Precautionary savings over the life cycle: efficient two-stage nonparametric instrumental variables estimation

JUAN M. RODRIGUEZ-POO AND ALEXANDRA SOBERON

Department of Economics, Universidad de Cantabria, Avda. Los Castros s/n, E-39005, Santander, Spain
E-mail: juan.rodriguez@unican.es, alexandra.soberon@unican.es

March 2, 2015

Abstract

This paper considers the problem of estimating a structural model of optimal life-cycle savings, controlling both for uncertainty about health-care expenditures and household risk aversion. To avoid misleading inference, we propose a new estimation technique that compared to those already proposed in the literature enables us to deal simultaneously with the presence of (i) unobserved cross-sectional heterogeneity, (ii) flexible unknown form in the Euler equation and (iii) endogenous covariates. We propose one-step backfitting estimators based on first differencing techniques that are combined through a minimum distance estimation approach. The resulting estimators are shown to be oracle efficient and they exhibit the optimal rate of convergence for this smoothness problem. To illustrate the feasibility and possible gains of this method, we present an application about household's precautionary savings over the life-cycle. From this empirical application, we obtain (i) households accumulate wealth at least in two periods in life. In younger stages, in order to guard against uncertainty about potential income downturns, and when they become older due to other reasons such as retirement and bequests. (ii) Public health programs have a negative impact on precautionary savings. The paper concludes with a Monte Carlo simulation where the small sample properties of the estimators. [1]

JEL codes: C01, C14, C33
Key Words: Panel data, varying coefficient models, endogeneity, fixed effects, marginal integration, life-cycle hypothesis model, precautionary savings, backfitting, minimum distance estimator.

---

1 Introduction

This paper is focused on the evaluation of the potential impact of both the age of the individuals and their risk aversion on household savings. In order to maintain a constant utility level in all periods, the life-cycle hypothesis model (LCH henceforth) of Modigliani and Brumberg (1954) states that, there is a hump-shaped age-saving profile since individuals tend to save from the middle of their life until retirement and dissave in younger and older ages. In this context, savings turn on a protection tool against unforeseen adverse conditions and as the result of the prudent behavior of households. Then, we note that households save primarily for two reasons, to finance expenses after retirement (life cycle motive) and to protect consumption against unexpected shocks (precautionary motive).

In the past two decades, there have been 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 errors.

The sources of misspecification that are typically ignored can be grouped into three major weaknesses. First, the effect of the individual preferences is a relevant aspect when we try to model the household behavior. In the regression model, it usually appears in the form of unobserved heterogeneity, but unfortunately most of these studies omit this topic. Second, uncertainty about household's out-of-pocket medical expenses is a relevant issue for preventive savings, as it is noted in Palumbo (1999), and the decision of how much to spend depends on some social and demographic features of the household. Considering this variable as exogenous can lead to inconsistent estimators of the parameters (functions) of interest. Finally, these models are usually based on a log-linearized Euler equation that is captured in the form of a fully parametric model. However, as it is noted in Attanasio et al. (1999) this rigid functional form can be a poor approximation to the marginal utility (smooth) function and more flexible procedures such as nonparametric and semi-parametric models are more suitable; see Chou et al. (2004), Maynard and Qiu (2009), Gao et al. (2012) or Kuan and Chen (2013). Unfortunately, these latter studies do not consider either individual heterogeneity nor endogeneity.

The aim of this paper is to contribute to the literature on precautionary savings providing a solution to the main weaknesses discussed previously. For this, we propose to extend a semiparametric LCH model as the one developed in Chou et al. (2004) to the analysis of flexible panel data models under the following particularities: (i) presence of fixed effects; (ii) varying coefficient form in the Euler equation allowed to capture the potential relationship between the endogenous and the explanatory variable; and (iii) health-care spending endogenously determined. To our knowledge, the estimation procedure that we develop in this paper is the first that enables us to estimate directly the impact of different types of uncertainty on the household precautionary savings in a panel data context with fixed effects and endogenous regressors without resorting to restrictive assumptions about functional

forms, such as it is common with the well-known log-linearized version of the Euler equation.

Specifically, the general idea that we propose with this new procedure is to estimate the unknown functions of interest through a simple estimator based on differencing techniques. Also, to cope with the endogeneity issue we use as explanatory variables the predicted values of the (possibly) endogenous variables generated from the nonparametric estimation of the reduced form equation. We establish that the resulting nonparametric local constant IV estimator is consistent and asymptotically normal but achieves a slower rate of convergence. The reason is that a bivariate kernel weight is needed to control the distance between the fixed term of the approximation of the unknown function and all the values of the sample but the variance of the estimator is enlarged.

In order to avoid the previous difficulties, following the ideas of Fan and Zhang (1999), we develop a one-step backfitting procedure. It turns out that the resulting estimator is oracle efficient and it exhibits an optimal rate of convergence. However, because of the additive form of the regression model two alternative consistent estimators for the same unknown function are obtained. With the aim of improving the efficiency we combine both estimators through a minimum distance estimation technique and hence, the resulting estimator is oracle efficient and achieves the optimal rate of convergence.

Direct nonparametric estimation of differencing panel data models has been considered as rather cumbersome in the literature; see Su and Ullah (2011). The reason is that, for each individual, we have an additive function with the same functional form at different times. In Henderson et al. (2008) the problem is solved using profile likelihood techniques, in Su and Lu (2013) the estimator comes out as the solution of a second order Friedholm integral equation whereas in Rodriguez-Poo and Soberon (2014, 2015) a direct strategy is proposed. Unfortunately, all these estimators are asymptotically biased in the presence of endogeneity. On its part, some IV methods have been proposed in the context of nonparametric panel data varying coefficient models with random effects. In Cai and Li (2008) it is proposed to estimate the nonparametric functions using the so-called nonparametric generalized method of moments. However, this method does not control for heterogeneity when it is correlated with some explanatory variables, and hence it renders to asymptotically biased estimators when fixed effects are present.

To show the feasibility and possible gains of this new procedure, we first study the optimal consumption decision problem of the Spanish household from 1985 to 1996. Later, we investigate the finite sample properties of the estimators. Nowadays, in the empirical studies there is a substantial interest in investigating the determinants of population welfare. Therefore, allowing health-care expenditures to have a different impact on household behavior depending on their age-group is an issue of great importance specially to political considerations.

The structure of this paper is as follows. Section 2 lays out a theoretical framework for consumer maximization. In Section 3 we set up the econometric model and the estimation procedures. In Section 4 we study the statistical properties of the previous estimators. Section 5 provides more efficient estimators such as one-step backfitting and minimum distance estimators. In Section 6 we

apply our proposed estimation procedures to a standard LCH model and we present some simulation results to investigate their finite sample performance. Finally, Section 7 concludes the paper. The proofs of the main results are collected in the Appendix.

2   Conceptual framework and econometric model

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. This significant impact of health-care expenditures on household's wealth coupled with their persistent and increasing behavior with the age of the individuals make them a significant part of this uncertainty. 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).

The aim of this section is to analyze how unexpected health expenditures influence households' savings through a stochastic LCH model, see Blanchard and Fisher (1989) and Deaton (1992). Specifically, we solve the basic problem of the consumer i at the time t in the presence of uncertainty and some endogenous covariates. We assume that individuals live T periods, work during the first $T-1$ periods and at each work period t they receive a stochastic income I_{it} and incur in an out-of-pocket health-care expenditure W_{it}. If W_{it} were known, households would decide how to spend their income between consumption C_{it} and financial wealth A_{it} by maximizing an additively time-separable utility that has a positive third-order derivative ($U'''(\cdot) > 0$), see Chou et al. (2004) for further details. Thus, according to Caballero (1990) we use a negative exponential utility function assuming that the degree of absolute risk aversion and absolute prudence are both constant and equal to α.

Therefore, the i-th household maximizes the following problem at time $t = 0$,

$$\max_{C_{it}} E_0 \left[\sum_{i=1}^{N} \sum_{t=0}^{T-1} \left(-\frac{1}{\alpha_i} \right) exp(-\alpha_i C_{it}) \right], \tag{2.1}$$

subject to

$$\begin{aligned} A_{i(t+1)} &= A_{it} + I_{it} - W_{it} - C_{it}, \\ W_{it} &= W_{i(t-1)} + \epsilon_{it} \quad ; \quad \epsilon_{it} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right), \end{aligned} \tag{2.2}$$

where health-care expenditures are modeled as a random walk. For the sake of simplicity, we assume there do not exist liquidity constraints so the discount and the interest rate are equal to zero. Then, taking first-order conditions in (2.1) the expected consumption is

$$C_{i(t+1)} = C_{it} + \frac{\alpha_i \sigma_\epsilon^2}{2} + \epsilon_{i(t+1)}, \tag{2.3}$$

4

and combining this result with the inter-temporal budget restriction (2.2) the optimal level of consumption is

$$C_{it} = \frac{1}{T-t}A_{it} + (I_{it} - W_{it}) - \frac{\alpha_i(T-t-1)\sigma_\epsilon^2}{4}, \qquad (2.4)$$

where $(I_{it} - W_{it} - C_{it})$ is the level of precautionary savings and $\alpha_i(T-t-1)\sigma_\epsilon^2/4$ reflects the effect of uncertainty. We do not enter in the debate about the importance of retirement versus bequests reasons, so the measurement of retirement that we use implicity includes both.

Analyzing in detail the optimal consumption expression (2.4), we can highlight that increases in uncertainty about future medical expenses σ_ϵ^2 or in the degree of risk aversion α_i, together with a broad horizon of life $(T-t-1)$, will lead to smaller consumption and greater precautionary savings so households turn on buffer-stock agents against potential adversities. On the other hand, if we focus on the expected consumption, equation (2.3), we observe that higher risk about unforeseen medical expenses σ_ϵ^2 or larger risk aversion α_i causes a steeper consumption path. Also, in those countries where national health programs exist σ_ϵ^2 is close to zero and therefore consumption profile upward proportionally to $(T-t-1)/4$. In other words, younger households will increase their consumption level since their incentives to save seem to be only related to potential income downturns.

Based on the previous findings we notice that households' consumption path varies in a nonlinear way with the age of individuals, however there are other issues such as uncertainty about unforeseen medical expenses, σ_ϵ^2, and risk aversion, α_i, that are relevant; see, e.g., Hubbard et al. (1994), Carroll (1997) or Deaton (1992), among others. More precisely, in our model we will assume that both σ_ϵ^2 and α_i are age-dependent parameters since the risk of incur in an out-of pocket health-care expenditure increases with the age of the individuals, whereas the wealth profile against age is usually hump-shaped over the life cycle with some peaks close to retirement. Thus, we can highlight that household consumption profile over the life cycle varies mainly for two reasons (i) for a life cycle motive relates with the age of the households (ii) and for precautionary savings.

Therefore, in order to determine the effect of uncertainty on households precautionary savings we extend the specification of Chou et al. (2004) to a panel data context with endogenous covariates so the model to estimate in this paper is

$$\begin{aligned} Y_{it} &= \alpha(Z_{it}) + W_{it}^\top m_1(Z_{it}) + U_{it}^\top m_2(Z_{it}) + \mu_i + v_{it}, \\ W_{it} &= g(X_{it}) + \zeta_i + \xi_{it}, \end{aligned} \qquad (2.5)$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where v_{it} and ξ_{it} are the idiosyncratic error terms whereas μ_i and ζ_i reflects the unobserved individual heterogeneity. Also, precautionary savings Y (i.e., income I minus consumption C) and health-care expenditures W are endogenously determined through the age of the household head Z, some financial wealth measurement U, and a vector of their demographic features X (i.e., age of the household head, number of children and so on). In this sense, household savings are characterized by the risk aversion of the family related with the age, $\alpha(\cdot)$, uncertainty about future health-care expenses, $m_1(\cdot)$, and uncertainty about income downturns, $m_2(\cdot)$.

The previous model can be extended to allow for several endogenous covariates. Assume that Z_{it}, W_{it}, U_{it} and X_{it} are vectors of covariates of dimension $q \times 1$, $(M-1) \times 1$, $a \times 1$ and $b \times 1$, respectively. Furthermore, $\alpha(Z)$ is an unknown function, $m_1(Z)$ and $g(X)$ are $(M-1) \times 1$ dimensional vectors and $m_2(Z)$ is a $a \times 1$ vector of unknown forms to estimate. 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, without loss of generality we assume

$$E(v_{it}|\mathbf{Z}, \mathbf{U}) = 0, \qquad E(\xi_{it}|\mathbf{X}) = 0, \quad E(v_{it}|\xi_{it}) \neq 0, \tag{2.6}$$

$$E(\mu_i|\mathbf{Z}, \mathbf{W}, \mathbf{U}) \neq 0, \quad E(\zeta_i|\mathbf{X}) \neq 0. \tag{2.7}$$

The econometric model specified in (2.5) is well-known in the literature of households precautionary behavior however the presence of both endogenous variables and unobserved individual heterogeneity is not considered. There have been many empirical studies that have tried to examine the relationship between household precautionary savings and uncertainty but without achieving conclusive results. Gourinchas and Parker (2002) confirm the patterns established by the LCH model. They find that early in life U.S. households behave as buffers-stock agents and accumulate wealth to face to unexpected income downturns, whereas around the age forty these savings are mainly for retirement and legacy. Also, in Cagetti (2003) the role of precautionary savings is determined to explain the behavior of household wealth. 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. Also, different papers have considered the impact of specific sources of uncertainty. In this way, in Starr-McCluer (1996) and Egen and Gruber (2001) it was found out that a reduction in the level of uncertainty, for example through unemployment insurance or public health programs, has a negative impact on households savings. On the other hand, in Gruber (1997) and Gertler and Gruber (2002) it was established that these type of programs smooth individual consumption. Recently, in Chou et al. (2004) it was shown that public health programs do have a negative effect on household savings. Furthermore, in Kuan and Chen (2013) it is shown that public health programs do have more impact on those household with higher incomes.

However, as we state previously, despite the interesting results of these studies they ignore some sources of misspecification that can override their conclusions, such as the endogeneity issue of the household's consumption decisions or the unobserved individual heterogeneity. Therefore, the aim of the method that we propose in this paper is to overcome such problems in order to estimate the impact of both types of uncertainty on household's precautionary savings; i.e., household risk aversion and unforseen health-care expenses.

3 Estimation procedure

Any attempt to estimate directly the unknown functions of (2.5) through standard nonparametric estimation techniques will render to inconsistent estimators of the underlying curves because of two reasons. On one hand, since μ_i is allowed to be correlated with Z_{it}, W_{it} and/or U_{it}, and ζ_i with X_{it}, we get $E(\mu_i|Z_{it} = z, W_{it} = w, U_{it} = u) \neq 0$ and $E(\zeta_i|X_{it} = x) \neq 0$. On the other hand, $E(v_{it}|\xi_{it}) \neq 0$ is assumed so the endogeneity problem of the model comes from the correlation between the idiosyncratic error terms of the M equations of (2.5).

In order to overcome the problem of the statistical dependence between the unobserved individual heterogeneity and the explanatory variables, we take the standard solution of removing the fixed effects by the first difference transformation, i.e.,

$$
\begin{aligned}
\Delta Y_{it} &= \alpha(Z_{it}, Z_{i(t-1)}) + \left(W_{it}^\top m_1(Z_{it}) - W_{i(t-1)}^\top m_1(Z_{i(t-1)})\right) + \left(U_{it}^\top m_2(Z_{it}) - U_{i(t-1)}^\top m_2(Z_{i(t-1)})\right) \\
&\quad + \Delta v_{it}, \\
\Delta W_{it} &= g(X_{it}, X_{i(t-1)}) + \Delta \xi_{it},
\end{aligned} \tag{3.1}
$$

for $i = 1, \cdots, N$ and $t = 2, \cdots, T$, where $\alpha(\cdot)$ and $g(\cdot)$ are $\mathbb{R}^{2q} \to \mathbb{R}$ and $\mathbb{R}^{2b} \to \mathbb{R}$ additive functions to estimate, $\alpha(Z_{it}, Z_{i(t-1)}) = \alpha(Z_{it}) - \alpha(Z_{i(t-1)})$ and $g(X_{it}, X_{i(t-1)}) = g(X_{it}) - g(X_{i(t-1)})$, respectively.

As it is noted in Su and Ullah (2011), direct nonparametric estimation of these unknown functions has been considered as rather cumbersome since the conditional expectation of ΔY_{it} over $(Z_{it}, Z_{i(t-1)})$ contains linear combinations of $W_{it}^\top m_1(Z_{it})$ and $U_{it}^\top m_2(Z_{it})$ for different t. Therefore, the kernel estimation requires some iterative procedures such as the backfitting algorithm or the marginal integration method and the asymptotic analysis of the resulting estimator is harder. In order to overcome this problem, several alternatives have been proposed. In Rodriguez-Poo and Soberon (2014, 2015) two-stage differencing techniques are proposed to obtain consistent and asymptotically optimal estimators of both functions and parameters of interest. Unfortunately, their proposal fail when endogenous covariates are considered in the econometric specification.

To illustrate our new proposal, consider the simplest case with $q = (M - 1) = a = b = 1$. The Taylor approximation in (3.1) for any $z \in \mathcal{A}$, being $\mathcal{A}$ a compact subset in a nonempty interior of $\mathbb{R}$, implies

$$
\begin{aligned}
\alpha(Z_{it}, Z_{i(t-1)}) &\approx \alpha'(z)\Delta Z_{it} + \frac{1}{2}\alpha''(z)\left((Z_{it} - z)^2 - (Z_{i(t-1)} - z)^2\right) + \cdots + \\
&\quad + \frac{1}{p!}\alpha^{(p)}(z)\left((Z_{it} - z)^p - (Z_{i(t-1)} - z)^p\right), \\
W_{it}m_1(Z_{it}) - W_{i(t-1)}m_1(Z_{i(t-1)}) &\approx \Delta W_{it}m_1(z) + \left(W_{it}(Z_{it} - z) - W_{i(t-1)}(Z_{i(t-1)} - z)\right)m_1'(z) + \\
&\quad + \frac{1}{2}\left(W_{it}(Z_{it} - z)^2 - W_{i(t-1)}(Z_{i(t-1)} - z)^2\right)m_1''(z) + \cdots + \\
&\quad + \frac{1}{p!}\left(W_{it}(Z_{it} - z)^p - W_{i(t-1)}(Z_{i(t-1)} - z)^p\right)m_1^{(p)}(z),
\end{aligned}
$$

7

and similarly for $U_{it}m_2(Z_{it}) - U_{i(t-1)}m_2(Z_{i(t-1)})$. Looking at these approximations we realize that we can estimate $\alpha'(z), \cdots, \alpha^{(p)}(z), m_1(z), m_1'(z), \cdots, m_1^{(p)}(z), m_2(z), m_2'(z), \cdots, m_2^{(p)}(z)$ by regressing ΔY_{it} on the terms of the right-hand side of these approximations with kernel weights. Clearly, $m(\cdot)$ is identified but $\alpha(\cdot)$ is not. This is due to the additive structure of the differencing procedure and it leads us to estimate these components by separate.

In this situation, the quantities of interest can be estimated using the following local constant regression, also denoted as a Nadaraya-Watson estimator,

$$\sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta Y_{it} - \Delta W_{it}^{\top} \beta_1 - \Delta U_{it}^{\top} \beta_2 \right)^2 K_H(Z_{it} - z) K_H(Z_{i(t-1)} - z), \tag{3.2}$$

where H is a $q \times q$ symmetric positive definite bandwidth matrix and K is a q-variate such that

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

see Nadaraya (1964), Watson (1964) and Fan and Gijbels (1995) for further details. Let $\widehat{\beta}_1$ and $\widehat{\beta}_2$ be the minimizers of (3.2). The above exposition suggests as estimator for $\beta_1 = m_1(z)$ and $\beta_2 = m_2(z)$, $\widehat{m}_1(z, z; H_2) = \widehat{\beta}_1$ and $\widehat{m}_2(z, z; H_2) = \widehat{\beta}_2$.

Unfortunately, despite the resulting estimator of (3.2) is robust to fixed effects (see Rodriguez-Poo and Soberon (2014, 2015) for further details), it is still biased due to endogeneity. When $E(v_{it}|\xi_{it}) \neq 0$ we cannot estimate consistently the unknown functions of the structural equations (3.1) 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. In order to solve this problem, we propose to use a $b \times 1$ vector of IV; i.e., $E(\Delta W_{it}|X_{it}, X_{i(t-1)}) = g(X_{it}, X_{i(t-1)}) = g_{it,i(t-1)}$.

To simplify notation, let us denote $\widetilde{W}_{it}^{\top} = (W_{it}^{\top} \quad U_{it}^{\top})$ and $m(Z_{it}) = \left(m_1(Z_{it})^{\top} \quad m_2(Z_{it})^{\top} \right)^{\top}$ as d-dimensional vectors, where $d = (M-1) + a$. Similar definitions are used for $\widetilde{W}_{i(t-1)}$ and $m(Z_{i(t-1)})$. Thus, rearranging terms, (3.1) can be written as

$$\begin{aligned} \Delta Y_{it} &= \alpha(Z_{it}, Z_{i(t-1)}) + \widetilde{W}_{it}^{\top} m(Z_{it}) - \widetilde{W}_{i(t-1)}^{\top} m(Z_{i(t-1)}) + \Delta v_{it} \\ \Delta W_{it} &= g(X_{it}, X_{i(t-1)}) + \Delta \xi_{it}. \end{aligned} \tag{3.3}$$

Let $\Delta \widetilde{W}_{g,it} = \left(g_{it,i(t-1)}^{\top} \quad \Delta U_{it}^{\top} \right)^{\top}$ be a $d \times 1$ vector. Multiplying both sides of (3.3) by $\Delta \widetilde{W}_{g,it}$ and taking conditional expectations on $(Z_{it} = z, Z_{i(t-1)} = z)$ if $E \left[\Delta \widetilde{W}_{g,it} \Delta \widetilde{W}_{it}^{\top} | Z_{it} = z, Z_{i(t-1)} = z \right]$ is positive definite we obtain

$$m(z) = E \left[\Delta \widetilde{W}_{g,it} \Delta \widetilde{W}_{it}^{\top} | Z_{it} = z, Z_{i(t-1)} = z \right]^{-1} E \left[\Delta \widetilde{W}_{g,it} \Delta Y_{it} | Z_{it} = z, Z_{i(t-1)} = z \right]. \tag{3.4}$$

Note that it is also necessary for identification that both $E[\Delta U_{it} \Delta U_{it}^{\top} | Z_{it} = z, Z_{i(t-1)} = z]$ and $E[g_{it,i(t-1)} g_{it,i(t-1)}^{\top} | Z_{it} = z, Z_{i(t-1)} = z]$ are positive definite. As the reader can realize, 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.

To obtain a feasible estimator for $m(z)$ we can replace $g_{it,i(t-1)}$ in (3.4) by a consistent estimator, i.e., $\widehat{g}(X_{it}, X_{i(t-1)}; H_1)$, where H_1 is a bandwidth matrix. This estimator can be any nonparametric smoother such as a local linear regression estimator, a spline smoothing or a sieve estimator. Later, in Section 3 we will discuss some minimal set of assumptions that this estimator must fulfill in order to guarantee that the resulting IV estimator exhibits the right asymptotic properties. For the sake of simplicity let us denote $\widehat{g}_{it,i(t-1)} = \widehat{g}(X_{it}, X_{i(t-1)}; H_1)$.

Let $\Delta \widetilde{W}_{\widehat{g},it} = \left(\widehat{g}_{it,i(t-1)}^\top \quad \Delta U_{it}^\top \right)^\top$ be a $d \times 1$ vector, the resulting local constant IV regression estimator is

$$
\widehat{m}_{\widehat{g}}(z, z; H_2) = \left(\begin{array}{c} \widehat{m}_{1\widehat{g}}(z, z; H_2) \\ \widehat{m}_{2\widehat{g}}(z, z; H_2) \end{array} \right) = \left(\sum_{i=1}^N \sum_{t=2}^T K_{H_2}(Z_{it} - z) K_{H_2}(Z_{i(t-1)} - z) \Delta \widetilde{W}_{\widehat{g},it} \Delta \widetilde{W}_{it}^\top \right)^{-1}
$$
$$
\times \sum_{i=1}^N \sum_{t=2}^T K_{H_2}(Z_{it} - z) K_{H_2}(Z_{i(t-1)} - z) \Delta \widetilde{W}_{\widehat{g},it} \Delta Y_{it}, \quad (3.5)
$$

where H_2 is the corresponding bandwidth matrix.

Despite the standard nonparametric regression techniques, this estimator exhibits the peculiarity that the kernel weights are related to both Z_{it} and $Z_{i(t-1)}$. If we would have 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 does not vanish asymptotically. Therefore, the asymptotic bias would also be non-negligible. This phenomena was already pointed out in Mundra (2005) and Lee and Mukherjee (2008), but it was solved in Rodriguez-Poo and Soberon (2014, 2015) in another context.

Once the estimators of the functional coefficients are proposed, let us now turn to the estimation of $\alpha(\cdot)$. To this end, define

$$
\Delta \widehat{Y}_{it} = \Delta Y_{it} - \Delta \widetilde{W}_{\widehat{g},it}^\top \widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}; H_2), \quad (3.6)
$$

where $\widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}; H_2)$ is the local constant IV regression estimator defined in (3.5). To simplify notation let us write $\widehat{m}_{\widehat{g}}(z, z; H_2) = \widehat{m}_{\widehat{g}}(z; H_2)$. If we substitute ΔY_{it} in (3.6) by (3.3) and take a local constant approximation around $m(\cdot)$, it is possible to show that after rearranging terms we get

$$
\Delta \widehat{Y}_{it} = \alpha(Z_{it}, Z_{i(t-1)}) + \Delta \widehat{v}_{it}, \qquad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \quad (3.7)
$$

where

$$
\Delta \widehat{v}_{it} = \Delta v_{it} - \left(\Delta \widetilde{W}_{\widehat{g},it} - \Delta \widetilde{W}_{it} \right)^\top \widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}; H_2) - \Delta \widetilde{W}_{it}^\top \left(\widehat{m}_{\widehat{g}}(Z_{it}, Z_{i(t-1)}; H_2) - m(Z_{it}, Z_{i(t-1)}) \right).
$$

In this fully nonparametric context, we propose to estimate $\alpha(\cdot)$ using marginal integration techniques. See Linton and Nielsen (1995) and Newey (1994). Recall that, since $\alpha(\cdot)$ is unknown, when we take first differences in (3.3) the object that can be identified is $\alpha(Z_{it}, Z_{i(t-1)})$, although we know

that $\alpha(Z_{it}, Z_{i(t-1)}) = \alpha(Z_{it}) - \alpha(Z_{i(t-1)})$. Then, for any $z_1, z_2 \in \mathcal{A}$ we can estimate $\alpha(\cdot)$ using a multivariate kernel regression estimation technique, i.e.

$$\widehat{\alpha}(z_1, z_2; H_3) = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_3}(Z_{it} - z_1) K_{H_3}(Z_{i(t-1)} - z_2) \Delta \widehat{Y}_{it}}{\sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_3}(Z_{it} - z_1) K_{H_3}(Z_{i(t-1)} - z_2)}, \tag{3.8}$$

where H_3 is the bandwidth. Once obtained $\widehat{\alpha}(z_1, z_2; H_3)$, as in Qian and Wang (2012), we proceed to estimate $\widehat{\alpha}(z; H_3)$ by marginally integrating (3.8) obtaining

$$\widehat{\alpha}(z; H_3) = \int_{\mathbb{R}} \widehat{\alpha}(z, s; H_3) q(s) ds, \tag{3.9}$$

where $q(\cdot)$ is a pre-specified positive weighting function that satisfies $\int_{\mathbb{R}} q(s) ds = 1$. For model identification, following also Qian and Wang (2012) we assume that $\int_{\mathbb{R}} \alpha(z) q(z) dz = 0$. It is clear that under these conditions $\alpha(z)$ is uniquely identified. i.e.

$$\int_{\mathbb{R}} \alpha(z, s) q(s) ds = \alpha(z) \int_{\mathbb{R}} q(s) ds - \int_{\mathbb{R}} \alpha(s) q(s) ds = \alpha(z).$$

Then, if NT is large enough and $q(\cdot)$ is chosen as the density function of Z_{it} we can use the sample version of (3.9) and propose the following estimator

$$\widehat{\alpha}(z; H_3) = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} \widehat{\alpha}(z, Z_{i(t-1)}). \tag{3.10}$$

4 Statistical 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 4.1 *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.*

ASSUMPTION 4.2 *Let $f_{X_{1t}}(\cdot)$ and $f_{Z_{1t}}(\cdot)$ be the probability density functions of X_{it} and Z_{it}, respectively. Moreover, let $f_{X_{1t}, X_{1(t-1)}}(\cdot, \cdot)$ and $f_{Z_{1t}, Z_{1(t-1)}}(\cdot, \cdot)$ be the probability density functions of $(X_{1t}, X_{1(t-1)})$ and $(Z_{1t}, Z_{1(t-1)})$, 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 4.3 *The random errors, v_{it} and ξ_{it}, are independent and identically distributed, with zero mean and homoscedastic variances, $\sigma_v^2 < \infty$ and $\sigma_\xi^2 < \infty$. They are also independent of $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$ for all i and t, but $E(v_{it}\xi_{it}) \neq 0$. Furthermore, $E|v_{it}|^{2+\delta} < \infty$ and $E|\xi_{it}|^{2+\delta} < \infty$, for some $\delta > 0$.*

ASSUMPTION 4.4 *Let z be an interior point in the support of $f_{Z_{1t}}(\cdot)$. All second-order derivatives of $\alpha(\cdot)$ and $m_1(\cdot), \cdots, m_{d-1}(\cdot)$ are bounded and uniformly continuous and they satisfy a Lipschitz condition. Also, let (x_1, x_2) be interior points in the support of $f_{X_{1t}, X_{1(t-1)}}(\cdot, \cdot)$, all second-order derivatives of $g_1(\cdot, \cdot), \cdots, g_{M-1}(\cdot, \cdot)$ are bounded and continuous.*

ASSUMPTION 4.5 *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 $N \to \infty$, $N|H_1| \to \infty$, $N|H_2| \to \infty$, $N|H_1|/log(N) \to \infty$ and $N|H_2|/log(N) \to \infty$. Furthermore, $tr(H_1) = o_p(tr(H_2))$.*

ASSUMPTION 4.6 *Let $\|A\| = \sqrt{tr(A^\top A)}$, then $E\left[\|\widetilde{W}_{g,it}\widetilde{W}_{g,it}^\top\|^2|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ is bounded and uniformly continuous in its support. Also, let*

$$\mathcal{X}_{g,it} = \left(\widetilde{W}_{g,it}^\top, \quad \widetilde{W}_{g,i(t-1)}^\top\right)^\top \qquad and \qquad \Delta\mathcal{X}_{g,it} = \left(\Delta\widetilde{W}_{g,it}^\top \quad \Delta\widetilde{W}_{g,i(t-1)}^\top\right)^\top,$$

whereas $\mathcal{X}_{it}$ and $\Delta\mathcal{X}_{it}$ are defined similarly without the index g. Furthermore, the following matrix functions $E[\Delta\mathcal{X}_{g,it}|z_1, z_2]$, $E[\Delta\mathcal{X}_{g,it}\mathcal{X}_{it}^\top|z_1, z_2]$, $E\left[\Delta\mathcal{X}_{g,it}\Delta\mathcal{X}_{g,it}^\top|z_1, z_2\right]$, $E\left[\Delta\mathcal{X}_{g,it}\Delta\mathcal{X}_{it}^\top|z_1, z_2\right]$, $E[\Delta\mathcal{X}_{g,it}|z_1, z_2, z_3]$, $E\left[\Delta\mathcal{X}_{g,it}\mathcal{X}_{it}^\top|z_1, z_2, z_3\right]$, $E\left[\Delta\mathcal{X}_{g,it}\Delta\mathcal{X}_{g,it}^\top|z_1, z_2, z_3\right]$ and $E\left[\Delta\mathcal{X}_{g,it}\Delta\mathcal{X}_{it}^\top|z_1, z_2, z_3\right]$ are bounded and uniformly continuous in their support.

ASSUMPTION 4.7 *The kernel function K is the product of 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 4.8 *The following moment functions $E\left[\Delta U_{it}\Delta U_{it}^\top|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ and $E\left[g_{it,i(t-1)}g_{it,i(t-1)}^\top|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ are positive definite in any interior point (z_1, z_2) in the support of $f_{Z_{it}, Z_{i(t-1)}}(z_1, z_2)$.*

ASSUMPTION 4.9 *For some $\delta > 0$, the functions $E\left[|\Delta\widetilde{W}_{g,it}v_{it}|^{2+\delta}|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$, $E\left[|\Delta\widetilde{W}_{g,it}\xi_{it}^\top|^{2+\delta}|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ and $E\left[|\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^\top v_{it}\xi_{it}|^{1+\delta/2}|Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ are bounded and uniformly continuous in any point of their support. These results hold for $\xi_{i(t-1)}$ and $v_{i(t-1)}$.*

Assumption **4.1** is standard in the nonparametric panel data regression analysis and characterizes the data generating process. In particular, it states that the individuals are independent and, for a fixed individual, we allow for correlation along time. Other time-series structures can also be considered; see, e.g., Cai and Li (2008) or Cai et al. (2009). Also, for the estimation of the fully

nonlinear part in the one-step backfitting algorithm we need to impose some further assumptions about the density functions than the usual Lipschitz continuity. Thus, Assumption **4.2** 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 $f_{X_{1t}, X_{1(t-1)}, X_{1(t-2)}}(\cdot, \cdot, \cdot)$ and $f_{Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)}}(\cdot, \cdot, \cdot)$, being the probability density functions of $(X_{it}, X_{i(t-1)}, X_{i(t-2)})$, and $(Z_{it}, Z_{i(t-1)}, Z_{i(t-2)})$, respectively. Assumption **4.3** 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 **4.4**-**4.7** 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 **4.5** contains a standard bandwidth condition for smoothing techniques and some uniform converge results. Also, with $\text{tr}(H_1) = o_p(\text{tr}(H_2))$ we assume that $\widehat{m}_{\widehat{g}}(z; H_2)$ is not sensitive to the choice of H_1 and also we impose that the bandwidth H_1 should be chosen small enough or, at least, smaller than H_2. Thus, the fitted model in the first-stage is under-smoothed and its bias is not too large. Therefore, the conditions in **4.5** are enough to show the point-wise consistency of the local constant IV estimator, see Cai (2002a,b) for further details.

Furthermore, by the smoothness and boundedness conditions established in Assumptions **4.2** and **4.4**-**4.8** for the kernel function, conditional moments and densities we are allowed to claim the uniform convergence results established in Masry (1996, Theorem 6). Assumption **4.8** is a generalization of the usual rank condition for identification of simultaneous equation systems in a parametric context and it implies that $E\left[\Delta \widetilde{W}_{g,it} \Delta \widetilde{W}_{it}^{\top} | Z_{it} = z, Z_{i(t-1)} = z\right]$ is definite positive. Assumption **4.9** is required to show that the Lyapunov condition holds.

Under these assumptions we can establish the asymptotic normality of the local constant IV estimator $\widehat{m}_{\widehat{g}}(z; H_2)$ for the standard case in which $\mu_2(K_u) = \mu_2(K_v)$. The proof is relegated to the Appendix.

THEOREM 4.1 *Under Assumptions 4.1-4.9, as N tends to infinity and T is fixed*

$$\sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z; H_2) - m(z) - B(z; H_2)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, V(z; H_2)\right),$$

where

$$
\begin{aligned}
B(z; H_2) &= \mu_2(K)\left[diag_d\left(D_f(z) H_2 \sqrt{N|H_2|} D_{m_r}(z)\right) \imath_d f_{Z_{it}, Z_{i(t-1)}}^{-1}(z, z)\right. \\
&\left. + \frac{1}{2} diag_d\left(tr\left(\mathcal{H}_{m_r}(z) H_2 \sqrt{N|H_2|}\right)\right) \imath_d \right],
\end{aligned}
$$

$$
\begin{aligned}
V(z; H_2) &= 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^{\top} \Sigma_{\Delta\xi\Delta\xi} m_1(z) + \Sigma_{\Delta v \Delta\xi} m_1(z)\right) \\
&\times \mathcal{B}_{\Delta\widetilde{W}_g \Delta\widetilde{W}}^{-1}(z, z) \mathcal{B}_{\Delta\widetilde{W}_g \Delta\widetilde{W}_g}(z, z) \mathcal{B}_{\Delta\widetilde{W}_g \Delta\widetilde{W}}^{-1}(z, z),
\end{aligned}
$$

$\Sigma_{\Delta\xi\Delta\xi} = E(\Delta\xi_{it}\Delta\xi_{it}^\top)$ is a $(M-1) \times (M-1)$ matrix and $\Sigma_{\Delta v\Delta\xi} = E(\Delta v_{it}\Delta\xi_{it}^\top)$ is a vector of dimension $1 \times (M-1)$. Moreover, for $r = 1, \cdots, d$, $D_{m_r}(z)$ is the first order derivative vector of the rth component of $m(\cdot)$, $\mathcal{H}_{m_r}(z)$ its Hessian matrix, $D_f(z)$ the first order derivative vector of the density function and

$$\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z) = E\left[\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z),$$

$$\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z) = E\left[\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top | Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z).$$

Furthermore, $diag_d\left(tr(\mathcal{H}_{m_r}(z)H_2)\right)$ and $diag_d(D_f(z)H_2D_{m_r}(z))$ stand for a diagonal matrix of elements of $tr(\mathcal{H}_{m_r}(z)H_2)$ and $D_f(z)H_2D_{m_r}(z)$, respectively, being $\imath_d$ a $d \times 1$ unitary vector.

The previous results exhibit some remarkable differences with respect to standard local constant estimators (i.e., Nadaraya-Watson estimator). Meanwhile the asymptotic bias exhibits the same expression as the corresponding for the Nadaraya-Watson, the order of the variance is different. More precisely, the bias term depends mainly on the smoothness of $m(\cdot)$ but, as it could be expected, it is also going to depend on the smoothness of $g(\cdot)$. However, by introducing a condition that relates the bandwidth matrices, i.e. $tr(H_2) = o_p(tr(H_1))$, the dependence on $g(\cdot)$ becomes asymptotically negligible.

With respect to the variance term, we point out two relevant issues. On one hand, the variance is composed by three elements: the first one related to the variation of the measurement error of the structural equation, the second one addresses for the variability of the estimated reduced form, and the last one accounts for the covariance among the measurement error of all equations in the system. On the other hand, and most important, the rate of convergence of the variance is suboptimal. In this smoothness class the lower rate of convergence for this type of estimators is $N|H_2|^{1/2}$. For this reason, in the next section we will propose a one-step backfitting algorithm that will make the rate of convergence of our estimator optimal.

Focus now on the study of the asymptotic behavior of the marginal integration estimator $\widehat{\alpha}(z_1; H_3)$ we need the following additional assumptions.

ASSUMPTION 4.10 *Let $f_Z(u)$ be the marginal density of Z, that is twice continuously differentiable. Furthermore, let $q(\cdot)$ be a positive weighting function defined on the compact support of $f_Z(u)$. It holds that*

$$\int q(u)du = 1 \quad ; \quad \int \alpha(u)q(u)du = 0.$$

ASSUMPTION 4.11 *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 $N \to \infty$, $N|H_3| \to \infty$, $N|H_3|/log(N) \to \infty$.*

As it was already explained in the previous section, Assumption **4.10** is a standard condition in this literature to identify $\alpha(z)$ up to a multiplicative constant. Assumption **4.11** contains necessary

conditions to show the point-wise consistency of this marginal integration estimator under uniform convergence results. Thus, for the standard case where $\mu_2(K_u) = \mu_2(K_v)$ we obtain the following result,

LEMMA 4.1 *Under Assumptions 4.1-4.5, 4.7 and 4.10-4.11, as N tends to infinity and T is fixed*

$$\sqrt{N|H_3|^{1/2}}\left(\widehat{\alpha}(z; H_3) - \alpha(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(B(z; H_3), V(z; H_3)\right),$$

where

$$
\begin{aligned}
B(z; H_3) &= \mu_2(K)\left[\frac{1}{2}tr(\mathcal{H}_\alpha(z)H_3) - \frac{1}{2}\int tr(\mathcal{H}_\alpha(s)H_3)q_{Z_{i(t-1)}}(s)ds \right. \\
&\quad \left. + D_\alpha(z)H_3D_f(z)\int \frac{q_{Z_{i(t-1)}}(s)}{f_{Z_{it}, Z_{i(t-1)}}(z, s)}ds - \int D_\alpha(s)H_3D_f(s)\frac{q_{Z_{i(t-1)}}(s)}{f_{Z_{it}, Z_{i(t-1)}}(z, s)}ds\right], \\
V(z; H_3) &= 2\sigma_v^2 R(K_u)R(K_v)\int \frac{q_{Z_{it, i(t-1)}}^2(s)}{f_{Z_{it}, Z_{i(t-1)}}(z, s)}ds.
\end{aligned}
$$

The proof of this result follows exactly the same lines as in Qian and Wang (2012) and we omit it to avoid redundances. Nevertheless, as it is noted also in Qian and Wang (2012) when Z_{it} is accurately predictable by $Z_{i(t-1)}$ the joint distribution function $f_{Z_{it}, Z_{i(t-1)}}(z_1, z_2)$ is close to zero and it increases the asymptotic variance. Also, this estimator can be improved upon since the structure of the covariance of the error term is ignored.

5 More efficient estimators

5.1 One-step backfitting and minimum distance estimators

In this section we first propose a one-step backfitting algorithm that will enable us to achieve optimal nonparametric rates of convergence for the estimators of $m(\cdot)$. Note that the estimator already proposed in Section 3 for $m(\cdot)$ was suboptimal (see Theorem 4.1). It turns out that, as it will be detailed further in the section, the backfitting procedure generates two alternative estimators for $m(\cdot)$. That is why, further in this section, we propose a minimum distance estimator to obtain an unique more efficient solution.

Starting with the one-step backfitting technique let us define $\Delta\widetilde{Y}_{1it}$ as

$$\Delta\widetilde{Y}_{1it} = \Delta Y_{it} - \alpha(Z_{it}) + \alpha(Z_{i(t-1)}) + \widetilde{W}_{i(t-1)}^\top m(Z_{i(t-1)}), \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \quad (5.1)$$

and substituting the structural equation of (3.3) into (5.1) we obtain

$$\Delta\widetilde{Y}_{1it} = \widetilde{W}_{it}^\top m(Z_{it}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \quad (5.2)$$

However, despite (5.2) exhibits a low dimensional problem where $m(\cdot)$ could be estimated using kernel weights that are related only to Z_{it}; the quantities $m(Z_{i(t-1)})$, $\alpha(Z_{it})$ and $\alpha(Z_{i(t-1)})$ are not

observed. Thus, any standard nonparametric procedure generates unfeasible estimators. To overcome this situation, we propose to replace in (5.1) these unknown functions by their corresponding estimators in (3.5) and (3.10), i.e., $\widehat{m}_{\widehat{g}}(z; H_2)$, $\widehat{\alpha}(z_1; H_3)$ and $\widehat{\alpha}(z_2; H_3)$.

Let $\Delta \widehat{Y}_{1it} = \Delta Y_{it} - \widehat{\alpha}(Z_{it}; H_3) + \widehat{\alpha}(Z_{i(t-1)}; H_3) + \widetilde{W}_{i(t-1)}^{\top} \widehat{m}_{\widehat{g}}(Z_{i(t-1)}, H_2)$. The regression problem turns out to be

$$\Delta \widehat{Y}_{1it} = \widetilde{W}_{it}^{\top} m(Z_{it}) + \Delta \widehat{v}_{1it}, \tag{5.3}$$

where the composite error term is of the form

$$\Delta \widehat{v}_{1it} = \Delta v_{it} + \widetilde{W}_{i(t-1)}^{\top} \left(\widehat{m}_{\widehat{g}}(Z_{i(t-1)}; H_2) - m(Z_{i(t-1)})\right) - \left(\widehat{\alpha}(Z_{it}; H_3) - \alpha(Z_{it})\right) + \left(\widehat{\alpha}(Z_{i(t-1)}; H_3) - \alpha(Z_{i(t-1)})\right).$$

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

$$\sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta \widehat{Y}_{1it} - \widetilde{W}_{it}^{\top} \gamma_1\right)^2 K_{H_4}(Z_{it} - z), \tag{5.4}$$

where H_4 is the bandwidth matrix at this stage. We denote by $\widehat{\gamma}_1$ the minimizer of (5.4).

However, the resulting estimator of (5.4) is also biased due to the endogeneity problem and, in order to solve it, we resort again to the IV method. Let $\widetilde{W}_{\widehat{g},it} = (\widehat{g}_{it,i(t-1)}^{\top} \quad U_{it}^{\top})$ be a $d \times 1$ vector of IV, the one-step IV backfitting estimator, $\widehat{m}_{\widehat{g}}^{(1)}(z; H_4)$, of $m(\cdot)$ has the following closed form

$$\widehat{m}_{\widehat{g}}^{(1)}(z; H_4) = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_4}(Z_{it} - z)\widetilde{W}_{\widehat{g},it}\widetilde{W}_{it}^{\top}\right)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_4}(Z_{it} - z)\widetilde{W}_{\widehat{g},it}\Delta \widehat{Y}_{1it}. \tag{5.5}$$

Before analyzing the main statistical properties of this estimator note that this backfitting procedure provides two different estimators for the same function $m(\cdot)$. The second estimator will be obtained by defining $\Delta \widehat{Y}_{2it} = \Delta Y_{it} - \widehat{\alpha}(Z_{it}; H_3) + \widehat{\alpha}(Z_{i(t-1)}; H_3) - \widetilde{W}_{it}^{\top} \widehat{m}_{\widehat{g}}(Z_{it}; H_2)$. Then, if and substitute (3.3) into $\Delta \widehat{Y}_{2it}$ the regression becomes

$$-\Delta \widehat{Y}_{2it} = \widetilde{W}_{i(t-1)}^{\top} m(Z_{i(t-1)}) + \Delta \widehat{v}_{2it}, \tag{5.6}$$

where the composite error term is

$$\Delta \widehat{v}_{2it} = \widetilde{W}_{it}^{\top} \left(\widehat{m}_{\widehat{g}}(Z_{it}; H_2) - m(Z_{it})\right) + \left(\widehat{\alpha}(Z_{it}; H_3) - \alpha(Z_{it})\right) - \left(\widehat{\alpha}(Z_{i(t-1)}; H_3) - \alpha(Z_{i(t-1)})\right) - \Delta v_{it}.$$

Following the same procedure as before and letting $\widetilde{W}_{\widehat{g},i(t-1)} = \left(\widehat{g}_{it,i(t-1)}^{\top} \quad U_{i(t-1)}^{\top}\right)$ a $d \times 1$ vector of IV, the corresponding one-step IV backfitting estimator is

$$\widehat{m}_{\widehat{g}}^{(2)}(z; H_4) = -\left(\sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_4}(Z_{i(t-1)} - z)\widetilde{W}_{\widehat{g},i(t-1)}\widetilde{W}_{i(t-1)}^{\top}\right)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} K_{H_4}(Z_{i(t-1)} - z)\widetilde{W}_{\widehat{g},i(t-1)}\Delta \widehat{Y}_{2it}. \tag{5.7}$$

Note that we have available two different estimators, $\widehat{m}_{\widehat{g}}^{(1)}(z; H_4)$ and $\widehat{m}_{\widehat{g}}^{(2)}(z; H_4)$, for $m(z)$. A natural idea to combine both in an efficient way would be to obtain such an estimator by minimizing the following criterion function

$$\left(\left(\begin{array}{c} \widehat{m}_{\widehat{g}}^{(1)}(z; H_4) \\ \widehat{m}_{\widehat{g}}^{(2)}(z; H_4) \end{array} \right) - \left(\begin{array}{c} m(z) \\ m(z) \end{array} \right) \right)^{\top} \mathcal{W}_m^{-1} \left(\left(\begin{array}{c} \widehat{m}_{\widehat{g}}^{(1)}(z; H_4) \\ \widehat{m}_{\widehat{g}}^{(2)}(z; H_4) \end{array} \right) - \left(\begin{array}{c} m(z) \\ m(z) \end{array} \right) \right). \tag{5.8}$$

As a weighting matrix, $\mathcal{W}_m$, we propose the variance-covariance matrix of

$$\left((\widehat{m}_{\widehat{g}}^{(1)}(z; H_4) - m(z))^{\top} \quad (\widehat{m}_{\widehat{g}}^{(2)}(z; H_4) - m(z))^{\top} \right)^{\top},$$

i.e. Ω_m. Let $\widetilde{m}(z; H_4)$ be the minimizer of (5.8). It is easy to verify that

$$\widetilde{m}(z; H_4) = \left(\Omega_{m_{11}}^{(1)} + 2\Omega_{m_{21}}^{(1)} + \Omega_{m_{22}}^{(1)} \right)^{-1} \left(\left(\Omega_{m_{11}}^{(1)} + \Omega_{m_{21}}^{(1)} \right) \widehat{m}_{\widehat{g}}^{(1)}(z; H_4) + \left(\Omega_{m_{12}}^{(1)} + \Omega_{m_{22}}^{(1)} \right) \widehat{m}_{\widehat{g}}^{(2)}(z; H_4) \right), \tag{5.9}$$

where the $d \times d$ matrix $\Omega_{m_{ij}}^{(1)}$ is the (i, j)-th component of the following block partitioned matrix

$$\Omega_m^{-1} = \left(\begin{array}{cc} \Omega_{m_{11}}^{(1)} & \Omega_{m_{12}}^{(1)} \\ \Omega_{m_{21}}^{(1)} & \Omega_{m_{22}}^{(1)} \end{array} \right).$$

We now proceed to analyze the asymptotic properties of both the backfitting and the minimum distance estimator.

5.2 Asymptotic properties

In order to show that the resulting one-step backfitting estimator achieves optimal rates of convergence i.e. $1/N|H_4|^{1/2}$, we need Assumptions **4.1**-**4.3** and the smoothness and boundedness conditions already established in Assumptions **4.4**-**4.7** and **4.10**-**4.11**. Furthermore, since we use the estimators obtained in Section 3, to cancel asymptotically the additive terms in (3.1), we need to ensure that the bias rate of these estimators, i.e. $\widehat{m}_{\widehat{g}}(z; H_2)$ and $\widehat{\alpha}(z; H_3)$, is uniform. For this, following Masry (1996) we impose some additional assumptions about the bandwidth H_4 and its relationship with the bandwidths of the previous steps.

ASSUMPTION 5.1 *The bandwidth matrix H_4 is symmetric and strictly positive definite. Also, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|H_4| \to \infty$.*

ASSUMPTION 5.2 *The bandwidth matrices H_2, H_3 and H_4 must fulfill that $N|H_2||H_4|/log(N) \to \infty$ and $N|H_3||H_4|/log(N) \to \infty$, whereas $tr(H_2)/tr(H_4) \to 0$ and $tr(H_3)/tr(H_4) \to 0$ as $N \to \infty$.*

Under these assumptions we get the following asymptotic distribution of $\widehat{m}_{\widehat{g}}^{(j)}(z; H_4)$, for $j = 1, 2$.

16

THEOREM 5.1 *Assume conditions 4.1-4.8, 4.11 and 5.1-5.2 hold, then as $N \to \infty$ and T is fixed we obtain*

$$\sqrt{N|H_4|^{1/2}} \left(\widehat{m}_{\widehat{g}}^{(j)}(z; H_4) - m(z) - B(z; H_4) \right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, V^{(j)}(z; H_4) \right),$$

where for $j = 1$ we get

$$B(z; H_4) = \mu_2(K) \left(diag_d \left(D_f(z) H_4 \sqrt{N|H_4|^{1/2}} D_{m_r}(z) \right) \imath_d f_{Z_{it}}(z)^{-1} + \frac{1}{2} diag_d \left(tr \left(\mathcal{H}_{m_r}(z) H_4 \sqrt{N|H_4|^{1/2}} \right) \right) \imath_d \right),$$

$$V^{(1)}(z; H_4) = 2R(K_u) \left(\sigma_v^2 + m_1(z)^\top \Sigma_{\Delta\xi\Delta\xi} m_1(z) + \Sigma_{\Delta v\Delta\xi} m_1(z) \right) \mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z) \mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z) \mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z),$$

whereas for $j = 2$

$$V^{(2)}(z; H_4) = 2R(K_u) \left(\sigma_v^2 + m_1(z)^\top \Sigma_{\Delta\xi\Delta\xi} m_1(z) + \Sigma_{\Delta v\Delta\xi} m_1(z) \right) \mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}^{-1}(z) \mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{g-1}}(z) \mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}^{-1}(z),$$

and

$$\mathcal{B}_{\widetilde{W}_g\widetilde{W}}(z) = E\left[\widetilde{W}_{g,it} \widetilde{W}_{it}^\top | Z_{it} = z \right] f_{Z_{it}}(z), \quad \mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z) = E\left[\widetilde{W}_{g,it} \widetilde{W}_{g,it}^\top | Z_{it} = z \right] f_{Z_{it}}(z),$$

$$\mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}(z) = E\left[\widetilde{W}_{g,i(t-1)} \widetilde{W}_{i(t-1)}^\top | Z_{i(t-1)} = z \right] f_{Z_{i(t-1)}}(z),$$

$$\mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{g-1}}(z) = E\left[\widetilde{W}_{g,i(t-1)} \widetilde{W}_{g,i(t-1)}^\top | Z_{i(t-1)} = z \right] f_{Z_{i(t-1)}}(z).$$

The proof of this result is done in the Appendix.

We focus now on the asymptotic properties of the minimum distance estimator, $\widetilde{m}(z; H_4)$, getting the following result,

THEOREM 5.2 *Assume conditions 4.1-4.8 and 5.1-5.2 hold, then as $N \to \infty$ and T is fixed*

$$\sqrt{N|H_4|^{1/2}} \left(\widetilde{m}(z; H_4) - m(z) - B(z; H_4) \right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \left(V^{(1)}(z; H_4)^{-1} + V^{(2)}(z; H_4)^{-1} \right)^{-1} \right).$$

The proof of this result is relegated to the Appendix.

Finally, focusing on Theorem 5.2 it can be highlighted that the asymptotic biases of the minimum distance estimators are similar as the previous estimators. Nevertheless, the asymptotic variance is the sum of the variance of the combined estimators and both exhibit the optimal rate of convergence of this type of problems. Therefore, it is proved that this technique enables to obtain more efficient estimators for $m(\cdot)$ and, at the same time, to achieve optimality.

6 Empirical results and Monte Carlo simulations

In order to determine the varying effects of unexpected health-care expenditures and households risk aversion on their savings over the life cycle, in this section we first focus on the optimal consumption

decision problem of the Spanish households. Later, we corroborate the results obtained via a Monte Carlo experiment.

Although the theory of precautionary savings states that uncertainty has a negative impact on household consumption and a positive effect on their savings, several empirical studies attempted to establish this relationship but without achieving conclusive results; see Carroll and Samwick (1997). Nevertheless, because the optimal consumption choice depends on life-time resources, 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 household.

In this context and with the aim of showing the feasibility and possible gains of the method proposed in the previous sections, in the following we consider a simulated example and analyze a stochastic dynamic model of precautionary saving based on the LCH model of Modigliani and Brumberg (1954). Then, using a random sample of Spanish households, we are able to establish to what extend the precautionary behavior of the households affects to their consumption/saving decisions, without imposing restrictive assumptions on either the functional forms or the unobserved individual heterogeneity, something that to our knowledge is completely new.

6.1   Data

The data used in this analysis are obtained from the ECPF elaborated by the INE for the period 1985(I)-1996(IV) where each household is interviewed for eight consecutive quarters. Traditionally, the household precautionary behavior to unexpected changes in income has been measured via spending changes in non-durable goods. However, some authors, as in Aaronson et al. (2012), have shown that spending in durable goods are more sensitive to these shocks than spending in non-durable goods. 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). 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. Thus, each definition is the result of reducing households 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 those households which answer to the eight quarters and provide full information about their incomes and expenses. For sample size reasons and to avoid having to specify an inheritance function, households whose head is aged under 26 or over 65 years old are excluded.

As the reader can see in Table 1, in this sample there is a large proportion of households with negative savings; i.e., 60.36% of the entire sample. Then, we must be specially careful when we define the saving variables. In Chou et al. (2004) 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 invalidate our conclusions. To overcome it and 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.

Table 1. Distribution of households with negative savings by population groups.

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%

Note: 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.

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 expenditure 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 household disposable income and households expenditures of the sample is collected in Table 2.

Analyzing the figures in Table 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 about 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 3, we collect the distribution of household income and expenditures by educational level.

Looking at the figures in Table 3 we can state that, as expected, both the revenue and expenditure from the group with a high level of education are larger. Therefore, in the next subsection we first estimate the model specified previously without considering the educational level. Next, we reestimate the model considering different education levels so we can analyze the heterogeneity between these population groups.

Table 2. Distribution of the household disposable income and household 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.

Table 3. Distribution of household 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 al least a high-school diploma, whereas household with low educational level are those whose head is illiterate, have no education or first degree studies.

6.2 Empirical results

The system of equations that we estimate is

$$
\begin{aligned}
Y_{it} &= \alpha(Z_{it}) + W_{it} m_1(Z_{it}) + U_{it} m_2(Z_{it}) + \mu_i + v_{it}, \\
W_{it} &= g(X_{it}) + \zeta_i + \xi_{it},
\end{aligned}
\qquad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \quad (6.1)
$$

where i index the household, t the time, Z_{it} is the age of the household head, W_{it} the health-care expenditures (log), Y_{it} the savings, U_{it} the permanent income (log), X_{it} a vector that contains the age of the householdhead of the i-th household and Z_{it} the number of children under 14 years old.

Note that household 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 to the productivity rate of growth and 65 years old is the maximum age at which people works, the permanent earnings at age τ_0 can be calculated as

$$
Y(\tau_0) = X^\top \beta + (65 - \tau_0 + 1)^{-1} \sum_{\tau = \tau_0}^{65} f(\tau),
$$

where $f(\tau)$ is the estimated quadratic function of age, Y_{it} the household income and X_{it} a vector of demographic characteristics.

As we state in previous sections, bandwidth selection is an important issue. There are many standard procedures in the literature for optimal bandwidth selection but for methodological simplicity the

bandwidth H_4 is chosen according to Silverman's rule-of-thumb; i.e., $\widehat{H}_4 = \widehat{h}_4 I = \widehat{\sigma}_z n^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of Z. Also, remember that in order to obtain the desirable properties of the proposed estimators it is necessary that the biases of $\widehat{g}_{it,i(t-1)}$, $\widehat{m}_{\widehat{g}}(z; H_2)$ and $\widehat{\alpha}(z_1; H_3)$ are not too large. Therefore, H_1, H_2 and H_3 must be chosen as smallest as possible or, at least, smaller than H_4.

Estimation results are shown in Figures 1-3. The estimated curves are plotted against the age variable jointly with 95% pointwise confidence intervals calculated adapting the wild bootstrap technique of Härdle et al. (2004) to this context. Figure 1 shows the results for the sample without considering educational levels. On the other side, in Figures 2 and 3 we show the estimation results distinguishing between those who have lower educational level (Figure 2) and those who have higher education level (Figure 3).

Figures 1-3 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., $\widetilde{\alpha}(\cdot)$. Panels B exhibit the corresponding elasticity to changes in health-care expenditures, i.e., $\widetilde{m}_1(\cdot)$, whereas Panels C show the precautionary savings elasticity to changes in household income; i.e., $\widetilde{m}_2(\cdot)$. In addition, 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, whereas Panel A-3 compares the estimated curves when endogeneity is not considered. This structure is maintained for Panels B and C.

Focusing on the results in Figure 1, we can note that when we control for household risk aversion (Panels A) the saving rate is hump shaped. Younger households (26-37) behave as bugger-stock agents, when savings exhibit a downward path. Meanwhile, when we control for income uncertainty (Panels C) savings decrease from age 26 to age 34, when savings increase significantly corroborating thus 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 (26-33) exhibit a declining savings rate, following by a constant path till the age of 40, where the hump-shaped appears again.

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 whereas in Spain at 45 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 elasticities for the different savings results, we can note that consumption of durable goods react more to unexpected changes in income, whereas consumptions on non-durable goods is more sensitive to both household risk aversion and potential health-care payout. This holds specially for households over 45 years old.

Finally, in order to evaluate the empirical relevance of the endogeneity problem we compare the results of our technique (lines grey and black) against those obtained without considering endogeneity (lines grey and black); see Panels 3 of Figure 1. By looking at these results, there are some significant differences: when we control for uncertainty about health-care expenditures households accumulate assets in the middle of their life, whereas when endogeneity is not taken into account

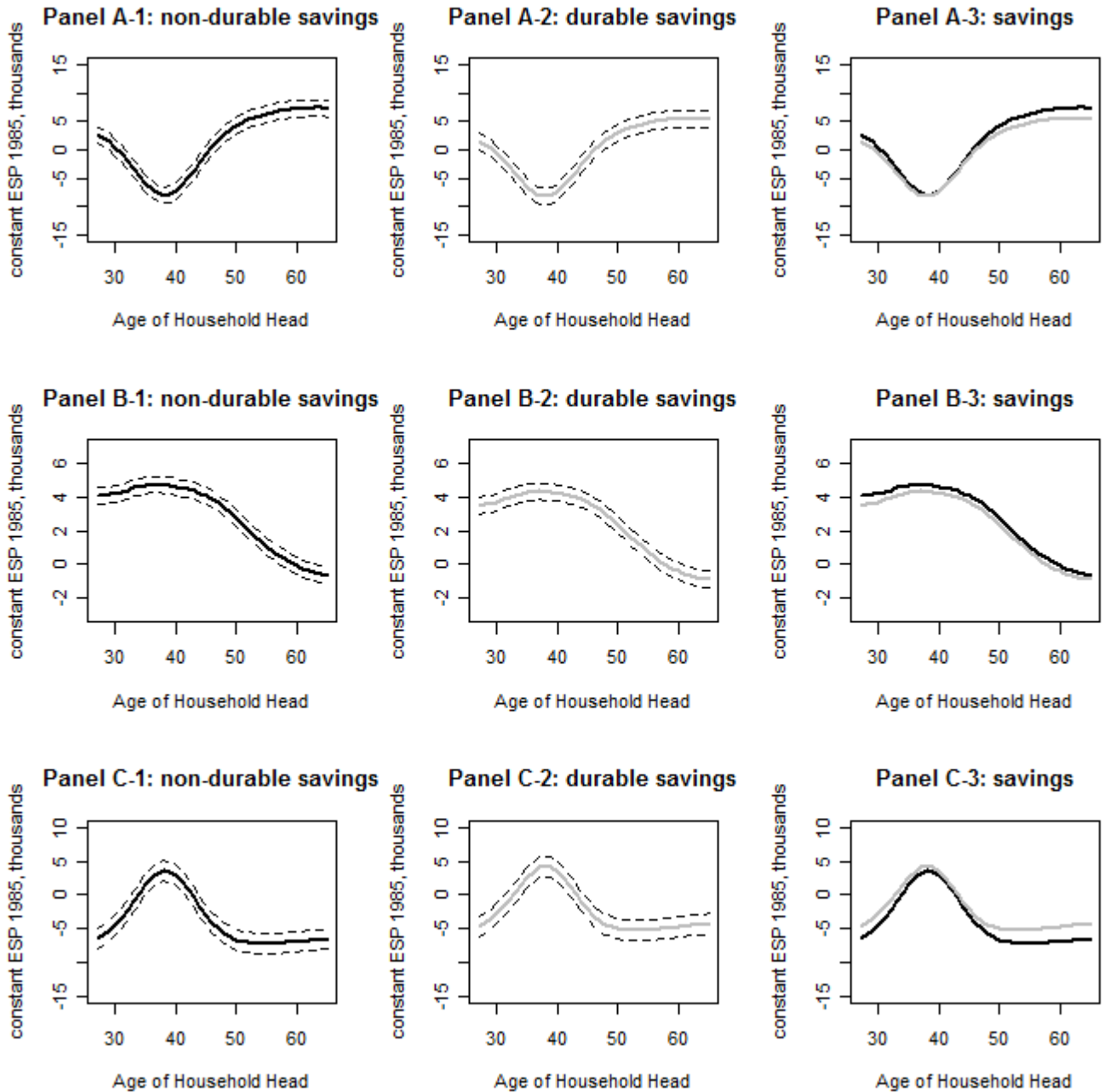there is a more or less constant path over the life cycle.

Figure 1. Household savings over the life-cycle
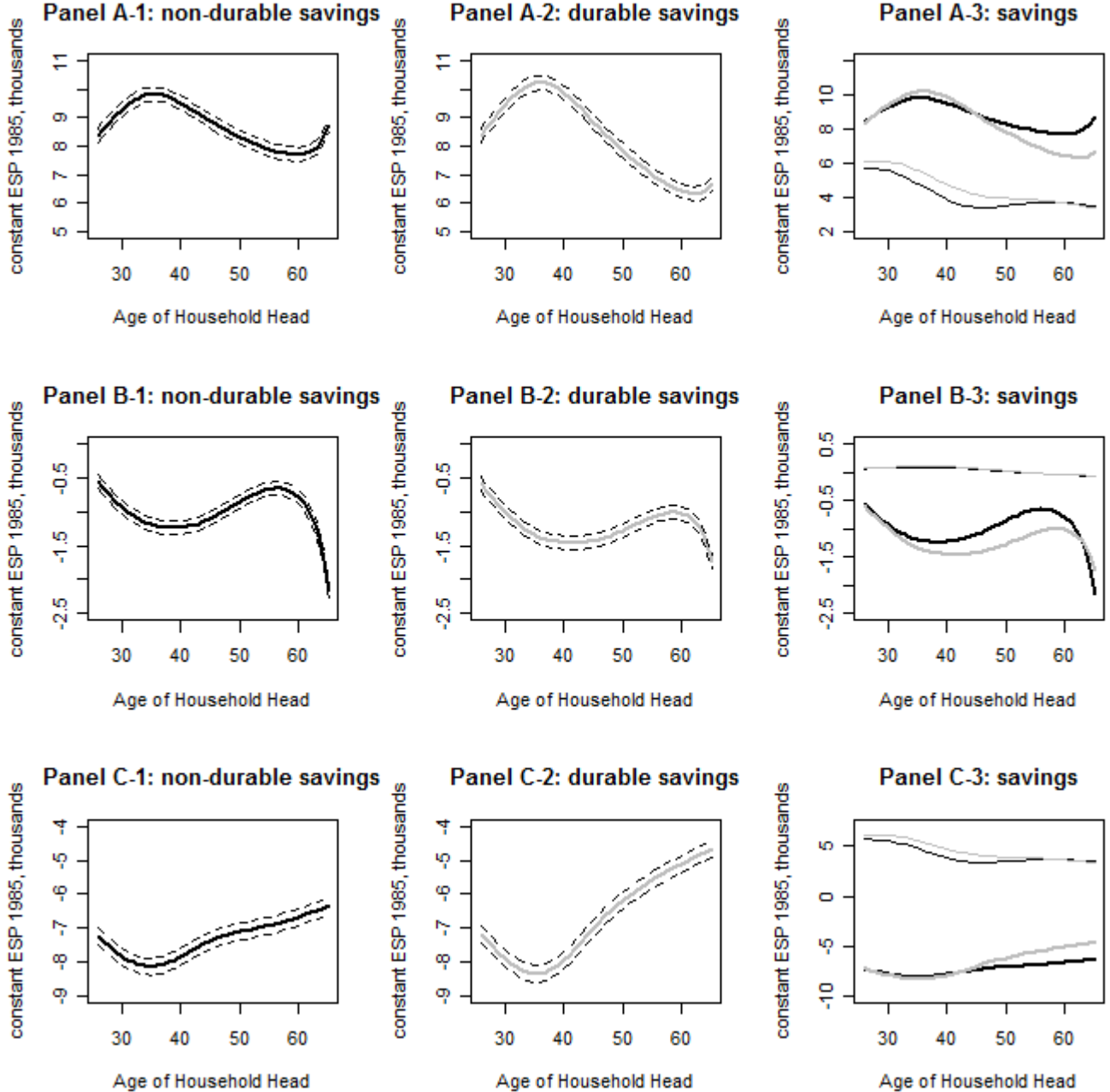
Notes: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

Now, we turn to analyze the impact of the different types of uncertainty on the household precautionary behavior when facing different educational levels (high or low). Controlling for household risk aversion we find out that the saving rate is different according to educational level. As we can

Figure 2. Household savings over the life-cycle by education level: low education

Notes: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

Figure 3. Household savings over the life-cycle: high education
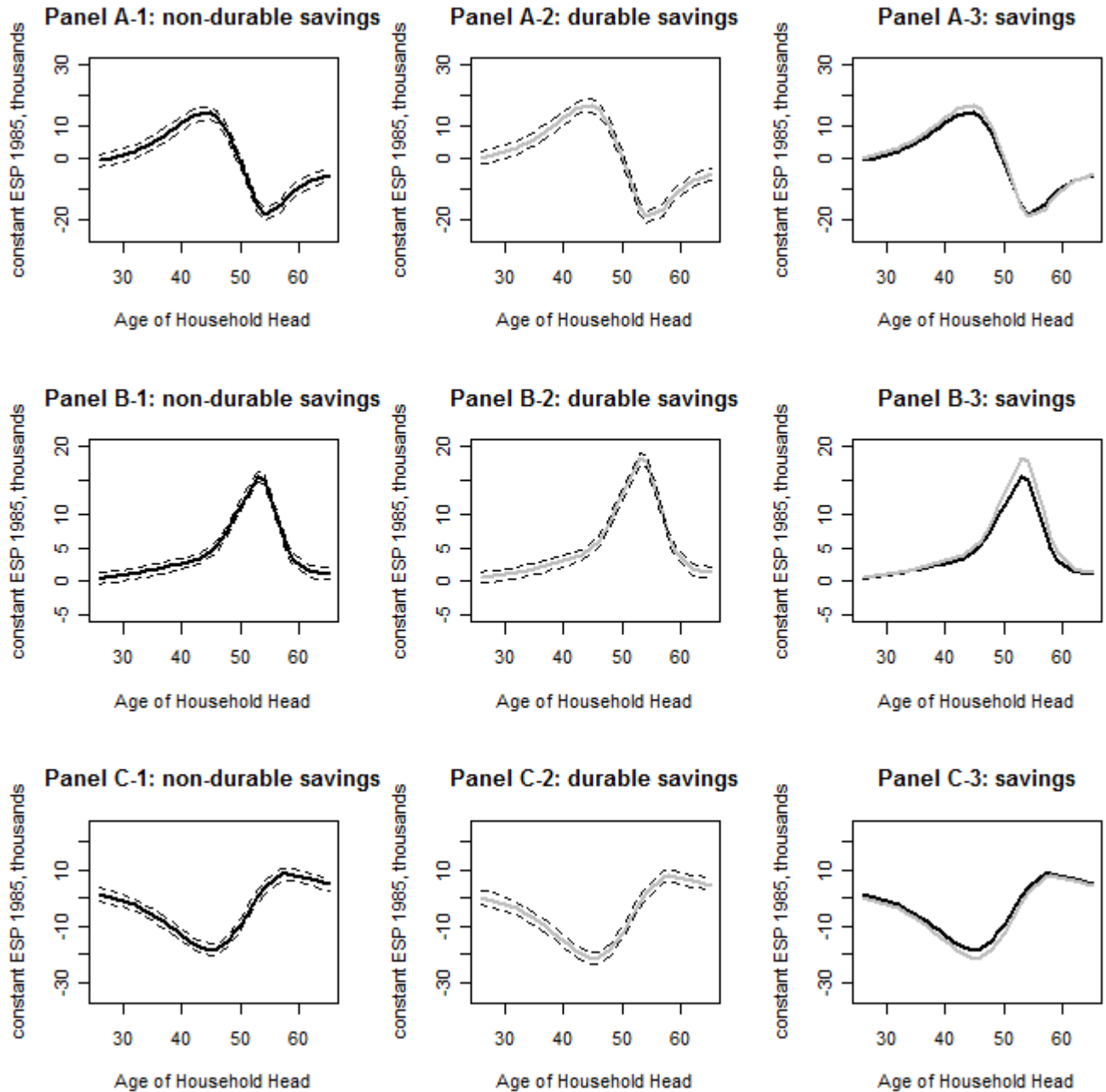
Notes: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

24

see in Figures 2 and 3, higher educational agents are risk averse tending to save during the early stages of work (26-42). Afterwards, they increase their consumption. On the contrary, households with less education show a smaller degree of risk aversion. They exhibit an inverted hump-shaped from age 29 to age 45 followed by a rather steady savings path along time. This result extends the findings that appear in Cagetti (2003), where it is pointed out that households with less education exhibit a lower degree of risk aversion.

Focusing on the precautionary savings elasticity to changes in unexpected health-care expenses we obtain that households with a higher educational level show a more cautious behavior regarding to unforseen health-care expenditures. Finally, analyzing this elasticity to unexpected income changes we appreciate a completely different behavior between these populations. Households with low education exhibit the hump-shaped established by the LCH model during the early stages of work (30-46), maintaining a constant saving rate in adulthood, whereas household with a higher educational level have a declining savings rate until age 47 when savings rate increase exponentially due to retirement or legacy reasons.

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 attractive. In addition, combining the particularities of this model jointly with the estimation strategy proposed in this paper enables us to determine household's consumption/savings decisions without having to resort to further strong assumptions about functional forms or densities.

6.3   Monte Carlo experiment

To investigate the small sample properties of our estimator, $\widetilde{m}(\cdot)$, we perform a Monte Carlo simulation. As a measure of accuracy we propose the following averaged mean squared error (AMSE)

$$AMSE\left(\widetilde{m}(z)\right) = \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{r=1}^{d}\left(\widetilde{m}_{\ell_r}(z) - m_{\ell_r}(z)\right)\right)^2,$$

where ℓ is the ℓ-th replication and L the number of replications.

Observations are generated from the following semi-parametric panel data regression system:

$$Y_{it} = \alpha(Z_{it}) + W_{it}m_1(Z_{it}) + U_{it}m_2(Z_{it}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T,$$

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 $(0,\pi/2)$ 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. $\mathcal{N}(0,1)$). 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 the individual effects are assumed to be correlated with the nonparametric covariates. Specifically, 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, whereas, for $i = 2, \cdots, N$, $\overline{Z}_{i\cdot} = T^{-1}\sum_{t=1}^{T} Z_{it}$ and $\overline{X}_{i\cdot} = T^{-1}\sum_{t=1}^{T} X_{it}$.

To verify the asymptotic theory of previous sections, 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 1000 Monte Carlo replications L and a Gaussian kernel. Following the necessary assumptions to obtain nonparametric estimators with a suitable asymptotic behavior, we propose to obtain the bandwidth matrix of H_4 by the Silverman's rule-of-thumb, whereas H_1, H_2, H_3 are chosen to be under-smoothing. Thus, $\widehat{H}_4 = \widehat{h}I = \widehat{\sigma}_z(NT)^{-1/5}$, where $\widehat{\sigma}_x$ and $\widehat{\sigma}_z$ are the sample standard deviation of X_{it} and Z_{it}, respectively. The estimated bias, standard deviation (Std), and AMSE of the estimators suggested above are reported in Table 4.

Table 4. Simulation results for empirical sizes.

T	N	Result	$\widetilde{m}(z)$	$\widetilde{\alpha}(z)$
3	50	Bias	0.032	-0.808
		Std	0.043	0.048
		AMSE	0.044	0.701
3	75	Bias	0.053	-0.814
		Std	0.033	0.036
		AMSE	0.036	0.698
3	100	Bias	0.047	-0.814
		Std	0.024	0.028
		AMSE	0.026	0.691

As we can see in Table 4, both minimum distance estimators, $\widetilde{m}(z)$ and $\widetilde{\alpha}(z)$, perform quite well. For all T, as N increases, the AMSEs of both estimators are lower, as we expected from their asymptotic properties described in the previous section. Furthermore, the biases are more or less steady for the different empirical sizes studied but the standard deviations are not. For $\widetilde{m}(z)$ the standard deviation is reduced from 0.043 when $N = 50$ to 0.024 when $N = 100$, whereas for $\widetilde{\alpha}(z)$ it passes from 0.048 to 0.028. Therefore, based on these results we can conclude that the consistency of the previous empirical results is proved.

7   Conclusion

This paper considers the estimation of a flexible structural model of optimal life-cycle savings, controlling both uncertainty about health-care expenditures and household risk aversion. To avoid misleading inferences, a locally constant IV estimator can be proposed. However, since a bivariate kernel weight is necessary to avoid the non-negligible asymptotic bias, the variance is enlarged. In order to achieve the optimal rate of convergence of this type of problems, a new one-step backfitting algorithm is proposed based on first differencing techniques that are combined through minimum distance estimation approach. The resulting estimators are shown to be oracle efficient and they exhibit the optimal rate of convergence for this smoothness problem. Compared to the estimation procedures already proposed in the literature, this new technique provides nonparametric estimators that enable us to deal simultaneously with several estimation problems (i) unobserved cross-sectional heterogeneity arbitrarily correlated with the covariates in an unknown way, (ii) varying parameters of unknown form in the Euler equation and (iii) endogenous explanatory variables. To illustrate the feasibility and possible gains of this method, we present an application about household's precautionary savings over the life-cycle. From this empirical application, we obtain that households accumulated wealth at least in two periods of life. In younger stages, to guard against uncertainty about unforeseen income downturns and when they become older due to other reasons related to retirement and bequests. Also it is shown the negative impact of public health programs on precautionary savings. The paper concludes with a Monte Carlo simulation.

8   Appendix

Proof of Theorem 4.1

Remember that throughout the paper we denoted the following $d \times 1$ vectors,

$$\Delta \widetilde{W}_{it} = \left(\Delta W_{it}^\top \quad \Delta U_{it}^\top \right)^\top, \qquad \Delta \widetilde{W}_{g,it} = \left(g_{it,i(t-1)}^\top \quad \Delta U_{it}^\top \right)^\top, \qquad \Delta \widetilde{W}_{\widehat{g},it} = \left(\widehat{g}_{it,i(t-1)}^\top \quad \Delta U_{it}^\top \right)^\top$$

$$\widetilde{W}_{it} = \left(W_{it}^\top \quad U_{it}^\top \right)^\top, \qquad \widetilde{W}_{i(t-1)} = \left(W_{i(t-1)}^\top \quad U_{i(t-1)}^\top \right)^\top.$$

In order to obtain the desired results of Theorem 4.1 we define

$$\widehat{m}_g(z; H_2) = \left(\sum_{i=1}^{T} \sum_{t=2}^{T} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \Delta \widetilde{W}_{it}^\top \right)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \Delta Y_{it},$$

where

$$K_{it} = \frac{1}{|H_2|^{1/2}} K \left(H_2^{-1/2}(Z_{it} - z) \right) \qquad ; \qquad K_{i(t-1)} = \frac{1}{|H_2|^{1/2}} K \left(H_2^{-1/2}(Z_{i(t-1)} - z) \right).$$

Clearly, the two-step weighted locally constant least-squares estimator (3.5) can be written as

$$\widehat{m}_{\widehat{g}}(z; H_2) = \left(\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2) \right) + \widehat{m}_g(z; H_2). \tag{A.1}$$

According to (A.1), to prove the Theorem 4.1 all that we need to show is that, under the conditions established in the Theorem 4.1, we get that as N tends to infinity and T is fixed,

$$\sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2)\right) = o_p(1), \qquad \text{uniformly in} \quad z$$

$$\sqrt{N|H_2|}\left(\widehat{m}_g(z; H_2) - m(z) - B(z; H_2)\right) \xrightarrow{d} \mathcal{N}\left(0, V(z, H_2)\right),$$

where

$$B(z; H_2) = \mu_2(K)\left(diag_d\left(D_f(z)H_2\sqrt{N|H_2|}D_{m_r}(z)\right)\imath_d f^{-1}_{Z_{it}, Z_{i(t-1)}}(z, z) + \frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_r}(z)H_2\sqrt{N|H_2|}\right)\right)\imath_d\right)$$

and

$$V(z; H_2) = 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z, z)\mathcal{B}_{\Delta\widetilde{W}\Delta\widetilde{W}}(z, z)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z, z).$$

These results are proved in Lemmas 8.1 and 8.2.

LEMMA 8.1 *Under conditions of Theorem 4.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2)\right) = o_p(1) \qquad \text{uniformly in} \quad z.$$

Proof of Lemma 8.1.

Throughout this appendix we use the following notation

$$\widehat{S}_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{\widehat{g},it}\Delta\widetilde{W}_{it}^\top \quad ; \qquad \widehat{T}_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{\widehat{g},it}\Delta Y_{it}.$$

Let us write the first element of (A.1) as

$$\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2) = \widehat{S}_n^{-1}\widehat{T}_n - S_n^{-1}T_n, \tag{A.2}$$

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 $\widehat{g}_{it,i(t-1)}$. At this situation, we first show that as N tends to infinity,

$$\widehat{S}_n^{-1} = \mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z, z) + o_p(\|H_2^{1/2}\|), \tag{A.3}$$

where

$$\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z, z) = E\left[\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top|Z_{it} = z, Z_{i(t-1)} = z\right]f_{Z_{it}, Z_{i(t-1)}}(z, z).$$

For this, let us denote

$$\widehat{S}_n = S_n + I\!I_{1n}, \tag{A.4}$$

where as

$$S_n = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top,$$

$$I\!I_{1n} = (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\left(\Delta\widetilde{W}_{\widehat{g},it} - \Delta\widetilde{W}_{g,it}\right)\Delta\widetilde{W}_{it}^\top,$$

so we have to analyze each term separately to prove (A.3). To this end, we follow the usual Taylor expansion; i.e.,

$$f(z + H_2^{1/2}v) = f(z) + D_f^\top(z)H_2^{1/2}v + o_p(\|H_2^{1/2}\|), \quad as \quad \|H_2\| \to 0.$$

Then, given that Z_{it} and v_{it} are *i.i.d.* across i and because the stationary assumption, when N tends to infinity and by the law of iterated expectations it implies

$$
\begin{aligned}
E(S_n) &= \int \int E\left[\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top | Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v\right] \\
&\times \ f_{Z_{it},Z_{i(t-1)}}\left(Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v\right)K(u)K(v)dudv.
\end{aligned}
$$

Under Assumption **4.1**,

$$
\begin{aligned}
Var(S_n) &= Var\left(K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top\right) \\
&+ \frac{1}{T}\sum_{t=3}^{T}(T-t)Cov\left(K_{i2}K_{i1}\Delta\widetilde{W}_{g,i2}\Delta\widetilde{W}_{i2}^\top, K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top\right),
\end{aligned}
$$

where, under conditions **4.7-4.8**, it holds

$$Var\left(K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top\right) = O_p\left(\frac{1}{N|H_2|}\right)$$

and

$$Cov\left(K_{i2}K_{i1}\Delta\widetilde{W}_{g,i2}\Delta\widetilde{W}_{i2}^\top, K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{it}^\top\right) = O_p\left(\frac{1}{N|H_2|}\right).$$

If $N|H_2|$ tends to infinity, this variance term tends to zero and it is proved

$$S_n = \mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z, z)(1 + o_p(1)). \tag{A.5}$$

Now, focus on the behavior of $\mathbb{II}_{1n}$, by Assumptions **4.2** and **4.5-4.8** we obtain

$$\mathbb{II}_{1n} = \begin{pmatrix} (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}\right)\Delta\widetilde{W}_{it}^\top \\ \\ (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\left(\Delta U_{it} - \Delta U_{it}\right)\Delta\widetilde{W}_{it}^\top \end{pmatrix} = o_p(1) \tag{A.6}$$

given that, using the uniform convergence results as the ones established in Theorem 6 of Masry (1996),

$$
\begin{aligned}
& (NT)^{-1}\sum_{it} K_{it}K_{i(t-1)}\left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}\right)\Delta\widetilde{W}_{it}^\top \\
\leq \ & (NT)^{-1}\sum_{it} \sup_{\{X_{it},X_{i(t-1)}\}} |\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}| |K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}|^\top \\
= \ & o_p(1), \tag{A.7}
\end{aligned}
$$

since it is straightforward to show $(NT)^{-1}\sum_{it} |K_{it}K_{i(t-1)}\Delta\widetilde{W}_{it}| = O_p(1)$.

By (A.4) we know $\widehat{S}_n = S_n + I\!\!I_{1n}$ and following the rule of the inverse matrix and the Taylor's theorem we get

$$\widehat{S}_n^{-1} = S_n^{-1} + S_n^{-1} I\!\!I_{1n} S_n^{-1} + o_p(\|H_2^{1/2}\|).$$

Replacing these previous results here we get

$$\widehat{S}_n^{-1} = \mathcal{B}_{\Delta\widetilde{W}_g \Delta\widetilde{W}}^{-1}(z,z) + o_p(\|H_2^{1/2}\|) \tag{A.8}$$

so the result (A.3) is proved.

On the other hand, replacing (A.3) and (A.5) in (A.2) and by the Cramer-Wold device we get

$$\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2) = \mathcal{B}_{\Delta\widetilde{W}_g \Delta\widetilde{W}}^{-1}(z,z)(\widehat{T}_n - T_n) + o_p(1). \tag{A.9}$$

Focus now on the behavior of $\widehat{T}_n - T_n$ we claim that following the same procedure as in (A.6) we get

$$\widehat{T}_n - T_n \;\;=\;\; (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)}(\Delta\widetilde{W}_{\widehat{g},it} - \Delta\widetilde{W}_{g,it})\Delta Y_{it} = o_p(1). \tag{A.10}$$

Using the uniform convergence of $\widehat{g}_{it,i(t-1)}$,

$$(NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \left(\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}\right) \Delta Y_{it}$$
$$\leq \;\; (NT)^{-1} \sum_{it} \sup_{\{X_{it}, X_{i(t-1)}\}} |\widehat{g}_{it,i(t-1)} - g_{it,i(t-1)}||K_{it} K_{i(t-1)}\Delta Y_{it}|$$
$$= \;\; o_p(1)$$

since it is straightforward to show that $(NT)^{-1} \sum_{it} |K_{it} K_{i(t-1)}\Delta Y_{it}| = O_p(1)$. Then, by (A.9) and (A.10) we obtain that as $N|H_2| \to \infty$

$$\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2) = o_p\left(\frac{1}{\sqrt{N|H_2|}}\right),$$

so the Lemma 8.1 is proved.

∎

In order to prove the asymptotic distribution of $\widehat{m}_{\widehat{g}}(z; H_2)$ we can write (A.1) as

$$\sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z; H_2) - m(z)\right) = \sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z; H_2) - \widehat{m}_g(z; H_2)\right) + \sqrt{N|H_2|}\left(\widehat{m}_g(z; H_2) - m(z)\right) \tag{A.11}$$

and need the following Lemma.

LEMMA 8.2 *Under the conditions established in Theorem 4.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{N|H_2|}\left(\widehat{m}_g(z; H_2) - m(z) - B(z, H_2)\right) \;\; \xrightarrow{d} \;\; \mathcal{N}\left(0, V(z, H_2)\right).$$

Proof of Lemma 8.2.

The proof of this lemma is structured as follows. First, the asymptotic bias of the estimator is analyzed. Later, we focus on the variance term and we 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 $o_p\left(tr(H_2)\right)$, if we replace ΔW_{it} by $(g(X_{it}, X_{i(t-1)}) + \Delta\xi_{it})$ in (3.3) the Taylor's approximation of the smooth functions implies

$$
\begin{aligned}
\Delta Y_{it} \;=\; & \Delta\widetilde{W}_{it}^\top m(z) + \Delta Z_{it}^\top D_\alpha(z) + \left(\widetilde{W}_{it} \otimes (Z_{it} - z) - \widetilde{W}_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right)^\top D_m(z) \\
& + \frac{1}{2}\left((Z_{it} - z)^\top \mathcal{H}_\alpha(z)(Z_{it} - z) - (Z_{i(t-1)} - z)^\top \mathcal{H}_\alpha(z)(Z_{i(t-1)} - z)\right) \\
& + \frac{1}{2}\left(\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z) - \widetilde{W}_{i(t-1)}^\top \otimes (Z_{i(t-1)} - z)^\top \mathcal{H}_m(z)(Z_{i(t-1)} - z)\right) \\
& + \Delta v_{it} + \Delta\xi_{it}^\top m_1(z) + o_p(1), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{(A.12)}
\end{aligned}
$$

let $D_m(z)$ be a $(d-1)q \times 1$ vector and $D_\alpha(z)$ a $q \times 1$ vector, for $D_m(z) = vec(\partial m(z)/\partial z^\top)$ and $D_\alpha(z) = vec(\partial\alpha(z)/\partial z^\top)$ being the corresponding first-order derivatives vector of $m(\cdot)$ and $\alpha(\cdot)$, respectively. Also, $\mathcal{H}_m(z)$ is a $(d-1)q \times q$ matrix and $\mathcal{H}_\alpha(z)$ is a $q \times q$ matrix, for $\mathcal{H}_m(z) = \partial^2 m(z)/\partial zz^\top$ and $\mathcal{H}_\alpha = \partial^2\alpha(z)/\partial z\partial z^\top$ being the corresponding Hessian matrix of $m(\cdot)$ and $\alpha(\cdot)$, respectively. Also, for the sake of simplicity we denote

$$
\begin{aligned}
\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top D_m(z) \;=\; & \left(W_{it}^\top \otimes (Z_{it} - z)^\top D_{m_1}(z), \quad U_{it}^\top \otimes (Z_{it} - z)^\top D_{m_2}(z)\right), \\
\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z) \;=\; & \left(W_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_{m_1}(z)(Z_{it} - z), \quad U_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_{m_2}(z)(Z_{it} - z)\right)
\end{aligned}
$$

and similarly for $(\widetilde{W}_{i(t-1)} \otimes (Z_{i(t-1)} - z))^\top D_m(z)$ and $(\widetilde{W}_{i(t-1)} \otimes (Z_{i(t-1)} - z))^\top \mathcal{H}_m(z)(Z_{i(t-1)} - z)$.

Combining (A.12) with the second element of (A.11) we can write $\widehat{m}_g(z; H_2)$ as

$$
\begin{aligned}
& \widehat{m}_g(z; H_2) - m(z) = \\
& \left(\sum_{i=1}^N \sum_{t=2}^T K_{it} K_{i(t-1)} \Delta\widetilde{W}_{g,it} \Delta\widetilde{W}_{it}^\top\right)^{-1} \sum_{i=1}^N \sum_{t=2}^T K_{it} K_{i(t-1)} \Delta\widetilde{W}_{g,it}\left(G_{it} + \Delta v_{it} + m_1(z)^\top \Delta\xi_{it}\right), \quad \text{(A.13)}
\end{aligned}
$$

where

$$
\begin{aligned}
G_{it} \;=\; & \Delta Z_{it}^\top D_\alpha(z) + \left(\widetilde{W}_{it} \otimes (Z_{it} - z) - \widetilde{W}_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right)^\top D_m(z) \\
& + \frac{1}{2}\left((Z_{it} - z)^\top \mathcal{H}_\alpha(z)(Z_{it} - z) - (Z_{i(t-1)} - z)^\top \mathcal{H}_\alpha(z)(Z_{i(t-1)} - z)\right) \\
& + \frac{1}{2}\left(\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z) - \widetilde{W}_{i(t-1)}^\top \otimes (Z_{i(t-1)} - z)^\top \mathcal{H}_m(z)(Z_{i(t-1)} - z)\right) \\
& - m_1(z)^\top \Delta\xi_{it}.
\end{aligned}
$$

For the sake of simplicity, let us denote

$$
\widehat{m}_g(z; H_2) - m(z) = S_n^{-1}\left(B_n + U_n + R_n\right), \qquad\qquad\qquad\qquad\qquad \text{(A.14)}
$$

where

$$B_n = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} G_{it},$$

$$U_n = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \Delta v_{it},$$

$$R_n = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} m_1(z)^\top \Delta \xi_{it}.$$

Thus, to complete the proof of Theorem 4.1 it is enough to show

$$\sqrt{N|H_2|}\left(\widehat{m}_g(z; H_2) - m(z)\right) - \sqrt{N|H_2|} S_n^{-1} B_n = \sqrt{N|H_2|} S_n^{-1} \left(U_n + R_n\right), \qquad (A.15)$$

where we will demonstrate that $S_n^{-1} B_n$ contributes to the asymptotic bias, whereas the two terms of the right-hand side of (A.14) are asymptotically normal.

Focus first on the asymptotic behavior of the bias term, we can decompose B_n into five different terms that we have to analyze separately. Specifically, B_n can be rewritten as

$$B_n = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} G_{it} = \left(B_n^{(1)} + B_n^{(2)} + B_n^{(3)} + B_n^{(4)} - B_n^{(5)}\right), \qquad (A.16)$$

where

$$B_n^{(1)} = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \Delta Z_{it}^\top D_\alpha(z),$$

$$B_n^{(2)} = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \left(\widetilde{W}_{it} \otimes (Z_{it} - z) - \widetilde{W}_{i(t-1)} \otimes \left(Z_{i(t-1)} - z\right)\right)^\top D_m(z),$$

$$B_n^{(3)} = (2NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \left((Z_{it} - z)^\top \mathcal{H}_\alpha(z) (Z_{it} - z) - \left(Z_{i(t-1)} - z\right)^\top \mathcal{H}_\alpha(z) \left(Z_{i(t-1)} - z\right)\right),$$

$$B_n^{(4)} = (2NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \left(\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z) (Z_{it} - z) - \widetilde{W}_{i(t-1)}^\top \otimes \left(Z_{i(t-1)} - z\right)^\top\right.$$
$$\left. \times \mathcal{H}_m(z) \left(Z_{i(t-1)} - z\right)\right),$$

$$B_n^{(5)} = (NT)^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} m_1(z)^\top \Delta \xi_{it}.$$

For the standard case $\mu_2(K_u) = \mu_2(K_v)$ we get that by the law of iterated expectations and the stationarity condition,

$$\begin{aligned}
E(B_n^{(1)}) &= E\left[K_{it} K_{i(t-1)} \Delta \widetilde{W}_{g,it} \Delta Z_{it}^\top D_\alpha(z)\right] \\
&= \int K_{it} K_{i(t-1)} E(\Delta \widetilde{W}_{g,it} | Z_{it}, Z_{i(t-1)}) \Delta Z_{it}^\top D_\alpha(z) f(Z_{it}, Z_{i(t-1)}) dZ_{it} dZ_{i(t-1)} \\
&= E(\Delta \widetilde{W}_{g,it} | Z_{it} = z, Z_{i(t-1)} = z) \left(\mu_2(K_u) - \mu_2(K_v)\right) D_f(z) H_2 D_\alpha(z) \\
&= o_p(1). \qquad (A.17)
\end{aligned}$$

Similarly, by iterated expectations

$$
\begin{aligned}
E(B_n^{(2)}) &= E\Big[K_{it}K_{i(t-1)}\Big(E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it},Z_{i(t-1)})\otimes(Z_{it}-z)^\top \\
&\quad - E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{i(t-1)}^\top|Z_{it},Z_{i(t-1)})\otimes\big(Z_{i(t-1)}-z\big)^\top\Big)D_m(z)\Big] \\
&= \int\Big(E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it}=z,Z_{i(t-1)}=z)D_f(z)(H_2^{1/2}u)\Big)\otimes(H_2^{1/2}u)^\top D_m(z)K(u)K(v)dudv \\
&\quad - \int\Big(E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{i(t-1)}^\top|Z_{it}=z,Z_{i(t-1)}=z)D_f(z)(H_2^{1/2}v)\Big)\otimes(H_2^{1/2}v)^\top D_m(z)K(u)K(v)dudv \\
&= \mu_2(K)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)diag_d\big(D_f(z)tr(H_2)D_{m_r}(z)\big)\imath_d f_{Z_{it},Z_{i(t-1)}}^{-1}(z,z)+o_p(tr(H_2)). \qquad (A.18)
\end{aligned}
$$

On its part, following a similar procedure as in (A.17) it is straightforward to show

$$
\begin{aligned}
E(B_n^{(3)}) &= \frac{1}{2}E\Big[K_{it}K_{i(t-1)}E(\Delta\widetilde{W}_{g,it}|Z_{it},Z_{i(t-1)})\Big((Z_{it}-z)^\top\mathcal{H}_\alpha(z)(Z_{it}-z) \\
&\quad - (Z_{i(t-1)}-z)^\top\mathcal{H}_\alpha(z)(Z_{i(t-1)}-z)\Big)\Big] \\
&= \frac{1}{2}E(\Delta\widetilde{W}_{g,it}|Z_{it}=z,Z_{i(t-1)}=z)(\mu_2(K_u)-\mu_2(K_v))tr(\mathcal{H}_\alpha(z)H_2)+o_p(tr(H_2)), \quad (A.19)
\end{aligned}
$$

where $diag_d(tr(\mathcal{H}_\alpha(z)H_2))$ stands for a diagonal matrix of element $tr(\mathcal{H}_\alpha(z)H_2)$ whereas following the proof of (A.18) and (A.19),

$$
\begin{aligned}
E(B_n^{(4)}) &= \frac{1}{2}E\Big[K_{it}K_{i(t-1)}\Big(E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it},Z_{i(t-1)})\otimes(Z_{it}-z)^\top\mathcal{H}_m(z)(Z_{it}-z) \\
&\quad - E(\Delta\widetilde{W}_{g,it}\widetilde{W}_{i(t-1)}^\top|Z_{it},Z_{i(t-1)})\otimes(Z_{i(t-1)}-z)^\top\mathcal{H}_m(z)(Z_{it}-z)\Big)\Big] \\
&= \frac{1}{2}\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)diag_d(tr(\mathcal{H}_{m_r}(z)H_2))\imath_d\mu_2(K)+o_p(tr(H_2)). \qquad (A.20)
\end{aligned}
$$

Also, it is straightforward to show that $E(B_n^{(5)})=o_p(1)$ since we assume $E(\xi_{it}|Z_{it},Z_{i(t-1)})=0$.

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 8.1 and assuming $H_2\to 0$ and $N|H_2|\to\infty$. Then, replacing (A.17)-(A.20) in B_n, using (A.7) and applying the Cramer-Wold device we get that the bias term of this local constant IV regression estimator (3.5) is

$$
\begin{aligned}
S_n^{-1}B_n &= \mu_2(K)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)^{-1}\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\Big[diag_d(D_f(z)tr(H_2)D_{m_r}(z))\imath_d f_{Z_{it},Z_{i(t-1)}}^{-1}(z,z) \\
&\quad + \frac{1}{2}diag_d(tr(\mathcal{H}_{m_r}(z)H_2))\imath_d\Big]+o_p(tr(H_2)), \qquad (A.21)
\end{aligned}
$$

so the first part of the proof is done.

On the other hand, to obtain the asymptotic variance of the right part of (A.15) 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 $\Delta v=(\Delta v_1,\cdots,\Delta v_N)$ as the $N(T-1)\times 1$-vector for $\Delta v_i=(\Delta v_{i2},\cdots,\Delta v_{iT})^\top$,

$$
E(\Delta v_i\Delta v_i^\top|X_{it},X_{i(t-1)},Z_{it},Z_{i(t-1)},U_{it},U_{i(t-1)}) = \begin{cases} 2\sigma_v^2, & for \quad i=i', \quad t=t', \\ -\sigma_v^2, & for \quad i=i', \quad |t-t'|<2, \quad (A.22) \\ 0, & for \quad i=i', \quad |t-t'|\geq 2. \end{cases}
$$

33

When we analyze U_n we claim that by the law of iterated expectations and Assumptions **4.1**, **4.2** and **4.4-4.8**

$$
\begin{aligned}
N|H_2|Var(U_n) &= |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta\widetilde{W}_{g,it}E(\Delta v_{it}\Delta v_{i't'}|X_{it},X_{i(t-1)},Z_{it},Z_{i(t-1)},U_{it},U_{i(t-1)})\right.\\
&\qquad\left.\times\Delta\widetilde{W}_{g,i't'}^{\top}K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right]\\
&= 2\sigma_v^2 R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)(1+o_p(1)).
\end{aligned}
\tag{A.23}
$$

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 **4.1**, **4.2** and **4.4-4.8**, by standard kernel methods we obtain

$$
|H_2|T^{-1}\sum_{t=2}^{T}E\left[\Delta\widetilde{W}_{g,it}E(\Delta v_{it}^2|X_{it},X_{i(t-1)},Z_{it},Z_{i(t-1)},U_{it},U_{i(t-1)})\Delta\widetilde{W}_{g,it}^{\top}K_{it}^2K_{i(t-1)}^2\right]
$$
$$
= 2\sigma_v^2|H_2|E\left[E(\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^{\top}|Z_{it},Z_{i(t-1)})K_{it}^2K_{i(t-1)}^2\right] = 2\sigma_v^2 R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)(1+o_p(1)).
$$

Meanwhile, for $t\neq t'$, we proceed in the same way as in the previous equation so if we consider again the stationary assumption, **4.1**, we get

$$
2|H_2|T^{-1}\sum_{t=3}^{T}(T-t)E\left[\Delta\widetilde{W}_{g,i2}E(\Delta v_{i2}\Delta v_{it}|X_{it},X_{i(t-1)},X_{i(t-2)},Z_{it},Z_{i(t-1)},Z_{i(t-2)},U_{it},U_{i(t-1)},U_{i(t-2)})\right.
$$
$$
\left.\times\Delta\widetilde{W}_{g,it}^{\top}K_{i2}K_{i1}K_{it}K_{i(t-1)}\right]
$$
$$
= -\sigma_v^2|H_2|^{1/2}R(K_u)E[\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^{\top}|Z_{i1}=z,Z_{i2}=z,Z_{i3}=z]f_{Z_{i1},Z_{i2},Z_{i3}}(z,z,z)(1+o_p(1)).
$$

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 (A.22).

Second, we focus on the behavior of R_n and follow a similar procedure as in (A.23). Let $\Sigma_{\Delta\xi\Delta\xi} = E(\Delta\xi_{it}\Delta\xi_{it}^{\top})$ be,

$$
\begin{aligned}
N|H_2|Var(R_n) &= |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta\widetilde{W}_{g,it}m_1(z)^{\top}E(\Delta\xi_{it}\Delta\xi_{it}^{\top}|X_{it},X_{i(t-1)},Z_{it},Z_{i(t-1)},U_{it},U_{i(t-1)})\right.\\
&\qquad\left.\times m_1(z)\Delta\widetilde{W}_{g,i't'}^{\top}K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right]\\
&= 2R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)m_1(z)^{\top}\Sigma_{\Delta\xi\Delta\xi}m_1(z)(1+o_p(1)).
\end{aligned}
\tag{A.24}
$$

Similarly, let $\Sigma_{\Delta v\Delta\xi} = E(\Delta v_{it}\Delta\xi_{it}^{\top})$ be we get

$$
\begin{aligned}
N|H_2|Cov(U_n,R_n) &= |H_2|(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta\widetilde{W}_{g,it}E(\Delta v_{it}\Delta\xi_{it}^{\top}|X_{it},X_{i(t-1)},Z_{it},Z_{i(t-1)},U_{it},U_{i(t-1)})\right.\\
&\qquad\left.\times m_1(z)\Delta\widetilde{W}_{g,i't'}K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right]\\
&= R(K_u)R(K_v)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)\Sigma_{\Delta v\Delta\xi}m_1(z)(1+o_p(1)).
\end{aligned}
\tag{A.25}
$$

Then, applying the Cramer-Wold device and using (A.7) and (A.23)-(A.25), as $N|H_2| \to \infty$,

$$N|H_2|Var\left(S_n^{-1}(U_n + R_n)\right) = 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^\top \Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right)$$
$$\times \mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)(1+o_p(1)) \quad \text{(A.26)}$$

Since we believe that the conditions established on H_1 and H_2 are enough to show that the other terms are $o_p(1)$, to complete the proof of Theorem 4.1 is necessary to show that as $N \to \infty$,

$$\sqrt{N|H_2|}\left(\widehat{m}_g(z;H_2) - m(z)\right) \xrightarrow{d} \mathcal{N}\left(0, 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^\top \Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right)\right.$$
$$\left.\times \mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\right) \quad \text{(A.27)}$$

In order to show that, we check the Lyapunov condition. As the reader can appreciate, the Assumption **4.1** states that the variables are $i.i.d.$ in i but not in t, so we have independent random variables heterogeneously distributed. To overcome this situation we can define $\lambda_{n,i}^* = T^{-1/2}\sum_{it}\lambda_{it}$, which means that $\lambda_{n,i}^*$ is an independent random variable for T fixed. Therefore, in order to prove (A.27) we focus on the asymptotic normality of the local constant IV estimator (3.5) and we can write

$$\sqrt{N|H_2|}(NT)^{-1}\sum_{it}K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\left(\Delta v_{it} + \Delta\xi_{it}^\top m_1(z)\right) = \frac{1}{\sqrt{NT}}\sum_{it}\lambda_{it}, \quad \text{(A.28)}$$

where

$$\lambda_{it} = K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\left(\Delta v_{it} + \Delta\xi_{it}^\top m_1(z)\right)|H_2|^{1/2}, \qquad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T.$$

By Theorem 4.1 and previous proofs, we can state that as $H_2 \to 0$

$$Var(\lambda_{it}) = 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z)\right)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)$$
$$\times \mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)(1+o_p(1)),$$
$$Cov(\lambda_{i1},\lambda_{it}) = R(K_u)R(K_v)\Sigma_{\Delta v\Delta\xi}m_1(z)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)\mathcal{B}^{-1}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}(z,z)(1+o_p(1)),$$

and by the Minkowski inequality we get

$$E|\lambda_{n,i}^*|^{2+\delta} \le CT^{-(2+\delta)}E|\lambda_{it}|^{2+\delta}.$$

In this way, λ_{it} can be split up into two components; i.e., λ_{1it} and λ_{2it}. Analyzing separately each of these terms,

$$E|\lambda_{1it}|^{2+\delta} \le |H_2|^{(2+\delta)/2}E|K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}\Delta v_{it}|^{2+\delta}$$
$$= |H_2|^{-\delta/2}\int E\left(|\Delta\widetilde{W}_{g,it}\Delta v_{it}|^{2+\delta}|Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v\right)$$
$$\times f_{Z_{it},Z_{i(t-1)}}(z + H_2^{1/2}u, z + H_2^{1/2}v)K^{2+\delta}(u)K^{2+\delta}(v)dudv$$
$$= |H_2|^{-\delta/2}E\left(|\Delta\widetilde{W}_{g,it}\Delta v_{it}|^{2+\delta}|Z_{it} = z, Z_{i(t-1)} = z\right)f_{Z_{it},Z_{i(t-1)}}(z,z)$$
$$\times \int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2}).$$

Similarly,

$$
\begin{aligned}
E|\lambda_{2it}|^{2+\delta} &\leq |H_2|^{(2+\delta)/2} E|K_{it}K_{i(t-1)}\Delta\widetilde{W}_{g,it}m_1(z)^\top\Delta\xi_{it}|^{2+\delta} \\
&= |H_2|^{1+\delta/2} E\left[E\left(|\Delta\widetilde{W}_{g,it}m_1(z)^\top\Delta\xi_{it}|^{2+\delta}|Z_{it},Z_{i(t-1)}\right)K_{it}^{2+\delta}K_{i(t-1)}^{2+\delta}\right] \\
&= |H_2|^{-\delta/2} E\left(|\Delta\widetilde{W}_{g,it}\Delta\xi_{it}^\top|^{2+\delta}|Z_{it}=z,Z_{i(t-1)}=z\right)m_1(z)^{2+\delta}f_{Z_{it},Z_{i(t-1)}}(z,z) \\
&\times \int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2})
\end{aligned}
$$

and

$$
\begin{aligned}
E|\lambda_{1it}\lambda_{2it}^\top|^{1+\delta/2} &\leq |H_2|^{(2+\delta)/2} E|K_{it}^2 K_{i(t-1)}^2\Delta\widetilde{W}_{g,it}\Delta v_{it}\Delta\xi_{it}^\top m_1(z)\Delta\widetilde{W}_{g,it}^\top|^{1+\delta/2} \\
&= -|H_2|^{1+\delta/2} E\left[E\left(|\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^\top\Delta v_{it}\Delta\xi_{it}^\top|^{1+\delta/2}|Z_{it},Z_{i(t-1)}\right)m_1(z)^{(1+\delta/2)}K_{it}^{2+\delta}K_{i(t-1)}^{2+\delta}\right] \\
&= -|H_2|^{1+\delta/2} E\left(|\Delta\widetilde{W}_{g,it}\Delta\widetilde{W}_{g,it}^\top\Delta v_{it}\Delta\xi_{it}^\top|^{1+\delta/2}|Z_{it}=z,Z_{i(t-1)}=z\right)m_1(z)^{(1+\delta/2)} \\
&\times f_{Z_{it},Z_{i(t-1)}}(z,z)\int K^{2+\delta}(u)K^{2+\delta}(v)dudv + o_p(|H_2|^{-\delta/2}).
\end{aligned}
$$

Then, it is proved that

$$
E|\lambda_{n,i}|^{2+\delta} = E\left|N^{-1/2}\sum_i \lambda_{n,i}^*\right|^{2+\delta} \leq CO_p\left((N|H_2|)^{-\delta/2}\right)
$$

and, as N tends to infinity, $N|H_2| \to \infty$. Therefore, since the Lyapunov condition holds, we resort to the Lyapunov Central Limit Theorem to verify (A.27) and the Lemma 8.2 is proved.

Finally, using the results of Lemmas 8.1 and 8.2 in (A.14) we get

$$
\begin{aligned}
\sqrt{N|H_2|}\left(\widehat{m}_{\widehat{g}}(z;H_2) - m(z) - B(z,H_2)\right) \xrightarrow{d} \; &\mathcal{N}\left(0, 2R(K_u)R(K_v)\left(\sigma_v^2 + m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right)\right. \\
&\left.\times \mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}^{-1}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}_g}(z,z)\mathcal{B}_{\Delta\widetilde{W}_g\Delta\widetilde{W}}^{-1}(z,z)\right)
\end{aligned}
$$

so the proof of Theorem 4.1 is done.

$\blacksquare$

Proof of Theorem 5.1

The proof of this result follows the same lines as in the proof of Theorem 4.1. Remember that we denoted

$$
\widetilde{W}_{\widehat{g},it}^\top = \left(\widehat{g}_{it,i(t-1)}^\top \quad U_{it}^\top\right)^\top, \quad \widetilde{W}_{g,i(t-1)}^\top = \left(g_{it,it(t-1)}^\top \quad U_{i(t-1)}^\top\right)^\top
$$

as vectors of $d \times 1$ dimension.

Proceeding as before, we define

$$
\widehat{m}_g^{(1)}(z;H_4) = \left(\sum_{i=1}^N\sum_{t=2}^T K_{it}\ddot{W}_{g,it}\ddot{W}_{it}^\top\right)^\top \sum_{i=1}^N\sum_{t=2}^T K_{it}\ddot{W}_{g,it}\Delta\widetilde{Y}_{1it}
$$

where now

$$K_{it} = \frac{1}{|H_4|^{1/2}} K\left(H_4^{-1/2}(Z_{it} - z)\right)$$

and the one-step backfitting estimator can be written as

$$\widehat{m}_{\widehat{g}}^{(1)}(z; H_4) = \left(\widehat{m}_{\widehat{g}}^{(1)}(z; H_4) - \widehat{m}_g^{(1)}(z; H_4)\right) + \widehat{m}_g^{(1)}(z; H_4). \tag{A.29}$$

Therefore, in order to prove Theorem 5.1 is necessary to show the results of the following Lemmas 8.3 and 8.4.

LEMMA 8.3

$$\sqrt{N|H_4|^{1/2}} \left(\widehat{m}_{\widehat{g}}^{(1)}(z; H_4) - \widehat{m}_g^{(1)}(z; H_4)\right) = o_p(1), \quad \textit{uniformly in } z$$

To proof this lemma we follow the same line as in the proof of Lemma 8.1.

LEMMA 8.4

$$\sqrt{N|H_4|^{1/2}} \left(\widehat{m}_g^{(1)}(z; H_4) - m(z) - B(z; H_4)\right) \xrightarrow{d} \mathcal{N}\left(0, V^{(1)}(z; H_4)\right),$$

where

$$B(z; H_4) = \mu_2(K_u)\left(diag_d\left(D_f(z)H_4\sqrt{N|H_4|^{1/2}}D_{m_r}(z)\right) \imath_d f_{Z_{it}}^{-1}(z) + \frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_r}(z)H_4\sqrt{N|H_4|^{1/2}}\right)\right)\imath_d\right),$$

$$V^{(1)}(z; H_4) = 2R(K_u)\left(\sigma_v^2 + m_1(z)^\top \Sigma_{\Delta\xi\Delta\xi} m_1(z) + \Sigma_{v\Delta\xi} m_1(z)\right)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z).$$

Proof of Lemma 8.4.

In order to show the results of Lemma 8.4 we need to prove the asymptotic behavior of (A.29). To this end, we first focus on the asymptotic bias of the one-step backfitting estimator and later on the corresponding variance. Using the results of Lemma 8.3 we know that the first element of the right-hand side is asymptotically cancel so we focus on the second one.

After substituting $W_{it} = g(X_{it}) + \zeta_i + \xi_{it}$ the Taylor's approximation of the smooth functions implies

$$\begin{aligned}
\Delta\widehat{Y}_{1it} &= \widetilde{W}_{it}^\top m(z) + \left(\widetilde{W}_{it} \otimes (Z_{it} - z)\right)^\top D_m(z) + \frac{1}{2}\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z) \\
&\quad + \widetilde{W}_{i(t-1)}^\top\left(\widehat{m}_{\widehat{g}}(Z_{i(t-1)}; H_2) - m(Z_{i(t-1)})\right) + \left(\widehat{\alpha}(Z_{i(t-1)}; H_3) - \alpha(Z_{i(t-1)})\right) \\
&\quad - \left(\widehat{\alpha}(Z_{it}; H_3) - \alpha(Z_{it})\right) + \Delta v_{it} + \xi_{it}^\top m_1(z) + o_p(1),
\end{aligned} \tag{A.30}$$

let $D_m(z)$, $\mathcal{H}_m(z)$ and $\widetilde{W}_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z)$ are defined in (A.12).

Combining (A.29) with (A.30) and substituting $W_{it} = g(X_{it}) + \zeta_i + \xi_{it}$ we can write $\widehat{m}_g^{(1)}(z; H_4)$ as

$$\widehat{m}_g^{(1)}(z; H_4) - m(z) = \left(\sum_{i=1}^N \sum_{t=2}^T K_{it}\widetilde{W}_{g,it}\widetilde{W}_{it}^\top\right)^{-1} \sum_{i=1}^N \sum_{t=2}^T K_{it}\widetilde{W}_{g,it}\left(G_{it}^{(1)} + Q_{it} + \Delta v_{it} + \Delta\xi_{it}^\top m_1(z)\right) \tag{A.31}$$

where now $K_{it} = |H_4|^{-1/2} K\left(H_4^{-1/2}(Z_{it} - z)\right)$,

$$G_{it}^{(1)} = \left(\widetilde{W}_{it} \otimes (Z_{it} - z)\right)^\top D_m(z) + \frac{1}{2}\left(\widetilde{W}_{it} \otimes (Z_{it} - z)\right)^\top \mathcal{H}_m(z)(Z_{it} - z) - \Delta\xi_{it}^\top m_1(z),$$

$$Q_{it} = \widetilde{W}_{i(t-1)}^\top \left(\widehat{m}_{\widehat{g}}(Z_{i(t-1)}; H_2) - m(Z_{i(t-1)})\right) - (\widehat{\alpha}(Z_{it}; H_3) - \alpha(Z_{it})) + \left(\widehat{\alpha}(Z_{i(t-1)}; H_3) - \alpha(Z_{i(t-1)})\right).$$

For the sake of simplicity, let us denote

$$\widehat{m}_g^{(1)}(z; H_4) - m(z) = \widetilde{S}_{n_1}^{-1}\left(\widetilde{B}_{n_1} + \widetilde{M}_{n_1} + \widetilde{U}_{n_1} + \widetilde{R}_{n_1}\right), \tag{A.32}$$

where

$$\begin{array}{ll}
\widetilde{B}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}G_{it}^{(1)}, & \widetilde{M}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}Q_{it}, \\
\widetilde{U}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\Delta\widetilde{W}_{g,it}\Delta v_{it}, & \widetilde{R}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}\Delta\xi_{it}^\top m_1(z), \\
\ddot{S}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}\widetilde{W}_{it}^\top. &
\end{array}$$

Therefore, to complete the proof of this lemma we have to show

$$\sqrt{N|H_4|^{1/2}}\left(\widehat{m}_g^{(1)}(z; H_4) - m(z)\right) - \sqrt{N|H_4|}\widetilde{S}_{n_1}^{-1}(\widetilde{B}_{n_1} + \widetilde{M}_{n_1}) = \sqrt{N|H_4|^{1/2}}\widetilde{S}_{n_1}^{-1}\left(\widetilde{U}_{n_1} + \widetilde{R}_{n_1}\right). \tag{A.33}$$

To obtain the bias term we first focus on the inverse term of (A.33) and later we analyze the behavior of $\widetilde{B}_{n_1}$ and $\widetilde{M}_{n_1}$. Then, following the same reasoning as in (A.3), we can show that as N tends to infinity

$$\widetilde{S}_{n_1}^{-1} = \mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z) + o_p(\|H_4^{1/2}\|), \tag{A.34}$$

since

$$\widetilde{S}_{n_1} = E\left[K_{it}\widetilde{W}_{g,it}\widetilde{W}_{it}^\top\right] = \mathcal{B}_{\widetilde{W}_g\widetilde{W}}(z) + o_p(1),$$

where

$$\mathcal{B}_{\widetilde{W}_g\widetilde{W}}(z) = E\left[\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it} = z\right]f_{Z_{it}}(z).$$

Focus now on $\ddot{B}_n$ we can split it up into three terms, i.e.

$$\widetilde{B}_{n_1} = (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}G_{it}^{(1)} = \widetilde{B}_{n_1}^{(1)} + \widetilde{B}_{n_1}^{(2)} - \widetilde{B}_{n_1}^{(3)}, \tag{A.35}$$

where

$$\begin{aligned}
\widetilde{B}_{n_1}^{(1)} &= (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}(\widetilde{W}_{it} \otimes (Z_{it} - z))^\top D_m(z), \\
\widetilde{B}_{n_1}^{(2)} &= (2NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}(\widetilde{W}_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)(Z_{it} - z), \\
\widetilde{B}_{n_1}^{(3)} &= (NT)^{-1}\sum_{it} K_{it}\widetilde{W}_{g,it}\Delta\xi_{it}^\top m_1(z).
\end{aligned}$$

Analyzing each of these terms separately we obtain that under standard properties of kernel density estimators, under the assumptions of Theorem 5.1 and as $N \to \infty$,

$$
\begin{aligned}
E(\widetilde{B}_{n_1}^{(1)}) &= E\left[K_{it}E(\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it}) \otimes (Z_{it}-z)^\top D_m(z)\right] \\
&= \mu_2(K_u)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}(z,z)diag_d\left(D_f(z)tr(H_4 D_{m_r}(z))\right)\imath_d f_{Z_{it}}^{-1}(z) + o_p(tr(H_4)), \quad \text{(A.36)} \\
E(\widetilde{B}_{n_1}^{(2)}) &= \frac{1}{2}E\left[K_{it}E(\widetilde{W}_{g,it}\widetilde{W}_{it}^\top|Z_{it}) \otimes (Z_{it}-z)^\top \mathcal{H}_{m_r}(z)(Z_{it}-z)\right] \\
&= \frac{1}{2}\mu_2(K_u)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}(z)diag_d\left(tr(\mathcal{H}_{m_r}(z)H_4)\right)\imath_d + o_p(tr(H_4)), \quad \text{(A.37)}
\end{aligned}
$$

whereas under similar reasoning it is straightforward to show $E(\widetilde{B}_{n_1}^{(3)}) = o_p(1)$. Also, assuming $H_4 \to 0$ and $N|H_4|^{1/2} \to \infty$ it is easy to prove that any component of the variance of $\widetilde{B}_{n_1}$ converge to zero.

To complete the proof of the asymptotic bias we have to analyze $\widetilde{M}_{n_1}$. Under assumptions of Theorem 4.1 we have proved that the bias rate of $\widehat{m}_{\widehat{g}}(z; H_2)$ is $o_p(tr(H_2))$, whereas in Qian and Wang (2012) is shown that this rate is $o_p(tr(H_3))$ for $\widehat{\alpha}(z_1; H_3)$. Following Masry (1996), under Assumptions **4.11** and **5.1** these rates are uniform in z and z_1, respectively, so under the same reasoning as the proof of Lemma 8.1 it is straightforward to show that as N tends to infinity

$$
\begin{aligned}
E(\widetilde{M}_{n_1}) &= E\left[K_{it}\widetilde{W}_{g,it}\left(\widetilde{W}_{i(t-1)}^\top(\widehat{m}_{\widehat{g}}(z; H_2) - m(z; H_2)) - (\widehat{\alpha}(z_1; H_3) - \alpha(z_1)) + (\widehat{\alpha}(z_2; H_3) - \alpha(z_2))\right)\right] \\
&= o_p(tr(H_2)) + o_p(tr(H_3)) + o_p(tr(H_3)), \quad \text{(A.38)}
\end{aligned}
$$

since $(NT)^{-1}\sum_{it}|K_{it}\widetilde{W}_{g,it}\widetilde{W}_{i(t-1)}^\top| = O_p(1)$ and $(NT)^{-1}\sum_{it}|K_{it}\widetilde{W}_{g,it}| = O_p(1)$.

Then, using the fact that $tr(H_2) \to 0$, $tr(H_3) \to 0$ and $tr(H_2) \to 0$ in the sense that $N|H_2| \to \infty$, $N|H_3| \to \infty$ and $N|H_4| \to \infty$, if we substitute the asymptotic results of (A.34) and (A.36)-(A.38) into the second term of (A.32) by the Cramer-Wold device we get the following expression for the bias of the one-step backfitting estimator

$$
\begin{aligned}
\widetilde{S}_{n_1}^{-1}(\widetilde{B}_{n_1} + \widetilde{M}_{n_1}) &= \mu_2(K_u)\left(diag_d\left(D_f(z)H_4 D_{m_r}(z)\right)\imath_d f_{Z_{it}}^{-1}(z) + \frac{1}{2}diag_d\left(tr(\mathcal{H}_{m_r}(z)H_4)\right)\imath_d\right) \\
&\quad + o_p(tr(H_4)). \quad \text{(A.39)}
\end{aligned}
$$

Therefore, it is proved that the bias rate of the one-step backfitting estimator (5.5) is the same as the corresponding of the local constant IV estimator (3.5), as we expected.

Focus now on the asymptotic variance we have that under assumptions of Theorem 5.1 and by the law of iterated expectations we obtain

$$
\begin{aligned}
N|H_4|^{1/2}Var(\widetilde{U}_{n_1}) &= |H_4|^{1/2}(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\widetilde{W}_{g,it}\widetilde{W}_{g,i't'}^\top \Delta v_{it}\Delta v_{i't'}K_{it}K_{i't'}\right] \\
&= 2\sigma_v^2 R(K_u)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)(1 + o_p(1)), \quad \text{(A.40)}
\end{aligned}
$$

where

$$
\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z) = E\left[\widetilde{W}_{g,it}\widetilde{W}_{g,it}^\top|Z_{it} = z\right]f_{Z_{it}}(z).
$$

39

Also, denote $\Sigma_{\Delta\xi\Delta\xi} = E(\Delta\xi_{it}\Delta\xi_{it}^\top)$ a $(M-1) \times (M-1)$ matrix and $\Sigma_{b_1\Delta v\Delta\xi} = E(\Delta v_{it}\Delta\xi_{it}^\top)$ as a $1 \times (M-1)$ vector we get that as $N \to \infty$,

$$
\begin{aligned}
N|H_4|^{1/2}Var(\widetilde{R}_{n_1}) &= |H_4|^{1/2}(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\widetilde{W}_{g,it}m_1(z)^\top\Delta\xi_{it}\Delta\xi_{i't'}^\top m_1(z)\widetilde{W}_{g,i't'}^\top K_{it}K_{i't'}\right] \\
&= 2R(K_u)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z)(1+o_p(1)), \qquad (A.41)
\end{aligned}
$$

and

$$
\begin{aligned}
N|H_4|^{1/2}Cov(\widetilde{U}_{n_1},\widetilde{R}_{n_1}) &= |H_4|^{1/2}(NT)^{-1}\sum_{ii'}\sum_{tt'}E\left[\widetilde{W}_{g,it}\Delta v_{it}\Delta\xi_{i't'}^\top m_1(z)\widetilde{W}_{g,i't'}^\top K_{it}K_{i't'}\right] \\
&= R(K_u)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)\Sigma_{\Delta v\Delta\xi}m_1(z)(1+o_p(1)). \qquad (A.42)
\end{aligned}
$$

Then, substituting (A.33) and (A.39)-(A.41) into the right-hand side of (A.32) and by the Cramer-Wold device we obtain that as $N|H_4| \to \infty$,

$$
\begin{aligned}
N|H_4|Var(\widetilde{S}_{n_1}^{-1}(\widetilde{U}_{n_1}+\widetilde{R}_{n_1})) &= 2R(K_u)\left(\sigma_v^2 + m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right) \\
&\quad \times\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z). \qquad (A.43)
\end{aligned}
$$

Finally, to prove the asymptotic normality of the one-step backfitting estimator (5.5) is necessary to check the Lyapunov condition. For it, we can write

$$
\sqrt{N|H_4|^{1/2}}(NT)^{-1}\sum_{it}K_{it}\widetilde{W}_{g,it}\left(\Delta v_{it} + \Delta\xi_{it}^\top m_1(z)\right) = \frac{1}{\sqrt{NT}}\sum_{it}\lambda_{b_1it}, \qquad (A.44)
$$

where $\lambda_{b_1it} = K_{it}\widetilde{W}_{g,it}\left(\Delta v_{it}\Delta\xi_{it}^\top m_1(z)\right)|H_4|^{1/2}$. Then, following the same structure as in the proof of Theorem 4.1 and by Assumption **5.2**

$$
\sqrt{N|H_4|^{1/2}}\left(\widehat{n}_g^{(1)}(z;H_4) - m(z) - B(z,H_4)\right) \xrightarrow{d} \qquad (A.45)
$$
$$
\mathcal{N}\left(0, 2R(K_u)\left(\sigma_v^2 + m_1(z)^\top\Sigma_{\Delta\xi\Delta\xi}m_1(z) + \Sigma_{\Delta v\Delta\xi}m_1(z)\right)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_g}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z)\right)
$$

and the proof is done.

■

Proof of Theorem 5.2

By substracting in both terms of (5.9) the quantity $m(z)$ and noting that $\mathcal{G}^{-1}(\mathcal{J}_1 + \mathcal{J}_2) = I$ we obtain

$$
\begin{aligned}
\sqrt{N|H_4|^{1/2}}(\widetilde{m}(z;H_4) - m(z)) &= \mathcal{G}^{-1}\mathcal{J}_1\sqrt{N|H_4|^{1/2}}\left(\widehat{m}_g^{(1)}(z;H_4) - m(z)\right) \\
&\quad + \mathcal{G}^{-1}\mathcal{J}_2\sqrt{N|H_4|^{1/2}}\left(\widehat{m}_g^{(2)}(z;H_4) - m(z)\right), \qquad (A.46)
\end{aligned}
$$

where for the sake of simplicity,

$$
\begin{aligned}
\mathcal{G} &= \Omega_{m_{11}}^{(1)} + 2\Omega_{m_{21}}^{(1)} + \Omega_{m_{22}}^{(1)}, \qquad (A.47) \\
\mathcal{J}_1 &= \Omega_{m_{11}}^{(1)} + \Omega_{m_{21}}^{(1)}, \qquad (A.48) \\
\mathcal{J}_2 &= \Omega_{m_{12}}^{(1)} + \Omega_{m_{22}}^{(1)}. \qquad (A.49)
\end{aligned}
$$

40

If we can prove that $\Omega_{m_{12}} = o(1)$, as N tends to infinity, applying Theorem 4.1 and Theorem A (Serfling (1980), p. 122) the desired result is shown.

Now we prove that $\Omega_{m_{12}} = o(1)$, as N tends to infinity. If we combine the righthand side of (A.32) for $\widehat{m}_{\widehat{g}}^{(1)}(z; H_4)$ with the corresponding expression for $\widehat{m}_{\widehat{g}}^{(2)}(z; H_4)$ we obtain that

$$\Omega_{m_{12}} = N|H_4|^{1/2} Cov\left(\widetilde{S}_{n_1}\left(\widetilde{U}_{n_1} + \widetilde{R}_{n_1}\right), -\widetilde{S}_{n_2}\left(\widetilde{U}_{n_2} + \widetilde{R}_{n_2}\right)\right), \tag{A.50}$$

where for $K_{it} = |H_4|^{-1/2} K\left(H_4^{-1/2}(Z_{it} - z)\right)$ and $K_{i(t-1)} = |H_4|^{-1/2} K\left(H_4^{-1/2}(Z_{i(t-1)} - z)\right)$,

$$\widetilde{S}_{n_2} = (NT)^{-1}\sum_{it} K_{i(t-1)}\widetilde{W}_{g,i(t-1)}\widetilde{W}_{i(t-1)}^{\top}, \qquad \widetilde{U}_{n_2} = -(NT)^{-1}\sum_{it} K_{i(t-1)}\widetilde{W}_{g,i(t-1)}\Delta v_{it},$$
$$\widetilde{R}_{n_2} = -(NT)^{-1}\sum_{it} K_{i(t-1)}\widetilde{W}_{g,i(t-1)}\Delta\xi_{it}^{\top}m_1(z).$$

If we focus on the middle term of (A.50), under assumptions of Theorem 5.1 we can prove that by the law of iterated expectations,

$$N|H_4|^{1/2}Cov(\widetilde{U}_{n_1}, \widetilde{U}_{n_2}) = -|H_4|^{1/2}(NT)^{-1}\sum_{it} E\left[K_{it}K_{i(t-1)}\widetilde{W}_{g,it}\Delta v_{it}^2\widetilde{W}_{g,i(t-1)}^{\top}\right]$$
$$= -2\sigma_v^2|H_4|^{1/2}\mathcal{B}_{\widetilde{W}_g\widetilde{W}_{g-1}}(z,z)(1 + o_p(1)), \tag{A.51}$$

$$N|H_4|^{1/2}Cov(\widetilde{U}_{n_1}, \widetilde{R}_{n_2}) = |H_4|^{1/2}(NT)^{-1}\sum_{it} E\left[K_{it}K_{i(t-1)}\widetilde{W}_{g,it}\Delta v_{it}m_1(z)^{\top}\Delta\xi_{it}\widetilde{W}_{g,i(t-1)}^{\top}\right]$$
$$= |H_4|^{1/2}\mathcal{B}_{\widetilde{W}_g\widetilde{W}_{g-1}}(z,z)\Sigma_{\Delta v\Delta\xi}m_1(z)(1 + o_p(1)), \tag{A.52}$$

where

$$\mathcal{B}_{\widetilde{W}_g\widetilde{W}_{g-1}}(z,z) = E\left[\widetilde{W}_{g,it}\widetilde{W}_{g,i(t-1)}^{\top}|Z_{it} = z, Z_{(t-1)} = z\right]f_{Z_{it},Z_{i(t-1)}}(z,z).$$

Similarly,

$$N|H_4|^{1/2}Cov(\widetilde{R}_{n_1}, \widetilde{R}_{n_2}) = |H_4|^{1/2}(NT)^{-1}\sum_{it} E\left[K_{it}K_{i(t-1)}\widetilde{W}_{g,it}\Delta\xi_{it}^{\top}m_1(z)m_1(z)^{\top}\Delta\xi_{it}\widetilde{W}_{g,i(t-1)}^{\top}\right]$$
$$= |H_4|^{1/2}\mathcal{B}_{\widetilde{W}_g\widetilde{W}_{g-1}}(z,z)m_1(z)^{\top}\Sigma_{\Delta\xi\Delta\xi}m_1(z)(1 + o_p(1)). \tag{A.53}$$

In addition, under the same reasoning as in (A.34)

$$\ddot{S}_{n_2}^{-1} = \mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}^{-1}(z) + o_p(\|H_4^{1/2}\|), \tag{A.54}$$

where

$$\mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}(z) = E\left[\widetilde{W}_{g,i(t-1)}\widetilde{W}_{i(t-1)}^{\top}|Z_{i(t-1)} = z\right]f_{Z_{i(t-1)}}(z).$$

If we substitute (A.34) and (A.51)-(A.54) into (A.50) and by the Cramer-Wold device we obtain that as $N \to \infty$

$$\Omega_{m_{12}} = |H_4|^{1/2}\left(2\sigma_v^2 - m_1(z)^{\top}\Sigma_{\Delta\xi\Delta\xi}m_1(z)\right)\mathcal{B}_{\widetilde{W}_g\widetilde{W}}^{-1}(z)\mathcal{B}_{\widetilde{W}_g\widetilde{W}_{g-1}}(z,z)\mathcal{B}_{\widetilde{W}_{g-1}\widetilde{W}_{-1}}^{-1}(z)(1 + o_p(1)).$$

And the proof is done.

■

41

References

Aaronson, D., S. Agarwal, and E. French (2012). "The spending and debt response to minimum wage hikes". *American Economic Review 102*(7), 3111–3139.

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.

Blanchard, O. 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), 859–971.

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. 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.

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.

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.

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. and I. Gijbels (1995). *"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.

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.

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.

Härdle, W., S. Huet, E. Mammen, and S. Sperlich (2004). "Bootstrap inference in semiparametric generalized additive models". *Econometric Theory 20*, 265–300.

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.

Hubbard, G., J. S. Skinner, and S. P. Zeldes (1994). "The importance of precautionary motives for explaining individual and aggreagate 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.

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*.

Linton, O. and J. P. Nielsen (1995). "A kernel method of estimating structured nonparametric regression based on marginal integration". *Biometrika 82*, 93–101.

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.

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.

Palumbo, M. G. (1999). "Uncertain medical expenses and precautionary saving: near the end of the life cycle". *Review of Economic Studies 66*, 395–422.

Qian, J. and L. Wang (2012). "Estimating semiparametric panel data models by marginal integration". *Journal of Econometrics 167*, 483–493.

Rodriguez-Poo, J. M. and A. Soberon (2014). "Direct semi-parametric estkmation of fixed effects panel data varying coefficient models". *Econometrics Journal 17*, 107–138.

Rodriguez-Poo, J. M. and A. Soberon (2015). "Nonparametric estimation of fixed effects panel data varying coefficient models". *Journal of Multivariate Analysis 133*, 95–122.

Ruppert, D. and M. P. Wand (1994). "Multivariate locally weighted least squares regression". *The Annals of Statistics 22*(3), 1346–1370.

Serfling, R. J. (1980). *"Approximation theorems of mathematical statistics"*. John Wiley and Sons, Inc.

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 (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.

Watson, G. S. (1964). "Smooth regression analysis". *Sankhya: The Indian Journal of Statistics 26*(4), 359–372.