Lecture 10: Dynamic Spatial Panel Data Models

Zhenlin Yang

School of Economics, Singapore Management University
zlyang@smu.edu.sg
http://www.mysmu.edu.sg/faculty/zlyang/

ECON747 Spatial Econometric Models and Methods
Term I 2024-25

The materials presented in this chapter are drawn from Yang (2018a, JOE) and its `Supplement` Yang (2018b). To simplify the notation, we suppress the subscript n in a vector or a matrix. Consider the following dynamic spatial panel data (DSPD) model of the form:

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t'\beta + Z\gamma + \mu + \alpha_t 1_n + u_t, \\
u_t &= \lambda_3 W_3 u_t + v_t, \quad t = 1, \ldots, T,
\end{aligned} \tag{10.1}
$$

where for $r = 1, 2, 3$, W_r are the given $n \times n$ spatial weight matrices, λ_r are, respectively, the spatial lag (`SL`), space-time lag (`STL`), and spatial error (`SE`) parameter,

- y_t: $n \times 1$ vector of response
- X_t: $n \times p$ matrix of time-varying regressors,
- μ: $n \times 1$ vector of individual-specific effects,
- v_t: $n \times 1$ vector of idiosyncratic errors, $iid(0, \sigma_v^2)$.

Fixed-effects DSPD models

We focus on the fixed-effects (FE) DSPD model, i.e, μ is allowed to be correlated with X_t in an arbitrary manner, leading to an

- FE-DSPD model with `SL`, `STL`, and `SE` dependence, or simply `STLE`;
- or FE-DSPD model with `SL` and `SE`, or `SLE`, by setting $\lambda_2 = 0$;
- or FE-DSPD model with `SL` and `STL` by setting $\lambda_3 = 0$;
- or FE-DSPD model with only `SL` by setting $\lambda_2 = \lambda_3 = 0$;
- or FE-DSPD model with only `SE` by setting $\lambda_1 = \lambda_3 = 0$.

We consider the large-n and small-T setting (**short panels**), and assume:

1. data collection starts from time point $t = 0$,
2. **initial observations**, $\{y_0, X_0, Z\}$, are available,
3. process started m periods before $t = 0$ from positions y_{-m}, treated as **exogenous**, with m being finite (unknown) or infinite.

Note:

1. When $m = 0$, y_0 is exogenous; when $m \geq 1$, y_0 is endogenous, and when $m = \infty$, processes have become stationary.
2. Model (1) can be extended to allow for higher-order spatial lags;
3. It can also be reduced to some specific models:
   - Model with λ_1 and λ_2 only: Yu et al. (2008), large n and large T;
   - Model with λ_3 only: Su and Yang (2015), large n and small T;
   - Model with λ_1 only: Elhorst (2010), large n and small T.
4. The methods introduced in this lecture can be applied to all of the above models, and are not restricted to small T.

The methods provide a unified framework for estimating short DSPD models, free from the initial conditions and robust against nonnormality.

1. **Incidental parameters problem:** number of parameters increases with sample size;
2. **initial values problem:** the distribution of the vector of initial observations depends on the past unobservables.

Pros and cons with QML and GMM:

- QML estimation is more efficient than GMM estimation;
- QML estimation needs the distribution of y_0, or Δy_1 for setting up the unconditional (quasi) likelihood;
- The distribution of y_0 or Δy_1 involves unobservables (e.g., $X_{-1}, X_{-2}, \ldots, y_{-m}$), $\Rightarrow$ a proper 'model' for y_0 or Δy_1 is required;
- Such a model may depend on unknown process start time '$-m$', and may need stronger conditions on X_t;
- The traditional modeling strategy for y_0 or Δy_1 may not work for models with spatial lags.

For dynamic panel data (DPD) models:

- Anderson, T. W., Hsiao, C. (1981, `JASA`).
- Anderson, T. W., Hsiao, C. (1982, `JOE`).
- Bhargava, A., Sargan, J. D. (1983, `Econometrica`).
- Hsiao, C., et al. (2002, `JOE`), and references therein.
- Binder, M., Hsiao, C., Pesaran, M. H. (2005, `ET`)
- Hayakawa and Pesaran (2015, `JOE`);
- Hayakawa, et al. (2020, `WP`).
- And many more . . .

For dynamic spatial panel data (DSPD) models:

- Yu, J., de Jone, R. and Lee, L. F. (2008, JOE).
- Elhorst, J. P. (2010, RSUE; 2012, JGS).
- Lee, L.-F., Yu, J. (2010, RSUE).
- Su, L., Yang, Z. L. (2015, JOE).
- Lee, L.-F., Yu, J. (2015, JAE).
- Yang, Z. L. (2018a, JOE; 2018b, Supplement).
- Kuersteiner, G. M., Prucha, I. R. (2020a, Econometrica; 2018b, Supplement).
- Baltagi, B. H., Pirotte, A. and Yang, Z. L. (2021, JOE).
- Li, L. Y. and Yang, Z. L. (2020, RSUE).
- Yang, Z. L. (2021, Empirical Economics).

This lecture,

1. introduces a unified method for estimating the FE-DSPD models with short panels, the *M-estimation*, which is free from the specifications of the initial conditions and robust against nonnormality of errors;
2. presents results for consistency and asymptotic normality of the M-estimators;
3. introduces robust method of estimating the VC matrix of the M-estimators, free initial-conditions and allowing nonnormal errors;
4. presents results for consistency of the VC-matrix estimator;
5. presents Monte Carlo results for the finite sample performance of the methods introduced;
6. presents an empirical application to illustrate the proposed method.

Recall: M-Estimator or Zero-Estimator. The term **M-estimation** was coined by Huber (1964) to mean the maximum likelihood type estimation. It can be defined as either

1. the solution of a maximization problem:

$$\hat{\psi}_n = \arg\max\{Q_n(\psi)\};$$

2. or the root of a set of estimating equations:

$$\hat{\psi}_n = \arg\{S_n(\psi) = 0\}.$$

The latter is also called **zero estimator** as it makes the estimating equation zero. See van der Vaart (1998, *Asymptotic Statistics*).

See also Huber (1981, *Robust Statistics*).

Taking first-difference of Model (10.1) to eliminate the fixed effects μ:

$$\begin{aligned}\Delta y_t = \; & \rho \Delta y_{t-1} + \lambda_1 W_1 \Delta y_t + \lambda_2 W_2 \Delta y_{t-1} + \Delta X_t \beta + \Delta u_t, \\ & \Delta u_t = \lambda_3 W_3 \Delta u_t + \Delta v_t, \quad t = 2, 3, \cdots, T.\end{aligned} \tag{10.2}$$

Note: time-invariant variables Z are also eliminated, and the terms corresponding to α_t are merged into X_t, as T is fixed.

Stacking these vectors, the model is written in matrix form:

$$\begin{aligned}\Delta Y \; & = \rho \Delta Y_{-1} + \lambda_1 \mathbf{W}_1 \Delta Y + \lambda_2 \mathbf{W}_2 \Delta Y_{-1} + \Delta X \beta + \Delta u, \\ & \Delta u = \lambda_3 \mathbf{W}_3 \Delta u + \Delta v,\end{aligned} \tag{10.3}$$

- $\mathbf{W}_r = I_{T-1} \otimes W_r, r = 1, 2, 3$; $\otimes$: Kronecker product; I_k: identity matrix.

Denote: $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$, $\quad \lambda = (\lambda_1, \lambda_2, \lambda_3)'$, $\quad \theta = (\beta', \rho, \lambda_1, \lambda_2)'$.

$B_1(\lambda_1) = I_n - \lambda_1 W_1$, $B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$, $B_3(\lambda_3) = I_n - \lambda_3 W_3$.

$\mathbf{B}_r(\lambda_r) = I_{T-1} \otimes B_r(\lambda_r)$, $r = 1, 3$, $\quad \mathbf{B}_2(\rho, \lambda_2) = I_{T-1} \otimes B_2(\rho, \lambda_2)$.

The **quasi Gaussian loglikelihood** of ψ, as if Δy_1 is exogenous is:

$$\ell(\psi) = -\frac{n(T-1)}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| \\ - \frac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega(\lambda_3)^{-1}\Delta u(\theta), \qquad (10.4)$$

- $\Delta u(\theta) = \mathbf{B}_1(\lambda_1)\Delta Y - \mathbf{B}_2(\rho, \lambda_2)\Delta Y_{-1} - \Delta X\beta$,
- $\Omega(\lambda_3) = \frac{1}{\sigma_v^2}\mathrm{Var}(\Delta u) = C \otimes [B_3'(\lambda_3)B_3(\lambda_3)]^{-1}$, and

$$C = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

- Maximizing $\ell(\psi)$ in (10.4) leads to the conditional QMLE $\tilde{\psi}$ of ψ.
- However, $\tilde{\psi}$ cannot be consistent if T is small and fixed,
- as the information about parameters contained in Δy_1 is ignored.

Let $\theta_1 = (\beta', \rho, \lambda_2)'$. Given λ_1 and λ_3, (10.4) is maximized at

$$\tilde{\theta}_1(\lambda_1, \lambda_3) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega\mathbf{B}_1(\lambda_1)\Delta Y, \quad (10.5)$$

$$\tilde{\sigma}_v^2(\lambda_1, \lambda_3) = \frac{1}{n(T-1)}\Delta\tilde{u}'(\lambda_1, \lambda_3)\Omega^{-1}\Delta\tilde{u}(\lambda_1, \lambda_3), \quad (10.6)$$

- $\tilde{u}(\lambda_1, \lambda_3) = \mathbf{B}_1(\lambda_1)\Delta Y - \Delta\mathbb{X}\tilde{\theta}(\lambda_1, \lambda_3)$; $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1}, \mathbf{W}_2\Delta Y_{-1})$.

Substituting $\tilde{\theta}_1(\lambda_1, \lambda_3)$ and $\tilde{\sigma}_v^2(\lambda_1, \lambda_3)$ back into (10.4) gives the concentrated conditional quasi loglikelihood of (λ_1, λ_3),

$$\ell^c_{\texttt{STLE}}(\lambda_1, \lambda_3) = -\frac{n(T-1)}{2}\log[\tilde{\sigma}_v^2(\lambda_1, \lambda_3)] - \frac{1}{2}\log|\Omega(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)|, \quad (10.7)$$

where the constant term is dropped.

Maximizing $\ell^c_{\texttt{STLE}}(\lambda_1, \lambda_3)$ gives the conditional QML (CQML) estimators $\tilde{\lambda}_1$ and $\tilde{\lambda}_3$ of λ_1 and λ_3.

The CQML estimators of θ_1 and σ_v^2 are thus $\tilde{\theta}_1 \equiv \tilde{\theta}_1(\tilde{\lambda}_1, \tilde{\lambda}_3)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1, \tilde{\lambda}_3)$.

Note: the $\ell_{\texttt{STLE}}(\psi)$ is a quasi Gaussian loglikelihood both in

- the traditional sense that $\{v_{it}\}$ are not exactly Gaussian but Gaussian likelihood is used,
- and the sense that Δy_1 is not exogenous but is treated as exogenous.
- The latter causes inconsistency of the CQMLEs when T is small.

Furthermore,

- we see from the results presented below that even if T increases with n, the CQMLEs may encounter an asymptotic bias;
- we introduce a method that not only gives a consistent estimator of the model parameters when T is small, but also eliminates the asymptotic bias when T is large.

Details on these important points follow.

First, to simplify the notation,

- a parametric quantity (scalar, vector or matrix) evaluated at the general values of the parameters is denoted by dropping its arguments, e.g., $B_1 \equiv B_1(\lambda_1)$, $\mathbf{B}_1 \equiv \mathbf{B}_1(\lambda_1)$, $\Omega \equiv \Omega(\lambda_3)$, and similarly for B_r and $\mathbf{B}_r, r = 2, 3$;
- a parametric quantity evaluated at the true values of the parameters is denoted by dropping its argument and then adding a subscript 0, e.g., $B_{10} \equiv B_1(\lambda_{10})$, $\Omega_0 \equiv \Omega(\lambda_{30})$.
- Let $\mathbf{C} = C \otimes I_n$.
- Denote $\Delta u \equiv \Delta u(\theta_0)$.
- The usual expectation, variance and covariance operators, 'E', 'Var' and 'Cov', correspond to the true parameter values.

The conditional quasi score function $S(\psi) = \frac{\partial}{\partial \psi}\ell(\psi)$ has the form:

$$S(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}, \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3), \end{cases} \tag{10.8}$$

where $A_3 = W_3'B_3 + B_3'W_3$ and $G_3 = W_3B_3^{-1}$.

- Under mild conditions, maximizing the conditional loglikelihood $\ell_{\texttt{STLE}}(\psi)$ is equivalent to solving the estimating equation $S_{\texttt{STLE}}(\psi) = 0$;
- The QML type estimation is special case of M-estimation;
- A necessary condition for the M-estimators to be consistent is that the probability limit of the estimating function at the true parameter value is zero (see, e.g., van der Vaart, 1998).

For the estimation of the FE-DSPD model, this condition becomes,

$$\lim_{n\to\infty} \frac{1}{nT} S_{\text{STLE}}(\psi_0) \xrightarrow{p} 0.$$

However, as shown below this is not the case. Thus,

1. CQMLEs are not consistent unless $T \to \infty$.
2. Further, even if T goes to infinity with n (proportionally), the CQMLEs encounter a bias of order $O(T^{-1})$, giving the **asymptotic bias:**
   - if $\frac{1}{nT}\mathrm{E}[S_{\text{STLE}}(\psi_0)] = O(\frac{1}{T})$, then $\frac{1}{\sqrt{nT}}\mathrm{E}[S_{\text{STLE}}(\psi_0)] = O((\frac{n}{T})^{\frac{1}{2}})$,
   - implying $\mathrm{E}[\sqrt{nT}(\tilde{\psi} - \psi_0)] = O((\frac{n}{T})^{\frac{1}{2}})$, and
   - $\sqrt{nT}(\tilde{\psi} - \psi_0)$ converges to a non-centered normal if $\frac{n}{T} \to c > 0$.
3. If $\frac{n}{T} \to 0$ (large T case), the asymptotic bias vanishes, but this would not be a case of interest to a spatial panel model.
4. To overcome this major problem, we first derive $\mathrm{E}[S_{\text{STLE}}(\psi_0)]$, and then adjust $S_{\text{STLE}}(\psi)$ so that the **adjusted quasi score** (AQS) vector, say $S^*_{\text{STLE}}(\psi)$, is such that $\mathrm{plim}_{n\to\infty} \frac{1}{nT} S^*_{\text{STLE}}(\psi_0) = 0$.

In contrast with Hsiao et al . (2002), Elhorst (2010), and Su and Yang (2015), we only need to have very minimum knowledge about the processes in the past.

Assumption A: *(i) the processes started m periods before the start of data collection, the* 0*th period, and (ii) if* $m \geq 1$*,* Δy_0 *is independent of future errors* $\{v_t, t \geq 1\}$*; if* $m = 0$*,* y_0 *is independent of future errors* $\{v_t, t \geq 1\}$.

- The proposed method requires neither $\{y_s, s = -m, \ldots, -1\}$ to follow the same processes as $\{y_t, t = 0, 1, \ldots, T\}$, nor $\{x_{it}\}$ to be trend-stationary or first-difference stationary;
- has a much weaker requirement on processes starting positions y_m.

To derive the results, the reduced form of (10.2) is important:

$$\Delta y_t = \mathcal{B}_0 \Delta y_{t-1} + B_{10}^{-1} \Delta X_t \beta_0 + B_{10}^{-1} B_{30}^{-1} \Delta v_t, \quad t = 2, \ldots, T, \qquad (10.9)$$

where $\mathcal{B} \equiv \mathcal{B}(\rho, \lambda_1, \lambda_2) = B_1^{-1}(\lambda_1) B_2(\rho, \lambda_2)$.

Lemma. *Suppose Assumption A holds. Assume further that, for* $i = 1, \ldots, n$ *and* $t = 0, 1, \ldots, T$, *(i) the idiosyncratic errors* $\{v_{it}\}$ *are iid across* i *and* t *with mean 0 and variance* σ^2_{v0}, *(ii) the time-varying regressors* X_t *are exogenous, and (iii) both* B_{10}^{-1} *and* B_{30}^{-1} *exist. Then,*

$$\mathrm{E}(\Delta Y_{-1}\Delta v') = -\sigma^2_{v0}\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}, \tag{10.10}$$

$$\mathrm{E}(\Delta Y \Delta v') = -\sigma^2_{v0}\mathbf{D}_{0}\mathbf{B}_{30}^{-1}, \tag{10.11}$$

where $\mathbf{D}_{-1} \equiv \mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)$ *and* $\mathbf{D} \equiv \mathbf{D}(\rho, \lambda_1, \lambda_2)$, *given as,*

$$\mathbf{D}_{-1} = \begin{pmatrix} I_n, & 0, & \ldots & 0, & 0 \\ \mathcal{B} - 2I_n, & I_n, & \ldots & 0, & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}^{T-4}(I_n - \mathcal{B})^2, & \mathcal{B}^{T-5}(I_n - \mathcal{B})^2, & \ldots & \mathcal{B} - 2I_n, & I_n \end{pmatrix} \mathbf{B}_1^{-1},$$

$$\mathbf{D} = \begin{pmatrix} \mathcal{B} - 2I_n, & I_n, & \ldots & 0 \\ (I_n - \mathcal{B})^2, & \mathcal{B} - 2I_n, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{B}^{T-3}(I_n - \mathcal{B})^2, & \mathcal{B}^{T-4}(I_n - \mathcal{B})^2, & \ldots & \mathcal{B} - 2I_n \end{pmatrix} \mathbf{B}_1^{-1}.$$

The results of the Lemma lead immediately to

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\Delta Y_{-1}) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \tag{10.12}$$

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\mathbf{W}_1\Delta Y) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \tag{10.13}$$

$$\mathrm{E}(\Delta u'\Omega_0^{-1}\mathbf{W}_2\Delta Y_{-1}) = -\sigma_{v0}^2 \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}\mathbf{W}_2), \tag{10.14}$$

$\Rightarrow$ the $(\rho, \lambda_1, \lambda_2)$ elements of $\mathrm{E}[S(\psi_0)]$ are of order $O(n)$, and hence $\mathrm{plim}_{n\to\infty}\frac{1}{n(T-1)}S(\psi_0) \neq 0$,

$\Rightarrow$ at least, the conditional QMLEs $\tilde{\rho}, \tilde{\lambda}_1$, and $\tilde{\lambda}_2$ are inconsistent.

Under an interesting special case where $\lambda_1 = \lambda_2 = 0$, the FE-DSPD model with SE only considered by Su and Yang (2015), we have

$$\mathrm{plim}\frac{1}{n(T-1)}\frac{\partial}{\partial\rho}\ell_{\mathtt{STLE}}(\psi_0) = \frac{1-\rho_0^T}{T^2(1-\rho_0)^2} - \frac{1}{T(1-\rho_0)},$$

- which is not zero, and thus $\tilde{\rho}$ is not consistent;
- even if $T \to \infty$ with n, Bias$(\tilde{\rho}) = O(\frac{1}{T})$.
- This corresponds to the well-known Nickel bias (Nickel, 1981).

It is expected that this result would hold for the general model, and the bias in $\tilde{\rho}$ would spill over to the other CQMLEs.

The bias terms in (10.12)-(10.14) are functions of parameters and are free from the initial conditions, and thus give a set of *adjusted quasi score* (AQS) functions:

$$S^*(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3), \end{cases} \tag{10.15}$$

which are a set of unbiased and consistent estimating functions, i.e., $\mathrm{E}[S^*(\psi_0)] = 0$, and $\mathrm{plim}_{n\to\infty}\frac{1}{n(T-1)}S^*(\psi_0) = 0$, even when T is fixed.

Solving $S^*(\psi) = 0$ gives the *M*-estimator $\hat{\psi}_{\mathrm{M}}$ of ψ!

This root-finding process can be simplified:

Given $\delta = (\rho, \lambda')'$, the constrained *M*-estimators of β and σ_v^2 are:

$$\hat{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1}), \tag{10.16}$$

$$\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), \tag{10.17}$$

where $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1, \lambda_2)$.

Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the δ-components of $S^*(\psi)$ given in (10.15), we obtain the concentrated AQS functions:

$$S_{\mathrm{C}}^*(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2), \\ \frac{1}{2\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\mathrm{tr}(G_3). \end{cases} \tag{10.18}$$

Solving $S_{\mathrm{C}}^*(\delta) = 0$ gives the *M*-estimators $\hat{\delta}_{\mathrm{M}}$ of δ, and hence the *M*-estimators of β and σ_v^2: $\hat{\beta}_{\mathrm{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathrm{M}})$ and $\hat{\sigma}_{v,\mathrm{M}}^2 \equiv \hat{\sigma}_v^2(\hat{\delta}_{\mathrm{M}})$.

Notation: Recall: ψ_0 = true value of ψ, $B_1 = B_1(\lambda_1)$, $B_{10} = B_0(\lambda_{10})$, $\Omega \equiv \Omega(\lambda_3)$, $\Omega_0 \equiv \Omega(\lambda_{30})$, etc.; $\Delta u \equiv \Delta u(\theta_0)$. Further,

(i) $\boldsymbol{\Delta}$ is the parameter space of $\delta = (\rho, \lambda')'$;

(ii) 'E' and 'Var' correspond to the true parameter values ψ_0;

(iii) $\text{tr}(\cdot)$, $|\cdot|$, $\|\cdot\|$: trace, determinant, Frobenius norm;

(iv) $\gamma_{\max}(A)$, $\gamma_{\min}(A)$: largest and smallest eigenvalues of a real symmetric matrix A.

(v) $\text{diag}(a_k)$ forms a diagonal matrix using the elements $\{a_k\}$, $\text{blkdiag}(A_k)$ forms a block-diagonal matrix using matrices $\{A_k\}$.

Assumption B: *The innovations v_{it} are iid for all i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \sigma_v^2$, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption C: *The space $\boldsymbol{\Delta}$ is compact, and the true parameter δ_0 lies in its interior.*

Assumption D: *The time-varying regressors $\{X_t,\ t = 0, 1, \ldots, T\}$ are exogenous, their values are uniformly bounded, and $\lim_{n\to\infty} \frac{1}{nT} \Delta X' \Delta X$ exists and is nonsingular.*

Assumption E: *(i) For $r = 1, 2, 3$, the elements $w_{r,ij}$ of W_r are at most of order h_n^{-1}, uniformly in all i and j, and $w_{r,ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_r, r = 1, 2, 3\}$ and $\{B_{r0}^{-1}, r = 1, 3\}$ are uniformly bounded in both row and column sum norms; (iv) For $r = 1, 3$, $\{B_r^{-1}\}$ are uniformly bounded in either row or column sum norms, uniformly in λ_r in a compact parameter space $\boldsymbol{\Lambda}_r$, and*

$$0 < \underline{c}_r \leq \inf_{\lambda_r \in \boldsymbol{\Lambda}_r} \gamma_{\min}(B_r' B_r) \leq \sup_{\lambda_r \in \boldsymbol{\Lambda}_r} \gamma_{\max}(B_r' B_r) \leq \bar{c}_r < \infty.$$

Assumption F: *For an $n \times n$ matrix Φ uniformly bounded in either row or column sums, with elements of uniform order h_n^{-1}, and an $n \times 1$ vector ϕ with elements of uniform order $h_n^{-1/2}$,*

(i) $\frac{h_n}{n}\Delta y_1'\Phi\Delta y_1 = O_p(1)$ and $\frac{h_n}{n}\Delta y_1'\Phi\Delta v_2 = O_p(1)$;

(ii) $\frac{h_n}{n}(\Delta y_1 - \mathrm{E}(\Delta y_1))'\phi = o_p(1)$;

(iii) $\frac{h_n}{n}[\Delta y_1'\Phi\Delta y_1 - \mathrm{E}(\Delta y_1'\Phi\Delta y_1)] = o_p(1)$, and

(iv) $\frac{h_n}{n}[\Delta y_1'\Phi\Delta v_2 - \mathrm{E}(\Delta y_1'\Phi\Delta v_2)] = o_p(1)$.

Define $\bar{S}^*(\psi) = \mathrm{E}[S^*_{\mathtt{STLE}}(\psi)]$, the **population counter part** of $S^*(\psi)$ given in (10.15). Given δ, $\bar{S}^*_{\mathtt{STLE}}(\psi) = 0$ is partially solved at

$$\bar{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\mathrm{E}\Delta Y - \mathbf{B}_2\mathrm{E}\Delta Y_{-1}), \tag{10.19}$$

$$\bar{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta\bar{u}(\delta)], \tag{10.20}$$

where $\Delta\bar{u}(\delta) = \Delta u(\theta)|_{\beta=\bar{\beta}(\delta)} = \mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1} - \Delta X\bar{\beta}(\delta)$.

These lead to the population counter part of $S_c^*(\psi)$ given in (10.18), upon substituting $\bar{\beta}(\delta)$ and $\bar{\sigma}_v^2(\delta)$ back into the δ-component of $\bar{S}^*(\psi)$:

$$\bar{S}_c^*(\delta) = \begin{cases} \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta Y_{-1}] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y_{-1}] + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_1), \\ \frac{1}{2\bar{\sigma}_v^2(\delta)}\mathrm{E}[\Delta\bar{u}(\delta)'(C^{-1}\otimes A_3)\Delta\bar{u}(\delta)] - (T-1)\mathrm{tr}(G_3). \end{cases} \tag{10.21}$$

Idea for consistency: By Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathrm{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{\sqrt{n(T-1)}}\left\|S_c^*(\delta) - \bar{S}_c^*(\delta)\right\| \xrightarrow{p} 0$, and the following identification condition holds.

Assumption G: $\inf_{\delta:\ d(\delta,\delta_0)\geq\varepsilon}\left\|\bar{S}_c^*(\delta)\right\| > 0$ *for every* $\varepsilon > 0$*, where* $d(\delta,\delta_0)$ *is a measure of distance between* δ_0 *and* δ.

Theorem

Suppose Assumptions A-G hold. Assume further that

(*i*) $\gamma_{\max}[\mathrm{Var}(\Delta Y)]$ *and* $\gamma_{\max}[\mathrm{Var}(\Delta Y_{-1})]$ *are bounded, and*

(*ii*) $\inf_{\delta\in\Delta}\gamma_{\min}\big(\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\big) \geq \underline{c}_y > 0$.

We have, as $n\to\infty$, $\hat{\psi}_{\mathrm{M}} \xrightarrow{p} \psi_0$.

For **asymptotic normality**, let $\Delta\mathbf{y}_1 = 1_{T-1}\otimes\Delta y_1$. We have,

$$\Delta Y = \mathbb{R}\,\Delta\mathbf{y}_1 + \boldsymbol{\eta} + \mathbb{S}\Delta v, \tag{10.22}$$

$$\Delta Y_{-1} = \mathbb{R}_{-1}\Delta\mathbf{y}_1 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\Delta v, \tag{10.23}$$

where $\mathbb{R} = \mathrm{blkdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^{T-1})$, $\mathbb{R}_{-1} = \mathrm{blkdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-2})$, $\boldsymbol{\eta} = \mathbb{B}\mathbf{B}_{10}^{-1}\Delta X\beta_0$, $\boldsymbol{\eta}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\Delta X\beta_0$, $\mathbb{S} = \mathbb{B}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$, $\mathbb{S}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$,

$$\mathbb{B} = \begin{pmatrix} I_n & 0 & \ldots & 0 & 0 \\ \mathcal{B}_0 & I_n & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \ldots & \mathcal{B}_0 & I_n \end{pmatrix}, \quad \mathbb{B}_{-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ I_n & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-3} & \mathcal{B}_0^{T-4} & \ldots & I_n & 0 \end{pmatrix}.$$

Using (10.22) and (10.23), the AQS function can be represented as:

$$S^*(\psi_0) = \begin{cases} \Delta v'\Pi_1, \\ \Delta v'\Phi_1\Delta v - \frac{n(T-1)}{2\sigma_{v0}^2}, \\ \Delta v'\Psi_1\Delta \mathbf{y}_1 + \Delta v'\Pi_2 + \Delta v'\Phi_2\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \\ \Delta v'\Psi_2\Delta \mathbf{y}_1 + \Delta v'\Pi_3 + \Delta v'\Phi_3\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \\ \Delta v'\Psi_3\Delta \mathbf{y}_1 + \Delta v'\Pi_4 + \Delta v'\Phi_4\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}\mathbf{W}_2), \\ \Delta v'\Phi_5\Delta v - (T-1)\mathrm{tr}(G_{30}), \end{cases} \tag{10.24}$$

$\Pi_1 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\Delta X$, $\Pi_2 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_1\boldsymbol{\eta}$, $\Pi_4 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_2\boldsymbol{\eta}_{-1}$, $\Phi_1 = \frac{1}{2\sigma_{v0}^4}\mathbf{C}^{-1}$, $\Phi_2 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbb{S}_{-1}$, $\Phi_3 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_1\mathbb{S}$, $\Phi_4 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_2\mathbb{S}_{-1}$, $\Phi_5 = \frac{1}{\sigma_{v0}^2}[C^{-1} \otimes (G_{30}' + G_{30})]$, $\Psi_1 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbb{R}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_1\mathbb{R}$, $\Psi_3 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_b\mathbf{W}_2\mathbb{R}_{-1}$, and $\mathbb{C}_b = C^{-1} \otimes B_{30}$.

- By the CLT for bilinear-quadratic forms (Yang 2018, Appendix), one shows the asymptotic normality of $S^*(\psi_0)$,
- and hance the asymptotic normality of the M-estimator $\hat{\psi}_{\mathrm{M}}$.

Theorem

Assume Assumptions A-G hold. We have, as $n \to \infty$,

$$\sqrt{n(T-1)}(\hat{\psi}_{\mathrm{M}} - \psi_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty} \Sigma^{*-1}(\psi_0)\Gamma^*(\psi_0)\Sigma^{*-1}(\psi_0)\big],$$

where $\Sigma^*(\psi_0) = -\frac{1}{n(T-1)}\mathrm{E}[\frac{\partial}{\partial\psi'}S^*(\psi_0)]$ *and* $\Gamma^*(\psi_0) = \frac{1}{n(T-1)}\mathrm{Var}[S^*(\psi_0)]$*, both assumed to exist and* $\Sigma^*(\psi_0)$ *to be positive definite, for sufficiently large n.*

- In practical applications, one needs to estimate $\Sigma^*(\psi_0)$ and $\Gamma^*(\psi_0)$ to get standard errors of the *M*-estimators.
- As $\Sigma^*_{\mathtt{STLE}}(\psi_0)$ is the expected negative modified Hessian, its observed counter part immediately offers a consistent estimate, i.e.,

$$\Sigma^*(\hat{\psi}_{\mathrm{M}}) = -\frac{1}{n(T-1)}\frac{\partial}{\partial\psi'}S^*(\psi)\Big|_{\psi=\hat{\psi}_{\mathrm{M}}}. \tag{10.25}$$

- Note that $\frac{\partial}{\partial\psi'}S^*(\psi_0)$ is not symmetric. Its expression is given in Yang (2018, Appendix), but the asymmetric parts were missing. See Yang (2021) for a discussion on this.

However,

- Estimation of $\Gamma^*(\psi_0)$ runs into problems.
- From (10.24), AQS function $S^*(\psi_0)$ contains three types of elements:

$$\Pi'\Delta v, \quad \Delta v'\Phi\Delta v, \quad \Delta v'\Psi\Delta \mathbf{y}_1.$$

- $\mathrm{Var}(\Delta v'\Psi\Delta \mathbf{y}_1)$, etc, depend on unobservables.

Idea of the Proposed Method: write the above quantities as sums of martingale difference (MD) sequences. Then, *outer-product-of-MDs* (OPMD) gives a consistent estimate of the VC matrix.

- For a square matrix A, decompose $A = A^u + A^l + A^d$, sum of upper-triangular, lower-triangular, and diagonal matrix.
- Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ partitioned according to $t, s = 2, \ldots, T$.

First, for the terms linear in Δv:

$$\begin{aligned}\Pi'\Delta v &= \sum_{t=2}^{T} \Pi_t'\Delta v_t \\ &= \sum_{t=2}^{T}\sum_{i=1}^{n} \Pi_{it}'\Delta v_{it} \\ &= \sum_{i=1}^{n}\sum_{t=2}^{T} \Pi_{it}'\Delta v_{it} \equiv \sum_{i=1}^{n} g_{1i}.\end{aligned}$$

Then, for the terms quadratic in Δv: $\mathrm{E}(\Delta v'\Phi\Delta v) = \sigma_{v0}^2\mathrm{tr}(\mathbf{C}\Phi)$, and

$$\begin{aligned}&\Delta v'\Phi\Delta v - \mathrm{E}(\Delta v'\Phi\Delta v) \\ &= \sum_t\sum_s \Delta v_t'\Phi_{ts}\Delta v_s - \sigma_{v0}^2\mathrm{tr}(\mathbf{C}\Phi) \\ &= \sum_t\sum_s \Delta v_t'(\Phi_{ts}^u + \Phi_{ts}^l + \Phi_{ts}^d)\Delta v_s - \sigma_{v0}^2\mathrm{tr}(\mathbf{C}\Phi) \\ &= \sum_t\sum_s \Delta v_t'(\Phi_{st}^{u\prime} + \Phi_{ts}^l + \Phi_{ts}^d)\Delta v_s - \sigma_{v0}^2\mathrm{tr}(\mathbf{C}\Phi) \\ &= \sum_t \Delta v_t'\Delta\xi_t + \sum_t \Delta v_t'\Delta v_t^* - \sigma_{v0}^2\mathrm{tr}(\mathbf{C}\Phi) \\ &= \sum_{i=1}^{n}\sum_t(\Delta v_{it}\Delta\xi_{it} + \Delta v_{it}\Delta v_{it}^* - \sigma_{v0}^2 d_{it}) \equiv \sum_{i=1}^{n} g_{2i},\end{aligned}$$

whee $\Delta\xi_t = \sum_{s=2}^{T}(\Phi_{st}^{u\prime} + \Phi_{ts}^l)\Delta v_s$; $\Delta v_t^* = \sum_{s=2}^{T} \Phi_{ts}^d\Delta v_s$; $\{d_{it}\} = \mathrm{diag}(\mathbf{C}\Phi)$.

Finally, for bilinear terms: $\Delta v' \Psi \Delta \mathbf{y}_1$, define

- $\Psi_{t+} = \sum_{s=2}^{T} \Psi_{ts}, \quad t = 2, \ldots, T,$
- $\Delta y_1^{\circ} = B_{30} B_{10} \Delta y_1; \qquad \Delta y_{1t}^{*} = \Psi_{t+} \Delta y_1,$
- $\Theta = \Psi_{2+} (B_{30} B_{10})^{-1}; \qquad \{\Theta_{ii}\} = \text{diag}(\Theta),$
- $\{\Delta \zeta_{1i}\} = \Delta v_2' \Theta^{u}; \qquad \{\Delta \zeta_{2i}\} = \Theta^{l} \Delta y_1^{\circ}.$

We obtain:

$$
\begin{aligned}
& \Delta v' \Psi \Delta \mathbf{y}_1 - \mathrm{E}(\Delta v' \Psi \Delta \mathbf{y}_1) \\
= & \sum_{i=1}^{n} (\Delta \zeta_{1i} \Delta y_{1i}^{\circ} + \Delta v_{2i} \Delta \zeta_{2i} + \Theta_{ii} (\Delta v_{2i} \Delta y_{1i}^{\circ} + \sigma_{v0}^2) + \sum_{t=3}^{T} \Delta v_{it} \Delta y_{1it}^{*}) \\
\equiv & \sum_{i=1}^{n} g_{3i}
\end{aligned}
$$

The $\{g_{ri}, \mathcal{F}_{n,i}\}$ are M.D. sequences, for $r = 1, 2, 3$!

where $\{\mathcal{F}_{n,i}\}_{i=1}^{n}$ is an increasing sequence of σ-fields generated by $v_0, \Delta y_0, (v_{j1}, \ldots, v_{jT}, j = 1, \ldots, i), i = 1, \ldots, n$.

Applying the above results to the elements of $S^*(\psi_0)$ in (10.24):

- For each Π_r term in (10.24), define g_{1ri}, $r = 1, 2, 3, 4$,
- For each Φ_r term in (10.24), define g_{2ri}, $r = 1, 2, 3, 4, 5$,
- For each Ψ_r term in (10.24), define g_{3ri}, $r = 1, 2, 3$.

Define

$$g_i = \begin{cases} g_{11i}, \\ g_{21i}, \\ g_{31i} + g_{12i} + g_{22i}, \\ g_{32i} + g_{13i} + g_{23i}, \\ g_{33i} + g_{14i} + g_{24i}, \\ g_{25i}. \end{cases}$$

Then, $S^*(\psi_0) = \sum_{i=1}^n g_i$,

- $\{g_i\}$ form a vector M.D. sequence, and hence
- $\mathrm{Var}[S^*(\psi_0)] = \sum_{i=1}^n \mathrm{E}(g_i g_i')$.

The 'average' of the outer products of the estimated $g_i's$, i.e.,

$$\widehat{\Gamma}^* = \frac{1}{n(T-1)} \sum_{i=1}^n \hat{g}_i \hat{g}_i', \tag{10.26}$$

thus gives a consistent estimator of the variance of $\Gamma^*_{\mathtt{STLE}}(\psi_0)$, where $\hat{g}_i$ is obtained by replacing ψ_0 in g_i by $\hat{\psi}_{\mathrm{M}}$ and Δv in g_i by its observed counterpart $\widehat{\Delta} v$, noting that Δy_1 is observed.

We have the following theorem.

Theorem

Under the assumptions of Theorem (1), we have, as $n \to \infty$,

$$\widehat{\Gamma}^* - \Gamma^*(\psi_0) = \frac{1}{n(T-1)} \sum_{i=1}^n \left[\hat{g}_i \hat{g}_i' - \mathrm{E}(g_i g_i')\right] \xrightarrow{p} 0,$$

and hence, $\Sigma^{*-1}(\hat{\psi}_{\mathrm{M}})\widehat{\Gamma}^*\Sigma^{*-1}(\hat{\psi}_{\mathrm{M}}) - \Sigma^{*-1}(\psi_0)\Gamma^*(\psi_0)\Sigma^{*-1}(\psi_0) \xrightarrow{p} 0.$

10.5. M-Estimation for Submodels

Certain submodels deserve some special attention. We concentrate on the submodels that contain spatial dependence, namely,

- the FE-DSPD model with only `SE` dependence,
- the FE-DSPD model with only `SL` dependence,
- the FE-DSPD model with both `SL` and `STL` dependence, and
- the FE-DSPD model with both `SL` and `SE` dependence.

We are particularly interested in comparing our approach with the standard small T or large T approaches, to demonstrate that

- when T is small our approach provides results that are comparable with the standard full QML approach when the initial model is correctly specified.
- However, our approach provides results that are more robust against misspecification of the initial model than does the full QML approach.
- When T is large, our approach provides results that are less biased compared with the conditional QML approach.

Setting $\lambda_1 = \lambda_2 = 0$, Model (10.2) reduces to an FE-DSPD with only SE dependence of a SAR form,

- which has been rigorously treated in Su and Yang (2015) based on a full QML approach where the initial differences are modeled.
- It would be certainly interesting to see how the proposed approach compares with this full QML approach.

The conditional quasi Gaussian loglikelihood (10.4) simplifies to:

$$\ell_{\mathrm{SE}}(\psi) = -\frac{n(T-1)}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta), \qquad (10.27)$$

where $\psi = \{\beta', \sigma_v^2, \rho, \lambda_3\}'$, $\theta = (\beta', \rho)'$, and $u(\theta) = \Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$.

Given λ_3, $\ell_{\mathrm{SE}}(\psi)$ is maximized at

$$\tilde{\theta}(\lambda_3) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega\Delta Y,$$
$$\tilde{\sigma}_v^2(\lambda_3) = \frac{1}{n(T-1)}\Delta\tilde{u}'(\lambda_3)\Omega^{-1}\Delta\tilde{u}(\lambda_3),$$

where $\Delta\tilde{u}(\lambda_3) = \Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_3)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$.

1. Substituting $\tilde{\theta}(\lambda_3)$ and $\tilde{\sigma}_v^2(\lambda_3)$ back into $\ell_{\text{SE}}(\psi)$ gives the concentrated quasi loglikelihood function of λ_3,

$$\ell_{\text{SE}}^c(\lambda_3) = -\frac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_3)) - \frac{1}{2}\log|\Omega|. \tag{10.28}$$

2. Maximizing $\ell_{\text{SE}}^c(\lambda_3)$ gives the CQMLE $\tilde{\lambda}_3$ of λ_3, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\tilde{\lambda}_3)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_3)$ of β and σ_v^2, respectively.

Now, the quasi score $S_{\text{SE}}(\psi) = \frac{\partial}{\partial\psi}\ell_{\text{SE}}(\psi)$ is:

$$S_{\text{SE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}, \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3). \end{cases}$$

Only the ρ-element of $\mathrm{E}[S_{\text{SE}}(\psi_0)]$ is non-zero,

$$\sigma_{v0}^{-2}\mathrm{E}(\Delta u'\Omega^{-1}Y_{-1}) = -n\mathrm{tr}[C^{-1}D(\rho_0)], \tag{10.29}$$

where the $(T-1)\times(T-1)$ matrix $D(\rho)$ has the following expression, noting that $\mathbf{D}_{-1}$ in the Lemma reduces to $D(\rho)\otimes I_n$:

$$D(\rho)=\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \rho-2 & 1 & \cdots & 0 & 0 \\ (1-\rho)^2 & \rho-2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-5}(1-\rho)^2 & \rho^{T-6}(1-\rho)^2 & \cdots & 1 & 0 \\ \rho^{T-4}(1-\rho)^2 & \rho^{T-5}(1-\rho)^2 & \cdots & \rho-2 & 1 \end{pmatrix}.$$

It is easy to see that, when $|\rho|<1$,

$$\mathrm{tr}[C^{-1}D(\rho)]=\tfrac{1}{1-\rho}-\tfrac{1-\rho^T)}{T(1-\rho)^2},$$

which is a result that has appeared in the literature of non-spatial dynamic panel data models (e.g., Nickell, 1981; Lancaster, 2002; and Alvarez and Arellano, 2004), and was derived from different angles.

The result suggests that

- the ρ-element of the conditional quasi score function is such that $\text{plim}_{n\to\infty} \frac{1}{nT\sigma_v^2} \Delta u' \Omega^{-1} \Delta Y_{-1} \neq 0$, unless T also approaches ∞.
- A necessary condition for consistency is violated, and hence the conditional QMLE of ρ is inconsistent when T is fixed.
- It also suggests that even under the large n and large T set up, the conditional QMLE of ρ would incur a bias of order $O(T^{-1})$ as shown in Hahn and Kuersteiner (2002) for the regular DPD model.

With (10.29) and the fact that other score elements have zero expectation, the adjusted quasi score becomes

$$S^*_{\mathrm{SE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \Delta u(\theta), \\ \frac{1}{2\sigma_v^4} \Delta u(\theta)' \Omega^{-1} \Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2} \Delta u(\theta)' \Omega^{-1} \Delta Y_{-1} + n\mathrm{tr}(C^{-1} D(\rho)), \\ \frac{1}{2\sigma_v^2} \Delta u(\theta)' (C^{-1} \otimes A_3) \Delta u(\theta) - (T-1)\mathrm{tr}(G_3). \end{cases} \tag{10.30}$$

- Solving $S^*_{\mathrm{SE}}(\psi) = 0$ leads to the *M*-estimator $\hat{\psi}_{\mathrm{M}}$ of ψ.
- This root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_3)'$, resulting in the constrained *M*-estimators of β and σ_v^2 as

$$\begin{aligned}\hat{\beta}(\delta) &= (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y(\rho)\\ \hat{\sigma}_v^2(\delta) &= \tfrac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta),\end{aligned}$$

  where $\Delta Y(\rho) = \Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho)$.
- Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the last two components of the AQS function in (10.30) gives the concentrated AQS functions:

$$S^{*c}_{\mathrm{SE}}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + n\mathrm{tr}(C^{-1}D(\rho)), \\ \frac{1}{2\hat{\sigma}^2_{v,\mathrm{M}}(\delta)}\Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\mathrm{tr}(G_3). \end{cases} \quad (10.31)$$

- Solving the resulted concentrated estimating equations, $S^{*c}_{\mathrm{SE}}(\delta) = 0$, we obtain the unconstrained *M*-estimators $\hat{\delta}_{\mathrm{M}} = (\hat{\rho}_{\mathrm{M}}, \hat{\lambda}_{3,\mathrm{M}})'$ of δ.

- The unconstrained *M*-estimators of β and σ_v^2 are thus $\hat{\beta}_{\mathrm{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathrm{M}})$ and $\hat{\sigma}^2_{v,\mathrm{M}} \equiv \hat{\sigma}^2_v(\hat{\delta}_{\mathrm{M}})$.
- Compared with the full QML estimation of Su and Yang (2015), the proposed *M*-estimation, though slightly less efficient, is much simpler as it is free from the specification of the initial conditions, and is thus robust against misspecifications of initial conditions.
- In contrast, the full QML estimation requires that the process starting time *m* is known a priori and that the processes evolve in the same manner before and after the data collection.
- Our Monte Carlo results and those in Su and Yang (2015) confirm these points.

Setting $\lambda_2 = \lambda_3 = 0$ gives the FE-DSPD model with only $\texttt{SL}$. Now, $\psi = (\beta', \sigma_v^2, \rho, \lambda_1)'$. The conditional quasi loglikelihood of ψ reduces to:

$$\ell_{\texttt{SL}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \tfrac{1}{2}\log|\mathbf{C}| - \tfrac{1}{2\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta), \quad (10.32)$$

where $\theta = (\theta_1', \lambda_1)'$, $\theta_1 = (\beta', \rho)'$, and $v(\theta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$.

- Given λ_1, $\ell_{\texttt{SL}}(\psi)$ is maximized at

$$\tilde{\theta}_1(\lambda_1) = |(\Delta\mathbb{X}'\mathbf{C}^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\mathbf{C}^{-1}\mathbf{B}_1\Delta Y$$
$$\tilde{\sigma}_v^2(\lambda_1) = \tfrac{1}{n(T-1)}\Delta\tilde{v}'(\lambda_1)\mathbf{C}^{-1}\Delta\tilde{v}(\lambda_1),$$

  where $\Delta\tilde{v}(\lambda_1) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_1)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$.
- Substituting $\tilde{\theta}_1(\lambda_1)$ and $\tilde{\sigma}_v^2(\lambda_1)$ back into $\ell_{\texttt{SL}}(\psi)$ gives the concentrated conditional loglikelihood function of λ_1,

$$\ell_{\texttt{SL}}^c(\lambda_1) = \log|\mathbf{B}_1| - \tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_1)) - \tfrac{1}{2}\log|\mathbf{C}|. \quad (10.33)$$

- Maximizing $\ell_{\texttt{SL}}^c(\lambda_1)$ gives the CQMLE $\tilde{\lambda}_1$ of λ_1, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\hat{\lambda}_1)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1)$ of θ and σ_v^2, respectively.

Now, the CQS function $S_{\mathrm{SL}}(\psi)$ has the form:

$$S_{\mathrm{SL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y - \mathrm{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1). \end{cases}$$

The expectations of the first two components of $S_{\mathrm{SL}}(\psi_0)$ are zero, but these of the last two are not as by the Lemma,

$$\mathrm{E}(\Delta v'\mathbf{C}^{-1}\Delta Y_{-1}) = -\sigma_{v0}^2\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \text{ and} \quad (10.34)$$

$$\mathrm{E}(\Delta v'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y) = -\sigma_{v0}^2\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \quad (10.35)$$

where $\mathbf{D}_{-1}$ and $\mathbf{D}$ are from the general model, but $\mathcal{B}$ simplifies to ρB_1^{-1}.

- These show that the last two elements of $\text{plim}_{n\to\infty}\frac{1}{nT}S_{\text{SL}}(\psi_0)$ are not zero, showing that the CQMLEs of the SL model are inconsistent.
- Even when T grows with n, it can be shown that the CQMLE of ρ has a bias of order $O(T^{-1})$ instead of the desired order $O((nT)^{-1})$.
- Some modifications are thus necessary whether T is fixed or not.

The adjusted quasi score function is,

$$S^*_{\text{SL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1). \end{cases} \qquad (10.36)$$

The M-estimator for the FE-DSPD-SLD model is thus defined as

$$\hat{\psi}_{\text{M}} = \arg\{S^*_{\text{SL}}(\psi) = 0\}.$$

- The root-finding process can be simplified by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_1)'$, leading to the constrained *M*-estimators for β and σ_v^2:

$$\hat{\beta}(\delta) = (\Delta X'\mathbf{C}^{-1}\Delta X)^{-1}\Delta X'\mathbf{C}^{-1}\Delta Y(\delta),$$
$$\hat{\sigma}_v^2(\delta) = \tfrac{1}{n(T-1)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\Delta\tilde{v}(\delta),$$

  where $\Delta Y(\delta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{v}(\delta) = \Delta v(\hat{\beta}(\delta), \delta)$.

- Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ into the last two components of (10.36) gives the concentrated AQS function of δ:

$$S_{\mathrm{SL}}^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_{v,\mathrm{M}}^2(\delta)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1). \end{cases} \quad (10.37)$$

- Solving the concentrated equations, $S_{\mathrm{SL}}^{*c}(\delta) = 0$, gives the unconstrained *M*-estimator $\hat{\delta}_{\mathrm{M}}$ of δ. The unconstrained *M*-estimators of β and σ_v^2 are thus $\hat{\beta}_{\mathrm{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathrm{M}})$ and $\hat{\sigma}_{v,\mathrm{M}}^2 \equiv \hat{\sigma}_v^2(\hat{\delta}_{\mathrm{M}})$.

Setting $\lambda_3 = 0$ gives the FE-DSPD model with $\texttt{SL}$ and $\texttt{STL}$ dependence. Now, $\psi = (\beta', \sigma_v^2, \rho, \lambda_1, \lambda_2)'$, and the conditional quasi loglikelihood of ψ:

$$\ell_{\texttt{STL}}(\psi) = -\frac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \frac{1}{2}\log|\mathbf{C}| - \frac{1}{2\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta), \tag{10.38}$$

where $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$, and $v(\theta) = \mathbf{B}_1\Delta Y - (\rho I_n + \lambda_2\mathbf{W}_2)\Delta Y_{-1} - \Delta X\beta$.

- Let $\theta_1 = (\beta', \rho, \lambda_2)'$. Given λ_1, $\ell_{\texttt{STL}}(\psi)$ is maximized at

$$\tilde{\theta}_1(\lambda_1) = (\Delta\mathbb{X}'\mathbf{C}^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\mathbf{C}^{-1}\mathbf{B}_1\Delta Y,$$
$$\tilde{\sigma}_v^2(\lambda_1) = \frac{1}{n(T-1)}\Delta\tilde{v}'(\lambda_1)\mathbf{C}^{-1}\Delta\tilde{v}(\lambda_1),$$

  where $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1}, \mathbf{W}_2\Delta Y_{-1})$ and $\Delta\tilde{v}(\lambda_1) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda_1)$.
- The concentrated conditional quasi loglikelihood function of λ_1 is,

$$\ell^c_{\texttt{STL}}(\lambda_1) = +\log|\mathbf{B}_1| - \frac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda_1)) - \frac{1}{2}\log|\mathbf{C}|. \tag{10.39}$$

- Maximizing $\ell^c_{\texttt{STL}}(\lambda_1)$ gives the CQMLE $\tilde{\lambda}_1$, and thus the CQMLEs $\tilde{\theta}_1 \equiv \tilde{\theta}_1(\hat{\lambda}_1)$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda}_1)$.

The CQS function $S_{\texttt{STL}}(\psi)$ becomes:

$$S_{\texttt{STL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y - \mathrm{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1}. \end{cases}$$

It is easy to see that the last components of $\mathrm{E}[S_{\texttt{STL}}(\psi_0)]$ are not zero, and are obtained from (10.12)-(10.14). The adjusted quasi score function is,

$$S^*_{\texttt{STL}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\mathbf{C}^{-1}\Delta v(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v(\theta)'\mathbf{C}^{-1}\Delta v(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta v(\theta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2). \end{cases} \quad (10.40)$$

- The M-estimator for the FE-DSPD-SLD model is thus defined as $\hat{\psi}_{\mathtt{SL}} = \arg\{S^*_{\mathtt{M}}(\psi) = 0\}$.
- It can be found by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda_1, \lambda_2)'$, leading to the constrained M-estimators

$$\hat{\beta}(\delta) = (\Delta X'\mathbf{C}^{-1}\Delta X)^{-1}\Delta X'\mathbf{C}^{-1}\Delta Y(\delta),$$
$$\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{v}(\delta)'\mathbf{C}^{-1}\Delta\hat{v}(\delta),$$

where $\Delta Y(\delta) = \mathbf{B}_1\Delta Y - (\rho I_n + \lambda_2\mathbf{W}_2)\Delta Y_{-1}$ and $\Delta\hat{v}(\delta) = \Delta v(\hat{\beta}(\delta), \delta)$;

- and then solving the concentrated estimating equations, $S^{*c}_{\mathtt{SL}}(\delta) = 0$, where the concentrated AQS function of δ has the form:

$$S^{*c}_{\mathtt{STL}}(\delta) = \begin{cases} \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\tilde{v}(\delta)'\mathbf{C}^{-1}\Delta\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{W}_1\mathbf{C}^{-1}\mathbf{D}), \\ \frac{1}{\tilde{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\tilde{v}(\delta)'\mathbf{C}^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2). \end{cases} \quad (10.41)$$

- The M-estimators of β and σ_v^2 are, thus, $\hat{\beta}_{\mathtt{M}} \equiv \hat{\beta}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}^2_{v,\mathtt{M}} \equiv \hat{\sigma}^2_v(\hat{\delta}_{\mathtt{M}})$.

Setting $\lambda_2 = 0$ gives the FE-DSPD model with SL and SE dependence. The conditional quasi loglikelihood of $\psi = (\beta', \sigma_v^2, \rho, \lambda_1, \lambda_3)'$ is,

$$\ell_{\text{SLE}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) + \log|\mathbf{B}_1| - \tfrac{1}{2}\log|\Omega| - \tfrac{1}{2\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta), \tag{10.42}$$

where $\theta = (\beta', \rho, \lambda_1)'$ and $\Delta u(\theta) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\beta$.

- $\ell_{\text{SLE}}(\psi)$ is partially maximized at

$$\tilde{\theta}(\lambda) = (\Delta\mathbb{X}'\Omega^{-1}\Delta\mathbb{X})^{-1}\Delta\mathbb{X}'\Omega^{-1}\mathbf{B}_1\Delta Y,$$
$$\tilde{\sigma}_v^2(\lambda) = \tfrac{1}{n(T-1)}\Delta\tilde{u}'(\lambda)\Omega^{-1}\Delta\tilde{u}(\lambda),$$

  where $\Delta\tilde{u}(\lambda) = \mathbf{B}_1\Delta Y - \Delta\mathbb{X}\,\tilde{\theta}(\lambda)$, and $\Delta\mathbb{X} = (\Delta X, \Delta Y_{-1})$.

- Maximizing the concentrated loglikelihood function of λ,

$$\ell_{\text{SLE}}^{c}(\lambda) = \log|\mathbf{B}_1| - \tfrac{n(T-1)}{2}\log(\tilde{\sigma}_v^2(\lambda)) - \tfrac{1}{2}\log|\Omega|. \tag{10.43}$$

- gives the CQMLE $\tilde{\lambda}$, and thus the CQMLEs $\tilde{\theta} \equiv \tilde{\theta}(\hat{\lambda})$ and $\tilde{\sigma}_v^2 \equiv \tilde{\sigma}_v^2(\tilde{\lambda})$.

The CQS function $S_{\texttt{SLE}}(\psi)$ has the form:

$$S_{\texttt{SLE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3). \end{cases}$$

The ρ and λ_1 components of $\mathrm{E}[S_{\texttt{SLE}}(\psi_0)]$ are not zero, as seen from the Lemma given for the general model:

$$\mathrm{E}(\Delta u'\Omega^{-1}\Delta Y_{-1}) = -\sigma_{v0}^2\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}),$$
$$\mathrm{E}(\Delta u'\Omega^{-1}\mathbf{W}_1\Delta Y) = -\sigma_{v0}^2\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1),$$

which are of identical forms as those for the SLD model.

The existence of SE, parameters in the error terms, does not effect the adjustments on the conditional quasi score!

The results show that the CQMLEs are not consistent unless T is also large. The conditional quasi score function should be modified as:

$$S^*_{\mathrm{SLE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{2\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes A_3)\Delta u(\theta) - (T-1)\mathrm{tr}(G_3). \end{cases} \tag{10.44}$$

The M-estimator of the FE-DPD-SLE model is defined as

$$\hat{\psi}_{\mathrm{M}} = \arg\{S^*_{\mathrm{SLE}}(\psi) = 0\}.$$

which can be found by first solving for β and σ_v^2 to give:

$$\hat{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y(\rho, \lambda_1),$$
$$\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta),$$

where $\Delta Y(\rho, \lambda_1) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1}$ and $\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1)$.

Then, solve the concentrated estimating equations $S^{*c}_{\texttt{SLE}}(\delta) = 0$ to give the *M*-estimator $\hat{\delta}_{\text{M}}$ of δ, where the concentrated AQS function is:

$$S^{*c}_{\texttt{SLE}}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{v,\text{M}}(\delta)} \Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \text{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}^2_{v,\text{M}}(\delta)} \Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \text{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{2\hat{\sigma}^2_{v,\text{M}}(\delta)} \Delta\hat{u}(\delta)'(C^{-1}\otimes A_3)\Delta\hat{u}(\delta) - (T-1)\text{tr}(G_3), \end{cases} \tag{10.45}$$

obtained by substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}^2_v(\delta)$ into the last three components of $S^{*}_{\texttt{SLE}}(\psi)$ given by (10.44).

The *M*-estimators of β and σ^2_v are thus:

$$\hat{\beta}_{\text{M}} \equiv \hat{\beta}(\hat{\delta}_{\text{M}}) \quad \text{and} \quad \hat{\sigma}^2_{v,\text{M}} \equiv \hat{\sigma}^2_v(\hat{\delta}_{\text{M}}).$$

Letting $\hat{\psi}_{\text{M}} = (\hat{\beta}'_{\text{M}}, \hat{\sigma}^2_{v,\text{M}}, \hat{\delta}_{\text{M}})'$.

- The consistency and asymptotic normality of $\hat{\psi}_{\text{M}}$ for the SLE model are implied by the results for the general model.
- The estimate of robust VC matrix is obtained using the relevant submatrices for the general model.

Monte Carlo experiments are carried out to investigate

(*i*) the finite sample performance of the *M*-estimators of the FE-DSPD models,

(*ii*) the finite sample performance of the proposed OPMD estimates of the robust standard errors,

(*iii*) the performance of proposed methods relative to existing ones.

We use the following five models:

$$
\begin{aligned}
\texttt{SE}: \quad & y_t = \rho y_{t-1} + \beta_0 \iota_n + X_t\beta_1 + Z\gamma + \mu + u_t, \quad u_t = \lambda_3 W_3 u_t + v_t, \\
\texttt{SL}: \quad & y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \beta_0 \iota_n + X_t\beta_1 + Z\gamma + \mu + v_t, \\
\texttt{SLE}: \quad & y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \beta_0 \iota_n + X_t\beta_1 + Z\gamma + \mu + u_t, \\
& u_t = \lambda_3 W_3 u_t + v_t, \\
\texttt{STL}: \quad & y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + \beta_0 \iota_n + X_t\beta_1 + Z\gamma + \mu + v_t, \\
\texttt{STLE}: \quad & y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + \beta_0 \iota_n + X_t\beta_1 + Z\gamma + \mu + u_t, \\
& u_t = \lambda_2 W_3 u_t + v_t.
\end{aligned}
$$

Table 10.1a. Empirical Mean(sd) of CQMLE, FQMLE and M-Estimator, SE Model, $T = 3$, $m = 5$

		$n = 50$			$n = 200$		
err	ψ	CQMLE	FQMLE	M-Est	CQMLE	FQMLE	M-Est
1	1	1.0152(.096)	1.0017(.100)	1.0015(.100)	1.0109(.050)	1.0021(.052)	1.0020(.053)
	1	.9154(.135)	.9678(.148)	.9719(.154)	.9080(.065)	.9960(.079)	.9962(.080)
	.5	.3605(.055)	.4995(.065)	.5015(.066)	.2869(.033)	.5009(.043)	.5013(.044)
	.5	.4702(.107)	.4761(.093)	.4793(.105)	.4775(.073)	.4877(.060)	.4907(.070)
2	1	1.0142(.098)	1.0007(.102)	1.0002(.102)	1.0099(.050)	1.0015(.053)	1.0014(.053)
	1	.9176(.266)	.9662(.284)	.9785(.307)	.9045(.128)	.9920(.152)	.9935(.155)
	.5	.3610(.066)	.4975(.069)	.5023(.078)	.2876(.041)	.5002(.047)	.5018(.052)
	.5	.4701(.106)	.4770(.092)	.4803(.104)	.4741(.075)	.4844(.063)	.4883(.072)
3	1	1.0133(.099)	1.0001(.103)	.9997(.103)	1.0090(.047)	1.0003(.049)	1.0003(.049)
	1	.9192(.198)	.9678(.212)	.9771(.227)	.9060(.099)	.9938(.119)	.9947(.121)
	.5	.3585(.059)	.4953(.066)	.4992(.071)	.2881(.036)	.5018(.046)	.5029(.048)
	.5	.4681(.110)	.4736(.093)	.4786(.106)	.4741(.075)	.4852(.062)	.4884(.073)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_3)'$; err=1 (normal), 2 (normal mixture), and 3 (chi-square).

X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 1, .5)$, as in Footnote 1.

W_3 is generated according to Group Interaction scheme as in Footnote 2.

Table 10.1b. Empirical Mean(sd) of CQMLE, FQMLE and M-Estimator, SE Model, $T = 7$, $m = 5$

		$n = 50$			$n = 100$		
err	ψ	CQMLE	FQMLE	M-Est	CQMLE	FQMLE	M-Est
1	1	1.0248(.044)	1.0015(.044)	1.0013(.044)	1.0231(.033)	1.0018(.033)	1.0017(.033)
	1	.9771(.081)	.9888(.083)	.9893(.083)	.9821(.059)	.9949(.060)	.9956(.061)
	.5	.4456(.028)	.4987(.029)	.4990(.029)	.4407(.021)	.4990(.022)	.4994(.022)
	.5	.4928(.057)	.4920(.055)	.4947(.056)	.4931(.047)	.4904(.044)	.4953(.046)
2	1	1.0247(.045)	1.0012(.045)	1.0010(.045)	1.0232(.033)	1.0020(.033)	1.0019(.033)
	1	.9776(.183)	.9887(.186)	.9899(.187)	.9806(.129)	.9931(.132)	.9942(.133)
	.5	.4461(.028)	.4988(.029)	.4992(.029)	.4412(.022)	.4991(.022)	.4996(.022)
	.5	.4919(.058)	.4914(.055)	.4941(.057)	.4906(.048)	.4882(.046)	.4928(.047)
3	1	1.0250(.044)	1.0015(.044)	1.0013(.044)	1.0214(.033)	1.0003(.033)	1.0002(.033)
	1	.9751(.130)	.9863(.133)	.9872(.134)	.9779(.095)	.9908(.097)	.9915(.097)
	.5	.4458(.028)	.4986(.029)	.4990(.029)	.4413(.020)	.4996(.021)	.5000(.021)
	.5	.4903(.057)	.4896(.056)	.4923(.057)	.4919(.048)	.4898(.045)	.4940(.047)

Table 10.1c. Empirical sd and average of estimated standard errors of M-Estimator
SE Model, $T = 3$, $m = 5$, Parameter configurations as in Table 10.1a.

		$n = 50$				$n = 100$				$n = 200$			
dgp	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.100	.112	.099	.096	.071	.073	.070	.069	.053	.053	.051	.051
	1	.154	.165	.150	.146	.113	.114	.110	.109	.080	.081	.079	.080
	.5	.066	.068	.064	.065	.059	.054	.054	.056	.044	.040	.042	.044
	.5	.105	.111	.099	.096	.083	.086	.081	.080	.070	.070	.068	.068
2	1	.102	.124	.099	.093	.071	.078	.069	.068	.053	.055	.051	.050
	1	.307	.117	.152	.263	.209	.076	.110	.198	.155	.050	.079	.147
	.5	.078	.071	.064	.070	.065	.053	.054	.063	.052	.037	.042	.051
	.5	.104	.126	.099	.090	.089	.095	.082	.078	.072	.074	.068	.067
3	1	.103	.117	.099	.095	.070	.075	.069	.069	.049	.053	.051	.051
	1	.227	.133	.151	.203	.162	.089	.110	.153	.121	.061	.079	.113
	.5	.071	.070	.064	.066	.062	.053	.054	.060	.048	.039	.042	.047
	.5	.106	.118	.099	.093	.088	.091	.082	.079	.073	.072	.068	.067

Table 10.2a. Empirical Mean(sd) of CQMLE and M-Estimator, SL Model, $T = 3$, $m = 5$, $rho = 0.5$

		$n = 50$		$n = 100$		$n = 200$	
err	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0190(.053)	.9992(.055)	1.0024(.035)	1.0004(.036)	1.0112(.025)	.9998(.025)
	1	.9365(.133)	.9695(.143)	.9620(.096)	.9855(.100)	.9657(.068)	.9949(.072)
	.5	.4279(.042)	.5015(.047)	.4467(.024)	.5007(.026)	.4310(.020)	.4995(.022)
	.2	.2331(.060)	.1933(.064)	.2114(.049)	.1953(.052)	.2048(.039)	.1980(.040)
2	1	1.0193(.052)	.9992(.054)	1.0005(.034)	.9984(.034)	1.0116(.026)	1.0003(.026)
	1	.9391(.260)	.9743(.280)	.9558(.184)	.9797(.194)	.9635(.137)	.9929(.145)
	.5	.4289(.045)	.5031(.048)	.4474(.027)	.5012(.028)	.4318(.022)	.5000(.023)
	.2	.2335(.061)	.1938(.065)	.2124(.050)	.1967(.053)	.2030(.037)	.1962(.039)
3	1	1.0180(.055)	.9980(.056)	1.0019(.035)	.9998(.036)	1.0111(.024)	.9997(.025)
	1	.9388(.203)	.9730(.218)	.9581(.147)	.9817(.155)	.9623(.102)	.9913(.108)
	.5	.4277(.043)	.5018(.047)	.4461(.027)	.5000(.029)	.4319(.021)	.4999(.023)
	.2	.2354(.060)	.1960(.064)	.2121(.050)	.1962(.052)	.2054(.037)	.1986(.039)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_1)'$; err = 1 (normal), 2 (normal mixture), and 3 (chi-square).

X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$, as in Footnote 1.

W_1 is generated according to Queen Contiguity scheme.

Table 10.2b. Empirical Mean(sd) of CQMLE and M-Estimator, SL Model, $T = 3$, $m = 5$, $rho = -0.5$

		$n = 50$		$n = 100$		$n = 200$	
err	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0261(.057)	.9978(.059)	1.0188(.037)	.9993(.037)	1.0226(.027)	.9992(.027)
	1	.9600(.136)	.9753(.141)	.9761(.099)	.9891(.102)	.9797(.068)	.9917(.070)
	-.5	-.5577(.046)	-.4976(.050)	-.5543(.031)	-.4989(.034)	-.5513(.022)	-.4991(.024)
	.2	.1898(.096)	.1841(.097)	.1855(.059)	.1978(.059)	.1824(.044)	.1975(.044)
2	1	1.0254(.056)	.9971(.058)	1.0179(.037)	.9985(.037)	1.0228(.028)	.9993(.028)
	1	.9557(.281)	.9719(.290)	.9745(.201)	.9878(.206)	.9857(.140)	.9980(.143)
	-.5	-.5557(.048)	-.4958(.052)	-.5542(.033)	-.4990(.035)	-.5518(.023)	-.4995(.024)
	.2	.1985(.093)	.1920(.094)	.1853(.058)	.1974(.058)	.1815(.043)	.1966(.043)
3	1	1.0261(.057)	.9978(.059)	1.0184(.037)	.9989(.037)	1.0225(.028)	.9990(.028)
	1	.9487(.198)	.9640(.204)	.9752(.155)	.9884(.159)	.9808(.105)	.9929(.108)
	-.5	-.5562(.048)	-.4971(.052)	-.5547(.034)	-.4992(.037)	-.5514(.023)	-.4991(.025)
	.2	.1938(.097)	.1877(.098)	.1824(.058)	.1943(.058)	.1827(.043)	.1974(.044)

Table 10.2c. Empirical sd and average of estimated standard errors of M-Estimator

SL Model, $T = 3$, $m = 5$, Parameter configurations as in Table 10.2a.

		$n = 50$				$n = 100$				$n = 200$			
err	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.055	.060	.054	.052	.036	.036	.035	.034	.025	.026	.026	.025
	1	.143	.158	.142	.138	.100	.106	.101	.100	.072	.075	.073	.072
	.5	.047	.049	.044	.044	.026	.028	.026	.026	.022	.021	.021	.021
	.2	.064	.070	.063	.061	.052	.048	.052	.059	.040	.037	.039	.042
2	1	.054	.066	.054	.052	.034	.039	.034	.034	.026	.027	.026	.025
	1	.280	.105	.143	.255	.194	.063	.101	.190	.145	.042	.073	.141
	.5	.048	.052	.044	.046	.028	.029	.026	.027	.023	.021	.021	.022
	.2	.065	.077	.063	.061	.053	.051	.052	.058	.039	.039	.039	.042
3	1	.056	.062	.054	.052	.036	.037	.034	.034	.025	.027	.026	.025
	1	.218	.122	.143	.196	.155	.078	.101	.144	.108	.053	.072	.105
	.5	.047	.050	.044	.045	.029	.028	.026	.027	.023	.021	.021	.022
	.2	.064	.074	.063	.060	.052	.049	.052	.058	.039	.038	.039	.042

Table 10.3a. Empirical Mean(sd) of CQMLE and M-Estimator, SLE Model, $T = 3$, $m = 5$

		$n = 50$		$n = 100$		$n = 200$	
err	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0028(.051)	1.0012(.052)	.9894(.035)	.9990(.036)	1.0130(.028)	1.0001(.028)
	1	.9268(.133)	.9542(.141)	.9491(.093)	.9800(.100)	.9629(.070)	.9911(.075)
	.5	.4332(.041)	.4993(.044)	.4285(.029)	.4999(.032)	.4337(.019)	.5010(.021)
	.2	.2064(.078)	.1930(.083)	.2190(.069)	.1905(.075)	.1871(.063)	.1938(.066)
	.2	.1383(.185)	.1489(.183)	.1561(.148)	.1582(.146)	.1781(.122)	.1723(.120)
2	1	1.0006(.050)	.9989(.051)	.9908(.036)	1.0005(.037)	1.0123(.027)	.9995(.028)
	1	.9219(.263)	.9505(.280)	.9495(.188)	.9813(.200)	.9626(.138)	.9910(.146)
	.5	.4355(.043)	.5011(.045)	.4291(.032)	.5006(.034)	.4332(.022)	.5001(.023)
	.2	.2016(.078)	.1881(.083)	.2199(.067)	.1921(.072)	.1881(.064)	.1940(.067)
	.2	.1411(.175)	.1525(.171)	.1597(.148)	.1634(.145)	.1778(.122)	.1733(.120)
3	1	1.0001(.052)	.9984(.053)	.9920(.036)	1.0015(.037)	1.0121(.028)	.9993(.028)
	1	.9247(.199)	.9527(.212)	.9461(.143)	.9771(.152)	.9596(.102)	.9875(.108)
	.5	.4345(.042)	.5006(.046)	.4287(.031)	.4996(.034)	.4324(.021)	.4991(.022)
	.2	.2038(.080)	.1901(.086)	.2209(.071)	.1925(.076)	.1898(.063)	.1955(.066)
	.2	.1394(.186)	.1510(.182)	.1598(.148)	.1629(.145)	.1745(.121)	.1695(.118)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_3)'$; err = 1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$, as in Footnote 1.
W_1 and W_3 are from Group Interaction scheme, and not equal; see Footnote 2.

Table 10.3b. Empirical sd and average of estimated standard errors of M-Estimator
SLE Model, $T = 3$, $m = 5$, Parameter configurations as in Table 10.3a.

		$n = 50$				$n = 100$				$n = 200$			
err	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.052	.056	.051	.050	.036	.038	.036	.035	.028	.029	.028	.028
	1	.141	.157	.140	.137	.100	.108	.102	.100	.075	.075	.072	.072
	.5	.044	.045	.042	.042	.032	.032	.031	.031	.021	.021	.021	.021
	.2	.083	.072	.080	.098	.075	.065	.072	.086	.066	.056	.064	.076
	.2	.183	.179	.160	.163	.146	.143	.135	.138	.120	.116	.114	.117
2	1	.051	.062	.050	.049	.037	.041	.036	.035	.028	.030	.028	.027
	1	.280	.107	.140	.247	.200	.067	.102	.192	.146	.043	.072	.141
	.5	.045	.049	.041	.042	.034	.033	.031	.033	.023	.021	.021	.022
	.2	.083	.080	.079	.095	.072	.070	.072	.085	.067	.059	.064	.074
	.2	.171	.208	.160	.153	.145	.155	.134	.134	.120	.122	.113	.114
3	1	.053	.058	.051	.050	.037	.039	.036	.035	.028	.029	.028	.028
	1	.212	.123	.140	.190	.152	.080	.101	.144	.108	.054	.072	.105
	.5	.046	.046	.041	.043	.034	.031	.031	.033	.022	.021	.021	.022
	.2	.086	.076	.080	.097	.076	.067	.072	.085	.066	.057	.064	.075
	.2	.182	.191	.160	.159	.145	.147	.134	.136	.118	.119	.114	.115

Table 10.4a. Empirical Mean(sd) of CQMLE and M-Estimator, STL Model, $T = 7$, $m = 5$

		$n = 50$		$n = 100$		$n = 200$	
err	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0045(.025)	1.0000(.025)	1.0073(.017)	.9998(.017)	1.0078(.012)	1.0002(.012)
	1	.9809(.079)	.9855(.080)	.9897(.058)	.9940(.058)	.9923(.041)	.9966(.041)
	.5	.4780(.018)	.4996(.019)	.4807(.012)	.5001(.012)	.4812(.008)	.5001(.009)
	.2	.1904(.046)	.1950(.046)	.1971(.031)	.1994(.031)	.1974(.023)	.1984(.023)
	.2	.2280(.041)	.2043(.042)	.2208(.030)	.2006(.030)	.2200(.021)	.2013(.021)
2	1	1.0040(.026)	.9994(.026)	1.0078(.017)	1.0003(.018)	1.0074(.012)	.9998(.012)
	1	.9920(.182)	.9968(.184)	.9884(.130)	.9927(.131)	.9934(.090)	.9977(.091)
	.5	.4780(.018)	.4999(.019)	.4805(.013)	.4999(.013)	.4816(.009)	.5005(.009)
	.2	.1908(.046)	.1954(.046)	.1962(.032)	.1985(.032)	.1986(.023)	.1995(.023)
	.2	.2276(.042)	.2038(.042)	.2216(.031)	.2014(.031)	.2186(.022)	.1999(.022)
3	1	1.0050(.025)	1.0006(.025)	1.0076(.018)	1.0001(.018)	1.0075(.012)	.9999(.012)
	1	.9815(.135)	.9861(.137)	.9912(.095)	.9955(.096)	.9903(.067)	.9945(.068)
	.5	.4783(.018)	.4999(.018)	.4805(.012)	.4999(.012)	.4810(.008)	.4999(.009)
	.2	.1905(.048)	.1950(.048)	.1954(.031)	.1977(.031)	.1978(.023)	.1988(.023)
	.2	.2278(.042)	.2041(.042)	.2226(.030)	.2024(.030)	.2198(.022)	.2012(.022)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2)'$; err = 1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$, as in Footnote 1.
W_1 and W_2 are from Queen Contiguity, and equal.

Table 10.4b. Empirical sd and average of estimated standard errors of M-Estimator STL Model, $T = 7$, $m = 5$, Parameter configurations as in Table 10.4a.

		$n = 50$				$n = 100$				$n = 200$			
err	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.025	.027	.025	.024	.017	.018	.017	.017	.012	.012	.012	.012
	1	.080	.088	.081	.080	.058	.060	.058	.057	.041	.042	.041	.041
	.5	.019	.019	.018	.019	.012	.012	.012	.013	.009	.008	.009	.009
	.2	.046	.049	.047	.048	.031	.032	.032	.032	.023	.023	.023	.024
	.2	.042	.044	.042	.044	.030	.030	.030	.032	.021	.021	.022	.023
2	1	.026	.029	.025	.025	.018	.019	.017	.017	.012	.013	.012	.012
	1	.184	.046	.082	.174	.131	.029	.058	.126	.091	.020	.041	.091
	.5	.019	.021	.019	.019	.013	.013	.012	.013	.009	.008	.009	.009
	.2	.046	.052	.047	.048	.032	.033	.031	.032	.023	.024	.023	.024
	.2	.042	.047	.042	.044	.031	.031	.030	.032	.022	.021	.022	.023
3	1	.025	.028	.025	.024	.018	.018	.017	.017	.012	.013	.012	.012
	1	.137	.060	.081	.126	.096	.039	.058	.092	.068	.026	.041	.066
	.5	.018	.020	.018	.019	.012	.012	.012	.013	.009	.008	.009	.009
	.2	.048	.050	.046	.048	.031	.033	.032	.032	.023	.023	.023	.024
	.2	.042	.045	.042	.044	.030	.030	.030	.032	.022	.021	.022	.023

Table 10.5a. Empirical Mean(sd) of CQMLE and M-Estimator, STLE Model, $T = 3$, $m = 5$

		$n = 50$		$n = 100$		$n = 200$	
err	ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
1	1	1.0076(.031)	.9995(.031)	1.0092(.024)	.9999(.025)	1.0057(.016)	.9999(.016)
	1	.9287(.133)	.9437(.138)	.9622(.099)	.9746(.102)	.9723(.070)	.9841(.071)
	.3	.2578(.033)	.3000(.035)	.2663(.021)	.2992(.022)	.2685(.014)	.2999(.015)
	.2	.1966(.073)	.1957(.075)	.1877(.059)	.1967(.060)	.2030(.037)	.1987(.038)
	.2	.2209(.084)	.2055(.091)	.2240(.047)	.2037(.049)	.2019(.037)	.2005(.039)
	.2	.1463(.185)	.1459(.191)	.1846(.132)	.1745(.135)	.1785(.092)	.1838(.094)
2	1	1.0084(.031)	1.0003(.032)	1.0091(.025)	.9998(.025)	1.0055(.016)	.9997(.016)
	1	.9288(.264)	.9448(.273)	.9591(.195)	.9717(.201)	.9713(.140)	.9832(.144)
	.3	.2564(.035)	.2986(.036)	.2661(.022)	.2988(.022)	.2685(.015)	.2998(.015)
	.2	.1989(.073)	.1983(.076)	.1895(.060)	.1985(.060)	.2043(.036)	.1999(.037)
	.2	.2168(.082)	.2008(.089)	.2211(.047)	.2008(.050)	.2015(.036)	.2003(.039)
	.2	.1367(.184)	.1358(.190)	.1805(.131)	.1702(.134)	.1799(.090)	.1855(.092)
3	1	1.0066(.031)	.9985(.031)	1.0084(.025)	.9991(.025)	1.0054(.016)	.9995(.016)
	1	.9318(.201)	.9476(.208)	.9644(.148)	.9769(.152)	.9727(.102)	.9846(.105)
	.3	.2590(.034)	.3014(.036)	.2679(.022)	.3008(.022)	.2689(.015)	.3003(.015)
	.2	.1983(.071)	.1978(.073)	.1879(.060)	.1969(.060)	.2046(.037)	.2003(.038)
	.2	.2193(.081)	.2035(.087)	.2207(.047)	.2003(.049)	.2009(.037)	.1998(.040)
	.2	.1412(.184)	.1403(.190)	.1852(.133)	.1750(.135)	.1794(.093)	.1849(.095)

Note: Par = $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; err = 1 (normal), 2 (normal mixture), and 3 (chi-square).
X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, \mathbf{3}, 1)$, as in Footnote 1.
W_1, W_2 and W_3 are all from Queen Contiguity, and equal.

Table 10.5b. Empirical sd and average of estimated standard errors of M-Estimator
STLE Model, $T = 3$, $m = 5$, Parameter configurations as in Table 10.5a.

		$n = 50$				$n = 100$				$n = 200$			
err	ψ	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$	sd	$\widetilde{se}$	$\widehat{se}$	$\widehat{rse}$
1	1	.031	.035	.031	.030	.025	.026	.025	.025	.016	.017	.016	.016
	1	.138	.156	.136	.132	.102	.106	.099	.098	.071	.073	.071	.070
	.3	.035	.038	.035	.035	.022	.023	.021	.022	.015	.015	.015	.015
	.2	.075	.080	.071	.071	.060	.062	.059	.060	.038	.039	.038	.038
	.2	.091	.079	.085	.108	.049	.043	.049	.060	.039	.033	.039	.050
	.2	.191	.209	.184	.186	.135	.141	.132	.132	.094	.096	.093	.094
2	1	.032	.039	.031	.030	.025	.028	.025	.024	.016	.017	.016	.016
	1	.273	.109	.137	.239	.201	.065	.099	.187	.144	.042	.071	.139
	.3	.036	.042	.034	.035	.022	.024	.021	.022	.015	.016	.015	.015
	.2	.076	.089	.070	.069	.060	.067	.059	.059	.037	.041	.038	.038
	.2	.089	.088	.083	.104	.050	.047	.049	.059	.039	.034	.039	.049
	.2	.190	.243	.184	.177	.134	.154	.132	.130	.092	.102	.093	.092
3	1	.031	.037	.031	.030	.025	.027	.025	.025	.016	.017	.016	.016
	1	.208	.124	.137	.187	.152	.079	.099	.142	.105	.052	.071	.104
	.3	.036	.040	.035	.035	.022	.023	.021	.022	.015	.015	.015	.015
	.2	.073	.084	.070	.069	.060	.065	.059	.059	.038	.040	.038	.038
	.2	.087	.084	.084	.106	.049	.045	.049	.059	.040	.033	.039	.050
	.2	.190	.224	.183	.181	.135	.147	.131	.130	.095	.098	.093	.094

Public Capital Productivity. To facilitate the practical applications of the proposed methods, we provide an empirical illustration using the well known data set on public capital productivity of Munnell (1990).

- The dataset gives indicators related to public capital productivity for 48 US states observed over 17 years (1970-1986).
- The dataset can be downloaded from http://pages.stern.nyu.edu/∼wgreene/Text/Edition6/tablelist6.htm
- This dataset has been extensively used for illustrating the applications of the regular panel data models (see, e.g., Baltagi, 2013).
- In the spatial framework, it was used by Millo and Piras (2012) for illustrating the QML and GMM estimation of fixed effects and random effects spatial panel data models,
- and by Yang et al. (2016) for illustrating the bias-correction and refined inferences for fixed effects spatial panel data models.

In Munnel (1990), the empirical model specified is a Cobb-Douglas production function of the form:

$$\ln Y = \beta_0 + \beta_1 \ln K_1 + \beta_2 \ln K_2 + \beta_3 \ln L + \beta_4 \text{Unemp} + \epsilon,$$

with state specific fixed effects, where

- Y is the gross social product of a given state,
- K_1 is public capital,
- K_2 is private capital,
- L is labour input and
- Unemp is the state unemployment rate.

This model is now extended by adding the dynamic effect and one or more spatial effects. The spatial weights matrix W takes a contiguity form with its (i, j)th element being 1 if states i and j share a common border, otherwise 0. The final W is row normalised. For models with more than one spatial term, the corresponding $W's$ are taken to be the same.

Each of the five models discussed in this lecture is estimated using

- (a) full data,
- (b) data from the last six years ($T + 1 = 6$),
- and (c) data from first six years.

Table 10.6a summarize the CQMLE, FQMLE, M-Est and the standard error of the M-Est for the SE model, as for this model the full QMLE is available (Su and Yang, 2015). From the results we see that

- (i) the dynamic and SE effects are highly significant in all models,
- (ii) three methods give quite different estimates of dynamic effect,
- and (iii) the FQMLE of ρ improves over CQMLE in that it is much closer to the M-estimate in particular when T is small.
- FQMLE uses $m = 6$, and the time mean of the regressors as the predictor for the initial differences. The results are quite robust to the value of m, but not quite to the choice of the predictors.

Table 10.6a. CQMLE, FQMLE, , M-Est and its *t*-Ratio based on Munnell Data: SE Model

	Full Data				Last 6 Years				First 6 Years			
	CQM	FQM	M-Est	*t*-ratio	CQM	FQM	M-Est	*t*-ratio	CQM	FQM	M-Est	*t*-ratio
β_1	-.0433	-.0234	-.0467	-1.877	-.1008	-.1124	-.0852	-2.440	-.0851	-.0922	-.0810	-1.136
β_2	-.0393	-.0309	-.0702	-2.796	-.0305	-.0336	-.0501	-1.373	.0644	.0106	-.0714	-.639
β_3	.2644	.2008	.1654	3.329	.7840	.6504	.5971	5.526	.4192	.3532	.3161	2.353
β_4	-.0024	-.0026	-.0028	-5.306	-.002	-.0018	-.0021	-3.590	-.0028	-.0031	-.0031	-4.389
σ_v^2	.0001	.0001	.0001	5.931	.0000	.0000	.0000	5.366	.0000	.0000	.0000	3.998
ρ	.7772	.8283	.9140	17.222	.4409	.5728	.6265	7.162	.4594	.5942	.6521	4.018
λ_3	.7592	.7550	.7697	20.665	.7133	.7460	.7638	14.021	.7114	.7120	.7155	13.842

Table 10.6b summarize the results for the other four models. The results show that, for any model estimated and data used,

- (*i*) the dynamic effect is alway significant,
- (*ii*) there is alway at least one spatial effect that is significant,
- and (*iii*) the CQMLE is always significantly smaller than the corresponding M-estimate.
- The empirical results are consistent with the theories.

Table 10.6b. CQMLE, M-Est and its *t*-Ratio based on Munnell Data: Other Models

	Full Data			Last 6 Years			First 6 Years		
	CQMLE	M-Est	*t*-ratio	CQMLE	M-Est	*t*-ratio	CQMLE	M-Est	*t*-ratio
SL Model									
β_1	-0.0620	-0.0598	-1.8194	-0.1850	-0.1692	-2.5069	-0.0165	-0.0079	-0.1005
β_2	0.0296	0.0105	0.3514	-0.0365	-0.0540	-1.1542	-0.1081	-0.2194	-2.7020
β_3	0.3045	0.2480	3.1542	0.9917	0.9012	10.4729	0.3916	0.2369	1.2416
β_4	-0.0025	-0.0027	-4.0988	-0.0016	-0.0019	-2.5384	-0.0018	-0.0018	-2.5330
σ_v^2	0.0001	0.0001	9.5094	0.0001	0.0001	8.6974	0.0001	0.0001	3.5254
ρ	0.5333	0.6132	7.0194	0.1625	0.2448	4.4754	0.2849	0.4801	2.8386
λ_1	0.2131	0.2046	4.3797	0.2077	0.1991	4.4475	0.3767	0.4134	4.0345
SLE Model									
β_1	-0.0412	-0.0454	-1.6237	-0.0888	-0.0755	-1.8749	-0.1023	-0.0829	-1.1113
β_2	-0.0364	-0.0675	-1.3981	-0.0197	-0.0373	-0.8777	0.4341	0.0429	0.1011
β_3	0.2649	0.1685	1.4418	0.7585	0.5904	5.3430	0.4201	0.3343	2.2261
β_4	-0.0024	-0.0027	-3.9247	-0.0021	-0.0023	-3.5416	-0.0025	-0.0031	-3.8491
σ_v^2	0.0001	0.0001	5.1623	0.0000	0.0000	4.8590	0.0000	0.0000	2.9107
ρ	0.7752	0.9092	5.9496	0.4515	0.6189	8.0173	0.3754	0.6123	2.9549
λ_1	-0.0235	-0.0123	-0.3143	-0.0804	-0.0789	-0.8565	-0.3615	-0.1289	-0.4139
λ_3	0.7753	0.7757	17.6446	0.7800	0.8015	10.7070	0.8878	0.7789	4.2353

Table 4.6b. Cont'd

	Full Data			Last 6 Years			First 6 Years		
	CQMLE	M-Est	*t*-ratio	CQMLE	M-Est	*t*-ratio	CQMLE	M-Est	*t*-ratio
STL Model									
β_1	-0.0383	-0.0343	-1.2882	-0.1367	-0.1072	-3.0105	-0.0791	-0.0727	-0.8560
β_2	0.0215	0.0040	0.1641	-0.0158	-0.0262	-0.6303	0.1456	0.0937	0.8758
β_3	0.2414	0.1844	2.9434	0.7215	0.5669	5.5058	0.4769	0.4040	4.3346
β_4	-0.0011	-0.0012	-3.4687	-0.0014	-0.0017	-2.8457	-0.0017	-0.0018	-3.1086
σ_v^2	0.0001	0.0001	6.1872	0.0000	0.0000	5.0666	0.0000	0.0000	4.9172
ρ	0.7547	0.8474	12.1490	0.4757	0.6365	7.2715	0.4258	0.5700	4.6003
λ_1	0.6662	0.681	15.2637	0.4890	0.5409	7.9038	0.5533	0.5565	10.9247
λ_2	-0.6350	-0.6747	-11.3723	-0.466	-0.5797	-6.4991	-0.5343	-0.5775	-4.5748
STLE Model									
β_1	-0.0399	-0.0432	-1.7639	-0.1255	-0.1071	-2.8461	-0.0657	-0.0322	-0.2979
β_2	-0.0370	-0.0617	-1.3938	-0.0180	-0.0264	-0.5836	0.1254	0.0584	0.5115
β_3	0.2146	0.1353	1.2129	0.7684	0.5690	3.7925	0.4517	0.3512	2.5418
β_4	-0.0023	-0.0026	-3.5825	-0.0017	-0.0017	-2.3548	-0.0015	-0.0012	-1.1755
σ_v^2	0.0000	0.0001	4.5221	0.0000	0.0000	5.0517	0.0000	0.0000	4.2264
ρ	0.7973	0.9164	6.2388	0.4484	0.6349	5.3390	0.4367	0.6001	3.8399
λ_1	-0.5538	-0.5566	-5.3667	0.4137	0.5381	3.6888	0.5976	0.6711	3.9109
λ_2	0.4985	0.5331	4.8853	-0.4138	-0.5770	-3.6064	-0.5514	-0.6536	-3.4999
λ_3	0.9074	0.9059	31.9162	0.2058	0.0078	0.0237	-0.1215	-0.3409	-0.6752

Please see "Computing Lab 6" for details.

References

- Anderson, T. W., Hsiao, C., 1981. Estimation of dynamic models with error components. *Journal of American Statistical Association* 76, 598-606.
- Anderson, T. W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47-82.
- Bhargava, A., Sargan, J. D., 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635-1659.
- Elhorst, J. P., 2010. Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40, 272-282.
- Elhorst, J. P., 2012a. Dynamic spatial panels: models, methods, and applications. *Journal of Geographical Systems* 14, 5-28.
- Hayakawa, K., Pesaran, M.H. 2015. Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* 188, 111-134.
- Hayakawa, K., Pesaran, M.H., Smith, L. V., 2020. Short T Dynamic Panel Data Models with Individual and Interactive Time Effects. *Working Paper*.

- Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short panels. *Journal of Econometrics* 109, 107-150.
- Kuersteiner, G. M., Prucha, I. R., 2020. Dynamic panel data models: networks, common shocks, and sequential exogeneity. *Econometrica* 88, 2109-2146.
- Lee, L. F., Yu, J., 2010c. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40, 255-271.
- Lee, L. F., Yu, J., 2016. Identification of spatial Durbin panel models. *Journal of Applied Econometrics* 31, 133-162.
- Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.
- Yang, Z. L., 2018a. Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics* 205, 423-447.
- Yang, Z. L., 2018b. A supplement to “Unified M-estimation of fixed-effects spatial dynamic models with short panels”. *Working paper, Singapore Management University*, http://www.mysmu.edu/faculty/zlyang/.

- Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.
- van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge: Cambridge University Press.
- Baltagi, B. H., Pirotte, A. and Yang, Z. L. (2021). Diagnostic tests for homoscedasticity in spatial cross-sectional or panel models. *Journal of Econometrics* 224, 245-270.
- Li, L. Y. and Yang, Z. L. (2020). Estimation of Fixed Effects Spatial Dynamic Panel Data Models with Small T and Unknown Heteroskedasticity. *Regional Science and Urban Economics* 81, 103520.
- Li, L. Y. and Yang, Z. L. (2021). Spatial Dynamic Panel Data Models with Correlated Random Effects. *Journal of Econometrics* 221, 424-454.
- Yang, Z. L. (2021). Joint tests for dynamic and spatial effects in short dynamic panel data models with fixed effects and heteroskedasticity. Empirical Economics 60, 51-92.
- Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.

- Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.
- Lancaster, T., 2002. Orthogonal parameters and panel data. *Review of Economic Studies* 69, 647-666.
- Hahn J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.