of the unknown functions $m(\cdot)$ that minimize the corresponding criterion function of (5.8) are grouped into the following vector,

$$\widehat{m}_{\widehat{g}}(z_1; H_2) = \begin{pmatrix} \widehat{m}_{1\widehat{g}}(z_1; H_2) \\ \widehat{m}_{2\widehat{g}}(z_1; H_2) \end{pmatrix} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_2}(Z_{it} - z_1) K_{H_2}(Z_{i(t-1)} - z_1) \Delta\widehat{W}_{it} \Delta\widehat{W}_{it}^{\top} \right)^{-1} \times \sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_2}(Z_{it} - z_1) K_{H_2}(Z_{i(t-1)} - z_1) \Delta\widehat{W}_{it} \Delta Y_{it}, \quad (5.9)$$

where H_2 is the $q \times q$ symmetric positive-definite bandwidth matrix of this second stage and K is a q-variable such that

$$K_H(u) = \frac{1}{|H|^{1/2}} K\left(H^{-1/2} u\right).$$

The intuition behind this transformation is the following. When $E\left(\Delta v_{it} | \mathbf{W}\right) \neq 0$, we cannot estimate consistently the unknown functions of the structural equation (5.7) by projecting ΔY_{it} over

$$\alpha(Z_{it}, Z_{i(t-1)}) + \left(W_{it}^{\top} m_1(Z_{it}) - W_{i(t-1)}^{\top} m_1(Z_{i(t-1)})\right) + \left(U_{it}^{\top} m_2(Z_{it}) - U_{i(t-1)}^{\top} m_2(Z_{i(t-1)})\right)$$

in the $\mathcal{L}_2(\mathbf{Z}, \mathbf{W}, \mathbf{U})$ projection space. Then, a standard solution is to use a $(M-1) \times 1$ vector of IV; i.e., $E(\Delta W_{it} | X_{it}, X_{i(t-1)}) = g_{it,i(t-1)}$, and multiply both sides of (5.1) by $g_{it,i(t-1)}$.

Taking conditional expectations over $(Z_{it}, Z_{i(t-1)})$, evaluated both in z_1, and rearranging terms we obtain

$$m(z_1) = E\left[\Delta\widetilde{W}_{it} g_{it,i(t-1)}^{\top} | Z_{it} = z_1, Z_{i(t-1)} = z_1\right]^{-1} E\left[g_{it,i(t-1)} \Delta Y_{it} | Z_{it} = z_1, Z_{i(t-1)} = z_1\right], \quad (5.10)$$

where we denote by $\Delta\widetilde{W}_{it} = (g_{it,i(t-1)}^{\top} \quad \Delta U_{it}^{\top})^{\top}$ and $m(z_1) = (m_1(z_1)^{\top} \quad m_2(z_1)^{\top})$ $d \times 1$ vectors and we assume that the matrix functions $E\left[\Delta U_{it} g_{it,i(t-1)}^{\top} | Z_{it} = z_1, Z_{i(t-1)} = z_1\right]$ and $E\left[g_{it,i(t-1)} g_{it,i(t-1)}^{\top} | Z_{it} = z_1, Z_{i(t-1)} = z_1\right]$ are definite positive.

As the reader can appreciate, this invertibility condition is a generalization of the well-known rank condition of parametric models with endogenous covariates that guarantees that $m(\cdot)$ is identified. In this way, it is easy to see why we propose the estimator in (5.10). Since the vector functions $g(X_{it}, X_{i(t-1)})$ is unknown to the researcher it has to be replaced by a consistent estimator; i.e., $\widehat{g}_{it,i(t-1)}$. Doing that and replacing the population moments by their corresponding sample moments we obtain directly (5.10), that is the same estimator that we would obtain if we substitute in (5.8) the endogenous variable ΔW_{it} by $\widehat{g}_{it,i(t-1)}$ and minimize the resulting criterion function with respect to $\beta = (m_1(z_1)^\top \quad m_2(z_1)^\top)^\top$. Note that these are the standard equivalence results between the optimal IV estimators and the so-called two-stage least-squares estimators in the fully linear parametric context that nicely appear again in this semi-parametric framework.

Once the estimators of the functional coefficients are proposed, let us now turn to the estimation of $\alpha(z_1)$. In order to do so, we define

$$\Delta\widetilde{Y}_{it} = \Delta Y_{it} - \Delta\widehat{W}_{it}^\top \widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}),$$

where $\widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)})$ is the two-stage local constant regression estimator defined in (5.9). By the equation (5.7) we obtain

$$\Delta\widetilde{Y}_{it} = \alpha(Z_{it}) - \alpha(Z_{i(t-1)}) + \Delta\widetilde{v}_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{5.11}$$

where

$$\Delta\widetilde{v}_{it} = \Delta v_{it} - \Delta\widehat{W}_{it}^\top \left(\widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}) - m(Z_{it}, Z_{i(t-1)})\right).$$

In this situation, we propose to estimate $\alpha(\cdot)$ using marginal integration techniques such as the developed in Hastie and Tibshirani (1990), Linton and Nielsen (1995) and Newey (1994).

We consider $\alpha(Z_{it}, Z_{i(t-1)}) = \alpha(Z_{it}) - \alpha(Z_{i(t-1)})$, so for any $z_1, z_2 \in \mathcal{A}$ we can estimate $\alpha(\cdot)$ by multivariate kernel smoothing methods obtaining

$$\widehat{\alpha}(z_1, z_2; H_3) = \frac{\sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{it} - z_1) K_{H_3}(Z_{i(t-1)} - z_2)\Delta\widetilde{Y}_{it}}{\sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{it} - z_1) K_{H_3}(Z_{i(t-1)} - z_2)}. \tag{5.12}$$

Finally, once obtained $\widehat{\alpha}(z_1, z_2; H_3)$, we obtain $\widehat{\alpha}\,(z_{11}; H_3)$ by the marginal integration of (5.12), i.e.,

$$\widehat{\alpha}(z_1; H_3) = \int_{\mathbb{R}} \widehat{\alpha}(z_1, z_2; H_3) q(z_2) dz_2. \tag{5.13}$$

If NT is large enough and $q(\cdot)$ is chosen as the density function of Z_{it}, we can use the sample version of (5.13) and then we can propose

$$\widehat{\alpha}(z_1; H_3) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\alpha}(z_1, Z_{it}; H_3). \tag{5.14}$$

Note that this technique has been already considered in the context of panel data models in Qian and Wang (2012).

5.3.1 Asymptotic properties

In this section, we investigate some asymptotic properties of the estimators proposed in the previous section. For this, we need the following assumptions.

Assumption 5.2 *Let $f_{X_{1t}}(\cdot)$ and $f_{Z_{1t}}(\cdot)$ be the probability density function of X_{it} and Z_{it}, respectively, all density functions are uniformly continuous in all their arguments and are bounded from above and below in any point of their support.*

Assumption 5.3 *The random errors, v_{it} and ξ_{it}, are independent and identically distributed, with zero mean and homoscedastic variances, $\sigma_v^2 < \infty$ and $E(\xi_{it}\xi_{it}^\top) = \sigma_\xi^2 I_{(M-1)} < \infty$. In addition, $E\left(\Delta v_{it} \Delta \xi_{it}\right) = \sigma_{v\xi}\imath_{(M-1)} < \infty$ is a $(M-1) \times 1$ covariates vector of the error terms of the M equations of the system. They are also independent of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{W}$, $\boldsymbol{U}$ for all i and t, but $E(v_{it}|\,\boldsymbol{W}) \neq 0$. Furthermore, $E|v_{it}|^{2+\delta} < \infty$ and $E|\xi_{it}|^{2+\delta} < \infty$, for some $\delta > 0$.*

Assumption 5.4 *Let z be an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $\alpha(\cdot)$ and $m_1(\cdot), \cdots, m_{d-1}(\cdot)$ are bounded and uniformly continuous and satisfy the Lipschitz condition. Furthermore, let (x_1, x_2) be interior points in the support of $f_{X_{1t}}$, all second-order derivatives of $g_1(\cdot,\cdot), \cdots, g_{M-1}(\cdot,\cdot)$ are continuous.*

Assumption 5.5 *The bandwidth matrices H_1 and H_2 are symmetric and strictly definite positive. Also, let h_1 and h_2 be each entry of the matrices H_1 and H_2, respectively, $h_1 \to 0$ and $h_2 \to 0$. As $N \to \infty$, $N|H_1| \to \infty$, $N|H_2| \to \infty$ and $N|H_1|/log(N) \to \infty$. Furthermore, $H_1 = o_p(H_2)$.*

Assumption 5.6 *let $\|A\| = \sqrt{tr\,(A^\top A)}$, then $E\left[\|\widetilde{W}_{it}\widetilde{W}_{it}^\top\|^2|Z_{it} = z_1, Z_{i(t-1)} = z_1\right]$ is bounded and uniformly continuous in its support. Furthermore, let*

$$\mathcal{X}_{it} = \begin{pmatrix} \widetilde{W}_{it}^\top & \widetilde{W}_{i(t-1)}^\top \end{pmatrix}^\top \qquad \text{and} \qquad \Delta\mathcal{X}_{it} = \begin{pmatrix} \Delta\widetilde{W}_{it}^\top & \Delta\widetilde{W}_{i(t-1)}^\top \end{pmatrix}^\top.$$

Also, the matrix functions $E\left[\mathcal{X}_{it}\mathcal{X}_{it}^\top|z_1, z_2\right]$, $E\left[\Delta\mathcal{X}_{it}\Delta\mathcal{X}_{it}^\top|z_1, z_2\right]$, $E\left[\mathcal{X}_{it}\mathcal{X}_{it}^\top|z_1, z_2, z_3\right]$ and $E\left[\Delta\mathcal{X}_{it}\Delta\mathcal{X}_{it}^\top|z_1, z_2, z_3\right]$ are bounded and uniformly continuous in their support.

Assumption 5.7 *The Kernel function K is a Gaussian kernel product based on univariate kernels, symmetric around zero and compactly supported. Also, the kernel is bounded such that $\int uu^\top K(u)du = \mu_2(K)I$ and $\int K^2(u)du = R(K)$, where $\mu_2(K)$ and $R(K)$ are scalars and I the identity matrix. In addition, all odd-order moments of K vanish, that is $\int u_1^{\iota_1} \cdots u_q^{\iota_q} K(u)du = 0$, for all nonnegative integers $\iota_1, \cdots, \iota_q$ such that their sum is odd.*

Assumption 5.8 *The moment function $E\left[\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^\top|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ is definite positive in any interior point (z_1, z_2) in the support of $f_{Z_{it},Z_{i(t-1)}}(z_1, z_2)$.*

Assumption 5.9 $E\left[|Y_{it}|^{2+\delta}|X_{it}, X_{i(t-1)}\right] < \infty$ *and* $E\left[|W_{it}|^{2+\delta}|X_{it}, X_{i(t-1)}\right]^{2+\delta} < \infty$, *for some $\delta > 0$.*

Assumption 5.10 *For some $\delta > 0$, $E\left[|\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^\top\Delta v_{it}\Delta\xi_{it}|^{1+\delta/2}|Z_{it} = z, Z_{i(t-1)} = z\right]$, $E\left[|\Delta\widetilde{W}_{it}\Delta v_{it}|^{2+\delta}|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, and $E\left[|\Delta\widetilde{W}_{it}\Delta\xi_{it}^\top|^{2+\delta}|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ are bounded and uniformly continuous in any point of their support.*

For the estimation of the fully nonlinear part of the third stage of the procedure, we need to impose further strong assumptions about the density functions than the usual Lipschitz continuity. Thus, Assumption **5.2** states that density functions are bounded from above and below and at least first-order partially differentiable with a Lipschitz-continuous remainder. In addition, it holds for $f_{X_{1t},X_{1(t-1)}}(\cdot,\cdot)$, $f_{X_{1t},X_{1(t-1)},X_{1(t-2)}}(\cdot,\cdot,\cdot)$, $f_{Z_{1t},Z_{1(t-1)}}(\cdot,\cdot)$

or $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(\cdot,\cdot,\cdot)$, being the probability density functions of $\left(X_{it}, X_{i(t-1)}\right)$, $(Z_{it}, Z_{i(t-1)})$, $\left(X_{it}, X_{i(t-1)}, X_{i(t-2)}\right)$, and $\left(Z_{it}, Z_{i(t-1)}, Z_{i(t-2)}\right)$, respectively. Assumption **5.3** combines standard conditions for simultaneous equation systems allowing for correlation along time and between the error terms of the different equations of the system.

Assumptions **5.4-5.7** are standard in the literature of local linear regression estimates, for which the Nadaraya-Watson estimator is the local constant approximation; see Ruppert and Wand (1994). Assumption **5.5** contains a standard bandwidth condition for smoothing techniques. With $H_1 = o_p(H_2)$ we impose that the bandwidth of the first-stage H_1 should be chosen as small as possible and, above all, much smaller than H_2. Thus, the fitted model of the first-stage is undersmoothing and the corresponding bias can be ignored as we wish; see Cai (2002a,b) for further details. Furthermore, by the smoothness and boundedness conditions established in Assumptions **5.2** and **5.4-5.8** for the kernel function and conditional moments and densities, we may use uniform convergence results as the ones established in Masry (1996, Theorem 6).

Assumption **5.8** is a generalization of the usual rank condition for the identification of simultaneous equation systems in the parametric context. Assumption **5.9** states the conditions to obtain consistent estimators in the presence of IV, as it is done in Cai and Li (2008), Cai et al. (2006) or Cai and Xiong (2012), while Assumption **5.10** is required to show that the Lyapunov condition holds.

Under these assumptions, we can establish the asymptotic normality of our estimator for the standard case in which $\mu_2(K_u) = \mu_2(K_v)$. The proofs are relegated to the Appendix 3.

Theorem 5.1 *Under Assumptions 5.1 and 5.2-5.10, as N tends to infinity and T is fixed*

$$\sqrt{NT|H_2|}\left(\widehat{m}_{\widehat{g}}(z_1; H_2) - m(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b(z_1), \upsilon(z_1)\right),$$

where

$$\begin{aligned} b(z_1) &= \mu_2(K)\left[diag_d\left(D_f(z_1)H_2\sqrt{NT|H_2|}D_{m_r}(z_1)\right)\imath_d f^{-1}_{Z_{it},Z_{i(t-1)}}(z_1,z_1)\right. \\ &+ \left.\frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_r}(z_1)H_2\sqrt{NT|H_2|}\right)\right)\imath_d\right], \\ \upsilon(z_1) &= 2R(K_u)R(K_v)\left(\sigma_v^2 + \sigma_\xi^2 m(z_1)^\top m(z_1) + \sigma_{v\xi}m(z_1)^\top \imath_{(M-1)}\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1), \end{aligned}$$

and, for $r = 1, \cdots, d$, $D_{m_r}(z_1)$ is the first-order derivative vector of the rth component of $m(\cdot)$, $\mathcal{H}_{m_r}(z_1)$ its Hessian matrix, $D_f(z_1)$ the first order derivative vector of the density function and

$$\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) = E\left[\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}|Z_{it} = z_1, Z_{i(t-1)} = z_1\right] f_{Z_{it}, Z_{i(t-1)}}(z_1, z_1).$$

Furthermore, $diag_d\left(tr(\mathcal{H}_{m_r}(z_1)H_2)\right)$ and $diag_d\left(D_f(z_1)H_2D_{m_r}(z_1)\right)$ stand for a diagonal matrix of elements of $tr(\mathcal{H}_{m_r}(z_1)H_2)$ and $D_f(z_1)H_2D_{m_r}(z_1)$, respectively, being $\imath_d$ a $d \times 1$ unitary vector.

From this result, we can appreciate that whereas the bias term of this new estimator is exactly the same as the Nadaraya-Watson estimator without endogenous regressors; see Pagan and Ullah (1999), it exhibits some differences with respect to the asymptotic variance. In particular, the asymptotic bias emerges mainly from the first and second derivatives of the function $m(\cdot)$ due to the approximation errors of the smooth functions $m(\cdot)$, while the asymptotic variance exhibits two additional terms besides the usual: one related to the behavior of the measurement error of the reduced forms of the system and another one with the correlation between the error terms of all equations of the system. However, note that the additional covariance terms do not appear in other IV estimators; see Newey (1994) and Cai et al. (2006) for more details. Thus, it is proved that the proposed estimators are consistent and asymptotically normal with a convergence rate that depends on the sample size and the second stage bandwidth H_2 but not the first stage bandwidth H_1, because the condition $H_1 = o_p(H_2)$ is verified.

Just to illustrate the asymptotic behavior of the proposed estimators, we give a result for the univariate case when $M = 2$, $a = q = 1$ and $H_2 = h^2 I$. In this particular case, the previous result can be written as

Corollary 5.1 *Assume conditions 5.1 and 5.2-5.10 hold, then if $h \to 0$ in such a way that $Nh^2 \to \infty$ and N tends to infinity and T is fixed, we obtain*

$$\sqrt{NTh^2}\left(\widehat{m}_{\widehat{g}}(z_1; h) - m(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b_h(z_1), v_h(z_1)\right),$$

where

$$
\begin{aligned}
b_h(z_1) &= h^2\mu_2(K)\left(\frac{f'_{Z_{it},Z_{i(t-1)}}(z_1)}{f_{Z_{it},Z_{i(t-1)}}(z_1,z_1)}D_m(z_1)+\frac{1}{2}\mathcal{H}_m(z_1)\imath_d\right)\sqrt{NTh^2},\\
v_h(z_1) &= 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m(z_1)^\top m(z_1)+\sigma_{v\xi}m(z_1)^\top\imath_{(M-1)}\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1).
\end{aligned}
$$

Note that the asymptotic bias remains to be the same as of the standard Nadaraya-Watson estimator and again, even in the simplest case, and the variance term is still what makes the difference between the estimator that solves the endogeneity problem and which not.

Furthermore, there are some empirical analysis in which all regressors are correlated with the error term in an unknown way. If this is the case, ($a = 0$) and the Theorem 5.1 can be written as follows.

Corollary 5.2 *Let Assumptions* 5.1 *and* 5.2-5.10*, as N tends to infinity and T is fixed*

$$
\sqrt{NT|H_2|}\left(\widehat{m}_{1\widehat{g}}(z_1;H_2)-m_1(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b_1(z_1),v_1(z_1)\right).
$$

where

$$
\begin{aligned}
b_1(z_1) &= \mu_2(K)\left[diag_{(M-1)}\left(D_f(z_1)H_2\sqrt{NT|H_2|}D_{m_{1r}}(z_1)\right)\imath_{(M-1)}f^{-1}_{Z_{it},Z_{i(t-1)}}(z_1,z_1)\right.\\
&+ \left.\frac{1}{2}diag_{(M-1)}\left(tr\left(\mathcal{H}_{m_{1r}}(z_1)H_2\sqrt{NT|H_2|}\right)\right)\imath_{(M-1)}\right],\\
v_1(z_1) &= 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m_1(z_1)^\top m_1(z_1)+\sigma_{v\xi}m_1(z_1)^\top\imath_{(M-1)}\right)\mathcal{B}^{-1}_{gg}(z_1,z_1),
\end{aligned}
$$

where $\mathcal{H}_{m_{1r}}(z_1)$ is the Hessian matrix of $m_1(\cdot)$ and

$$
\mathcal{B}_{gg}(z_1,z_1)=E\left[g_{it,i(t-1)}g^\top_{it,i(t-1)}|Z_{it}=z_1,Z_{i(t-1)}=z_1\right]f_{Z_{it},Z_{i(t-1)}}(z_1,z_1).
$$

On the other hand, to obtain the asymptotic behavior of the marginal integration estimator we need the following additional assumptions.

Assumption 5.11 $q(\cdot)$ *is a positive weighting function defined on the compact support of $f_{Z_{it}}$, twice continuously differentiable and holds*

$$
\int q(u)du = 1 \qquad ; \qquad \int f(u)q(u)du = 0.
$$

Assumption 5.12 *The bandwidth matrix H_3 is symmetric and strictly definite positive. Also, let h_3 be each entry of the matrix H_3, $h_3 \to 0$ and as $N \to \infty$, $N|H_3| \to \infty$.*

Assumption **5.11** is a standard condition in this literature to identify $f_{Z_{it}}$ up to a multiplicative constant. However, if we do not impose $\int f(u)q(u)du = 0$, $f_{Z_{it}}$ can also be identified up to an additive constant. Then, for the standard case that $\mu_2(K_u) = \mu_2(K_v)$ we obtain,

Corollary 5.3 *Under Assumptions 5.1, 5.2-5.5, 5.7 and 5.11-5.12, as N tends to infinity and T is fixed*

$$
\begin{aligned}
bias\left[\widehat{\alpha}(z_1;H_3)\right] &= \mu_2(K)\left[\frac{1}{2}tr(\mathcal{H}_\alpha(z_{11})H_3) - \frac{1}{2}\int tr(\mathcal{H}_\alpha(z_2)H_3)q_{Z_{i(t-1)}}(z_2)dz_2\right. \\
&+ D_\alpha(z_1)H_3D_f(z_1)\int \frac{q_{Z_{i(t-1)}}(z_2)}{f_{Z_{it},Z_{i(t-1)}}(z_1,z_2)}dz_2 \\
&- \left.\int D_\alpha(z_2)H_3D_f(z_2)\frac{q_{Z_{i(t-1)}}(z_2)}{f_{Z_{it},Z_{i(t-1)}}(z_1,z_2)}dz_2\right] + o_p\left(tr(H_3)\right) + O_p\left(\frac{1}{\sqrt{NTH_3}}\right), \\
Var\left(\widehat{\alpha}(z_1;H_3)\right) &= \frac{2\sigma_v^2R(K_u)R(K_v)}{NT|H_3|}\int \frac{q^2_{Z_{i(t-1)}}(z_2)}{f_{Z_{it},Z_{i(t-1)}}(z_1,z_2)}dz_2(1+o_p(1)).
\end{aligned}
$$

The proof of this result is straightforward following the proof of Qian and Wang (2012, Theorem 2, pp. 489-492)

5.3.2 Confidence Intervals

In order to obtain $(1-\alpha)$-confidence intervals for the empirical application, we propose to follow the wild bootstrap technique of Härdle et al. (2004). Thus, with this method we pretend to obtain the following statistic

$$
S_{m_j} = sup_z\left|\widehat{m}_j(z_1;H_2) - m_j(z_1;H_2)\right|\widehat{\sigma}_v(z_1;H_2)^{-1}, \quad j = 1,\cdots,d,
$$

where $\widehat{\sigma}_v^2$ is the estimated variance of $\widehat{m}_j(z_1;H_2)$. To obtain this statistic, it is necessary to determine the distribution of S_{m_j} via the calculus of

$$
S^*_{m_j} = sup_z\left|\widehat{m}^*_j(z_1;H_2) - E\left[\widehat{m}^*_j(z_1;H_2)\right]\right| se\left(\widehat{m}^*_j(z_1;H_2)\right)^{-1},
$$

where $E[\cdot]$ is the expectation over the bootstrap samples estimators, $se(\cdot)$ the corresponding standard deviation, and $*$ indicates that we are referring to the bootstrap samples. As it is emphasized in Härdle et al. (2004), to determine the distribution of S we can use the estimated variance of the estimators in question, $\widehat{\sigma}_v^2(z)$, or the estimated variance of the estimators of the bootstrap samples, $\widehat{\sigma}_v^{*2}(z)$. However, since the bootstrap theory suggests that the second option provides a greater higher order adequacy to bootstrap, we chose the latter one.

To determine the distribution of the statistic S it is necessary to obtain the bootstrap sample, for which we have to calculate the bootstrap residuals as

$$\Delta\widehat{\xi}_{it}^* = \Delta\widehat{\xi}_{it}\epsilon_i \left(\frac{N(T-1)}{N(T-1)-1}\right)^{1/2} \quad \text{and} \quad \Delta\widehat{v}_{it}^* = \widehat{\rho}\Delta\widehat{\xi}_{it}^*,$$

where $\Delta\widehat{\xi}_{it}$ are the residuals of the original estimation and $\widehat{\rho}$ the linear correlation coefficient of the error terms of the different equations of the system.

Let ϵ_i be a random error term between individuals that follows a Rademacher distribution and satisfies $E(\epsilon_i) = 0$, $E(\epsilon_i^2) = 1$ and $P[\epsilon_i = 1] = P[\epsilon_i = -1] = 1/2$, the bootstrap samples are generated such as

$$\Delta Y_{it}^* = \widehat{\alpha}(Z_{it}, Z_{i(t-1)}) + \Delta W_{it}^{*\top}\widehat{m}_{1\widehat{g}}(Z_{it}, Z_{i(t-1)}) + \Delta U_{it}^{\top}\widehat{m}_{2\widehat{g}}(Z_{it}, Z_{i(t-1)}) + \Delta\widehat{v}_{it}^*,$$

where ΔW_{it}^* is the vector of bootstrap dependent variables of the $M-1$ equations of the system that are obtained using the nonparametric estimators $\widehat{g}(X_{it}, X_{i(t-1)})$ and $\Delta\xi_{it}^*$.

Therefore, in order to obtain uniform confidence bands we use

$$\left[\widehat{m}_j(z_1; H_2) - s_{m_j}^*\widehat{se}(\widehat{m}_j^*(z_1; H_2)), \widehat{m}_j(z_1; H_2) + s_{m_j}^*\widehat{se}(\widehat{m}_j^*(z_1; H_2))\right],$$

for each point z_1, where $s_{m_j}^*$ is the $(1-\delta)$-quantile of $S^*(\cdot)$ and $\delta \in (0,1)$. This procedure is similar for the estimator $\widehat{\alpha}(\cdot)$.

5.4 Empirical application

Although the theory of precautionary savings states that uncertainty has a negative impact on household consumption and positive on their savings; see Carroll and Samwick (1997), several empirical studies attempt to establish this relationship but without achieving conclusive results. However, because the optimal consumption choice depends on both the life-time resources and the expected rate of income growth and household's health-care spending, there is some consensus in considering that household's consumption/saving decisions vary systematically with the age of the house.

In this context and with the aim of showing the feasibility and possible gains of the proposed method, in the following we consider a simulated example and analyze a stochastic model of precautionary saving based on the LCH model of Modigliani and Brumberg (1954). Thus, using a random sample of Spanish households, we establish to what extend the precautionary behavior of the households affects to their consumption/saving decisions, without imposing restrictive assumptions on the functional forms or unobserved individual heterogeneity, something that to our knowledge is completely new.

5.4.1 Data sample and empirical application

The microdata used in this analysis are obtained from the ECPF elaborated by the INE for the period 1985(I)-1996(IV). The ECPF is a rotating quarterly survey of representative samples of the Spanish population and each household is interviewed for eight consecutive quarters.. The survey contains not only information about demographic characteristics, economic status and industrial sector of employment for each household, but also about the detailed categories of income and consumption expenditure. Total disposable households income includes earnings of self-employment and employment in public or private sectors, investment income and property, regular transfers (i.e., pensions, unemployment insurance and other regular transfers) and other cash income (extraordinary and non-broken down). Total households expenditures include durable, non-durable and other miscellaneous expenditures.

Traditionally, and based on Hall and Mishkin (1982), the household precautionary behavior to unexpected changes in income has been measured by spending changes in non-durable goods. However, recently authors like Aaronson et al. (2012) show that durable goods react to these shocks much more than non-durable. Therefore, in order to determine the household behavior to this type of adversity is often convenient to work with two different savings variables; see Attanasio et al. (1999) or Chou et al. (2004), among others. Specifically, the first saving variable excludes consumption in durable goods from the calculation; i.e., furniture and household equipment and paid or imputed rent of the house, whereas the second one takes into account such expenses. In this way, each definition is the result of reducing household's disposable income by the corresponding expenditure variable.

The number of observations initially available in the ECPF is $148,679$, but in order to work with a balanced panel as complete as possible we only consider information from those households that answer to the eight quarters and provide information about their incomes and expenses. For reasons of sample size and to avoid having to specify an inheritance function, households whose head is aged under 26 or over 65 years old are excluded. As we can see in Table 5.1, in this sample there is a large proportion of households with negative savings; i.e., 60.36% of the entire sample, so we have to take care when we define the saving variables. In Table 5.1 we summarize the distributions of the households with negative savings.

Table 5.1. Distribution of households with negative savings

Population Group	Total Obs	Negative Obs	% total
Total	30,000	18,107	60.36%
26-35 age	5,314	3,153	10.51%
36-45 age	7,494	4,787	15.96%
46-55 age	7,764	4,744	15.81%
56-65 age	9,428	5,423	18,08%

Notes: This saving variable is defined as the difference between total household disposable income and total expenditures. % total is the proportion of observations with negative saving in the entire sample. Obs = observations

In order to calculate the saving variables, in Chou et al. (2004) it is proposed to follow the usual choice of taking $S = ln(I - C)$ as dependent variable, where I is the income and C the consumption. However, this expression excludes those households with negative savings and the omission of a such considerable proportion of the random sample causes a serious sample selection problem that can invalidates our conclusions. To overcome this problem, following Deaton and Paxson (1994) we use an approximate saving rate as dependent variable; i.e. $S = ln(I) - ln(C)$. In this way, this technique enables us to take into account the information of the entire sample, including those households with negative savings.

Since data are large enough, we focus our attention to a sample restricted to married couples with one or two children that own a unique property. In addition, to remove income and expense outliers we eliminate the 2.5% in the upper and lower tail of the income distribution of the households of the sample, whereas for the expenses in non-durable the 1% of the upper and lower tail is removed. Thus, we work with a final sample of $1,856$ observations; i.e., 232 families. The distribution of both household's disposable income and household's expenditures of the working sample is collected in Table 5.2.

Table 5.2. Distribution of household's disposable income and expenditures by age-group

	Total household disposable income				Total household disposable income			
	26-35	36-45	46-55	56-65	26-35	36-45	46-55	56-65
Mean	1,165,321	4,781.7	1,133,457	1,070,904	1,687,246	1,653,106	1,623,759	1,606,457
Std	468,277.1	2,530.2	448,960.6	517,507.8	842,219.2	839,472.8	714,451.7	825,215.2
Obs	586	503	353	414	586	503	353	414

Notes: Both revenues and expenses are measured in constant 1985 pesetas. Std = Standard deviation.

Analyzing the figures in Table 5.2, we can note that, in average, total expenditures are higher in the younger group and as the household age increases they become smaller. On its part, the figures of total disposable income do not exhibit a clear trend and it might be necessary to consider other features of the household in order to obtain a better understanding of its evolution by groups of age. More precisely, we propose to use educational level of the household as a feature to explain total disposable income.

In Table 5.3, we collect the distribution of household's income and expenditures by educational level.

Table 5.3. Distribution of household's disposable income and expenditures by educational level

	Total household disposable income		Total household disposable income	
	Low education	High education	Low education	High education
Mean	1,078,119	1,188,709	1,622,239	1,663,152
Std	481,438.6	471,254.8	784,840.8	832,175.1
Obs	1,128	592	1,128	592

Note: As household with high education we consider those with at least a high-school diploma, whereas household with low educational level are those whose head is illiterate, have no education or first degree studies.

Analyzing the figures in Table 5.3, we can see that, as expected, both revenue and expenditure from the group with a high level of education are bigger. Therefore, in the next subsection we first estimate the model specified in Section 5.3 without considering the education level. Next, we reestimate the model considering different education levels and, in this way, we can analyze the heterogeneity between these populations groups.

5.4.2 Empirical results

As we state in Section 5.2, based on the LCH model and avoiding restrictive assumptions about functional forms and unobserved cross-sectional heterogeneity, the system of equations that we estimate is

$$\begin{array}{rcl} Y_{it} & = & \alpha(Z_{it}) + W_{it}m_1(Z_{it}) + U_{it}m_2(Z_{it}) + \mu_i + v_{it}, \\ W_{it} & = & g(X_{it}) + \zeta_i + \xi_{it}, \end{array} \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{5.15}$$

where i index the household, t the time, Z_{it} is the age of the household head, W_{it} is the health-care expenditures (log), Y_{it} the savings, U_{it} the permanent income (log), and X_{it} is a vector that contains the age of the main breadwinner of the house and the number of children under 14 years old.

Note that household's permanent income is not directly observable. In order to approximate this variable we follow the proposal in Chou et al. (2004). Thus, assuming that the interest rate equals the productivity rate of growth and 65 years old is the maximum age at which people works, the permanent earnings at age τ_0 are the fitted values that results from the estimation of the following fixed effects model

$$Y_{it}(\tau_0) = X_{it}^{\top}\beta + (65 - \tau_0 + 1)^{-1}\sum_{\tau=\tau_0}^{65} f(\tau_{it}) + \widetilde{\mu}_i + \widetilde{v}_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T,$$

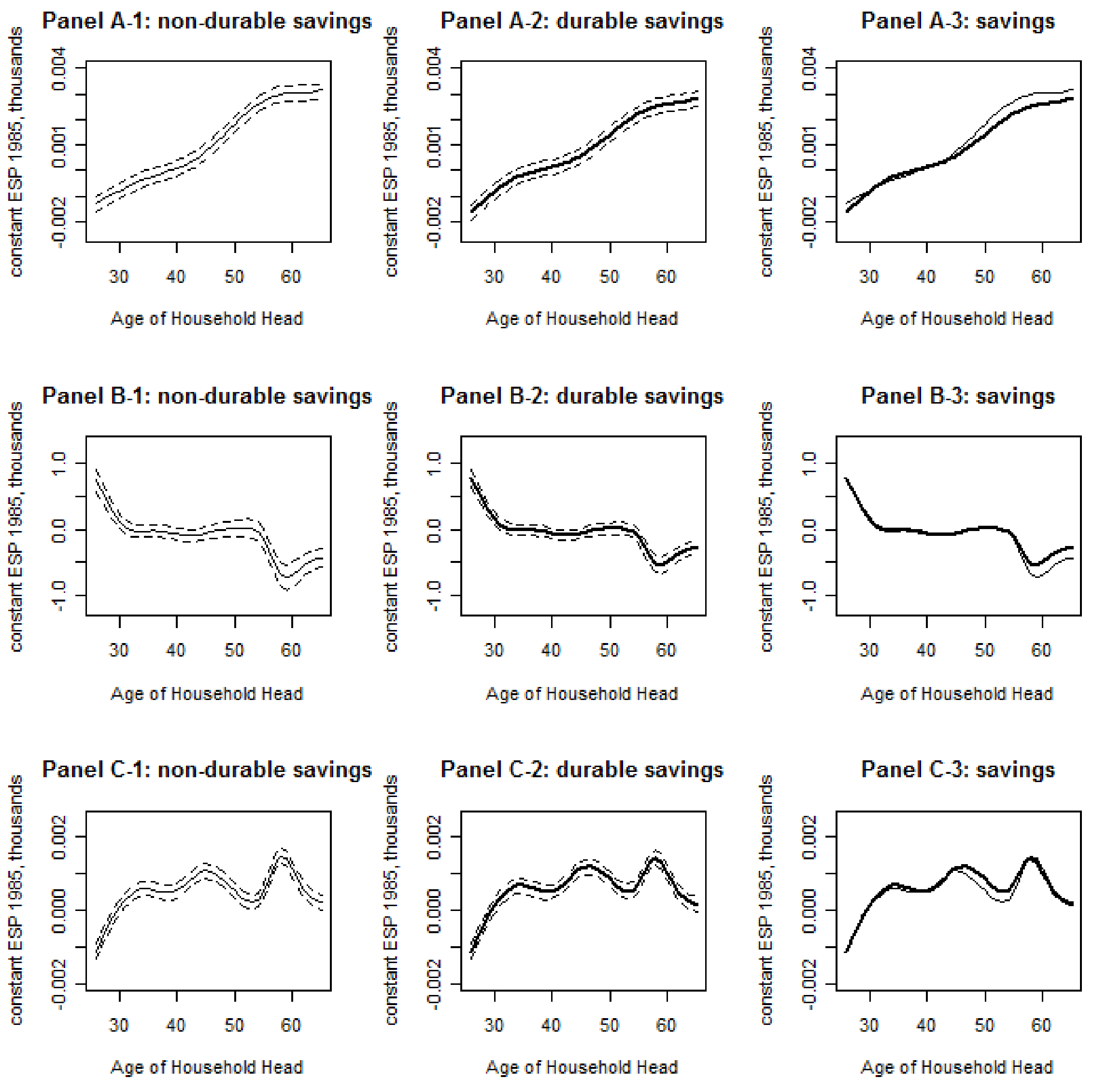
where $f(\tau)$ is the estimated quadratic function of age, Y_{it} the household income, X_{it} a vector of demographic characteristics, $\widetilde{\mu}_i$ is the unobserved fixed effects, and $\widetilde{v}_{it}$ is the random error term.

Before to show the estimation results, we make a brief discussion about the choice of different bandwidths. As it is well-known, there are many standard procedures for optimal bandwidth selection, such as plug-in, cross-validation criteria, and so on. For methodological simplicity, the bandwidth of the second-step estimation procedure is chosen using the rule-of-thumb; i.e., $\widehat{H}_2 = \widehat{h}_2 I = \widehat{\sigma}_z n^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of Z. The same is done to compute $\widehat{H}_3$. For the selection of $\widehat{H}_1$ note that in order to accomplish with the conditions establish in Theorem 5.1 we should choose a $\widehat{H}_1$ that clearly undersmooth the optimal; i.e., $\widehat{H}_1 = \widehat{h}_1 I = \widehat{\sigma}_x n^{-1/3}$, where $\widehat{\sigma}_x$ is the sample standard deviation of X.

Estimation results are shown in Figures 5.1-5.4. The estimated curves are plotted against the age variable jointly with 95% pointwise confidence intervals. Figure 5.1 shows results for the entire working sample, without considering educational levels. On the other side, in Figures 5.2 and 5.3 we show the estimation results distinguishing between sample of those who have higher educational level (Figure 5.2) and those who have low education level (Figure 5.3). Finally, in Figure 5.4 we show the estimation results when endogeneity is taking into account.

Figures 5.1-5.4 have the same structure. They are divided into three panels, A, B and C. Panels A show the precautionary savings elasticity to changes in households risk aversion; i.e., $\widehat{\alpha}(\cdot)$. Panels B exhibit the corresponding elasticity to changes in health-care expenditures, i.e., $\widehat{m}_1(\cdot)$, whereas Panels C show the precautionary savings elasticity to changes in household income; $\widehat{m}_2(\cdot)$. Specifically, Panel A-1 shows the estimated curves when durable goods are not taken into account. Panel A-2 focuses on the second definition of savings, while Panel A-3 makes a comparison between both results. Note that this structure is maintained for Panels B and C.

Figure 5.1. Household's savings over the life-cycle



Note: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

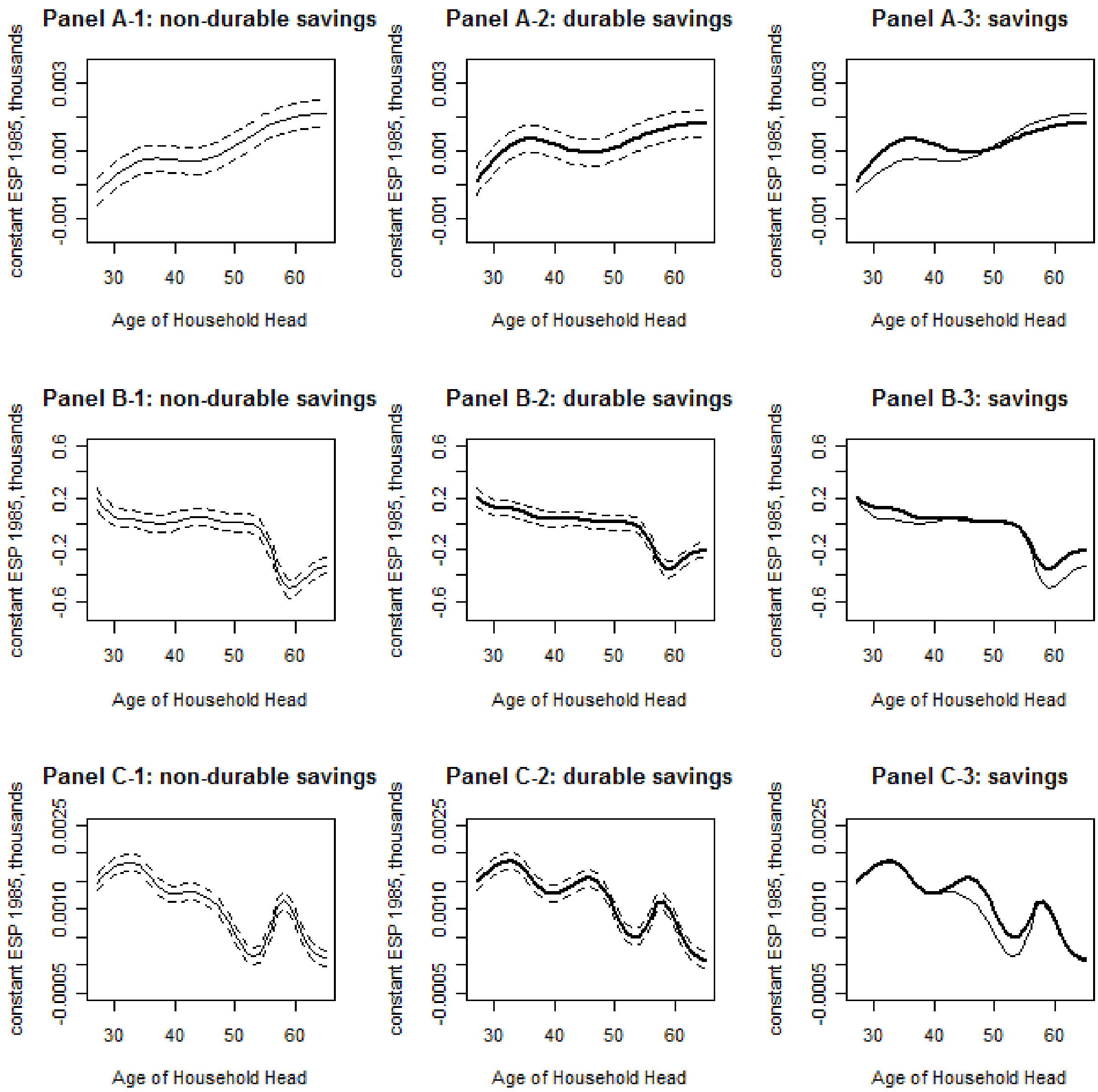
Focusing on the results in Figure 5.1, we can note that when we control for household risk aversion (Panels A) there is a positive relationship between the savings rate of households and their age. This is especially stronger after age 47, presumably for retirement or inheritance reasons. Meanwhile, when we control for income uncertainty (Panels C) we find that younger households (26−46) behave as buffer-stock agents and close to retirement they begin to accumulate wealth, corroborating in this way the results in Gourinchas and Parker (2002) and Cagetti (2003). When we control for uncertainty about health-care expenditures (Panels B) we see that younger households (26 − 31) exhibit a declining savings rate, followed by a constant path till the age of 56, where the hump-shaped is inverted.

If we combine these results with the delay in the wealth accumulation process of the Spanish households (note that in the U.S. it begins around 40 age while in Spain at 55 age) we realize the negative impact that public health programs have on precautionary savings, confirming the results in Chou et al. (2004). Finally, comparing the behavior of the elasticity for the different savings results, we can note that consumption of durable goods reacts much more to unexpected changes in either income or potential health-care payout, whereas consumption in non-durable goods is more sensitive to household risk aversion. This holds especially for households over 50 years old.

Now, we turn to analyze the impact of the different types of uncertainty on the household precautionary behavior when facing different educational level (high or low). Controlling for household risk aversion in Figures 5.2 and 5.3 we find out that there is a positive relationship between the saving rate of the households and their age for both samples. As we can see in Figures 5.2 and 5.3, when we control for household preferences we observe that less educated agents are risk averse tending to save during the early stages of work (26 − 38). Afterwards, they increase their consumption. On the contrary, households with more education show a smaller degree of risk aversion. Furthermore, these household show a rather steady savings path along time. This change dramatically at the age of 48, where savings rate increase exponentially, may be due to retirement or legacy reasons. This result extends the findings that appear in Cagetti (2003), where he points out that households with less education exhibit a lower risk aversion degree.

Focusing on the precautionary savings elasticity to changes in unexpected health-care expenses in Figures 5.2 and 5.3 we obtain a savings rate with a similar path until the age of 50. Thereafter, households with a high educational level show a more cautious behavior regarding to unforeseen health-care expenditures. Finally, analyzing this elasticity to unexpected income changes we appreciate a completely different behavior between these

Figure 5.2. Household's savings over the life-cycle by education level: low education

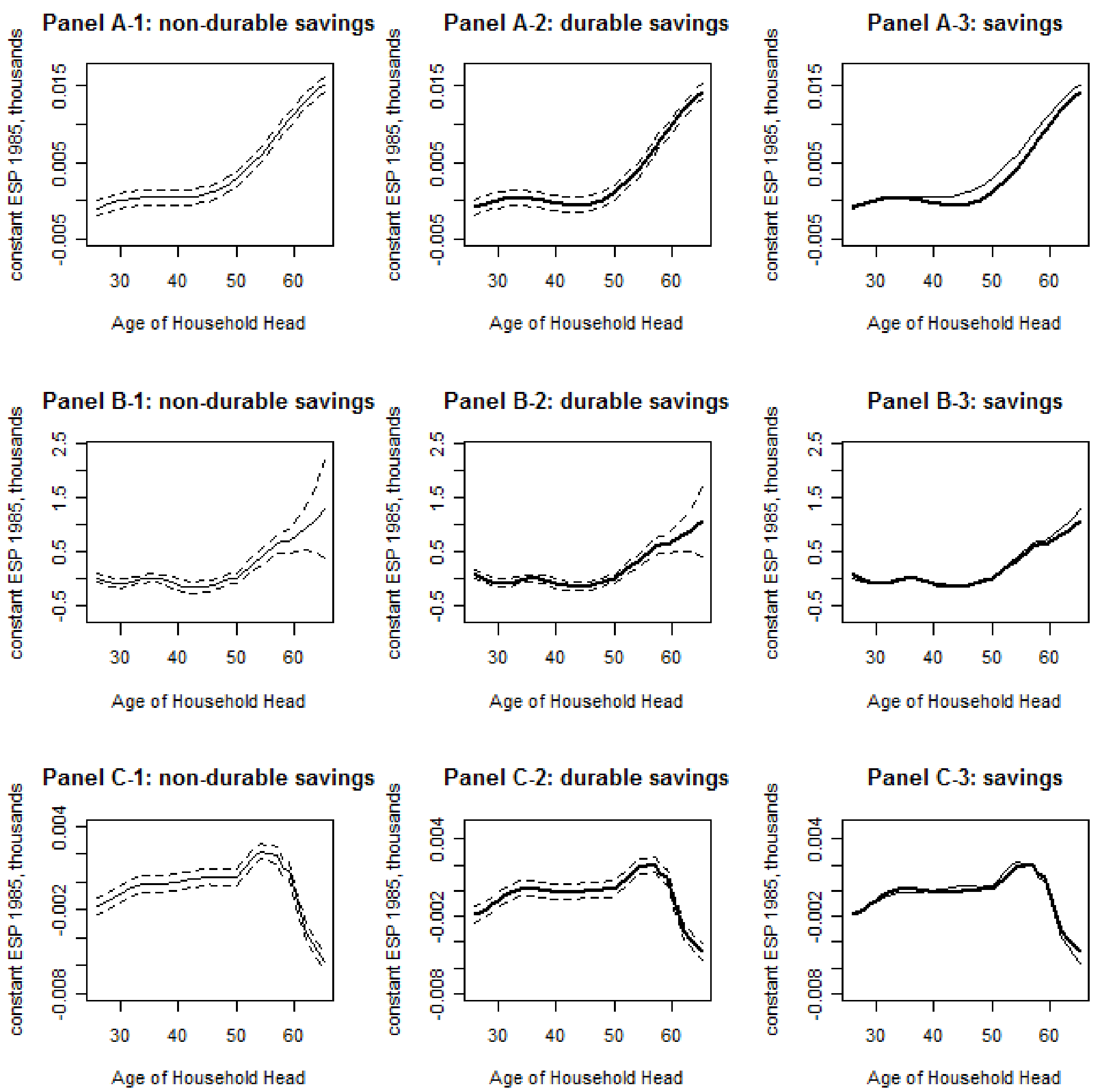


Note: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

populations. While households with low education exhibit a declining savings rate until age 50, when we observe the hump-shaped established by the LCH model, households with a higher educational level behave as buffer-stock agents until age 39, maintaining a constant saving rate in adulthood and later a hump-shaped from the 55 to the age of retirement.

Finally, in order to evaluate the empirical relevance of the endogeneity problem we compare

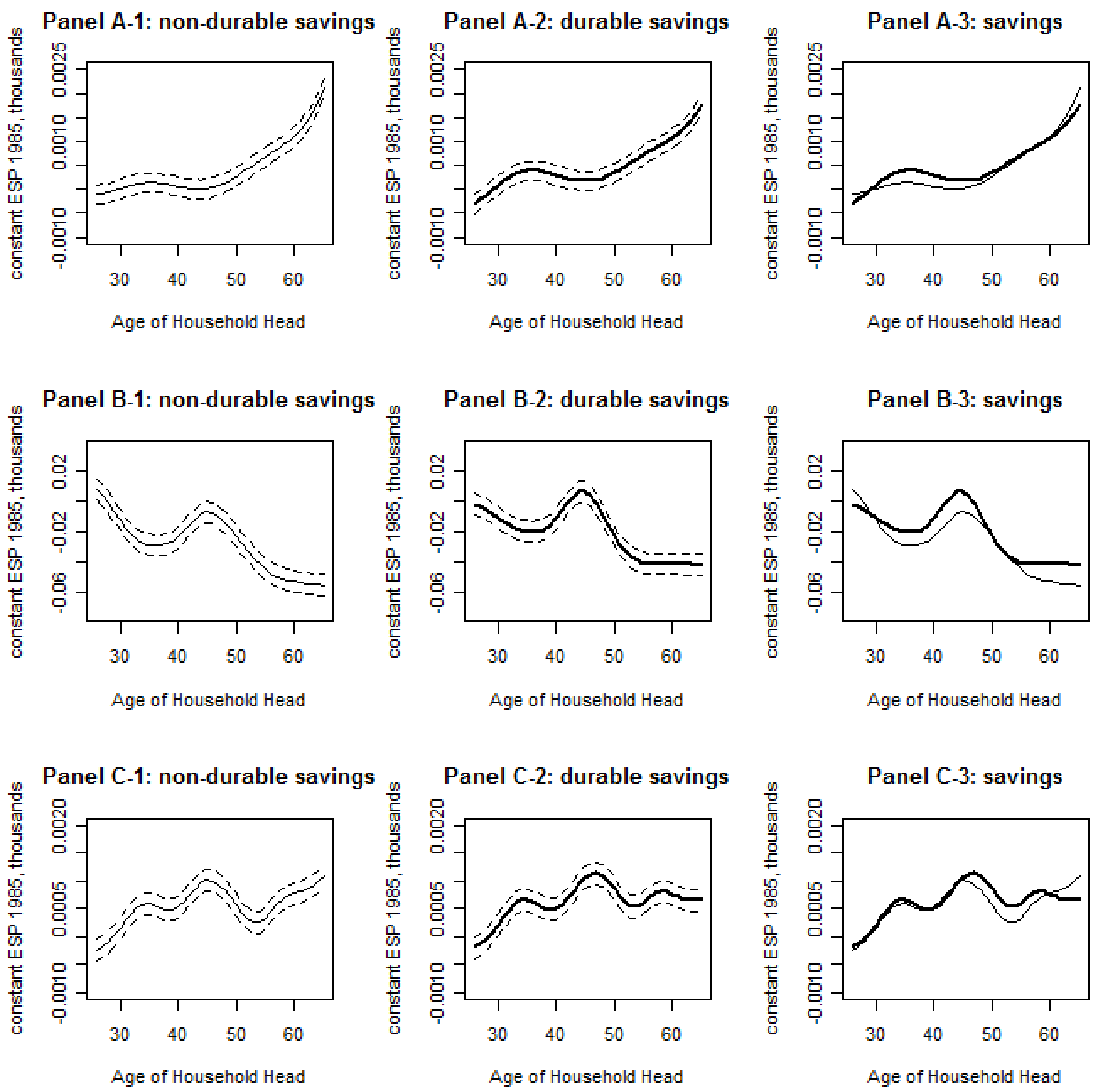
Figure 5.3. Household's savings over the life-cycle by education level: high education



Note: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

the results in Figure 5.1, that is, where the endogeneity of the health-care expenditures has been treated as IV, against the results obtained when the endogeneity problem is omitted; see Figure 5.4. We note that the empirical results are rather similar for the case of exogenous variables, but there is a significant difference when we analyze the estimates related to the potentially endogenous variables. By correcting by endogeneity through the use of IV we observe that younger households do not accumulate assets, whereas when endogeneity is not taken into account younger households behave as buffer-stock agents.

Figure 5.4. Household's savings over the life-cycle with endogeneity



Notes: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

In summary, these results confirm what is obtained in other papers of this literature. All these results indicate that an extension of the standard life cycle model that takes into account households' preventive motif linked to uncertainty of both labor market and life expectancy is very appealing. In addition, combining the particularities of this model jointly with the estimation strategy proposed in this chapter enables us to determine household's consumption/savings decisions without having to resort to further strong assumptions about functional forms or densities.

5.4.3 Monte Carlo experiment

In this part, we use the simulated data to examine the performance of the proposed estimator in finite samples under different semi-parametric specifications of the model discussed in Section 5.2. In this way, in order to corroborate the consistency of the empirical results obtained in Section 5.4.2 we use the mean squared error (MSE) as a measure of the estimation accuracy of the proposed estimators.

Let φ be the φth replication and Q the number of replications, the MSE of the two-step weighted locally constant least-squares estimators (5.9) may be approximated by the averaged mean squared error (AMSE), so for the two-step weighted locally constant least-squares estimator we have

$$AMSE(\widehat{m}(z_1;H_2)) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{r=1}^{d}\left(\widehat{m}_{\varphi_r}(z_1;H_2) - m_{\varphi_r}(z_1;H_2)\right)\Delta\widehat{W}_{it,\varphi_r}\right)^2,$$

while the corresponding AMSE for the marginal integration estimator is

$$AMSE(\widehat{\alpha}(z_1;H_3)) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{\alpha}_{\varphi}(z_1;H_3) - \alpha_{\varphi}(z_1;H_3)\right)^2.$$

Observations are generated from the following semi-parametric panel data regression system:

$$Y_{it} = \alpha(Z_{it}) + W_{it}m_1(Z_{it}) + U_{it}m_2(Z_{it}) + \mu_i + v_{it}, \quad i=1,\cdots,N \quad ; \quad t=1,\cdots,T, \tag{5.16}$$

where W_{it} is an endogenous variable constructed as $W_{it} = g(X_{it}) + \zeta_i + \xi_{it}$. Also, Z_{it}, U_{it} and X_{it} are random variables generated such that $Z_{it} = \omega_{it} + \omega_{i(t-1)}$ (ω_{it} is an i.i.d. uniform distributed (0,Π/2) random variable), $U_{it} = 0.25U_{i(t-1)} + \psi_{1it}$, and $X_{it} = 0.5X_{i(t-1)} + \psi_{2it}$ (ψ_{1it} and ψ_{2it} are *i.i.d.* $\mathcal{N}(0,1)$).

The error distributions of v_{it} and ξ_{it} are generated as

$$v_{it} = 0.65v_{i(t-1)} + \vartheta_{it} \qquad \text{and} \qquad \xi_{it} = \epsilon_{it} + \rho v_{it},$$

where $\epsilon_{it} = 0.5\epsilon_{i(t-1)} + \vartheta^*_{it}$ (ϑ_{it} and ϑ^*_{it} are *i.i.d.* $\mathcal{N}(0,1)$ random variables). Clearly, v_{it} is independent of Z_{it} and U_{it} so that $E(v_{it}|\mathbf{Z},\mathbf{U}) = 0$ and $E(v_{it}) = 0$. However, $E(v_{it}|\mathbf{W}) \neq 0$ because v_{it} and ξ_{it} are correlated through the parameter $\rho = 0.5$, that is responsible for this correlation.

In addition, to allow the presence of cross-sectional heterogeneity in the form of fixed effects we assume that the individual effects are correlated with the nonparametric covariates. In particular, the dependence between these terms is imposed by generating $\mu_i = 0.5\overline{Z}_{i\cdot} + u_i$ and $\zeta_i = 0.5\overline{X}_{i\cdot} + u_i$, where u_i is an *i.i.d.* $\mathcal{N}(0,1)$ random variable. For $i = 2, \cdots, N$, $\overline{Z}_{i\cdot} = T^{-1}\sum_{t=1}^{T} Z_{it}$ and $\overline{X}_{i\cdot} = T^{-1}\sum_{t=1}^{T} X_{it}$.

In the experiment, we consider two different cases of study. First, we analyze (5.16) where the varying coefficients can be related to both endogenous and exogenous covariates. Later, we focus on a specification in which the nonparametric elements are the only exogenous regressors; i.e.,

$$Y_{it} = \alpha(Z_{it}) + W_{it}m_1(Z_{it}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{5.17}$$

In this context, to verify the asymptotic theory of Section 5.3, the number of period T is fixed at 3, while the number of cross-sections N is varied between 50, 75 and 100. We use 500 Monte Carlo replications Q and a Gaussian kernel. Following the assumptions necessary to obtain estimators with a suitable asymptotic behavior, we propose to obtain the bandwidth matrix of H_2 and H_3 by the rule-of-thumb, whereas H_1 is chosen to be undersmoothing. Thus, $\widehat{H}_1 = \widehat{h}_1 I = \widehat{\sigma}_x (NT)^{-1/3}$, $\widehat{H}_2 = \widehat{h}_2 I = \widehat{\sigma}_z (NT)^{-1/5}$ and $\widehat{H}_3 = \widehat{h}_3 I = \widehat{\sigma}_z (NT)^{-1/5}$, where $\widehat{\sigma}_x$ and $\widehat{\sigma}_z$ are the sample standard deviation of X_{it} and Z_{it}, respectively.

In the following tables, we report the simulations results through the estimated bias, standard deviation (Std), and AMSE for the two estimators of both specifications. The latter is computed as it is previously set, whereas the bias is obtained via

$$Bias\left(\widehat{m}(z_1; H_2)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{r=1}^{d}\left(E(\widehat{m}_{\varphi r}(z_1; H_2)) - m_{\varphi r}(z_1)\right)\Delta\widehat{W}_{it,\varphi r}\right),$$

and the standard variance such that

$$sd\left(\widehat{m}(z_1; H_2)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\left(\widehat{m}_{\varphi r}(z_1; H_2) - E(\widehat{m}_{\varphi r}(z_1; H_2))\right)\Delta\widehat{W}_{it,\varphi r}\right)^2\right)^{1/2}$$

Table 5.4. Simulation results. AMSE for $M = 2$ and $a = 0$

		Two-step locally constant least-squares estimator			Marginal integration estimator		
		Bias	Std	AMSE	Bias	Std	AMSE
	N = 50	0.01138	0.51785	0.63383	-0.82138	0.42625	0.91673
T=3	N = 75	0.00118	0.47344	0.45622	-0.80465	0.42282	0.87155
	N = 100	0.00739	0.44615	0.41289	-0.80889	0.40361	0.85723

Table 5.5. Simulation results. AMSE for $M = 2$ and $a = 1$

		Two-step locally constant least-squares estimator			Marginal integration estimator		
		Bias	Std	AMSE	Bias	Std	AMSE
	N = 50	1.65751	0.83162	2.41698	0.55222	0.46183	0.95029
T=3	N = 75	1.57539	0.84969	1.99571	0.54179	0.44837	0.91023
	N = 100	1.57340	0.71259	1.99565	0.53444	0.43619	0.89058

As we can see in Tables 5.4 and 5.5, the simulations provide the expected results. For all T, as the sample size increases the bias estimated of the two-stage weighted locally constant least-squares estimator, $\widehat{m}(z_1; H_2)$, for both $a = 0$ (Table 5.4) and $a = 1$ (Table 5.5) shrinks from 0.01138 to 0.00739 and from 1.65751 to 1.57340, respectively. Meanwhile, the values of the standard deviations show a similar pattern going from 0.51785 to 0.44615 when $a = 0$ and from 0.83162 to 0.71259 when $a = 1$. In turn, this asymptotic behavior is also holds for the integration marginal estimator. Thus, when the sample size increase we see that the bias estimator of $\widehat{\alpha}(z_1; H_3)$ decreases from -0.82138 to -0.80888 in the first specification and from 0.58222 to 0.53444 in the second one, whereas the standard deviations pass from 0.42625 to 0.40361 and from 0.46183 to 0.43619, respectively.

Finally, focus on the behavior of AMSE (Tables 5.4 and 5.5) we can highlight how, maintaining T fixed, the AMSEs decrease as N increases. In this way, all these results enable us to state that the estimators proposed in this chapter are consistent corroborating the results of their theoretical properties established in the previous sections. Therefore, the consistency of the previous empirical results is proved.

5.5 Conclusions

This chapter considers the nonparametric estimation of a structural model of optimal life-cycle savings, controlling for both uncertainty about health-care expenditures and household risk aversion. The main attraction of this estimator is that, compared to those already proposed in the literature, it allows to deal simultaneously with different problems such as unobserved cross-sectional heterogeneity, varying parameters of unknown form in the Euler equation and endogenous covariates. The estimator of the function of interest turns out to have the simple form of a two-step weighted locally constant least-squares estimator. Additionally, some marginal integration technique is needed to compute a subset of the functionals of interest. In the chapter, we establish the asymptotic properties of these estimators. To illustrate the feasibility and possible gains of this method, the chapter concludes with an application about household's precautionary savings over the life-cycle. From this empirical application, we obtain the following conclusions: Households accumulate wealth at least in two periods in life. In younger stages, household's savings are devoted to guard against uncertainty about potential income downturn, whereas when households become older their savings are intended for retirement and bequests. Furthermore, public health programs have a negative impact on precautionary savings. Finally, by comparing educational levels we obtain that households with low education levels are more risk averse than those with higher levels.

Conclusiones

Resultados

La disponibilidad de datos de panel ha permitido aumentar considerablemente la complejidad de los modelos econométricos. Sin embargo, en muchas áreas científicas como la economía, medicina, etc., la estimación de modelos de datos de panel no paramétricos o semi-paramétricos puede llevarnos a resultados infrasuavizados si ignoramos posibles relaciones no lineales entre las variables explicativas o si no se permite que ciertos coeficientes de la regresión varíen en función de ciertas covariables propuestas por la teoría económica, ver por ejemplo Card (2001) o Kottaridi and Stengos (2010). En este sentido, los modelos de coeficientes variables de datos de panel que nos permiten explotar posibles características dinámicas ocultas en el conjunto de datos han sido objeto de una investigación intensiva en los últimos años, tanto desde el punto de vista teórico como metodológico.

En este contexto, esta tesis doctoral persigue un doble objetivo. Por un lado, desarrollar nuevas técnicas de estimación que nos permitan obtener estimadores consistentes para las funciones de interés suaves de un modelo de coeficientes variables de datos de panel en el cual la heterogeneidad individual no observable está correlacionada con ciertas variables explicativas. Por otro lado, demostrar las posibles ganancias empíricas proporcionadas por las estrategias de estimación propuestas en los distintos capítulos de este trabajo. Resaltar que aunque en este caso nosotros nos centramos en un caso empírico concreto, el comportamiento de los ahorros preventivos de los hogares a lo largo del ciclo vital, estas estrategias de estimación pueden ser fácilmente aplicadas a estudios empíricos de diversa naturaleza.

En esta sección, y a modo de conclusión, recogemos los principales objetivos perseguidos en cada uno de los capítulos que conforman esta tesis doctoral así como los principales resultados obtenidos.

En el Capítulo 1, se realiza una revisión intensiva sobre la literatura econométrica de modelos de datos de panel semi-paramétricos y totalmente no paramétricos. Primero, analizamos los modelos de datos de panel totalmente no paramétricos tanto con efectos fijos como aleatorios. Posteriormente, se repasan los modelos parcialmente lineales bajo tres especificaciones distintas: efectos fijos y aleatorios, y presencia de covariables endógenas. Concluimos con una revisión de los modelos de coeficientes variables. Para cada una de estas áreas se discuten las características del modelo a estimar así como la metodología propuesta. Además, también se analizan las principales propiedades asintóticas de los estimadores resultantes.

En el Capítulo 2, se desarrolla una estrategia de estimación que nos permite estimar las funciones de interés de un modelo de coeficientes aleatorios en un contexto de datos de panel en el que se asume que la heterogeneidad individual no observable está correlacionada con algunas/todas las covariables del modelo de regresión. Para evitar el problema de dependencia estadística entre la heterogeneidad no observada de sección cruzada y las covariables, se recurre a la trasformación en primeras diferencias. Sin embargo, dado que esta transformación nos proporciona un modelo de regresión que puede ser considerado como una función aditiva con la misma forma funcional pero evaluada en distintos períodos de tiempo, seguimos la propuesta de Yang (2002). De este modo, presentamos un estimador basado en una aproximación local lineal y en el uso de una ponderación de kernel de mayor dimensión.

Como se señala en Lee and Mukherjee (2008), la aplicación directa de técnicas de aproximación lineal a este tipo de modelos de datos de panel en diferencias nos conduce a estimadores que exhiben un sesgo que no desaparece incluso en muestras grandes. Sin embargo, el uso de ponderaciones de kernel de mayor orden nos permiten resolver este problema. Además de la habitual distancia entre el término fijo alrededor del cual se realiza la aproximación y un valor de la muestra, esta técnica nos permite considerar también la suma de estas distancias respecto a los distintos valores de la muestra. Desafortunadamente, analizando las principales propiedades asintóticas de este estimador comprobamos que el problema del sesgo efectivamente desaparece, pero al coste de generar un incremento del término de la varianza. Por lo tanto, este estimador alcanza una tasa de convergencia no paramétrica más lenta. Con el objeto de alcanzar optimalidad, en este capítulo proponemos reducir el problema de la dimensionalidad recurriendo a un suavizado adicional a través de un algoritmo de backfitting de una etapa. Analizando las propiedades asintóticas de este estimador queda demostrado que un suavizado adicional nos propociona estimadores que alcanzan la tasa óptima de convergencia y que exhiben la propiedad de eficiencia oráculo, como se señala en Fan and Zhang (1999). Asimismo, y dado el papel fundamental que juega la matriz de anchos de banda a la hora de alcanzar el equilibrio entre sesgo y varianza, proporcionamos un método que nos permite calcular esta matriz de manera empírica. Finalmente, a través de un experimento de Monte Carlo corroboramos el buen comportamiento de este estimador en muestras finitas. En concreto, para T fijo, a medida que aumenta el tamaño muestral el valor del AMSE es menor, tal y como esperábamos a la vista de sus propiedades asintóticas.

En el Capítulo 3, nuestro objetivo se centra en la obtención de estimadores que exhiban las propiedades asintóticas estándar de los estimadores no paramétricos bajo el supuesto de que los términos de error idiosincráticos son *i.i.d.* Para ello, recurrimos a una transformación en desviaciones respecto de la media para eliminar la dependencia estadística existente entre la heterogeneidad individual y las covariables en un modelo de coeficientes variables de datos de panel. De nuevo, para evitar el problema del sesgo asintótico no insignificante proponemos un estimador de regresión local lineal con una ponderación de kernels de dimensión T. Dado que el estimador resultante alcanza subóptimas tasas de convergencia no paramétricas, proponemos combinar este procedimiento con un algoritmo de backfitting de una etapa que nos permite cancelar asintóticamente todos los términos aditivos del modelo de regresión en diferencias. Posteriormente, demostramos que este estimador de dos etapas es asintóticamente normal y alcanza una tasa de convegencia óptima dentro de este tipo de funciones suaves. Además exhibe la propiedad de eficiencia oráculo. Finalmente, a través de un estudio de Monte Carlo confirmamos los resultados teóricos y el buen comportamiento de los estimadores en muestras finitas.

Una vez presentados los estimadores en diferencias propuestos (primeras diferencias y efectos fijos), en el Capítulo 4 se realiza un estudio comparativo sobre el comportamiento de ambos estimadores. Analizando las propiedades asintóticas de los dos estimadores de regresión local lineal se aprecia que, aunque ambos mantienen el mismo orden de magnitud del término de sesgo, presentan límites asintóticos distintos para cada término de varianza. De este modo, ambos estimadores alcanzan tasas subóptimas de convergencia no paramétricas. Por su parte, para los dos estimadores de backfitting de una etapa se demuestra que, bajo condiciones bastante generales, en ambos casos se alcanza la optimalidad y se exhibe la propiedad de eficiencia oráculo. Dado que los dos estimadores de backfitting son asintóticamente equivalentes, el análisis sobre su comportamiento en tamaños muestrales pequeños es muy interesante.

En un contexto totalmente paramétrico, se ha demostrado que bajo supuestos de exogeneidad estricta la elección entre los estimadores en diferencias depende de la estructura estocástica de los términos de error idiosincráticos. Sin embargo, en el análisis asintótico de los estimadores no paramétricos obtenemos que además de la estructura estocástica existen otros factores como la dimensionalidad de las covariables y el tamaño muestral que son de gran interés. En concreto, para analizar el comportamiento de estos estimadores en muestras finitas nos centramos en la evolución del promedio de su error cuadrático medio (AMSE) bajo distintos escenarios del término de error.

A la vista de los resultados obtenidos de la simulación, podemos destacar que el AMSE de los dos estimadores de regresión local lineal tienden a converger bajo distintas especificaciones del término de error. Además, para valores pequeños de T, el AMSE del estimador en primeras diferencias tiende a dominar en términos del AMSE del estimador de efectos fijos. Así, el estimador de efectos fijos es preferido, tal y como esperábamos. Por su parte, cuando T aumenta el comportamiento de los dos estimadores empeora considerablemente, tal y como establecen los límites asintóticos de sus términos de varianza. Por otro lado, en cuanto a los dos estimadores de backfitting se obtiene que para los distintos tipos de error considerados el estimador de dos etapas presenta un mejor comportamiento que el estimador de regresión local lineal dado que nos permite evitar el problema de la dimensionalidad. Asimismo, bajo estructuras *i.i.d.* del término de error o procesos estacionarios AR(1) se aprecia que el estimador de backfitting de una etapa de efectos fijos presenta un mejor comportamiento que el de primeras diferencias. Cuando el término de error sigue un paseo aleatorio el comportamiento del estimador es mejor en el sentido contrario. Así, en síntesis podemos afirmar que los resultados de la simulación confirman las propiedades teóricas establecidas previamente, tanto para los estimadores de regresión local lineal como para los de backfitting de una etapa. Asimismo, encontramos que los estimadores de efectos fijos son bastante sensibles al tamaño del número de observaciones temporales por individuo.

Finalmente, en el Capítulo 5 y con el objetivo de demostrar la viabilidad empírica que reportan estos nuevos procedimientos para el análisis empírico, se considera la estimación no paramétrica de un modelo estructural sobre los ahorros preventivos de los hogares españoles, para el período $1985-1996$, motivado por el modelo del ciclo vital de Modigliani and Brumberg (1954). A partir de esta especificación, estimamos un modelo en el cual los ahorros de los hogares se relacionan tanto con la incertidumbre sobre futuros gastos médicos imprevistos como con la aversión al riesgo de los hogares. Con el objetivo de contribuir a la literatura sobre los ahorros preventivos, en este capítulo tratamos de hacer frente a los errores de especificación más habituales en este tipo de estudios. Para ello, extendemos el modelo semi-paramétrico desarrollado en Chou et al. (2004) al análisis de modelos de datos de panel. Así, lo que nos proponemos es estimar un modelo del ciclo vital que nos permita determinar el comportamiento de los hogares bajo las siguientes particularidades: (i) heterogeneidad de sección cruzada no observable correlacionada con las variables explicativas; (ii) existencia de funciones desconocidas en la ecuación de Euler que relacionan variables endógenas y exógenas y que deben ser estimadas; y (iii) gasto en productos sanitarios determinado de manera endógena.

Partiendo entonces de un modelo de regresión de coeficientes variables de datos de panel con efectos fijos, el estimador que proponemos resuelve el problema de endogeneidad utilizando los valores predichos de las variables endógenas generados en la estimación no paramétrica de la ecuación en forma reducida. De este modo, el estimador resultante tiene la forma de un estimador simple de mínimos cuadrados de dos etapas localmente ponderado. Asimismo, para la estimación de un subconjunto de funciones de interés suaves ciertas técnicas de integración marginal son necesarias. Para determinar el comportamiento de estos estimadores obtenemos sus principales propiedades asintóticas. Finalmente, el capítulo concluye con un experimento de Monte Carlo que corrobora el buen comportamiento del estimador propuesto en muestras finitas y con una aplicación empírica sobre los ahorros preventivos de los hogares españoles durante el periodo $1985 - 1996$. A la vista de los resultados obtenidos para esta aplicación empírica podemos concluir que los hogares acumulan riqueza al menos en dos períodos de su vida. Por un lado, los hogares más jóvenes ahorran para protegerse contra posibles adversidades económicas imprevistas. Por otro lado, los hogares más adultos ahorran principalmente por motivos de jubilación o de legado, corroborando lo establecido de manera teórica por el modelo del ciclo vital de Modigliani and Brumberg (1954). Además, la reducción de la incertidumbre a través de programas sanitarios públicos tiene un impacto negativo sobre los ahorros. Asimismo, comparando el comportamiento de los hogares por nivel educativo obtenemos que los hogares con menor niveles educativo son más aversos al riesgo.

Futuras líneas de investigación

A lo largo del presente trabajo, y dadas las diversas ventajas que los modelos de datos de panel de coeficientes variables de forma desconocida ofrecen para estudios empíricos de diversa tipología, ciertas líneas futuras de investigación han ido surgiendo. A este respecto, y tal y como se demuestra en Card (2001), a la hora de analizar el rendimiento educativo de los individuos es necesario permitir que ciertas variables explicativas evolucionen con otras variables con el fin de obtener resultados consistentes. De este modo, un primer ejercicio puede consistir en la extensión del modelo de regresión analizado en Card (2001) a un modelo de datos de panel que puede ser estimado recurriendo a los procedimientos desarrollados en esta tesis doctoral.

Otra línea de investigación relevante es la elaboración de una serie de test no paramétricos. Por un lado, en los distintos capítulos de esta tesis doctoral ha quedado demostrado que los modelos de coeficientes variables son de gran utilidad dado que nos permiten explotar las posibles características del conjunto de datos, sin tener que recurrir a supuestos tan restrictivos sobre la especificación del modelo como los modelos totalmente paramétricos. Sin embargo, también hemos demostrado que esta flexibilidad tiene un coste añadido que es la obtención de estimadores no paramétricos con unas tasas de convergencia más lentas. Con el objetivo de poder contrastar de manera empírica la correcta especificación del modelo a analizar, resulta conveniente desarrollar un test de especificación que contraste la modelización paramétrica versus la semi-paramétrica. Para ello, siguiendo la idea de Henderson et al. (2008), un test basado en la técnica de bootstrap puede ser interesante.

Asimismo, sabemos que cuando la heterogeneidad individual no observable está correlacionada con las covariables sólo los estimadores en primeras diferencias proporcionan resultados consistentes, mientras que cuando los efectos individuales son independientes de los regresores tanto los estimadores en diferencias como los de efectos aleatorios son consistentes. Sin embargo, dado que trabajar con modelos de coeficientes variables en diferencias es una tarea laboriosa resulta interesante poder contrastar con qué tipo de efectos individuales estamos trabajando. De este modo, un test no paramétrico de efectos fijos versus efectos aleatorios que siga el estilo del bien conocido test de Hausman puede ser de gran utilidad.

Finalmente, en el Capítulo 5 se ha demostrado cómo las técnicas de estimación de regresión local lineal pueden ser fácilmente extensibles al contexto de modelos de coeficientes funcionales con variables endógenas. Sin embargo, en economía y otras ciencias sociales es común encontrarnos con situaciones en las cuales la covariable no paramétrica se determina de manera endógena; por ejemplo, en el problema estándar de elección del consumidor. En esta situación uno podría asumir que existe un vector de variables instrumentales que nos permite resolver este problema de endogeneidad. Sin embargo, en los modelos no paramétricos y semi-paramétricos esto no es tan sencillo dado que podemos incurrir en el problema de la inversa mal planteada.

Conclusions

Results

The availability of panel data has enriched greatly the complexity of econometric models. However, in many scientific areas such as economics, medicine, and others, the estimation of nonparametric and semi-parametric panel data models can lead to undersmoothing results when we ignore possible non-linear relationship between explanatory variables or some regression coefficients vary depending on certain covariates proposed by economic theory; see Card (2001) or Kottaridi and Stengos (2010), for example. Thus, varying coefficient panel data models that allow us to explode possible features hidden in the data set have been object of an intensive research in the last years, both from a theoretical and methodological point of view.

In this context, the objective of this doctoral dissertation is twofold. On one hand, to develop new estimation techniques that enable to obtain consistent estimators of the smooth functions of interest of a varying coefficient panel data model where the unobserved individual heterogeneity is correlated with some explanatory variables. On the other hand, to show the potential gains for empirical analysis provided by the proposed estimation strategies in the different chapters of the doctoral thesis. Note that although in this case we focus in a particular empirical problem, i.e., the behavior of household's precautionary savings over the life-cycle, these procedures can be easily applied to empirical studies of various kinds.

In this section, and as conclusion, we summarize the main objectives pursued in each of the chapters of this dissertation and the main results obtained.

In Chapter 1, an intensive review of the econometric literature on semi-parametric and fully nonparametric panel data models is performed. First, fully nonparametric panel data model with both random and fixed effects are analyzed. Second, we survey partially linear models under three different specifications: fixed and random effects, and presence of endogenous covariates. We conclude with a review of panel data varying coefficient models. For each of these areas, we discuss both the basic model to estimate and the proposed methodology. We also analyze the main asymptotic properties of the resulting estimators.

In Chapter 2, we develop a new estimation strategy that enables us to estimate the functions of interest of a varying coefficient panel data model where it is assumed that the unobserved individual heterogeneity is correlated with some/all the covariates of the regression model. In order to avoid the statistical dependence problem between the unobserved cross-sectional heterogeneity and the covariates, we use the first differences transformation. However, since this transformation provides a regression model that can be considered as an additive function with the same functional form but evaluated in different periods we follow the proposal in Yang (2002). Thus, we present an estimator based on a local linear approximation and on the use of a higher-dimensional kernel weight.

As it is noted in Lee and Mukherjee (2008), direct application of linear approximation techniques to such differencing panel data models leads us to estimators that exhibit a bias that does not disappear even in large samples. However, the use of a higher-dimensional kernel weight enables to overcome this problem. Besides the usual distance between the fixed term around which the approximation is performed and a value of the sample, this technique also allows us to consider the sum of these distances with respect to the different values of the sample. Unfortunately, when we analyze the main asymptotic properties of this estimator we obtain that the bias problem actually disappear, but at the cost of generating an increase in the variance term. Therefore, this estimator achieves slower nonparametric rates of convergence. With the aim of achieving optimality, in this chapter we propose to ameliorate the dimensionality problem resorting to an additional smoothing through a one-step backfitting algorithm. Analyzing the asymptotic properties of this estimator it is shown that an additional smoothing provides estimators that achieve the optimal rate of convergence of this type of problems and exhibit the oracle efficiency property, as it is noted in Fan and Zhang (1999). Furthermore, because of the relevant role of the bandwidth matrix in the trade-off bias and variance, we provide a method that allows us to compute this matrix empirically. Finally, through a Monte Carlo experiment we corroborate the good performance of this estimator in finite samples. In particular, for T fixed, as far as the sample size increase the value of the AMSE is lower, as we expected from its asymptotic properties.

In Chapter 3, the main goal that we pursue is to provide estimators that exhibit the standard asymptotic properties of the nonparametric estimators under the assumption of the idiosyncratic error terms are *i.i.d.* To do it, we use a deviation from the mean transformation with the aim of removing the statistical dependence problem between the fixed effects and the explanatory variables of a varying coefficient panel data model. Again, to avoid the non-negligible asymptotic bias we propose a local linear regression estimator with a kernels function of dimension T. Since the resulting estimator achieves suboptimal rates of convergence, we propose to combine this procedure with a one-step backfitting algorithm that allows us to cancel asymptotically all the additive terms of the regression model in differences. Later, we show that this two-step estimator is asymptotically normal and achieves the optimal rate of convergence of this type of smooth functions. Also, it exhibits the oracle efficiency property. Finally, a Monte Carlo experiment enables us to confirm the theoretical findings and the good performance of the estimator in finite samples.

Once presented the differencing estimators that we propose in this doctoral dissertation (first-differences and fixed effects), in Chapter 4 we perform a comparative analysis about the behavior of both estimators. Analyzing the asymptotic properties of the two local linear regression estimators we can see that, although both maintain the same order of magnitude of the bias term, they show different asymptotic limits for the variance term. In this way, both estimators achieve suboptimal rates of convergence. On the contrary, for the two one-step backfitting estimators it is shown that, under fairly general conditions, both achieve optimality and are oracle efficient. Since both backfitting estimators are asymptotically equivalent, the analysis of their behavior in small sample sizes is very interesting.

In a fully parametric context, it is shown that under assumptions of strict exogeneity the choice between differencing estimators depends on the stochastic structure of the idiosyncratic error terms. However, in the asymptotic analysis of the nonparametric estimators we observe that apart from the stochastic structure, there exist other factors such as the dimensionality of the covariates and the sample size are of great interest. Specifically, to analyze the performance of these estimators in small sample sizes we focus on the evolution of the average mean square error (AMSE) under fairly different scenarios of the error term.

In view of the simulation results obtained, we can highlight that the AMSE of the two local linear regression estimators tends to converge under different specifications of the error term. Also, for small values of T, the AMSE of the first-differences estimator tends to dominate in terms of the AMSE of the fixed effects estimator. In this way, the fixed effects estimator is preferred, as we expected. On the contrary, when T increase the behavior of the two estimators worsens considerably, as the asymptotic limits of their variance terms set. On the other hand, regarding to the backfitting estimators is obtained that for the various types of errors considered, the two-step estimator has a better performance than the local linear regression estimator since it allows us to avoid the dimensionality problem. In addition, under *i.i.d.* structures of the error terms or AR(1) stationary process we see that the one-step backfitting fixed effects estimator has a better performance than the first-differences. When the error term follows a random walk the behavior of the estimator is better in the opposite sense. Thus, in short, we can say that the simulation results confirm the theoretical findings established previously for both local linear regression and one-step backfitting estimators. Also, we find that fixed effects estimators are quite sensitive to the size of the number of time observations per individual.

Finally, in Chapter 5 and with the aim of showing the empirical feasibility that these new procedures report for empirical analysis, we consider the nonparametric estimation of a structural model on Spanish household's precautionary savings, for the period $1985-1996$, motivated by the life-cycle hypothesis model of Modigliani and Brumberg (1954). Starting from this specification, we estimate a model where household's savings are related to both uncertainty about future healthcare expenses and household risk aversion. In order to contribute to the literature on precautionary savings, in this chapter we try to address the most common misspecification problems in such studies. To this end, we extend the semi-parametric model developed in Chou et al. (2004) to the analysis of panel data models. In this way, what we propose is to estimate a life-cycle hypothesis model that enables to state the household behavior under the following peculiarities: (i) unobserved cross-sectional heterogeneity correlated with the explanatory variables; (ii) unknown functions in the Euler equation that relates endogenous and exogenous covariates and that have to be estimated; and (iii) expense in healthcare products determined endogenously.

Starting from a varying coefficient panel data regression model with fixed effects, the estimator that we propose solve the endogeneity problem using the predicted values of the endogenous covariates generated in the nonparametric estimation of the reduced form equation. In this way, the resulting estimator has the simple form of a two-step weighted locally constant least-squares estimator. Also, certain marginal integration techniques are necessary to estimate a subset of the functionals of interest. To determine the behavior of both estimators we obtain their main asymptotic properties. Finally, the chapter conclude with a Monte Carlo experiment that corroborate the good performance of the proposed estimator in finite sample and an empirical application on Spanish household's precautionary savings for the period $1985-1996$. In view of the results, we obtain that from this empirical application we can conclude that households accumulate wealth at least in two periods of their life. On one hand, younger households save to guard to uncertainty about potential income downturn. On the other hand, older households save primarily for retirement or legacy reasons, corroborating what is established in the life-cycle hypothesis model of Modigliani and Brumberg (1954). In addition, the uncertainty reduction through public health programs has a negative impact on the savings. Finally, comparing the behavior of both estimators by educational level we obtain that households with low education levels are more risk averse than those with higher levels.

Future research

Throughout this work, and given the different gains offered by the varying coefficient panel data models of unknown form for several empirical studies, some future lines of research have arisen. In this regard, as it is shown in Card (2001), when we analyze the returns of educational is necessary to allow certain explanatory variables to evolve with other variables to obtain consistent results. In this way, a first study can consist on the extension of the regression model used in Card (2001) to a panel data model that can be estimate using the procedures developed in this doctoral dissertation.

Another line of relevant research is the development of some nonparametric tests. On one hand, in the chapters of this dissertation it has been shown that varying coefficient models are very useful because they allow us to exploit existing features in the data set, without resorting to assumptions so restrictive on model specification as fully parametric models. However, we have also shown that this flexibility has an added cost because it provides nonparametric estimators with slower rates of convergence. With the aim of testing empirically the correct specification of the model to analyze, it is desirable to develop a specification test that check parametric versus semi-parametric modeling. For this, following the idea of Henderson et al. (2008) a test based on the bootstrap technique can be very interesting.

We also know that when the unobserved individual heterogeneity is correlated with the covariates only first-differences estimators provide consistent results, whereas when individual effects are independent of the regressors both differencing estimators and random effects are consistent. However, working with varying coefficient models in differences is a cumbersome task so it is interesting to test which type of individual effects we are working with. In this way, a nonparametric test of fixed effects versus random effects that follows the style of the well-known Hausman test can be useful.

Finally, in Chapter 5 we have shown that local linear regression estimation techniques can be easily extended to the context of functionals coefficient models with endogenous variables. However, in economics and other social sciences it is very common to deal with situations in which the nonparametric covariate is determined endogenously; i.e., in the standard consumer selection problem. In this situation, we can assume that there is a vector of instrumental variables that enables us to solve this endogeneity problem. However, in nonparametric and semi-parametric models this is not so easy since we can incur in the ill-posed inverse problem.

Bibliography

Aaronson, D., S. Agarwal, and E. French (2012). "The spending and debt response to minimum wage hikes". *American Economic Review 102*(7), 3111–3139.

Ai, C. and Q. Li (2008). "Semiparametric and nonparametric methods in panel data models". In L. Mátyás and P. Semester (Eds.), *The Econometrics of Panel Data: Fundamentals and Revent Developments in Theory and Practice* (3rd ed.)., pp. 451–478. Sprinker.

Ai, C., J. You, and Y. Zhou (2013). "Estimation of fixed effects panel data partially linear additive regression models". Forthcoming article in *The Econometrics Journal.*

Arellano, M. (2003). *"Panew data econometrics"*. Oxford University Press.

Attanasio, O. A., J. Banks, C. Meghir, and G. Weber (1999). "Humps and bumps in lifetime consumption". *Journal of Business and Economic Statistics 17*(1), 22–35.

Balestra, P. and M. Nerlove (1966). "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas". *Econometrica 34*, 585–612.

Baltagi, B. (2013). *"Econometrics analysis of panel data"* (2nd ed.). Wiley, New York.

Baltagi, B. H. and Q. Li (2002). "Series estimation of partially linear panel data models with fixed effects". *Annals of Economic and Finance 3*, 103–116.

Blanchard, O. J. and S. Fisher (1989). *"Lecture on Macroeconomics"*. MIT Press: Cambridge, MA.

Caballero, R. (1990). "Consumption puzzles and precautionary savings". *Journal of Monetary Economics 25*(4), 859–871.

Cagetti, M. (2003). "Wealth accumulation over the life cycle and precautionary savings". *Journal of Business and Economics Statistics 21*(3), 339–353.

Cai, Z. (2002a). "A two-stage approach to additive time series models". *Statistica Neerlandica 56*, 415–433.

Cai, Z. (2002b). "Two-step likelihood estimation procedure for varying-coefficient models". *Journal of Multivariate Analysis 81*, 189–209.

Cai, Z., L. Cheng, and Y. Fang (2013). "Semiparametric estimation of partially varying-coefficient dynamic panel data models". Forthcoming article in *Econometric Reviews.*

Cai, Z., M. Das, H. Xiong, and X. Wu (2006). "Functional-coefficient instrumental variables models". *Journal of Econometrics 133*, 207–241.

Cai, Z. and Q. Li (2008). "Nonparametric estimation of varying coefficient dynamic panel data models". *Econometric Theory 24*, 1321–1342.

Cai, Z., Q. Li, and J. Y. Park (2009). "Functional-coefficient models for nonstationary time series data". *Journal of Econometrics 148*, 101–113.

Cai, Z. and H. Xiong (2012). "Partially varying coefficient instrumental variables models". *Statistica Neerlandica 66*, 85–110.

Card, D. (2001). "Estimating the return to schooling: progress on some persistent econometric problems". *Econometrica 69*, 1127–1160.

Carroll, C. (1997). "Buffer-stock saving and the life cycle/permanent income hypothesis". *Quarterly Journal of Economics 112*, 1–55.

Carroll, C. D. and A. A. Samwick (1997). "The nature of precautionary wealth". *Journal of Monetary Economics 40*, 41–71.

Chen, X. (2007). "Large sample sieve estimation of semi-nonparametric models". In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, pp. 5549–5632. North Holland, Amsterdam.

Chou, S., J. Liu, and K. Hammit (2003). "National health insurance and precautionary saving: evidence from Taiwan". *Journal of Public Economics 87*, 1873–1894.

Chou, S. Y., J. Liu, and C. Huang (2004). "Health insurance and savings over the life cycle: a semiparametric smooth coefficient estimation". *Journal of Applied Econometrics 19*, 295–322.

Cleveland, W., E. Grosse, and W. Chyu (1991). "Local regression model". In J. M. Chambers and T. J. Hastie (Eds.), *Statistical Models in S*, pp. 309–376. Wandsworth and Brooks, Pacific Grove.

Das, M. (2005). "Instrumental variables estimators of nonparametric models with discrete endogenous regressors". *Journal of Econometrics 124*, 33–361.

Deaton, A. S. (1992). *"Understanding consumption"*. Oxford University Press: New York.

Deaton, A. S. and C. Paxson (1994). "Saving, growth and aging in Taiwan". In D. A. Wise (Ed.), *Studies in the Economics of Aging*. University of Chicago Press, Chicago.

Egen, E. M. and J. Gruber (2001). "Unemployment insurance and precautionary saving". *Journal of Monetary Economics 47*, 545–579.

Fan, J. (1993). "Local linear regression smoothers and their minimax efficiencies". *Annals of Statistics 21*, 196–216.

Fan, J. and I. Gijbels (1992). "Variable bandwidth and local linear regression smoothers". *Annals of Statistics 20*, 2008–2036.

Fan, J. and I. Gijbels (1995a). "Date-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation". *Journal of the Royal Statistical Society 57*(Series B), 371–394.

Fan, J. and I. Gijbels (1995b). *"Local polynomial modelling and its applications"*. Chapman & Hall.

Fan, J. and W. Zhang (1999). "Statistical estimation in varying coefficient models". *Annals of Statistics 27*, 1491–1518.

Fan, R., Y. Fang, and S. Park (2010). "Resource abundance and economic growth in China". Working paper.

Friedman, M. (1957). *"A theory of the consumption function"*. Princeton: Princeton University Press.

Gao, Q., J. Gu, and P. Hernandez-Verme (2012). "A semiparametric time trend varying coefficients model: with an application to evaluate credit rationing in U.S. credit market". *Annals of Economics and Finance 13*, 189–210.

Gao, Y. and K. Li (2013). "Nonparametric estimation of fixed effects panel data models". *Journal of Nonparametric Statistics 25*(3), 679–693.

Gertler, P. and J. Gruber (2002). "Insuring consumption against illness". *American Economic Review 92*, 51–70.

Gourinchas, P. and J. Parker (2002). "Consumption over the life cycle". *Econometrica 70*, 47–89.

Gruber, J. (1997). "The consumption smoothing benefits of unemployment insurance". *American Economic Review 87*, 192–205.

Hall, P., S. J. Sheather, M. C. Jones, and J. S. Marron (1991). "On optimal data-based bandwidth selection in kernel density estimation". *Biometrika 78*, 263–271.

Hall, R. and F. Mishkin (1982). "The sensitivity of consumption to transitory income: estimates from panel data on households". *Econometrica 50*, 461–481.

Härdle, W. (1990). *"Applied nonparametric regression"*. Cambridge University Press.

Härdle, W., S. Huet, E. Mammen, and S. Sperlich (2004). "Bootstrap inference in semiparametric generalized additive models". *Econometric THenry 20*, 265–300.

Hastie, T. and R. Tibshirani (1990). *"Generalized additive models"*. Chapman and Hall, New York.

Henderson, D. J., R. J. Carroll, and Q. Li (2008). "Nonparametric estimation and testing of fixed effects panel data model". *Journal of Econometrics 144*, 257–275.

Henderson, D. J. and A. Ullah (2005). "A nonparametric random effects estimator". *Economics Letters 88*, 403–407.

Henderson, D. J. and A. Ullah (2012). "Nonparametric estimation in a one-way error component model: a Monte Carlo analysis". In A. Sengupta, T. Samanta, and A. Basu (Eds.), *Statistical Paradigms: Revent Advances and Reconciliations*, Volume 14 of *Statistical Science and Interdisciplinary Research*. Wold Scientific Publishing Company Incorporated.

Hengartner, N. W. and S. Sperlich (2005). "Rate optimal estimation with the integration method in the presence of many covariates". *Journal of Multivariate Analysis 95*, 246–272.

Hjellvik, V. and D. Tjøstheim (1999). "Modelling panels of intercorrelated autoregressive time series". *Biometrika 86*, 573–590.

Hsiao, C. (2003). *"Analysis of panel data"* (2nd ed.). Econometric Society Monographs.

Hubbard, G., J. S. Skinner, and S. P. Zeldes (1994). "The importance of precautionary motives for explaining individual and aggregate saving". In A. H. Meltzer and C. I. Plosser (Eds.), *Carnegie-Rochester Conference Series on Public Policy*, Volume 40, pp. 59–125. Amsterdam: North Holland.

Kottaridi, C. and T. Stengos (2010). "Foreign direct investment, human capital and nonlinearities in economic growth". *Journal of Macroeconomics 32*(3), 858–871.

Kuan, C. and C. Chen (2013). "Effects of national health insurance on precautionary saving: new evidence from Taiwan". *Empirical Economics 44*, 942–943.

Lee, Y. and D. Mukherjee (2008). "New nonparametric estimation of the marginal effects in fixed-effects panel models: an application on the environmental Kuznets curve". Revise and resubmit invited from *Journal of Econometrics*.

Li, G., H. Peng, and T. Tong (2013). "Simultaneous confidence band for nonparametric fixed effects panel data models". *Economics Letters 19*, 229–232.

Li, Q. and T. Stengos (1996). "On the root-n-consistent semiparametric estimation of partially linear models". *Journal of Econometrics 71*, 389–397.

Li, Q. and A. Ullah (1998). "Estimating partially linear panel data models with one-way error components". *Econometric Reviews 17*(2), 145–166.

Lin, X. and R. J. Carroll (2000). "Nonparametric function estimation for clustered data when the predictor is measured without/with error". *Journal of the American Statistical Association 95*, 520–534.

Lin, Z., Q. Li, and Y. Sun (2014). "A consistent nonparametric test of parametric regression functional form in fixed effects panel data models". *Journal of Econometrics 178*(1), 167–179.

Linton, O. and J. P. Nielsen (1995). "A kernel method of estimating structured nonparametric regression based on marginal integration". *Biometrika 82*, 93–101.

Maddala, G. S. (1987). "Revent developments in the econometrics of panel data analysis". *Transportation Research 21*, 303–326.

Mammen, E., O. Linton, and J. P. Nielsen (1999). "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions". *Annals of Statistics 27*, 1443–1490.

Mammen, E., B. Støve, and D. Tjøstheim (2009). "Nonparametric additive models for panels of time series". *Econometric Theory 25*, 442–481.

Martins-Filho, C. and F. Yao (2009). "Nonparametric regression estimation with general parametric error covariance". *Journal of Multivariate Analysis 100*, 309–333.

Masry, E. (1996). "Multivariate local polynomial regression for time series: uniform strong consistency rate". *Journal of Time Series Analysis 17*, 571–599.

Maynard, A. and J. Qiu (2009). "Public insurance and private savings: who is affected and by how much?". *Journal of Applied Econometrics 24*, 282–308.

Modigliani, F. and R. Brumberg (1954). Utility and the consumption function: an interpretation of cross-section data. In K. Kurihara (Ed.), *"Postkeynesian Economics"*, pp. 388–436. Rutgers University Press, News Runswick.

Müller, H. G. and U. Stadtmüller (1987). "Variable bandwidth kernel estimators of regression curves". *Annals of Statistics 15*, 182–201.

Mundra, K. (2005). Nonparametric slope estimation for fixed-effect panel data. Working paper, Department of Economics, San Diego State University.

Nadaraya, E. A. (1964). "On estimating regression". *Theory of Probability and its Applications 9*(1), 141–142.

Newey, W. K. (1994). "The asymptotic variance of semiparametric estimators". *Econometrica 62*, 1349–1682.

Neyman, J. and E. Scott (1948). "Consistent estimates based on partially consistent observations". *Econometrica 16*, 1–32.

Pagan, A. and A. Ullah (1999). *"Nonparametric econometrics"*. New York: Cambridge University Press.

Palumbo, M. G. (1999). "Uncertain medical expenses and precautionary saving: near the end of the life cycle". *Review of Economic Studies 66*, 395–422.

Powell, J. L., J. H. Stock, and T. M. Stocker (1989). "Semiparametric estimation of the index coefficients". *Econometrica 57*, 1043–1430.

Qian, J. and L. Wang (2012). "Estimating semiparametric panel data models by marginal integration". *Journal of Econometrics 167*, 483–493.

Robinson, P. M. (1988). "Root-n-consistent semiparametric regression". *Econometrica 56*(4), 931–954.

Ruckstuhl, A. F., A. H. Welsh, and R. J. Carroll (2000). "Nonparametric function estimation of the relationship between two repeatedly measured variables". *Statistica Sinica 10*, 51–71.

Ruppert, D., S. J. Sheather, and M. P. Wand (1995). "An effective bandwidth selector for local least-squares regression". *Journal of the American Statistical Association 90*, 1257–1270.

Ruppert, D. and M. P. Wand (1994). "Multivariate locally weighted least squares regression". *The Annals of Statistics 22*(3), 1346–1370.

Sachs, J. D. and A. M. Warner (2001). "The curse of natural resources". *European Economic Review 45*, 827–838.

Schultz, T. P. (2003). "Human capital, schooling and health". *Economics and Human Biology 1*, 207–221.

Sperlich, S., D. T. Tjøstheim, and L. Yang (2002). "Nonparametric estimation and testing of interaction in additive models". *Econometric Theory 18*, 197–251.

Starr-McCluer, M. (1996). "Health insurance and precautionary saving". *American Economic Review 86*, 285–295.

Su, L. and X. Lu (2013). "Nonparametric dynamic panel data models: Kernel estimation and specification testing". *Journal of Econometrics 176*, 112–133.

Su, L. and A. Ullah (2006a). "More efficient estimation in nonparametric regression with nonparametric autocorrelated errors". *Econometric Theory 22*, 98–126.

Su, L. and A. Ullah (2006b). "Profile likelihood estimation of partially linear panel data models with fixed effects". *Economics Letters 92*, 75–81.

Su, L. and A. Ullah (2007). "More efficient estimation of nonparametric panel data models with random effects". *Economics Letters 96*, 375–380.

Su, L. and A. Ullah (2011). "Nonparametric and semiparametric panel econometric models: estimation and testing". In A. Ullah and D. E. Giles (Eds.), *Handbook of Empirical Economics and Finance*, pp. 455–497. New York: Taylor and Francis Group.

Sun, Y., R. J. Carroll, and D. Li (2009). "Semiparametric estimation of fixed effects panel data varying coefficient models". *Advances in Econometrics 25*, 101–130.

Tran, K. C. and E. G. Tsiomas (2010). "Local GMM estimation of semiparametric panel data with smooth coefficient models". *Econometric Reviews 29*, 39–61.

Ullah, A. and N. Roy (1998). "Nonparametric and semiparametric econometrics of panel data". In A. Ullah and D. E. Giles (Eds.), *Handbook of Applied Economic Statistics*, pp. 579–604. New York: Marcel Dekker.

Watson, G. S. (1964). "Smooth regression analysis". *Sankhya: The Indian Journal of Statistics 26*(4), 359–372.

Wooldridge, J. (2005). "Fixed effects and related estimators for correlated random-coefficient and treatment effect panel data models". *Review of Economics and Statistics 87*, 385–390.

Wooldridge, J. M. (2003). *"Econometric analysis of cross section and panel data models"*. The MIT Press, Cambridge, Massachusetts.

Yang, L. (2002). "Direct estimation in an additive model when the components are proportional". *Statistica Sinica 12*, 801–821.

You, J., X. Zhou, and Y. Zhou (2010). "Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors". *Journal of Multivariate Analysis 101*, 1079–1101.

Zhan-Qian, L. (1996). "Multivariate locally weighted polynomial fitting and partial derivative estimation". *Journal of Multivariate Analysis 59*, 187–205.

Zhang, J., S. Feng, G. Li, and H. Lian (2011). "Empirical likelihood inference for partially linear panel data models with fixed effects". *Economics Letters 113*, 165–167.

Zhang, W. and S. Lee (2000). "Variable bandwidth selection in varying-coefficient models". *Journal of Multivariate Analysis 74*, 116–134.

Zhou, B., J. You, Q. Xu, and G. Chen (2010). "Weighted profile least squares estimation for a panel data varying-coefficient partially linear model". *Chinese Annals of Mathematics 31B*(2), 247–272.

Appendix

Appendix 1

Proof of Theorem 2.1

Taking Assumption **A.2** conditional expectations in (2.14) and noting that

$$E\left[v_{it}|\mathbb{X},\mathbb{Z}\right]=0,\qquad i=1,\cdots,N\qquad;\qquad t=2,\cdots,T,$$

then

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right]=e_1^{\top}\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}\widetilde{Z}^{\top}WM \tag{A1.1}$$

where $M=\left(X_{12}^{\top}m(Z_{12})-X_{11}^{\top}m(Z_{11}),...,X_{NT}^{\top}m(Z_{NT})-X_{N(T-1)}^{\top}m(Z_{N(T-1)})\right)^{\top}$.

Taylor's Theorem implies that

$$M=\widetilde{Z}\left[\begin{array}{c} m(z)\\ \mathrm{vec}(D_m(z))\end{array}\right]+\frac{1}{2}Q_m(z)+R(z), \tag{A1.2}$$

where

$$Q_m(z)=S_{m1}(z)-S_{m2}(z), \tag{A1.3}$$

$$\begin{aligned} S_{m1}(z) &= \left(S_{m1,12}^{\top}(z),\cdots,S_{m1,NT}^{\top}(z)\right)^{\top},\\ S_{m2}(z) &= \left(S_{m2,11}^{\top}(z),\cdots,S_{m2,N(T-1)}^{\top}(z)\right)^{\top} \end{aligned}$$

and

$$\begin{aligned} S_{m1,it}(z) &= \left((X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{H}_m(z)(Z_{it}-z)\right),\\ S_{m2,i(t-1)}(z) &= \left((X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^{\top}\mathcal{H}_m(z)(Z_{i(t-1)}-z)\right). \end{aligned}$$

We denote by

$$\mathcal{H}_m(z) = \begin{pmatrix} \mathcal{H}_{m1}(z) \\ \mathcal{H}_{m2}(z) \\ \vdots \\ \mathcal{H}_{md}(z) \end{pmatrix},$$

a $dq \times q$ matrix such that $\mathcal{H}_{md}(z)$ is the Hessian matrix of the dth component of $m(\cdot)$.

The remainder term can be written as

$$R(z) = R_1(z) - R_2(z), \tag{A1.4}$$

$$\begin{aligned} R_1(z) &= \left(R_{1,12}^\top(z), ..., R_{1,NT}^\top(z)\right)^\top, \\ R_2(z) &= \left(R_{2,11}^\top(z), ..., R_{2,N(T-1)}^\top(z)\right)^\top \end{aligned}$$

and

$$\begin{aligned} R_{1,it}(z) &= \left((X_{it} \otimes (Z_{it} - z))^\top \mathcal{R}\left(Z_{it}; z\right)(Z_{it} - z)\right), \\ R_{2,i(t-1)}(z) &= \left((X_{i(t-1)} \otimes (Z_{i(t-1)} - z))^\top \mathcal{R}\left(Z_{i(t-1)}; z\right)\right). \end{aligned}$$

We denote by

$$\mathcal{R}(Z_{it}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{it}; z) \\ \mathcal{R}_2(Z_{it}; z) \\ \vdots \\ \mathcal{R}_d(Z_{it}; z) \end{pmatrix}, \mathcal{R}(Z_{i(t-1)}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{i(t-1)}; z) \\ \mathcal{R}_2(Z_{i(t-1)}; z) \\ \vdots \\ \mathcal{R}_d(Z_{i(t-1)}; z) \end{pmatrix},$$

and

$$\begin{aligned} \mathcal{R}_d(Z_{it}; z) &= \int_0^1 \left(\frac{\partial^2 m_d}{\partial z \partial z^\top}(z + \omega(Z_{it} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^\top}(z)\right)(1-\omega)d\omega, \\ \mathcal{R}_d(Z_{i(t-1)}; z) &= \int_0^1 \left(\frac{\partial^2 m_d}{\partial z \partial z^\top}(z + \omega(Z_{i(t-1)} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^\top}(z)\right)(1-\omega)d\omega. \end{aligned} \tag{A1.5}$$

First, we analyze the bias term. In order to do this, we substitute (A1.2) into (A1.1) and noting that $vec(D_m(z))$ in (A1.2) vanishes because

$$e_1^\top(\widetilde{Z}^\top W\widetilde{Z})^{-1}\widetilde{Z}^\top W\widetilde{Z}\begin{bmatrix} m(z) \\ vec(D_m(z)) \end{bmatrix} = e_1^\top \begin{bmatrix} m(z) \\ vec(D_m(z)) \end{bmatrix} = m(z), \qquad \text{(A1.6)}$$

then,

$$\begin{aligned} E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) &= \frac{1}{2}e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WQ_m(z) \\ &+ e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WR(z). \qquad \text{(A1.7)} \end{aligned}$$

We first analyze the asymptotic behavior of $\frac{1}{2}e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WQ_m(z)$. Later in the appendix, we do the same with the second term. For the sake of simplicity, let us denote

$$K_{it} = \frac{1}{|H|^{1/2}}K\left(H^{-1/2}(Z_{it}-z)\right),$$

now, we define the symmetric block matrix

$$(NT)^{-1}\widetilde{Z}^\top W\widetilde{Z} = \begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{21} & \mathcal{A}_{NT}^{22} \end{pmatrix} \qquad \text{(A1.8)}$$

where,

$$\begin{aligned} \mathcal{A}_{NT}^{11} &= (NT)^{-1}\sum_{it}\Delta X_{it}\Delta X_{it}^\top K_{it}K_{i(t-1)}, \\ \mathcal{A}_{NT}^{12} &= (NT)^{-1}\sum_{it}\Delta X_{it}\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^\top K_{it}K_{i(t-1)}, \\ \mathcal{A}_{NT}^{21} &= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)\Delta X_{it}^\top K_{it}K_{i(t-1)}, \\ \mathcal{A}_{NT}^{22} &= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \\ &\times\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^\top K_{it}K_{i(t-1)}. \end{aligned}$$

We show that as N tends to infinity

$$\mathcal{A}_{NT}^{11} = \mathcal{B}_{\Delta X\Delta X}(z,z) + o_p(1). \qquad \text{(A1.9)}$$

where,

$$\mathcal{B}_{\Delta X\Delta X}(z,z) = E\left[\Delta X_{it}\Delta X_{it}^\top|Z_{it}=z, Z_{i(t-1)}=z\right]f_{Z_{it},Z_{i(t-1)}}(z,z).$$

In order to do so, note that under the stationarity assumption and using iterated expectations

$$\begin{aligned} E\left(\mathcal{A}_{NT}^{11}\right) &= \int\int E\left[\Delta X_{it}\Delta X_{it}^{\top}|Z_{it}=z+H^{1/2}u, Z_{i(t-1)}=z+H^{1/2}v\right] \\ &\quad \times f_{Z_{it},Z_{i(t-1)}}\left(Z_{it}=z+H^{1/2}u, Z_{i(t-1)}=z+H^{1/2}v\right)K(u)K(v)dudv. \end{aligned}$$

Furthermore, under Assumptions **2.1** and **2.4** and a Taylor expansion, as N tends to infinity, (A1.9) holds. All what we need to close the proof is to show that Var $\left(\mathcal{A}_{NT}^{11}\right)\to 0$, as the sample size tends to infinity. Note that under Assumption **2.1**,

$$\begin{aligned} \text{Var}\left(\mathcal{A}_{NT}^{11}\right) &= \frac{1}{NT}\text{Var}\left(\Delta X_{it}\Delta X_{it}^{\top}K_{it}K_{i(t-1)}\right) \\ &+ \frac{1}{NT^2}\sum_{t=3}(T-t)\text{Cov}\left(\Delta X_{i2}\Delta X_{i2}^{\top}K_{i2}K_{i1}, \Delta X_{it}\Delta X_{it}^{\top}K_{it}K_{i(t-1)}\right). \end{aligned}$$

Under Assumptions **2.4-2.6**

$$\text{Var}\left(\Delta X_{it}\Delta X_{it}^{\top}K_{it}K_{i(t-1)}\right)\leq\frac{C}{NT|H|}.$$

and

$$\text{Cov}\left(\Delta X_{i2}\Delta X_{i2}^{\top}K_{i2}K_{i1}, \Delta X_{it}\Delta X_{it}^{\top}K_{it}K_{i(t-1)}\right)\leq\frac{C'}{N|H|}.$$

Then, if both $NT|H|$ and $N|H|$ tend to infinity the variance tends to zero and (A1.9) holds.

Similarly, we can show that

$$\mathcal{A}_{NT}^{12}=\mathcal{DB}_{\Delta XX}(z,z)\left(I_d\otimes\mu_2\left(K_u\right)H\right)-\mathcal{DB}_{\Delta XX_{-1}}(z,z)\left(I_d\otimes\mu_2\left(K_v\right)H\right)+o_p(H). \tag{A1.10}$$

Here, $\mathcal{DB}_{\Delta XX}(Z_1,Z_2)$ and $\mathcal{DB}_{\Delta XX_{-1}}(Z_1,Z_2)$ are $d\times dq$ gradient matrices. We define $\mathcal{DB}_{\Delta XX}(Z_1,Z_2)$ as

$$\mathcal{DB}_{\Delta XX}(Z_1,Z_2)=\begin{pmatrix} \frac{\partial b_{11}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^{\top}} & \cdots & \frac{\partial b_{1d}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^{\top}} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{d1}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^{\top}} & \cdots & \frac{\partial b_{dd'}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^{\top}} \end{pmatrix},$$

and

$$b_{dd'}^{\Delta XX}(Z_1,Z_2)=E\left[\Delta X_{dit}X_{d'it}|Z_{it}=Z_1,Z_{i(t-1)}=Z_2\right]f_{Z_{it},Z_{i(t-1)}}\left(Z_1,Z_2\right).$$

The $\mathcal{DB}_{\Delta XX_{-1}}(Z_1, Z_2)$ gradient matrix is

$$\mathcal{DB}_{\Delta XX_{-1}}(Z_1, Z_2) = \begin{pmatrix} \frac{\partial b_{11}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^\top} & \dots & \frac{\partial b_{1d}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^\top} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{d1}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^\top} & \dots & \frac{\partial b_{dd}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^\top} \end{pmatrix},$$

and

$$b_{dd'}^{\Delta XX_{-1}}(Z_1, Z_2) = E\left[\Delta X_{dit} X_{d'i(t-1)} | Z_{it} = Z_1, Z_{i(t-1)} = Z_2\right] f_{Z_{it}, Z_{i(t-1)}}(Z_1, Z_2).$$

Finally,

$$\mathcal{A}_{NT}^{22} = \mathcal{B}_{XX}(z,z) \otimes \mu_2(K_u)H + \mathcal{B}_{X_{-1}X_{-1}}(z,z) \otimes \mu_2(K_v)H + o_p(H), \tag{A1.11}$$

where

$$\begin{aligned} \mathcal{B}_{XX}(z,z) &= E\left[X_{it}X_{it}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z), \\ \mathcal{B}_{X_{-1}X_{-1}}(z,z) &= E\left[X_{i(t-1)}X_{i(t-1)}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z). \end{aligned}$$

Using the results shown in (A1.9), (A1.10) and (A1.11), we obtain

$$NT\left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{21} & \mathcal{C}_{22} \end{pmatrix}, \tag{A1.12}$$

where

$$\begin{aligned} \mathcal{C}_{11} &= \mathcal{B}_{\Delta X \Delta X}^{-1}(z,z) + o_p(1), \\ \mathcal{C}_{12} &= -\mathcal{B}_{\Delta X \Delta X}^{-1}(z,z)\left(\mathcal{DB}_{\Delta XX}(z,z)\left(I_d \otimes \mu_2\left(K_u\right)H\right) - \mathcal{DB}_{\Delta XX_{-1}}(z,z)\left(I_d \otimes \mu_2\left(K_v\right)H\right)\right) \\ &\quad \times \left(\left(\mathcal{B}_{XX}(z,z) \otimes \mu_2\left(K_u\right)H + \mathcal{B}^{X_{-1}X_{-1}}(z,z) \otimes \mu_2\left(K_v\right)H\right)\right)^{-1} + o_p(1), \\ \mathcal{C}_{21} &= \left(\left(\mathcal{B}_{XX}(z,z) \otimes \mu_2\left(K_u\right)H + \mathcal{B}_{X_{-1}X_{-1}}(z,z) \otimes \mu_2\left(K_v\right)H\right)\right)^{-1} \\ &\quad \times \left(\mathcal{DB}_{\Delta XX}(z,z)\left(I_d \otimes \mu_2\left(K_u\right)H\right) - \mathcal{DB}_{\Delta XX_{-1}}(z,z)\left(I_d \otimes \mu_2\left(K_v\right)H\right)\right)^\top \mathcal{B}_{\Delta X \Delta X}^{-1}(z,z) \\ &\quad + o_p(1), \\ \mathcal{C}_{22} &= \left(\left(\mathcal{B}_{XX}(z,z) \otimes \mu_2(K_u)H + \mathcal{B}_{X_{-1}X_{-1}}(z,z) \otimes \mu_2(K_v)H\right)\right)^{-1} + o_p\left(H^{-1}\right). \end{aligned}$$

Also, it is straightforward to show that the terms in

$$(NT)^{-1}\widetilde{Z}^{\top}WS_{m1}(z)$$

$$= \begin{pmatrix} (NT)^{-1}\sum_{it}\Delta X_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)(Z_{it}-z)K_{it}K_{i(t-1)} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z) \\ \qquad\times(Z_{it}-z)K_{it}K_{i(t-1)} \end{pmatrix},$$

are asymptotically equal to

$$\begin{aligned} & (NT)^{-1}\sum_{it}\Delta X_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)(Z_{it}-z)K_{it}K_{i(t-1)} \\ = \; & \mu_2\left(K_u\right)E\left[\Delta X_{it}X_{it}^{\top}|Z_{it}=z,Z_{i(t-1)}=z\right]f_{Z_{it},Z_{i(t-1)}}(z,z)\,diag_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)H)\right)i_d \\ + \; & o_p\left(\mathrm{tr}(H)\right), \end{aligned} \tag{A1.13}$$

where $\mathrm{diag}_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)H)\right)$ stands for a diagonal matrix of elements $\mathrm{tr}(\mathcal{H}_{m_r}(z)H)$, for $r=1,\cdots,d$, and i_d is a $d\times 1$ unit vector, and

$$\begin{aligned} & (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z) \\ & \qquad\times(Z_{it}-z)K_{it}K_{i(t-1)} \\ = \; & \int\mathcal{B}_{XX}(z,z)\otimes(H^{1/2}u)(H^{1/2}u)^{\top}\mathcal{H}_m(z)(H^{1/2}u)K(u)K(v)dudv \\ - \; & \int\mathcal{B}_{X_{-1}X}(z,z)\otimes(H^{1/2}v)(H^{1/2}u)^{\top}\mathcal{H}_m(z)(H^{1/2}u)K(u)K(v)dudv+o_p(H^{3/2}) \\ = \; & O_p(H^{3/2}). \end{aligned} \tag{A1.14}$$

Finally, the terms in

$$(NT)^{-1}\widetilde{Z}^{\top}WS_{m2}(z)=$$

$$\begin{pmatrix} (NT)^{-1}\sum_{it}\Delta X_{it}\left(X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^{\top}\mathcal{H}_m(z)(Z_{i(t-1)}-z)K_{it}K_{i(t-1)} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)\left(X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^{\top}\mathcal{H}_m(z) \\ \qquad\times(Z_{i(t-1)}-z)K_{it}K_{i(t-1)} \end{pmatrix}$$

are of order

$$\begin{aligned} & (NT)^{-1}\sum_{it}\Delta X_{it}\left(X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^{\top}\mathcal{H}_m(z)(Z_{i(t-1)}-z)K_{it}K_{i(t-1)} \\ = \; & \mu_2\left(K_v\right)E\left[\Delta X_{it}X_{i(t-1)}^{\top}|Z_{it}=z,Z_{i(t-1)}=z\right]f_{Z_{it},Z_{i(t-1)}}(z,z)diag_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)H)\right)i_d \\ + \; & o_p\left(\mathrm{tr}(H)\right), \end{aligned} \tag{A1.15}$$

and

$$
\begin{aligned}
& (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)\left(X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^{\top}\mathcal{H}_m(z) \\
& \qquad \times(Z_{i(t-1)}-z)K_{it}K_{i(t-1)} \\
= & \int\mathcal{B}_{XX_{-1}}(z,z)\otimes(H^{1/2}u)(H^{1/2}v)^{\top}\mathcal{H}_m(z)(H^{1/2}v)K(u)K(v)dudv \\
- & \int\mathcal{B}_{X_{-1}X_{-1}}(z,z)\otimes(H^{1/2}v)(H^{1/2}v)^{\top}\mathcal{H}_m(z)(H^{1/2}v)K(u)K(v)dudv+o_p(H^{3/2}) \\
= & O_p(H^{3/2}). \qquad\qquad (A1.16)
\end{aligned}
$$

The second term for the bias expression is $e_1^{\top}\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}\widetilde{Z}^{\top}WR(z)$. We already know what is the asymptotic expression for $\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}$, so now we proceed to analyze the asymptotic behavior of $\widetilde{Z}^{\top}WR(z)$. According to (A1.4) and (A1.5), note that

$$
(NT)^{-1}\widetilde{Z}^{\top}WR(z)=\begin{pmatrix}\mathcal{E}_1(z)\\ \mathcal{E}_2(z)\end{pmatrix},
$$

where

$$
\begin{aligned}
& \mathcal{E}_1(z)=\frac{1}{NT}\sum_{it}\Delta X_{it} \\
& \times\Big((X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)-(X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^{\top}\mathcal{R}(Z_{i(t-1)};z)(Z_{i(t-1)}-z)\Big) \\
& \times K_{it}K_{i(t-1)} \qquad\qquad (A1.17)
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathcal{E}_2(z)=\frac{1}{NT}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \\
& \times\Big((X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)-(X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^{\top}\mathcal{R}(Z_{i(t-1)};z)(Z_{i(t-1)}-z)\Big) \\
& \times K_{it}K_{i(t-1)}. \qquad\qquad (A1.18)
\end{aligned}
$$

Note that

$$
\mathcal{E}_1(z)=\mathcal{E}_{11}(z)+\mathcal{E}_{12}(z), \qquad (A1.19)
$$

where

$$
\begin{aligned}
& \mathcal{E}_{11}(z)=\frac{1}{NT}\sum_{it}K_{it}K_{i(t-1)}\Delta X_{it} \qquad\qquad (A1.20) \\
& \times\Big((X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)-(X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{i(t-1)}-z)\Big)
\end{aligned}
$$

and

$$\begin{aligned}\mathcal{E}_{12}(z) &= \frac{1}{NT}\sum_{it} K_{it}K_{i(t-1)}\Delta X_{it} \qquad (A1.21)\\ &\times \left((X_{i(t-1)} \otimes (Z_{i(t-1)}-z))^\top (\mathcal{R}(Z_{it};z)-\mathcal{R}(Z_{i(t-1)};z))(Z_{i(t-1)}-z)\right).\end{aligned}$$

Now, we show that, as N tends to infinity,

$$E[\mathcal{E}_1(z)] = o_p\left(\mathrm{tr}(H)\right). \qquad (A1.22)$$

In order to prove this, note that

$$\begin{aligned}&E[\mathcal{E}_{11}(z)]\\ =& \int\int K(u)K(v)\left(\mathcal{B}_{\Delta XX}(z+H^{1/2}u, z+H^{1/2}v)\otimes (H^{1/2}u)^\top\right)\mathcal{R}(z+H^{1/2}u;z)(H^{1/2}u)dudv\\ -& \int\int K(u)K(v)\left(\mathcal{B}_{\Delta XX_{-1}}(z+H^{1/2}u, z+H^{1/2}v)\otimes (H^{1/2}v)^\top\right)\mathcal{R}(z+H^{1/2}u;z)(H^{1/2}v)dudv.\end{aligned}$$

By (A1.5) and Assumption **2.7**,

$$|\mathcal{R}_d(z+H^{1/2}u;z)| \leq \int_0^1 \varsigma(\omega\|H^{1/2}u\|)(1-\omega)d\omega, \forall d,$$

where $\varsigma(\eta)$ is the modulus of continuity of $(\partial^2 m_d/\partial z_i\partial z_j)(z)$. Hence, by boundedness of f, $\mathcal{B}_{\Delta XX}$ and $\mathcal{B}_{\Delta XX_{-1}}$

$$\begin{aligned}|E[\mathcal{E}_{11}(z)]| &\leq C_1\int\int\int_0^1 |(H^{1/2}u)^\top||\varsigma(\omega\|H^{1/2}u\|)||H^{1/2}u|K(u)K(v)d\omega dudv\\ &+ C_2\int\int\int_0^1 |(H^{1/2}v)^\top||\varsigma(\omega\|H^{1/2}u\|)||H^{1/2}v|K(u)K(v)d\omega dudv.\end{aligned}$$

Also, $E[\mathcal{E}_{11}(z)] = o_p\left(\mathrm{tr}(H)\right)$ follows by dominated convergence.

Similarly,

$$\begin{aligned}E[\mathcal{E}_{12}(z)] &= \int\int K(u)K(v)\left(\mathcal{B}_{\Delta XX_{-1}}(z+H^{1/2}u, z+H^{1/2}v)\otimes (H^{1/2}v)^\top\right)\\ &\times\left(\mathcal{R}(z+H^{1/2}u;z)-\mathcal{R}(z+H^{1/2}v;z)\right)(H^{1/2}v)dudv.\end{aligned}$$

Therefore,

$$\begin{aligned}|E[\mathcal{E}_{12}(z)]| &\leq C_3\int\int\int_0^1 |(H^{1/2}v)^\top||\varsigma(\omega\|H^{1/2}u\|)||H^{1/2}v|K(u)K(v)d\omega dudv\\ &+ C_4\int\int\int_0^1 |(H^{1/2}v)^\top||\varsigma(\omega\|H^{1/2}v\|)||H^{1/2}v|K(u)K(v)d\omega dudv.\end{aligned}$$

Then, proceeding as in the proof of the previous result, we also find that $E[\mathcal{E}_{12}(z] = o_p\left(\mathrm{tr}(H)\right)$.

Now, for $\mathcal{E}_2(z)$, note that

$$\mathcal{E}_2(z) = \mathcal{E}_{21}(z) + \mathcal{E}_{22}(z), \tag{A1.23}$$

where

$$\begin{aligned}\mathcal{E}_{21}(z) &= \frac{1}{NT}\sum_{it}\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \qquad \text{(A1.24)}\\ &\times \Big((X_{it}\otimes(Z_{it}-z))^\top \mathcal{R}(Z_{it};z)(Z_{it}-z) - (X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^\top \mathcal{R}(Z_{it};z)(Z_{i(t-1)}-z)\Big).\end{aligned}$$

and

$$\begin{aligned}\mathcal{E}_{22}(z) &= \frac{1}{NT}\sum_{it}\left(X_{it}\otimes(Z_{it}-z) - X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \qquad \text{(A1.25)}\\ &\times \Big((X_{i(t-1)}\otimes(Z_{i(t-1)}-z))^\top (\mathcal{R}(Z_{it};z) - \mathcal{R}(Z_{i(t-1)};z))(Z_{i(t-1)}-z)\Big).\end{aligned}$$

Following the same lines as for the proof of (A1.22), it is easy to show that

$$E[\mathcal{E}_2(z)] = o_p(H^{3/2}). \tag{A1.26}$$

Substituting (A1.12), (A1.13), (A1.15) and (A1.22) into (A1.7), the asymptotic bias can be written as

$$\begin{aligned}& E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z)\\ = \;& \frac{1}{2}e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\left(S_{m_1}(z) - S_{m_2}(z)\right)\\ = \;& \frac{1}{2}\mathcal{B}^{-1}_{\Delta X\Delta X}(z,z)\left(\mu_2(K_u)\mathcal{B}_{\Delta XX}(z,z) - \mu_2(K_v)\mathcal{B}_{\Delta XX_{-1}}(z,z)\right) diag_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)H)\right) i_d\\ + \;& o_p(\mathrm{tr}(H)).\end{aligned}$$

To obtain an asymptotic expression for the variance, let us first define the $(N(T-1)\times 1)$-vector $\Delta v = (\Delta v_1, \cdots, \Delta v_N)^\top$ where $\Delta v_i = (\Delta v_{i2}, \cdots, \Delta v_{iT})^\top$ and let $E\left(\Delta v \Delta v^\top|\mathbb{X},\mathbb{Z}\right) = \mathcal{V}$ be a $N(T-1)\times N(T-1)$ matrix that contains the V_{ij}'s matrices

$$V_{ij} = E(\Delta v_i \Delta v_j^\top|\mathbb{X},\mathbb{Z}) = \begin{cases} 2\sigma_v^2, & \text{for} \quad i=j, \quad t=s\\ -\sigma_v^2, & \text{for} \quad i=j, \quad |t=s|<2.\\ 0, & \text{for} \quad i=j, \quad |t=s|\geq 2.\end{cases} \tag{A1.27}$$

Then, taking into account that

$$\widehat{m}(z;H) - E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \Delta v, \tag{A1.28}$$

the variance of $\widehat{m}(z;H)$ can be written as

$$Var\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \mathcal{V} W^\top \widetilde{Z} \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} e_1. \tag{A1.29}$$

Based on Assumption **2.2** and by the fact that the v_{it} are i.i.d., then the upper left entry of $(1/NT)\widetilde{Z}^\top W \mathcal{V} W^\top \widetilde{Z}$ is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT} \sum_{it} \Delta X_{it} \Delta X_{it}^\top K_{it}^2 K_{i(t-1)}^2 - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^\top K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \Delta X_{i(t-2)}^\top K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} \\
= \quad & \frac{2\sigma_v^2 R(K_u) R(K_v)}{|H|} \mathcal{B}_{\Delta X \Delta X}(z,z)(1+o_p(1)),
\end{aligned} \tag{A1.30}$$

because

$$\frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^\top K_{it} K_{i(t-1)}^2 K_{i(t-2)} = \frac{\sigma_v^2 R(K_v)}{|H|^{1/2}} \mathcal{B}_{\Delta X \Delta X_{-1}}(z,z,z)(1+o_p(1)), \tag{A1.31}$$

and

$$\frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \Delta X_{i(t-2)}^\top K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} = \sigma_v^2 R(K_v) \mathcal{B}_{\Delta X \Delta X_{-2}}(z,z,z,z)(1+o_p(1)). \tag{A1.32}$$

The upper right block is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT} \sum_{it} \Delta X_{it} \left(X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right)^\top K_{it}^2 K_{i(t-1)}^2 \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \left(X_{i(t-1)} \otimes (Z_{i(t-1)} - z) - X_{i(t-2)} \otimes (Z_{i(t-2)} - z)\right)^\top K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \left(X_{i(t-2)} \otimes (Z_{i(t-2)} - z) - X_{i(t-3)} \otimes (Z_{i(t-3)} - z)\right)^\top K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} \\
& = \mathbf{I}_1 - \mathbf{I}_2 - \mathbf{I}_3.
\end{aligned} \tag{A1.33}$$

$$\begin{aligned}
\mathbf{I}_1 &= \frac{\sigma_v^2}{|H|}\int\Big(\mathcal{B}_{\Delta XX}(z+H^{1/2}u,z+H^{1/2}v)\otimes(H^{1/2}u)^\top \\
&- \mathcal{B}_{\Delta XX_{-1}}(z+H^{1/2}u,z+H^{1/2}v)\otimes(H^{1/2}v)^\top\Big)K^2(u)K^2(v)dudv(1+o_p(1)) \\
&= O_p(|H|), \qquad\qquad (A1.34)\\
\mathbf{I}_2 &= \frac{\sigma_v^2}{|H|^{1/2}}\int\Big(\mathcal{B}_{\Delta XX_{-1}}(z+H^{1/2}u,z+H^{1/2}v,z+H^{1/2}w)\otimes(H^{1/2}v)^\top \\
&- \mathcal{B}_{\Delta XX_{-2}}(z+H^{1/2}u,z+H^{1/2}v,z+H^{1/2}w)\otimes(H^{1/2}w)^\top\Big)K(u)K^2(v)K(w) \\
&\quad \times dudvdw(1+o_p(1)) \\
&= O_p(|H|^{1/2}), \qquad\qquad (A1.35)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{I}_3 &= \sigma_v^2\int\Big(\mathcal{B}_{\Delta XX_{-2}}(z+H^{1/2}u,z+H^{1/2}v,z+H^{1/2}w,z+H^{1/2}s)\otimes(H^{1/2}w)^\top \\
&- \mathcal{B}_{\Delta XX_{-3}}(z+H^{1/2}u,z+H^{1/2}v,z+H^{1/2}w,z+H^{1/2}s)\otimes(H^{1/2}s)^\top\Big) \\
&\quad \times K(u)K(v)K(w)K(s)dudvdwds(1+o_p(1)) \\
&= O_p(1). \qquad\qquad (A1.36)
\end{aligned}$$

Note that we denote,

$$\begin{aligned}
&\frac{\mathcal{B}_{\Delta XX}(z,z)}{f_{Z_{it},Z_{i(t-1)}}(z,z)} = E\left[\Delta X_{it}X_{it}^\top|Z_{it}=z,Z_{i(t-1)}=z\right], \\
&\frac{\mathcal{B}_{\Delta XX_{-1}}(z,z)}{f_{Z_{it},Z_{i(t-1)}}(z,z)} = E\left[\Delta X_{it}X_{i(t-1)}^\top|Z_{it}=z,Z_{i(t-1)}=z\right], \\
&\frac{\mathcal{B}_{\Delta XX_{-1}}(z,z,z)}{f_{Z_{it},Z_{i(t-1)},Z_{i(t-2)}}(z,z,z)} = E\left[\Delta X_{it}X_{i(t-1)}^\top|Z_{it}=z,Z_{i(t-1)}=z,Z_{i(t-2)}=z\right], \\
&\frac{\mathcal{B}_{\Delta XX_{-2}}(z,z,z)}{f_{Z_{it},Z_{i(t-1)},Z_{i(t-2)}}(z,z,z)} = E\left[\Delta X_{it}X_{i(t-2)}^\top|Z_{it}=z,Z_{i(t-1)}=z,Z_{i(t-2)}=z\right] \\
&\frac{\mathcal{B}_{\Delta XX_{-2}}(z,z,z,z)}{f_{Z_{it},Z_{i(t-1)},Z_{i(t-2)},Z_{i(t-3)}}(z,z,z,z)} \\
&\qquad = E\left[\Delta X_{it}X_{i(t-2)}^\top|Z_{it}=z,Z_{i(t-1)}=z,Z_{i(t-2)}=z,Z_{i(t-3)}=z\right] \\
&\frac{\mathcal{B}_{\Delta XX_{-3}}(z,z,z,z)}{f_{Z_{it},Z_{i(t-1)},Z_{i(t-2)},Z_{i(t-3)},Z_{i(t-4)}}(z,z,z,z)} \\
&\qquad = E\left[\Delta X_{it}X_{i(t-3)}^\top|Z_{it}=z,Z_{i(t-1)}=z,Z_{i(t-2)}=z,Z_{i(t-3)}=z,Z_{i(t-4)}=z\right].
\end{aligned}$$

Finally, the lower-right block is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \\
& \times\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right)^{\top}K_{it}^2K_{i(t-1)}^2 \\
- & \frac{\sigma_v^2}{NT}\sum_{i}\sum_{t=3}^{T}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \\
& \times\left(X_{i(t-1)}\otimes(Z_{i(t-1)}-z)-X_{i(t-2)}\otimes(Z_{i(t-2)}-z)\right)^{\top}K_{it}K_{i(t-1)}^2K_{i(t-2)} \\
- & \frac{\sigma_v^2}{NT}\sum_{i}\sum_{t=4}^{T}\left(X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z)\right) \\
& \times\left(X_{i(t-2)}\otimes(Z_{i(t-2)}-z)-X_{i(t-3)}\otimes(Z_{i(t-3)}-z)\right)^{\top}K_{it}K_{i(t-1)}K_{i(t-2)}K_{i(t-3)} \\
= & \mathbf{I}_1-\mathbf{I}_2-\mathbf{I}_3, \qquad \text{(A1.37)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{I}_1 &= \frac{2\sigma_v^2\mu_2(K^2)R(K_v)}{|H|}\mathcal{B}_{XX}(z,z)\otimes H+\frac{2\sigma_v^2\mu_2(K^2)R(K_u)}{|H|}\mathcal{B}_{X_{-1}X_{-1}}(z,z)\otimes H \\
&\quad +o_p(|H|^{-1}H), \\
\mathbf{I}_2 &= \frac{\sigma_v^2\mu_2(K^2)}{|H|^{1/2}}\mathcal{B}_{X_{-1}X_{-1}}(z,z,z)\otimes H+o_p(|H|^{-1/2}H), \\
\mathbf{I}_3 &= o_p(H).
\end{aligned}$$

So now, substituting (A1.12), (A1.30), (A1.33) and (A1.37) into (A1.29) we obtain

$$Var\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right)=\frac{2\sigma_v^2R(K_u)R(K_v)}{NT|H|}\mathcal{B}_{\Delta X\Delta X}^{-1}(z,z)(1+o_p(1)).$$

∎

Proof of Theorem 2.2

Let

$$\begin{aligned}
\widehat{m}(z;H)-m(z) &= (\widehat{m}(z;H)-E[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}])+(E[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}]-m(z)) \\
&\equiv \mathbf{I}_1+\mathbf{I}_2.
\end{aligned}$$

Now, we show that

$$\sqrt{NT\,|H|}\mathbf{I}_1 \quad \xrightarrow{d} \quad \mathcal{N}\left(0,2\sigma_v^2R(K_u)R(K_v)\mathcal{B}_{\Delta X\Delta X}^{-1}(z,z)\right), \qquad \text{(A1.38)}$$

as N tends to infinity.

In order to show this, let

$$\widehat{m}(z;H) - E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \Delta v, \tag{A1.39}$$

where $\Delta v = [\Delta v_{11}, \cdots, \Delta v_{NT}]^\top$. We are going to show the asymptotic normality of

$$\frac{1}{\sqrt{NT}} \widetilde{Z}^\top W \Delta v. \tag{A1.40}$$

Because (A1.40) is a multivariate vector, we define a unit vector $d \in \mathbb{R}^{d(1+q)}$ in such a way that

$$\frac{1}{\sqrt{NT}} d^\top \widetilde{Z}^\top W \Delta v = \frac{1}{\sqrt{NT}} \sum_i \sum_t \lambda_{it}, \tag{A1.41}$$

where

$$\lambda_{it} = |H|^{1/2} d^\top \widetilde{Z}_{it} K_{it} K_{i(t-1)} \Delta v_{it}, \qquad i = 1, \cdots, N \qquad ; \qquad t = 2, \ldots, T$$

and

$$\widetilde{Z}_{it} = \begin{pmatrix} \Delta X_{it} \\ \\ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \end{pmatrix}. \tag{A1.42}$$

By Theorem 2.1 and conditions thereof, we find that

$$\begin{aligned} \operatorname{Var}(\lambda_{it}) \;=\; & 2\sigma_v^2 d^\top \begin{pmatrix} R(K_u)R(K_v)\mathcal{B}_{\Delta X \Delta X}(z,z) & 0 \\ 0 & \mu_2(K^2)R(K_v)\mathcal{B}_{XX}(z,z) \otimes H + \mu_2(K^2) \\ & \times R(K_u)\mathcal{B}_{X_{-1}X_{-1}}(z,z) \otimes H \end{pmatrix} \\ & \times d(1 + o_p(1)) \end{aligned} \tag{A1.43}$$

and

$$\sum_t |\operatorname{Cov}(\lambda_{i1}, \lambda_{it})| = o_p(1).$$

Now we define $\lambda_{n,i}^* = T^{-1/2} \sum_{t=1}^T \lambda_{it}$. For fixed T, the $\left\{\lambda_{n,i}^*\right\}$ are independent random variables. Therefore, to show (A1.37), it suffices to check the Liapunov condition. Using the Minkowski inequality,

$$E|\lambda_{n,i}^*|^{2+\delta} \leq C T^{\frac{2+\delta}{2}} E|\lambda_{it}|^{2+\delta}.$$

Because of (A1.41), we split λ_{it} into two components, λ_{1it} and λ_{2it}, we analyze them separately:

$$
\begin{aligned}
& E|\lambda_{1it}|^{2+\delta} \\
\leq\ & |H|^{\frac{2+\delta}{2}} E|d^\top \Delta X_{it} K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} \\
=\ & |H|^{\frac{2+\delta}{2}} E\left[E\left[|d^\top \Delta X_{it} \Delta v_{it}|^{2+\delta} | Z_{it}, Z_{i(t-1)}\right] K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta}\right] \\
=\ & |H|^{-\delta/2} \int E\left[|d^\top \Delta X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z + H^{1/2}u, Z_{i(t-1)} = z + H^{1/2}v\right] \\
& \times f_{Z_{it}, Z_{i(t-1)}}\left(z + H^{1/2}u, z + H^{1/2}v\right) K^{2+\delta}(u) K^{2+\delta}(v) du dv \\
=\ & |H|^{-\delta/2} E\left[|d^\top \Delta X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \int K^{2+\delta}(u) K^{2+\delta}(v) du dv \\
& + o_p(|H|^{-\delta/2})
\end{aligned}
$$

and

$$
\begin{aligned}
& E|\lambda_{2it}|^{2+\delta} \\
\leq\ & |H|^{\frac{2+\delta}{2}} E|d^\top (X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)) K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} \\
\leq\ & |H|^{\frac{2+\delta}{2}} E\left[|d^\top X_{it} \otimes (Z_{it} - z) K_{it} K_{i(t-1)} \Delta v_{it}|\right]^{2+\delta} \\
& + |H|^{\frac{2+\delta}{2}} E\left[|d^\top X_{i(t-1)} \otimes (Z_{i(t-1)} - z) K_{it} K_{i(t-1)} \Delta v_{it}|\right]^{2+\delta} \\
=\ & |H|^{\frac{2+\delta}{2}} E\left[E\left[|d^\top X_{it} \Delta v_{it}|^{2+\delta} | Z_{it}, Z_{i(t-1)}\right] \otimes |Z_{it} - z|^{2+\delta} K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta}\right] \\
& + |H|^{\frac{2+\delta}{2}} E\left[E\left[|d^\top X_{i(t-1)} \Delta v_{it}|^{2+\delta} | Z_{it}, Z_{i(t-1)}\right] \otimes |Z_{i(t-1)} - z|^{2+\delta} K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta}\right] \\
=\ & |H|\, E\left[|d^\top X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \otimes \int |u|^{2+\delta} K^{2+\delta}(u) K^{2+\delta}(v) du dv \\
& + |H| E\left[|d^\top X_{i(t-1)} \Delta v_{it}|^{2+\delta} | Z_{it}, Z_{i(t-1)}\right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \otimes \int |v|^{2+\delta} K^{2+\delta}(u) K^{2+\delta}(v) du dv \\
& + o_p(|H|).
\end{aligned}
$$

Therefore, $(NT)^{-\frac{2+\delta}{2}} \sum_{i=1}^N E|\lambda_{n,i}^*|^{2+\delta} \leq C(N|H|)^{-\delta/2}$. This, indeed, tends to zero when $N|H| \to \infty$ and therefore the Lyapunov condition holds and (A1.37) follows. We have already defined $\mathcal{D}$ in (A1.43). Finally, using (A1.12) and applying the Cramer-Wold device, the proof is done.

Using the bias expression computed in Theorem 2.1, we can then write

$$
\begin{aligned}
& E\left[\widehat{m}\left(z; H\right) | \mathbb{X}, \mathbb{Z}\right] - m(z) \\
& = \frac{1}{2} \mu_2(K_u) \mathrm{diag}_d \left(\mathrm{tr}(\mathcal{H}_{m_r}(z) H)\right) i_d + O_p(H^{3/2}) + o_p(\mathrm{tr}(H)). \qquad \text{(A1.44)}
\end{aligned}
$$

Note that, by the law of iterated expectations,

$$E\left[\widehat{m}(z;H)\right] = \int E\left[\widehat{m}(z;H)|X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right]dF\left(X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right).$$

The leading term in (A1.44) does not depend on the sample, and then the proof is closed.

∎

Proof of Theorem 2.3

The proof of this result follows the same lines as in the proof of Theorem 2.1. Let

$$\widetilde{m}(z;\widetilde{H}) = e_1^\top \left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\Delta\widetilde{Y}^b. \tag{A1.45}$$

Then proceeding as before in the proof of Theorem 2.1 we obtain

$$E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\left(M^{(1)}+M^{(2)}\right), \tag{A1.46}$$

where

$$\begin{aligned} M^{(1)} &= \left((X_{12}^\top m(Z_{12}))^\top,\cdots,(X_{NT}^\top m(Z_{NT}))^\top\right)^\top, \\ M^{(2)} &= \Bigg(\left(X_{11}^\top(E[\widehat{m}(Z_{11};H)|\mathbb{X},\mathbb{Z}]-m(Z_{11}))\right)^\top, \\ &\qquad \cdots,\left(X_{N(T-1)}^\top(E[\widehat{m}(Z_{N(T-1)};H)|\mathbb{X},\mathbb{Z}]-m(Z_{N(T-1)}))\right)^\top\Bigg)^\top \end{aligned}$$

are $N(T-1)\times 1$ vectors. We can approximate $M^{(1)}$ through a Taylor's expansion, i.e.,

$$M^{(1)} = \widetilde{Z}^b\begin{bmatrix} m(z) \\ vec(D_m(z)) \end{bmatrix} + \frac{1}{2}Q_m^b(z) + R^b(z),$$

where,

$$Q_m^b(z) = \left(S_{m,12}^{b\top}(z),\cdots,S_{m,NT}^{b\top}(z)\right)^\top,$$

and

$$S_{m,it}^b(z) = \left(X_{it}\otimes(Z_{it}-z)\right)^\top\mathcal{H}_m(z)(Z_{it}-z).$$

Using Assumption **2.1** we obtain

$$e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b R^b(z) = o_p(\mathrm{tr}(\widetilde{H})),$$

and therefore,

$$E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right] - m(z) = e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b \left(\frac{1}{2} Q_m^b(z) + M^{(2)}\right) + o_p(\mathrm{tr}(\widetilde{H})). \tag{A1.47}$$

To obtain an asymptotic expression for the bias we first calculate

$$\frac{1}{NT} \widetilde{Z}^{b\top} W^b \widetilde{Z}^b =$$

$$\begin{pmatrix} (NT)^{-1} \sum_{it} X_{it} X_{it}^\top K_{it} & (NT)^{-1} \sum_{it} X_{it} \left(X_{it} \otimes (Z_{it} - z)\right)^\top K_{it} \\ (NT)^{-1} \sum_{it} \left(X_{it} \otimes (Z_{it} - z)\right) X_{it}^\top K_{it} & (NT)^{-1} \sum_{it} \left(X_{it} \otimes (Z_{it} - z)\right) \left(X_{it} \otimes (Z_{it} - z)\right)^\top K_{it} \end{pmatrix}.$$

Using standard properties of kernel density estimators, under Assumptions **2.1** to **2.9** and as N tends to infinity,

$$\begin{aligned} (NT)^{-1} \sum_{it} X_{it} X_{it}^\top K_{it} &= \mathcal{B}_{XX}(z) + o_p(1), \\ (NT)^{-1} \sum_{it} X_{it} \left(X_{it} \otimes (Z_{it} - z)\right)^\top K_{it} &= \mathcal{DB}_{XX}(z) \left(I_d \otimes \mu_2(K_u)\widetilde{H}\right) + o_p(\widetilde{H}), \\ (NT)^{-1} \sum_{it} \left(X_{it} \otimes (Z_{it} - z)\right) \left(X_{it} \otimes (Z_{it} - z)\right)^\top K_{it} &= \mathcal{B}_{XX}(z) \otimes \mu_2(K_u)\widetilde{H} + o_p(\widetilde{H}). \end{aligned}$$

Note that $\mathcal{B}_{XX}(z)$ and $\mathcal{DB}_{XX}(z)$ are defined as in the proof of Theorem 2.1 but the moment functions now are taken conditionally only to $Z_{it} = z$.

Using the previous results,

$$NT \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} = \begin{pmatrix} \mathcal{C}_{11}^{(1)} & \mathcal{C}_{12}^{(1)} \\ \mathcal{C}_{21}^{(1)} & \mathcal{C}_{22}^{(1)} \end{pmatrix}, \tag{A1.48}$$

where

$$\begin{aligned} \mathcal{C}_{11}^{(1)} &= \mathcal{B}_{XX}^{-1}(z) + o_p(1), \\ \mathcal{C}_{12}^{(1)} &= -\mathcal{B}_{XX}^{-1}(z) \left(\mathcal{DB}^{XX}(z)\right) \left(\mathcal{B}_{XX}^{-1}(z) \otimes I_q\right) + o_p(1), \\ \mathcal{C}_{22}^{(1)} &= (\mathcal{B}_{XX}(z) \otimes \mu_2(K_u)\widetilde{H})^{-1} + o_p(\widetilde{H}^{-1}). \end{aligned}$$

Furthermore, the terms in

$$
\begin{aligned}
&(NT)^{-1}\widetilde{Z}^{b\top}W^bQ_m^b(z) \\
&= \begin{pmatrix} (NT)^{-1}\sum_{it} X_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^\top \mathcal{H}_m(z)(Z_{it}-z)K_{it} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^\top \mathcal{H}_m(z)(Z_{it}-z)K_{it} \end{pmatrix}
\end{aligned} \tag{A1.49}
$$

are of order

$$
\mu_2(K_u)E\left[X_{it}X_{it}^\top|Z_{it}=z\right]f_{Z_{it}}(z)\mathrm{diag}_d\left(\mathrm{tr}(\widetilde{\mathcal{H}}_{m_r}(z))\right)i_d+o_p(\mathrm{tr}(\widetilde{H}))
$$

and $O_p(\widetilde{H}^{3/2})$, respectively. In order to evaluate the asymptotic bias of the last term, we have to calculate

$$
\begin{aligned}
&(NT)^{-1}\widetilde{Z}^{b\top}W^bM^{(2)} \\
&= \begin{pmatrix} (NT)^{-1}\sum_{it} X_{it}X_{i(t-1)}^\top\left(E\left[\widehat{m}(Z_{i(t-1)})|\mathbb{X},\mathbb{Z}\right]-m(Z_{i(t-1)})\right)K_{it} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)X_{i(t-1)}^\top\left(E\left[\widehat{m}(Z_{i(t-1)})|\mathbb{X},\mathbb{Z}\right]-m(Z_{i(t-1)})\right)K_{it} \end{pmatrix}.
\end{aligned} \tag{A1.50}
$$

It is straightforward to show that

$$
(NT)^{-1}\sum_{it} X_{it}X_{i(t-1)}^\top\left(E\left[\widehat{m}(Z_{i(t-1)})|\mathbb{X},\mathbb{Z}\right]-m(Z_{i(t-1)})\right)K_{it}=o_p(\mathrm{tr}(\widetilde{H})),
$$

as N tends to infinity, and

$$
\begin{aligned}
(NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)X_{i(t-1)}^\top\left(E\left[\widehat{m}(Z_{i(t-1)})|\mathbb{X},\mathbb{Z}\right]-m(Z_{i(t-1)})\right)K_{it}& \\
=o_p(\mathrm{tr}(H)\mathrm{tr}(\widetilde{H}))&,
\end{aligned}
$$

as N tends to infinity. Under Assumptions **2.1-2.9**, the bias is $o_p(tr(H))$, and the rate is uniform in z; see Masry (1996) for details.

Now we substitute the asymptotic expressions for (A1.48), (A1.49) and (A1.50) into (A1.47) and apply $\mathrm{tr}(H)\to 0$ $\mathrm{tr}(\widetilde{H})\to 0$ in such way that $N|H|\to\infty$, $N|\widetilde{H}|\to\infty$. Thus, we have shown that the asymptotic bias in $\widetilde{m}(z;\widetilde{H})$ is of the same order as it was in the first step.

For the variance term, recall that

$$\begin{aligned}\widetilde{m}^b(z;\widetilde{H}) - E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right] &= e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\Delta v \\ &+ e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\widehat{v},\end{aligned}$$

where $\widehat{v} = (\widehat{v}_1, \cdots, \widehat{v}_N)^\top$ is a $(N(T-1)\times 1)$-vector, such that

$$\widehat{v}_i = \left((X_{i0}^\top r(Z_{i0};H))^\top, \cdots, (X_{i(T-1)}^\top r(Z_{i(T-1)};H))^\top\right)^\top,$$

$i = 1, \cdots, N$, and

$$r(Z_{i(t-1)};H) = \widehat{m}(Z_{i(t-1)};H) - E\left[\widehat{m}(Z_{i(t-1)};H)|\mathbb{X},\mathbb{Z}\right],$$

for $i = 1, \cdots, N$ and $t = 2, \cdots, T$.

Then, the variance of $\widetilde{m}^b(z;\widetilde{H})$ takes the form

$$\begin{aligned}&\text{Var}\left(\widetilde{m}^b(z;H)|\mathbb{X},\mathbb{Z}\right) \\ &= e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\mathcal{V}W^{b\top}\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1 \\ &+e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE\left[\widehat{v}\widehat{v}^\top|\mathbb{X},\mathbb{Z}\right]W^{b\top}\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1 \\ &+2e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE\left[\widehat{v}\Delta v^\top|\mathbb{X},\mathbb{Z}\right]W^{b\top}\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1. \\ &\equiv \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3.\end{aligned}$$

Following exactly the same lines as in the proof of the variance term in Theorem 2.1, as N tends to infinity, we obtain

$$\mathbf{I}_1 = \frac{2\sigma_v^2 R(K_u)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{XX}^{-1}(z)(1+o_p(1)). \tag{A1.51}$$

In order to calculate the asymptotic order of $\mathbf{I}_2$, we just need to calculate

$$\frac{1}{NT}\widetilde{Z}^{b\top}W^bE\left[\widehat{v}\widehat{v}^\top|\mathbb{X},\mathbb{Z}\right]W^{b\top}\widetilde{Z}^b. \tag{A1.52}$$

The upper-left entry is

$$(NT)^{-1}\sum_i\sum_{ts}X_{it}X_{i(t-1)}^\top E\left[r(Z_{i(t-1)};H)r(Z_{i(s-1)};H)^\top|\mathbb{X},\mathbb{Z}\right]X_{i(s-1)}X_{is}^\top K_{it}K_{is}. \tag{A1.53}$$

Applying the Cauchy-Schwarz inequality for covariance matrices, then (A1.53) is bounded by

$$(NT)^{-1}\sum_i\sum_{ts} X_{it}X_{i(t-1)}^\top \mathrm{vec}^{1/2}\left(\mathrm{diag}\left(E\left[r\left(Z_{i(t-1)};H\right)r\left(Z_{i(t-1)};H\right)^\top|\mathbb{X},\mathbb{Z}\right]\right)\right)$$
$$\times\mathrm{vec}^{1/2}\left(\mathrm{diag}\left(E\left[r\left(Z_{i(s-1)};H\right)r\left(Z_{i(s-1)};H\right)^\top|\mathbb{X},\mathbb{Z}\right]\right)\right)^\top X_{i(s-1)}X_{is}^\top K_{it}K_{is}.$$

Now, note that under the conditions of Theorem 2.1

$$\mathrm{vec}\left(\mathrm{diag}\left(E\left[r(z;H)r(z;H)^\top|\mathbb{X},\mathbb{Z}\right]\right)\right)=O_p\left(\frac{\log NT}{NT|H|}\right),$$

uniformly in z, and therefore (A1.53) is of order

$$O_p\left(\frac{\log NT}{NT|H||\widetilde{H}|^{1/2}}\right).$$

Following the same lines, it is easy to show that the upper-right entry of (A1.52) is

$$(NT)^{-1}\sum_i\sum_{ts} X_{it}X_{i(t-1)}^\top E\left[r(Z_{i(t-1)};H)r(Z_{i(s-1)};H)^\top|\mathbb{X},\mathbb{Z}\right]X_{i(s-1)}\left(X_{is}\otimes(Z_{is}-z)\right)^\top$$
$$\times K_{it}K_{is}$$
$$=o_p\left(\frac{\log NT}{NT|H||\widetilde{H}|^{1/2}}\right),$$

and, finally, the lower-right entry of (A1.52) is

$$(NT)^{-1}\sum_i\sum_{ts}\left(X_{it}\otimes(Z_{it}-z)\right)X_{i(t-1)}^\top E\left[r(Z_{i(t-1)};H)r(Z_{i(s-1)};H)^\top|\mathbb{X},\mathbb{Z}\right]$$
$$\times X_{i(s-1)}\left(X_{is}\otimes(Z_{is}-z)\right)^\top K_{it}K_{is}=O_p\left(\frac{\log NT}{NT|H||\widetilde{H}|^{1/2}}\right).$$

Now, combining results in (A1.48) and (A1.52), we show that

$$\mathbf{I}_2=o_p\left(\frac{\log NT}{NT|H||\widetilde{H}|^{1/2}}\right).$$

Finally, a standard Cauchy-Schwarz inequality is enough to show that

$$\mathbf{I}_3=o_p\left(\frac{\log NT}{NT|H||\widetilde{H}|}\right),$$

and then the proof of the result is closed.

∎

Appendix 2

Proof of Theorem 3.1

We first focus on the analysis of the conditional bias of the local weighted linear least-squares estimator, $\widehat{m}(z;H)$, and later on the behavior of its conditional variance-covariance matrix. We follow the standard proofs scheme as in Appendix 1.

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT})$ and $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the observed covariates vectors. By the particularities of the idiosyncratic error term collected in the Assumption **3.2** we know $E(v_{it}|\mathbb{X},\mathbb{Z}) = 0$, so the conditional expectation on (3.14) provides

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W M, \tag{A2.1}$$

where

$$M = \left[\begin{array}{ccc} X_{11}^\top m(Z_{11}) - T^{-1}\sum_{s=1}^T X_{1s}^\top m(Z_{1s}), & \cdots, & X_{NT}^\top m(Z_{NT}) - T^{-1}\sum_{s=1}^T X_{Ns}^\top m(Z_{Ns}) \end{array}\right]^\top.$$

Approximating M using the multivariate Taylor's theorem we obtain

$$M = \widetilde{Z}\left[\begin{array}{c} m(z) \\ vec(D_m(z)) \end{array}\right] + \frac{1}{2}Q_m(z) + R(z), \tag{A2.2}$$

where

$$Q_m(z) = S_m(z) - \overline{S}_m(z), \tag{A2.3}$$

$$\begin{aligned} S_m(z) &= \left[S_{m_{11}}^\top(z), \cdots, S_{m_{NT}}^\top(z)\right]^\top, \\ \overline{S}_m(z) &= \left[\overline{S}_{m_{11}}^\top(z), \cdots, \overline{S}_{m_{NT}}^\top(z)\right]^\top. \end{aligned}$$

The corresponding entries of these vectors are

$$\begin{aligned} S_{m_{it}}(z) &= \left[(X_{it}\otimes(Z_{it}-z))^\top \mathcal{H}_m(z)(Z_{it}-z)\right], \\ \overline{S}_{m_{it}}(z) &= \left[\frac{1}{T}\sum_{s=1}^T (X_{is}\otimes(Z_{is}-z))^\top \mathcal{H}_m(z)(Z_{is}-z)\right], \end{aligned}$$

where we denote by

$$\mathcal{H}_m(z) = \begin{pmatrix} \mathcal{H}_{m1}(z) \\ \mathcal{H}_{m2}(z) \\ \vdots \\ \mathcal{H}_{md}(z) \end{pmatrix}$$

a $dq \times d$ matrix such that $\mathcal{H}_{md}(z)$ is the Hessian matrix of the dth component of $m(\cdot)$.

On the other hand, the remainder term of the Taylor approximation can be written as

$$R(z) = R_m(z) - \overline{R}_m(z), \tag{A2.4}$$

where

$$\begin{aligned} R_m(z) &= \left[R_{m_{11}}^{\top}(z), \cdots, R_{m_{NT}}^{\top}(z)\right]^{\top}, \\ \overline{R}_m(z) &= \left[\overline{R}_{m_{11}}^{\top}(z), \cdots, \overline{R}_{m_{NT}}^{\top}(z)\right]^{\top}, \end{aligned}$$

and the corresponding entry of each vector are

$$\begin{aligned} R_{m_{it}}(z) &= \left[(X_{it} \otimes (Z_{it} - z))^{\top} \mathcal{R}(Z_{it}; z)(Z_{it} - z)\right], \\ \overline{R}_{m_{it}}(z) &= \left[\frac{1}{T}\sum_{s=1}^{T}(X_{is} \otimes (Z_{is} - z))^{\top} \mathcal{R}(Z_{is}; z)(Z_{is} - z)\right]. \end{aligned}$$

We denote by

$$\mathcal{R}(Z_{it}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{it}; z) \\ \mathcal{R}_2(Z_{it}; z) \\ \vdots \\ \mathcal{R}_d(Z_{it}; z) \end{pmatrix}, \mathcal{R}(Z_{is}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{is}; z) \\ \mathcal{R}_2(Z_{is}; z) \\ \vdots \\ \mathcal{R}_d(Z_{is}; z) \end{pmatrix},$$

and

$$\mathcal{R}_d(Z_{it}; z) = \int_0^1 \left[\frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z + \omega(Z_{it} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z)\right](1-\omega)\, d\omega, \tag{A2.5}$$

$$\mathcal{R}_d(Z_{is}; z) = \int_0^1 \left[\frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z + \omega(Z_{is} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z)\right](1-\omega)\, d\omega, \tag{A2.6}$$

where ω is a weight function.

If we replace (A2.2) in (A2.1) we obtain the conditional bias expression consisting in the following two additive terms

$$\begin{aligned} E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]-m(z) &= \frac{1}{2}e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WQ_m(z) \\ &+ e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WR(z), \end{aligned} \tag{A2.7}$$

where we can appreciate that the $vec(D_m(z))$ term of (A2.2) vanishes by the effect of e_1.

As this bias expression has two additive terms, to obtain the conditional bias of this estimator we must analyze both terms of (A2.7) separately. Focus first on the analysis of $e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WQ_m(z)$, we show that this is the leading term of the expression of bias and which actually sets the order of this estimator. Later, we study the behavior of $e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WR(z)$ and explain why this term is asymptotically negligible. For the sake of simplicity let us denote

$$K_{i\ell}=\frac{1}{|H|^{1/2}}K\left(H^{-1/2}\left(Z_{i\ell}-z\right)\right) \qquad \text{for} \quad \ell=1,\cdots,\mathrm{T}.$$

The inverse term of (A2.7) can be rewritten as the following symmetric block matrix

$$(NT)^{-1}\widetilde{Z}^\top W\widetilde{Z}=\begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{21} & \mathcal{A}_{NT}^{22} \end{pmatrix} \tag{A2.8}$$

where,

$$\begin{aligned} \mathcal{A}_{NT}^{11} &= (NT)^{-1}\sum_{it}\ddot{X}_{it}\ddot{X}_{it}^\top\prod_{\ell}K_{i\ell} \\ \mathcal{A}_{NT}^{12} &= (NT)^{-1}\sum_{it}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\prod_{\ell}K_{i\ell}, \\ \mathcal{A}_{NT}^{21} &= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\ddot{X}_{it}^\top\prod_{\ell}K_{i\ell}, \\ \mathcal{A}_{NT}^{22} &= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right) \\ &\quad\times\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\prod_{\ell}K_{i\ell}. \end{aligned}$$

Analyzing each of these terms, we first show that as N tends to infinity

$$\mathcal{A}_{NT}^{11} = \mathcal{B}_{\ddot{X}_t\ddot{X}_t}\left(z,\cdots,z\right) + o_p(1), \tag{A2.9}$$

where

$$\mathcal{B}_{\ddot{X}_t\ddot{X}_t}\left(z,\cdots,z\right) = E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{i1}=z,\cdots,Z_{iT}=z\right] f_{Z_{i1},\cdots,Z_{iT}}\left(z,\cdots,z\right).$$

With the aim of showing this result, under the stationarity assumption and by the law of iterated expectations we obtain

$$\begin{aligned} E\left(\mathcal{A}_{NT}^{11}\right) &= \int E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{i1}=z+H^{1/2}u_1,\cdots,Z_{iT}=z+H^{1/2}u_T\right] \\ &\quad \times f_{Z_{i1},\cdots,Z_{iT}}\left(Z_{i1}=z+H^{1/2}u_1,\cdots,Z_{iT}=z+H^{1/2}u_T\right)\prod_{\ell=1}^{T}K\left(u_\ell\right)du_\ell \end{aligned}$$

and by the Taylor expansion of the unknown functions and Assumptions **3.1** and **3.4** the expression (A2.9) holds. However, note that to conclude this proof is necessary to turn to a law of large numbers. Therefore, we have to show that $Var\left(\mathcal{A}_{NT}^{11}\right) \to 0$, as N tends to infinity. Under the Assumption **3.1**,

$$\begin{aligned} Var\left(\mathcal{A}_{NT}^{11}\right) &= \frac{1}{NT}Var\left(\ddot{X}_{it}\ddot{X}_{it}^{\top}\prod_{\ell=1}^{T}K_{i\ell}\right) \\ &+ \frac{1}{NT^2}\sum_{t=3}^{T}\left(T-t\right)Cov\left(\ddot{X}_{i2}\ddot{X}_{i2}^{\top}\prod_{\ell=2}^{T}K_{i\ell},\ddot{X}_{it}\ddot{X}_{it}^{\top}\prod_{\ell=3}^{T}K_{i\ell}\right). \end{aligned}$$

Then, under Assumptions **3.4** and **3.6** we can show that the first element is

$$Var\left(\ddot{X}_{it}\ddot{X}_{it}^{\top}\prod_{\ell=1}^{T}K_{i\ell}\right) \leq \frac{C}{NT|H|}$$

while the second one is

$$Cov\left(\ddot{X}_{i2}\ddot{X}_{i2}^{\top}\prod_{\ell=2}^{T}K_{i\ell},\ddot{X}_{it}\ddot{X}_{it}^{\top}\prod_{\ell=3}^{T}K_{i\ell}\right) \leq \frac{C'}{N|H|}.$$

Then, if both $NT|H|$ and $N|H|$ tends to infinity the variance term tends to zero and (A2.9) follows.

Following a similar procedure we obtain

$$\mathcal{A}_{NT}^{12} = \mathcal{DB}_{\ddot{X}_tX_t}(z,\cdots,z)\,(I_d\otimes\mu_2(K_{u_\tau})H) - \frac{1}{T}\sum_{s=1}^{T}\mathcal{DB}_{\ddot{X}_tX_s}(z,\cdots,z)\,(I_d\otimes\mu_2(K_{u_s})H)$$
$$+ \quad o_p(H). \tag{A2.10}$$

This is because using the same reasoning,

$$\begin{aligned}
& E\left(\mathcal{A}_{NT}^{12}\right)\\
= \quad & \int E\left(\ddot{X}_{it}X_{it}^\top|Z_{i1}=z+H^{1/2}u_1,\cdots,Z_{iT}=z+H^{1/2}u_T\right) f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z)\otimes(H^{1/2}u_\tau)^\top\\
& \times\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell\\
- \quad & \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\int E\left(\ddot{X}_{it}X_{is}^\top\middle| Z_{i1}=z+H^{1/2}u_1,\cdots,Z_{iT}=z+H^{1/2}u_T\right) f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z)\\
& \otimes(H^{1/2}u_s)^\top\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell
\end{aligned}$$

and as N tends to infinity, $Var\left(\mathcal{A}_{NT}^{12}\right)\to 0$. Then, (A2.6) is shown.

Note that $\mathcal{DB}_{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)$, for $s=1,\cdots,T$, is defined in a similar way as in Appendix 1. Thus, $\mathcal{DB}_{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)$ is a $d\times dq$ gradient matrix of the form

$$\mathcal{DB}_{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)=\begin{pmatrix}\frac{\partial b_{11}^{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)}{\partial Z_1^\top} & \cdots & \frac{\partial b_{1d}^{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)}{\partial Z_1^\top}\\ \vdots & \ddots & \vdots\\ \frac{\partial b_{d1}^{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)}{\partial Z_1^\top} & \cdots & \frac{\partial b_{dd'}^{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)}{\partial Z_1^\top}\end{pmatrix},$$

and

$$b_{dd'}^{\ddot{X}_tX_t}(Z_1,\cdots,Z_T)=E\left[\ddot{X}_{dit}X_{d'it}|Z_{i1}=Z_1,\cdots,Z_{iT}=Z_T\right]f_{Z_{i1},\cdots,Z_{iT}}(Z_1,\cdots,Z_T).$$

Finally, we obtain that as N tends to infinity

$$\mathcal{A}_{NT}^{22}=\mathcal{B}_{X_tX_t}(z,\cdots,z)\otimes\mu_2(K_{u_\tau})H-\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}(z,\cdots,z)\otimes\mu_2(K_{u_s})H+o_p(H), \tag{A2.11}$$

where

$$\mathcal{B}(z,\cdots,z)=\mathcal{B}_{X_tX_s}(z,\cdots,z)+\mathcal{B}_{X_sX_t}(z,\cdots,z)-\frac{1}{T}\mathcal{B}_{X_sX_s}(z,\cdots,z)$$

and

$$\mathcal{B}_{X_tX_s}(z,\cdots,z) \;=\; E\left[X_{it}X_{is}^\top|Z_{i1}=z,\cdots,Z_{iT}=z\right]f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z).$$

Then, using the results of (A2.9)-(A2.11) in (A2.8) we obtain

$$NT\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}=\begin{pmatrix}\mathcal{C}_{11} & \mathcal{C}_{12}\\ \mathcal{C}_{21} & \mathcal{C}_{22}\end{pmatrix}, \tag{A2.12}$$

where

$$\begin{aligned}
\mathcal{C}_{11} &= \mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\cdots,z)+o_p(1),\\
\mathcal{C}_{12} &= -\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\cdots,z)\\
&\quad\times\left(\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\cdots,z)\left(I_d\otimes\mu_2\left(K_{u_\tau}\right)H\right)-\frac{1}{T}\sum_{s=1}^{T}\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\cdots,z)\left(I_d\otimes\mu_2\left(K_{u_s}\right)H\right)\right)\\
&\quad\times\left(\mathcal{B}_{X_tX_t}(z,\cdots,z)\otimes\mu_2\left(K_{u_\tau}\right)H-\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}(z,\cdots,z)\otimes\mu_2\left(K_{u_s}\right)H\right)^{-1}+o_p(1),\\
\mathcal{C}_{21} &= -\left(\mathcal{B}_{X_tX_t}(z,\cdots,z)\otimes\mu_2\left(K_{u_\tau}\right)H-\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}(z,\cdots,z)\otimes\mu_2\left(K_{u_s}\right)H\right)^{-1}\\
&\quad\times\left(\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\cdots,z)\left(I_d\otimes\mu_2\left(K_{u_\tau}\right)H\right)-\frac{1}{T}\sum_{s=1}^{T}\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\cdots,z)\left(I_d\otimes\mu_2\left(K_{u_s}\right)H\right)\right)^\top\\
&\quad\times\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\cdots,z)+o_p(1),\\
\mathcal{C}_{22} &= \left(\mathcal{B}_{X_tX_t}(z,\cdots,z)\otimes\mu_2\left(K_{u_\tau}\right)H-\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}(z,\cdots,z)\otimes\mu_2\left(K_{u_s}\right)H\right)^{-1}+o_p\left(H^{-1}\right).
\end{aligned}$$

On the other hand, following the same technique we can show that

$$(NT)^{-1}\widetilde{Z}^\top WS_m(z)$$

$$=\begin{pmatrix}(NT)^{-1}\sum_{it}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^\top\mathcal{H}_m(z)\left(Z_{it}-z\right)\prod_\ell K_{i\ell}\\ \\ (NT)^{-1}\sum_{it}(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z))\left(X_{it}\otimes(Z_{it}-z)\right)^\top\mathcal{H}_m(z)\\ \times\left(Z_{it}-z\right)\prod_\ell K_{i\ell}\end{pmatrix}$$

are asymptotically equal to

$$\begin{aligned}
&(NT)^{-1}\sum_{it}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^\top\mathcal{H}_m(z)\left(Z_{it}-z\right)\prod_\ell K_{i\ell}\\
&=\mu_2\left(K_{u_\tau}\right)\mathcal{B}_{\ddot{X}_tX_t}(z,\cdots,z)\,diag_d(tr(\mathcal{H}_{m_r}(z)H))\imath_d+o_p\left(tr(H)\right),
\end{aligned} \tag{A2.13}$$

where

$$\mathcal{B}_{\ddot{X}_t X_t}(z,\cdots,z) = E\left[\ddot{X}_{it} X_{it}^\top | Z_{i1} = z, \cdots, Z_{iT} = z\right] f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z),$$

$diag_d\,(tr(\mathcal{H}_{m_r}(z)H))$ stands for a diagonal matrix of elements $tr(\mathcal{H}_{m_r}(z)H)$, for $r = 1,\cdots,d$, and $\imath_d$ is a $d\times 1$ unit vector. In addition,

$$\begin{aligned}
& (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T} X_{is}\otimes(Z_{is}-z)\right)(X_{it}\otimes(Z_{it}-z))^\top \mathcal{H}_m(z) \\
& \times (Z_{it}-z)\prod_{\ell} K_{i\ell} \\
= \quad & \int \mathcal{B}_{X_t X_t}(z,\cdots,z)\otimes(H^{1/2}u_\tau)(H^{1/2}u_\tau)^\top \mathcal{H}_m(z)(H^{1/2}u_\tau)\prod_{\ell=1}^{T} K(u_\ell)\,du_\ell \\
& -\frac{1}{T}\sum_{s=1}^{T}\int \mathcal{B}_{X_s X_t}(z,\cdots,z)\otimes(H^{1/2}u_s)(H^{1/2}u_\tau)^\top \mathcal{H}_m(z)(H^{1/2}u_\tau)\prod_{\ell=1}^{T} K(u_\ell)\,du_\ell \\
= \quad & O_p(H^{3/2}). \qquad \text{(A2.14)}
\end{aligned}$$

Furthermore, the terms of

$$(NT)^{-1}\widetilde{Z}^\top W \overline{S}_m(z)$$

$$= \begin{pmatrix} (NT^2)^{-1}\sum_{its}\ddot{X}_{it}(X_{is}\otimes(Z_{is}-z))^\top \mathcal{H}_m(z)(Z_{is}-z)\prod_{\ell=1}^{T} K_{i\ell} \\ \\ (NT^2)^{-1}\sum_{its}\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T} X_{is}\otimes(Z_{is}-z)\right)(X_{is}\otimes(Z_{is}-z))^\top \mathcal{H}_m(z) \\ \times(Z_{is}-z)\prod_{\ell=1}^{T} K_{i\ell} \end{pmatrix}$$

are of order

$$\begin{aligned}
& \left(NT^2\right)^{-1}\sum_{its}\ddot{X}_{it}(X_{is}\otimes(Z_{is}-z))^\top \mathcal{H}_m(z)(Z_{is}-z)\prod_{\ell=1}^{T} K_{i\ell} \\
= \quad & \frac{1}{T}\sum_{s=1}^{T}\mu_2(K_{u_s})\,\mathcal{B}_{\ddot{X}_t X_s}(z,\cdots,z)\,diag_d(tr(\mathcal{H}_{m_r}(z)H))\imath_d + o_p(tr(H)), \qquad \text{(A2.15)}
\end{aligned}$$

where

$$\mathcal{B}_{\ddot{X}_t X_s}(z,\cdots,z) = E\left[\ddot{X}_{it} X_{is}^\top | Z_{i1} = z, \cdots, Z_{iT} = z\right] f_{Z_{i1},\cdots,Z_{iT}}(z,\cdots,z),$$

and under the stationarity assumption, when $N \to \infty$ and T remains to be fixed we obtain

$$\begin{aligned}
& \left(NT^2\right)^{-1} \sum_{its} \left(X_{it} \otimes (Z_{it} - z) - T^{-1} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z) \right) \left(X_{is} \otimes (Z_{is} - z) \right)^\top \mathcal{H}_m(z) \\
& \times (Z_{is} - z) \prod_{\ell=1}^{T} K_{i\ell} \qquad \text{(A2.16)} \\
= \quad & \int \mathcal{B}_{X_t X_s}(z, \cdots, z) \otimes (H^{1/2} u_\tau)(H^{1/2} u_s)^\top \mathcal{H}_m(z)(H^{1/2} u_s) \prod_{\ell=1}^{T} K(u_\ell) du_\ell \\
- \quad & \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathcal{B}_{X_s X_s}(z, \cdots, z) \otimes (H^{1/2} u_s)(H^{1/2} u_s)^\top \mathcal{H}_m(z)(H^{1/2} u_s) \prod_{\ell=1}^{T} K(u_\ell) du_\ell \\
= \quad & O_p(H^{3/2}). \qquad \text{(A2.17)}
\end{aligned}$$

Then, replacing (A2.13)-(A2.16) into (A2.3), we can conclude

$$Q_m = \begin{pmatrix} \mu_2(K) \left(\mathcal{B}_{\ddot{X}_t X_t}(z, \cdots, z) - \frac{1}{T} \sum_{s=1}^{T} \mathcal{B}_{\ddot{X}_t X_s}(z, \cdots, z) \right) diag_d(tr(\mathcal{H}_{mr}(z) H)) \imath_d \\ + o_p\left(tr(H)\right) \\ \\ O_p\left(H^{3/2}\right) \end{pmatrix}. \qquad \text{(A2.18)}$$

Focus now on the residual term of (A2.7), we use the notation of the beginning of the appendix in order to write

$$(NT)^{-1} \widetilde{Z}^\top W R(z) = \begin{pmatrix} \varepsilon_1(z) \\ \varepsilon_2(z) \end{pmatrix}, \qquad \text{(A2.19)}$$

where

$$\begin{aligned}
\varepsilon_1(z) \quad = \quad & (NT)^{-1} \sum_{it} \ddot{X}_{it} \\
\times \quad & \left[\left(X_{it} \otimes (Z_{it} - z) \right)^\top \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} \left(X_{is} \otimes (Z_{is} - z) \right)^\top \mathcal{R}(Z_{is}; z)(Z_{is} - z) \right] \\
\times \quad & \prod_{\ell} K_{i\ell} \qquad \text{(A2.20)}
\end{aligned}$$

and

$$
\begin{aligned}
\varepsilon_2(z) &= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\\
&\times \left[(X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}(X_{is}\otimes(Z_{is}-z))^{\top}\mathcal{R}(Z_{is};z)(Z_{is}-z)\right]\\
&\times \prod_{\ell}K_{i\ell}. \qquad \text{(A2.21)}
\end{aligned}
$$

Note that $\varepsilon_1(z)$ can be decomposed into the following two terms

$$
\begin{aligned}
\varepsilon_1(z) &= (NT)^{-1}\sum_{it}\ddot{X}_{it}\Big[(X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)\\
&- T^{-1}\sum_{s=1}^{T}(X_{is}\otimes(Z_{is}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{is}-z)\Big]\prod_{\ell}K_{i\ell}\\
&+ \left(NT^2\right)^{-1}\sum_{its}\ddot{X}_{it}(X_{is}\otimes(Z_{is}-z))^{\top}(\mathcal{R}(Z_{it};z)-\mathcal{R}(Z_{is};z))(Z_{is}-z)\prod_{\ell}K_{i\ell}\\
&= \varepsilon_{11}(z)+\varepsilon_{12}(z). \qquad \text{(A2.22)}
\end{aligned}
$$

We want to show that as $N\to\infty$,

$$
E(\varepsilon_1(z)) = o_p(tr(H)) \qquad \text{(A2.23)}
$$

so, in order to do it, we have to analyze each term of $\varepsilon_1(z)$ separately. Starting from $\varepsilon_{11}(z)$ and by the strict stationarity we have

$$
\begin{aligned}
&E(\varepsilon_{11}(z))\\
&= \mathcal{B}_{\ddot{X}_tX_t}(z+H^{1/2}u_1,\cdots,z+H^{1/2}u_T)\otimes(H^{1/2}u_\tau)^{\top}\mathcal{R}(z+H^{1/2}u_\tau;z)(H^{1/2}u_\tau)\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell\\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{B}_{\ddot{X}_tX_s}(z+H^{1/2}u_1,\cdots,z+H^{1/2}u_T)\otimes(H^{1/2}u_s)^{\top}\mathcal{R}(z+H^{1/2}u_\tau;z)(H^{1/2}u_s)\\
&\times\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell.
\end{aligned}
$$

By definition (A2.5) and Assumption **3.7**,

$$\left|\mathcal{R}_d(z+H^{1/2}u_\tau;z)\right| \leq \int_0^1 \varsigma(\omega\|H^{1/2}u_\tau\|)\,(1-\omega)\,d\omega, \qquad \forall d,$$

where $\varsigma\,(\eta)$ is the modulus of continuity of $\frac{\partial^2 m_r}{\partial z_i \partial z_j}(z)$. Hence, by boundedness of f and $\mathcal{B}_{\ddot{X}_t X_t}$, and Assumption **3.4**, for all t we obtain

$$\begin{aligned} E\,|\varepsilon_{11}(z)| \;&\leq\; C_1 \int\int_0^1 |(H^{1/2}u_\tau)^\top||\varsigma(\omega\|H^{1/2}u_\tau\|)||H^{1/2}u_\tau|d\omega \prod_\ell K(u_\ell)du_\ell \\ &+\; \frac{C_2}{T}\sum_s \int\int_0^1 |(H^{1/2}u_s)^\top||\varsigma(\omega\|H^{1/2}u_\tau\|)||H^{1/2}u_s|d\omega \prod_\ell K\,(u_\ell)\,du_\ell \end{aligned}$$

and $E(\varepsilon_{11}(z)) = o_p\,(tr(H))$ follows by dominated convergence.

Similarly, analyzing the second term of (A2.22) and by strict stationarity we have

$$\begin{aligned} E(\varepsilon_{12}(z)) \;&=\; \frac{1}{T}\sum_{s=1}^{T}\int \Big(\mathcal{B}_{\ddot{X}_t X_s}(z+H^{1/2}u_1,\cdots,z+H^{1/2}u_T)\otimes(H^{1/2}u_s)^\top\Big) \\ &\times\; \Big(\mathcal{R}(z+H^{1/2}u_\tau;z)-\mathcal{R}(z+H^{1/2}u_s;z)\Big)\,(H^{1/2}u_s)\prod_{\ell=1}^{T} K\,(u_\ell)\,du_\ell, \end{aligned}$$

where, as previously, we can show

$$\begin{aligned} &|E(\varepsilon_{12}(z))| \\ &\leq \frac{C_3}{T}\sum_s \int\int_0^1 |(H^{1/2}u_s)^\top||\varsigma(\omega\|H^{1/2}u_\tau\|-\omega\|H^{1/2}u_s\|)||H^{1/2}u_s|\prod_\ell K\,(u_\ell)\,du_\ell. \end{aligned}$$

Then, proceeding as previously we have that by dominated convergence $E(\varepsilon_{12}(z)) = o_p\,(tr(H))$.

Once this result (A2.23) has been verified, our interest focuses on the second term of (A2.22), $\varepsilon_2(z)$, with the aim of showing that as $N\to\infty$,

$$E(\varepsilon_2(z)) = O_p(H^{3/2}). \tag{A2.24}$$

In order to prove this result, we follow the same lines as the proof of (A2.23) and $\varepsilon_2(z)$ can be decomposed in two terms

$$\varepsilon_2(z) \;=\; \varepsilon_{21}(z)+\varepsilon_{22}(z), \tag{A2.25}$$

where

$$
\begin{aligned}
&\varepsilon_{21}(z)\\
&= (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\\
&\times\left[(X_{it}\otimes(Z_{it}-z))^{\top}\mathcal{R}(Z_{it};z)(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}(X_{is}\otimes(Z_{is}-z))^{\top}\mathcal{R}(Z_{is};z)(Z_{is}-z)\right]\\
&\times\prod_{\ell}K_{i\ell} \qquad (A2.26)
\end{aligned}
$$

and

$$
\begin{aligned}
\varepsilon_{22}(z) &= \left(NT^{2}\right)^{-1}\sum_{its}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)(X_{is}\otimes(Z_{is}-z))^{\top}\\
&\quad\times(\mathcal{R}(Z_{it};z)-\mathcal{R}(Z_{is};z))(Z_{is}-z)\prod_{\ell}K_{i\ell}. \qquad (A2.27)
\end{aligned}
$$

Applying the same arguments as for the proof of (A2.23), it is straightforward to show that

$$E(\varepsilon_2(z)) = o_p(H^{3/2}). \qquad (A2.28)$$

Then, replacing (A2.23) and (A2.24) in (A2.19) we obtain

$$(NT)^{-1}\widetilde{Z}^{\top}WR(z) = \begin{pmatrix} o_p(tr(H)) \\ O_p(H^{3/2}) \end{pmatrix} \qquad (A2.29)$$

and substituting (A2.12), (A2.18) and (A2.29) in (A2.7), the asymptotic bias can be written as

$$
\begin{aligned}
& E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]-m(z)\\
=\ & \frac{1}{2}e_1^{T}\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}\widetilde{Z}^{\top}W(S_m(z)-\overline{S}_m(z))\\
=\ & \frac{1}{2}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\cdots,z)\left(\mu_2(K_{u_\tau})\mathcal{B}_{\ddot{X}_tX_t}(z,\cdots,z)-\frac{1}{T}\sum_{s=1}^{T}\mu_2(K_{u_s})\mathcal{B}_{\ddot{X}_tX_s}(z,\cdots,z)\right)\\
\times\ & diag_d\left(tr(\mathcal{H}_{m_r}(z)H)\right)\imath_d+o_p(tr(H)).
\end{aligned}
$$

For the asymptotic expression of the variance term let us define the NT vector $v = (v_1,\cdots,v_N)^{\top}$, where $v_i = (v_{i1},\cdots,v_{iT})^{\top}$. Furthermore, let $E\left(vv^{\top}|\mathbb{X},\mathbb{Z}\right) = \mathcal{V}$ be a $NT\times NT$ matrix that contains the V_{ij}'s matrices. By Assumption **3.2** we obtain

$$V_{ij} = E(v_iv_j^{\top}|\mathbb{X},\mathbb{Z}) = \sigma_v^2 I_T. \qquad (A2.30)$$

Denote by $Q_T = I_T - \imath_T \left(\imath_T^\top \imath_T\right)^{-1} \imath_T^\top$ a $T \times T$ symmetric and idempotent matrix with rank $T-1$, where I_T is a $T \times T$ identity matrix and $\imath_T$ a $T \times 1$ unitary vector. Furthermore, let $Q = I_N \otimes Q_T$ be an $NT \times NT$ matrix. It is clear that, $\widetilde{Z} = Q\widetilde{Z}^b$ and $\ddot{v} = Qv$.

Then, substituting the previous equalities into

$$\widehat{m}(z;H) - E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \ddot{v}, \tag{A2.31}$$

we obtain

$$\widehat{m}(z;H) - E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} Z_0^\top Q^\top W Q v. \tag{A2.32}$$

Since Q is an idempotent matrix, the variance term of $\widehat{m}(z;H)$ can be written as

$$Var\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \mathcal{V} W \widetilde{Z} \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} e_1. \tag{A2.33}$$

As by Assumption **3.2** the v_{it}'s are *i.i.d.* in the subscript i, the upper left entry of $(NT)^{-1}\widetilde{Z}^\top W\mathcal{V}W\widetilde{Z}$ is

$$\begin{aligned} &\frac{\sigma_v^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\ddot{X}_{it}^\top\prod_{\ell=1}^{T}K_{i\ell}^2 \\ &= \frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{|H|^{T/2}}\mathcal{B}_{\ddot{X}_t\ddot{X}_t}\left(z,\cdots,z\right)\left(1+o_p(1)\right). \end{aligned} \tag{A2.34}$$

The upper right block is

$$\begin{aligned} & \frac{\sigma_v^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\prod_{\ell=1}^{T}K_{i\ell}^2 \\ = \;& \frac{\sigma_v^2}{|H|^{T/2}}\int\Bigg(\mathcal{B}_{\ddot{X}_tX_t}(z+H^{1/2}u_1,\cdots,z+H^{1/2}u_T)\otimes(H^{1/2}u_\tau)^\top \\ & -\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}_{\ddot{X}_tX_s}(z+H^{1/2}u_1,\cdots,z+H^{1/2}u_T)\otimes(H^{1/2}u_s)^\top\Bigg)\prod_{\ell=1}^{T}K^2\left(u_\ell\right)du_\ell\left(1+o_p(1)\right) \\ = \;& O_p(|H|^{-T/2}). \end{aligned} \tag{A2.35}$$

Finally, the lower-right block is

$$\begin{aligned}
&\frac{\sigma_v^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\\
&\qquad\times\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^{\top}\prod_{\ell=1}^{T}K_{i\ell}^2\\
= \;&\frac{\sigma_v^2\mu_2\left(K_{u_\tau}^2\right)}{|H|^{T/2}}\mathcal{B}_{X_tX_t}(z,\cdots,z)\otimes H-\frac{\sigma_v^2}{T\,|H|^{T/2}}\sum_{s=1}^{T}\mu_2\left(K_{u_s}^2\right)\mathcal{B}(z,\cdots,z)\otimes H\\
+ \;&O_p(|H|^{-T/2}H). \qquad \text{(A2.36)}
\end{aligned}$$

Then, substituting (A2.12), (A2.34), (A2.35) and (A2.36) into (A2.33) we obtain the following conditional covariance matrix result,

$$Var\left(\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right)=\frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{NT|H|^{T/2}}\mathcal{B}_{\ddot{X}_t\ddot{X}_t}^{-1}(z,\cdots,z)\left(1+o_p(1)\right).$$

∎

Proof of Theorem 3.2

With the aim of obtaining the asymptotic distribution of the local weighted linear least squares estimator $\widehat{m}(z;H)$ we follow a similar proof scheme as in Appendix 1. For this, let us denote

$$\widehat{m}(z;H)-m(z)=\left(\widehat{m}(z;H)-E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]\right)+\left(E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]-m(z)\right)\equiv\mathbf{I}_1+\mathbf{I}_2,$$

so in order to obtain the asymptotic distribution of this estimator we must to show that as $N\to\infty$ it holds

$$\sqrt{NT|H|^{T/2}}\mathbf{I}_1 \quad\xrightarrow{d}\quad N\left(0,\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)\mathcal{B}_{\ddot{X}_t\ddot{X}_t}^{-1}(z,\cdots,z)\right) \qquad \text{(A2.37)}$$

and

$$E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2}\mu_2\left(K_{u_\tau}\right)diag_d\left(tr(\mathcal{H}_{m_r}(z)H)\right)\imath_d+O_p(H^{3/2})+o_p(tr(H)). \qquad \text{(A2.38)}$$

By Assumption **3.1** we state that the variables are *i.i.d.* in the subscript i but not in T, so the Lindeberg condition cannot be verified directly. Thus, in order to show (A2.37) it suffices to check the Lyapunov condition. We have shown that

$$\widehat{m}(z;H) - E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W v. \tag{A2.39}$$

The behavior of the inverse term has been analyzed previously, with the aim of proving the result (A2.39) we must to focus on the asymptotic normality of

$$\frac{1}{\sqrt{NT}} \widetilde{Z}^\top W v. \tag{A2.40}$$

As (A2.40) is a multivariate vector, with the sake of simplicity we can define a unit vector $d \in I\!R^{d(1+q)}$ in such a way that

$$\frac{1}{\sqrt{NT}} d^T \widetilde{Z}^\top W v = \frac{1}{\sqrt{NT}} \sum_i \sum_t \lambda_{it}, \tag{A2.41}$$

where

$$\lambda_{it} = |H|^{T/4} d^\top \widetilde{Z}_{it} W_{it} v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T.$$

Following Assumption **3.8**, we have that $R(K) = \int K^2(u)\,du = \left(2\Pi^{1/2}\right)^{-1}$ and $R(K_{u_1}) = \cdots = R(K_{u_T})$, so $\prod_{\ell=1}^T R(K_{u_\ell}) = R(K)^T$ holds. Combining these conditions with the results of Theorem 3.1 we can write

$$Var(\lambda_{it}) =$$

$$\sigma_v^2 d^\top \begin{pmatrix} R(K)^T \mathcal{B}_{\ddot{X}_t \ddot{X}_t}(z, \cdots, z) & 0 \\ 0 & \begin{array}{c} \mu_2\left(K_{u_\tau}^2\right) \mathcal{B}_{X_t X_t}(z, \cdots, z) \otimes H \\ -\frac{1}{T}\sum_{s=1}^T \mu_2\left(K_{u_s}^2\right) \mathcal{B}(z, \cdots, z) \otimes H \end{array} \end{pmatrix} d\,(1 + o_p(1)), \tag{A2.42}$$

whereas

$$\sum_{t=1}^T |Cov(\lambda_{i1}, \lambda_{it})| = o_p(1). \tag{A2.43}$$

In order to check the Lyapunov condition let us denote $\lambda_{n,i}^{*} = T^{-1/2}\sum_{t=1}^{T}\lambda_{it}$ as independent random variables for T fixed and $n = NT$. Then, by the Minknowski inequality and the matrix structure of $\widetilde{Z}_{it}$ we obtain

$$E\left|\lambda_{n,i}^{*}\right|^{2+\delta} \leq CT^{\frac{(2+\delta)}{2}} E\left|\lambda_{it}\right|^{2+\delta} = CT^{\frac{(2+\delta)}{2}} E\left|\lambda_{1it} + \lambda_{2it}\right|^{2+\delta}.$$

Analyzing each term separately we obtain

$$\begin{aligned}
& E\left|\lambda_{1it}\right|^{2+\delta} \\
\leq\ & E|\,|H|^{T/4}\, d^{\top}\ddot{X}_{it}v_{it}\prod_{\ell=1}^{T}K_{i\ell}|^{2+\delta} = |H|^{T(2+\delta)/4}\, E\left[E\left(|d^{\top}\ddot{X}_{it}v_{it}|^{2+\delta}|\mathbb{X},\mathbb{Z}\right)\prod_{\ell=1}^{T}K_{i\ell}^{2+\delta}\right] \\
=\ & \frac{1}{|H|^{T\delta/4}}\int E\left(|d^{\top}\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1} = z + H^{1/2}u_{1}, \cdots, Z_{iT} = z + H^{1/2}u_{T}\right) \\
& \times f_{Z_{i1},\cdots,Z_{iT}}(z + H^{1/2}u_{1}, \cdots, z + H^{1/2}u_{T})\prod_{\ell=1}^{T}K^{2+\delta}\left(u_{\ell}\right)du_{\ell} \\
=\ & |H|^{-T\delta/4}\, E\left(|d^{\top}\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1} = z, \cdots, Z_{iT} = z\right) f_{Z_{i1},\cdots,Z_{iT}}\left(z, \cdots, z\right)\prod_{\ell=1}^{T}\int K^{2+\delta}\left(u_{\ell}\right)du_{\ell} \\
+\ & o_{p}(|H|^{-T\delta/4}),
\end{aligned}$$

whereas

$$\begin{aligned}
& E|\lambda_{2it}|^{2+\delta} \\
\leq\ & E\left|\,|H|^{T/4}\, d^{\top}\left(X_{it}\otimes(Z_{it} - z) - T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is} - z)\right)v_{it}\prod_{\ell=1}^{T}K_{i\ell}\right|^{2+\delta} \\
\leq\ & |H|^{T(2+\delta)/4}E\left[E\left(|d^{\top}\ddot{X}_{it}v_{it}|Z_{i1}, \cdots, Z_{iT}\right)\otimes|Z_{it} - z|^{2+\delta}\prod_{\ell=1}^{T}K_{i\ell}^{2+\delta}\right] \\
+\ & |H|^{T(2+\delta)/4}\frac{1}{T}\sum_{s=1}^{T}E\left[E\left(|d^{\top}X_{is}v_{it}|Z_{i1}, \cdots, Z_{iT}\right)\otimes|Z_{is} - z|^{2+\delta}\prod_{\ell=1}^{T}K_{i\ell}^{2+\delta}\right] \\
=\ & |H|^{-T\delta/4}E\left(|d^{\top}\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1} = z, \cdots, Z_{iT} = z\right) f_{Z_{i1},\cdots,Z_{iT}}\left(z, \cdots, z\right) \\
& \otimes\int|H^{1/2}u|^{2+\delta}\prod_{\ell=1}^{T}K^{2+\delta}\left(u_{\ell}\right)du_{\ell} \\
+\ & |H|^{-T\delta/4}\frac{1}{T}\sum_{s=1}^{T}E\left(|d^{\top}\ddot{X}_{is}v_{it}|^{2+\delta}|Z_{i1} = z, \cdots, Z_{iT} = z\right) f_{Z_{i1},\cdots,Z_{iT}}\left(z, \cdots, z\right) \\
& \otimes\int|H^{1/2}u_{s}|^{2+\delta}\prod_{\ell=1}^{T}K^{2+\delta}\left(u_{\ell}\right)du_{\ell} + o_{p}(|H|^{1-(T-2)\delta/4}).
\end{aligned}$$

In this way, we can write

$$(NT)^{-\frac{2+\delta}{2}} \sum_{i=1}^{N} E\left|\lambda_{n,i}^{*}\right|^{2+\delta} \leq C(N|H|^{T/2})^{-\delta/2}, \tag{A2.44}$$

and given that when $N|H| \to \infty$ this term tends to zero it is proved that the Lyapunov condition holds. Then, using (A2.12), (A2.34), (A2.35), (A2.36) and the Crammer-Wold device the proof of the result (A2.37) is done.

On the other hand, focus on the proof of (A2.38) we know that by the law of iterated expectations

$$E\left[\widehat{m}\left(z;H\right)\right] = \int E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right] dF\left(\mathbb{X}\right).$$

Then, we can turn to the bias expression of the estimator collected in the Theorem 3.1 and the proof is closed.

∎

Proof of Theorem 3.3

The proof of this theorem follows the pattern set by the Theorem 3.1. The estimator to analyze is

$$\widetilde{m}(z;\widetilde{H}) = e_1^{\top}\left(\widetilde{Z}^{b\top}W^{b}\widetilde{Z}^{b}\right)^{-1}\widetilde{Z}^{b\top}W^{b}\ddot{Y}^{b}, \tag{A2.45}$$

we can write

$$E\left[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right] = e_1^{\top}\left(\widetilde{Z}^{b\top}W^{b}\widetilde{Z}^{b}\right)^{-1}\widetilde{Z}^{b\top}W^{b}\left[M^{(1)}+M^{(2)}\right], \tag{A2.46}$$

where

$$\begin{aligned}
M^{(1)} &= \left[\left(X_{11}^{\top}m\left(Z_{11}\right)\right)^{\top},\cdots,\left(X_{NT}^{\top}m\left(Z_{NT}\right)\right)^{\top}\right]^{\top}, \\
M^{(2)} &= \left[\left(T^{-1}\sum_{s=1}^{T}X_{1s}^{\top}\left(E\left[\widehat{m}\left(Z_{1s};H\right)|\mathbb{X},\mathbb{Z}\right]-m\left(Z_{1s}\right)\right)\right)^{\top},\right. \\
&\qquad \left.\cdots,\left(T^{-1}\sum_{s=1}^{T}X_{Ns}^{\top}\left(E\left[\widehat{m}\left(Z_{Ns};H\right)|\mathbb{X},\mathbb{Z}\right]-m\left(Z_{Ns}\right)\right)\right)^{\top}\right]^{\top}.
\end{aligned}$$

Taylor Theorem implies that we can approximate $M^{(1)}$ as

$$M^{(1)} = \widetilde{Z}^b \begin{bmatrix} m(z) \\ vec(D_m(z)) \end{bmatrix} + \frac{1}{2} Q_m^b(z) + R^b(z). \tag{A2.47}$$

Following a similar nomenclature as in Theorem 3.1,

$$\begin{aligned} Q_m^b(z) &= \left[S_{m_{11}}^{b\top}, \cdots, S_{m_{NT}}^{b\top} \right]^\top, \\ R^b(z) &= \left[R_{m_{11}}^{b\top}(z), \cdots, R_{m_{NT}}^{b\top}(z) \right]^\top, \end{aligned}$$

where $R^b(z)$ is the remainder term of this approximation. Then, the corresponding entries of these vectors are

$$\begin{aligned} S_{m_{it}}^b &= \left[\left(X_{it} \otimes (Z_{it} - z)\right)^\top \mathcal{H}_m(z) \left(Z_{it} - z\right) \right] \\ R_{it}^b(z) &= \left[\left(X_{it} \otimes (Z_{it} - z)\right)^\top \mathcal{R}\left(Z_{it}; z\right) \left(Z_{it} - z\right) \right], \end{aligned}$$

where $\mathcal{R}\left(Z_{it}; z\right)$ has already been defined in (A2.5).

If we replace (A2.47) in (A2.46) the bias expression is then

$$\begin{aligned} & E\left[\widetilde{m}(z; \widetilde{H}) | \mathbb{X}, \mathbb{Z}\right] - m(z) \\ & = \frac{1}{2} e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b \right)^{-1} \widetilde{Z}^{b\top} W^b Q_m^b(z) + e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b \right)^{-1} \widetilde{Z}^{b\top} W^b M^{(2)} \\ & + o_p(tr(\widetilde{H})), \end{aligned} \tag{A2.48}$$

given that following to Ruppert and Wand (1994) and the Assumption **3.1**,

$$e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z} \right)^{-1} \widetilde{Z}^{b\top} W^b R^b(z) = O_p(tr(\widetilde{H})).$$

As you can see in (A2.48), this bias expression is formed by two additive terms. The first one refers to the approximation error of the estimates, whereas the second one reflects the potential estimation error dragged from the first stage. Within this context, our aim is to show that this second term converges in probability to zero, so it is the first element which provides the asymptotic distribution of the backfitting estimator. For the sake of simplicity let us denote

$$K_{it} = \frac{1}{|\widetilde{H}|^{1/2}} K\left(\widetilde{H}^{-1/2} \left(Z_{it} - z\right) \right).$$

Focus first in the behavior of the inverse term of (A2.48) we analyze

$$(NT)^{-1}\, \widetilde{Z}^{b\top} W^b \widetilde{Z}^b =$$

$$\begin{pmatrix} (NT)^{-1}\sum_{it} X_{it}X_{it}^{\top}K_{it} & (NT)^{-1}\sum_{it} X_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}K_{it} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)X_{it}^{\top}K_{it} & (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}K_{it} \end{pmatrix}$$

and as it is proved in the Appendix 1, using standard properties of kernel density estimators, conditions **3.1** to **3.3** and **3.4** to **3.10**, as $N \to \infty$ we obtain

$$NT\left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} = \tag{A2.49}$$

$$\begin{pmatrix} \mathcal{B}_{X_tX_t}^{-1}(z) + o_p(1) & -\mathcal{B}_{X_tX_t}^{-1}(z)\left[\mathcal{D}\mathcal{B}_{X_tX_t}(z)\right]\left(\mathcal{B}_{X_tX_t}^{-1}(z)\otimes I_q\right) + o_p(1) \\ \\ -\left(\mathcal{B}_{X_tX_t}^{-1}(z)\otimes I_q\right)^{\top}\left[\mathcal{D}\mathcal{B}_{X_tX_t}(z)\right]^{\top}\mathcal{B}_{X_tX_t}^{-1}(z) + o_p(1) & \left(\mathcal{B}_{X_tX_t}(z)\otimes\mu_2(K)\widetilde{H}\right)^{-1} + o_p\left(\widetilde{H}^{-1}\right) \end{pmatrix},$$

where $\mathcal{B}_{X_tX_t}(z)$ and $\mathcal{D}\mathcal{B}_{X_tX_t}(z)$ has been already defined in the proof of Theorem 3.1 conditioning only to $Z_{it} = z$.

Furthermore,

$$(NT)^{-1}\widetilde{Z}^{b\top} W^b Q_m^b(z) = \tag{A2.50}$$

$$\begin{pmatrix} (NT)^{-1}\sum_{it} X_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)K_{it} \\ \\ (NT)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)K_{it} \end{pmatrix}$$

are of order

$$\mu_2\left(K_u\right)\mathcal{B}_{X_tX_t}(z)\, diag_d\left(tr(\mathcal{H}_{m_r}(z)\widetilde{H})\right)\imath_d + o_p(tr(\widetilde{H}))$$

and $O_p(\widetilde{H}^{3/2})$, respectively. Substituting these latter results and (A2.49) in the first term of (A2.48) we obtain

$$\begin{aligned} & \frac{1}{2}e_1^{\top}\left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top} W^b Q_m^b(z) \\ = \quad & \frac{1}{2}\mu_2\left(K\right)\mathcal{B}_{X_tX_t}^{-1}(z)\,\mathcal{B}_{X_tX_t}(z)\, diag_d\left(tr(\mathcal{H}_{m_r}(z)\widetilde{H})\right)\imath_d + o_p(tr(\widetilde{H})). \end{aligned} \tag{A2.51}$$

Focus now on the behavior of the second term of (A2.48),

$$(NT)^{-1}\widetilde{Z}^{b\top}W^{b}M^{(2)} = \tag{A2.52}$$

$$\begin{pmatrix} \left(NT^2\right)^{-1}\sum_{its} X_{it}X_{is}^{\top}\left(E\left[\widehat{m}\left(Z_{is}\right)\middle|\,\mathbb{X},\mathbb{Z}\right]-m\left(Z_{is}\right)\right)K_{it} \\ \\ \left(NT^2\right)^{-1}\sum_{its}\left(X_{it}\otimes\left(Z_{it}-z\right)\right)X_{is}^{\top}\left(E\left[\widehat{m}\left(Z_{is}\right)\middle|\,\mathbb{X},\mathbb{Z}\right]-m\left(Z_{is}\right)\right)K_{it} \end{pmatrix}$$

and analyzing both terms separately we can show that as N tends to infinity

$$\left(NT^2\right)^{-1}\sum_{its} X_{it}X_{is}^{\top}\left(E\left[\widehat{m}\left(Z_{is}\right)\middle|\,\mathbb{X},\mathbb{Z}\right]-m\left(Z_{is}\right)\right)K_{it} = o_p(tr(\widetilde{H}))$$

and

$$\left(NT^2\right)^{-1}\sum_{its}\left(X_{it}\otimes\left(Z_{it}-z\right)\right)X_{is}^{\top}\left(E\left(\widehat{m}\left(Z_{is}\right)\middle|\,\mathbb{X},\mathbb{Z}\right)-m\left(Z_{is}\right)\right)K_{it} = o_p(tr(H)tr(\widetilde{H})).$$

Under Assumptions **3.1** to **3.3**, **3.10** and **3.12**, this latter expression is $o_p\left(tr(H)\right)$ and the rate is uniform in z; see Masry (1996) for more details.

Replacing these results in the second term of (A2.48),

$$e_1^{\top}\left(\widetilde{Z}^{b\top}W^{b}\widetilde{Z}^{b}\right)^{-1}\widetilde{Z}^{b\top}W^{b}M^{(2)} = o_p(tr(\widetilde{H})). \tag{A2.53}$$

Finally, substituting (A2.51) and (A2.53) in (A2.48) the proof of the conditional bias is done. Also, it is proved that the asymptotic bias of $\widetilde{m}(z;\widetilde{H})$ is the same order as $\widehat{m}(z;H)$, given that $tr(H)\to 0$, $tr(\widetilde{H})\to 0$ in such a way that $N|H|\to\infty$ and $N|\widetilde{H}|\to\infty$.

From the standpoint of the variance, let us denote $\widehat{v}=\left(\widehat{v}_1,\cdots,\widehat{v}_N\right)^{\top}$ as a NT-dimensional vector such that

$$\widehat{v}_i = \left(T^{-1}\sum_{s=1}^{T}\left(X_{is}^{\top}r(Z_{is};H)\right)^{\top},\cdots,T^{-1}\sum_{s=1}^{T}\left(X_{is}^{\top}r(Z_{is};H)\right)^{\top}\right)^{\top},$$

where

$$r\left(Z_{is};H\right)=\widehat{m}\left(Z_{is};H\right)-E\left[\widehat{m}\left(Z_{is};H\right)\middle|\,\mathbb{X},\mathbb{Z}\right].$$

As we know, the conditional variance-covariance matrix of the estimator has the following form

$$Var\left(\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right) = E\left[\left(\widetilde{m}(z;\widetilde{H}) - E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]\right)\left(\widetilde{m}(z;\widetilde{H}) - E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]\right)^\top |\mathbb{X},\mathbb{Z}\right]$$

where

$$\widetilde{m}(z;\widetilde{H}) - E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}] = e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\ddot{v} + e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\widehat{v}.$$

Remember that $\ddot{v}_i = Q_T v_i$ and it is straightforward to show that $Q_T\widetilde{Z}_i^b = \widetilde{Z}_i$. Thus, the previous equation can be rewritten as

$$\widetilde{m}(z;\widetilde{H}) - E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}] = e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{\top}W^b v + e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\widehat{v}.$$

Taking into account that let $E\left(vv^\top|\mathbb{X},\mathbb{Z}\right) = \mathcal{V}$ be a $NT \times NT$ matrix whose ijth have the form of (A2.30), the variance term of $\widetilde{m}(z;\widetilde{H})$ has the form

$$\begin{aligned}
Var\left(\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}\right) &= e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{\top}W^b\mathcal{V}W^b\widetilde{Z}\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1 \\
&+ e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b E\left(\widehat{v}\widehat{v}^\top|\mathbb{X},\mathbb{Z}\right)W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1 \\
&+ 2e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b E\left(\widehat{v}\ddot{v}^\top|\mathbb{X},\mathbb{Z}\right)W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1 \\
&= \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3. \qquad \text{(A2.54)}
\end{aligned}$$

Then, with the aim of obtaining the asymptotic order of the variance of $\widetilde{m}(z;\widetilde{H})$ we have to analyze each of these terms separately. Following the same procedure as in (A2.33) to analyze the behavior of $\widetilde{Z}^\top W^b\mathcal{V}W^b\widetilde{Z}$. Under Assumptions **3.1** to **3.3** and **3.4** to **3.10**, using the result (A2.49) and the Crammer-Wold device it is straightforward to show that as $N\to\infty$

$$\mathbf{I}_1 = \frac{\sigma_v^2 R(K)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{X_tX_t}(z)^{-1}\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)\,\mathcal{B}_{X_tX_t}(z)^{-1}(1+o_p(1)), \qquad \text{(A2.55)}$$

while

$$\mathbf{I}_2 = o_p\left(\frac{logNT}{NT\,|H|^{T/2}\,|\widetilde{H}|^{1/2}}\right). \qquad \text{(A2.56)}$$

In order to prove this latter result we have to analyze the behavior of the following expression

$$(NT)^{-1}\widetilde{Z}^{b\top}W^{b}E\left(\widehat{v}\widehat{v}^{\top}|\mathbb{X},\mathbb{Z}\right)W^{b}\widetilde{Z}^{b}. \tag{A2.57}$$

Thus, the upper left entry is

$$(NT^{3})^{-1}\sum_{i}\sum_{tt'}\sum_{ss'}X_{it}X_{is}^{\top}E\left(r\left(Z_{is};H\right)r\left(Z_{is'};H\right)^{\top}|\mathbb{X},\mathbb{Z}\right)X_{is'}X_{it'}^{\top}K_{it}K_{it'} \tag{A2.58}$$

and by the Cauchy-Schwarz inequality for variance-covariance matrices (A2.58) is bounded by

$$\begin{aligned}
&(NT^{3})^{-1}\sum_{i}\sum_{tt'}\sum_{ss'}X_{it}X_{is}^{\top}vec^{1/2}\left(diag\left(E(r\left(Z_{is};H\right)r\left(Z_{is};H\right)^{\top}|\mathbb{X},\mathbb{Z})\right)\right)\\
\times\quad & vec^{1/2}\left(diag\left(E(r\left(Z_{is'};H\right)r\left(Z_{is'};H\right)^{\top}|\mathbb{X},\mathbb{Z})\right)\right)X_{is'}X_{it'}^{\top}K_{it}K_{it'}\\
=\quad & O_{p}\left(\frac{logNT}{NT|H|^{T/2}}\right),
\end{aligned} \tag{A2.59}$$

given that under the conditions of Theorem 3.1 and following Masry (1996),

$$vec\left(diag\left(E(r\left(z;H\right)r\left(z;H\right)^{\top}|\mathbb{X},\mathbb{Z})\right)\right)=O_{p}\left(\frac{logNT}{NT|H|^{T/2}}\right),$$

uniformly in z.

Following the same lines, the upper right entry of (A2.57) is

$$\begin{aligned}
&(NT^{2})^{-1}\sum_{i}\sum_{tt'}\sum_{ss'}X_{it}X_{is}^{\top}E\left(r\left(Z_{is};H\right)r\left(Z_{is'};H\right)^{\top}|\mathbb{X},\mathbb{Z}\right)\left(X_{it'}\otimes\left(Z_{it'}-z\right)\right)^{\top}K_{it}K_{it'}\\
&=o_{p}\left(\frac{logNT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}\right)
\end{aligned} \tag{A2.60}$$

and the lower right entry of (A2.57) is

$$\begin{aligned}
&(NT^{2})^{-1}\sum_{i}\sum_{tt'}\sum_{ss'}\left(X_{it}\otimes\left(Z_{it}-z\right)\right)X_{is}^{\top}E\left(r\left(Z_{is};H\right)r\left(Z_{is'};H\right)^{\top}|\mathbb{X},\mathbb{Z}\right)\\
&\times\left(X_{it'}\otimes\left(Z_{it'}-z\right)\right)^{\top}K_{it}K_{it'}\\
=\quad & o_{p}\left(\frac{logNT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}\right).
\end{aligned} \tag{A2.61}$$

Then, combining the results (A2.59)-(A2.61) with (A2.49) and by the Crammer-Wold device the proof of (A2.56) is done. Finally, focus on $\mathbf{I}_3$ the Cauchy-Schwarz inequality is enough to show that

$$\mathbf{I}_3 = o_p\left(\sqrt{\frac{logNT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}}\right) \tag{A2.62}$$

and the proof is done.

∎

Appendix 3

Proof of Theorem 5.1

In order to obtain the desired results of Theorem 5.1 we denote by

$$\widehat{m}_g(z_1;H_2) \;=\; \left(\sum_{i=1}^{T}\sum_{t=2}^{T} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta Y_{it},$$

where

$$K_{it}=\frac{1}{|H_2|^{1/2}}K\left(H_2^{-1/2}(Z_{it}-z_1)\right) \qquad ; \qquad K_{i(t-1)}=\frac{1}{|H_2|^{1/2}}K\left(H_2^{-1/2}(Z_{i(t-1)}-z_1)\right).$$

Clearly, the two-step weighted locally constant least-squares estimator (5.9) can be written as

$$\widehat{m}_{\widehat{g}}(z_1;H_2)=\left(\widehat{m}_{\widehat{g}}(z_1;H_2)-\widehat{m}_g(z_1;H_2)\right)+\widehat{m}_g(z_1;H_2). \tag{A3.1}$$

According to (A3.1), to prove Theorem 5.1 all that we need to show is that, under the conditions established in Theorem 5.1, we obtain

$$\sqrt{NT|H_2|}\left(\widehat{m}_{\widehat{g}}(z_1;H_2)-\widehat{m}_g(z_1;H_2)\right)=o_p(1), \qquad \text{uniformly in} \quad z_1$$

$$\sqrt{NT}|H_2|\left(\widehat{m}_g(z_1;H_2)-m(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b(z_1),\upsilon(z_1)\right),$$

where

$$\begin{aligned} b(z_1) \;&=\; \mu_2(K)\left(diag_d\left(D_f(z_1)H_2\sqrt{NT|H_2|}D_{m_r}(z_1)\right)\imath_d f^{-1}_{Z_{it},Z_{i(t-1)}}(z_1,z_1)\right.\\ &+\; \left.\frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_r}(z_1)H_2\sqrt{NT|H_2|}\right)\right)\imath_d\right) \end{aligned}$$

and

$$\upsilon(z_1) \;=\; 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m(z_1)^{\top}m(z_1)+\sigma_{v\xi}m(z_1)^{\top}\imath_{(M-1)}\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)$$

as N tends to infinity and T is fixed. These results are proved in Lemmas 5.1 and 5.2.

Lemma 5.1 *Under conditions of Theorem 5.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{NT|H_2|}\left(\widehat{m}_{\widehat{g}}(z_1;H_2) - \widehat{m}_g(z_1;H_2)\right) = o_p(1), \qquad \text{uniformly in} \quad z_1.$$

In order to prove the asymptotic distribution of $\widehat{m}_{\widehat{g}}(\cdot)$ we can write (A3.1) as

$$\begin{aligned}\sqrt{NT|H_2|}\left(\widehat{m}_{\widehat{g}}(z_1;H_2) - m(z_1)\right) &= \sqrt{NT|H_2|}\left(\widehat{m}_{\widehat{g}}(z_1;H_2) - \widehat{m}_g(z_1;H_2)\right) \\ &+ \sqrt{NT|H_2|}\left(\widehat{m}_g(z_1;H_2) - m(z_1)\right) \qquad \text{(A3.2)}\end{aligned}$$

and need the following Lemma.

Lemma 5.2 *Under the conditions established in Theorem 5.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{NT|H_2|}\left(\widehat{m}_g(z_1;H_2) - m(z_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b(z_1), v(z_1)\right).$$

∎

Proof of Lemma 5.1.

Throughout this appendix we use the following notation

$$\widehat{S}_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widehat{W}_{it}\Delta\widehat{W}_{it}^{\top} \quad ; \qquad \widehat{T}_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widehat{W}_{it}\Delta Y_{it}.$$

Let us write the first element of (A3.1) as

$$\widehat{m}_{\widehat{g}}(z_1;H_2) - \widehat{m}_g(z_1;H_2) = \widehat{S}_n^{-1}\widehat{T}_n - S_n^{-1}T_n, \qquad \text{(A3.3)}$$

where $n = NT$ and S_n and T_n are the corresponding expressions of $\widehat{S}_n$ and $\widehat{T}_n$, respectively, with $g(Z_{it}, Z_{i(t-1)})$ instead of $\widehat{g}(Z_{it}, Z_{i(t-1)})$. At this situation, we first show that as N tends to infinity,

$$\widehat{S}_n^{-1} = \mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) + o_p(\|H_2^{1/2}\|), \qquad \text{(A3.4)}$$

where, remember that $\Delta\widetilde{W}_{it} = (g_{it,i(t-1)}^{\top} \quad \Delta U_{it}^{\top})^{\top}$, and

$$\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) = E\left[\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}|Z_{it} = z_1, Z_{i(t-1)} = z_1\right] f_{Z_{it},Z_{i(t-1)}}(z_1, z_1).$$

For this, let us denote

$$\widehat{S}_n = S_n + (2\mathbb{I}_{1n} + \mathbb{I}_{2n}), \tag{A3.5}$$

where as

$$\begin{aligned}
S_n &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top},\\
\mathbb{I}_{1n} &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}(\Delta\widehat{W}_{it}-\Delta\widetilde{W}_{it})^{\top},\\
\mathbb{I}_{2n} &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{it}-\Delta\widetilde{W}_{it})(\Delta\widehat{W}_{it}-\Delta\widetilde{W}_{it})^{\top},
\end{aligned}$$

so we have to analyze each term separately to prove (A3.4). For this end, we follow the usual Taylor expansion; i.e.,

$$f(z+H^{1/2}v) = f(z) + D_f^{\top}(z)H^{1/2}v + o_p(\|H^{1/2}\|), \quad as \quad \|H\|\to 0.$$

Then, given that Z_{it}, v_{it} is $i.i.d.$ across i and because the stationary assumption, when N tends to infinity and by the law of iterated expectations it implies

$$\begin{aligned}
E(S_n) &= \int\int E\left[\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}|Z_{it}=z_1+H_2^{1/2}u, Z_{i(t-1)}=z_1+H_2^{1/2}v\right]\\
&\quad \times f_{Z_{it},Z_{i(t-1)}}\left(Z_{it}=z_1+H_2^{1/2}u, Z_{i(t-1)}=z_1+H_2^{1/2}v\right)K(u)K(v)dudv.
\end{aligned}$$

Under Assumption **5.1**,

$$\begin{aligned}
Var(S_n) &= Var\left(K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\right)\\
&+ \frac{1}{T}\sum_{t=3}^{T}(T-t)Cov\left(K_{i2}K_{i1}\Delta\widetilde{W}_{i2}\Delta\widetilde{W}_{i2}^{\top}, K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\right),
\end{aligned}$$

where, under Assumptions **5.7-5.9**, it holds

$$Var\left(K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\right) = O_p\left(\frac{1}{NT|H_2|}\right)$$

and

$$Cov\left(K_{i2}K_{i1}\Delta\widetilde{W}_{i2}\Delta\widetilde{W}_{i2}^{\top}, K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\right) = O_p\left(\frac{1}{N|H_2|}\right).$$

Then if both $NT|H_2|$ and $N|H_2|$ tends to infinity, this variance term tends to zero and it is proved

$$S_n = \mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1)(1 + o_p(1)). \tag{A3.6}$$

Now, focus on the behavior $I\!I_{1n}$, by Assumptions **5.2** and **5.5-5.9** we obtain

$$I\!I_{1n} = \left(\begin{array}{c} (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta\widetilde{W}_{it} \left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)} \right)^\top \\ \\ (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta\widetilde{W}_{it} \left(\Delta U_{it} - \Delta U_{it} \right)^\top \end{array} \right) = o_p(1) \tag{A3.7}$$

given that, using the uniform convergence results as the ones established in Masry (1996, Theorem 6),

$$\begin{aligned} & (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta\widetilde{W}_{it} \left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)} \right)^\top \\ \leq \;& (NT)^{-1} \sum_{it} |K_{it} K_{i(t-1)} \Delta\widetilde{W}_{it}| \sup_{\{Z_{it}, Z_{i(t-1)}\}} |\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}|^\top \\ = \;& o_p(1), \end{aligned} \tag{A3.8}$$

since it is straightforward to show $(NT)^{-1} \sum_{it} |K_{it} K_{i(t-1)} \Delta\widetilde{W}_{it}| = O_p(1)$. Similarly, we obtain $I\!I_{2n} = o_p(1)$.

By (A3.5) we know $\widehat{S}_n = S_n + (2I\!I_{1n} + I\!I_{2n})$ and

$$\widehat{S}_n^{-1} = S_n^{-1} + S_n^{-1} \left(2I\!I_{1n} + I\!I_{2n} \right) S_n^{-1} + o_p(\|H_2^{1/2}\|).$$

Replacing these previous results here we obtain

$$\widehat{S}_n^{-1} = \mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) + o_p(\|H_2^{1/2}\|) \tag{A3.9}$$

so the result (A3.4) is proved.

On the other hand, replacing (A3.4) and (A3.6) in (A3.3) and by the Crammer-Wald device we obtain

$$\widehat{m}_{\widehat{g}}(z_1; H_2) - \widehat{m}_g(z_1; H_2) = \mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1)(\widehat{T}_n - T_n) + o_p(1). \tag{A3.10}$$

Focus now on the behavior of $\widehat{T}_n - T_n$, we claim that following the same procedure as in (A3.7) we obtain

$$\widehat{T}_n - T_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{it} - \Delta\widetilde{W}_{it})\Delta Y_{it} = o_p(1), \qquad \text{(A3.11)}$$

since by Assumption **5.9** and using the uniform convergence of $\widehat{g}_{it,i(t-1)}$ we obtain that since $(NT)^{-1}\sum_{it}|K_{it}K_{i(t-1)}\Delta Y_{it}| = O_p(1)$,

$$\begin{aligned} & (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}\right)\Delta Y_{it} \\ \leq\ & (NT)^{-1}\sum_{it}|K_{it}K_{i(t-1)}\Delta Y_{it}| \sup_{\{Z_{it},Z_{i(t-1)}\}} |\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}| \\ =\ & o_p(1) \end{aligned}$$

Then, as $N|H_2| \to \infty$ by (A3.10) and (A3.11) we obtain

$$\widehat{m}_{\widehat{g}}(z_1; H_2) - \widehat{m}_g(z_1; H_2) = o_p\left(\frac{1}{\sqrt{NT|H_2|}}\right),$$

so the Lemma 5.1 is proved.

∎

Proof of Lemma 5.2

The proof of this proposition is structured as follows. First, the asymptotic bias of the estimator is analyzed. Later, we focus on the variance term and conclude the proof with the asymptotic normality of the estimator, once confirmed that the Lyapunov condition holds.

Since by the regularity conditions the Taylor's remainder term is $o_p\left(tr(H_2)\right)$, the approximation of the smooth functions of (5.7) by the Taylor theorem implies

$$\Delta Y_{it} = \Delta\widetilde{W}_{it}^{\top} m(z_1) + G_{it} + \Delta v_{it} + \Delta\xi_{it}^{\top} m_2(z_1) + o_p(1) \qquad \text{(A3.12)}$$

once replaced ΔW_{it} by $g(X_{it}, X_{i(t-1)}) + \Delta\xi_{it}$ in the resulting expression, where

$$\begin{aligned} G_{it} =\ & \left(\widetilde{W}_{it} \otimes (Z_{it} - z_1) - \widetilde{W}_{i(t-1)} \otimes (Z_{i(t-1)} - z_1)\right)^{\top} D_m(z_1) + \Delta Z_{it}^{\top} D_{\alpha}(z_1) \\ +\ & \frac{1}{2}\left(\widetilde{W}_{it}^{\top} \otimes (Z_{it} - z_1)^{\top}\mathcal{H}_m(z_1)(Z_{it} - z_1) - \widetilde{W}_{i(t-1)}^{\top} \otimes (Z_{i(t-1)} - z_1)^{\top}\mathcal{H}_m(z_1)(Z_{i(t-1)} - z_1)\right) \\ +\ & \frac{1}{2}\left((Z_{it} - z_1)^{\top}\mathcal{H}_{\alpha}(z_1)(Z_{it} - z_1) - (Z_{i(t-1)} - z_1)^{\top}\mathcal{H}_{\alpha}(z_1)(Z_{i(t-1)} - z_1)\right), \end{aligned}$$

let $D_m(z_1)$ be a $(d-1)q\times 1$ vector and $D_\alpha(z_1)$ a $q\times 1$ vector, for $D_m(z_1) = vec(\partial m(z_1)/\partial z_1^\top)$ and $D_\alpha(z_1) = vec(\partial \alpha(z_1)/\partial z_1^\top)$ being the corresponding first-order derivatives vector of $m(\cdot)$ and $\alpha(\cdot)$, respectively. Also, $\mathcal{H}_m(z_1)$ is a $(d-1)q \times q$ matrix and $\mathcal{H}_\alpha(z_1)$ a $q \times q$ matrix, for $\mathcal{H}_m(z_1) = \partial^2 m(z_1)/\partial z_1 z_1^\top$ and $\mathcal{H}_\alpha(z_1) = \partial^2 \alpha(z_1)/\partial z_1 z_1^\top$ being the corresponding Hessian matrix of $m(\cdot)$ and $\alpha(\cdot)$, respectively.

Combining (A3.12) with the second element of (A3.2), we can write $\widehat{m}_g(\cdot)$ as

$$\widehat{m}_g(z_1; H_2) - m(z_1) = S_n^{-1}\left(U_n + B_n + R_n\right) \tag{A3.13}$$

where, for the sake of simplicity, let us denote

$$\begin{aligned} U_n &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta v_{it}, \\ R_n &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\xi_{it}^\top m_2(z_1), \\ B_n &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}G_{it}. \end{aligned}$$

Thus, to complete the proof of Theorem 5.1 it is enough to show

$$\sqrt{NT|H_2|}(\widehat{m}_g(z_1; H_2) - m(z_1)) - \sqrt{NT|H_2|}S_n^{-1}B_n = \sqrt{NT|H_2|}S_n^{-1}\left(U_n + R_n\right), \tag{A3.14}$$

where we will demonstrate that $S_n^{-1}B_n$ contributes to the asymptotic bias and the two terms of the right part of (A3.14) are asymptotically normal.

Focus first in the asymptotic behavior of the bias term, we can decompose B_n into four different terms that we have to analyze separately. In particular, for the standard case $\mu_2(K_u) = \mu_2(K_v)$ we obtain

$$B_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}G_{it} = B_n^{(1)} + B_n^{(2)} + B_n^{(3)} + B_n^{(4)}, \tag{A3.15}$$

where

$$\begin{aligned} B_n^{(1)} &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\left(\widetilde{W}_{it}\otimes(Z_{it}-z_1) - \widetilde{W}_{i(t-1)}\otimes(Z_{i(t-1)}-z_1)\right)^\top D_m(z_1), \\ B_n^{(2)} &= (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta Z_{it}^\top D_\alpha(z_1), \\ B_n^{(3)} &= (2NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\left(\widetilde{W}_{it}^\top\otimes(Z_{it}-z_1)^\top\mathcal{H}_m(z_1)(Z_{it}-z_1)\right. \\ &- \left.\widetilde{W}_{i(t-1)}^\top\otimes(Z_{i(t-1)}-z_1)^\top\mathcal{H}_m(z_1)(Z_{i(t-1)}-z_1)\right), \end{aligned}$$

and

$$
\begin{aligned}
B_n^{(4)} &= (2NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{it} \left((Z_{it} - z_1)^\top \mathcal{H}_\alpha(z_1)(Z_{it} - z_1) \right. \\
&- \left. (Z_{i(t-1)} - z_1)^\top \mathcal{H}_\alpha(z_1)(Z_{i(t-1)} - z_1) \right).
\end{aligned}
$$

For stationary and using iterated expectations, we denote $\mathcal{B}_{\widetilde{W}} = E(\widetilde{W}_{it}|X_{it}, X_{i(t-1)})$ and $\mathcal{B}_{\widetilde{W}_{(-1)}} = E(\widetilde{W}_{i(t-1)}|X_{it}, X_{i(t-1)})$ obtaining

$$
\begin{aligned}
E(B_n^{(1)}) &= E\left[K_{it} K_{i(t-1)} \left(E\left(\Delta \widetilde{W}_{it} E(\widetilde{W}_{it}^\top | X_{it}, X_{i(t-1)}) | Z_{it}, Z_{i(t-1)} \right) \otimes (Z_{it} - z_1)^\top \right.\right. \\
&- \left.\left. E\left(\Delta \widetilde{W}_{it} E(\widetilde{W}_{i(t-1)}^\top | X_{it}, X_{i(t-1)}) | Z_{it}, Z_{i(t-1)} \right) \otimes (Z_{i(t-1)} - z_1)^\top \right) \right] \\
&= \int \left(E\left(\Delta \widetilde{W}_{it} \mathcal{B}_{\widetilde{W}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_1 \right) D_f(z_1)(H_2^{1/2} u) \right) \otimes (H_2^{1/2} u)^\top D_m(z_1) K(u) K(v) du dv \\
&- \int \left(E\left(\Delta \widetilde{W}_{it} \mathcal{B}_{\widetilde{W}_{(-1)}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_1 \right) D_f(z_1)(H_2^{1/2} v) \right) \otimes (H_2^{1/2} v)^\top D_m(z_1) K(u) K(v) du dv \\
&= \mu_2(K) \mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) diag_d \left(D_f(z_1) H_2 D_{m_r}(z_1) \right) \imath_d f^{-1}_{Z_{it}, Z_{i(t-1)}}(z_1, z_1) + o_p(H_2), \qquad \text{(A3.16)}
\end{aligned}
$$

for $r = 1, \cdots, d$ and being $\imath_d$ a $d \times 1$ unitary vector.

Similarly, by the law of iterated expectations

$$
\begin{aligned}
E(B_n^{(2)}) &= E\left[K_{it} K_{i(t-1)} E(\Delta \widetilde{W}_{it} | Z_{it}, Z_{i(t-1)}) \Delta Z_{it}^\top D_\alpha(z_1) \right] \\
&= \int K_{it} K_{i(t-1)} E(\Delta \widetilde{W}_{it} | Z_{it}, Z_{i(t-1)}) \Delta Z_{it}^\top D_\alpha(z_1) f\left(Z_{it}, Z_{i(t-1)} \right) dZ_{it} dZ_{i(t-1)} \\
&= E(\Delta \widetilde{W}_{it} | Z_{it} = z_1, Z_{i(t-1)} = z_1) \left(\mu_2(K_u) - \mu_2(K_v) \right) D_f(z_1) H_2 D_\alpha(z_1) \\
&= o_p(1). \qquad \text{(A3.17)}
\end{aligned}
$$

On its part, following a similar procedure as in (A3.16) it is straightforward to show

$$
\begin{aligned}
E(B_n^{(3)}) &= \frac{1}{2} E\left[K_{it} K_{i(t-1)} \left(E\left(\Delta \widetilde{W}_{it} \mathcal{B}_{\widetilde{W}}^\top | Z_{it}, Z_{i(t-1)} \right) \otimes (Z_{it} - z_1)^\top \mathcal{H}_m(z_1)(Z_{it} - z_1) \right.\right. \\
&- \left.\left. E\left(\Delta \widetilde{W}_{it} \mathcal{B}_{\widetilde{W}_{(-1)}}^\top | Z_{it}, Z_{i(t-1)} \right) \otimes (Z_{i(t-1)} - z_1)^\top \mathcal{H}_m(z_1)(Z_{i(t-1)} - z_1) \right) \right] \\
&= \frac{1}{2} \mu_2(K) \mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1) diag_d \left(tr(\mathcal{H}_{m_r}(z_1) H_2) \right) \imath_d + o_p\left(tr(H_2) \right), \qquad \text{(A3.18)}
\end{aligned}
$$

where $diag_d \left(tr(\mathcal{H}_{m_r}(z_1) H_2) \right)$ stands for a diagonal matrix of element $tr(\mathcal{H}_{m_r}(z_1) H_2)$ while, following the procedure of (A3.17) and (A3.18),

$$\begin{aligned}
E(B_n^{(4)}) &= \frac{1}{2}E\left[K_{it}K_{i(t-1)}E(\Delta\widetilde{W}_{it}|Z_{it},Z_{i(t-1)})\left((Z_{it}-z_1)^\top\mathcal{H}_\alpha(z_1)(Z_{it}-z_1)\right.\right.\\
&- \left.\left.(Z_{i(t-1)}-z_1)^\top\mathcal{H}_\alpha(z_1)(Z_{i(t-1)}-z_1)\right)\right]\\
&= \frac{1}{2}E\left(\Delta\widetilde{W}_{it}|Z_{it}=z_1,Z_{i(t-1)}=z_1\right)(\mu_2(K_u)-\mu_2(K_v))tr\left(\mathcal{H}_\alpha(z_1)H_2\right)+o_p\left(tr(H_2)\right).
\end{aligned} \tag{A3.19}$$

Furthermore, it is easy to prove that any component of the variance of B_n converges to zero following a similar procedure as in the proof of Lemma 5.1 and assuming $H_2 \to 0$ and $N|H_2| \to \infty$. Then, replacing (A3.16)-(A3.19) in B_n, using (A3.8) and applying the Crammer-Wald device we obtain that the bias term of this two-state least-squares constant regression estimator (5.9) is

$$\begin{aligned}
S_n^{-1}B_n &= \mu_2(K)\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)\\
&\times \left[diag_d\left(D_f(z_1)H_2D_{m_r}(z_1)\right)\imath_d f^{-1}_{Z_{it},Z_{i(t-1)}}(z_1,z_1)+\frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_r}(z_1)H_2\right)\right)\imath_d\right]\\
&+ o_p\left(tr(H_2)\right),
\end{aligned} \tag{A3.20}$$

so the first part of the proof is done.

On the other hand, to obtain the asymptotic variance of the right part of (A3.14) we have to analyze the variance of U_n and R_n as well as the covariance between both terms. For this, let us denote by $\Delta v = (\Delta v_1, \cdots, \Delta v_N)$ the $N(T-1)\times 1$-vector for $\Delta v_i = (\Delta v_{i2}, \cdots, \Delta v_{iT})^T$,

$$E(\Delta v_i \Delta v_{i'}^\top|\mathbf{Z}) = \begin{cases} 2\sigma_v^2, & \text{for} \quad i=i', \quad t=t' \\ -\sigma_v^2, & \text{for} \quad i=i', \quad |t-t'|<2, \\ 0, & \text{for} \quad i=i', \quad |t-t'|\geq 2, \end{cases} \tag{A3.21}$$

Replacing in (A3.21) σ_v^2 by σ_ξ^2, a similar definition we obtain for $E(\Delta\xi_i\Delta\xi_{i'}^\top|\mathbf{Z})$.

We first analyze U_n and take into account the fact that $E(\Delta v_{it}|Z_{it},Z_{i(t-1)})=0$. Thus, by the law of iterated expectations and Assumptions **5.1**, **5.2** and **5.4-5.9**, we claim that

$$\begin{aligned}
NT|H_2|Var(U_n) &= |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[E\left(\Delta\widetilde{W}_{it}E(\Delta v_{it}\Delta v_{i't'}|X_{it},X_{i(t-1)},X_{i't'},X_{i'(t'-1)})\right.\right.\\
&\times \left.\left.\Delta\widetilde{W}_{i't'}^\top|Z_{it},Z_{i(t-1)},Z_{it'},Z_{i(t'-1)}\right)K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right]\\
&= 2\sigma_v^2R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)(1+o_p(1)).
\end{aligned} \tag{A3.22}$$

To show this result, note that the covariance between different individuals are clearly zero by the independence condition. Therefore, for $i = i'$ we consider two different cases: $t = t'$ and $t \neq t'$. For $t = t'$ and Assumptions **5.1**, **5.2** and **5.4-5.9**, by the standard kernel methods we obtain

$$\begin{aligned} & |H_2|T^{-1}\sum_{t=2}^{T} E\left[E\left(\Delta\widetilde{W}_{it}E(\Delta v_{it}^2|X_{it}, X_{i(t-1)})\Delta\widetilde{W}_{it}^\top|Z_{it}, Z_{i(t-1)}\right)K_{it}^2K_{i(t-1)}^2\right] \\ = \quad & 2\sigma_v^2|H_2|E\left[E\left(\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^\top|Z_{it}, Z_{i(t-1)}\right)K_{it}^2K_{i(t-1)}^2\right] \\ = \quad & 2\sigma_v^2R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1)(1 + o_p(1)). \end{aligned}$$

Meanwhile, for $t \neq t'$, and proceeding in the same way as in the preceding equation, if we consider again the stationarity Assumption **5.1**, we obtain

$$\begin{aligned} & 2|H_2|T^{-1}\sum_{t=3}^{T}(T-t)E\left[E\left(\Delta\widetilde{W}_{i2}E(\Delta v_{i2}\Delta v_{it}|X_{i2}, X_{i1}, X_{it}, X_{i(t-1)})\Delta\widetilde{W}_{it}^\top|Z_{i2}, Z_{i1}, Z_{it}, Z_{i(t-1)}\right)\right. \\ & \qquad \left.\times K_{i2}K_{i1}K_{it}K_{i(t-1)}\right] \\ & = -2\sigma_v|H_2|^{1/2}R(K_u)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1, z_1)(1 + o_p(1)), \end{aligned}$$

where

$$\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1, z_1) = E\left[\Delta\widetilde{W}_{i2}\Delta\widetilde{W}_{i3}^\top|Z_{i1} = z_1, Z_{i2} = z_1, Z_{i3} = z_1\right]f_{Z_{i1},Z_{i2},Z_{i3}}(z_1, z_1, z_1).$$

Note that only those terms of the variance-covariance matrix in which $|t - t'| < 2$ holds are nonzero. The remaining terms of this matrix are zero by the structure of the error term in first differences established in (A3.21).

Second, we focus on the behavior of R_n and follow a similar procedure as in (A3.22),

$$\begin{aligned} NT|H_2|Var(R_n) \quad = \quad & |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta\widetilde{W}_{it}m(z_1)^\top E(\Delta\xi_{it}\Delta\xi_{i't'}^\top|X_{it}, X_{i(t-1)}, X_{i't'}, X_{i'(t'-1)})\right. \\ & \left.\times\, m(z_1)\Delta\widetilde{W}_{i't'}^\top K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right] \\ = \quad & 2R(K_u)R(K_v)\sigma_\xi^2 m(z_1)^\top m(z_1)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1)(1 + o_p(1)). \end{aligned} \tag{A3.23}$$

For the last term, we obtain

$$\begin{aligned} NT|H_2|Cov(U_n, R_n) \quad = \quad & |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta\widetilde{W}_{it}m(z_1)^\top E(\Delta v_{it}\Delta\xi_{i't'}|X_{it}, X_{i(t-1)}, X_{i't'}, X_{i'(t'-1)})\right. \\ \times \quad & \left.\Delta\widetilde{W}_{i't'}^\top K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right] \\ = \quad & R(K_u)R(K_v)\sigma_{v\xi}m(z_1)^\top\imath_{(M-1)}\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1, z_1)(1 + o_p(1)), \end{aligned} \tag{A3.24}$$

following the same procedure as in (A3.22).

Applying the Crammer-Wald device and using (A3.8) and (A3.22)-(A3.24), as $N|H_2| \to \infty$ we obtain

$$\begin{aligned} NT|H_2|Var\left(S_n^{-1}\left(U_n+R_n\right)\right) &= 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m(z_1)^\top m(z_1)+\sigma_{v\xi}m(z_1)^\top \imath_{(M-1)}\right) \\ &\times \mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)(1+o_p(1)). \end{aligned} \quad \text{(A3.25)}$$

Finally, once established the main asymptotic properties of the two-stage least-squares local constant regression estimator (5.9), to prove Theorem 5.1 is necessary to show that as $N \to \infty$,

$$\begin{aligned} &\sqrt{NT|H_2|}\left(\widehat{m}_g(z_1;H_2)-m(z_1)\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m(z_1)^\top m(z_1)+\sigma_{v\xi}m(z_1)^\top \imath_{(M-1)}\right)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)^{-1}\right). \end{aligned} \quad \text{(A3.26)}$$

In order to show that, we check the Lyapunov condition. As the reader can appreciate, the above assumptions state that the variables are *i.i.d.* in the subscript i but not in t, so we have independent random variables heterogeneously distributed. To overcome this situation, we may define $\lambda^*_{n,i} = T^{-1/2}\sum_{it}\lambda_{it}$, what means that $\lambda^*_{n,i}$ is an independent random variable for T fixed. Therefore, in order to show (A3.26) it suffices to analyze the asymptotic normality of the two-stage least-squares local constant regression estimator (5.9),

$$\frac{1}{\sqrt{NT}}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}(\Delta v_{it}+\Delta\xi_{it}^\top m(z_1)) = \frac{1}{\sqrt{NT}}\sum_{it}\lambda_{it}, \quad \text{(A3.27)}$$

where

$$\lambda_{it} = K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\left(\Delta v_{it}+\Delta\xi_{it}^\top m(z_1)\right)|H_2|^{1/2}, \quad i=1,\cdots,N \quad ; \quad t=2,\cdots,T.$$

By Theorem 5.1 and previous proofs, we can state that as $H_2 \to 0$,

$$Var(\lambda_{it}) = 2R(K_u)R(K_v)\left(\sigma_v^2+\sigma_\xi^2 m(z_1)^\top m(z_1)\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)(1+o_p(1))$$

and

$$Cov(\lambda_{i1},\lambda_{it}) = R(K_u)R(K_v)\sigma_{v\xi}m(z_1)^\top \imath_{(M-1)}\mathcal{B}^{-1}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z_1,z_1)(1+o_p(1)).$$

By the Minkowski inequality we obtain

$$E|\lambda_{n,i}^*|^{2+\delta} \leq CT^{\frac{2+\delta}{2}} E|\lambda_{it}|^{2+\delta},$$

so λ_{it} can be divided into two components; i.e., λ_{1it} and λ_{2it}. Analyzing separately each of those terms, we obtain

$$\begin{aligned}
E|\lambda_{1it}|^{2+\delta} &\leq |H_2|^{(2+\delta)/2} E|K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta v_{it}|^{2+\delta} \\
&= |H_2|^{-\delta/2}\int E\left(|\Delta\widetilde{W}_{it}\Delta v_{it}|^{2+\delta}|Z_{it}=z_1+H_2^{1/2}u, Z_{i(t-1)}=z_1+H_2^{1/2}v\right) \\
&\quad \times f_{Z_{it},Z_{i(t-1)}}\left(z_1+H_2^{1/2}u, z_1+H_2^{1/2}v\right)K^{2+\delta}(u)K^{2+\delta}(v)dudv \\
&= |H_2|^{-\delta/2}E\left(|\Delta\widetilde{W}_{it}\Delta v_{it}|^{2+\delta}|Z_{it}=z_1, Z_{i(t-1)}=z_1\right)f_{Z_{it},Z_{i(t-1)}}(z_1,z_1) \\
&\quad \times\int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
E|\lambda_{2it}|^{2+\delta} &\leq |H_2|^{(2+\delta)/2}E|K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}\Delta\xi_{it}^{\top}m(z_1)|^{2+\delta} \\
&= |H_2|^{1+\delta/2}E\left[E\left(|\Delta\widetilde{W}_{it}\Delta\xi_{it}^{\top}|^{2+\delta}|Z_{it},Z_{i(t-1)}\right)m(z_1)^{2+\delta}K_{it}^{2+\delta}K_{i(t-1)}^{2+\delta}\right] \\
&= |H_2|^{-\delta/2}E\left(|\Delta\widetilde{W}_{it}\Delta\xi_{it}^{\top}|^{2+\delta}|Z_{it}=z_1,Z_{i(t-1)}=z_1\right)m(z_1)^{2+\delta}f_{Z_{it},Z_{i(t-1)}}(z_1,z_1) \\
&\times \int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2})
\end{aligned}$$

and

$$\begin{aligned}
E|\lambda_{1it}\lambda_{2it}^{\top}|^{1+\delta/2} &\leq |H_2|^{(2+\delta)/2}E|K_{it}^2K_{i(t-1)}^2\Delta\widetilde{W}_{it}\Delta v_{it}m(z_1)^{\top}\Delta\xi_{it}\Delta\widetilde{W}_{it}^{\top}|^{1+\delta/2} \\
&= -|H_2|^{1+\delta/2}E\left[E\left(|\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\Delta v_{it}\Delta\xi_{it}|^{1+\delta/2}|Z_{it},Z_{i(t-1)}\right)(m(z_1)^{\top})^{(1+\delta/2)}K_{it}^{2+\delta}K_{i(t-1)}^{2+\delta}\right] \\
&= -|H_2|^{1+\delta/2}E\left(|\Delta\widetilde{W}_{it}\Delta\widetilde{W}_{it}^{\top}\Delta v_{it}\Delta\xi_{it}|^{1+\delta/2}|Z_{it}=z_1,Z_{i(t-1)}=z_1\right)(m(z_1)^{\top})^{(1+\delta/2)} \\
&\quad \times f_{Z_{it},Z_{i(t-1)}}(z_1,z_1)\int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2}).
\end{aligned}$$

Then, it is proved that

$$E|\lambda_{n,i}|^{2+\delta} = E|N^{-1/2}\sum_i \lambda_{n,i}^*|^{2+\delta} \leq CO_p((N|H_2|)^{-\delta/2})$$

and, as N tends to infinity, $N|H_2| \to \infty$. Since the Lyapunov condition holds, we resort to the Lyapunov Central Limit Theorem to verify (A3.26) and Lemma 5.2 is proved.

Finally, using the results of Lemmas 5.1 and 5.2 in (A3.13) we obtain

$$\sqrt{NT}|H_2|\left(\widehat{m}_{\widehat{g}}(z_{11};H_2)-m(z_{11})\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(b(z_{11};H_2),v(z_{11};H_2)\right)$$

so the proof of Theorem 5.1 is done.

∎

Appendix 4

Finally, in this appendix we show the computational programs that we develop for estimates. They are implemented in R, a well-known computing statistical program.

In the following, n is the sample size in differences and we denote by DY a $N(T-1)\times 1$ vector of dependent variables in first differences, whereas X and Z are $N(T-1)\times d$ and $N(T-1)\times q$ matrices of covariates. Also, $Xlag$ and $Zlag$ contains the first lag of X and Z, respectively. H is the bandwidth obtained by any standard nonparametric procedure as the rule-of-thumb or the cross-validation technique. We also denote by $mlocalFD$ the corresponding $\widehat{m}_F(Z;H)$ estimator.

Algorithm 1 $\widehat{m}_F(z;H)$ estimator when $q=1$ and $d=1$

```
localFD < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Ztilde < − cbind( (X - Xlag), (X ∗ (Z - a) - Xlag ∗ (Zlag - a)))
e < − rbind (1, 0)
W < − diag( (1/H∧(2∗ q)) ∗ ( ( (1/sqrt(2∗ 3.14159)) ∗ exp ( -(1/2) ∗ ((Z - a) / H)∧2 ))
∗ ( (1/sqrt(2∗ 3.14159))∗ exp ( -(1/2) ∗ ((Zlag - a) / H)∧2 )) ))
mlocalFD < − t(e) % ∗ % solve ( t(Ztilde) % ∗ % W % ∗ % Ztilde ) % ∗ % t(Ztilde) % ∗ %
W % ∗ % DY
mod[[paste(”run”,i,sep=””)]] < − mlocalFD
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 2 $\widehat{m}_F(z; H)$ estimator when $q = 2$ and $d = 1$

```
localFD <- function(){
mod <- list()
for(i in 1:n){
a <- Z[i, 1]
b <- Z[i, 2]
Ztilde <- cbind( (X - Xlag), (X * (Z[,1] - a) - Xlag * (Zlag[,1] - b)), (X * (Z[,2] - a) -
Xlag * (Zlag[,2] - b)))
e <- rbind (1, 0, 0)
W <- diag( (1/H∧(2*q)) * ( ( (1/sqrt(2*3.14159)) * exp ( -(1/2) * ((1/H) * sqrt( (Z[,1] -
a)∧2 + (Z[,2] - b)∧2))∧2)) * ( (1/sqrt(2*3.14159)) * exp ( -(1/2) * ((1/H) * sqrt( (Zlag[,1]
- a)∧2 + (Zlag[,2] - b)∧2))∧2)) ))
mlocalFD <- t(e) %*% solve ( t(Ztilde) %*% W %*% Ztilde ) %*% t(Ztilde) %*%
W %*% DY
mod[[paste("run",i,sep="")]] <- mlocalFD
}
results <- list()
for(i in 1:n) results [[ names(mod)[i] ]] <- mod[i]
data.frame(results)
}
```

Algorithm 3 $\widehat{m}_F(z;H)$ estimator when $q=1$ and $d=2$

```
localFD < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Ztilde < − cbind( (X - Xlag), (X ∗ (Z - a) - Xlag ∗ (Zlag - a)))
e < − matrix(c(1,0,0,0,0,1,0,0), nc=d)
W < − diag( (1/H∧(2∗q)) ∗ ( ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Z1D - a) / H)∧2 ))
∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Zlag - a) / H)∧2 )) ))
mlocalFD < − t(e) %∗% solve ( t(Ztilde) %∗% W %∗% Ztilde ) %∗% t(Ztilde) %∗%
W %∗% DY
mod[[paste(”run”,i,sep=””)]] < − mlocalFD
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 4 $\widetilde{m}_F(z; H)$ estimator when $q = 1$ and $d = 1$

```
backFD < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Ztildeb < − cbind( X, (X ∗ (Z - a)) )
DYb < − DY + Xlag ∗ mlocalFD
e < − rbind( 1, 0)
Wb < − diag( (1/H∧q) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Z - a) / H)∧2 ) ))
mbackFD < − t(e) %∗% solve(t(Ztildeb) %∗% Wb %∗% Ztildeb) %∗% t(Ztildeb) %∗%
Wb %∗% DYb
mod[[paste(”run”,i,sep=””)]] < − mbackFD
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 5 $\widetilde{m}_F(z; H)$ estimator when $q = 2$ and $d = 1$

```
backFD <- function(){
mod <- list()
for(i in 1:n){
a <- Z[i, 1]
b <- Z[i, 2]
Ztildeb <- cbind( X, (X * (Z[,1] - a)), (X * (Z[,2] - b)))
DYb <- DY + Xlag * mlocalFD
e <- rbind( 1, 0, 0)
Wb <- diag( (1/H^q) * ((1/sqrt(2*3.14159)) * exp ( -(1/2) * ((1/H) * sqrt( (Zlag[,1] -
a)^2 + (Z[,2] - b)^2)^2) )))
mbackFD <- t(e) %*% solve(t(Ztildeb) %*% Wb %*% Ztildeb) %*% t(Ztildeb) %*%
Wb %*% DYb
mod[[paste("run",i,sep="")]] <- mbackFD
}
results <- list()
for(i in 1:n) results [[ names(mod)[i] ]] <- mod[i]
data.frame(results)
}
```

Algorithm 6 $\widetilde{m}_F(z;H)$ estimator when $q = 1$ and $d = 2$

```
backFD < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Ztildeb < − cbind( X, (X ∗ (Z - a)) )
DYb < − DY + Xlag ∗ mlocalFD
e < − matrix(c(1,0,0,0,0,1,0,0), nc=d)
Wb < − diag( (1/H∧q) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Z - a) / H)∧2 ) ))
mbackFD < − t(e) %∗% solve(t(Ztildeb) %∗% Wb %∗% Ztildeb) %∗% t(Ztildeb) %∗%
Wb %∗% DYb
mod[[paste("run",i,sep="")]] < − mback
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Let n be the sample size in differences and denote by $Ydot = \ddot{Y}$ a $NT \times 1$ vector of time-demeaned dependent variable. X and Z are of dimension $NT \times d$ and $NT \times q$, respectively. $Xlag$ and $Zlag$ contains the first lag of X and Z, respectively, whereas $Xlaglag$ and $Zlaglag$ are of dimension $NT \times d$ and $NT \times q$ that contains the second lag of X and Z, respectively. H is the bandwidth matrix and T is the number of time observations for each individual (in this case we assume $T = 3$). We denote by $mlocalFE$ the corresponding $\widehat{m}_w(z; H)$ estimator.

Algorithm 7 $\widehat{m}_w(z; H)$ estimator when $q = 1$, $d = 1$ and $T = 3$

```
localFE <- function(){
mod <- list()
for(i in 1:n){
a <- Z[i]
XZ <- X * (Z - a)
XZtilde <- as.vector(rep(tapply(XZ, id, mean), each=T))
Ztilde <- cbind( (X - Xtilde), (X * (Z - a) - XZtilde))
e <- rbind (1, 0)
W <- diag( (1/H^(T*q)) * ( ( (1/sqrt(2*3.14159)) * exp ( -(1/2) * ((Z - a) / H)^2 )) *
( (1/sqrt(2*3.14159)) * exp ( -(1/2) * ((Zlag - a) / H)^2 )) * ( (1/sqrt(2*3.14159)) * exp
( -(1/2) * ((Zlaglag - a) / H)^2 )) ))
mlocalFE <- t(e) %*% solve ( t(Ztilde) %*% W %*% Ztilde ) %*% t(Ztilde) %*%
W %*% Ydot
mod[[paste("run",i,sep="")]] <- mlocalFE
}
results <- list()
for(i in 1:n) results [[ names(mod)[i] ]] <- mod[i]
data.frame(results)
}
```

Algorithm 8 $\widehat{m}_w(z; H)$ estimator when $q = 2$, $d = 1$ and $T = 3$

```
localFE < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i, 1]
b < − Z[i, 2]
XZ1 < − X ∗ (Z[,1] - a)
XZ2 < − X ∗ (Z[,2] - b)
XZ1tilde < − as.vector(rep(tapply(XZ1, id, mean), each=T))
XZ2tilde < − as.vector(rep(tapply(XZ2, id, mean), each=T))
Ztilde < − cbind( (X - Xtilde), (X ∗ (Z[,1] - a) - XZ1tilde), (X ∗ (Z[,2] - b) - XZ2tilde))
e < − rbind (1, 0, 0)
W < − diag( (1/H∧(T∗q)) ∗ ( ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((1/H) ∗ sqrt( (Z[,1] - a)∧2 + (Z[,2] - b)∧2))∧2)) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((1/H) ∗ sqrt( (Zlag[,1] - a)∧2 + (Zlag[,2] - b)∧2))∧2)) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((1/H) ∗ sqrt( (Zlaglag[,1] - a)∧2 + (Zlaglag[,2] - b)∧2))∧2)) ))
mlocalFE < − t(e) % ∗ % solve ( t(Ztilde) % ∗ % W % ∗ % Ztilde ) % ∗ % t(Ztilde) % ∗ % W % ∗ % Ydot
mod[[paste("run",i,sep="")]] < − mlocalFE
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 9 $\widehat{m}_w(z; H)$ estimator when $q = 1$, $d = 2$ and $T = 3$

```
localFE < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
X1Z < − X[,1] ∗ (Z - a)
X2Z < − X[,2] ∗ (Z - a)
X1Ztilde < − as.vector(rep(tapply(X1Z, id, mean), each=T))
X2Ztilde < − as.vector(rep(tapply(X2Z, id, mean), each=T))
Ztilde < − cbind( (X - Xtilde), (X[,1] ∗ (Z - a) - X1Ztilde), (X[,2] ∗ (Z - a) - X2Ztilde))
e < − matrix(c(1,0,0,0,0,1,0,0), nc=d)
W < − diag( (1/H∧(T∗q)) ∗ ( ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Z - a) / H)∧2 )) ∗
( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Zlag - a) / H)∧2 )) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp
( -(1/2) ∗ ((Zlaglag - a) / H)∧2 )) ))
mlocalFE < − t(e) %∗% solve ( t(Ztilde) %∗% W %∗% Ztilde ) %∗% t(Ztilde) %∗%
W %∗% Ydot
mod[[paste("run",i,sep="")]] < − mlocalFE
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 10 $\widetilde{m}_w(z;H)$ estimator when $q = 1$, $d = 1$ and $T = 3$

```
backFE < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Ztildeb < − cbind( X, (X ∗ (Z - a)) )
Xmlocal < − X∗ mlocalFE
Xmhattilde < − as.vector(rep(tapply(Xmlocal,id,mean),each=T))
Ydotb < − Ydot + Xmhattilde
e < − rbind( 1, 0)
Wb < − diag( (1/H∗q) ∗ ( (1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((Z - a) / H)∧2 ) ))
mbackFE < − t(e) %∗% solve(t(Ztildeb) %∗% Wb %∗% Ztildeb) %∗% t(Ztildeb) %∗%
Wb %∗% Ydotb
mod[[paste(”run”,i,sep=””)]] < − mbackFE
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 11 $\widetilde{m}_w(z;H)$ estimator when $q=2$, $d=1$ and $T=3$

```
backFE < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i, 1]
b < − Z[i, 2]
Ztildeb < − cbind( X, (X ∗ (Z[,1] - a)), (X ∗ (Z[,2] - b)) )
Xmlocal < − X∗ mlocalFE
Xmhattilde < − as.vector(rep(tapply(Xmlocal,id,mean),each=T))
Ydotb < − Ydot + Xmhattilde
e < − rbind( 1, 0, 0)
Wb < − diag( (1/H∧q) ∗ ( ((1/sqrt(2∗3.14159)) ∗ exp ( -(1/2) ∗ ((1/H) ∗ sqrt( (Z[,1] -
a)∧2 + (Z[,2] - b)∧2))∧2) )))
mbackFE < − t(e) %∗% solve(t(Ztildeb) %∗% Wb %∗% Ztildeb) %∗% t(Ztildeb) %∗%
Wb %∗% Ydotb
mod[[paste(”run”,i,sep=””)]] < − mbackFE
}
results < − list()
for(i in 1:n) results [[ names(mod)[i] ]] < − mod[i]
data.frame(results)
}
```

Algorithm 12 $\widetilde{m}_w(z;H)$ estimator when $q=1$, $d=2$ and $T=3$

```
backFE <- function(){
mod <- list()
for(i in 1:n){
a <- Z[i]
Ztildeb <- cbind( X, (X * (Z - a)) )
X1mlocal <- X[,1] * mlocalFE[,1]
X2mlocal <- X[,2] * mlocalFE[,2]
X1mhattilde <- as.vector(rep(tapply(X1mlocal, id, mean), each=T))
X2mhattilde <- as.vector(rep(tapply(X2mlocal, id, mean), each=T))
Ydotb <- Ydot + X1mhattilde + X2mhattilde
e <- matrix(c(1,0,0,0,0,1,0,0), nc=d)
Wb <- diag( (1/H^q) * ( (1/sqrt(2*3.14159)) * exp ( -(1/2) * ((Z - a) / H)^2 ) ))
mbackFE <- t(e) %*% solve(t(Ztildeb) %*% Wb %*% Ztildeb) %*% t(Ztildeb) %*%
Wb %*% Ydotb
mod[[paste("run",i,sep="")]] <- mbackFE
}
results <- list()
for(i in 1:n) results [[ names(mod)[i] ]] <- mod[i]
data.frame(results)
}
```

Let us denote by $DWHAT = \widehat{g}_{it,i(t-1)}$ the nonparametric estimator obtained previously through standard techniques such as the Nadaraya-Watson estimator of the local linear regression. Let n be the sample size in differences and $ntot$ the full sample size, DU is a $N(t-1) \times a$ matrix whereas Z and $Zlag$ are of dimension $N(T-1) \times q$. We also denote by H_2 the bandwidth that is obtained through standard nonparametric techniques and that holds $H_1 = o_p(H_2)$, where H_1 is the bandwidth of the first-stage of the procedure. Also, H_3 is the bandwidth of the last stage of the procedure. Finally, ind denote the number of individuals of the sample, whereas $t1$ is the number of temporal observations per individual.

Algorithm 13 $\widetilde{m}_{\widehat{g}}(z_1; H_2)$ estimator when $q = 1$ and $M_1 = 1$

```
Nad < − function(){
mod < − list()
for(i in 1:n){
a < − Z[i]
Xtilde < − cbind(DWHAT,DU)
W < − diag((1/H2∧(2∗q)) ∗ ( ((1/sqrt(2∗3.14159)) ∗ exp(-(1/2) ∗ ((Z - a)/H2)∧2)) ∗
((1/sqrt(2∗3.14159)) ∗ exp(-(1/2) ∗ ((Zlag - a)/H2)∧2)) ))
mnad < − solve(t(Xtilde) % ∗ % W % ∗ % Xtilde) % ∗ % t(Xtilde) % ∗ % DY
m1nad < − cbind(1,0) % ∗ % mnad
m2nad < − cbind(0,1) % ∗ % mnad
mod[[paste(”run”, i, sep=””)]] < − cbind(m1nad,m2nad)
}
results < − do.call(rbind,mod)
data.frame(results)
}
```

Algorithm 14 $\widetilde{\alpha}(z_1; H_x)$ estimator when $q = 1$ and $M_1 = 1$

```
alphamarg <- function(){
mod <- list()
for(j in 1:n){
alpha <- matrix(0, ntot, n)
for(i in 1:ntot){
I <- cbind(rep(1, n))
W <- diag((1/H3^(2*q)) * ( ((1/sqrt(2*3.14159)) * exp(-(1/2) * ((Z - Z[j])/H3)^2)) *
((1/sqrt(2*3.14159)) * exp(-(1/2) * ((Zlag - Z[i])/H3)^2)) ))
alpha[i,] <- solve(t(I) %*% W %*% I) %*% t(I) %*% W
}
alphaNad <- (1/(ind*t1)) * apply(alpha, 2, sum)
mod[[paste("run", j, sep="")]] <- cbind(alphaNad)
}
results <- do.call(rbind, mod)
data.frame(results)
}
DYHAT <- as.numeric(DY - DWHAT*m1nad - DW2*m2nad)
alphahat <- Salpha %*% DYHAT
```