Nonparametric Estimation in Large Panels with Cross Sectional Dependence

Xiao Huang*
Department of Economics, Finance and Quantitative Analysis
Kennesaw State University
October 2010

Abstract
In this paper we consider nonparametric estimation in panel data under cross sectional dependence. Both the number of cross sectional units ($N$) and the time dimension of the panel ($T$) are assumed to be large, and the cross sectional dependence has a multifactor structure. Local linear regression is used to filter the unobserved cross sectional factors and to estimate the nonparametric conditional mean. A Monte Carlo study shows that the proposed estimator yields satisfactory finite sample properties.

Keywords: Cross sectional dependence; mixing process; large panels; local linear regression.

JEL Classification: C14; C23

*The author thanks the editor and two referees for useful comments that significantly improved the paper. The author also thanks Aman Ullah, Brett Katzman, Timothy Mathews and seminar participants at the 2006 Econometric Society Summer Meetings for helpful comments. All remaining errors are mine. Financial support from Kennesaw State University and U.C. Riverside is greatly acknowledged. Correspondence address: Department of Economics, Finance and Quantitative Analysis, M.D. 403, Kennesaw State University, Kennesaw, GA 30144, U.S.A. Phone: (770)423-6318. Fax: (770)499-3209. E-mail: xhuang3@kennesaw.edu
1 Introduction

Cross sectional dependence is common in economic data. For example, such dependence would arise in a cross country study of oil price changes or regional financial crises with global effects. Likewise, this issue would be of concern in within country, cross industry studies, due to common changes in macroeconomic policy. Several recent studies have focused on the estimation of panel data under cross sectional dependence. Coakley et al. (2002) use a principal component method to extract cross sectional common shocks and use them as augmented regressors in estimation; Huang (2008) studies panel vector autoregression under cross sectional dependence; Phillips and Sul (2003, 2007) study estimation and its bias in dynamic panels under cross sectional dependence. More recently, Bai (2009) proposes an iterative procedure for joint estimation of mean parameters and common factors in large panels.

In addition to considering cross sectional dependence, correct specification in the conditional mean is also important in panel data modeling. While previous research focuses on linear panels under cross sectional dependence, it is always interesting to consider a nonparametric specification in the conditional mean. This paper extends the method in Pesaran (2006) to nonparametric estimation in large panels under multifactor cross sectional dependence. The common correlated effects estimator (CCE) in Pesaran (2006) uses "means of cross-section aggregates" to filter unobserved common factors in estimation. Using local linear regression, we further show that means of time dimension aggregates can be used to filter the unobserved fixed-effects. Our model nests the usual random-effects specification, and our estimator is applicable
to both static and dynamic panels. Further, Monte Carlo simulation shows that the proposed estimator yields satisfactory results for processes with high degrees of heterogeneity and dynamics.

The paper is organized as follows. A nonparametric multifactor model and a filtering method are introduced in Section 2. Assumptions and the estimation procedure are detailed in Section 3. The proposed estimator is shown to be asymptotically normal in Section 4. Monte Carlo simulation results are reported in Section 5, and Section 6 concludes. All proofs and two lemmas are relegated to the Appendix.

2 A nonparametric multifactor model

We introduce the method to filter cross sectional dependence in nonparametric panels in this section.

2.1 A general approach

Equation (1) generalizes the model in Pesaran (2006) to a nonparametric framework with multifactor errors

$$y_{it} = m(x_{it}) + \gamma'_{1t}f_{1t} + \gamma'_{2t}f_{2t} + \varepsilon_{it},$$

(1)

where $y_{it}$ is a scalar, $x_{it}$ is a $d_x \times 1$ vector of observed regressors, $f_{1t}$ is a $d_1 \times 1$ vector of observed common factors, $f_{2t}$ is a $d_2 \times 1$ vector of unobserved common factors, $\gamma'_{1t}$ and $\gamma'_{2t}$ are factor loadings, $\varepsilon_{it}$ is a scalar error term, and $m(x_{it})$ is a smooth, unknown function of $x_{it}$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. 

2
Both $f_{1t}$ and $f_{2t}$ are sources of cross sectional dependence. Note that fixed-effects and random-effects models are both nested in (1) when $f_{1t} = 1$.

The basic idea in Pesaran (2006) is to replace the unobserved factors with observed data and factors in the panel regression model. This is achieved by representing the unobserved common factors, $f_{2t}$, as a function of the observed data and observed factors after cross sectional averaging. In the following, we show how, with modifications, the method of Pesaran (2006) can be applied to a nonparametric regression. We illustrate the idea with a local linear regression and higher order local polynomial fitting employs the same procedure.

Consider a first order Taylor series expansion of $m(x_{it})$ at the point $x$

$$y_{it} = m(x) + (x_{it} - x)' \beta(x) + R(x_{it}, x) + \gamma_{1t}f_{1t} + \gamma_{2t}f_{2t} + \varepsilon_{it}, \quad (2)$$

where $m(x)$ is the local conditional mean of $y_{it}$ and $\beta(x)$ is the local first order derivative of $m(x_{it})$ w.r.t. $x$. $R(x_{it}, x)$ denotes the remainder in the Taylor series expansion which is order $O(h^2)$ where $h$ is the bandwidth. Applying the idea in Pesaran (2006), we can average (2) over $i$ to obtain

$$\bar{y}_t = m(x) + (\bar{x}_t - x)' \beta(x) + \frac{1}{n} \sum_{i=1}^{n} R(x_{it}, x) + \gamma_{1t}f_{1t} + \gamma_{2t}f_{2t} + \bar{\varepsilon}_t, \quad (3)$$

where $\bar{y}_t = \frac{1}{n} \sum_{i=1}^{n} y_{it}$, $\bar{x}_t = \frac{1}{n} \sum_{i=1}^{n} x_{it}$, $\bar{\gamma}_{1t} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{1i}$, $\bar{\gamma}_{2t} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{2i}$, and $\bar{\varepsilon}_t = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{it}$ are all cross sectional averages. Rewriting (3) and expressing the unobserved factors, $f_{2t}$, as a function of observed data, we have

$$f_{2t} = (\gamma_{2}' \gamma_{2})^{-1} \gamma_{2}(\bar{y}_t - m(x) - (\bar{x}_t - x)' \beta(x) - \frac{1}{n} \sum_{i=1}^{n} R(x_{it}, x) - \gamma_{1t}f_{1t} - \bar{\varepsilon}_t). \quad (4)$$
Both $m(x)$ and $\beta(x)$ are local constants, and the term $\frac{1}{n} \sum_{i=1}^{n} R(x_{it}, x)$ is order $O(h^2)$. We also have $\tilde{\varepsilon}_i \to 0$ as $n \to \infty$ by law of large numbers. This preliminary analysis suggests that $f_{2t}$ in (4) can be approximated by a linear function of observed variables such as $\bar{y}_t$, $\bar{x}_t$, $f_{1t}$, and a local constant. Although constants such as $m(x)$ appear in (4), they are not needed as augmented regressors when approximating $f_{2t}$. The conditional mean of $f_{2t}$ given $x$ will be captured by $m(x)$ in (2) if it is not equal to zero, and adding a constant as an augmented regressor will create an identification problem for $m(x)$. This further implies we can replace $f_{2t}$ in (1) and (2) with the augmented regressors $\bar{y}_t$ and $\bar{x}_t$ ($f_{1t}$ is not augmented since it already appears in (1) and (2)). A similar idea is used to derive the CCE for the linear panel regression model in Pesaran (2006). Hence, after replacing $f_{2t}$ with a linear function of $\bar{y}_t$ and $\bar{x}_t$, we may write (1) as

$$y_{it} = m(x_{it}) + \gamma_{1i}^* f_{1t} + \beta_1^* \bar{y}_t + \beta_2^* \bar{x}_t + \varepsilon_{it}^*, \quad (5)$$

where $\gamma_{1i}^*$ is the revised factor loading for $f_{1t}$, $\varepsilon_{it}^*$ consists of both $\varepsilon_{it}$ and the approximation error from (4), and $m(x_{it})$ in (5) can be estimated using a local linear regression.

### 2.2 The fixed-effects model

The model in (1) nests both fixed-effects and random-effects models. In contrast, the approach in (5) is inapplicable to the special case when $f_{1t} = 1$ or one element in $f_{1t}$ is equal to one, which corresponds to the fixed-effects model if $\gamma_{1i}$ is a constant.
It is helpful to first show how the method in (5) breaks down in this special case. Consider the case when both $\gamma_{1i}$ and $f_{1it}$ are scalars, and let $f_{1it} = 1$. Equation (1) reduces to

$$y_{it} = m(x_{it}) + \gamma_{1i} + \gamma_{2i}f_{2i} + \varepsilon_{it}. \quad (6)$$

In a local linear regression with first order Taylor series expansion, we can write (5) as

$$y_{it} = m(x) + \gamma_{1i}^* + (x_{it} - x)' \beta(x) + R(x_{it}, x) + \beta_1^* \bar{y}_i + \beta_2^* \bar{x}_i + \varepsilon_i^* \quad (7)$$

where $\gamma_{1i}^* = \gamma_{1i}$ in this case. Note that the intercept in (7) is $m(x) + \gamma_{1i}^*$ and $m(x)$ is not identified unless a restriction such as $\bar{\gamma}_1^* = 0$ (equivalent to $\bar{\gamma}_1 = 0$) is imposed, where $\bar{\gamma}_1^* = \frac{1}{n} \sum_{i=1}^n \gamma_{1i}^*$. A similar restriction is commonly imposed in one-way or two-way error component models in order to identify the intercept (see for example Hsiao (2003)). However, such impositions are usually purely made for identification purposes without any economic explanation or justification. Instead of arbitrarily imposing such a restriction, we show an alternative way to resolve this issue in large panels.

Averaging (7) over $t$ for cross sectional unit $i$ under first order Taylor series expansion yields

$$\bar{y}_i = m(x) + (\bar{x}_i - x)' \beta(x) + \frac{1}{T} \sum_{t=1}^T R(x_{it}, x) + \gamma_{1i}^* + \beta_1^* \bar{y}_i + \beta_2^* \bar{x}_i + \varepsilon_i^*, \quad (8)$$

where $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$, $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$, $\bar{y} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{it}$, $\bar{x} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T x_{it}$, and $\bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}^*$. Rearranging terms in (8) and writing $\gamma_{1i}^*$ as a function
of the rest of the variables, we have

$$
\gamma_{1i}^* = \tilde{y}_i - m(x) - (\bar{x}_i - x)' \beta(x) - \frac{1}{T} \sum_{t=1}^{T} R(x_{it}, x) - \beta_1^* \tilde{y} - \beta_2^* \bar{x} - \tilde{\varepsilon}_i^*.
$$

(9)

Note that both $\tilde{y}$ and $\bar{x}$ converge to some non-stochastic constants under suitable conditions as $n, T \to \infty$. By reasoning similar to that which was given immediately after (4) (describing how $f_{2t}$ can be approximated), $\gamma_{1i}^*$ in (9) can be approximated with a linear function of $\tilde{y}_i$ and $\bar{x}_i$ as $n, T \to \infty$.

More generally, when $f_{1t}$ is a vector with one element equal to one, e.g. $f_{1t,1} = 1$, (9) becomes

$$
\gamma_{1i,1}^* = \tilde{y}_i - m(x) - (\bar{x}_i - x)' \beta(x) - \frac{1}{T} \sum_{t=1}^{T} R(x_{it}, x) - \gamma_{1i}^{*-1} \tilde{f}_1 - \beta_1^* \tilde{y} - \beta_2^* \bar{x} - \tilde{\varepsilon}_i^*,
$$

(10)

where $\tilde{f}_1 = \frac{1}{T} \sum_{t=1}^{T} f_{1t}$, and $\gamma_{1i}^{*-1}$ and $f_{1t}$ are vectors excluding $\gamma_{1i,1}^*$ and $f_{1t,1}$, respectively. We can again approximate $\gamma_{1i,1}^*$ with a linear function of $\tilde{y}_i$ and $\bar{x}_i$ by noting that observed factors $f_{1t}$ already appear in equation (1).

Combining the results in (4) and (9), we conclude that $f_{2t}$ and $\gamma_{1i}^*$ in (6) and (7) can be replaced by a linear function of $\tilde{y}_i$, $\bar{x}_i$, $\tilde{y}_t$, and $\bar{x}_t$ as $n, T \to \infty$ when $f_{1t} = 1$. This leads to the following nonparametric panel data model with augmented regressors, on which the rest of the paper focuses

$$
y_{it} = m(x_{it}) + \beta_1^* \tilde{y}_t + \beta_2^* \bar{x}_t + \beta_3^* \tilde{y}_i + \beta_4^* \bar{x}_i + e_{it},
$$

(11)

where $e_{it}$ contains both idiosyncratic and approximation errors. Asymptotic normality of $\hat{m}(x)$ is derived in Section 4.
2.3 Remarks

Remark 1. In addition to the nonparametric specification, the main difference between the current method and the CCE in Pesaran (2006) is the filtering method in (9) and (10), where we replace the fixed-effects with a linear function of \( \bar{y}_i \) and \( \bar{x}_i \) in order to identify \( m(x) \). This identification issue appears only in nonparametric panel models in (1) and (6), but not in linear panel models. As discussed in Subsection 2.2, an alternative way to address the identification issue would be to assume \( \gamma_1^i = 0 \), an assumption which not only lacks economic intuition but also is hard to verify in practice.

Remark 2. The method in Pesaran (2006) further assumes a multifactor structure for \( x_{it} \)

\[ x_{it} = \Gamma_{1i}'f_{1t} + \Gamma_{2i}'f_{2t} + v_{it}, \tag{12} \]

where \( \Gamma_{1i} \) and \( \Gamma_{2i} \) are \( d_1 \times d_x \) and \( d_2 \times d_x \)-matrix factor loadings, respectively, and \( v_{it} \) is a \( d_x \times 1 \) vector of specific component for \( x_{it} \), which is distributed independently of \( f_{1t} \) and \( f_{2t} \) for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \).

Equation (4) in Pesaran (2006) stacks \( y_{it} \) and \( x_{it} \) and thus, unobserved factors can be solved from this system of equations. In this paper, we make no assumption equivalent to (12), but rather solve for \( f_{2t} \) based upon equation (1) or (6). Under either approach certain rank conditions must be satisfied. Namely, the outer product of the average of the unobserved factor loading matrix in Pesaran (2006) or the average of \( \gamma_{2i} \) in (4) must have full rank. Compared to the full rank assumption for the outer product \( \bar{\gamma}_2 \bar{\gamma}'_2 \) in (4), the rank condition assumption in equation (21) of Pesaran (2006) is stronger since it also imposes restrictions on \( \Gamma_{2i} \). For example, under this assumption
the columns in $\Gamma'_2$ are restricted to be nonzero, which implies that all un-
observed factors $f_{2t}$ affect all elements of $x_{it}$ at time $t$. Restrictions like this
are unnecessary in our model since we do not assume a multifactor structure
for $x_{it}$ in (12) and we extract $f_{2t}$ from (1) or (6). If the rank condition fails
to hold, augmenting the original equation with cross sectional averages may
still reduce the impact of cross sectional dependence (see Section 3 in Pe-
saran (2006) for a discussion). Recently, Jin and Su (2008) show how a sieve
estimator such as that in Baltagi and Li (2002) can be used for estimation,
employing an approach similar to that in Pesaran (2006) but focusing exclu-
sively on (12): solving for $f_{2t}$ from the factor model (12) and augmenting (1)
with $f_{1t}$ and $\bar{x}_t$.

Remark 3. The augmented nonparametric equation (11) is obtained
from a first order Taylor series expansion although different orders of Taylor
expansion could be used. In its simplest form, we can use a local constant
approach so that (2) becomes

$$y_{it} = m(x) + R(x_{it}, x) + \gamma'_1 f_{1t} + \gamma'_2 f_{2t} + \varepsilon_{it},$$

and equation (4) reduces to

$$f_{2t} = (\bar{\gamma}'_2 \bar{\gamma}'_2)^{-1} \bar{\gamma}'_2 (\bar{y}_t - m(x)) - \frac{1}{n} \sum_{i=1}^n R(x_{it}, x) - \bar{\gamma}'_1 f_{1t} - \bar{\varepsilon}_t.$$ 

The augmented equation when $f_{1t} = 1$ is

$$y_{it} = m(x_{it}) + \beta_1^* \bar{y}_t + \beta_2^* \bar{y}_t + \epsilon_{it}. \tag{13}$$
Alternatively, higher order Taylor series expansion could be used in filtering. In the case of \( d_x = 1 \) with a second order Taylor series expansion, we have

\[
y_{it} = m(x) + (x_{it} - x) \beta(x) + (x_{it} - x)^2 \tilde{\beta}(x)/2 + R(x_{it}, x) + \gamma_1 f_{1t} + \gamma_2 f_{2t} + \varepsilon_{it},
\]

where \( \tilde{\beta}(x) \) is the local second derivative, and the augmented equation becomes

\[
y_{it} = m(x_{it}) + \beta_1^* \tilde{y}_t + \beta_2^* \bar{x}_t + \beta_3^* \bar{x}_t^2 + \beta_4^* \bar{y}_i + \beta_5^* \bar{x}_i + \beta_6^* \bar{x}_i^2 + e_{it}. \tag{14}
\]

If \( x_{it} \) is a vector, products between elements of \( \bar{x}_t \) and between elements of \( \bar{x}_i \) may appear in (14).

**Remark 4.** It might be possible to modify the approach in Henderson et al. (2008) to allow for the presence of cross sectional dependence. However, the fixed-effects would be eliminated by taking a first order difference where additional care would be required in exploring the additive structure in the differenced equation.

## 3 Assumptions and the estimation

We explain the assumptions and describe the estimation technique in this section.

### 3.1 Assumptions

The following assumptions are made for the model in (6) for all \( i = 1, \cdots, n, \)
\( t = 1, \cdots, T, \) and some positive constants \( C_1 \) and \( C_2. \) Let \( \alpha(j) \) denote the
mixing coefficient, $\|\cdot\|$ denote the Euclidean norm, and $L_1$ be the set of integrable functions of order one. Let $w_{it}$ denote all regressors in (11), e.g., $w_{it} = (x'_{it}, y_{it}, y_{it}, \bar{y}_t, \bar{x}'_t)$.

**Assumption 1**

(i) For each $i$, the process $\{y_{it}, x_{it}\}$ is stationary, strongly mixing with $\sum_{j=1}^{\infty} j^a [\alpha(j)]^{1-2/v} < \infty$ for some $v > 2$, $a > 1 - 2/v$, and a mixing size of $-r/(r-1)$ with $r > 1$. Both $y_{it}$ and $x_{it}$ satisfy $E|y_{it}|^{r+\delta} < \infty$ and $E|x_{it}|^{r+\delta} < \infty$ for some $\delta > 0$.

(ii) Common factors: $f_{2t}$ is strongly mixing with $\sum_{j=1}^{\infty} j^a [\alpha(j)]^{1-2/v} < \infty$ for some $v > 2$ and $a > 1 - 2/v$ and is independent of $\varepsilon_{it}$.

(iii) Factor loadings: both $\gamma_{1i}$ and $\gamma_{2i}$ are either bounded constants or i.i.d. random variables with a finite first moment, distributed independently of $f_{2t}$ and $\varepsilon_{it}$. $\text{Rank}(\tilde{\gamma}_2 \tilde{\gamma}'_2) = d_2 > 0$ for all $n$.

(iv) $\varepsilon_{it}$ is i.i.d. and independent of $x_{it}$ and $f_{2t}$. $E|\varepsilon_{it}| < \infty$ and $E(\varepsilon_{it}) = 0$.

In Assumption 1(i), we allow the process $\{y_{it}, x_{it}\}$ to be both individually and jointly dependent for each $i$, which covers both static and dynamic panels in (6). If model (1) is considered, $f_{1t}$ may also be assumed to be mixing. In Assumption 1(iii), we do not rule out the possibility of dependence or correlation between factor loadings and regressors, which is consistent with the usual specification in fixed-effects models. The rank condition in Assumption 1(iii) is needed in (4) to filter $f_{2t}$. This condition is not as strict as the assumption in equation (21) in Pesaran (2006). When $f_{1t} = 1$ and $\gamma_{1i}$ is random, $y_{it}$ is not mixing for each $i$, contradicting Assumption 1(i). $\gamma_{1i}$ needs to be interpreted as a fixed effect in this case.

**Assumption 2**
(i) The kernel function $K \in L_1$ is defined on a compact set in $\mathbb{R}^{dx}$ and is bounded. In addition, we have $\|u\|^2 K(u) \in L_1$.

(ii) The conditional density $f_{w_{i1},w_{il}}(u,v|y_{i1},y_{il}) < C_1$ for all $l > 1$.

(iii) For the strongly mixing process in Assumption 1(i), $E [|y_{i1}|^v |w_{i1}] < C_2$ for $v > 2$.

(iv) $h \to 0$, $n\pi_T h \to 0$, $nT h^{dx} \to \infty$, $n \sum_{l=\pi_T}^{\infty} t^a[\alpha(l)]^{1-2/v} \to 0$ with $\pi_T \to \infty$ and there exists a sequence of positive integers $\{v_T\}$ such that $v_T \to \infty$, $v_T = o((T h^{dx})^{1/2})$, and $(T/h^{dx})^{1/2} \alpha(v_T) \to 0$ as $n, T \to \infty$.

(v) The conditional density of $y_{it}$ given $w_{it}$ is continuous.

The assumption $\|u\|^2 K(u) \in L_1$ is used to ensure the existence of integrals in Lemma 1. A stronger assumption is needed for higher order polynomial fitting. Assumption 2 is modified from Conditions 1 to 4 in Masry (1996a). Assumption 2(i) is needed for $h^{-dx} u^{k(j)} K(u/h)$ to be an approximation of the identity, which is used for the convergence result of $E (s_n T, jq)$ in Lemma 1 (stated in Section 4). Assumption 2(ii) is used in proving Lemma 2(b) (stated in the Appendix) for strongly mixing processes, similar to condition (2b) in Masry (1996a). Assumption 2(iii) is the same as Condition 2(c) in Masry (1996a) and is needed for the convergence result in Lemma 2(b).

Assumption 2(iv) is complicated and needs more explanation. $h \to 0$ is the usual requirement for a shrinking bandwidth in nonparametric regression. Note that $n \pi_T h \to 0$ implies $nh \to 0$, $n\pi_T h^{dx} \to 0$, and $\pi_T h^{dx} \to 0$. $n\pi_T h^{dx} \to 0$ and $nT h^{dx} \to \infty$ imply that $T$ grows much faster than does $\pi_T$. The assumption $n \sum_{l=\pi_T}^{\infty} t^a[\alpha(l)]^{1-2/v} \to 0$ is needed to prove the convergence result in Lemma 2(b), which requires $\pi_T$ to grow fast enough so that $\sum_{l=\pi_T}^{\infty} t^a[\alpha(l)]^{1-2/v}$ converges to zero faster than does $1/n$. The assumption
about \( v_T \) is used to prove the existence of a "\( q_n \)" term in equation (3.12) of Masry (1996a), which is further used to construct the block size in the proof of asymptotic normality of the estimator. In simulations or other applications, there is no need to construct a sequence of \( v_T \) or \( \pi_T \) since they are only used to derive asymptotic normality. As for \( n \) and \( T \), we show in simulations that the proposed method has good finite sample properties even when \( n \) and \( T \) grow at the same rate with moderately large sample sizes.

3.2 The estimation

In the following, we consider local linear filtering and estimation in (6) based on the framework in Masry (1996a,b). The method in Robinson (1983) is also useful and nested in our results. The advantages of using local linear regression are discussed in Fan (1992a,b) and Fan and Gijbels (1996). Also see Pagan and Ullah (1999) for a summary. Simulation results show that a local linear estimator performs better than a local constant estimator.

In (4), we first note \( \bar{e}_t \to 0 \) as \( n \to \infty \) under Assumption 1(iv); \( \frac{1}{n} \sum_{i=1}^{n} R(x_{it}, x) \) is order \( O(h^2) \) and \( O(h^2) \to 0 \) under Assumption 2(iv). Both \( m(x) \) and \( x \beta(x) \) in (4) are included in the local constant of the local linear regression for \( m(x_{it}) \). This suggests \( f_{2t} \) in (4) can be approximated by some linear function of \( \bar{y}_t, \bar{x}_t, \) and \( f_{1t} \) plus a term of order \( o_p(1) \). When \( f_{1t} = 1, \bar{y}_1 \) becomes a constant as \( n \to \infty \) under Assumption 1(iii) and will also be included in the local constant. Similar analysis of (9) suggests \( \gamma_{1i}^* \) can be approximated by a linear function of \( \bar{y}_i, \bar{x}_i, \bar{y}_t, \) and \( \bar{x}_i \) plus an \( o_p(1) \) term. This further implies
that (11) may be more precisely written as

\[
y_{it} = m(x_{it}) + \beta_1^* \bar{y}_t + \beta_2^* \bar{x}_t + \beta_3^* \bar{y}_i + \beta_4^* \bar{x}_i + o_p(1) + e_{it},
\]

where \( o_p(1) \) captures possible approximation error for \( f_{2t} \) and \( \gamma_{1i}^* \) in (4) and (7). For ease of exposition, we omit the \( o_p(1) \) term in the following analysis, while keeping in mind the objective function for nonparametric regression in (18) should include an additional \( o_p(1) \) term inside the square brackets. All proofs can be easily modified to incorporate \( o_p(1) \), leading to the same asymptotic result in Theorem 1.

Let \( x_{it} = (x_{it,1}, \cdots, x_{it,d_x})' \) and \( x = (x_1, \cdots, x_{d_x})' \). We approximate \( m(x_{it}) \) in (11) at \( x \) by a multivariate local polynomial of order one

\[
m(x_{it}) \approx \sum_{0 \leq |k| \leq 1} D^k m(x) (x_{it} - x)^k
\]

where \( k = (k_1, \cdots, k_{d_x}) \) with \( k_1, \cdots, k_{d_x} = 0 \) or 1 is a \( d_x \)-tuple taking the following \( d_x + 1 \) different values

\[
\begin{pmatrix}
(0,0,\cdots,0,0) \\
(0,0,\cdots,0,1) \\
(0,0,\cdots,1,0) \\
\vdots \\
(1,0,\cdots,0,0)
\end{pmatrix}_{(d_x+1) \times d_x}
\]
\[ |k| = \sum_{i=1}^{d_x} k_i, \quad (x_{it} - x)^k = (x_{it,1} - x_1)^{k_1} \times \cdots \times (x_{it,d_x} - x_{d_x})^{k_{d_x}}, \] and

\[ D^k m (x) = \frac{\partial^k m (x)}{\partial x_1^{k_1} \cdots \partial x_{d_x}^{k_{d_x}}}. \]

Hence, there are \( d_x + 1 \) terms on the r.h.s. of (16). With the augmented regressors in (11), there are a total of \( N = 3d_x + 3 \) terms in the local linear regression. Similar to Masry (1996a), the objective function can be written as

\[ \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \sum_{q=1}^{d_x+1} b_q h^{k(q)} W_{it,q} - \sum_{q=d_x+2}^{N} b_q W_{it,q})^2 K ((x_{it} - x) / h), \quad (18) \]

where the coefficient \( b_q = D^{k(q)} m (x) \) when \( 1 \leq q \leq d_x + 1 \), \( K (u) \) is a kernel function on \( R^{d_x} \),

\[ W_{it,q} = \begin{cases} \frac{(x_{it} - x)^{k(q)}}{h} & \text{if } 1 \leq q \leq d_x + 1, \\ \text{elements of } (\tilde{y}_t, \tilde{x}_t, \tilde{y}_i, \tilde{x}_i)' & \text{if } d_x + 2 \leq q \leq N, \end{cases} \quad (19) \]

and \( k(q) \) is the \( q \)th \( d_x \)-tuple in (17).

For \( 1 \leq j, q \leq N \), the f.o.c.s for minimizing (18) w.r.t. \( b_q \) can be written as

\[ t_{nT,j} = \sum_{q=1}^{N} \dot{\beta}_{nT,q} s_{nT,jq}, \quad (20) \]
where

\[ t_{nT,j} = \frac{1}{nT} \sum_{i=1}^{nT} \sum_{t=1}^{T} y_{it} W_{it,j} K_h (x_{it} - x), \quad (21) \]

\[ s_{nT,jq} = \frac{1}{nT} \sum_{i=1}^{nT} \sum_{t=1}^{T} W_{it,j} W_{it,q} K_h (x_{it} - x), \quad (22) \]

\[ \hat{\beta}_{nT,q} = \begin{cases} 
  h^{k(q)} \hat{b}_q & \text{if } 1 \leq q \leq d_x + 1, \\
  \hat{b}_q & \text{if } d_x + 2 \leq q \leq N,
\end{cases} \quad (23) \]

\[ K_h (u) = \frac{1}{h^d_x} K (u/h), \]

\[ W_{it,j} \] is defined similarly in (19), \(|k(q)|\) is the sum of all elements in \(k(q)\) when \(1 \leq q \leq d_x + 1\), and \(\hat{\beta}_{nT,q}\) is the estimator.

Note \(|k(q)| = 0\) if \(q = 1\) and \(|k(q)| = 1\) if \(1 < q \leq d_x + 1\). The expressions in (21) and (22) are used to derive asymptotics for the parameter estimates.

### 4 Asymptotic results for local linear regression

Based on (21) and (23), we define the following three \(N \times 1\) vectors

\[
\begin{bmatrix}
  t_{nT,1} \\
t_{nT,2} \\
\vdots \\
t_{nT,N}
\end{bmatrix}, \quad \hat{\beta}_{nT} = \begin{bmatrix}
  \hat{\beta}_{nT,1} \\
  \hat{\beta}_{nT,2} \\
  \vdots \\
  \hat{\beta}_{nT,N}
\end{bmatrix}, \quad \beta = \begin{bmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_N
\end{bmatrix},
\]

\[ (24) \]

where the elements of \(t_{nT}\) and \(\hat{\beta}_{nT}\) are defined in (21) and (23), respectively, and elements of \(\beta, \beta_q\) are defined similarly to \(\hat{\beta}_{nT,q}\) in (23) by replacing \(\hat{b}_q\)
with \( b_q \). The \( \beta_1^*, \beta_2^*, \beta_3^*, \) and \( \beta_4^* \) in (11) correspond to the last \( 2d_x + 2 \) elements in \( \beta \). Also define the \( N \times N \) matrix

\[
S_{nT} = \begin{bmatrix}
S_{nT,11} & S_{nT,12} & \cdots & S_{nT,1N} \\
S_{nT,21} & S_{nT,22} & \cdots & S_{nT,2N} \\
\vdots & \vdots & \ddots & \vdots \\
S_{nT,N1} & S_{nT,N2} & \cdots & S_{nT,NN}
\end{bmatrix},
\]

(25)

where \( S_{nT,jq} = s_{nT,jq} \) in (22) for \( 1 \leq j, q \leq N \). This allows us to rewrite (20) in matrix form as \( t_{nT} = S_{nT} \hat{\beta}_{nT} \).

Next, we discuss the convergence of elements in \( S_{nT} \). Recall \( w \) is a \( N \times 1 \) vector of regressors in (11), and let \( w_j \) or \( w_q \) be an element of \( w \).

**Lemma 1.** Under Assumption 2, we have

\[
E(s_{nT,jq}) \to \mu_{jq} \text{ for } 1 \leq j, q \leq N,
\]

where

\[
\mu_{jq} = \begin{cases}
  \int f(x) \int u^{k_j+1} K(u) \, du & \text{if } 1 \leq j, q \leq d_x+1; \\
  \int u^{k_j} K(u) \, du \int w_q f(x, w_q) \, dw_q & \text{if } 1 \leq j \leq d_x+1 \text{ and } d_x+2 \leq q \leq N; \\
  \int w_j f(x, w_j) \, dw_j \int u^{k_q} K(u) \, du & \text{if } d_x+2 \leq j \leq N \text{ and } 1 \leq q \leq d_x+1; \\
  \int K(u) \, du \int w_j w_q f(x, w_j, w_q) \, dw_j \, dw_q & \text{if } d_x+2 \leq j, q \leq N.
\end{cases}
\]

\( f(\cdot) \) is the probability density function.

See Appendix for the proof.
Next, we define the $N \times N$ matrix
\[
M = \begin{bmatrix}
M_{11} & M_{12} & \cdots & M_{1N} \\
M_{21} & M_{22} & \cdots & M_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
M_{N1} & M_{N2} & \cdots & M_{NN}
\end{bmatrix},
\]
where $M_{jq} = \mu_{jq}$ for $j, q = 1, \ldots, N$. It immediately follows from Lemma 1 that
\[
S_{nT} \to M \text{ as } n, T \to \infty. \quad (26)
\]

Define
\[
t^*_{nT,j} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - m(x_{it}) - \beta'_1 \bar{y}_t - \beta'_2 \bar{x}_i - \beta'_3 \bar{y}_i - \beta'_4 \bar{x}_i)W_{it,j}K_h(x_{it} - x). \quad (27)
\]
Subtracting (27) from (21) yields
\[
t_{nT,j} - t^*_{nT,j} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (m(x_{it}) + \beta'_1 \bar{y}_t + \beta'_2 \bar{x}_i + \beta'_3 \bar{y}_i + \beta'_4 \bar{x}_i)W_{it,j}K_h(x_{it} - x). \quad (28)
\]
Replacing $m(x_{it})$ in the above equation with the following first order Taylor series expansion of $m(x_{it})$ at $x$
\[
m(x_{it}) = \sum_{0 \leq |k| \leq 1} D^k m(x)(x_{it} - x)^k + O_p(h^2)
\]
\[
= \sum_{q=1}^{d_x+1} h^{|k(q)|} D^{k(q)} m(x)\left(\frac{x_{it} - x}{h}\right)^{k(q)} + O_p(h^2)
\]
gives
\[ t_{nT,j} - t_{nT,j}^* = \frac{1}{nT} \sum_{t=1}^{nT} \sum_{t=1}^{T} \sum_{q=1}^{N} (\beta_q \hat{W}_{it,q} \hat{W}_{it,j} K_h (x_{it} - x) + O_p (h^2)), \quad (28) \]

where \( \beta_q \) is defined in (24). Using results in (20), (22), and (23), we solve for \( t_{nT,j}^* \) in (28) to obtain
\[ t_{nT,j}^* = \sum_{q=1}^{N} (\hat{\beta}_{nT,q} - \beta_q) s_{nT,jq} - O_p (h^2). \]

Using the definitions in (24) and (25), the above expression can be written in matrix form as
\[ t_{nT,j}^* = S_{nT} (\hat{\beta}_{nT} - \beta) - O_p (h^2). \quad (29) \]

Solving (29) for \( (\hat{\beta}_{nT} - \beta) \), we obtain
\[ \hat{\beta}_{nT} - \beta = S_{nT}^{-1} t_{nT}^* + S_{nT}^{-1} O_p (h^2). \quad (30) \]

This is essentially equation (2.15) in Masry (1996a). For simplicity, we do not include the specific bias term in (30). However, (26) implies \( S_{nT}^{-1} O_p (h^2) \rightarrow MO_p (h^2) \) and the bias term is order \( O_p (h^2) \) as \( n, T \rightarrow \infty \).

**Theorem 1.** Under Assumptions 1 and 2, as \( n, T \rightarrow \infty \),
\[ (nT h^{d_x})^{1/2} (\hat{\beta}_{nT} - \beta) \overset{d}{\rightarrow} N(0, M^{-1} \Psi M^{-1}). \quad (31) \]

See Appendix for the proof. The first element of \( \hat{\beta}_{nT} \) is the local estimate for \( m(x) \) and the last \( 2d_x + 2 \) elements are estimates for \( \beta_1, \beta_2, \beta_3, \) and \( \beta_4 \) in (11).
Remark 5. Under the assumption of random slope coefficients in Pesaran (2006), CCE is $\sqrt{T}$-consistent, while the convergence rate of $\hat{\beta}_{nT}$ in this paper involves $\sqrt{nT}$. This difference occurs because the filtering of $\gamma_{1i}$ and identification of $m(x)$ in (7) uses information from all cross sectional units.

Remark 6. The asymptotic result in (31) is developed specifically for the model defined in (6) with a single observed cross sectional factor $f_{1t} = 1$ for all $t$. The same method can be easily extended to a multifactor model with multiple observed cross sectional factors in (1). The estimation procedure would be based on the filtering method in (10) instead of (9).

5 Simulation study

This section discusses Monte Carlo simulations that study the finite sample property of the proposed estimator. Two data generating processes (DGPs) are used in simulation. In the first, we use $\sin(x_{it})$ to generate $m(x_{it})$ in (6), and compare the nonparametric estimator to the mean group CCE (MGCCE) in Pesaran (2006). In the second, we let $m(x_{it})$ be a linear function of $x_{it}$, and study the efficiency loss of the nonparametric estimator in comparison to MGCCE.
5.1 Data generating process

First, we let \( m(x) = \sin(x) \), and further suppose the rest of the DGP follows that considered in Pesaran (2006). More specifically, we let

\[
y_{it} = \sin(x_{it}) + \gamma_{1i} + \gamma_{2i,1} f_{2t,1} + \gamma_{2i,2} f_{2t,2} + \varepsilon_{it}
\]
\[
x_{it} = \Gamma_{1i} + \Gamma_{2i,1} f_{2t,1} + \Gamma_{2i,3} f_{2t,3} + v_{it}
\]  

(32)

for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \).

Both \( x_{it} \) and \( y_{it} \) in (32) are scalars. We let the observed factor \( f_{1t} = 1 \), which appears in both \( y_{it} \) and \( x_{it} \), with \( \gamma_{1i} \) and \( \Gamma_{1i} \) being the factor loadings, respectively. The unobserved factor \( f_{2t} \) consists of three elements, \( f_{2t} = (f_{2t,1}, f_{2t,2}, f_{2t,3})' \), where \( (f_{2t,1}, f_{2t,2})' \) appear in \( y_{it} \) and \( (f_{2t,1}, f_{2t,3})' \) appear in \( x_{it} \). The unobserved common factors are independent of each other and follow stationary AR(1) processes with i.i.d. normal (IIDN) errors,

\[
f_{2t,j} = \rho_f f_{2(t-1),j} + v_{f2t,j}, \quad \text{for } j = 1, 2, 3,
\]
\[
v_{f2t,j} \sim \text{IIDN}(0, \rho_f^2), \quad \rho_f = 0.5 \text{ for } j = 1, 2, 3.
\]

The factor loading \( \gamma_{1i} \) in \( y_{it} \) is generated as \( \gamma_{1i} = 0.5 \bar{x}_i \) with \( \bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it} \) for \( i = 1, \ldots, n \). Factor loadings for \( f_{2t,1} \) and \( f_{2t,2} \) in \( y_{it} \) are generated as \( \gamma_{2i,1} \sim \text{IIDN}(0, 0.2) \) and \( \gamma_{2i,2} \sim \text{IIDN}(0, 1) \), and the error term is \( \varepsilon_{it} \sim \text{IIDN}(0, 1) \).
For the $x_{it}$ process, we let

$$
\Gamma_{1i} \sim IIDN(0, 0.5), \quad \Gamma_{2i,1} \sim IIDN(0.5, 0.5), \quad \text{and} \quad \Gamma_{2i,3} \sim IIDN(0, 0.5),
$$

$$
v_{it} = \rho_{v,i} v_{i,t-1} + e_{it},
$$

$$
e_{it} \sim IIDN(0, 1 - \rho^2_{v,i}) \quad \text{and} \quad \rho_{v,i} \sim IIDU[0.05, 0.95],
$$

where $IIDU$ is $i.i.d.$ uniform distribution.

In the second set of simulations, we let the conditional mean be a linear function of $x$

$$
y_{it} = x_{it} \beta + \gamma_{1i} \gamma_{2i,1} f_{2t,1} + \gamma_{2i,3} f_{2t,3} + \varepsilon_{it}
$$

$$
x_{it} = \Gamma_{1i} + \Gamma_{2i,1} f_{2t,1} + \Gamma_{2i,3} f_{2t,3} + v_{it}, \quad (33)
$$

where $\beta = 0.7$. The rest of the DGP in (33) is the same as that in (32).

### 5.2 Estimation results

We compare three estimators in the simulation. The first is the local constant regression for $m(x)$ in (13). The second is the local linear estimation of $m(x)$ in (11) with $\bar{y}_t, \bar{x}_t, \bar{y}_t$, and $\bar{x}_t$ as augmented regressors. The third is the MGCCE in Pesaran (2006), where the slope estimator is obtained by averaging estimates from all cross sectional units. By comparing the nonparametric estimator to the MGCCE, we can assess the efficiency gain or loss of the local linear regression for the DGPs in (32) and (33).

In simulations, we discard the first 100 observations and let $(n, T)$ take different values with $n = 50, 100, 200$ and $T = 50, 100, 200$. The number of
replications is 1000. Tables 1 and 2 report the average mean absolute error (AMAE) and average mean square error (AMSE), which are given by

\[
\text{AMAE} = \frac{1}{1000} \sum_{j=1}^{1000} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} |\hat{m}^{(j)}(x_{it}) - m^{(j)}(x_{it})|,
\]

\[
\text{AMSE} = \frac{1}{1000} \sum_{j=1}^{1000} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (\hat{m}^{(j)}(x_{it}) - m^{(j)}(x_{it}))^2,
\]

where \(\hat{m}^{(j)}(x_{it})\) and \(m^{(j)}(x_{it})\) correspond to the nonparametric estimator and the conditional mean in the \(j\)th replication. \(\hat{m}^{(j)}(x_{it})\) is obtained from MGCCE if Pesaran’s method is used. The bandwidth is chosen as \(h = 1.06\sigma_x T^{-1/5}\), where \(\sigma_x\) is the standard deviation of all observations of \(x_{it}\). This bandwidth satisfies Assumption 2(iv) if certain restrictions on the growth rates of \(n\) and \(T\) are imposed. The optimal bandwidth minimizing integrated mean squared error (MISE) is \(h \propto (nT)^{-1/5}\). However, it was found during simulation that a bandwidth such as \(h = 1.06\sigma_x(nT)^{-1/5}\) may produce kernel functions very close to zero and lead to difficulties in estimation. We note that the results in Tables 1 and 2 for nonparametric estimators are not optimal. A better choice of bandwidth using a more computation-intensive method (such as cross validation) may further improve AMAE and AMSE.

Table 1 reports AMAE and AMSE of the local constant estimator, the local linear estimator, and the MGCCE for the nonlinear DGP in (32). As expected, the MGCCE is outperformed by the two nonparametric estimators in all sample sizes considered in Table 1. The AMAE and AMSE of the MGCCE do not decrease as sample size increases while the efficiency gain of using nonparametric estimator increases with sample size. The local linear estimator also has a clear advantage over the local constant estimator in
terms of both AMAE and AMSE.

Table 2 reports AMAE and AMSE for the linear DGP in (33). The MGCCE is almost unbiased, while the bias of the local linear estimator is small and on the same scale as that of the MGCCE except when $n = 200$ and $T = 200$. The bias of the local linear regression also decreases as sample size increases. The AMSE for the local linear estimator is much larger than that of the MGCCE when sample size is small. However, the efficiency loss as measured by AMSE in the local linear regression decreases as sample size grows. When sample size is relatively large (e.g., $n = 200$ and $T = 200$), AMSE for the local linear regression decreases from 0.0206 to 0.003. The discussion about Assumption 2(iv) in Section 3.2 suggests the requirement of $T \gg n$ for consistency of the proposed nonparametric estimator. Table 2 nonetheless shows that the proposed method gives reasonably good results even when $T$ is comparable in size to $n$.

6 Conclusion

This paper studies nonparametric estimation in panel data under cross sectional dependence. The procedure we propose is an extension of the method in Pesaran (2006). The cross sectional shocks are extracted by averaging equations across different cross sectional units. We further develop a method to specifically handle the fixed-effects model. The result developed in this paper is useful for both estimation and specification test in panel data.

Monte Carlo simulation shows that the proposed method has good finite sample properties and that the efficiency loss of this nonparametric method
decreases as sample size increases.

One interesting direction for future research is to test for cross sectional dependence in nonparametric panel data, possibly using the similar method in Ng (2006). It would also be interesting to extend our method to the semiparametric panel data model in Fan et al. (1995) and Li and Stengos (1996). We leave these topics for future research.
Appendix

Proof of Lemma 1. Lemma 1 is similar to Theorem 1 in Masry (1996a). Note that the augmented regressors $\bar{y}_t$ and $\bar{x}_t$ in (11) have the same dependence structure as $x_{it}$ and $f_{2t}$ under Assumption 1(i) and 1(ii) when $n \to \infty$. For $\bar{y}_i$ and $\bar{x}_i$, they are constants as $T \to \infty$ under Assumption 1(i) and 1(ii).

Consider case 1 when $1 \leq j, q \leq d_x + 1$. We have

$$s_{nT,jq} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( \frac{x_{it} - x}{h} \right)^{k(j)+k(q)} K_h \left( x_{it} - x \right).$$

Observe that $h^{-d_x} (u/h)^{k(j)+k(q)} K(u/h)$ is an approximation of the identity defined in Section 9.2 of Wheeden and Zygmund (1977). Taking the expectation for $s_{nT,jq}$ gives the result in Lemma 1.

For case 2 when $1 \leq j \leq d_x + 1$ and $d_x + 2 \leq q \leq N$, consider the example when $q = d_x + 2$ so that

$$s_{nT,jq} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( \frac{x_{it} - x}{h} \right)^{k(j)} \bar{y}_t K_h \left( x_{it} - x \right).$$

Taking the expectation of $s_{nT,jq}$ gives

$$E(s_{nT,jq}) = \int \int \left( \frac{u - x}{h} \right)^{k(j)} w_q \frac{1}{h^{d_x}} K\left( \frac{u - x}{h} \right) f(u, w_q) du dw_q$$
$$= \int \left( \int \left( \frac{u - x}{h} \right)^{k(j)} \frac{1}{h^{d_x}} K\left( \frac{u - x}{h} \right) f(u|w_q) du \right) w_q f(w_q) dw_q$$
$$= \int u^{k(j)} K(u) du \int w_q f(x, w_q) dw_q,$$  \hspace{1cm} (34)

where we again use the approximation of the identity in the third equality in (34). Note that if the augmented regressor is $\bar{y}_i$ or $\bar{x}_i$, (34) reduces to
\( f(x)\tilde{w}_q \int u^{k(j)}K(u)du \), where \( \tilde{w}_q \) is the limit of \( \tilde{y}_i \) or elements of \( \bar{x}_i \) as \( T \to \infty \).

The results for case 3 and case 4 in Lemma 1 can be obtained in similar ways. \( \square \)

We next present Lemma 2 for the variance and covariance of \( t^*_{nT} \), which is similar to Theorem 2 in Masry (1996a). Consider a linear combination of elements of \( t^*_{nT} \)

\[
Q_{nT} = \sum_{j=1}^{N} c_j t^*_{nT,j} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - m(x_{it}) - \beta_1^* \bar{y}_t - \beta_2^* \bar{x}_t - \beta_3^* \bar{y}_i - \beta_4^* \bar{x}_i) C_h(x_{it} - x),
\]

where

\[
C_h(u) = \sum_{j=1}^{N} c_j W_{it,j} K_h(u) \equiv \frac{1}{h^d} C(u/h),
\]

and

\[
C(u) = \sum_{j=1}^{N} c_j W_{it,j} K(u). \tag{35}
\]

Defining

\[
Z_{it} = (y_{it} - m(x_{it}) - \beta_1^* \bar{y}_t - \beta_2^* \bar{x}_t - \beta_3^* \bar{y}_i - \beta_4^* \bar{x}_i) C_h(x_{it} - x), \tag{36}
\]

\( Q_{nT} \) can be written as

\[
Q_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{it}.
\]

Let \( \sigma^2(w) = \text{var}(y|w) \) and define the \( N \times N \) covariance matrix of \( t^*_{nT} \)

\[
\Psi = \begin{bmatrix}
\Psi_{11} & \Psi_{12} & \cdots & \Psi_{1N} \\
\Psi_{21} & \Psi_{22} & \cdots & \Psi_{2N} \\
\vdots & \vdots & & \vdots \\
\Psi_{N1} & \Psi_{N2} & \cdots & \Psi_{NN}
\end{bmatrix},
\]
where $\Psi_{jq} = \psi_{jq}$ with

$$
\psi_{jq} = \int \sigma^2(w) w_j w_q \frac{1}{h^d z} K^2(u - x) f(w)dw
$$

for $j, q = 1, \cdots, N$. Note that $u$ is part of the variables of integration.

**Lemma 2.** Under Assumption 2, for $j, q = 1, \cdots, N$ and $n, T \to \infty$, we have

(a) $h^d_{2} \text{var}(Z_{11}) \to \int \sigma^2(w) \frac{1}{h^d z} C^2\left(\frac{u - x}{h}\right) f(w) dw$;

(b) $h^d_{2} \sum_{i=1}^{n} \sum_{t=2}^{T} \text{cov}(Z_{i1}, Z_{it}) = o(1)$;

(c) $nT h^d_{2} \text{var}(Q_{nT}) \to \int \sigma^2(w) \frac{1}{h^d z} C^2\left(\frac{u - x}{h}\right) f(w) dw$;

(d) $\text{cov}((nT h^d_{2})^{1/2} t^{*}_{nT,j}, (nT h^d_{2})^{1/2} t^{*}_{nT,q}) \to \psi_{jq}$.

**Proof of Lemma 2.** The proof follows that of Theorem 2 in Masry (1996a).

Recall $\sigma^2(w)^2 = \text{var}(y|w)$ . By stationarity and $E(Z_{it}) = 0$ as $n, T \to \infty$, we have

$$
\text{var}(Z_{11}) = E\left(Z_{11}^2\right) = \frac{1}{h^d_{2}} \int_{R^d} \sigma^2(w) C^2\left(\frac{u - x}{h}\right) f(w) dw, \quad (37)
$$

which immediately gives result in Lemma 2(a)

$$
h^d_{2} \text{var}(Z_{11}) \to \int_{R^d} \sigma^2(w) \frac{1}{h^d z} C^2\left(\frac{u - x}{h}\right) f(w) dw. \quad (38)
$$

Note that the integration in (38) is taken w.r.t. to $w$, which includes $u$ and all augmented regressors such as $(\bar{y}_t, \bar{x}_t, \bar{y}_i, \bar{x}_i)'$. The expression of $Z_{it}$ in (36) includes both $x_{it}$ and $(\bar{y}_t, \bar{x}_t, \bar{y}_i, \bar{x}_i)'$, but $(\bar{y}_t, \bar{x}_t, \bar{y}_i, \bar{x}_i)'$ do not appear in the kernel function. Hence result for the approximation of the identity can not be applied here.
For $Q_{nT}$, we have

$$\text{var} (Q_{nT}) = \frac{1}{nT} \text{var} (Z_{11}) + \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \left( \frac{2}{T} \sum_{t=T}^{T} (1 - \frac{t}{T}) \text{cov} (Z_{i_11}, Z_{i_2t}) \right)$$}

(39)

Next we prove Lemma 2(b). Let

$$J = \sum_{t=2}^{T} \left| \text{cov} (Z_{i_11}, Z_{i_2t}) \right|$$

$$= \sum_{t=2}^{d_w} \left| \text{cov} (Z_{i_11}, Z_{i_2t}) \right| + \sum_{t=d_w+1}^{\pi_T} \left| \text{cov} (Z_{i_11}, Z_{i_2t}) \right| + \sum_{t=\pi_T+1}^{T} \left| \text{cov} (Z_{i_11}, Z_{i_2t}) \right|$$

$$= J_1 + J_2 + J_3.$$}

(40)

This is similar to equation (2.29) in Masry (1996a) and the convergence result for $J$ in (40) follows immediately from that in Masry (1996a). Based on equation (2.30), equation (2.31), and the proof for "$h_n^d |J_3|" on page 93 in Masry (1996a), we have

$$h_n^d J \leq O (h) + O(h^d \pi_T) + o (1).$$}

(41)

Combined with Assumption 2(iv), (41) implies that Lemma 2(b) holds. We note that the exponent for $h$ in (41) is $d_x$, not $d_w$ used in (40). This is because the asymptotic result in (41) is derived by the change of variables for the first $d_x$ variables in $w$ (see page 92 in Masry (1996) for more details).

Lemma 2(c) follows from (41) and Assumption 2(iv), and Lemma 2(d) follows from Lemma 2(c) and the definition of $t^*_nT,j$ in (27). □

Lemma 3 below is similar to Theorem 3 in Masry (1996a) and gives
asymptotic normality of \( Q_{nT} \). Define

\[
\theta^2(x) = \int \sigma^2(w) \frac{1}{h^d} C^2 \left( \frac{u - x}{h} \right) f(w) \, dw.
\]

**Lemma 3.** Under Assumption 2 as \( n, T \to \infty \), we have

\[
(nTH^d)^{1/2} Q_{nT} \to N(0, \theta^2(x))
\]

at continuity points \( w \) of \( \{\sigma^2(w), f(w)\} \).

**Proof of Lemma 3.** For \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \), let

\[
Z_{nT,it} = (h^d)^{1/2} Z_{it} \quad \text{and} \quad V_{nT} = \sum_{i=1}^n \sum_{t=1}^T Z_{nT,it},
\]

and

\[
(nTH^d)^{1/2} Q_{nT} = \frac{1}{\sqrt{nT}} V_{nT}.
\]

We will show that

\[
\frac{1}{\sqrt{nT}} V_{nT} \to N(0, \theta^2(x)).
\]

Define the following random variables

\[
\begin{align*}
\eta_j &= \sum_{i=1}^n \sum_{t=j(u+v)+u}^{(j+1)(u+v)+u} Z_{nT,it}, \quad 0 \leq j \leq k - 1, \\
\xi_j &= \sum_{i=1}^n \sum_{t=j(u+v)+\alpha+1}^{(j+1)(u+v)} Z_{nT,it}, \quad 0 \leq j \leq k - 1, \\
\zeta_k &= \sum_{i=1}^n \sum_{t=k(u+v)+1}^T Z_{nT,it}.
\end{align*}
\]
where $k$ is defined as

$$k_T = \left[ \frac{T}{u_T + v_T} \right]$$

and, similar to Masry (1996a), we partition $\{1, 2, \cdots, T\}$ into $2k_T + 1$ subsets with large blocks of size $u = u_T$ and small blocks of size $v = v_T$. Note the last block in this partition covers $[k_T (u + v), T]$.

Now write

$$V_{nT} = \sum_{j=0}^{k_T-1} \eta_j + \sum_{j=0}^{k_T-1} \xi_j + \zeta_{k_T}$$

$$= V'_{nT} + V''_{nT} + V'''_{nT}. \tag{42}$$

We first show that

$$\frac{1}{nT} E(V''_{nT})^2 \to 0.$$ 

For $E(V''_{nT})^2$, we have

$$E(V''_{nT})^2 = \text{var} \left( \sum_{j=0}^{k_T-1} \xi_j \right) = \sum_{j=0}^{k_T-1} \text{var} (\xi_j) + \sum_{j_1=0}^{k_T-1} \sum_{j_2=0}^{k_T-1} \sum_{j_1 \neq j_2} \text{cov} (\xi_{j_1}, \xi_{j_2})$$

$$= F_1 + F_2.$$ 

Consider $F_1$. By stationarity, we have

$$F_1 = \sum_{j=0}^{k_T-1} \sum_{i=1}^{n} v_T \text{var}(Z_{nT,i}) + \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} 2v_T \sum_{i=1}^{n} (1 - \frac{t}{v_T}) \text{cov}(Z_{nT,i_1}, Z_{nT,i_2})$$

$$= \sum_{j=0}^{k_T-1} v_T \theta^2 (x) (1 + o(1)) O(n)$$

$$= k_T v_T \theta^2 (x) (1 + o(1)) O(n) \sim o(T) O(n), \tag{43}$$

30
where the second equality follows from Lemmas 2(a) and 2(b) and the third equality follows from equation (3.17) in Masry (1996a). Note that a set of assumptions about block sizes similar to (3.13) and (3.14) in Masry (1996a) is also made for the convergence results in $F_1$ and $F_2$. Namely, we assume, as $n,T \to \infty$, $v_T/U_T \to 0$, $u_T/T \to 0$, $u_T/(nT h^{dz}) \to 0$, and $T \alpha (v_T) \to 0$.

Let $r_{j1} = j_1(u + v) + u$ and $r_{j2} = j_2(u + v) + u$. Consider $F_2$.

$$F_2 = \sum_{j_1=0}^{k_T-1} \sum_{j_2=0}^{k_T-1} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{l_1=1}^{v} \sum_{l_2=1}^{v} \text{cov}(Z_{nT,i_1(r_{j1}+l_1)}, Z_{nT,i_2(r_{j2}+l_2)}).$$

The same analysis in Masry (1996a) applies here: when $j_1 \neq j_2$, $|r_{j1} - r_{j2} + l_1 - l_2| \geq u$ and we have

$$|F_2| \leq \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} 2T \sum_{l_1=1}^{T-u} \sum_{l_2=l_1+u}^{T} |\text{cov}(Z_{nT,i_1l_1}, Z_{nT,i_2l_2})|.$$

By stationarity and Lemma 2(b), we have

$$|F_2| \leq \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} 2T \sum_{l=0}^{T} |\text{cov}(Z_{nT,i_1l_1}, Z_{nT,i_2l_2})| = o(T) n. \tag{44}$$

(43) and (44) suggest

$$\frac{1}{nT} E(V''_{nT})^2 \to 0 \text{ as } n, T \to \infty. \tag{45}$$
Similarly, we can prove, as \( n, T \to \infty \),

\[
\frac{1}{nT} E (V''_{nT})^2 \to 0, \tag{46}
\]

\[
\frac{1}{nT} \sum_{j=0}^{k_T-1} E(\eta_j^2) \to \theta^2(x), \tag{47}
\]

in (42). Finally, note that the proof of the following two results

\[
\left| E (\exp (itV'_{nT})) - \prod_{j=0}^{k_T-1} E(\exp (it\eta_j)) \right| \to 0, \tag{48}
\]

\[
\frac{1}{nT} \sum_{j=0}^{k_T-1} E(\eta_j^2 I\{\eta_j > C_3 \theta(x) \sqrt{nT}\}) \to 0, \tag{49}
\]

(where \( C_3 \) is a positive constant and \( I(\cdot) \) is an indicator function) follows the same procedure in Masry (1996a) and is not repeated here. (48) implies that \( \eta_j \) in \( V'_{nT} \) of (42) are independent across \( j \), and (49) is the usual Lindeberg-Feller condition in the Central Limit Theorem for a sequence of independent random variables. Results in (45), (46), (47), (48), and (49) imply that Lemma 3 holds. \( \square \)

**Proof of Theorem 1.** Lemma 3 implies \( (nTh\dot{d}^*)^{1/2}t^*_{nT} \to N(0, \Psi) \), where \( \Psi \) is defined before Lemma 2. Combining the results in (26), (30), and Lemma 3, it follows that Theorem 1 holds. \( \square \)
References


Table 1
Estimated AMAE and AMSE for the DGP in (32) with 1000 replications

<table>
<thead>
<tr>
<th>(n, T)</th>
<th>AMAE</th>
<th>AMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local constant</td>
<td>Local linear</td>
</tr>
<tr>
<td>(50, 50)</td>
<td>0.2453</td>
<td>0.1336</td>
</tr>
<tr>
<td>(50, 100)</td>
<td>0.2009</td>
<td>0.1072</td>
</tr>
<tr>
<td>(100, 50)</td>
<td>0.2513</td>
<td>0.1380</td>
</tr>
<tr>
<td>(100, 100)</td>
<td>0.2065</td>
<td>0.1127</td>
</tr>
<tr>
<td>(200, 200)</td>
<td>0.1629</td>
<td>0.1028</td>
</tr>
</tbody>
</table>

Table 2
Estimated AMAE and AMSE for the DGP in (33) with 1000 replications

<table>
<thead>
<tr>
<th>(n, T)</th>
<th>AMAE</th>
<th>AMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local constant</td>
<td>Local linear</td>
</tr>
<tr>
<td>(50, 50)</td>
<td>0.2241</td>
<td>0.0889</td>
</tr>
<tr>
<td>(50, 100)</td>
<td>0.1641</td>
<td>0.0747</td>
</tr>
<tr>
<td>(100, 50)</td>
<td>0.2379</td>
<td>0.0652</td>
</tr>
<tr>
<td>(100, 100)</td>
<td>0.1804</td>
<td>0.0563</td>
</tr>
<tr>
<td>(200, 200)</td>
<td>0.1599</td>
<td>0.0237</td>
</tr>
</tbody>
</table>