Some of the Recent Developments on Nonparametric Econometrics

Zongwu Cai
University of North Carolina at Charlotte,
Qi Li
Texas A&M University

1 Introduction

There is a growing literature in nonparametric econometrics.

This paper focuses on the following area:

(i) nonparametric estimation and testing of regression functions with mixed discrete and continuous covariates.

(ii) nonparametric estimation/testing with non-stationary data.

(iii) Nonparametric/semiparametric estimation with panel data models with fixed effects.

(iv) Nonparametric instrumental variable estimations.

(v) Nonparametric estimation of quantile regression models.
2 Models with Discrete and Continuous Covariates

2.1 Nonparametric Estimation of Regression Functions with Discrete and Continuous Covariates

We consider the following regression model

\[ Y_i = g(X_i) + u_i, \quad (i = 1, ..., n) \quad (1) \]

where \( X_i = (X^c_i, X^d_i) \).

\( X^c_i \in \mathcal{R}^q \) is a continuous \((q \geq 1)\).

\( X^d_i \) is a discrete of dimension \( r \) \((r \geq 0)\).

\( X^d_{is} \), the \( s^{th} \) component of \( X^d_i \), \( s = 1, ..., r \).

\( X^d_{is} \in \mathcal{D}_s = \{ a_1, a_2, ..., a_{cs} \} \) taking \( cs \) distinct different values.

The conventional frequency method.

Smoothing estimation of Discrete probabilities:

Aitchison and Aitken (1976)

Smoothing regression function estimation:

Hall, Racine and Li (2004), Racine and Li (2004), Hall, Li and Racine (2007).

The kernel function for un-ordered discrete variable \( X^d_{is} \)

\[ l(X^d_{is}, x^d_s, \lambda_s) = \begin{cases} 1, & \text{if } X^d_{is} = x^d_s, \\ \lambda_s, & \text{if } X^d_{is} \neq x^d_s. \end{cases} \]
If $X_{is}^d$ is an ordered discrete variable, we use the following kernel function:

$$l(X_{is}^d, x_s^d, \lambda_s) = \lambda_s |X_{is}^d - x_s^d|.$$  

When $\lambda_s = 0$,

$$l(X_{is}^d, x_s^d, 0) = 1(X_{is}^d = x_s^d),$$ an indicator function.

$\lambda_s = 1$,

$$l(X_{is}^d, x_s^d, 1) \equiv 1$$ is a constant function.

$\lambda_s \in [0, 1]$ for all $s = 1, ..., r$.

The product kernel is

$$L(X_i^d, x_i^d, \lambda) = \prod_{s=1}^r l(X_{is}^d, x_s^d, \lambda_s).$$

For the continuous variables $x_c = (x_{c1}, \ldots, x_{cq})$ the product kernel is

$$W_h(x_c, X_i^c) = \prod_{s=1}^p h^{-1}_s w\left(\frac{x_c^s - X_{is}^c}{h_s}\right).$$

The kernel function for the mixed regressor $x = (x_c, x^d)$ is $K_{\gamma}(x, X_i) = W_h(x_c, X_i^c)L(x_i^d, x_i^d, \lambda)$, where $\gamma = (h, \lambda)$.

We estimate $g(x) = E(Y|X = x)$ by the local constant method:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_\gamma(x, X_i)}{\sum_{i=1}^n K_\gamma(x, X_i)}.$$  \hspace{1cm} (2)

If $\lambda_s = 0$ for all $s = 1, ..., r$,

then $L(X_i^d, x_i^d, 1) = 1(X_i^d = x_i^d)$.

$\hat{g}(x)$ reduces back to the conventional frequency estimator of $g(x)$.

If $\lambda_s = 1$ for some $s \in \{1, ..., r\}$, $l(X_{is}^d, x_s^d, 1) \equiv 1$,
\( \hat{g}(x) \) becomes unrelated to \( x^d_s \), 

i.e., \( x^d_s \) is completely removed from the regression model.

If \( h_s \) is sufficiently large, \( x^c_s \) is removed from the regression model.

Because \( \lim_{h_s \to \infty} w \left( \frac{X^c - x^c}{h_s} \right) = w(0) \), a constant.

We choose the smoothing parameters \( \gamma \equiv (h, \lambda) = (h_1, \ldots, h_p, \lambda_1, \ldots, \lambda_q) \) by minimizing the cross-validation function

\[
CV(h, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{g}_{-i}(X_i))^2 M(X_i),
\]

where

\[
\hat{g}_{-i}(X_i) = \frac{\sum_{j \neq i} Y_j K(X_i, X_j)}{\sum_{j \neq i} K(X_i, X_j)}
\]

is the leave-one-out kernel estimator of \( E(Y_i | X_i) \), \( M(\cdot) \) is a weight function.

N Hall, Li and Racine (2007), irrelevant regressors will be removed automatically.

Hall, Li and Racine (2007) requires ‘strong’ independence assumption.

\( (Y, \bar{X}) \perp \tilde{X} \)

Simulation results suggest ‘conditional’ independence is sufficient.

\( Y | \bar{X} \perp \tilde{X} \)

**Open question:** can one relax the independent assumption to conditional independent assumption?
2.2 Local Linear Estimation Method

One can also estimate \( g(x) \) by the local linear method. We choose \( a \) and \( b \) to minimize the following objective function:

\[
\min_{\{a,b\}} \sum_{j=1}^{n} \left[ Y_j - a - (X_j^c - x^c)'b \right]^2 K_\gamma(x, X_j).
\]

Let \( \hat{a} \) and \( \hat{b} \) denote the solution of the above minimization problem, it can be shown that \( \hat{a} = \hat{a}(x) \) estimates \( g(x) \) and \( \hat{b} = \hat{b}(x) \) estimates \( dg(x)/dx \).

\[
\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \left[ \sum_{j=1}^{n} \begin{pmatrix} 1 \\ (X_j - x)'(X_j - x)(X_j - x)' \end{pmatrix} K_\gamma(x, X_j) \right]^{-1} \left[ \sum_{j=1}^{n} \begin{pmatrix} 1 \\ (X_j - x)' \end{pmatrix} Y_j K_\gamma(x, X_j) \right]. \tag{4}
\]

We can choose the smoothing parameters by the least squares cross validation method:

\[
CV(h, \lambda) = \sum_{i=1}^{n} \left[ Y_i - \hat{g}_{-i}(X_i) \right]^2
\]

where \( \hat{g}_{-i}(X_i) \) is the local linear leave-one-out kernel estimator of \( g(X_i) \).

One replace \( x \) by \( X_i \) and replace \( \sum_{j=1}^{n} \) by \( \sum_{j\neq i}^{n} \) in (4), then \( \hat{a} \) becomes \( \hat{g}_{-i}(X_i) \)

Li and Racine (2003) examine the case that \( g(x) \) is nonlinear in \( x^c \) and that all \( x^d \) regressors are relevant ones.

They also study the case that \( g(x) \) is linear in some \( x^c_s \) an/or some \( x^d_s \) are irrelevant variables by simulations.
Simulation results show that:

\[ \hat{h}_s \rightarrow \infty \text{id}\ g(x) \text{ is linear in } x^c_s. \]

\[ \hat{\lambda}_s \rightarrow 1 \text{ if } x^d_s \text{ is an irrelevant regressor.} \]

**Open question:** can one prove the above result theoretically ?

How about \( x^c_s \) is an irrelevant variable ?

In practice, it is better to first use LC-CV method to remove irrelevant variables.

Then using the LL-CV method to detect linear components.

**Semiparametric Models**

Li and Racine (2007) extend it to the case of estimating a varying coefficient model.

**Weakly Dependent Data Case**

Li, Racine and Ouyang (2008), Su and Ullah (2008) extended the results of HLR to weakly dependent data case.

**All Regressors are Discrete**

Irrelevant discrete covariates will be smoothed out by the LSCV method with a positive probability, say \( \delta \).

Ouyang, Li and Racine (2008) argue that \( 0.5 < \delta < 1 \).

The simulation results reported in their paper suggest that \( \delta \in [0.6, 0.65] \).

R-package (np).

stata programs will be available soon.
2.3 Consistent Model Specification Tests

First consider a simple univariate regression model

\[ Y_i = g(X_i) + u_i, \]

where \( X_i \) is a univariate continuous random variable.

\[ H_0 : \quad P[E(Y|X) = \beta_0 + X\beta_1] = 1 \]

for some constant parameters \( \beta_0 \) and \( \beta_1 \).

Define \( u_i = Y_i - \beta_0 - X_i\beta_1 \).

\[ I = E[u_iE(u_i|X_i)f(X_i)] = E[(E(u_i|X_i))^2f(X_i)] \geq 0, \]

and \( I = 0 \) iff \( H_0 \) is true.

\( f(\cdot) \) is the density function of \( X_i \).

\[ I_n = \frac{1}{n} \sum_{i=1}^n \hat{u}_i \hat{E}_{-i}(u_i|X_i) \hat{f}(X_i) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \hat{u}_i \hat{u}_j K_h(X_x - x), \]

where \( K_h(v) = \frac{1}{h} K(v) \).

Under \( H_0 \), \( \hat{I}_n = O_p((nh^{1/2})^{-1}) \).

Under \( H_1 \) (i.e., \( H_0 \) is false), \( \hat{I}_n \) converges to a positive constant.

A standardized test is given by

\[ T_n = nh^{1/2}I_n/\hat{\sigma}_0, \]

where \( \hat{\sigma}_0^2 = 2[n(n-1)h]^{-1} \sum_{i=1}^n \sum_{j \neq i} \hat{u}_i^2 \hat{u}_j^2 K_h^2(X_i - x). \)
\[ T_n \xrightarrow{d} N(0, 1) \text{ under } H_0 \]

and

\[ T_n \text{ diverges to } +\infty \text{ at the rate of } n\sqrt{h}. \]

Need \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \).

Slowly versus fast changing function (high/low frequency data).

The power of a nonparametric test:

More accurate estimation of the unknown function is expected to lead to a better power of a test.

However, non-smoothing tests allow for \( h \) to be a constant (see Chapter 13 of Li and Racine).

Hsiao, Li and Racine (2007) suggest to use the LSCV method to select \((h, \lambda)\).

\[ H_0 : P\left[ E(Y_i|X_i) = m(X_i, \beta) \right] = 1 \text{ for some } \beta \in \mathcal{B}. \]  \hspace{1cm} (5)

\[ H_1 : P\left[ E(Y_i|X_i) = m(X_i, \beta) \right] < 1 \text{ for all } \beta \in \mathcal{B}. \]  \hspace{1cm} (6)

\[ I_n = n^{-1} \sum_{i=1}^{n} \hat{u}_i \hat{E}_{-i}(u_i|X_i)\hat{f}_{-i}(x_i) = n^{-1} \sum_{i=1}^{n} \hat{u}_i \left\{ n^{-1} \sum_{j=1, j \neq i}^{n} \hat{u}_j W_{h,ij} L_{\lambda,ij} \right\} \]

\[ = n^{-2} \sum_{i} \sum_{j \neq i} \hat{u}_i \hat{u}_j K_{\gamma,ij}, \]
where $K_{\gamma,ij} = W_{h,ij}L_{\lambda,ij}$ ($\gamma = (h, \lambda)$), $\hat{u}_i = y_i - m(x_i, \hat{\beta})$.

$$\hat{J}_n \overset{\text{def}}{=} n(\hat{h}_1 \ldots \hat{h}_q)^{1/2}\hat{I}_n/\sqrt{\hat{\Omega}} \to N(0,1)$$ in distribution under $H_0$,

where $\hat{\Omega} = [2(\hat{h}_1 \ldots \hat{h}_q)/n^2] \sum_i \sum_{j \neq i} \hat{u}_i^2 \hat{u}_j^2 W_{\hat{h},ij}^2 \hat{J}_{\lambda,ij}$.

$\hat{J}_n$ test diverges to $+\infty$ if $H_0$ is false, thus it is a consistent test.

Residual based bootstrap method performs well in finite sample applications.

$$\sup_{z \in \mathbb{R}} \left| P(\hat{J}_n^* \leq z|\{X_i, Y_i\}_{i=1}^n) - \Phi(z) \right| = o_p(1), \quad (7)$$

where $\hat{J}_n^* = n(\hat{h}_1 \ldots \hat{h}_q)^{1/2}\hat{I}_n^*/\sqrt{\hat{\Omega}_n^*}$, $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

### 2.4 Testing the Significance (Relevance) of Discrete Variables

$H_0 : E(Y|X, Z) = E(Y|X)$ almost everywhere (a.e.), \quad (8)

where $z$ is a discrete variable, while $x$ can contain both discrete and continuous components.

Assuming that $z$ takes values in $z \in \{0, 1, ..., c - 1\}$.

Under the null hypothesis $H_0$ is equivalent to: $m(X, z = l) = m(X, z = 0)$ for all $l \neq 0$.

$$I = \sum_{l=1}^{c-1} E \left\{ [m(X, z = l) - m(X, z = 0)]^2 \right\}. \quad (9)$$
\[ I \geq 0 \text{ and } I = 0 \text{ if and only if } H_0 \text{ is true.} \]

A feasible test statistic is given by

\[
\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{c-1} [\hat{m}(X_i, z = l) - \hat{m}(X_i, z = 0)]^2, \tag{10}
\]

where \( \hat{m}(X_i, z) \) is kernel estimator of \( m(X_i, z) = E(Y_i|X_i, z) \).

**A Bootstrap Procedure**

1. Randomly select \( Z^*_i \) from \( \{Z_j\}_{j=1}^{n} \) with replacement, and call \( \{Y_i, X_i, Z^*_i\}_{i=1}^{n} \) the bootstrap sample.

2. Use the bootstrap sample to compute the bootstrap statistic \( \hat{I}^*_n \).

3. Repeat steps 1 and 2 a large number of times to use the empirical distribution \( \{\hat{I}^*_n,j\}_{j=1}^{B} \) to approximate the distribution of \( \hat{I}_n \) under \( H_0 \).

The simulations reported in Racine, Hart and Li (2006) show that the above bootstrap procedure works well in finite sample applications.

**3 Nonparametric Regression Models With Non-stationary Data**

Related works:


Cai, Li and Park (2007), and Xiao (2007).

In this section we consider a semiparametric varying coefficient model

\[ Y_t = X'_t \beta(Z_t) + u_t, \]

where at least some (or all) components of \( X_t \) and \( Z_t \) are non-stationary.

### 3.1 Case (I) \( Z_t \) is \( I(0) \)

Cai, Li and Park (2007) consider the following model

\[ Y_t = X'_t \beta(Z_t) + u_t = X'_{t1} \beta_1(Z_t) + X'_{t2} \beta_2(Z_t) + u_t, \quad 1 \leq t \leq n, \quad (11) \]

where

\( X_{t1}, Z_t, u_t \sim I(0). \)

\( X_{t2} \sim I(1). \)

We allow for co-integration with varying co-integration coefficients \( \beta(Z_t) \).

Shorthand notation: \( A^{\otimes 2} = AA^T \ (A^{\otimes 1} = A) \) for a vector or matrix \( A \).

\[
\begin{pmatrix}
\hat{\beta}(z) \\
\hat{\beta}^{(1)}(z)
\end{pmatrix} = \left[ \sum_{t=1}^{n} \left( \begin{array}{c} X_t \\ (Z_t - z) X_t \end{array} \right) \right]^{\otimes 2} K \left( \frac{Z_t - z}{h} \right)^{-1} \sum_{t=1}^{n} \left( \begin{array}{c} X_t \\ (Z_t - z) X_t \end{array} \right) Y_t K \left( \frac{Z_t - z}{h} \right) .
\]

(12)

\[ X_{t2} - X_{t-1,2} = \eta_t, \quad \eta_t \sim I(0). \]

\[ X_{t,2}/\sqrt{n} \Rightarrow W_{\eta,2}(r). \]
Let $f_z(z)$ denote the marginal density of $Z_t$. Define, for $1 \leq k \leq 2$, 
$M_k(z) = E\left[X_t^{\otimes k} | Z_t = z\right].$

$S(z) = \begin{pmatrix} M_2(z) & M_1(z) W^{(1)}_{n,2}^T \\ W^{(1)}_{n,2} M_1(z)^T & W^{(2)}_{n,2} \end{pmatrix},$

Define $D_n = \begin{pmatrix} I_{d_1} & 0 \\ 0 & \sqrt{n} I_{d_2} \end{pmatrix}$, then under some regularity conditions, CLP show that

$$\sqrt{nh} D_n \left[ \hat{\beta}(z) - \beta(z) - \frac{1}{2} h^2 \mu_2(K) \beta^{(2)}(z) \right] \xrightarrow{d} MN(0, \Sigma_{\beta}(z)), \quad (13)$$

where $MN(0, \Sigma_{\beta}(z))$ is a mixed normal variable with mean zero and conditional covariance $\Sigma_{\beta}(z) = \sigma_u^2 \nu_0(K) S(z)^{-1} / f_z(z)$, $\mu_2(K) = \int_{-\infty}^{\infty} v^2 K(v) dv$ and $\nu_0(K) = \int_{-\infty}^{\infty} K^2(v) dv$.

Equation (13) implies that

$$\hat{\beta}_1(z) - \beta_1(z) = O_p(h^2 + (nh)^{-1/2}),$$

and

$$\hat{\beta}_1(z) - \beta_1(z) = O_p(h^2 + (n^2 h)^{-1/2}).$$

Both biases are $O(h^2)$.

$Var(\beta_1(z)) = O((nh)^{-1}).$

$Var(\beta_2(z)) = O((n^2 h)^{-1}).$

Optimal $h$ for $\hat{\beta}_1(z)$ is $h \sim n^{-1/5}$.

Optimal $h$ for $\hat{\beta}_2(z)$ is $h \sim n^{-2/5}$. 
\[ \hat{\beta}_1(z) - \beta(z) = O_p(n^{-2/5}). \]

\[ \hat{\beta}_2(z) - \beta_2(z) = O_p(n^{-4/5}). \]

In linear models, the estimated coefficient for \( I(1) \) variable is \( n \)-consistent,

The estimated coefficient for \( I(0) \) variable has the standard \( \sqrt{n} \) rate of convergence.

**3.2 Case (II): \( Z_t \) is \( I(1) \), \( X_t \) is \( I(0) \)**

\( X_t \sim I(0) \) and \( Z_t \sim I(1) \).

\[ Z_t = Z_{t-1} + v_t = Z_0 + \sum_{s=1}^{t} v_s, \text{ where } v_s \sim I(0). \]

For \( 0 \leq r \leq 1 \), \( Z_{[nr]}/\sqrt{n} \Rightarrow W_v(r), \)

where \( W_v(\cdot) \) is a Brownian motion on \([0, 1]\).

Cai et al (2007) derived the following result:

\[ \sqrt{n^{1/2} h} \left[ \hat{\beta}(z) - \beta(z) - h^2 B(z) \right] \xrightarrow{d} MN(0, \Sigma_1), \quad (14) \]

where \( B(z) = \mu_2(K) \beta^{(2)}(z)/2 \) and \( MN(0, \Sigma_1) \) is a mixed normal distribution with mean zero and conditional covariance \( \Sigma_1 = \sigma_u \nu_0(K) \left[ \text{E} \left( X_t X_t^T \right) L_{W(1, 0)} \right]^{-1}. \)

Equation (14) implies that

\[ \hat{\beta}(z) - \beta(z) = O_p \left( h^2 + \left( n^{1/4} h^{1/2} \right)^{-1} \right) \]
so that the optimal smoothing is that $h$ is proportional to $n^{-1/10}$.

$$\hat{\beta}(z) - \beta(z) = O_p(n^{-2/10}) = O_p(n^{-1/5}).$$

Thus, $h$ should converge to 0 at the fairly slow rate $n^{-1/10}$.

Because when $Z_t$ is $I(1)$, it returns to the fixed interval $[z - h, z + h]$ less often compared to the case when $Z_t$ is $I(0)$.

Therefore, one needs to let $h$ to converge to 0 slowly so as to balance the squared bias and the variance terms.

When $d = 1$ and $X_t \equiv 1$, the varying coefficient reduces to a simple regression model $Y_t = \beta(Z_t) + u_t$ ($Z_t$ is $I(1)$). The asymptotic variance in (14) simplifies to $\sigma^2_u L_W(1, 0)^{-1}$. It can be shown that $\hat{f}(z) \overset{d}{=} (nh)^{-1} \sum_{t=1}^n K(Z_t - z)/h$ consistently estimates $L_W(1, 0)$. Hence, in this case (14) can be equivalently written as

$$\left\{\hat{f}(z)/[\hat{\sigma}^2_u u_0]\right\}^{1/2} \sqrt{n^{1/2} h} \left[\hat{\beta}(z) - \beta(z) - h^2 B(z)\right] \overset{d}{\rightarrow} N(0, 1), \tag{15}$$

where $\hat{\sigma}^2_u = n^{-1} \sum_{i=1}^n [Y_t - \hat{\beta}(Z_t)]^2$ is a consistent estimator for $\sigma^2_u$. As expected, (15) is the same as given in Wang and Phillips (2006) who consider a nonparametric regression model with an $I(1)$ regressor.

### 3.3 Case (III): Both $X_t$ and $Z_t$ are $I(1)$

The same model. Now all $X_t$ and $Z_t$ are $I(1)$ variables, $u_t \sim I(0)$. 

14
Sun, Cai and Li (2008) show that
\[ \hat{\beta}(z) - \beta(z) = O_p(h^2 + (n^{3/2}h)^{-1/2}). \]

With optimal rate of \( h \sim n^{-3/10} \), we get
\[ \hat{\beta}(z) - \beta(z) = O_p(n^{-6/10}) = O_p(n^{-3/5}). \]

3.4 Other Type of Varying Coefficient Models

Bachmeier, Leelahanon and Li (2006) considered a the following semiparametric dynamic varying coefficient model:
\[ Y_t = \beta_1(Z_t) + Y_{t-1}\beta_2(Z_t) + u_t, \tag{16} \]
where \( Y_t \) is the rate of inflation, and \( Z_t \) is an I(1) variable ‘velocity of money’. Bachmier et al applied the above model to forecast U.S. inflation rate and show that the semiparametric varying coefficient dynamic model (with a non-stationary covariate) has smaller forecast mean squares error compared with the conventional linear model, or some nonparametric model using only stationary covariates.

Gu, Hernandez and Li (2008) consider the kernel-based estimation method of a varying coefficient model with a time trend variable appear as a parametric regressor:
\[ Y_t = X_{1t}^T\beta_1(Z_t) + t \beta_2(Z_t) + u_t, \tag{17} \]
\[ X_{1t}, \ Z_t \sim I(0). \]
\[ \hat{\beta}_1(z) - \beta_1(z) = O_p(h^2 + (nh)^{-1/2}) \]
\[ \hat{\beta}_1(z) - \beta_1(z) = O_p(h^2 + (n^3h)^{-1/2}), \]

Park and Hahn (1999)

Cai and ... (2008) consider similar time varying coefficient model as the one considered in Park and Hahn (1999) with non-stationary or near non-stationary (local to unit root) regressors \( X_t \). Cai and ... used local linear estimation method and derive the asymptotic distribution of their estimators.

### 3.5 Testing A Parametric Coefficient Functional Form

Sun, Cai and Li (2008) consider the problem of testing the null hypothesis

\[ H_0 : \ P(\beta(Z) = \beta_0) = 1 \]

\[ Y_t = X'\beta(Z_t) + u_t = X_{1t}'\beta_1(Z_t) + X_{2t}'\beta_2(Z_t) + u_t, \]

where \( X_{1t}, \ Z_t \) and \( u_t \) are \( I(0) \), and \( X_{2t} \) is \( I(1) \).

\[ \hat{I}_n = \frac{1}{n^3h} \sum_{t=1}^{n} \sum_{s \neq t} X'_t X_s \hat{u}_t \hat{u}_s K_{ts}, \]

(18)

here \( \hat{u}_t \) is the parametric residual.

Under some regularity conditions, Sun et al show that, under \( H_0 \),

\[ \hat{J}_n = n\sqrt{h} \hat{I}_n^{1/2} \sqrt{\hat{\sigma}^2} \xrightarrow{d} N(0, 1), \]
where $\hat{\sigma}_b^2 = \frac{1}{n^2h} \sum_{t=1}^{n} \sum_{s \neq t} \tilde{u}_t^2 \tilde{u}_s^2 [X_t'X_s]^2 K_{ts}^2$, $\tilde{u}_t = Y_t - X_t'\bar{\beta}(-t) (Z_t)$ is the leave-one-out nonparametric residual.

The power of the test.

Under some regularity conditions and under $H_1$,

(i) If $Pr[\beta_2(Z_t) = \beta_{20}] < 1$ for any $\beta_{20} \in B_2 \subset \mathbb{R}^{d_2}$, then

$$Pr[J_n > B_n] \rightarrow 1 \text{ as } n \rightarrow \infty \text{ for any non-stochastic sequence } B_n = o(n^2\sqrt{h}).$$

(ii) If $Pr[\beta_2(Z_t) = \beta_{20}] = 1$ for some $\beta_{20} \in B_2$, and $Pr[\beta_1(Z_t) = \beta_{10}] < 1$ for any $\beta_{10} \in B_1 \subset \mathbb{R}^{d_1}$, then

$$Pr[J_n^b > B_n] \rightarrow 1 \text{ as } n \rightarrow \infty \text{ for any non-stochastic sequence } B_n = o(n\sqrt{h}).$$

Sun et al show that when $\beta_1(z) = \beta_{10}$, and $\beta_2(z) \neq \beta_{20}$, the least squares estimator $\hat{\beta}_{10}$ diverges to $+\infty$ at the rate of $\sqrt{n}$.

It it is very important to test for the correction model specification when there exists integrated regressors.

3.6 Test for Co-Integration in Semiparametric Varying Coefficient Models

Sun and Li (2008) consider the problem of testing cointegration in a semiparametric varying coefficient models.

$$Y_t = X_t'\beta(Z_t) + u_t,$$
where $X_t$ is a $d \times 1$ vector of $I(1)$ variables, $Z_t$ is $I(0)$ scalar. $u_t$ follows an AR(1) process:

$$u_t = \rho u_{t-1} + \epsilon_t,$$

where $\epsilon_t$ is a zero mean stationary process.

The null hypothesis is $H_0$: $\rho = 0$ and the alternative hypothesis is $H_1$: $0 \leq \rho < 1$.

Alternatively, one can test whether the variance of $u_t$ is a constant, or it is proportional to $t$, Xiao (2007).

The null and the alternative are switched.

4 Nonparametric/Semiparametric Estimation of Panel Data Models

4.1 Nonparametric Panel Data Regression Model with Fixed Effects

Consider the following nonparametric panel data model with fixed effects

$$Y_{it} = g(X_{it}) + \mu_i + \nu_{it}, \quad (i = 1, \ldots, n; \ t = 1, \ldots, m) \quad (19)$$

$n$ is large and $m$ is fixed.

Henderson, Carroll and Li (2008)

$$\tilde{Y}_{it} \equiv Y_{it} - Y_{i1} = g(X_{it}) - g(X_{i1}) + \nu_{it} - \nu_{i1}. \quad (20)$$
Define $\tilde{\nu}_{it} = \nu_{it} - \nu_{i1}$ and $\tilde{\nu}_i = (\tilde{\nu}_{i2}, ..., \tilde{\nu}_{im})'$. The variance-covariance matrix of $\tilde{\nu}_i$, defined as $\Sigma = \text{cov}(\tilde{\nu}_i|Z_{i1}, ..., Z_{im}) = \text{cov}(\tilde{\nu}_i)$, is given by

$$
\Sigma = \sigma_\nu^2(I_{m-1} + e_{m-1}e_{m-1}')
$$

where $\sigma_\nu^2 = E(\nu_{it}^2)$, $I_{m-1}$ is an identity matrix of dimension $(m-1) \times (m-1)$, and $e_{m-1}$ is a $(m-1) \times 1$ vector of ones. It is easy to check that

$$
\Sigma^{-1} = \sigma_\nu^{-2}(I_{m-1} - e_{m-1}e_{m-1}'/m).
$$

A profile likelihood approach.

The criterion function for individual $i$ is given by

$$
\mathcal{L}_i(\cdot) = \mathcal{L}(Y, g_i) = -\frac{1}{2}(\tilde{Y}_i - g_i + g_{i1}e_{m-1})'\Sigma^{-1}(\tilde{Y}_i - g_i + g_{i1}e_{m-1}),
$$

where $\tilde{Y}_i = (\tilde{Y}_{i2}, ..., \tilde{Y}_{im})'$, $g_{it} = g(X_{it})$ and $g_i = (g_{i2}, ..., g_{im})'$. Defining $\mathcal{L}_{i,tg} = \partial \mathcal{L}_i(\cdot)/\partial g_{it}$, and $\mathcal{L}_{i,tsg} = \partial^2 \mathcal{L}_i(\cdot)/(\partial g_{it}\partial g_{is})$, from (21) we obtain

$$
\mathcal{L}_{i,1g} = -e_{m-1}'\Sigma^{-1}(\tilde{Y}_i - g_i + g_{i1}e_{m-1});
$$

and $\mathcal{L}_{i, tg} = c_{t-1}'\Sigma^{-1}(\tilde{Y}_i - g_i + g_{i1}e_{m-1})$ for $t \geq 2$, where $c_{t-1}$ is a vector of dimension $(m-1) \times 1$ with the $(t-1)$ element being 1 and all other elements being 0.

$$
K_h(v) = \Pi_{j=1}^q h_j^{-1}k(v_j/h_j).
$$

Denote by $(X_{it}-x)/h = \{(X_{it,1}-x_1)/h_1, ..., (X_{it,q}-x_q)/h_q\}'$ and $G_{it}(x, h) = [1, \{(X_{it}-x)/h\}]'$, $g^{(1)}(x) = \partial g(x)/\partial x$. 19
We estimate the unknown function \( g(x) \) by solving the first order condition

\[
0 = \sum_{i=1}^{n} \sum_{t=1}^{m} K_h(X_{it}, x)G_{it}(z, h) \mathcal{L}_{i, tg}[Y_i, \hat{g}(X_{i1}), \ldots, \hat{g}(x) + \{(X_{it} - x)/h\}\hat{g}^{(1)}(x), \ldots, \hat{g}(Z_{im})],
\]

(22)

where the argument \( \mathcal{L}_{i, tg} \) is \( \hat{g}(X_{is}) \) for \( s \neq t \) and \( \hat{g}(x) + \{(X_{it} - x)/h\}\hat{g}^{(1)}(x) \) when \( s = t \). The FOC (22) leads to the following iterative equation.

\[
\begin{pmatrix}
\hat{\alpha}_0(z) \\
\hat{\alpha}_1(z)
\end{pmatrix} = \begin{pmatrix}
\hat{\theta}_{[l]}(z) \\
\hat{\theta}_{[l]}^{(1)}(z)
\end{pmatrix} = D_1^{-1}(D_2 + D_3),
\]

where

\[
D_1 = n^{-1} \sum_{i=1}^{n} \left\{ e_{m-1}^T \Sigma^{-1} e_{m-1} K_h(Z_{i1} - z) G_{i1} G_{i1}^T + \sum_{t=2}^{m} c_t^T \Sigma^{-1} c_t K_h(Z_{it} - z) G_{it} G_{it}^T \right\}
\]

\[
D_2 = n^{-1} \sum_{i=1}^{n} \left\{ e_{m-1}^T \Sigma^{-1} e_{m-1} K_h(Z_{i1} - z) G_{i1} \hat{\theta}_{[l-1]}(Z_{i1}) \right. \\
+ \left. \sum_{t=2}^{m} c_t^T \Sigma^{-1} c_t K_h(Z_{it} - z) G_{it} \hat{\theta}_{[l-1]}(Z_{it}) \right\};
\]

\[
D_3 = n^{-1} \sum_{i=1}^{n} \left\{ \sum_{t=2}^{m} K_h(Z_{it} - z) G_{it} c_{t-1}^T \Sigma^{-1} H_{i,[l-1]} - K_h(Z_{i1} - z) G_{i1} e_{m-1}^T \Sigma^{-1} H_{i,[l-1]} \right\},
\]

\[
H_{i,[l-1]} = \begin{pmatrix}
Y_{i2} - \hat{\theta}_{[l-1]}(Z_{i2}) \\
\vdots \\
Y_{im} - \hat{\theta}_{[l-1]}(Z_{im})
\end{pmatrix} - \{Y_{i1} - \hat{\theta}_{[l-1]}(Z_{i1})\} e_{m-1}.
\]

It usually takes only three to four iterations to achieve convergence.

A gauss program is available from Henderson (email: djhender@binghamton.edu).

No close form expression for \( Bias(\hat{g}(x)) \), \( Bias(\hat{g}(x)) = O(h^2) \).

The asymptotic variance of \( \sqrt{nh_1 \ldots h_q} (\hat{g}(x) - g(x)) \) is

\[
\frac{(m - 1)\sigma^2}{f(x)}.
\]

20
**Open question** (conjecture): the estimator is asymptotically normally distributed.

**Other Semiparametric Models**

Sun and Carroll (2008) consider the varying coefficient model and suggest some non-iterative estimation procedures.

Su and Ullah (2006).
A Brownian motion is a random process, indexed by $r$, for $r \in [0, 1]$.

**Definition 4.1 [Brownian Motion]** We say $W(r)$ is a standard Brownian motion if $W(r)$, for $r \in [0, 1]$, satisfies the following conditions:

(i) $W(0) = 0$,

(ii) For all $r > s$, $W(r) - W(s) \sim N(0, r - s)$, and for any partition $0 \leq r_1 < t_2 < \ldots < t_k \leq 1$, the changes $[W(r_2) - W(r_1)], \ldots, [W(r_k) - W(r_{k-1})]$ are independent multivariate normal.

(iii) For any given realization (sample path), $W(r)$ is continuous in $r$ with probability one.

Let $Y_t$ follow an unit root process

$$Y_t = Y_{t-1} + u_t, \quad (t = 1, \ldots, n)$$

where $u_t$ is i.i.d. $(0, \sigma_u^2)$.

It can be shown that

$$\frac{X_t}{\sqrt{n}} \Rightarrow W_u(r) \equiv \sigma_u W(r),$$

where $W(r)$ is a standard Brownian motion.

$$X_t/\sqrt{n} = n^{-1/2}[u_t + u_{t-1} + \ldots + u_1] \sim N(0, (t/n)\sigma_u^2) \sim \sigma_u N(0, r) \sim \sigma_u W(r).$$

We now derive the limiting distribution of $\hat{\rho}$ under the assumption that
\[ \rho = 1. \]

\[ \hat{\rho} - 1 = \frac{\sum_{t=1}^{T} Y_{t-1} u_t}{\sum_{t=1}^{T} Y_{t-1}^2}. \quad (23) \]

In practice, the critical values for \( T(\hat{\rho} - 1) \) can be obtained by Monte Carlo simulations, see case 1 in Table B.5 (Hamilton, p.762). We can also use simulations to generate the critical values quite easily.

Note that

\[ \int_{0}^{1} g(x) dx = \int_{0}^{1} g(r) dr = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i)(x_i - x_{i-1}) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g\left(\frac{i}{n}\right). \]

When \( Y_t = Y_{t-1} + u_t \) with \( u_t \) is white noise process, we have the following useful results:

\[ \frac{Y_t}{\sqrt{T}} \sim W_u(r) = \sigma_u W(r), \quad \frac{u_t}{\sqrt{n}} \sim dW_u(r) = \sigma_u dW(r) \quad \frac{1}{n} \sim dr, \quad \sum_{t=1}^{n} \sim \int_{0}^{1}\ (24) \]

We already explained that \( Y_t/\sqrt{T} \sim \sigma W(r). \)

\[ u_t/\sqrt{n} = (Y_t - Y_{t-1})/\sqrt{n} = \Delta Y_t/\sqrt{n} \sim \Delta W_u(r) \sim dW_u(r). \]

\[ \int_{0}^{1} \sim \sum_{t=1}^{n} \text{ and } 1/n \sim dr \text{ should be obvious from the relationship between summation and integration.} \]

\( t = 1, \ldots, n \) and \( r = t/n \in [0, 1]. \)

Now we show that one can use the result of (24) to derive the asymptotic distribution of \( T(\hat{\rho} - 1) \) with great ease.

\[ n(\hat{\rho} - 1) = \frac{n^{-1} \sum_{t=1}^{n} Y_{t-1} u_t}{n^{-2} \sum_{t=1}^{n} Y_{t-1}^2} = \frac{\sum_{t=1}^{n} \frac{Y_{t-1} u_t}{\sqrt{n}}}{n^{-1} \sum_{t=1}^{n} \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2} \sim \frac{\int_{0}^{1} W(r) dW(r)}{\int_{0}^{1} W(r)^2 dr}. \quad (25) \]
**Case 2:** No constant term but estimated with a constant term.

The true process is

\[ Y_t = Y_{t-1} + u_t, \]

with \( u_t \) i.i.d. \((0, \sigma^2)\).

Suppose one estimate the following model by OLS

\[ Y_t = \alpha + \rho Y_{t-1} + u_t. \]

Then it can be shown that

\[ n(\hat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)d(W) - W(1) \int W(r)dr}{\int [W(r)]^2dr - [\int W(r)dr]^2}. \]  

(26)

\[ \hat{\rho} = \frac{\sum_t (Y_{t-1} - \bar{Y})(Y_t - \bar{Y})}{\sum_t (Y_{t-1} - Y)^2} \]

\[ n(\hat{\rho} - 1) = \frac{n^{-1} \sum_t (Y_{t-1} - \bar{Y})(u_t - \bar{u})}{n^{-2} \sum_t (Y_{t-1} - Y)^2}. \]

\[ n^{-2} \sum_t (Y_{t-1} - \bar{Y})^2 = n^{-2} \sum_{t=1}^n Y_t^2 - n^{-1}(\bar{Y})^2 \]

\[ = n^{-1} \sum_{t=1}^n (Y_t/\sqrt{n})^2 - (\bar{Y}/\sqrt{n})^2 \]

\[ \sim \int_0^1 W_u(r)^2dr - \left[ \int_0^1 W_u(r)dr \right]^2. \]

Since

\[ \bar{Y}/\sqrt{n} = n^{-1} \sum_{t=1}^n Y_t/\sqrt{n} \sim \int_0^1 W_u(r)dr. \]
Derive the asymptotic distribution of the above and check Hamilton’s book for the result.