

Nonparametric Regression with a Latent Time Series

Oliver Linton*
(London School of Economics)

Jens Perch Nielsen†
(Cass Business School)

Søren Feodor Nielsen‡
(University of Copenhagen)

DP No: EM 2009 538
October 2009

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6674

* Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk. Thanks to the ESRC and Leverhulme foundations for financial support. This paper was partly written while I was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support.

† Cass Business School, London. Email address: Jens.Nielsen.1@city.ac.uk

‡ Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark. E-mail address: feodor@stat.ku.dk

Abstract

In this paper we investigate a class of semiparametric models for panel datasets where the cross-section and time dimensions are large. Our model contains a latent time series that is to be estimated and perhaps forecasted along with a nonparametric covariate effect. Our model is motivated by the need to be flexible with regard to functional form of covariate effects but also the need to be practical with regard to forecasting of time series effects. We propose estimation procedures based on local linear kernel smoothing; our estimators are all explicitly given. We establish the pointwise consistency and asymptotic normality of our estimators. We also show that the effects of estimating the latent time series can be ignored in certain cases.

Keywords and Phrases: Kernel Estimation; Forecasting; Panel Data; Unit Roots

AMS (2000) subject classification: 62G07.

1 Introduction

Panel data are found in many contexts. Traditionally, it is associated with a series of household surveys conducted over time on the same individuals for which the cross-sectional dimension is large and the time series dimension is short. Parametric methods appropriate for this kind of data can be found in Hsiao (1986). There has also been some work on semiparametric models for such data, see for example Kyriazidou (1997), and nonparametric additive models, Porter (1996). The increase in the length of time series available for these data has led to some interest in the application of time series concepts, see for example Arellano (2003). More recently, there has been work on panel data with large cross-section and time series dimension, especially in finance where the datasets can be large along both dimensions and in macro where there are many series with modest length time series. Some recent works include Phillips and Moon (1999), Bai and Ng (2002), Bai (2003,2004), and Pesaran (2006). These authors have addressed a variety of issues including nonstationarity, estimation of unobserved factors, and model selection. They all work with essentially parametric models.

In this paper we investigate a class of semiparametric models for such datasets. Our model contains a latent time series that is to be estimated along with a nonparametric covariate effect. Our model is motivated by the need to be flexible with regard to functional form of covariate effects but also the need to be practical with regard to forecasting of time series effects. Our main contribution in fact is to provide results that support subsequent time series analysis on the latent time series, and for this purpose it is desirable and important to not require the latent time series to be stationary. Our framework is consistent with the influential model of Carter and Lee (1992) for US mortality. Some other related works in econometrics include Connor and Linton (2002), who applied a similar model to a large financial panel dataset. See also Fengler, Härdle and Mammen (2006) and Mammen, Støve, and Tjøstheim (2006). We propose estimates of the nonparametric component and the latent time series that are based on least squares objective functions and are defined in closed form. We establish the pointwise asymptotic distribution of our estimator of the nonparametric component and the joint distribution of the estimated latent time series in the case where the time series length is fixed. We then establish some properties in the case where the time series length increases to infinity at some rate. In many cases one wants to do further modelling of the latent time series with a view to forecasting future values. We prove that the estimated latent time series is close enough to the true latent time series such that the estimation error can be ignored in such future analysis.

We give an application on simulated data.

The paper is organized as follows. In section 2 we describe our model, while in section 3 we introduce our estimators of the key components. In section 4 we give the asymptotic properties of the estimates, while in section 5 we investigate the application of our estimates to further modelling strategies. We give some numerical evidence on the finite sample performance of our procedures in section 6, while section 7 concludes. All proofs are in the appendix.

2 Model

We suppose that the data are generated as an unbalanced panel:

$$Y_{i,t} = \theta_t + g(X_{i,t}) + u_{i,t}, \quad i = 1, \dots, n_t, t = 1, \dots, T, \quad (1)$$

where the unobserved errors $(u_{i,t})_{i,t}$ satisfy at least the conditional moment restriction $E[u_{i,t}|X_{i,t}, \theta_t] = 0$. Here, $(\theta_t)_t$ is an unobserved time series, while $(X_{i,t})_{i,t}$ are observed covariates. We shall assume throughout that $(\theta_t)_t$ is independent of the observed covariates and errors. The distribution theory requires additional conditions on the errors and the covariates to ensure that laws of large numbers and central limit theorems hold, we discuss this further below. The model is a semiparametric panel data model and some aspects of this have been discussed recently in for example Fan and Li (2004), Fan, Huang, and Li (2007), and Mammen, Støve, and Tjøstheim (2006), although our assumptions will be more general in some cases and our focus is different. In particular, the focus of our paper is on the latent time series $(\theta_t)_t$ itself. In practice we expect the distribution of observed and unobserved variables to change over time, and this is allowed for in our model. For example, we wish to allow the covariates to have potentially time-varying densities f_t , i.e.,

$$X_{i,t} \sim f_t, \quad i = 1, \dots, n_t. \quad (2)$$

This is different from most previous treatments of this model.

The model can also be thought of as an additive nonparametric regression model in covariates t/T and $X_{i,t}$ except that the function $t \mapsto \theta_t$ is not assumed to be smooth or even continuous, so most extant theory for additive regression models cannot be applied.

Our aim is to estimate the unknown smooth regression function $g(\cdot)$ and the time series $(\theta_t)_t$ from a sample $\{Y_{it}, X_{it}, i = 1, \dots, n_t, t = 1, \dots, T\}$. We allow the datasets to be unbalanced: the number

of observations in each time period, denoted n_t , and the number of time periods for each observation, denoted T_i , are allowed to vary freely but are assumed independent of all other randomness.

Observe that the mean of $Y_{i,t}$ is

$$E[Y_{i,t}] = E[\theta_t] + \int g(x)f_t(x)dx. \quad (3)$$

Without further restrictions, the mean of the latent process $\{\theta_t\}$ and the function $g(\cdot)$ are not separately identified. Clearly, we may subtract a constant from θ_t and add it to the function g without changing the distribution of the observed data. In the context of additive models, for example Linton and Nielsen (1995), it is common to assume that $E[g(X)] = 0$. However, since we wish to allow for the possibility that the covariate distribution is nonstationary, this is not an attractive assumption. One could instead assume that for example $E[g(X_{i,1})] = 0$, which would be consistent with nonstationary covariates. We instead put restrictions on the process $\{\theta_t\}$. A restriction on the mean of θ_t would effectively rule out nonstationarity in that component. Therefore, we shall impose that $\theta_1 = 0$ (one could choose an arbitrary initial value instead, if this has better interpretation). This is consistent with the process $\{\theta_t\}$ being a unit root process starting from the origin. It also allows the process $\{\theta_t\}$ to be asymptotically stationary. We remark that there is an air of arbitrariness in the decomposition between θ_t and $g(X_{i,t})$ and whatever restriction is imposed cannot get around this. The quantity $\varphi_t = E(Y_{it}|\theta_t, X_{it}) = \theta_t + g(X_{it})$ is invariant to the choice of identifying restriction. However, the quantity φ_t contains two sources of nonstationarity though, θ_t and the changing mean of g due to the changing covariate distribution. It is of interest to separate out these two sources of nonstationarity by examining separately θ_t and f_t .

We close this section with some motivation for considering the model (1). The model captures the general idea of an underlying and unobserved trend modifying the effect of a covariate on a response. For example, suppose that output of a firm Q is determined by inputs capital K and labour L but the production function F is subject to technological change a that affects all firms in the industry. This could be captured by the deterministic equation $Q = aF(K, L)$. Taking logs and adding a random error yields the specification (1) for $Y_{it} = \log Q_{it}$, $\theta_t = \log a_t$, and $g(\cdot) = \log F(\cdot)$. Note that $\partial \log Q / \partial \log a = 1$, and this specification imposes the so-called Hicks Neutral technical change. In this case, the Total Factor Productivity or Solow Residual is $\theta'(t)$, the part of growth not explainable by measurable changes in the inputs. In the popular special case where the production function is homothetic, one can replace $F(K_{it}, L_{it})$ by $f(X_{it})$, where X_{it} is the scalar capital to labour ratio.

Traditional econometric work chose particular functional forms for F like Cobb-Douglas or CES, and made θ_t a polynomial function of time. However, there is not general agreement on the form of production functions, see Jorgensen (1986), and so it is well motivated to treat g as a nonparametric function. Likewise it is restrictive to assume a particular form that underpins how the technology should change and so we do not restrict the relationship $t \mapsto \theta_t$. The model assumption that $\theta_1 = 0$ has a natural interpretation in this case as it corresponds to $a_1 = 1$, in which case $Q_{i1} = F(K_{i1}, L_{i1})$ is a baseline level of production.

3 Estimation

We next present several methods for estimation of the unknown quantities. All our methods are based on minimizing sample sums of squared residuals. This has several advantages: it leads to closed form estimators; it only requires the conditional moment restriction $E(u|X, \theta) = 0$ for consistency; it usually implies an efficient procedure under i.i.d. normal error terms as has been noted in earlier work. We also adopt the local linear regression paradigm because of its many advantages, Fan and Gijbels (1996). Extension to the local polynomial case is straightforward conceptually. Our estimation method is related to that considered in the paper of Mammen, Stove, and Tjostheim (2006) except that we consider different identification restrictions, which leads to a slightly different procedure. They consider a more general model with multiple covariates that enter in an additive fashion, which makes their procedure more complicated to describe. Also, they do not provide results for estimation of the latent time series, which is perhaps the main contribution of this paper.

We estimate $(g(x), g'(x))$ for each x in a set \mathcal{X} and $\theta = (\theta_t)_{t=2, \dots, T}$ by minimizing the following integrated weighted sum of squares:

$$\begin{aligned} & \int \sum_{t=1}^T \sum_{i=1}^{n_t} (Y_{i,t} - \theta_t - g(x) - g'(x)(X_{i,t} - x))^2 K\left(\frac{X_{i,t} - x}{h_t}\right) d\nu(x) \\ &= \int \left(Y - A\theta - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right)^\top K_{x,h} \left(Y - A\theta - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) d\nu(x), \end{aligned} \tag{4}$$

for some suitable measure ν concentrated on \mathcal{X} where $\theta_1 = 0$, $Y = (Y_{i,t})_{i=1, \dots, n_t, t=1, \dots, T}$, while A and B_x are suitable “design matrices” of dimension $N \times (T-1)$ and $N \times 2$ respectively, where $N = \sum_{t=1}^T n_t$. The rows in B_x are of the form $[1 \ X_{i,t} - x]$ and the typical row in A has a 1 in the

t -1st place and zeros elsewhere if the row corresponds to an observation $Y_{i,t}$ from the t th time period for $t = 2, \dots, T$; rows corresponding to observations $Y_{i,1}$ from the first time period have all element equal to 0. Finally, $K_{x,h}$ is the $N \times N$ diagonal matrix with diagonal elements $K((X_{i,t} - x)/h_t)$, where h_t is a bandwidth sequence and K is a kernel function. For any fixed value of θ the integrated sum of squares is minimized by minimizing the integrand. This leads to

$$\begin{bmatrix} \hat{g}(x) \\ \hat{g}'(x) \end{bmatrix} = (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} (Y - A\theta). \quad (5)$$

We note that this is just the (pooled) local linear regression of $Y - A\theta$ on X in the point x . Plugging this expression into (4) yields

$$\begin{aligned} & \int (Y - A\theta - W_{x,h}(Y - A\theta))^\top K_{x,h}(Y - A\theta - W_{x,h}(Y - A\theta)) d\nu(x) \\ &= (Y - A\theta)^\top \int K_{x,h}(I_N - W_{x,h}) d\nu(x) (Y - A\theta) \end{aligned}$$

where $W_{x,h} = B_x(B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h}$. Minimizing this as a function of θ yields the weighted least squares estimator

$$\begin{aligned} \hat{\theta} &= \left(A^\top \int K_{x,h}(I_N - W_{x,h}) d\nu(x) A \right)^{-1} A^\top \int K_{x,h}(I_N - W_{x,h}) d\nu(x) Y \\ &= \left(\int A^\top K_{x,h}(I_N - W_{x,h}) A d\nu(x) \right)^{-1} \int A^\top K_{x,h}(I_N - W_{x,h}) Y d\nu(x) \end{aligned} \quad (6)$$

The integrals in the matrices in (6) are one-dimensional and can be computed by standard numerical integration routines. Moreover, simple expressions can be given for the matrices $A^\top K_{x,h}(I_N - W_{x,h})A$ (see (21)) and $A^\top K_{x,h}(I_N - W_{x,h})Y$. Plugging (6) into (5) then gives the estimator of $g(x)$ (and $g'(x)$). It is worth noting that having derived the estimator of θ and the estimator of g as a solution to a least squares problem does not prevent us from using different x 's or another set of bandwidths (or even another choice of kernels) in the final estimation of g . This may be quite useful in some situations, perhaps especially when predicting future observations Y_{T+s} corresponding to a new covariate value x .

4 Asymptotic results

In this section we give some asymptotic properties of our estimators. Our main focus is in the estimation of the latent time series $(\theta_t)_t$ but we also provide results for the estimator of g . The

first properties we give hold for the large N and fixed T case. Here we give the joint asymptotic distribution of the estimation error for the time series. We then give some results for the case where both quantities grow. Here the focus is on sufficient conditions that allow us to apply standard asymptotic results from time series theory to the estimated time series.

4.1 Asymptotic results when T is fixed

We use the following regularity conditions, which as usual are sufficient but not necessary for our results.

ASSUMPTION A.

1. *Suppose that X_{it} are independent across i and t , and identically distributed across i , while $u_{i,t} = \sigma_t(X_{i,t})\epsilon_{i,t}$, where $\epsilon_{i,t}$ are i.i.d. with mean zero and variance one and independent of X_{it} .*
2. *Suppose that $d\nu(x) = \omega(x)dx$ for some density ω and that ν has compact support \mathcal{X} .*
3. *Suppose that g is twice differentiable on the compact set $\mathcal{X} \subset \cap_t\{x : f_t(x) > 0\}$, and satisfies $|g''(x) - g''(y)| \leq C|x - y|$ for some constant C . The marginal densities f_t are (uniformly over t) continuous and strictly positive throughout \mathcal{X} . The conditional variance functions σ_t^2 are (uniformly over t) continuous and strictly positive throughout \mathcal{X} .*
4. *Suppose that K is a Lipschitz-continuous density function symmetric about zero (a second order kernel) with compact support. Define $\|K\|_2^2 = \int K(u)^2 du$ and $\mu_j(K) = \int K(u)u^j du$.*
5. *Suppose that $N = \sum_{s=1}^T n_s \rightarrow \infty$ such that $n_t/N \rightarrow \lambda_t \in [\underline{\lambda}, \bar{\lambda}] \subset (0, \infty)$ for each $t = 1, \dots, T$.*
6. *There exists a sequence $h = h(N)$ such that $h_t/h \rightarrow b_t$, where $b_t \in [\underline{b}, \bar{b}] \subset (0, \infty)$ for all t , while $h \rightarrow 0$ and $Nh^5 \rightarrow 0$.*

We have maintained strong assumptions with regard to the errors. In principle, one can allow both cross-sectional dependence and time series dependence in the errors and most of our results go through with some modification of the limiting variances in some cases. However, note that the model itself induces cross-sectional and time series dependence in $Y_{i,t}$. We are assuming that the number of observations in each time period is of similar magnitude; this can be weakened but at the expense of a more complicated theory. It seems like a reasonable assumption to make here. In assumption

A6 h may be chosen to be any of the bandwidths h_1, \dots, h_T . Note that since the distribution of covariates and errors may differ from time period to time period it may in practice be very useful to have different bandwidths in each time period. The other assumptions are quite standard in the nonparametric literature. In the setup of this section, where T is fixed, the uniformity in t required in assumption A3 is just an assumption for each t . However, we will use the assumption again in section 5 where $T \rightarrow \infty$ and here some sort of uniformity is required.

We need to define some quantities that are important in the results. Define the $(T-1) \times (T-1)$ matrix $D(x)$ with elements

$$D(x)_{t,t'} = \begin{cases} f_{t+1}(x) \left(1 - \frac{\lambda_{t+1} b_{t+1} f_{t+1}(x)}{\sum_{s=1}^T \lambda_s b_s f_s(x)} \right) & \text{if } t = t' \\ -\frac{\sqrt{\lambda_{t+1} b_{t+1} f_{t+1}(x)} \sqrt{\lambda_{t'+1} b_{t'+1} f_{t'+1}(x)}}{\sum_{s=1}^T \lambda_s b_s f_s(x)} & \text{if } t \neq t'. \end{cases} \quad (7)$$

Under assumptions (A3) and (A5) the matrix $D(x)$ is strictly positive definite for $x \in \mathcal{X}$: If we let v be the $(T-1)$ vector with elements v_2, v_3, \dots, v_T , where $v_t = \lambda_t b_t f_t(x)$ for $t = 1, \dots, T$ and let $V = \text{diag}(v)$, then $D(x)$ may be written as $\Lambda^{-1/2} B^{-1/2} \bar{D}(x) B^{-1/2} \Lambda^{-1/2}$, where $\Lambda = \text{diag}\{\lambda_2, \dots, \lambda_T\}$, $B = \text{diag}\{b_2, \dots, b_T\}$, and

$$\bar{D}(x) = V - \frac{1}{\sum_{s=1}^T v_s} v v^\top.$$

We note that

$$\bar{D}(x)^{-1} = V^{-1} + \frac{1}{v_1} i_{T-1} i_{T-1}^\top,$$

which can easily be checked, see Berry, Linton and Pakes (2004) for some results on this type of matrices. In particular, $\bar{D}(x)^{-1}$ and therefore also $\bar{D}(x)$ and $D(x)$ are strictly positive definite.

Define also the $(T-1) \times T$ -matrix

$$C(x) = [0 | I_{T-1}] - \frac{v i_T^\top}{\sum_{s=1}^T v_s}, \quad (8)$$

where $i_T = (1, 1, \dots, 1)^\top \in \mathbb{R}^T$. Then let $\bar{C}(x) = B^{-1/2} \Lambda^{-1/2} C(x) \bar{\Lambda}^{1/2} \bar{B}^{1/2}$, where $\bar{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_T\}$, $\bar{B} = \text{diag}\{b_1, \dots, b_T\}$, and define

$$\Omega(x) = \bar{C}(x) \Psi(x) \bar{C}(x)^\top, \quad (9)$$

where $\Psi(x) = \text{diag}\{\sigma_1^2(x) f_1(x), \dots, \sigma_T^2(x) f_T(x)\}$.

Let $\Delta_T = \text{diag}\{n_2, \dots, n_T\}$ and $H_T = \text{diag}\{h_2, \dots, h_T\}$.

Theorem 1 *Suppose that assumptions A1-A6 hold. Then*

$$\Delta_T^{1/2} (\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, \Phi^{-1} \Xi \Phi^{-1}), \quad (10)$$

where

$$\Phi = B^{-1/2} \int D(x) \omega(x) dx B^{1/2} \quad \text{and} \quad \Xi = \int \Omega(x) \omega(x)^2 dx.$$

The asymptotic variance is a bit unusual for a semiparametric quantity in that the bandwidth constant matrix B enters the limiting variance. This is due to the fact that we have allowed different bandwidths in each time period; with a single choice of bandwidth this term cancels out. We discuss the form of the limiting variance more below. Consistent standard errors can be obtained by estimating the unknown quantities in the asymptotic variance by consistent estimators. A simpler approach is to work off the leading terms in the asymptotic expansion of the estimator as follows. Let

$$\begin{aligned} \widehat{\Phi} &= H_T^{-1/2} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) A \omega(x) dx \Delta_T^{-1/2} H_T^{-1/2} \\ \widehat{\Xi} &= \widehat{\mathcal{W}} \text{diag}\{\widehat{u}_{it}^2\} \widehat{\mathcal{W}}^\top \\ \widehat{\mathcal{W}} &= H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) \omega(x) dx, \end{aligned}$$

where $\widehat{u}_{it} = Y_{i,t} - \widehat{\theta}_t - \widehat{g}(X_{i,t})$ are nonparametric residuals.

We conclude this section with a discussion of the limiting variance (10). Consider the special case where $\sigma_t^2(x) = \sigma^2(x)$ for all t , $f_t(x) = f(x)$ for all t , and $\lambda_t = 1/T$. Then

$$\Phi^{-1} \Xi \Phi^{-1} = \frac{\int \sigma^2(x) f(x) \omega(x)^2 dx}{(\int f(x) \omega(x) dx)^2} [I_{T-1} + i_{T-1} i_{T-1}^\top].$$

If we knew the function g , then we would estimate θ_t by

$$\tilde{\theta}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (Y_{it} - g(X_{it})), \quad t = 1, \dots, T, \quad (11)$$

which satisfies

$$\left[\sqrt{n_1}(\tilde{\theta}_1 - \theta_1), \dots, \sqrt{n_T}(\tilde{\theta}_T - \theta_T) \right]^\top \xrightarrow{\mathcal{D}} N(0, \Sigma),$$

where $\Sigma = \text{diag}(\int \sigma_t^2(x) f_t(x) dx)$. In the special case considered above, $\Sigma = \int \sigma^2(x) f(x) dx I_{T-1}$. Of course this is an unfair comparison in view of the identification issue. If instead of knowing g we

know g up to an additive constant, α , (11) would estimate $\theta_t + \alpha$ instead of θ_t . Assuming as above that $\theta_1 = 0$ we would estimate θ_t by

$$\tilde{\theta}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (Y_{it} - g(X_{it})) - \frac{1}{n_1} \sum_{i=1}^{n_1} (Y_{i1} - g(X_{i1})), \quad t = 2, \dots, T,$$

with asymptotic distribution

$$\Delta_T^{1/2} (\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N \left(0, \int \sigma^2(x) f(x) dx [I_{T-1} + i_{T-1} i_{T-1}^\top] \right)$$

Observe that we may get arbitrarily close to this asymptotic variance by choosing \mathcal{X} to be a large compact subset of $\{x : f(x) > 0\}$ and letting $\omega(x) = 1$ in Theorem 1. Thus, the lack of efficiency of our estimator of θ_t is more due to the unidentifiability than to the unknown regression function g .

It follows from Theorem 1 that we can write $\hat{\theta} = \theta + (\hat{\theta} - \theta)$, where the two terms on the right hand side are asymptotically independent and the latter term is asymptotically $N(0, \Delta_T^{-1/2} \Phi^{-1} \Xi \Phi^{1/2} \Delta_T^{-1/2})$ -distributed. Hence, when n_t is large we may either model the estimated time series and from this derive a model for the latent time series, or – if n_t is sufficiently large so that the prediction error is negligible – use the estimated time series as if it were the latent time series.

4.2 Asymptotics for the estimator of g

Theorem 2 *Suppose that assumptions A1-A6 hold. Then*

$$\sqrt{N}h \left(\hat{g}(x) - g(x) - \frac{h^2}{2} \mu_2 \frac{\sum_{t=1}^T \lambda_t b_t^3 f_t(x)}{\sum_{t=1}^T \lambda_t b_t f_t(x)} g''(x) \right) \xrightarrow{\mathcal{D}} N \left(0, \|K\|_2^2 \frac{\sum_{s=1}^T \lambda_s b_s \sigma_s^2(x) f_s(x)}{\left(\sum_{s=1}^T \lambda_s b_s f_s(x) \right)^2} \right),$$

provided $\sqrt{N}hh^3 \rightarrow 0$ and $\sqrt{N}hh^2 r_N \rightarrow 0$, where $r_N = \max_{s=1, \dots, T} (h_s + \sqrt{\log n_s / (n_s h_s)})$.

Consistent standard errors can be obtained by estimating the unknown quantities in the asymptotic variance in the usual way, Fan and Gijbels (1996) and Fan and Yao (2003). In particular we note that the constants b_1, \dots, b_T and $\lambda_1, \dots, \lambda_T$ in practice may be replaced by h_t/h and n_t/N respectively.

If we knew the process θ_t we would estimate the function g from the pooled nonparametric regression of $Y_{it} - \theta_t$ on X_{it} . This satisfies the same CLT. In the special case where $\sigma_t^2(x) = \sigma^2(x)$ for all t , $f_t(x) = f(x)$ for all t , and $\lambda_t = 1/T$, the asymptotic variance is $T \|K\|_2^2 \sigma^2(x) / f(x)$.

5 Time series analysis

If one observed the time series $\theta_t, t = 1, \dots, T$, where T is large, the usual econometric approach would be to specify a model for it, thereby enabling description and forecasting. For example, suppose that θ_t follows an ARIMA(p,d,q) process with slowly varying mean, $A(L)(1-L)^d\theta_t = \mu(t/T) + B(L)\sigma(t/T)\zeta_t$, where $\mu(\cdot)$ and $\sigma(\cdot)$ are smooth functions on $[0, 1]$, ζ_t is a white noise process, while $A(L) = \sum_{j=0}^p a_j L^j$ and $B(L) = \sum_{j=0}^q b_j L^j$ are lag polynomials with roots outside the unit circle. Here, d is an integer denoting the order of nonstationarity. This is a convenient class of models for forecasting; it is just one (quite general) class of discrete time models that allows a certain type of nonstationary behaviour, others can be contemplated. The properties of estimators in such models generally rely on a long time series so that $T \rightarrow \infty$.

Our previous results can be formally extended to this case, although in an extension of Theorem 1, one would have to consider finite dimensional linear combinations of the expanding parameter vector. Instead, we address the issue of the impact of estimating the time series θ_t on inference about the parameters that govern its dynamic evolution. Hansen, Nielsen, and Nielsen (2004) consider the general problem of using estimated values in time series models. They prove a general result that provided

$$\sum_{t=2}^T (\hat{\theta}_t - \theta_t)^2 \xrightarrow{P} 0 \quad (12)$$

as $T \rightarrow \infty$, then we may use the estimated time series as if it was the true unobserved time series for instance in estimation and unit root testing in the sense that using the estimated values leads to the same asymptotic distribution (for $T \rightarrow \infty$) as if the true values were used. It is understood that the limits here are taken pathwise so that N and T approach infinity at some rate.

We next show that this property also holds in our case with a nonparametric covariate effect. As we now consider the case of $T \rightarrow \infty$ and $\min_{s=1, \dots, T} n_s \rightarrow \infty$, we need additional assumptions. When $T \rightarrow \infty$ we must have $n_t/N \rightarrow 0$ if not for all then at least for some t . Thus we need to replace assumption A5. A natural assumption would be to let all ratios n_t/N go to 0 with the same rate. Hence we will assume:

ASSUMPTION B.

1. Suppose that $n_t \rightarrow \infty$ for each t and $T \rightarrow \infty$ such that there exists a sequence $\{\lambda_s^*\}$, bounded

away from zero and infinity, such that as $T \rightarrow \infty$

$$\sup_{s=1, \dots, T} \left| \frac{n_s}{N} - \frac{\lambda_s^*}{T} \right| = o(1/T). \quad (13)$$

2. For each x $\sum_{s=1}^T \frac{n_s}{N} f_s(x)$ has a limit, $f(x)$ say, as $N \rightarrow \infty$.

3. $\int \sum_{s=1}^T \sigma_s^2(x) \frac{n_s}{N} f_s(x) dx$ is bounded as $N \rightarrow \infty$.

Note that under assumption B1 the limit in B2 may be rewritten as

$$f(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T \lambda_s^* f_s(x)$$

Moreover, under assumption A3 $f(x) > 0$ for $x \in \mathcal{X}$. Under assumptions A3 and B2, a sufficient condition for assumption B3 is that $\sigma_t(x)$ is bounded (in t). This latter condition is almost implied by A3.

Theorem 3 *Suppose that assumptions A1-A4, A6, and B1-B3 hold and that $\log N/(Nh) = o(1)$, $Th^2 = o(1)$ and $T/(\sqrt{N}h) = o(1)$ as $N \rightarrow \infty$. Then (12) holds.*

This shows that the estimation of θ_t does not affect the limiting distribution of the estimators of the parameters of the time series process or the tests. This means that standard errors can be constructed as if the θ_t were observed. Furthermore, under the strong exogeneity assumption, we can factor the likelihood so that our two-step approach to estimation of the parameters of θ_t does not lose information. Note that our result does not make any assumptions about properties of the process θ_t .

REMARK. In this asymptotic framework, we can revise the result of Theorem 2. For any x such that $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T \lambda_s b_s f_s(x) > 0$, we have

$$\sqrt{nTh} \left(\hat{g}(x) - g(x) - \frac{h^2 \sum_{t=1}^T \lambda_t b_t^3 f_t(x)}{2 \sum_{t=1}^T \lambda_t b_t f_t(x)} \mu_2 g''(x) \right) \xrightarrow{\mathcal{D}} N \left(0, \|K\|_2^2 \frac{\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T \lambda_s b_s \sigma_s^2(x) f_s(x)}{\left(\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T \lambda_s b_s f_s(x) \right)^2} \right),$$

provided h is chosen to be of order $(NT)^{-1/5}$.

6 Numerical Results

In this section we present the results of a small simulation experiment. We generated data from the design

$$y_{it} = \theta_t + x_{it} + u_{it},$$

where $u_{it} \sim N(0, 1)$, $x_{it} \sim U[-1, 1]$, and $\theta_t = \theta_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1)$ and $\theta_1 = 0$, with all random variables mutually independent. This results in the regression function and the time varying component having similar scale in most cases, see below. We take $T \in \{20, 40, 80\}$ and $n = n_t \in \{50, 100, 200\}$. Bandwidth was chosen by a Silverman rule of thumb procedure, specifically $h = 1.06\hat{\sigma}(nT)^{-1/5}$, where $\hat{\sigma}$ was the sample standard deviation of the covariates. This bandwidth is exactly optimal for the integrated mean squared error of a kernel density estimator when the underlying density is Gaussian. Obviously, it is not optimal for the problem at hand. However, it is so widely used and simple to implement and also relatively robust, that we decided on using it here. This means that the performance we report can likely be improved on by using a more time consuming method like least squares cross-validation. We evaluate several performance measures:

$$L_{T2}(\hat{\theta}) = E \sum_{t=2}^T (\hat{\theta}_t - \theta_t)^2 \quad ; \quad L_{\infty}(\hat{\theta}) = E \max_{2 \leq t \leq T} |\hat{\theta}_t - \theta_t|$$

$$L_2(\hat{g}) = \frac{1}{J} \sum_{j=1}^J E (\hat{g}(U_j) - g(U_j))^2 \quad ; \quad L_{\infty}(\hat{g}) = E \max_{1 \leq j \leq J} |\hat{g}(U_j) - g(U_j)|,$$

where $U_j \sim U[-1, 1]$ independent of the data. The expectations are computed by averaging over 100 simulation draws. We also evaluate the performance of the least squares estimator of the autoregressive coefficient, $\hat{\rho} = \sum_t \hat{\theta}_t \hat{\theta}_{t-1} / \sum_t \hat{\theta}_{t-1}^2$; we show the standard deviation and bias. Our results are given in Table 1.

Table 1

n	T	$L_{T_2}(\hat{\theta})$	$L_\infty(\hat{\theta})$	$L_2(\hat{g})$	$L_\infty(\hat{g})$	$\text{bias}(\hat{\rho})$	$\text{std}(\hat{\rho})$
50	20	0.3181	0.2757	0.0024	0.0878	-0.3022	0.2601
	40	0.6870	0.3215	0.0016	0.0722	-0.1589	0.1832
	80	1.5581	0.3720	0.0009	0.0535	-0.1220	0.1356
100	20	0.1635	0.2045	0.0014	0.0667	-0.1434	0.2473
	40	0.3761	0.2336	0.0010	0.0544	-0.0931	0.1501
	80	0.8382	0.2811	0.0005	0.0415	-0.0466	0.0832
200	20	0.0932	0.1540	0.0009	0.0553	-0.0645	0.1913
	40	0.2184	0.1789	0.0005	0.0434	-0.0374	0.1173
	80	0.5151	0.2159	0.0003	0.0326	-0.0254	0.0696

The performance of $\hat{\theta}$ clearly improves with n and gets worse with T . Note however that $L_{T_2}(\hat{\theta})$ roughly doubles and $L_\infty(\hat{\theta})$ increases by a factor $\sqrt{2}$ whenever T doubles as Theorem 1 would predict. Our asymptotics in section 5 refer to the case where $T(n) \rightarrow \infty$ as $n \rightarrow \infty$ and so one should ideally choose a path through these numbers. Our impression is that the results roughly correspond to the predictions of our asymptotics. The performance of \hat{g} seems to be much better and it improves with both n and T . Regarding $\hat{\rho}$ performance seems to improve primarily with T (as expected) but also there is some improvement as n increases, which reflects the reduction of the estimation error associated with the first stage. Note that even when the time series is observed and not estimated as here, $\hat{\rho}$ is negatively biased in finite samples.

Figure 1 below shows a typical outcome:

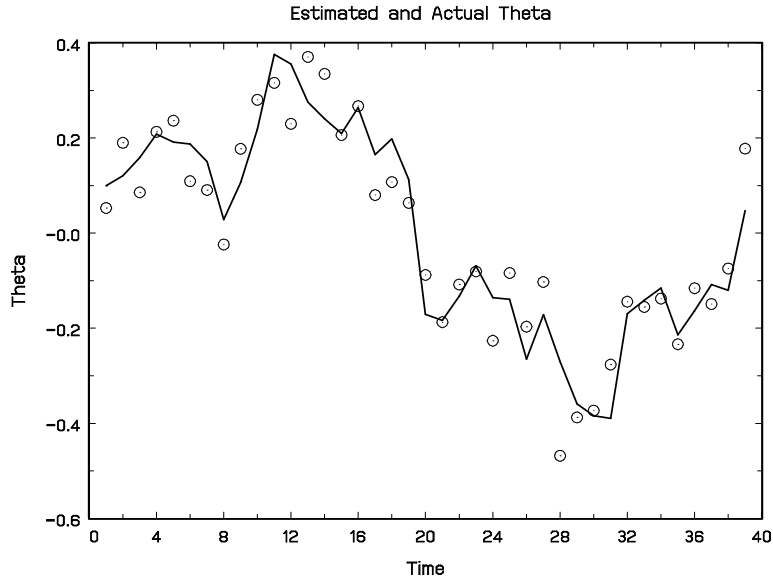


Figure 1. Shows actual time series (solid line) with estimated series (circles) for a case with $n = 200, T = 40$

7 Conclusions

We have established the theoretical properties of our estimation procedures for the quantities of interest in this semiparametric model for large panels. The simulation results generally support our asymptotic arguments.

The model can be extended in various ways. If the observed covariates X are multidimensional, our results go through provided we use multidimensional kernels and multidimensional local linear estimation. In some multivariate cases one may wish to impose additional structure on the function g such as additivity, index structure, or partial linearity. Our methodology provides consistent estimation of the unrestricted function; the additional structure may be imposed afterwards, see for example Linton and Nielsen (1995).

In some applications, one may also be concerned about individual effects, Hsiao (1986). For example, suppose that

$$Y_{i,t} = \alpha_i + \theta_t + g(X_{i,t}) + u_{i,t},$$

for some unobserved individual specific effect α_i . One can estimate the parameter vector $(\alpha_i)_i$ jointly with $(\theta_t)_t$ and $g(\cdot)$ by minimizing the re-defined sum of squared residuals in (4) subject to the constraint that $\sum_{i=1}^n \alpha_i = 0$. However, with a large cross-section this may be computationally demanding. Alternatively, either differencing or deviation from full mean eliminates the nuisance parameters and reduces the model to something very similar to (1).

A Appendix

A.1 Lemmas

We start by noting that

$$\widehat{\theta}_x = \theta + (A^\top K_{x,h}(I_N - W_{x,h})A)^{-1} (A^\top K_{x,h}(I_N - W_{x,h})Y^*),$$

where $Y^* = Y - A\theta$ is the vector with elements

$$\begin{aligned} Y_{i,t}^* &= g(X_{i,t}) + \sigma_t(X_{i,t})\epsilon_{i,t} \\ &= g(x) + g'(x)(X_{i,t} - x) + (g(X_{i,t}) - g(x) - g'(x)(X_{i,t} - x)) + \sigma_t(X_{i,t})\epsilon_{i,t}. \end{aligned}$$

Moreover as $(I_N - W_{x,h})B_x = 0$

$$A^\top K_{x,h}(I_N - W_{x,h})Y^* = A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) + A^\top K_{x,h}(I_N - W_{x,h})u \quad (14)$$

with $\mathbf{g} = (g(X_{i,t}))_{i,t}$. The first term on the far right is the ‘‘bias term’’, the second is the ‘‘variance term’’. Therefore,

$$\begin{aligned} &H_T^{1/2}\Delta_T^{1/2}(\widehat{\theta}_x - \theta) \\ &= \left(H_T^{-1/2}\Delta_T^{-1/2}A^\top K_{x,h}(I_N - W_{x,h})A\Delta_T^{-1/2}H_T^{-1/2} \right)^{-1} H_T^{-1/2}\Delta_T^{-1/2}A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \\ &+ \left(H_T^{-1/2}\Delta_T^{-1/2}A^\top K_{x,h}(I_N - W_{x,h})A\Delta_T^{-1/2}H_T^{-1/2} \right)^{-1} H_T^{-1/2}\Delta_T^{-1/2}A^\top K_{x,h}(I_N - W_{x,h})u. \end{aligned}$$

To prove our results we need the following two lemmas.

Lemma 1 *Suppose that assumption A holds. Then*

$$\sup_{x \in \mathcal{X}} \left\| H_T^{-1/2} \Delta_T^{-1/2} A^\top K_{x,h} (I_N - W_{x,h}) A \Delta_T^{-1/2} H_T^{-1/2} - D(x) \right\| = o_p(1).$$

where for a matrix W , $\|W\| = (\text{tr}(W^\top W))^{1/2}$.

Proof of Lemma 1. Letting

$$s_{j,t}(x) = \sum_{i=1}^{n_t} (X_{i,t} - x)^j K \left(\frac{X_{i,t} - x}{h_t} \right), \quad j \in \mathbb{N}_0, \quad t = 1, \dots, T,$$

it is well known (Fan and Yao (2003, Theorem 5.3)) that as $n_t \rightarrow \infty$ and $h_t \rightarrow 0$ such that $n_t h_t / \log n_t \rightarrow \infty$,

$$s_{j,t}(x) = N h^{j+1} \cdot b_t^{j+1} \lambda_t f_t(x) (\mu_j + O_P(r_N)) \quad (15)$$

with $r_N = \max_{s=1, \dots, T} (h_s + \sqrt{\log n_s / (n_s h_s)})$ the O_P -term is uniform in $x \in \mathcal{X}$. Note that by assumption A4 $\mu_0 = 1$ and $\mu_1 = 0$. Put $s_j(x) = \sum_{t=1}^T s_{j,t}(x)$, $j \in \mathbb{N}_0$.

It follows that the 2×2 -matrix $B_x^\top K_{x,h} B_x$ is

$$\begin{bmatrix} s_0(x) & s_1(x) \\ s_1(x) & s_2(x) \end{bmatrix} = \begin{bmatrix} N h \sum_{t=1}^T b_t \lambda_t f_t(x) (1 + O_P(r_N)) & N h^2 \sum_{t=1}^T b_t^2 \lambda_t f_t(x) O_P(r_N) \\ N h^2 \sum_{t=1}^T b_t^2 \lambda_t f_t(x) O_P(r_N) & \mu_2 N h^3 \sum_{t=1}^T b_t^3 \lambda_t f_t(x) (1 + O_P(r_N)) \end{bmatrix} \quad (16)$$

Next we see that for $t = 1, \dots, T-1$ the t' th row of $A^\top K_{x,h} B_x$ is

$$\begin{bmatrix} s_{0,t+1}(x) & s_{1,t+1}(x) \end{bmatrix} = N h f_{t+1}(x) \begin{bmatrix} b_{t+1} \lambda_{t+1} + O_P(r_N) & h O_P(r_N) \end{bmatrix}, \quad (17)$$

so that the t' th row of $A^\top K_{x,h} B_x (B_x^\top K_{x,h} B_x)^{-1}$ is

$$\begin{bmatrix} \frac{b_{t+1} \lambda_{t+1} f_{t+1}(x)}{\sum_{s=1}^T b_s \lambda_s f_s(x)} (1 + O_P(r_N)) & O_P(r_N/h) \end{bmatrix}, \quad (18)$$

Combining (17) and (18), $A^\top K_{x,h} W_{x,h} A$ is a $(T-1) \times (T-1)$ -matrix with (t, t') -element given by

$$N h \frac{b_{t+1} \lambda_{t+1} f_{t+1}(x) b_{t'+1} \lambda_{t'+1} f_{t'+1}(x)}{\sum_{s=1}^T b_s \lambda_s f_s(x)} (1 + O_P(r_N)). \quad (19)$$

The (t, t) element in the diagonal matrix $A^\top K_{x,h} A$ is $\sum_{i=1}^{n_{t+1}} K \left(\frac{X_{i,t+1} - x}{h_{t+1}} \right) = s_{0,t+1}(x)$. Hence the matrix $A^\top K_{x,h} (I_N - W_{x,h}) A$ is a $(T-1) \times (T-1)$ -matrix with diagonal-elements

$$N h b_{t+1} \lambda_{t+1} f_{t+1}(x) \left(1 - \frac{b_{t+1} \lambda_{t+1} f_{t+1}(x)}{\sum_{s=1}^T b_s \lambda_s f_s(x)} \right) (1 + O_P(r_N))$$

and off-diagonal elements given by (19). Pre and postmultiplying by $H_T^{-1/2} \Delta_T^{-1/2}$ gives the desired result. \blacksquare

Lemma 2 *Suppose that assumption A holds. Then the t 'th element of $H_T^{-1/2} \Delta_T^{-1/2} A^\top K_{x,h} (I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right)$ is $o_P(1)$ uniformly in $x \in \mathcal{X}$.*

Proof of Lemma 2. Under our assumptions

$$\begin{aligned} B_x^\top K_{x,h} \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) &= \frac{g''(x)}{2} \begin{bmatrix} s_2(x) \\ s_3(x) \end{bmatrix} + \begin{bmatrix} O(s_3^*(x)) \\ O(s_4(x)) \end{bmatrix} \\ &= Nh^3 \sum_{t=1}^T \lambda_t b_t^3 f_t(x) \left(\frac{g''(x)}{2} \begin{bmatrix} \mu_2 \\ hb_t \mu_3 \end{bmatrix} + \begin{bmatrix} hb_t \mu_3^* \\ h^2 b_t^2 \mu_4 \end{bmatrix} \right) \times (1 + O_P(r_N)) \end{aligned} \quad (20)$$

where $\mu_3^* = \int |z|^3 K(z) dz$ and

$$s_3^*(x) = \sum_{t=1}^T \sum_{i=1}^{n_t} |X_{i,t} - x|^3 K \left(\frac{X_{i,t} - x}{h_t} \right) = Nh^4 \sum_{t=1}^T \lambda_t b_t^4 f_t(x) \mu_3^* (1 + O(r_N)).$$

Combining (18) and (20) the t th element of the vector $A^\top K_{x,h} W_{x,h} \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right)$ is

$$Nh^3 \mu_2 \frac{g''(x)}{2} \lambda_{t+1} b_{t+1} f_{t+1}(x) \times \frac{\sum_{s=1}^T \lambda_s b_s^3 f_s(x)}{\sum_{s=1}^T \lambda_s b_s f_s(x)} \times (1 + O_P(r_N)).$$

Similarly, the t 'th element of $A^\top K_{x,h} \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right)$ is

$$\frac{g''(x)}{2} s_{2,t+1}(x) + O(s_{3,t+1}^*(x)) = Nh^3 \mu_2 \frac{g''(x)}{2} \lambda_{t+1} b_{t+1}^3 f_{t+1}(x) (1 + O_P(r_N)).$$

Therefore, the t th element of the vector $A^\top K_{x,h} (I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right)$ is

$$Nh^3 \mu_2 \frac{g''(x)}{2} \lambda_{t+1} b_{t+1} f_{t+1}(x) \left[b_{t+1}^2 - \frac{\sum_{s=1}^T \lambda_s b_s^3 f_s(x)}{\sum_{s=1}^T \lambda_s b_s f_s(x)} \right] (1 + O_P(r_N)).$$

Premultiplying by the diagonal matrix $H_T^{-1/2} \Delta_T^{-1/2}$ we get the desired result. \blacksquare

A.2 Proof of Theorem 1

First, write

$$\begin{aligned} \Delta_T^{1/2}(\widehat{\theta} - \theta) &= \left(H_T^{-1/2} H_T^{-1/2} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) A \omega(x) dx \Delta_T^{-1/2} H_T^{-1/2} H_T^{1/2} \right)^{-1} \\ &\quad \times H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \omega(x) dx \\ &+ \left(H_T^{-1/2} H_T^{-1/2} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) A \omega(x) dx \Delta_T^{-1/2} H_T^{-1/2} H_T^{1/2} \right)^{-1} \\ &\quad \times H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) u \omega(x) dx. \end{aligned}$$

From Lemma 1 and the hypothesis on ω , we have

$$H_T^{-1/2} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) A \omega(x) dx \Delta_T^{-1/2} H_T^{-1/2} = \int D(x) \omega(x) dx + o_p(1),$$

whence

$$H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) A \omega(x) dx \Delta_T^{-1/2} = B^{-1/2} \int D(x) \omega(x) dx B^{1/2} + o_p(1).$$

Moreover

$$H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \omega(x) dx = o_P(1)$$

by lemma 2. The t 'th element of $A^\top K_{x,h}(I_N - W_{x,h})u$ is (using (18))

$$\begin{aligned} \sum_{i=1}^{n_{t+1}} K \left(\frac{X_{i,t+1} - x}{h} \right) \sigma_{t+1}(X_{i,t+1}) \epsilon_{i,t+1} - \sum_{s=1}^T \sum_{i=1}^{n_s} K \left(\frac{X_{i,s} - x}{h_s} \right) (X_{i,s} - x) \sigma_s(X_{i,s}) \epsilon_{i,s} O_P(r_N/h) \\ - \frac{b_{t+1} \lambda_{t+1} f_{t+1}(x)}{\sum_{s=1}^T b_s \lambda_s f_s(x)} \sum_{s=1}^T \sum_{i=1}^{n_s} K \left(\frac{X_{i,s} - x}{h_s} \right) \sigma_s(X_{i,s}) \epsilon_{i,s} (1 + O_P(r_N)). \end{aligned}$$

It follows that

$$\begin{aligned} H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h}(I_N - W_{x,h}) \omega(x) dx \times u \\ = \int \bar{C}(x) \begin{bmatrix} \frac{1}{h_1 \sqrt{n_1}} \sum_{i=1}^{n_1} K \left(\frac{X_{i,1} - x}{h_1} \right) \sigma_t(X_{i,1}) \epsilon_{i,1} \\ \vdots \\ \frac{1}{h_T \sqrt{n_T}} \sum_{i=1}^{n_T} K \left(\frac{X_{i,T} - x}{h_T} \right) \sigma_t(X_{i,T}) \epsilon_{i,T} \end{bmatrix} \omega(x) dx \times (1 + O_P(r_N)). \end{aligned}$$

Define $(\bar{c}_t(x))_{t=1,\dots,T-1} = \gamma^\top \bar{C}(x)$ for arbitrary vectors $\gamma \in \mathbb{R}^{T-1}$. Then

$$\begin{aligned} & \sum_{t=1}^T \frac{1}{\sqrt{n_t}} \int \bar{c}_t(x) \frac{1}{h_t} \sum_{i=1}^{n_t} K\left(\frac{X_{i,t} - x}{h_t}\right) \omega(x) dx \times \sigma_t(X_{i,t}) \epsilon_{i,t} \\ &= \sum_{t=1}^T \frac{1}{\sqrt{n_t}} \sum_{i=1}^{n_t} \bar{c}_t(X_{i,t}) \omega(X_{i,t}) \sigma_t(X_{i,t}) \epsilon_{i,t} + o_p(1), \end{aligned}$$

by changing variables and dominated convergence. Using standard arguments the vector $Z_n = (Z_{n1}, \dots, Z_{nT})^\top$, where

$$Z_{nt} = \frac{1}{\sqrt{n_t \int \sigma_t^2(x) \bar{c}_t^2(x) \omega^2(x) f_t(x) dx}} \sum_{i=1}^{n_t} \bar{c}_t(X_{i,t}) \omega(X_{i,t}) \sigma_t(X_{i,t}) \epsilon_{i,t} \quad t = 1, \dots, T,$$

is jointly asymptotically normal with mean zero and identity variance covariance matrix. It follows that

$$\gamma^\top H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h} (I_N - W_{x,h}) \omega(x) dx \times u = i_T^\top \Psi_\gamma^{1/2} Z_n + o_p(1) \xrightarrow{\mathcal{D}} N(0, i_T^\top \Psi_\gamma i_T),$$

where $\Psi_\gamma = \text{diag}\{\int \sigma_t^2(x) \bar{c}_t^2(x) \omega^2(x) f_t(x) dx\}$. Hence $i_T^\top \Psi_\gamma i_T$ equals

$$\sum_{t=1}^T \int \sigma_t^2(x) \bar{c}_t^2(x) \omega^2(x) f_t(x) dx = \gamma^\top \int \bar{C}(x) \Psi(x) \bar{C}(x)^\top \omega^2(x) dx \gamma = \gamma^\top \int \Omega(x) \omega^2(x) dx \gamma$$

Therefore, by the Cramer-Wold device

$$H_T^{-1} \Delta_T^{-1/2} \int A^\top K_{x,h} (I_N - W_{x,h}) \omega(x) dx \times u \xrightarrow{\mathcal{D}} N\left(0, \int \Omega(x) \omega^2(x) dx\right)$$

The result follows. ■

A.3 Proof of Theorem 2

Let

$$\beta_T(x) = \frac{1}{2} \mu_2 \frac{\sum_{t=1}^T \lambda_t b_t^3 f_t(x)}{\sum_{t=1}^T \lambda_t b_t f_t(x)} g''(x).$$

We see that

$$\begin{aligned}
& \sqrt{Nh} (\hat{g}(x) - g(x) - h^2\beta_T(x)) \\
&= \sqrt{Nh} \left(\begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} (Y - A\theta) - h^2\beta_T(x) \right) \\
&\quad - \sqrt{Nh} \begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} A \Delta_T^{-1/2} \Delta_T^{1/2} (\hat{\theta} - \theta) \\
&= \sqrt{Nh} \left(\begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} (Y - A\theta) - h^2\beta_T(x) \right) + o_P(1)
\end{aligned}$$

using the results of the previous section and (18). Note that

$$\begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} (Y - A\theta)$$

is the pooled local linear regression estimator of x based on the independent data $Y_{i,t}^* = g(X_{i,t}) + \sigma_t(X_{i,t})\epsilon_{i,t}$ and the covariates $X_{i,t}$. We may rewrite this as

$$\begin{aligned}
\begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} Y^* &= \begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \\
&\quad + \begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \\
&\quad + \begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} B_x^\top K_{x,h} u
\end{aligned}$$

The first term is $g(x)$. To find the second term we note that by (15)

$$\begin{bmatrix} 1 & 0 \end{bmatrix} (B_x^\top K_{x,h} B_x)^{-1} = \left[\frac{1}{Nh \sum_1^T b_s \lambda_s f_s(x)} (1 + O_P(r_N)) \quad \frac{1}{Nh} O_P(r_N) \right].$$

Using this and (20) the second term becomes

$$\begin{aligned}
& \frac{Nh^3 \sum_{t=1}^T \lambda_t b_t^3 f_t(x) \frac{g''(x)}{2} \mu_2}{Nh \sum_1^T b_s \lambda_s f_s(x)} (1 + O_P(r_N)) + h^3 (1 + O_P(r_N)) + O_P\left(r_N \frac{Nh^4}{Nh}\right) \\
&= h^2 \beta_T(x) + o_P(1/\sqrt{Nh}).
\end{aligned}$$

The final term is

$$\frac{1}{\sum_{s=1}^T b_s \lambda_s f_s(x)} \cdot \frac{1}{Nh} \sum_{t=1}^T \sum_{i=1}^{n_t} K\left(\frac{X_{i,t} - x}{h}\right) \sigma_t(X_{i,t}) \epsilon_{i,t} (1 + O_P(r_N)) + O_P(r_N/\sqrt{Nh})$$

which using standard arguments is easily shown to be asymptotically normal with mean 0 and variance

$$\frac{1}{Nh} \|K\|_2^2 \frac{\sum_{s=1}^T \lambda_s b_s \sigma_s^2(x) f_s(x)}{\left(\sum_{s=1}^T \lambda_s b_s f_s(x)\right)^2}.$$

■

A.4 Proof of Theorem 3

First note that in the asymptotic set-up of Theorem 3

where

$$r_N = \max_{s=1, \dots, T} h_s + \sqrt{\frac{\log n_s}{n_s h_s}} = O\left(h + \sqrt{\frac{\log N}{Nh}}\right)$$

We have

$$\begin{aligned} \sum_{t=2}^T (\hat{\theta}_t - \theta_t)^2 &= \left(\int (Y^*)^\top (I_N - W_{x,h}) K_{x,h} A \omega(x) dx \right) \left(\int A^\top K_{x,h} (I_N - W_{x,h}) A \omega(x) dx \right)^{-1} \\ &\quad \times \left(\int A^\top K_{x,h} (I_N - W_{x,h}) A \omega(x) dx \right)^{-1} \left(\int A^\top K_{x,h} (I_N - W_{x,h}) Y^* \omega(x) dx \right) \\ &\leq \frac{1}{\zeta_N^2} \left(\int (Y^*)^\top (I_N - W_{x,h}) K_{x,h} A \omega(x) dx \right) \left(\int A^\top K_{x,h} (I_N - W_{x,h}) Y^* \omega(x) dx \right) \end{aligned}$$

where

$$\zeta_N = \inf_{z: z^\top z=1} z^\top \int A^\top K_{x,h} (I_N - W_{x,h}) A \omega(x) dx z$$

is the smallest eigenvalue of $\int A^\top K_{x,h} (I_N - W_{x,h}) A \omega(x) dx$. This may be bounded from below by $\int \zeta_N(x) \omega(x) dx$ with

$$\zeta_N(x) = \inf_{z: z^\top z=1} z^\top A^\top K_{x,h} (I_N - W_{x,h}) A z$$

the smallest eigenvalue of

$$A^\top K_{x,h} (I_N - W_{x,h}) A = \text{diag}(\underline{s}_0(x)) - [\underline{s}_0(x) \underline{s}_1(x)] \begin{bmatrix} s_0(x) & s_1(x) \\ s_1(x) & s_2(x) \end{bmatrix}^{-1} \begin{bmatrix} \underline{s}_0(x)^\top \\ \underline{s}_1(x)^\top \end{bmatrix} \quad (21)$$

where $\underline{s}_0(x)^\top = (s_{0,2}(x), s_{0,3}(x), \dots, s_{0,T}(x))$ and $\underline{s}_1(x)^\top = (s_{1,2}(x), s_{1,3}(x), \dots, s_{1,T}(x))$. Hence, we need to bound

$$\begin{aligned} z^\top A^\top K_{x,h}(I_N - W_{x,h})Az &= \frac{\sum_{t=2}^T z_t^2 s_{0,t}(x) \cdot s_0(x) - \left(\sum_{t=2}^T z_t s_{0,t}(x)\right)^2}{s_0(x) - s_1(x)^2/s_2(x)} \\ &+ \sum_{t=2}^T z_t^2 s_{0,t}(x) \cdot \frac{s_1(x)^2}{s_2(x)s_0(x) - s_1(x)^2} \\ &+ \frac{2s_1(x) \sum_{t=2}^T z_t s_{0,t}(x) \sum_{t=2}^T z_t s_{1,t}(x) - \left(\sum_{t=2}^T z_t s_{1,t}(x)\right)^2 s_0(x)}{s_2(x)s_0(x) - s_1(x)^2} \end{aligned} \quad (22)$$

away from 0. The first term of (22) may be re-written as

$$\begin{aligned} &\left(\sum_{s=2}^T s_{0,s}(x)\right)^2 \frac{\sum_{t=2}^T z_t^2 s_{0,t}(x) / \sum_{s=2}^T s_{0,s}(x) - \left(\sum_{t=2}^T z_t s_{0,t}(x) / \sum_{s=2}^T s_{0,s}(x)\right)^2}{s_0(x) - s_1(x)^2/s_2(x)} \\ &+ \frac{s_{0,1}(x)}{s_0(x) - s_1(x)^2/s_2(x)} \sum_{t=2}^T z_t^2 s_{0,t}(x) \end{aligned}$$

Of these two terms, the first one is non-negative and the second may be bounded from below by

$$\frac{s_{0,1}(x)}{s_0(x) - s_1(x)^2/s_2(x)} \min_{s=2,\dots,T} s_{0,s}(x) = \frac{Nh}{T} \frac{b_1 \lambda_1^* f_1(x)}{\frac{1}{T} \sum_{t=1}^T b_t \lambda_t^* f_t(x)} \min_{t=2,\dots,T} b_t \lambda_t^* f_t(x) (1 + O_P(r_N))$$

The second term of (22) is non-negative (and of smaller order than the first) whereas the third term is of order

$$\frac{(Nh^2/T \cdot O_P(r_N))^2 Nh/T \cdot (1 + O_P(r_N))}{N^2 h^4/T^2 \cdot (1 + O_P(r_N))} = O_P(Nhr_N^2/T)$$

It now follows that

$$\frac{T^2}{Nh} \zeta_N \geq \int \frac{b_1 \lambda_1^* f_1(x)}{\frac{1}{T} \sum_{t=1}^T b_t \lambda_t^* f_t(x)} \min_{t=2,\dots,T} b_t \lambda_t^* f_t(x) \cdot \omega(x) dx (1 + O_P(r_N)) + O_P(r_N^2)$$

which is bounded away from 0 by assumptions A3, A6, B1 and B2. Thus we need to show that

$$\left(\int \frac{T}{Nh} (Y^*)^\top (I_N - W_{x,h}) K_{x,h} A \omega(x) dx\right) \left(\int \frac{T}{Nh} A^\top K_{x,h} (I_N - W_{x,h}) Y^* \omega(x) dx\right) \quad (23)$$

is $o_P(1)$. Using (14) it suffices to bound

$$\begin{aligned} & \left(\frac{T}{Nh} \int A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \omega(x) dx \right)^\top \\ & \quad \times \left(\frac{T}{Nh} \int A^\top K_{x,h}(I_N - W_{x,h}) \left(\mathbf{g} - B_x \begin{bmatrix} g(x) \\ g'(x) \end{bmatrix} \right) \omega(x) dx \right) \end{aligned} \quad (24)$$

and

$$\left(\frac{T}{Nh} \int A^\top K_{x,h}(I_N - W_{x,h}) u \omega(x) dx \right)^\top \left(\frac{T}{Nh} \int A^\top K_{x,h}(I_N - W_{x,h}) u \omega(x) dx \right) \quad (25)$$

Applying lemma 2 we see that the t 'th element of first term of (??) (the ‘‘bias term’’) is $O_P(n_{t+1} h_{t+1}^2 r_N)$ so that (24) is

$$O_P \left(T^2 \frac{\sum_{t=2}^T (n_t h_t^2 r_N)^2}{N^2 h^2} \right) = O_P \left(T^2 h^2 \sum_{t=2}^T \left(\frac{n_t}{N} \right)^2 r_N^2 \right) = O_P(T h^2 r_N^2) = o_P(1)$$

For the vector containing the second term (??) (the ‘‘variance term’’) we write

$$\begin{aligned} \int A^\top K_{x,h}(I_N - W_{h,x}) u \omega(x) dx u &= \int C^*(x) \begin{bmatrix} \sum_{i=1}^{n_1} \sigma_1(X_{i,1}) K \left(\frac{X_{i,1}-x}{h_1} \right) \cdot \epsilon_{i,1} \\ \vdots \\ \sum_{i=1}^{n_T} \sigma_T(X_{i,T}) K \left(\frac{X_{i,T}-x}{h_T} \right) \cdot \epsilon_{i,T} \end{bmatrix} \omega(x) dx \\ & \quad \times (1 + O_P(r_N)) \end{aligned}$$

where $C^* = [0 | I_{T-1}] - \frac{v^* i_T^\top}{\sum_{s=1}^T v_s^*}$, with $v^* = (v_s^*)_{s=2, \dots, T}^\top$ and $v_s^* = \lambda_s^* b_s f_s(x)$ for $s = 1, \dots, T$. Ignoring the remainder term we get

$$\int \begin{bmatrix} \sum_{i=1}^{n_2} \sigma_2(X_{i,2}) K \left(\frac{X_{i,2}-x}{h_2} \right) \cdot \epsilon_{i,2} \\ \vdots \\ \sum_{i=1}^{n_T} \sigma_T(X_{i,T}) K \left(\frac{X_{i,T}-x}{h_T} \right) \cdot \epsilon_{i,T} \end{bmatrix} \omega(x) dx - \int v^* \frac{\sum_{t=1}^T \sum_{i=1}^{n_t} \sigma_t(X_{i,t}) K \left(\frac{X_{i,t}-x}{h_t} \right) \cdot \epsilon_{i,t}}{\sum_{s=1}^T v_s^*} \omega(x) dx$$

and since all terms have expectation 0, it suffices to show that

and

$$\frac{T^2}{N^2 h^2} \sum_{s=2}^T \text{Var} \left[\sum_{t=1}^T \sum_{i=1}^{n_t} \sigma_t(X_{i,t}) \epsilon_{i,t} \int \frac{v_s^*}{\sum_{s=1}^T v_s^*} K \left(\frac{X_{i,t}-x}{h_t} \right) \omega(x) dx \right] \quad (26)$$

go to 0. Here (??) may be bounded as follows:

$$\begin{aligned} & \frac{T^2}{N^2 h^2} \sum_{t=2}^T \text{Var} \left[\sum_{i=1}^{n_t} \sigma_t(X_{i,t}) \int K \left(\frac{X_{i,t} - x}{h_t} \right) \omega(x) dx \cdot \epsilon_{i,t} \right] \\ & \leq \text{const} \frac{T^2}{N h^2} \int \sum_{t=1}^T \sigma_t^2(x) \frac{n_t}{N} f_t(x) dx = O \left(\frac{T^2}{N h^2} \right) \end{aligned}$$

using assumption B3, whereas (26) may be bounded as follows:

$$\begin{aligned} & \frac{T^2}{N^2 h^2} \sum_{s=2}^T \text{Var} \left[\sum_{t=1}^T \sum_{i=1}^{n_t} \sigma_t(X_{i,t}) \epsilon_{i,t} \int \frac{v_s^*}{\sum_{s=1}^T v_s^*} K \left(\frac{X_{i,t} - x}{h_t} \right) \omega(x) dx \right] \\ & \leq \text{const} \frac{T}{N h^2} \int \sum_{t=1}^T \sigma_t^2(x) \frac{n_t}{N} f_t(x) dx = O \left(\frac{T}{N h^2} \right) \end{aligned}$$

■

References

- [1] Arellano, M. (2003). Panel Data Econometrics. Oxford University Press, Oxford.
- [2] Bai, J. 2003. Inferential theory for factor models of large dimension. *Econometrica* **71**, 135 – 171
- [3] Bai, J. 2004. Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* **122**, 137 – 183.
- [4] Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**, 191 – 221.
- [5] Berry, S., O.B. Linton, and A. Pakes (2004). Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems. *Review of Economic Studies* **71**, 613 – 654.
- [6] Carter, L.R. and Lee, R. D. (1992). Modelling and forecasting U.S mortality. *Journal of the American Statistical Association*. **87**, 659 - 671.
- [7] Connor, G. and O. Linton (2002). Semiparametric Estimation of a Characteristic-based Factor Model of Common Stock Returns. Forthcoming in *Journal of Empirical Finance*.

- [8] Fan, J. and R. Li (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* 99, 710-723.
- [9] Fan, J., T. Huang and R. Li (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* 102, 632-641.
- [10] Fan, J., and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications* Chapman and Hall.
- [11] Fan, J. and Q. Yao (2003). *Nonlinear Time Series Analysis*. Springer, Berlin.
- [12] Fengler M.R., W.K. Härdle and E. Mammen (2007). A semiparametric factor model for implied volatility surface dynamics *Journal of Financial Econometrics* 5, 189 – 218
- [13] Hansen, L.H., B. Nielsen and J.P. Nielsen (2004). Two sided analysis of variance with a latent time series. Nuffield College, Oxford University 2004-W25.
- [14] Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge University Press, Cambridge.
- [15] Jorgensen, D.W. (1986). Econometric methods for modelling producer behaviour. Handbook of Econometrics, vol 2. Eds, Z. Grilliches and M. Intrilligator.
- [16] Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica* 65, 1335 – 1364
- [17] Linton, O.B. and J.P. Nielsen, 1995, A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika* 82, 93 – 100.
- [18] Mammen, E., B. Støve, and D. Tjøstheim (2006). Nonparametric Additive Models for Panels of Time Series, manuscript.
- [19] Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure *Econometrica* 74, 967 – 1012.
- [20] Phillips. P.C.B., and H.R. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057 – 1113.

- [21] Porter, J. (1996). Nonparametric Regression Estimation for a Flexible Panel Data Model. PhD thesis, Department of Economics, MIT.