Using generalized estimating equations to estimate nonlinear models with spatial data

Cuicui Lu, Weining Wang, Jeffrey Wooldridge
Nanjing University Business School
Humboldt U zu Berlin and City, U of London
Michigan State U

Aug. 2019

Outline

- Introduction
- Econometric model
- The GEE estimation methods
- Two examples: 1. binary response data; 2. count data
- Monte Carlo simulations
- An empirical application to inflow FDI

Outline

- We study the estimation of nonlinear models with spatial (cross-sectional) correlated data.
- We estimate the parameters of interest within the Quasi-MLE framework, with a correct conditional mean assumption.
- Based on QMLE ignoring spatial dependence, we study possible efficiency gain of generalized estimating equations (GEE).
- We study the binary response data and count data with spatial correlation in an error term.
- For binary response data, we find no efficiency gain for the Probit GEE.
- For count data, we find Negative Binomial II GEE performs better than Poisson GEE.

Introduction

- In empirical economic, social studies and finance. There are examples of discrete data which exhibit spatial correlations due to the closeness of geographical locations of individuals or agents.
- Such discrete data includes binary response 0,1 and count data 0,1,2,...,1000,....
- **Example 1: Firm technology spillover.** The number of patents a firm received shows correlation with that received by other firms near by. This may be due to a technology spillover effect from other firms. E.g. Bloom et al (2013).
- **Example 2: Neighborhood effect.** There is a causal effect between the individual decision whether to own stocks and the average stock market participation of the individual's community. E.g. Brown et al (2008).
- The dependent variable has two characteristics: nonlinearity and spatial correlated (cross-sectionally) correlated.

Econometric Model

- The estimation of nonlinear models can be formed within **M-estimation** framework as a minimization or maximization problem. In particular, nonlinear least squares (NLS) and maximum likelihood (MLE) are two examples.
- Let $\{\mathbf{w}_N\} \equiv \{(\mathbf{x}_i, y_i), i = 1, 2, ..., N\}$ sampled on a lattice D_N. $\mathbf{w}_i$ is the observed data obtained at location s_i.
- We write the problem as an M-estimation of $\boldsymbol{\theta}_0$ given by minimizing the objective function Q_N as follows,

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_N\left(\mathbf{w}_N, \mathbf{D}_N; \boldsymbol{\theta}\right). \tag{1}$$

- In particular, $Q_N\left(\mathbf{w}_N, \mathbf{D}_N; \boldsymbol{\theta}\right)$ can be expressed as a sample average:

$$Q_N\left(\mathbf{w}_N, \mathbf{D}_N; \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N} q_i\left(\mathbf{w}_i, \mathbf{D}_N; \boldsymbol{\theta}\right), \tag{2}$$

where $q_i\left(\mathbf{w}_i, \mathbf{D}_N; \boldsymbol{\theta}\right)$ is some real valued function defined on $\boldsymbol{\Theta}$. For MLE, q_i is the (negative) log likelihood.

Maximum likelihood estimation (MLE)

- Maximum likelihood estimation (MLE) is a widely used method in estimating both linear and nonlinear models.
- A **full MLE** needs to specify the joint distributions of spatial random variables. This includes correctly specifying the marginal and the conditional distributions.
- However, given a spatial data set, the dependence structure is generally unknown. If the joint distribution of the variables is misspecified, MLE is generally not consistent.
- **Partial MLE** uses part of the joint distributions but needs to correctly specify the partial distribution. Wang, Iglesias and Wooldridge (2013) use a bivariate probit partial MLE to improve the estimation efficiency with a spatial probit model. Distribution of multivaraite probit is hard to compute.
- There is a quite related terminology, **composite likelihood**, whose motivation is to avoid computing or modelling the joint distributions of high dimensional random processes (Varin, Reid & Firth 2011).

QMLE in the LEF

- **Question:** Can we use less distributional assumption but improve estimation efficiency?
- QMLE requires only that we correctly specify the conditional mean and variance, and the relationship between conditional mean and variance of the dependent variable.
- Gourieroux, Monfort, and Trognon (1984): Using a density that belongs to a **linear exponential family** (LEF), QMLE is consistent if we correctly specify the conditional mean with other features of the density misspecified.
- Normal, Bernoulli, Poisson, exponential log likelihoods are all members of the linear exponential family.
- NLS is a QMLE based on the normal density function.
- Generalized linear models (GLM) with a link function to introduce nonlinearity, see McCullagh and Nelder (1989).

QMLE in the LEF

- We assume the conditional mean is correctly specified.

$$\mathrm{E}\left(y_i|\mathbf{x}_i\right) = m_i\left(\mathbf{x}_i, \boldsymbol{\theta}_0\right), i = 1, 2, ..., N \tag{3}$$

- Based on a conditional density $f\left(\cdot|\mathbf{x};\boldsymbol{\theta}\right)$ in the linear exponential family, the log likelihood for each individual is

$$l_i\left(\theta\right) \equiv \log f\left(y_i|\mathbf{x}_i;\boldsymbol{\theta}\right), i = 1, 2, ..., N. \tag{4}$$

  Ignoring any spatial dependence, the partial (or pooled) QMLE is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \sum_{i=1}^{N} \log f\left(y_i|\mathbf{x}_i;\boldsymbol{\theta}\right) \tag{5}$$

- One characterization of QMLE in LEF is that the individual score function has the following form:

$$\mathbf{s}_i\left(\boldsymbol{\theta}\right) = \nabla m_i\left(\mathbf{x}_i;\boldsymbol{\theta}\right)' \left[y_i - m_i\left(\mathbf{x}_i;\boldsymbol{\theta}\right)\right] / v_i\left(m_i\left(\mathbf{x}_i;\boldsymbol{\theta}\right)\right). \tag{6}$$

- QMLE is consistent. But it is not the most efficient.

Generalized estimating equations (GEE)

- The GEE approach was first extended to panel data by Zeger and Liang (1986), and then used in panel data cases. For e.g. Albert and McShane (1995).
- **A generalized estimating equation** is used to estimate the parameters of a generalized linear model with a possible unknown correlation between outcomes (Liang and Zeger 1986). GEE is characterized by its first order condition

$$\mathbf{S} = \nabla \mathbf{m}' (\boldsymbol{\theta}) \mathbf{W}^{-1} (\gamma, \boldsymbol{\theta}) [\mathbf{y} - \mathbf{m} (\boldsymbol{\theta})] , \tag{7}$$

  where $\nabla_{\theta} \mathbf{m}' (\boldsymbol{\theta})$ is the gradient of $\mathbf{m} (\boldsymbol{\theta})$.
- $\mathbf{W}$ is the working variance-covariance matrix. Other than just account for the variances, it specifies a working correlation matrix.

Implication from estimating linear models

- Bester, Conley and Hansen (2011) provide a cluster covariance matrix estimator using fixed-G asymptotics using a fixed small number of groups.
- Our asymptotic analysis based on large number of groups.
- Implication from linear regression model (Lu and Wooldridge 2017).
  - Groupwise GLS (QGLS) estimation can improve estimation efficiency compared to OLS.
  - It is computationally easier but does not loose much efficiency compared to GLS.
  - Robust standard errors can be provided.

GEE with spatial data

- We divide spatial data into G groups according to distance measures. Suppose each group has L observations. $N = G \times L$. Let $\{(\mathbf{X}_g, \mathbf{y}_g)\}$ be the observations for group g, $g = 1, 2, ..., G$. $\mathbf{X}_g$ is an $L \times K$ matrix and $\mathbf{y}_g$ is an $L \times 1$ vector.
- The conditional mean for group g is

$$\mathrm{E}\left(\mathbf{y}_g | \mathbf{X}_g; \boldsymbol{\theta}\right) = \mathbf{m}_g\left(\mathbf{X}_g; \boldsymbol{\theta}_0\right). \tag{8}$$

- The quasi-score equation, which is also can be called a first order condition for GEE is defined as follows:

$$\mathbf{s}_g\left(\boldsymbol{\theta}, \hat{\gamma}\right) = \nabla \mathbf{m}_g'\left(\boldsymbol{\theta}\right) \mathbf{W}_g^{-1}\left(\hat{\theta}, \hat{\gamma}\right)\left[\mathbf{y}_g - \mathbf{m}_g\left(\boldsymbol{\theta}\right)\right]. \tag{9}$$

$$S_G\left(\boldsymbol{\theta}, \hat{\gamma}\right) = \frac{1}{G}\sum_{g=1}^{G} \nabla \mathbf{m}_g'\left(\boldsymbol{\theta}\right) \mathbf{W}_g^{-1}\left(\hat{\theta}, \hat{\gamma}\right)\left[\mathbf{y}_g - \mathbf{m}_g\left(\boldsymbol{\theta}\right)\right]. \tag{10}$$

GEE with spatial data

- **First step**: use Quasi MLE to get a consistent estimators for the mean parameters. And use the first step residuals to estimate a working varaince-covariance matrix.
- **Second step**: Divide individuals into groups. Use the estimated working covariance matrix to do a feasible multivariate weighted nonlinear least squares (MWNLS). Use the correlation information within groups and ignore those between groups.
- A common way to model the working variance structure is to let $\hat{\mathbf{W}} = \hat{\mathbf{V}}^{1/2}\hat{\mathbf{R}}\hat{\mathbf{V}}^{1/2}$.
- The working variances is specified by the distribution:
  - For binary response $v_i = m_i\left(1 - m_i\right)$
  - For Poisson $v_i = m_i$
  - For Negative Binomial II $v_i = m_i + \eta \cdot m_i^2$

Working Correlation Matrix

In Stata, the working correlation matrix can have a few forms: independent, exchangeable, AR(1), unstructured and fixed matrix. For example, a 4×4 working correlation matrix can be

$$R_{ind} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; R_{EX} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Working Correlation Matrix

For example, a 4×4 working correlation matrix can be

$$R_{AR1} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}, \quad R_{fixed} = \begin{bmatrix} 1 & \rho/d_{12} & \rho/d_{13} & \rho/d_{14} \\ \rho/d_{21} & 1 & \rho/d_{23} & \rho/d_{24} \\ \rho/d_{31} & \rho/d_{32} & 1 & \rho/d_{34} \\ \rho/d_{41} & \rho/d_{42} & \rho/d_{43} & 1 \end{bmatrix}$$

Asymptotic Theory

- NED on mixing random field, more general notion of dependence
- Preserving after Lipschitz transformation, infinite input
- GEE with orthogonal score by construction.
  $\lim_{G\to\infty}\{\frac{1}{M_G|D_G|}\sum_g \mathrm{E}[\nabla_\gamma s_g\left((\tilde{\tilde{\theta}}, \tilde{\gamma}, \theta^0)\right)] = 0$

Asymptotic Theory

Bolthausen (1982), Conley (1999), Lee (2004), Jenish & Prucha (2009, 2012). We adopt notations and definitions in Jenish and Prucha (2009, 2012).

Theorem

(Consistency) Under A.1)-A.8) the GEE-estimator is consistent, that is, $\nu(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^0) \rightarrow_p 0$ *as* $G \rightarrow \infty$.

Theorem

Under A.1) - A.11), we have $AV(\theta^0) \stackrel{\text{def}}{=} \mathbf{H}_\infty^\top \mathbf{A}\mathbf{S}_\infty \mathbf{H}_\infty$.

$$\sqrt{G}AV(\theta^0)^{-1/2}(\hat{\theta} - \theta^0) \Rightarrow \mathbb{N}(0, I_p). \tag{11}$$

$$\hat{\mathbf{A}} = \frac{1}{|D_G|}\sum_g \nabla\hat{\mathbf{m}}_g^\top \hat{\mathbf{W}}_g^{-1}\nabla\hat{\mathbf{m}}_g, \tag{12}$$

$$\hat{\mathbf{B}} = \frac{1}{|D_G|}\sum_g\sum_{h\neq g} k(d_{gh})\nabla\hat{\mathbf{m}}_g^\top \hat{\mathbf{W}}_g^{-1}\hat{\mathbf{u}}_g\hat{\mathbf{u}}_h^\top\hat{\mathbf{W}}_h^{-1}\nabla\hat{\mathbf{m}}_h^\top, \tag{13}$$

where $\nabla\hat{\mathbf{m}}_g \equiv \nabla\hat{\mathbf{m}}_g\left(\hat{\boldsymbol{\theta}}\right)$, $\hat{\mathbf{W}}_g \equiv \hat{\mathbf{W}}_g(\hat{\gamma},\hat{\theta})$.

The estimator of $\mathrm{AV}\left(\boldsymbol{\theta}^0\right)$ which is robust to misspecification of the variance covariance matrix is

$$\begin{aligned}\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\theta}}\right) &= |D_G|\left(\sum_g \nabla\hat{\mathbf{m}}_g^\top\hat{\mathbf{W}}_g^{-1}\nabla\hat{\mathbf{m}}_g\right)^{-1} \\ &\quad \left(\sum_g\sum_{h(\neq g)} \nabla\hat{\mathbf{m}}_g^\top\hat{\mathbf{W}}_g^{-1}k\left(d_{gh}\right)\hat{\mathbf{u}}_g\hat{\mathbf{u}}_h^\top\hat{\mathbf{W}}_h^{-1}\nabla\hat{\mathbf{m}}_h\right) \\ &\quad \left(\sum_g \nabla\hat{\mathbf{m}}_g^\top\hat{\mathbf{W}}_g^{-1}\nabla\hat{\mathbf{m}}_g\right)^{-1},\end{aligned} \tag{14}$$

B.1) $\hat{\mathbf{u}}_g - \mathbf{u}_g = C_g \Delta_g$, where C_g is a $L \times p$, and Δ_g is a $p \times 1$ dimensional vector, with the condition that $|C_g|_2 = \mathcal{O}_p(1)$, and $|\Delta_g|_2 = \mathcal{O}_p((pG)^{-1/2})$.

B.2) The moment is bounded by a constant $\max_{h:\rho(h,g)\leq h_g} \mathrm{E}\,|Z_h|^{q'} \leq ML^2$, $q' \geq 1$, and M is a constant, where $Z_h \stackrel{\text{def}}{=} \nabla \mathbf{m}_h^\top(\theta^0)\mathbf{W}_h^{-1}(\theta^0, \gamma^0)\mathbf{u}_h$.

B.3) $|k(d_{gh}) - 1| \leq C_k |d_{gh}/h_g|^{\rho_K}$ for $d_{gh} \leq 1$ for some constant $\rho_k \geq 1$ and $0 < C_k < \infty$
$M_G^{-2}|D_G|^{-1} \sum_g \sum_h |\rho(g,h)/h_g|^{\rho_k} \|e_i^\top Z_g^\top\| \|Z_h e_j\| = \mathcal{O}(1)$.

B.4) Assume that $h_g^{d/q'}|D_G|^{-1}L^{d/q'}L^2 = \mathcal{O}(1)$,
$h_g^{2d} L^{2d} \sum_{r=1}^{\infty} r^{(d\tau^*+d)-1}\hat{\alpha}^{\delta/(2+\delta)}(r) = \mathcal{O}(G)$, and
$h_g^{2d} \sum_{r=1}^{\infty} L^{2d} r^{d-1}\psi((r-h_g)_+) = \mathcal{O}(G)$,
$((r-h_g)_+ = \max(r-h_g, 0))$ where δ is a constant and $\delta^* = \delta\tau/(2+\delta)$.

Theorem

Under assumption B.1)- B.4) and A.1) - A.8). The variance-covariance estimator is consistent. $\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\theta}}\right) \rightarrow_p \mathrm{AV}\left(\boldsymbol{\theta}^0\right)$.

Spatial Probit Model

- The spatial probit model can be formulated as follows:

$$\begin{aligned} y_i &= 1\left[y_i^* \geq 0\right], \\ y_i^* &= \mathbf{x}_i \boldsymbol{\beta} + e_i. \end{aligned} \tag{16}$$

- One specification as in Pinkse and Slade (1998) is

$$\begin{aligned} e &= \rho W e + u, \\ e &= (I - \rho W)^{-1} u. \end{aligned} \tag{17}$$

  where $W_{ii} = 0$ and $W_{ij} = 1/d_{ij}$ for $i \neq j$.
- For simplicity, we directly model the spatial correlation as a multivariate normal distribution:

$$e \tilde{} \mathrm{N}(0, \Omega), \tag{18}$$

  where $\Omega_{ii} = 1$ and $\Omega_{ij} = \rho/d_{ij}$.

Spatial Probit Model

- Bernoulli QMLE is obtained by maximizing the Probit log-likelihood.

$$\hat{\beta}_{PQMLE} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{N} y_i \log\Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right) + \sum_{i=1}^{N} (1-y_i)\log\left[1-\Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right)\right].$$

- The two-step GEE estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{\mathrm{GEE}} = \arg\min_{\beta} \sum_{g=1}^{G} \left(\mathbf{y}_g - \Phi\left(\mathbf{x}_g\boldsymbol{\beta}\right)\right)' \hat{\mathbf{W}}_g^{-1} \left(\mathbf{y}_g - \Phi\left(\mathbf{x}_g\boldsymbol{\beta}\right)\right). \quad (19)$$

The first order condition is

$$S_G = \sum_{g=1}^{G} \phi\left(\mathbf{x}_g\boldsymbol{\beta}\right)' \hat{\mathbf{W}}_g^{-1} \left(\mathbf{y}_g - \Phi\left(\mathbf{x}_g\boldsymbol{\beta}\right)\right). \quad (20)$$

Spatial Probit Model

$\hat{\mathbf{W}}_g = \hat{\mathbf{V}}_g{}^{1/2}\hat{\mathbf{R}}_g\hat{\mathbf{V}}_g^{1/2}$.
An estimator for the working variance matrix for each group is

$$\hat{\mathbf{V}}_g = \begin{pmatrix} \check{v}_1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \check{v}_2 & 0 & & & 0 \\ 0 & 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \check{v}_l & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & \check{v}_L \end{pmatrix}, \tag{21}$$

$$\check{v}_l = \Phi\left(\mathbf{x}_l\check{\beta}\right)\left[1 - \Phi\left(\mathbf{x}_l\check{\beta}\right)\right], \quad l = 1, ..., L. \tag{22}$$

- Can use an independent, exchangeable, AR(1) working correlation matrix.
- Can estimate a fixed working correlation matrix:

(1) Assume the working correlation matrix for $u_i = y_i - \Phi\left(\mathbf{x}_i \boldsymbol{\beta}\right)$ is $\mathbf{R}$ with the ijth element $\mathbf{R}_{ij} = \mathbf{C}_{ij}\left(d_{ij}, \lambda\right)$, where $\mathbf{C}_{ij}\left(d_{ij}, \lambda\right)$ is a function that increases in λ and decreases in d_{ij}.
(2) Let $\hat{r}_i = \check{u}_i / \sqrt{\check{v}_i}$, for $i = 1, 2, ..., N$, be the standardized residuals. $\hat{\mathbf{C}}_{ij}$ equals the sample correlation of $\check{u}_i / \sqrt{\check{v}_i}$ and $\check{u}_j / \sqrt{\check{v}_j}$. Let $\hat{\mathbf{R}} \equiv \mathbf{R}\left(\mathbf{D}_G, \hat{\lambda}\right)$ and $\hat{\mathbf{R}}_g$ stand for the correlation matrix $\mathbf{R}_g\left(\mathbf{D}_g, \hat{\lambda}\right)$ for the gth group. For example, use the minimum distance estimator:

$$\min_{\lambda} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\frac{\check{u}_i \check{u}_j}{\sqrt{\Phi\left(\mathbf{x}_i \check{\beta}\right)\left[1 - \Phi\left(\mathbf{x}_i \check{\beta}\right)\right]} \sqrt{\Phi\left(\mathbf{x}_j \check{\beta}\right)\left[1 - \Phi\left(\mathbf{x}_j \check{\beta}\right)\right]}} - \frac{\lambda}{d_{ij}} \right]^2 .$$

Monte Carlo Simulation

(1) Each individual resides on an intersection of the square lattice. Thus the pairwise distance can be calculated using the coordinates of each observation.
(2) According to pairwise distances, generate the pairwise correlations. In the simulation, each observation is correlated with all other observations.
(3) The spatial correlation in the error between observation i and j is $\frac{\rho}{d_{ij}}$, $\rho = 0, 0.2, 0.4, 0.6$.
(4) We compares QMLE estimator (GEE with independent correlation matrix), GEE with three correlation matrix: exchangeable, AR(1), and a specific fixed correlation matrix.

Monte Carlo Simulation-Probit

1. $x_{i1} = 1$; x_{i2} is a multivariate normal with mean 1 and variance 1 and covariance $\frac{\rho}{d_{ij}}$; $x_{i3} = 0.2x_{i2} - 1.2e_1, e_1 \sim \mathrm{N}(0,1)$; $\beta_1 = \beta_2 = \beta_3 = 1$.
2. $y_i^* = x_i\beta + e_i$, e_i is multivariate normal with $\mathrm{E}(e_i|x_i) = 0$, $\mathrm{Var}(e_i) = 1$ and $\mathrm{Cov}(e_i, e_j) = \frac{\rho}{d_{ij}}$; $\quad \rho = 0, 0.2, 0.4, 0.6$.
3. $y_i = 1$ if $y_i^* \geq 2.5$, $y_i = 0$ if $y_i^* < 2.5$.
4. Use $\mathrm{Corr}(y_i, y_j|\mathbf{x}_i, \mathbf{x}_j, \mathbf{D}_N) = \frac{\lambda}{d_{ij}}$ as the working correlation between the dependent variables.
5. Use the minimum distance estimator to estimate spatial parameter.

Table: Means and Standard Deviations for Probit, averaged over 1000 samples

		N=400, G=100, L=4		N=1600, G=400, L=4	
		Probit	GEE-probit	Probit	GEE-probit
$\rho = 0$	$\hat{\beta}_2$	1.076	1.033	1.016	1.007
	s.d.$(\hat{\beta}_2)$	0.230	**0.142**	0.103	**0.069**
	$\hat{\beta}_3$	1.070	1.031	1.016	1.018
	s.d.$(\hat{\beta}_3)$	0.205	**0.127**	0.084	**0.059**
	$\hat{\beta}_4$	1.069	1.021	1.019	1.011
	s.d.$(\hat{\beta}_4)$	0.304	**0.200**	0.136	**0.103**
$\rho = 0.5$	$\hat{\beta}_2$	1.310	1.252	1.229	1.213
	s.d.$(\hat{\beta}_2)$	0.293	**0.169**	0.124	**0.077**
	$\hat{\beta}_3$	1.310	1.256	1.229	1.214
	s.d.$(\hat{\beta}_3)$	0.259	**0.156**	0.111	**0.720**
	$\hat{\beta}_4$	1.297	1.243	1.227	1.213
	s.d.$(\hat{\beta}_4)$	0.364	**0.238**	0.165	**0.112**
$\rho = 1$	$\hat{\beta}_2$	1.254	1.203	1.180	1.167
	s.d.$(\hat{\beta}_2)$	0.280	**0.173**	0.121	**0.081**
	$\hat{\beta}_3$	1.236	1.192	1.176	1.164
	s.d.$(\hat{\beta}_3)$	0.236	**0.152**	0.105	**0.072**
	$\hat{\beta}_4$	1.238	1.196	1.175	1.165
	s.d.$(\hat{\beta}_4)$	0.356	**0.241**	0.156	**0.109**
$\rho = 1.5$	$\hat{\beta}_2$	1.022	0.982	0.963	0.949
	s.d.$(\hat{\beta}_2)$	0.230	**0.149**	0.102	**0.070**
	$\hat{\beta}_3$	1.013	0.979	0.966	0.953
	s.d.$(\hat{\beta}_3)$	0.196	**0.132**	0.086	**0.063**
	$\hat{\beta}_4$	1.003	0.963	0.968	0.953
	s.d.$(\hat{\beta}_4)$	0.309	**0.209**	0.139	**0.101**

Note: The estimates with smaller standard deviations are marked with bold.

Count Data

- A count variable is a variable that takes on nonnegative integer values. 0, 1, 2, ... For example, the number of patents applied for by a firm during a year, and the number of children under 18 in a household.
- Silva & Tenreyro (2006) model the gravity equation in the form of (24) and use PPML(Poisson QMLE).
- Count data can be characterized by a Poisson density in linear exponential family (LEF) and an exponential mean.

$$f\left(y|\mathbf{x}\right) = \exp\left[-\exp\left(\mathbf{x}\boldsymbol{\beta}\right)\right]\left[\exp\left(\mathbf{x}\boldsymbol{\beta}\right)\right]^{y}/y!, \tag{23}$$

where $y! = 1 \cdot 2 \cdot ... \cdot (y-1) \cdot y$ and $0! = 1$.
- Assume that the conditional mean function $\mathrm{E}(y_i|\mathbf{x}_i) = \exp\left(\mathbf{x}_i\boldsymbol{\beta}\right)$ is correctly specified and model a multiplicative error in the conditional mean,

$$\mathrm{E}\left(y_i|\mathbf{x}_i, v_i\right) = v_i \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right), \tag{24}$$

where v_i is the multiplicative spatial error term that has unobserved spatial correlation.

Count Data

A count data model with a multiplicative spatial error is characterized by the following assumptions:

(1) $y_i|\mathbf{x}_i, v_i \sim Poisson\left[v_i \exp\left(\mathbf{x}_i \boldsymbol{\beta}\right)\right]$

(2) y_i, y_j are independent conditional on $\mathbf{x}_i, \mathbf{x}_j, v_i, v_j, \qquad i \neq j$

(3) v_i is independent of $\mathbf{x}_i$, $\mathrm{E}(v_i) = 1$, $\mathrm{Var}\left(v_i\right) = \tau^2$, and $\mathrm{Cov}(v_i, v_j) = \tau^2 \cdot c\left(d_{ij}\right)$, where $c\left(d_{ij}\right)$ is the spatial correlation depending on the distance between observation i and j.

(4) By iterated expectation the conditional mean is assumed to be

$$\mathrm{E}\left(y_i|\mathbf{x}_i, \mathbf{D}_N\right) = \exp\left(\mathbf{x}_i \boldsymbol{\beta}_0\right). \tag{25}$$

The Poisson QMLE gives a consistent estimator for the mean parameters, which solves:

$$\hat{\boldsymbol{\beta}}_{PQMLE} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{N} l_i\left(\beta\right) = \sum_{i=1}^{N} y_i \mathbf{x}_i \boldsymbol{\beta} - \sum_{i=1}^{N} \exp\left(\mathbf{x}_i \boldsymbol{\beta}\right) - \sum_{i=1}^{N} \log\left(y_i!\right). \tag{26}$$

Count Data

- The GEE approach is:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \sum_{g=1}^{G} \left[\mathbf{y}_g - \exp\left(\mathbf{X}_g \boldsymbol{\beta}\right)\right]' \hat{\mathbf{W}}_g^{-1} \left[\mathbf{y}_g - \exp\left(\mathbf{X}_g \boldsymbol{\beta}\right)\right]. \tag{27}$$

- Working correlation matrix can be independent, exchangeable, or fixed.
- We can calculate the conditional variance and covariance of y on $\mathbf{x}$ and $\mathbf{D}$.

$$\begin{aligned} \mathrm{Var}\left(y_i | \mathbf{x}_i, \mathbf{D}_N\right) &= \exp\left(\mathbf{x}_i \boldsymbol{\beta}_0\right) + \exp\left(2\mathbf{x}_i \boldsymbol{\beta}_0\right) \cdot \tau^2 & (28) \\ \mathrm{Cov}\left(y_i, y_j | \mathbf{x}_i, \mathbf{x}_j, \mathbf{D}_N\right) &= \exp\left(\mathbf{x}_i \boldsymbol{\beta}_0\right) \exp\left(\mathbf{x}_j \boldsymbol{\beta}_0\right) \cdot \tau^2 \cdot c\left(d_{ij}\right) & (29) \end{aligned}$$

- Using the information above, τ^2 can be estimated by $\hat{\tau}^2$ as the coefficient by regressing $\check{u}_i^2 - \exp\left(\mathbf{x}_i \check{\beta}\right)$ on $\exp\left(2\mathbf{x}_i \check{\beta}\right)$. Obviously $\hat{\tau}^2$ does not depend on distances. If we assume $c\left(d_{ij}\right) = \frac{\lambda}{d_{ij}}$, λ can be estimated by regressing $\frac{\check{u}_i \check{u}_j}{\exp\left(\mathbf{x}_i \check{\beta}\right) \exp\left(\mathbf{x}_j \check{\beta}\right)}$ on $\hat{\tau}^2 / d_{ij}$.

NegBin II model

$$\mathrm{Var}\left(y_i|\mathbf{x}_i, \mathbf{D}_N\right) = \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right) + \exp\left(2\mathbf{x}_i\boldsymbol{\beta}_0\right) \cdot \tau^2. \tag{30}$$

- The traditional Poisson variance assumption is $\mathrm{Var}\left(y_i|\mathbf{x}_i\right) = \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right)$.
- The Poisson GLM variance assumption is $\mathrm{Var}\left(y_i|\mathbf{x}_i\right) = \sigma^2 \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right)$ with an overdispersion or underdispersion parameter σ^2, which is a constant.
- Obviously, there is over-dispersion since $\exp\left(2\mathbf{x}_i\boldsymbol{\beta}_0\right) \cdot \tau^2 \geq 0$, and the over-dispersion parameter is $1 + \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right) \cdot \tau^2$, which is changing with $\mathbf{x}_i$.
- So both the traditional Poisson variance assumption and the GLM variance assumption fail.
- We consider **NegBin II model** of Cameron and Trivedi (1986) as a alternative model.
- With an exponential mean and $y_i|\mathbf{x}_i, v_i \sim$Poisson$[v_i \exp\left(\mathbf{x}_i\boldsymbol{\beta}_0\right)]$, $y_i|\mathbf{x}_i$ is shown to follow a negative binomial II distribution.

Now the log-likelihood function for observation i is

$$l_i = \left(\tau^2\right)^{-2} \log\left[\frac{\left(\tau^2\right)^{-2}}{\left(\tau^2\right)^{-2} + \exp\left(\mathbf{x}_i\boldsymbol{\beta}\right)}\right] + y_i \log\left[\frac{\exp\left(\mathbf{x}_i\boldsymbol{\beta}\right)}{\left(\tau^2\right)^{-2} + \exp\left(\mathbf{x}_i\boldsymbol{\beta}\right)}\right] \tag{31}$$
$$+ \log\left[\Gamma\left(y_i + \left(\tau^2\right)^{-2}\right)/\Gamma\left(\left(\tau^2\right)^{-2}\right)\right],$$

where $\Gamma(\cdot)$ is the gamma function defined for $r > 0$ by $\Gamma(r) = \int_0^\infty z^{r-1} \exp(-z)\, dz$. For fixed τ^2, the log likelihood equation is in the exponential family; see GMT (1984a). Thus the negative binomial QMLE is consistent under conditional mean assupmtion only, which is the same as the Poisson QMLE.

Monte Carlo Simulation-Count

1. v_i is simulated as a multivariate lognormal variable, exponentiating an underlying multivariate normal distribution $\mathrm{N}\left(-\frac{1}{2}, 1\right)$ using with correlation matrix W. $W_{ij} = \frac{\rho}{d_{ij}}$, $\rho = 0, 0.2, 0.4, 0.6, i \neq j; W_{ij} = 1, i = j$. The underlying normal distribution implies that v_i follows a multivariate lognormal distribution with $\mathrm{E}(v_i) = 1$, $\tau^2 \equiv \mathrm{Var}\,(v_i) = \mathrm{e} - 1 \approx 1.718$.(e is the mathematical constant)
2. $\alpha = -1, \beta_2 = 1, \beta_3 = 1, \beta_4 = 1$.
3. x_2 follows a multivraite normal distribution $\mathrm{N}(\mathbf{0}, W)$; $x_3 \sim \mathrm{Uniform}\,(0, 1)$; $x_5 \sim \mathrm{N}(0, 1)$; $x_4 = 1\,[x_5 > 0]$.
4. $m_i = v_i \exp\left(\alpha + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4\right)$;
5. $y_i \sim \mathrm{Poisson}\,(m_i)$.

Table: Means and Standard Deviations for Count, averaged over 1000 samples.

		N=400,G=100,L=4				N=1600,G=400,L=4			
		Poisson	GEE-poisson	NB II	GEE-nb2	Poisson	GEE-poisson	NB II	GEE-nb2
$\rho = 0$	$\hat{\beta}_2$	1.000	0.999	1.002	1.002	0.994	0.994	0.997	0.997
	s.d.$(\hat{\beta}_2)$	**0.259**	0.260	**0.227**	0.228	**0.160**	**0.160**	**0.136**	**0.136**
	$\hat{\beta}_3$	1.000	0.999	1.002	1.002	0.999	1.000	0.998	0.998
	s.d.$(\hat{\beta}_3)$	**0.259**	0.260	**0.227**	0.228	**0.137**	**0.137**	**0.121**	**0.121**
	$\hat{\beta}_4$	0.998	0.998	0.996	0.996	1.003	1.003	1.003	1.003
	s.d.$(\hat{\beta}_4)$	**0.146**	0.147	**0.137**	**0.137**	**0.071**	**0.071**	**0.067**	**0.067**
$\rho = 0.5$	$\hat{\beta}_2$	0.985	0.985	0.993	0.994	1.000	1.000	1.000	0.999
	s.d.$(\hat{\beta}_2)$	0.256	**0.255**	0.216	**0.215**	**0.127**	**0.127**	0.110	**0.109**
	$\hat{\beta}_3$	1.006	1.006	1.004	1.005	1.005	1.004	1.003	1.003
	s.d.$(\hat{\beta}_3)$	0.211	**0.210**	0.180	**0.179**	**0.106**	**0.106**	**0.092**	**0.092**
	$\hat{\beta}_4$	1.002	1.002	1.003	1.003	1.003	1.003	1.002	1.002
	s.d.$(\hat{\beta}_4)$	**0.117**	**0.117**	0.111	**0.110**	**0.058**	**0.058**	**0.054**	**0.054**
$\rho = 1$	$\hat{\beta}_2$	0.987	0.988	0.991	0.991	0.998	0.997	0.997	0.997
	s.d.$(\hat{\beta}_2)$	0.267	**0.259**	0.234	**0.226**	0.130	**0.127**	0.130	**0.128**
	$\hat{\beta}_3$	1.003	1.003	1.004	1.004	1.000	0.999	1.000	0.999
	s.d.$(\hat{\beta}_3)$	0.220	**0.214**	0.195	**0.190**	0.105	**0.102**	0.094	**0.091**
	$\hat{\beta}_4$	0.995	0.996	0.996	0.997	1.000	1.000	1.000	1.000
	s.d.$(\hat{\beta}_4)$	0.120	**0.119**	0.113	**0.111**	0.060	**0.058**	0.056	**0.054**
$\rho = 1.5$	$\hat{\beta}_2$	0.980	0.982	0.995	0.998	0.988	0.988	0.995	0.997
	s.d.$(\hat{\beta}_2)$	0.320	**0.302**	0.276	**0.261**	0.183	**0.173**	0.154	**0.145**
	$\hat{\beta}_3$	0.997	0.995	0.992	0.992	0.992	0.994	0.997	0.999
	s.d.$(\hat{\beta}_3)$	0.288	**0.271**	0.250	**0.234**	0.143	**0.136**	0.126	**0.120**
	$\hat{\beta}_4$	0.997	0.999	1.000	0.998	1.002	1.001	1.003	1.003
	s.d.$(\hat{\beta}_4)$	0.146	**0.139**	0.139	**0.131**	0.077	**0.072**	0.073	**0.068**

Note: The estimates with smaller standard deviations are marked with bold.

Estimating the FDI equation

We collect the inflow FDI data for 290 cities in 31 provinces in China in 2007

The pairwise distances of cities are calculated use the latitudes and longitudes of the city.

We use the provinces as natural division of groups.

The estimating equation is based on

$$\begin{aligned} \mathrm{E}\left(FDI_i|x\right) &= \exp[\beta_0 + \beta_1 \log(GDP) + \beta_2 \log\left(POP_i\right) + \beta_3 \log(WAGE_i) \\ &\quad + \beta_4 \log\left(EDUCEXP_i\right) + \beta_5 \log\left(SCIEXP_i\right) + \beta_6 SEASIDE_i \\ &\quad + \beta_7 BORDER_i]. \end{aligned} \tag{32}$$

The log-linearized model is

$$\begin{aligned} \log\left(FDI_i\right) &= \beta_0 + \beta_1 \log(GDP_i) + \beta_2 \log\left(POP_i\right) + \beta_3 \log(WAGE_i) \\ &\quad + \beta_4 \log\left(EDUCEXP_i\right) + \beta_5 \log\left(SCIEXP_i\right) + \beta_6 SEASIDE_i \\ &\quad + \beta_7 BORDER_i + u. \end{aligned} \tag{33}$$

Table: Descriptive statistics

Variables	Sample size	Average value	Standard deviation	Min	Max
FDI	284	43571.94	99369.96	0	791954
logFDI	275	9.28	1.81	3.14	13.58
logGDP	287	15.58	0.92	13.34	18.60
logPOP	287	5.83	0.48	2.90	8.08
logWAGE	287	9.93	0.25	9.16	10.81
logEDUCEXP	287	11.84	0.72	9.32	14.85
logSCIEXP	287	8.86	1.26	6.15	13.91
BORDER	290	0.07	0.25	0	1
SEASIDE	290	0.18	0.39	0	1

‘

Table: Estimating the FDI equation

	OLS	Poisson	GEE _poisson	NB	GEE _nb2
lnGDP	1.099***	0.705***	0.746***	1.071***	0.982***
	(0.188)	(0.151)	(0.132)	(0.205)	(0.176)
lnGDPPC	0.570***	0.747***	0.687***	0.610***	0.533***
	(0.219)	(0.134)	(0.122)	(0.157)	(0.172)
lnWAGE	-0.123	-0.726*	-1.013***	-0.146	-0.111
	(0.393)	(0.384)	(0.390)	(0.400)	(0.331)
lnSCIEXP	0.186	0.289***	0.311***	0.094	0.137
	(0.142)	(0.110)	(0.102)	(0.111)	(0.106)
BORDER	-0.192	-0.593***	-0.197*	-0.556**	-0.037
	(0.187)	(0.166)	(0.128)	(0.185)	(0.273)
_cons	-13.894***	-3.884	-1.238	-12.360***	-11.021***
	(3.670)	(3.094)	(3.011)	(3.130)	(2.863)
Observations	275	284	284	284	284
F(5, 269)	152.03				
Wald Chi2(5)		701.24	269.58	602.66	495.67
p value	0.000	0.000	0.000	0.000	0.000

Note: Robust standard errors are in parentheses.
***, ** and * indicate significance at the 1%, 5%, and 10% level separately.

Conclusions and future work

- QMLE is a good starting point to estimate nonlinear models.
- Instead of estimating the complicated spatial correlation , we use working correlation matrix as an approximation.
- We found that Probit GEE do not have efficiency gain compared to Bernulli QMLE.
- For a multiplicative count data model, the negative binomial II distribution has better performance than Poisson.
- This method can be used to estimate other models: the gravity equation; the patent equation.
- The observations in each group is not the same in most applications. We can refer to the treatment of unbanlanced panel data. See, for e.g. Wooldridge (2010).

Conditions

A.1) The lattice $D \subseteq \mathbb{R}^d, d \geq 1$, is infinitely countable. The distance $\rho(i,j)$ between any two different individual units i and j in D is at least larger than a positive constant, i.e., $\forall i,j \in D : \rho(i,j) \geq \rho_0$, w.l.o.g. we assume $\rho_0 > 1$

A.2) $\{y_i\}$ is L_4- uniformly NED on the $\alpha-$ mixing random field $\varepsilon = \{\varepsilon_i, i \in D_n\}$, where $\varepsilon_i = (x_i, \epsilon_i)$($\epsilon_i$s are some underlying innovation processes). With the $\alpha-$ mixing coefficient $\overline{\alpha}(u,v,r) \leq (u+v)^\tau \hat{\alpha}(r)$, and $\hat{\alpha}(r) \to 0$ as $r \to \infty$. Assume that $\sum_{r=1}^{\infty} r^{d-1}\hat{\alpha}(r) < \infty$. The NED constant is $d_{n,i}$, ($\sup_{n,i\in T_n} d_{n,i} < \infty$) and the NED coefficient is $\psi(s)$ with $\psi(s) \to 0$, where recall that L is the group size, and $\sum_{r=0}^{\infty} r^{d-1}\psi(r) \to 0$.

Asymptotic Theory

A.3) The parameter space $\boldsymbol{\Theta} \times \boldsymbol{\Gamma}$ is a compact subset on $\mathcal{R}^{p+q}$ with metric $\nu(.,.)$.

A.4) $q_g(\boldsymbol{\theta}, \gamma)$, $(s_g(\boldsymbol{\theta}, \gamma))$, $(h_g(\boldsymbol{\theta}, \gamma))$ are $\mathbf{R}^{p_w} \times \Theta \times \Gamma \to \mathbf{R}^1(\mathbf{R}^p), (\mathbf{R}^{p^2})$ measurable for each $\theta \in \Theta, \gamma \in \Gamma$, and Lipschitz continuous on $\boldsymbol{\Theta} \times \Gamma$.

A.5) $\mathrm{E}\sup_{\theta\in\Theta} |m_{g,i}|^r \le C_1$, $\mathrm{E}\,|w_{g,i,j}|^r \le C_2$, $\mathrm{E}\,|y_{g,i}|^r \le C_3$
$\mathrm{E}\sup_{\theta\in\Theta} |\nabla_\theta m_{g,i}|^r \le C_4$, where C_1, C_2, C_3, C_4 are constants, where $w_{g,i,j}, y_{g,i}, m_{g,i}$ is the elementwise component for $\mathbf{W}_g^{-1}(\theta, \gamma)$, $\mathbf{y}_g$, $\mathbf{m}_g(\theta, \gamma)$. $r > 4p'' \vee 4p'$. $m_{g,i}, w_{g,i,j}$ are continuously differentiable up to the third order derivatives, and its rth moment (the supreme over the parameter space) is bounded up to the second order derivatives. Define $d_g = \max_{i\in B_g} d_{n,i}$,
$M_G \stackrel{\text{def}}{=} \max_g d_g \vee c_{g,q} \vee c_{g,s} \vee c_{g,h}$. Also assume that $\sup_G \sup_g (c_{g,q} \vee c_{g,s} \vee c_{g,h})/d_g \le C_5$, where C_5 is a constant.

Asymptotic Theory

A.6) The $\alpha-$ mixing coefficients of the input field $\tilde{\varepsilon}$ satisfy $\tilde{\alpha}(u,v,r) \le \phi(uL,vL)\hat{\alpha}(r)$, with $\phi(uL,vL) = (u+v)^{\tau}L^{\tau}$ and for some $\hat{\alpha}(r)$, $\sum_{r=1}^{\infty} L^{\tau} r^{d-1}\hat{\alpha}(r) < \infty$.

A.7) We assume moment conditions on the objects involved to prove the NED property of $\mathbf{H}_G(\theta,\gamma)$.
$b_{ij} \stackrel{\text{def}}{=} e_i^{\top}(\mathbf{1}^{\top}\mathbf{W}_g^{-1}(\theta,\gamma) \otimes I_g)|\partial\mathrm{Vec}(\nabla\mathbf{m}_g(\theta))/\partial\theta|_a e_j$.
$c_{ij} = e_i^{\top}(\mathbf{1}^{\top} \otimes \nabla\mathbf{m}_g^{\top}(\theta))|\partial\mathrm{Vec}(\nabla\mathbf{m}_g(\theta))/\partial\theta|_a e_j$. $\|b_{ij}\|$ and $\|c_{ij}\|$ are finite.

A.8) (Identifiability)Let $\overline{Q}_G(\boldsymbol{\theta},\gamma) \stackrel{\text{def}}{=} \frac{1}{|M_G||D_G|}\sum_g \mathrm{E}(q_g(\boldsymbol{\theta},\gamma))$. Recall that $Q_{\infty}(\theta,\gamma) \stackrel{\text{def}}{=} \lim_{G\to\infty}\bar{Q}_G(\boldsymbol{\theta},\gamma)$. Assume that θ^0,γ^0 are identified unique in a sense that
$\liminf_{G\to\infty} \mathbf{inf}_{\theta\in\Theta:\nu(\theta,\theta^0)\ge\varepsilon} Q_G(\boldsymbol{\theta},\gamma) > c_0 > 0$, for any γ and a positive constant c_0.

$$AS_G = \quad (34)$$

$$\frac{1}{G}\sum_g \mathrm{E}\left[\nabla \mathbf{m}_g^\top\left(\boldsymbol{\theta}^0\right)\mathbf{W}_g^{-1}\left(\boldsymbol{\theta}^0,\gamma^0\right)\mathbf{u}_g\mathbf{u}_g^\top\mathbf{W}_g^{-1}\left(\boldsymbol{\theta}^0,\gamma^0\right)\nabla \mathbf{m}_g\left(\boldsymbol{\theta}^0\right)\right] \quad (35)$$

$$+\frac{1}{G}\sum_g\sum_{h,h\neq g} \mathrm{E}\left[\nabla \mathbf{m}_g^\top\left(\boldsymbol{\theta}^0\right)\mathbf{W}_g^{-1}\left(\boldsymbol{\theta}^0,\gamma^0\right)\mathbf{u}_g\mathbf{u}_h^\top\mathbf{W}_h^{-1}\left(\boldsymbol{\theta}^0,\gamma^0\right)\nabla \mathbf{m}_h\left(\boldsymbol{\theta}^0\right)\right]$$

and $AS_\infty = \lim_{G\to\infty} AS_G$.

A.9) The true point θ^0, γ^0 lies in the interior point of Θ, Γ. $\hat{\gamma}$ is estimated with $|\hat{\gamma}-\gamma^0|_2 = \mathcal{O}_p(G^{-1/2})$.

A.10) $c' < \lambda_{min}(M_G^{-2}\,\mathrm{E}\left(\nabla \mathbf{m}_g^\top(\theta^0)\mathbf{W}_g^{-1}(\theta^0,\gamma^0)\nabla \mathbf{m}_g(\theta^0)\right))$

Define $\mathbf{u}_g = \mathbf{y}_g - \mathbf{m}_g(\theta^0)$ and $\hat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{m}_g(\hat{\theta})$

$$\mathbf{S}_G\left(\boldsymbol{\theta},\hat{\gamma}\right) = \frac{1}{M_G|D_G|}\sum_g \nabla \mathbf{m}_g^\top\left(\boldsymbol{\theta}\right)\mathbf{W}_g^{-1}\left(\boldsymbol{\theta},\hat{\gamma}\right)\left[\mathbf{y}_g - \mathbf{m}_g\left(\boldsymbol{\theta}\right)\right]. \quad (36)$$

A.11)

$$\mathbf{S}_G\left(\hat{\gamma},\hat{\boldsymbol{\theta}}\right)=\mathcal{O}_p(1)$$

. $\inf_G |D_G|^{-1} M_G^{-2}\lambda_{min}(\mathbf{AS}_\infty)>0$, where $\mathbf{AS}_\infty$ is defined in equation (34). The mixing coefficients satisfy $\sum_{r=1}^{\infty} r^{(d\tau^*+d)-1} L^{\tau^*}\hat{\alpha}^{\delta/(2+\delta)}(r)<\infty$. $(\tau^*=\delta\tau/(4+2\delta))$.