Lecture 12: Tests of Hypotheses for DSPD Models

Zhenlin Yang

School of Economics, Singapore Management University

ECON747 Spatial Econometrics, Term I 2024-25

12.1 Introduction

This lecture is based on Yang (2021a, Empirical Economics).

Simple and reliable tests are proposed for testing the existence of dynamic and/or spatial effects in fixed-effects panel data models with small T and possibly heteroskedastic errors. The tests are constructed based on the *adjusted quasi scores* (AQS), which correct the conditional quasi scores given the initial differences to account for the effect of *initial values*.

To improve the finite sample performance, standardized AQS tests are also derived, which are shown to have much improved finite sample properties. All the proposed tests are robust against nonnormality, but some are not robust against cross-sectional heteroskedasticity (CH).

A different type of adjustments are made on the AQS functions, leading to a set of tests that are fully robust against unknown CH.

Monte Carlo results show excellent finite sample performance of the standardized versions of the AQS tests.

Panel data (PD) model has been an important tool for applied economics researchers over the past few decades.

However, there have been growing concerns on whether panel models are dynamic in nature due to the impacts from the past to the current and future ‘economic’ performance, and whether the models contain spatial dependence due to interactions among economic agents or social actors, e.g., neighbourhood, copy-catting, social network, and peer group effects.

In other words, there have been growing concerns from applied researchers on whether a dynamic spatial panel data model (DSPD) is more appropriate than a regular PD model, or a regular dynamic panel data (DPD) model, or a static spatial panel data (SPD) model.

Thus, it is highly desirable to device simple and reliable tests helping applied researchers to choose the most appropriate model.

Consider the following dynamic spatial panel data (DSPD) model studied by Yang (2018a) under large-n and small-T, presented in Lecture 10:

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t'\beta + Z\gamma + \mu + \alpha_t 1_n + u_t, \\ & u_t = \lambda_3 W_3 u_t + v_t, \quad t = 1, \ldots, T, \end{aligned} \tag{12.1}$$

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, $iid(0, \sigma_v^2)$ or $inid(0, \sigma_{vi}^2)$.

Note: When T is small, the $\alpha_t 1_n$ part can be merged into $X_t'\beta$.

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

- Model (12.1) is fairly general, embedding several important submodels popular in the literature.
- As T is fixed and small, the time specific effects $\{\alpha_t\}$ are always treated as fixed effects and are merged into X_t.
- The individual specific effects μ can be treated as fixed effects (FE), random effects (RE) or correlated random effects (CRE).
- Yang (2018a) present a unified, initial conditions free, *M*-estimation and inference method for the FE-DSPD model,
- Li and Yang (2020b) extend this *M*-estimation and inference strategy to allow for unknown CH in the model, and
- Li and Yang (2020a) present an *M*-estimation and inference method for CRE-DSPD model, including RE-DSPD model as a special case.

A question arises naturally: in practical applications, do we really need such a general and complicated model, or does a simpler model suffice as it gives easier interpretations of the results?

This suggests that before applying this general model, it is helpful to carry out some specification tests to identify a suitable model based on the data.

To be exact, the tests of interest concern the dynamic and spatial parameters $\delta = (\rho, \lambda_1, \lambda_2, \lambda_3)' = (\rho, \lambda')'$. They can be:

- marginal or joint tests: under the null, one or more elements of δ are set to zero and the rest are treated as free parameters;
- or conditional tests: under the null, one or more element of δ are set to zero, given the rest already being set to zero.

This lecture concerns with this general testing problem by focusing on the DSPD model with small T, fixed effects, and possibly unknown CH.

The specific tests of interest are as follows.

Joint test H_0^{PD}: $\delta = 0$, *(regular FE panel data (PD) model suffices)*.

When H_0^{PD} is not rejected, then one proceeds with the regular panel data model with FE and the decision is clear. When H_0^{PD} is rejected, then at least one element of δ is not zero and one does not know the exact cause of rejection and hence it would be necessary to carry out some sub-joint or marginal tests to identify the cause of such a rejection.

Joint test H_0^{DPD}: $\lambda = 0$, *(regular FE dynamic panel data (DPD) model suffices)*.

If H_0^{DPD} is not rejected, then the cause of rejecting H_0^{PD} is due to ρ being non-zero and the FE-DPD model is chosen; otherwise, one needs to proceed with the following test:

Marginal test H_0^{STPD}: $\rho = 0$, *(space-time panel data (STPD) model suffices)*.

If $H_0^{\mathtt{STPD}}$ is not rejected, then the cause of rejecting $H_0^{\mathtt{PD}}$ is that at least one element of λ is not zero. In this case, one may proceed further to identify which element of λ is not zero by carrying out *conditional tests* on one or two elements of λ, given $\rho = 0$.

If $H_0^{\mathtt{STPD}}$ is rejected after $H_0^{\mathtt{DPD}}$ has been rejected, it is clear that at least one element of λ is non-zero when ρ is treated as a free parameter, and the marginal tests on λ_r should be carried out, respectively, for $r = 1, 2, 3$:

Marginal test $H_0^{\mathtt{DSPD1}}$: $\lambda_1 = 0$, *(FE-DSPD model without λ_1 suffices)*.

Marginal test $H_0^{\mathtt{DSPD2}}$: $\lambda_2 = 0$, *(FE-DSPD model without λ_2 suffices)*.

Marginal test $H_0^{\mathtt{DSPD3}}$: $\lambda_3 = 0$, *(FE-DSPD model without λ_3 suffices)*.

Note that the marginal test $H_0^{\mathtt{DSPD3}}$ is quite interesting as the general model (12.1) reduces to a DSPD model with SL and STL effects under the null, which is the model considered by Yu, de Jong and Lee (2008) under large n and large T set-up, allowing fixed individual and time effects.

The marginal test $H_0^{\texttt{DSPD2}}$ is also interesting as the null model becomes a DSPD model with both $\texttt{SL}$ and $\texttt{SE}$ effects, popular in practical applications. Another pair of joint tests of particular interest are,

Joint test $H_0^{\texttt{DSPD4}}$: $\lambda_1 = \lambda_2 = 0$, *(FE-DSPD model with only* $\texttt{SE}$ *effect suffices)*.

Joint test $H_0^{\texttt{DSPD5}}$: $\lambda_2 = \lambda_3 = 0$, *(FE-DSPD model with only* $\texttt{SL}$ *effect suffices)*.

When $H_0^{\texttt{DSPD4}}$ is true, the general model given in (12.1) reduces to a DSPD model with only the $\texttt{SE}$ effect.

- This model is extensively studied by Su and Yang (2015) under large n and small T set-up, with either random or fixed individual effects.
- However, specification test from Model (12.1) to this reduced model has not been considered.
- When $H_0^{\texttt{DSPD5}}$ is true, Model (12.1) reduces to a DSPD model with only the $\texttt{SL}$ effect. This is perhaps the most popular DSPD model among applied researchers. However, a test for its adequacy is not available.

The last test that we would like to highlight is:

Joint test $H_0^{\texttt{SDP}}$: $\rho = \lambda_2 = 0$, *(FE spatial panel data (FE-SPD) model suffices)*.

Under $H_0^{\texttt{SDP}}$, the model reduces to a static spatial panel data model with $\texttt{SL}$ and SE (or $\texttt{SARAR}$) effects. QML estimation and inference for this model were given by Lee and Yu (2010), LM tests for the spatial effects are given by Debarsy and Ertur (2010), and LM-type tests robust against unknown CH are given by Baltagi and Yang (2013b).

Conditional tests might be of interest besides the joint or marginal tests introduced above. By conditional tests we mean tests for certain types of effects, give some other effect(s) are removed from the model.

- For example, given $H_0^{\texttt{DSPD2}}$ is not rejected, i.e., λ_2 is set to zero, one might be interested in testing further whether $\rho = 0$, i.e., whether the static $\texttt{SARAR}$ model suffices;
- given $H_0^{\texttt{STPD}}$ is not rejected, i.e., $\rho = 0$, one might be interested in testing further whether $\lambda_2 = 0$ and if so a static $\texttt{SARAR}$ model suffices.

Despite of interest, methods for testing the above hypotheses do not seem to be available, in particular, when T is small.

There are two related works, GMM gradient tests (Taspinar et al., 2017) and robust LM tests (Bera et al., 2019), but both require a large panel, concern parametric misspecifications, and do not allow for unknown CH.

In contrast, the literature on statistical tests for spatial regression models or static spatial panel data models is much bigger. See, among others,

Anselin et al. (1996), Anselin and Bera (1998), Anselin (2001), Kelejian and Prucha (2001), Yang (2010, 2015, 2018c), Born and Breitung (2011), Baltagi and Yang (2013a,b), Robinson and Rossi (2014, 2015a), Jin and Lee (2015, 2018), Liu and Prucha (2018) for spatial regression models; Baltagi et al. (2003), Baltagi et al. (2007), Debarsy and Ertur (2010), Baltagi and Yang (2013a,b), Robinson and Rossi (2015b), and Xu and Yang (2020) for static panel data models.

This lecture introduces a general and yet simple method, the *adjusted quasi score* (AQS) method, for constructing test statistics for various hypothesis concerning the DSPD models with fixed-effects, small T and possibly heteroskedastic errors.

1. The AQS tests is a score-type test, which is preferred as it requires only the estimation of the null model.
2. The initial constructions of the tests are based on the unified M-estimation method of Yang (2018a):

   first adjusting the conditional quasi score functions given the initial differences to achieve unbiasedness and consistency,

   and then developing a martingale difference representation of the AQS function to give a consistent estimate of the variance-covariance matrix of the AQS functions.

- The resulting AQS tests are shown to have standard asymptotic null behavior and are free from the specifications of the initial conditions.
- Further corrections are made on the concentrated AQS functions, giving a set of standardized AQS (SAQS) tests with much better finite sample properties.
- All the proposed tests are robust against nonnormality.
- Certain tests are fully robust against unknown CH; the others are not.
- For this, alternative modifications are made by following the *M*-estimation strategy of Li and Yang (2020b) to give tests that are fully robust against unknown CH.
- Monte Carlo results show excellent performance of the SAQS tests under homoskedasticity, and the full robustness of the last test against unknown CH.

This lecture presents

1. the AQS tests under homoskedasticity;
2. the standardized AQS tests with better finite sample performance;
3. the CH-robust AQS tests;
4. asymptotic properties of all these tests;
5. Monte Carlo results for the finite sample performance of the tests introduced;
6. empirical applications to illustrate the proposed tests.

The methodology we adopt in constructing tests statistics for testing various hypotheses requires the estimation of the null models. In certain cases, e.g., $H_0^{\mathbb{PD}}$, the null models are very simple, but in other cases they are not as the null models may still contain the dynamic parameter ρ and/or some of the spatial parameters. Also, the construction of the AQS tests requires the AQS function of the full model. Thus, it is necessary to outline the unified *M*-estimation method of Yang (2018a).

First-differencing Model (12.1) to eliminate μ, we have,

$$\Delta y_t = \rho \Delta y_{t-1} + \lambda_1 W_1 \Delta y_t + \lambda_2 W_2 \Delta y_{t-1} + \Delta X_t \beta + \Delta u_t, \quad (12.2)$$
$$\Delta u_t = \lambda_3 W_3 \Delta u_t + \Delta v_t, \mathrm{for} t = 2, 3, \ldots, T.$$

- The time-invariant variables Z are also eliminated;
- The terms corresponding to α_t are merged into X_t, as T is fixed;
- The parameters left in Model (12.2) are $\psi = \{\beta', \sigma_v^2, \rho, \lambda'\}'$.

Note that Δy_1 depends on both the initial observations y_0 and the first period observations y_1. Thus, even if y_0 is exogenous, y_1 and hence Δy_1 is not. Let ψ_0 be the true value of ψ and $\mathrm{E}(\cdot)$ correspond to ψ_0.

Yang's (2018a) *M*-estimation strategy goes as follows:

- formulate the conditional quasi likelihood function **as if** v_t are normal and Δy_1 is exogenous to give the conditional quasi score vector $S(\psi)$,
- then adjust $S(\psi)$ to give the AQS vector $S^*(\psi_0) = S(\psi_0) - \mathrm{E}[S(\psi_0)]$,
- and then estimate ψ by solving the AQS equations $S^*(\psi) = 0$.

Interestingly, this method finds root in Neyman and Scott (1948) on modified likelihood equations.

Chudik and Pesaran (2017) use similar ideas to give a bias-corrected method of moments estimation.

Stacking the vectors and matrices:

- $\Delta Y = \{\Delta y_2', \ldots, \Delta y_T'\}'$, $\Delta Y_{-1} = \{\Delta y_1', \ldots, \Delta y_{T-1}'\}'$,
- $\Delta X = \{\Delta X_2', \ldots, \Delta X_T'\}'$, $\Delta v = \{\Delta v_2', \ldots, \Delta v_T'\}$.

Let $\otimes$ be the Kronecker product and I_m an $m \times m$ identity matrix. Define

- $\mathbf{W}_r = I_{T-1} \otimes W_r, r = 1, 2, 3$,
- $\mathbf{B}_r(\lambda_r) = I_{T-1} \otimes B_r(\lambda_r)$, with $B_r(\lambda_r) = I_n - \lambda_r W_r$, for $r = 1$ and 3;
- $\mathbf{B}_2(\rho, \lambda_2) = I_{T-1} \otimes B_2(\rho, \lambda_2)$, with $B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$.

Denote $B_1 = B_1(\lambda_1)$ and $B_{10} = B_1(\lambda_{10})$; similarly for other quantities.

Assume (*i*) the errors $\{v_{it}\}$ are iid across i and $t > 0$, (*ii*) the regressors $\{X_t\}$ are exogenous with respect to $\{v_{it}\}$, (*iii*) both B_{10}^{-1} and B_{30}^{-1} exist; and (*iv*) the following 'knowledge' about the process in the past:

Assumption A. *Under Model* (12.1)*, (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* $\{v_t, t \geq 1\}$*.*

Yang (2018a) shows: $\mathrm{E}(\Delta Y_{-1}\Delta v') = -\sigma_{v0}^2 \mathbf{D}_{-10}\mathbf{B}_{30}^{-1}$ and $\mathrm{E}(\Delta Y \Delta v') = -\sigma_{v0}^2 \mathbf{D}_0 \mathbf{B}_{30}^{-1}$, where

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

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

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

These immediately lead to $\mathrm{E}[S(\psi_0)]$, and to $S^*(\psi_0) = S(\psi_0) - \mathrm{E}[S(\psi_0)]$, the AQS vector at ψ_0, which takes the form at a general ψ:

$$S^*(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta), \\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta) - \frac{N}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2), \\ \frac{1}{\sigma_v^2}\Delta u(\theta)'(C^{-1}\otimes\mathcal{A})\Delta u(\theta) - (T-1)\mathrm{tr}(G_3), \end{cases} \tag{12.3}$$

where $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$, $\Delta u(\theta) = \mathbf{B}_1(\lambda_1)\Delta Y - \mathbf{B}_2(\rho, \lambda_2)\Delta Y_{-1} - \Delta X\beta$, $G_3 = W_3 B_3^{-1}$, $\mathcal{A} = \frac{1}{2}(W_3' B_3 + B_3' W_3)$, $\Omega = C \otimes (B_3' B_3)^{-1}$, noting $B_3 = B_3(\lambda_3)$, and

$$C = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}_{(T-1)\times(T-1)}$$

The M-estimation

Solving $S^*(\psi) = 0$ leads to the M-estimator $\hat{\psi}_{\rm M}$ of ψ. This root-finding process can be simplified by first solving concentrated AQS equations, $S^*_{\rm c}(\delta) = 0$, with β and σ_v^2 being concentrated out from (12.3), to give the M-estimator $\hat{\delta}_{\rm M}$ of δ, where

$$S^*_{\rm c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_v^2(\delta)} \Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_v^2(\delta)} \Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1), \\ \frac{1}{\hat{\sigma}_v^2(\delta)} \Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2), \\ \frac{1}{\hat{\sigma}_v^2(\delta)} \Delta\hat{u}(\delta)'(C^{-1}\otimes\mathcal{A})\Delta\hat{u}(\delta) - (T-1)\mathrm{tr}(G_3), \end{cases} \tag{12.4}$$

$\Delta\hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1, \lambda_2)$, $\hat{\beta}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})$, and $\hat{\sigma}_v^2(\delta) = \frac{1}{N}\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta)$, where $N = n(T-1)$. The M-estimators of β and σ_v^2 are thus $\hat{\beta}_{\rm M} \equiv \hat{\beta}(\hat{\delta}_{\rm M})$ and $\hat{\sigma}^2_{v,\rm M} \equiv \hat{\sigma}_v^2(\hat{\delta}_{\rm M})$.

Yang (2018a) show that under regularity conditions the M-estimator $\hat{\psi}_{\rm M} = (\hat{\beta}'_{\rm M}, \hat{\sigma}^2_{v,\rm M}, \hat{\delta}'_{\rm M})'$ is $\sqrt{N}$-consistent and asymptotically normal.

The M-estimators under the constraints imposed by various hypotheses will remain to be $\sqrt{N}$-consistent and asymptotically normal.

It is important to note that the adjustments (non-stochastic terms in (12.3)) are free from the initial conditions, and hence the resulting AQS function and the *M*-estimators are free from the initial conditions.

The AQS functions given in (12.3) are the key elements in the construction of the AQS tests. In this section, we first formulate the AQS test in a unified manner, and then present some details for the tests defined in Sec. 12.1.

- The construction of the joint and marginal AQS tests depends critically on the availability of the variance-covariance (VC) matrix, $\Gamma^*(\psi_0) = \frac{1}{N}\text{Var}[S^*(\psi_0)]$.
- The dynamic nature of Model (12.1) makes such an estimation very difficult, as the expression of $\Gamma^*(\psi_0)$ encounters the initial values problem.

To overcome this difficulty, Yang (2018a) proposed to decompose the AQS function into a sum of martingale difference (MD) sequences so that the *outer-product-of-martingale-differences* (OPMD) consistently estimates $\Gamma^*(\psi_0)$, free from the initial conditions!

Notation. Let diag(A) form a diagonal matrix by the diagonal elements of a square matrix A and blkdiag(A_k) form a block-diagonal matrix by matrices $\{A_k\}$. The subscript '$_n$' is often dropped shall no confusion arise.

Yang (2018a) developed the representations: $\Delta Y = \mathbb{R}\,\Delta \mathbf{y}_1 + \boldsymbol{\eta} + \mathbb{S}\Delta v$ and $\Delta Y_{-1} = \mathbb{R}_{-1}\Delta \mathbf{y}_1 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\Delta v$, leading to the expression:

$$S^*(\psi_0) = \begin{cases} \Pi_1'\Delta v, \\ \Delta v'\Phi_1\Delta v - \frac{N}{2\sigma_{v0}^2}, \\ \Delta v'\Psi_1\Delta \mathbf{y}_1 + \Delta v'\Pi_2 + \Delta v'\Phi_2\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}), \\ \Delta v'\Psi_2\Delta \mathbf{y}_1 + \Delta v'\Pi_3 + \Delta v'\Phi_3\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_0\mathbf{W}_1), \\ \Delta v'\Psi_3\Delta \mathbf{y}_1 + \Delta v'\Pi_4 + \Delta v'\Phi_4\Delta v + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-10}\mathbf{W}_2), \\ \Delta v'\Phi_5\Delta v - (T-1)\mathrm{tr}(G_{30}), \end{cases} \tag{12.5}$$

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

Furthermore, $\Delta\mathbf{y}_1 = 1_{T-1} \otimes \Delta y_1$, $\mathbb{R} = \mathtt{blkdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^{T-1})$, $\mathbb{R}_{-1} = \mathtt{blkdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-2})$, $\boldsymbol{\eta} = \mathbb{B}\mathbf{B}_{10}^{-1}\Delta X\beta_0$, $\boldsymbol{\eta}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\Delta X\beta_0$, $\mathbb{S} = \mathbb{B}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$, $\mathbb{S}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$, and

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

The expression (12.5) is the key to the proof of asymptotic normality of $\frac{1}{\sqrt{N}}S^*(\psi_0)$, and to the development of OPMD estimate of the VC matrix of $S^*(\psi_0)$, so that an AQS test can be constructed.

Note that $S^*(\psi_0)$ contains three types of stochastic elements:

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

where Π, Φ and Ψ are nonstochastic matrices (depending on ψ_0) with Π being $N \times p$ or $N \times 1$, and Φ and Ψ being $N \times N$.

As noted in Yang (2018a), the closed form expressions for variances of $\Pi'\Delta v$ and $\Delta v'\Phi\Delta v$, and their covariance can readily be derived, but the closed-form expressions for the variance of $\Delta v'\Psi\Delta\mathbf{y}_1$ and its covariances with $\Pi'\Delta v$ and $\Delta v'\Phi\Delta v$ depend on the knowledge of the distribution of Δy_1, which is unavailable.

Yang (2018a) went on to give a unified method of estimating the VC matrix of AQS function, the OPMD estimate, which is summarized as follows.

- For a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A = A^u + A^l + A^d$.
- Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ, partitioned according to $t, s = 2, \ldots, T$.
- Define $\Psi_{t+} = \sum_{s=2}^{T} \Psi_{ts}$, $\Theta = \Psi_{2+}(B_{30}B_{10})^{-1}$, $\Delta y_1^{\circ} = B_{30}B_{10}\Delta y_1$, and $\Delta y_{1t}^{*} = \Psi_{t+}\Delta y_1$.

Define

$$g_{1i} = \sum_{t=2}^{T} \Pi_{it}' \Delta v_{it}, \quad (12.6)$$

$$g_{2i} = \sum_{t=2}^{T} (\Delta v_{it} \Delta \xi_{it} + \Delta v_{it} \Delta v_{it}^{*} - \sigma_{v0}^{2} d_{it}), \quad (12.7)$$

$$g_{3i} = \Delta v_{2i} \Delta \zeta_i + \Theta_{ii} (\Delta v_{2i} \Delta y_{1i}^{\circ} + \sigma_{v0}^{2}) + \sum_{t=3}^{T} \Delta v_{it} \Delta y_{1it}^{*}, \quad (12.8)$$

where for (12.7), $\xi_t = \sum_{s=2}^{T} (\Phi_{st}^{u\prime} + \Phi_{ts}^{l}) \Delta v_s$, $\Delta v_t^{*} = \sum_{s=2}^{T} \Phi_{ts}^{d} \Delta v_s$, and $\{d_{it}\}$ are the diagonal elements of $\mathbf{C}\Phi$; for (12.8), $\{\Delta \zeta_i\} = \Delta \zeta = (\Theta^{u} + \Theta^{l}) \Delta y_1^{\circ}$, and $\texttt{diag}\{\Theta_{ii}\} = \Theta^{d}$. Then,

$$\Pi' \Delta v = \sum_{i=1}^{n} g_{1i}, \quad (12.9)$$

$$\Delta v' \Phi \Delta v - \mathrm{E}(\Delta v' \Phi \Delta v) = \sum_{i=1}^{n} g_{2i}, \quad (12.10)$$

$$\Delta v' \Psi \Delta \mathbf{y}_1 - \mathrm{E}(\Delta v' \Psi \Delta \mathbf{y}_1) = \sum_{i=1}^{n} g_{3i}, \quad (12.11)$$

and $\{(g_{1i}', \ g_{2i}, \ g_{3i})', \mathcal{F}_{n,i}\}_{i=1}^{n}$ form an MD sequence, where $\mathcal{F}_{n,i} = \mathcal{F}_{n,0} \otimes \mathcal{G}_{n,i}$, a product σ-field, with $\mathcal{F}_{n,0}$ being the σ-field generated by $(v_0, \Delta y_0)$, and $\{\mathcal{G}_{n,i}\}$ an increasing sequence of σ-fields generated by

$(v_{j1}, \ldots, v_{jT}, j = 1, \ldots, i), i = 1, \ldots, n.$

Now, following the above results,

- for each Π_r, $r = 1, 2, 3, 4$, defined in (12.5), define g_{1ri} according to (12.6);
- for each Φ_r, $r = 1, \ldots, 5$, defined in (12.5), define g_{2ri} according to (12.7); and
- for each Ψ_r, $r = 1, 2, 3$, defined in (12.5), define g_{3ri} according to (12.8).

Define

$$\mathbf{g}_i = (g'_{11i},\ g_{21i},\ g_{31i} + g_{12i} + g_{22i},\ g_{32i} + g_{13i} + g_{23i},\ g_{33i} + g_{14i} + g_{24i},\ g_{25i})'. \quad (12.12)$$

Then, $S^*(\psi_0) = \sum_{i=1}^n \mathbf{g}_i$, where $\{\mathbf{g}_i, \mathcal{F}_{n,i}\}$ form a vector MD sequence. It follows that $\Gamma^*(\psi_0) = \mathrm{Var}[S^*(\psi_0)] = \sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}'_i)$, and therefore its sample analogue,

$$\widehat{\Gamma}^* = \sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i, \quad (12.13)$$

gives a consistent OPMD estimator of $\Gamma^*(\psi_0)$, i.e.,

$$\mathrm{plim}_{n\to\infty} \frac{1}{N} \sum_{i=1}^n [\hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i - \Gamma^*(\psi_0)] = 0,$$

where $\hat{\mathbf{g}}_i$ is obtained by replacing ψ_0 in $\mathbf{g}_i$ by $\hat{\psi}_{\mathrm{M}}$ and Δv by its observed counterpart $\widehat{\Delta} v$, noting that Δy_1 is observed.

To construct the AQS tests, the estimates $(\hat{\psi}_{\mathrm{M}}, \widehat{\Delta} v)$ of the full model are replaced by the constrained estimates at the null, $(\tilde{\psi}_{\mathrm{M}}, \widetilde{\Delta} v)$.

This allows us to develop the AQS test in a unified manner:

- let $\delta = (\pi', \varphi')'$ and the null hypothesis specifies $\varphi = 0$.
- Let $\vartheta = (\beta', \sigma^2, \pi')'$ and therefore $\psi = (\vartheta', \varphi')'$.
- Let $\Sigma^*(\psi_0) = -\mathrm{E}[\frac{\partial}{\partial \psi'} S^*(\psi_0)]$.
- Partition $\Sigma^*(\psi)$ and $\Gamma^*(\psi)$ according to ϑ and φ, and denote their submatrices by $\Sigma^*_{ab}(\psi)$ and $\Gamma^*_{ab}(\psi)$, $a = \vartheta, \varphi$, $b = \vartheta, \varphi$.
- Let $S^*(\psi) = (S^{*\prime}_{\vartheta}(\psi), S^{*\prime}_{\varphi}(\psi))'$ and $\mathbf{g}_i = (\mathbf{g}'_{i,\vartheta}, \mathbf{g}'_{i,\varphi})'$.

Clearly, the construction of the test of $\varphi = 0$ depends on $S^*_{\varphi}(\tilde{\vartheta}, 0)$ and its variance, where $\tilde{\vartheta}$ is the null estimate of ϑ.

Under mild conditions, a Taylor expansion leads to the following asymptotic MD representation:

$$\begin{aligned}\tfrac{1}{\sqrt{N}}S^*_{\varphi}(\tilde{\vartheta},0_k) &= \tfrac{1}{\sqrt{N}}S^*_{\varphi}(\vartheta_0,0_k) - \tfrac{1}{\sqrt{N}}\Sigma^*_{\varphi\vartheta}\Sigma^{*-1}_{\vartheta\vartheta}S^*_{\vartheta}(\vartheta_0,0_k) + o_p(1)\\ &= \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^n(\mathbf{g}_{i,\varphi} - \Lambda\mathbf{g}_{i,\vartheta}) + o_p(1), \end{aligned} \quad (12.14)$$

where $\Lambda = \Sigma^*_{\varphi\vartheta}\Sigma^{*-1}_{\vartheta\vartheta}$, and $k = \dim(\varphi)$. Clearly $\{\mathbf{g}_{i,\varphi} - \Lambda\mathbf{g}_{i,\vartheta}\}$ form a vector MD sequence with respect to $\mathcal{F}_{n,i}$. Therefore,

$$\mathrm{Var}[\tfrac{1}{\sqrt{N}}S^*_{\varphi}(\tilde{\vartheta},0_k)] = \tfrac{1}{N}\textstyle\sum_{i=1}^n[(\mathbf{g}_{i,\varphi} - \Lambda\mathbf{g}_{i,\vartheta})(\mathbf{g}_{i,\varphi} - \Lambda\mathbf{g}_{i,\vartheta})'] + o(1). \quad (12.15)$$

An AQS-based test for testing the hypothesis $H_0 : \varphi = 0$ is thus,

$$T_{\mathtt{M}} = S^{*\prime}_{\varphi}(\tilde{\vartheta},0_k)\big\{\textstyle\sum_{i=1}^n(\tilde{\mathbf{g}}_{i,\varphi} - \tilde{\Lambda}\tilde{\mathbf{g}}_{i,\vartheta})(\tilde{\mathbf{g}}_{i,\varphi} - \tilde{\Lambda}\tilde{\mathbf{g}}_{i,\vartheta})'\big\}^{-1}S^*_{\varphi}(\tilde{\vartheta},0_k), \quad (12.16)$$

where M = PD, DPD, DSPD1, $\cdots$, DSPD5, and SPD, associated with the null hypotheses defined in Sec. 1, $\Lambda = \Sigma^*_{\varphi\vartheta}\Sigma^{*-1}_{\vartheta\vartheta}$, and all 'tilde' quantities are the estimates of the corresponding quantities at the null.

The asymptotic distribution of $T^{\mathtt{M}}_{\mathtt{AQS}}$, i.e., χ^2_k, can be proved under some regularity conditions generic to all tests, and some additional regularity conditions specific for a given test. The generic conditions are:

Assumption B: *The idiosyncratic errors* $\{v_{it}\}$ *are independent across* $i = 1, \ldots, n$ *and* $t = 0, 1, \ldots, T$*, with* $E(v_{it}) = 0$*,* $Var(v_{it}) = \sigma^2_{v0}$*, and* $E|v_{it}|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$*.*

When homoskedasticity is in question, Assumption B is relaxed to:

Assumption B*: *The idiosyncratic errors* $\{v_{it}\}$ *are independent across* $i = 1, \ldots, n$ *and* $t = 0, 1, \ldots, T$*, with* $E(v_{it}) = 0$*,* $Var(v_{it}) = \sigma^2_{v0} h_{ni}$ *such that* $0 < h_{ni} \leq c < \infty$ *and* $\frac{1}{n}\sum_{i=1}^{n} h_{ni} = 1$*, and* $E|v_{it}|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$*.*

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

Assumption D: *(i) For* $r = 1, 2, 3$*, the elements* $w_{r,ij}$ *of* W_r *are at most of order* ι_n^{-1}*, uniformly in all* i *and* j*, and* $w_{r,ii} = 0$ *for all* i*; (ii)* $\iota_n / n \to 0$ *as* $n \to \infty$*; (iii)* $\{W_r, r = 1, 2, 3\}$ *are uniformly bounded in both row and column sums.*

Assumption D allows the degree of spatial dependence, e.g., the number of neighbors each spatial unit has, to grow with the sample size but in a lower speed.

- As a result, the convergence rate of the estimators of certain parameters, e.g., the spatial error parameter, may need to be adjusted down to $\sqrt{N/\iota_n}$.
- See Lee (2004), Liu and Yang (2015), Su and Yang (2015), and Yang (2018a) for more details.
- However, this feature is not explicitly reflected in the subsequent developments as the implementations of the tests do not require ι.

Additional conditions on the initial differences are necessary when the null model contains the dynamic term, and additional conditions on B_1 and B_3 are necessary when the null model contains λ_1 and λ_3 terms.

Assumption E: *For Φ, $n \times n$, uniformly bounded in either row or column sums with elements of uniform order ι_n^{-1}, and ϕ, $n \times 1$, with elements of uniform order $\iota_n^{-1/2}$, (i) $\frac{\iota_n}{n}\Delta y_1' \Phi \Delta y_1 = O_p(1)$, $\frac{\iota_n}{n}\Delta y_1' \Phi \Delta v_2 = O_p(1)$; (ii) $\frac{\iota_n}{n}[\Delta y_1 - \mathrm{E}(\Delta y_1)]'\phi = o_p(1)$; (iii) $\frac{\iota_n}{n}[\Delta y_1' \Phi \Delta y_1 - \mathrm{E}(\Delta y_1' \Phi \Delta y_1)] = o_p(1)$; (iv) $\frac{\iota_n}{n}[\Delta y_1' \Phi \Delta v_2 - \mathrm{E}(\Delta y_1' \Phi \Delta v_2)] = o_p(1)$.*

Assumption F: *B_1^{-1} and B_2^{-1} exist, and are uniformly bounded in both row and column sums in absolute value, for (λ_1, λ_3) in a neighborhood of $(\lambda_{10}, \lambda_{30})$.*

Theorem 12.1

Under Assumptions A-F, if $\tilde{\vartheta}$ is $\sqrt{N}$-consistent, we have under $H_0^{\mathbb{M}}$,

$$T_{\mathbb{M}} \xrightarrow{D} \chi_k^2, \quad as\ n \to \infty,$$

where $\mathbb{M}$ denotes a null model specified in Sec. 12.1.

Note that in a special case where $\Gamma^* \approx \Sigma^*$ at the null, i.e., the information matrix equality (IME) holds (asymptotically), the AQS test is asymptotically equivalent to $T_{\mathbb{M},0} = S^{*\prime}(\tilde{\psi})(\sum_{i=1}^n \tilde{\mathbf{g}}_i \tilde{\mathbf{g}}_i')^{-1} S^*(\tilde{\psi})$, where $\tilde{\psi} = (\tilde{\vartheta}', 0_k')'$. The cases under which the above can be true are those with the null model being a static panel data model (i.e., $\rho = \lambda_2 = 0$) and the errors are Gaussian.

To facilitate practical applications of the AQS tests, we present details for each of the hypothesis postulated in Sec. 12.1 so that a specific test can directly be applied without going through the complicated general theory.

More interestingly, we show that certain tests are valid under Assumption B*, i.e., robust against unknown cross-sectional heteroskedasticity (CH).

The first is the **Joint test** H_0^{PD}**:** $\delta = 0$. Under H_0^{PD}, the model DSPD(δ) is reduced to the simplest PD model, and the estimation of the model at the null is simply the ordinary least squares (OLS) estimation, i.e.,

$$\tilde{\beta} = (\Delta X' \mathbf{C}^{-1} \Delta X)^{-1} \Delta X' \mathbf{C}^{-1} \Delta Y \text{ and } \tilde{\sigma}_v^2 = \frac{1}{N} \Delta \tilde{v}' \mathbf{C}^{-1} \Delta \tilde{v},$$

where $\Delta \tilde{v} = \Delta Y - \Delta X \tilde{\beta}$, leading to $\tilde{\psi} = (\tilde{\beta}', \tilde{\sigma}_v^2, 0_4')'$.

Under H_0^{PD}, $B_1 = B_3 = I_n$, and $B_2 = \mathbf{0}_n$ where $\mathbf{0}_n$ denotes an $n \times n$ matrix of zeros.

It is easy to see that $\mathrm{E}[S^*(\psi_0)|_{H_0^{\text{PD}}}] = 0$ and that $\tilde{\beta}$ and $\tilde{\sigma}_v^2$ are robust against unknown CH, implying the AQS test of H_0^{PD}: $\delta = 0$ is robust against unknown CH!

Corollary 12.1

Under Assumptions A, B*, C and D, $T_{\mathrm{PD}}|_{H_0^{\mathrm{PD}}} \xrightarrow{D} \chi_4^2$, *as* $n \to \infty$.

The very attractive feature of this joint test is that it is robust against unknown CH as specified in Assumption B*, besides being robust against nonnormality of the idiosyncratic errors v_{it}.

The same goes to the conditional tests where under the null and the given 'condition' the model becomes a pure panel data model.

The second is the **joint test** $H_0^{\rm DPD}$**:** $\lambda = 0$. Under $H_0^{\rm DPD}$, $B_1 = B_3 = I_n$, and $B_2 = \rho I_n$. The constrained M-estimators of β and σ_v^2, given ρ, are

$$\tilde{\beta}(\rho) = (\Delta X' \mathbf{C}^{-1} \Delta X)^{-1} \Delta X' \mathbf{C}^{-1} (\Delta Y - \rho \Delta Y_{-1}) \text{ and } \tilde{\sigma}_v^2(\rho) = \tfrac{1}{N} \Delta \tilde{v}'(\rho) \mathbf{C}^{-1} \Delta \tilde{v}(\rho),$$

where $\Delta \tilde{v}(\rho) = \Delta Y - \rho \Delta Y_{-1} - \Delta X \tilde{\beta}(\rho)$.

The constrained M-estimator of ρ under $H_0^{\rm DPD}$ is

$$\tilde{\rho} = \arg\left\{ \tfrac{1}{\tilde{\sigma}_v^2(\rho)} \Delta \tilde{v}'(\rho) \mathbf{C}^{-1} \Delta Y_{-1} + n\left(\tfrac{1}{1-\rho} - \tfrac{1-\rho^T}{T(1-\rho)^2}\right) = 0 \right\}, \tag{12.17}$$

leading to the constrained M estimators of β and σ_v^2 as $\tilde{\beta} = \tilde{\beta}(\tilde{\rho})$ and $\tilde{\sigma}_v^2 = \tilde{\sigma}_v^2(\tilde{\rho})$.

The constrained M-estimator of ϑ is thus $\tilde{\vartheta} = (\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\rho})'$.

- The concentrated AQS function for ρ contained in (12.17) clearly shows that the M-estimator is not only consistent when T is fixed but also eliminates the bias of order $O(T^{-1})$.
- In contrast, the estimator based on the unadjusted score is inconsistent when T is fixed and has a bias of order $O(T^{-1})$ when T grows with n.
- See Hahn and Kuersteiner (2002), and Yang (2018a,b) for more discussions.

Joint AQS test of $H_0^{\rm DPD}$: $\lambda = 0$, cont'd

The following lemma shows that the restricted M-estimator $\tilde{\rho}$ defined in (12.17) is robust against unknown CH.

Lemma 12.1

Under Assumptions A, B, and C-E, if ρ_0 is in the interior of a compact parameter space, then for the DPD model, we have, as $n \to \infty$,*

$$\tilde{\vartheta} = (\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\rho})' \xrightarrow{p} \vartheta_0 \text{ and } \sqrt{N}(\tilde{\vartheta} - \vartheta_0) \xrightarrow{D} N(0, \Psi),$$

for a suitably defined Ψ.

Corollary 12.2

Under the assumptions of Lemma 12.1, $T_{\rm DPD}|_{H_0^{\rm DPD}} \xrightarrow{D} \chi_3^2$, as $n \to \infty$.

This gives another interesting result: $T_{\rm DPD}$ is robust against nonnormality and unknown CH, applicable to all tests with DPD model as null.

Under the null H_0^{STPD}: $\rho = 0$, $B_2 = \lambda_2 W_2$. The constrained M-estimator $\tilde{\lambda}$ of λ solves the following estimating equations:

$$\begin{cases} \frac{1}{\tilde{\sigma}_v^2(\lambda)} \Delta\tilde{u}(\lambda)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1) = 0, \\ \frac{1}{\tilde{\sigma}_v^2(\lambda)} \Delta\tilde{u}(\lambda)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2) = 0, \\ \frac{1}{\tilde{\sigma}_v^2(\lambda)} \Delta\tilde{u}(\lambda)'(C^{-1} \otimes \mathcal{A})\Delta\tilde{u}(\lambda) - (T-1)\mathrm{tr}(G_3) = 0, \end{cases}$$

where $\Delta\tilde{u}(\lambda) = \mathbf{B}_1\Delta Y - \lambda_2\mathbf{W}_2\Delta Y_{-1} - \Delta X\tilde{\beta}(\lambda)$, and $\tilde{\beta}(\lambda)$ and $\tilde{\sigma}_v^2(\lambda)$ are those given below (12.4) by setting $\rho = 0$.

Let $\tilde{\beta} = \tilde{\beta}(\tilde{\lambda})$, $\tilde{\sigma}_v^2 = \tilde{\sigma}_v^2(\tilde{\lambda})$, and $\tilde{\vartheta} = \{\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\lambda}'\}'$. Based on the result of Li and Yang (2020b), it is easy to see that $\mathrm{plim}_{n\to\infty} \frac{1}{n} S^*(\psi_0)|_{\rho=0} \neq 0$ under unknown CH. Therefore $\tilde{\vartheta}$ cannot be consistent under unknown CH and T_{STPD} is generally not robust against unknown CH. Sec. 12.4 presents a CH-robust version of this test.

With $r = 1$, or 2 or 3, we have three marginal tests corresponding one specific type of spatial effects. Among these three marginal tests, the test of $H_0^{\texttt{DSPD2}}$: $\lambda_2 = 0$ is the most interesting one as under $H_0^{\texttt{DSPD2}}$ the model is reduced to the popular DSPD model with $\texttt{SL}$ and $\texttt{SE}$ effects. We consider only this case as the others can be handled in the similar manner. Under $H_0^{\texttt{DSPD2}}$, $B_2 = \rho I_n$. The constrained M-estimators $(\tilde\rho, \tilde\lambda_1, \tilde\lambda_3)$ of $(\rho, \lambda_1, \lambda_3)$ solve the following estimating equations:

$$\begin{cases} \frac{1}{\tilde\sigma_v^2(\rho,\lambda_1,\lambda_3)}\Delta\tilde u(\rho,\lambda_1,\lambda_3)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}) = 0, \\ \frac{1}{\tilde\sigma_v^2(\rho,\lambda_1,\lambda_3)}\Delta\tilde u(\rho,\lambda_1,\lambda_3)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1) = 0, \\ \frac{1}{\tilde\sigma_v^2(\rho,\lambda_1,\lambda_3)}\Delta\tilde u(\rho,\lambda_1,\lambda_3)'(C^{-1}\otimes\mathcal{A})\Delta\tilde u(\rho,\lambda_1,\lambda_3) - (T-1)\mathrm{tr}(G_3) = 0, \end{cases}$$

where $\Delta\tilde u(\rho,\lambda_1,\lambda_3) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\tilde\beta(\rho,\lambda_1,\lambda_3)$, and $\tilde\beta(\rho,\lambda_1,\lambda_3)$ and $\tilde\sigma_v^2(\rho,\lambda_1,\lambda_3)$ are those given below (12.4) by setting $\lambda_2 = 0$.

Let $\tilde\beta = \tilde\beta(\tilde\rho,\tilde\lambda_1,\tilde\lambda_3)$, $\tilde\sigma_v^2 = \tilde\sigma_v^2(\tilde\rho,\tilde\lambda_1,\tilde\lambda_3)$, and $\tilde\psi = \{\tilde\beta', \tilde\sigma_v^2, \tilde\rho, \tilde\lambda_1, 0, \tilde\lambda_3\}'$. We obtain the AQS test statistic $T_{\texttt{SPDD2}}$ from (12.16).

This is an interesting test as under the null the model reduces to a popular DSPD model with spatial error only, which was studied by Su and Yang (2015) under fixed T with initial observations being modeled. In this case, $B_1 = I_n$ and $B_2 = \rho I_n$, and the constrained M-estimators $\tilde{\rho}$ and $\tilde{\lambda}_3$ solve:

$$\begin{cases} \frac{1}{\tilde{\sigma}_v^2(\rho,\lambda_3)} \Delta\tilde{u}(\rho,\lambda_3)' \Omega^{-1} \Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}) = 0, \\ \frac{1}{\tilde{\sigma}_v^2(\rho,\lambda_3)} \Delta\tilde{u}(\rho,\lambda_3)' (C^{-1} \otimes \mathcal{A}) \Delta\tilde{u}(\rho,\lambda_3) - (T-1)\mathrm{tr}(G_3) = 0, \end{cases}$$

where $\Delta\tilde{u}(\rho,\lambda_3) = \Delta Y - \rho\Delta Y_{-1} - \Delta X \tilde{\beta}(\rho,\lambda_1,\lambda_3)$, and $\tilde{\beta}(\rho,\lambda_3)$ and $\tilde{\sigma}_v^2(\rho,\lambda_3)$ are those given below (12.4) by setting $\lambda_1 = \lambda_2 = 0$.

Let $\tilde{\beta} = \tilde{\beta}(\tilde{\rho},\tilde{\lambda}_3)$, $\tilde{\sigma}_v^2 = \tilde{\sigma}_v^2(\tilde{\rho},\tilde{\lambda}_3)$, and $\tilde{\psi} = \{\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\rho}, 0, 0, \tilde{\lambda}_3\}'$. We obtain from (12.16) the AQS test $T_{\mathtt{DSPD4}}$ for testing $H_0^{\mathtt{DSPD4}}$.

Under the null hypothesis, the model reduces to another popular model, the DSPD model with only the spatial lag effect. In this case, $B_2 = \rho I_n$ and $B_3 = I_n$, and the constrained M-estimators $\tilde{\rho}$ and $\tilde{\lambda}_1$ solve:

$$\begin{cases} \frac{1}{\tilde{\sigma}_v^2(\rho,\lambda_1)} \Delta\tilde{v}(\rho,\lambda_1)'\Omega^{-1}\Delta Y_{-1} + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}) = 0, \\ \frac{1}{\tilde{\sigma}_v^2(\rho,\lambda_1)} \Delta\tilde{v}(\rho,\lambda_1)'\Omega^{-1}\mathbf{W}_1\Delta Y + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1) = 0, \end{cases}$$

where $\Delta\tilde{v}(\rho,\lambda_1) = \mathbf{B}_1\Delta Y - \rho\Delta Y_{-1} - \Delta X\tilde{\beta}(\rho,\lambda_1)$, and $\tilde{\beta}(\rho,\lambda_1)$ and $\tilde{\sigma}_v^2(\rho,\lambda_1)$ are those given below (12.4) by setting $\lambda_2 = \lambda_3 = 0$.

Let $\tilde{\beta} = \tilde{\beta}(\tilde{\rho},\tilde{\lambda}_1)$, $\tilde{\sigma}_v^2 = \tilde{\sigma}_v^2(\tilde{\rho},\tilde{\lambda}_1)$, and $\tilde{\psi} = \{\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\rho}, \tilde{\lambda}_1, 0, 0\}'$. We obtain from (12.16) the AQS test $T_{\texttt{DSPD5}}$ for testing $H_0^{\texttt{DSPD5}}$.

Under the null, $B_2 = 0$ and $\mathbf{D} = -\mathbf{CB}_1^{-1}$, and the model becomes the static SARAR model. The constrained M-estimators $\tilde{\lambda}_1$ and $\tilde{\lambda}_3$ of λ_1 and λ_3 solve the following estimating equations (see also Lee and Yu (2010)):

$$\begin{cases} \frac{1}{\tilde{\sigma}_v^2(\lambda_1,\lambda_3)} \Delta\tilde{u}(\lambda_1,\lambda_3)'\Omega^{-1}\mathbf{W}_1\Delta Y - (T-1)\mathrm{tr}(B_1^{-1}W_1) = 0, \\ \frac{1}{\tilde{\sigma}_v^2(\lambda_1,\lambda_3)} \Delta\tilde{u}(\lambda_1,\lambda_3)'(C^{-1} \otimes \mathcal{A})\Delta\tilde{u}(\lambda_1,\lambda_3) - (T-1)\mathrm{tr}(G_3) = 0, \end{cases}$$

where $\Delta\tilde{u}(\lambda_1,\lambda_3) = \mathbf{B}_1\Delta Y - \Delta X\tilde{\beta}(\lambda_1,\lambda_3)$, and $\tilde{\beta}(\lambda_1,\lambda_3)$ and $\tilde{\sigma}_v^2(\lambda_1,\lambda_3)$ are those given below (12.4) by setting $\rho = \lambda_2 = 0$.

Let $\tilde{\beta} = \tilde{\beta}(\tilde{\lambda}_1,\tilde{\lambda}_3)$, $\tilde{\sigma}_v^2 = \tilde{\sigma}_v^2(\tilde{\lambda}_1,\tilde{\lambda}_3)$, and $\tilde{\psi} = \{\tilde{\beta}', \tilde{\sigma}_v^2, 0, \tilde{\lambda}_1, 0, \tilde{\lambda}_3\}'$. We obtain from (12.16) the AQS test T_{SPD} for testing H_0^{SPD}.

Conditional tests are those for testing whether the model can be further reduced, given that it has already been reduced. For example,

- $H_0^{\mathtt{PD1}}$: $\lambda_1 = 0$, given $\lambda_2 = \lambda_3 = 0$;
- $H_0^{\mathtt{PD3}}$: $\lambda_3 = 0$, given $\lambda_1 = \lambda_2 = 0$;
- $H_0^{\mathtt{SPD}}$: $\rho = 0$, given $\lambda_2 = 0$.

The last conditional test says that based on the model without λ_2, we want to see further if $\rho = 0$, i.e., a regular SPD model suffices.

The conditional tests conditional upon $\rho = \lambda_2 = 0$ are the tests of model reduction for the regular SPD model, and the LM-type of tests are given by Debarsy and Erther (2010) and Baltagi and Yang (2013a) for models with homoskedastic models, and Born and Breitung (2011) and Baltagi and Yang (2013b) for models with heteroskedastic errors.

All these conditional tests can be easily developed based on the general methodology presented above.

Conditional AQS tests, cont'd

Some conditional tests are robust against unknown CH in light of Corollaries 12.1 and 12.2, and some can be made to be robust against unknown CH in light of Baltagi and Yang (2013b).

Given the fact that the OPMD estimator of the VC matrix of AQS functions are robust against unknown CH, any AQS or SAQS test can be made to be CH-robust, provided the AQS function is made so. A general CH-robut method is given in Sec. 12.4.

All the tests developed above can be implemented in a unified manner based on the general expressions of the AQS function given in (12.3) or (12.5), and the general OPMD estimate of its VC matrix given in (12.13). $\tilde{\Sigma}^*$ can be $\Sigma^*(\tilde{\psi})$ or $-\frac{\partial}{\partial \psi} S^*(\psi)|_{\psi=\tilde{\psi}_{\text{M}}}$.

For each specific test, all it is necessary is to change the definitions of the matrices $B_r, r = 1, 2, 3$ according to the null hypothesis, and modify the user-supplied function that does root-finding. Empirical applications with Matlab are presented in Sec. 12.5.

The joint and marginal AQS tests presented above are simple but may not be satisfactory when n is not large enough. The reason is that the variability from the estimation of β and σ_v^2 are not taken into account when constructing the test statistics.

It is thus desirable to find simple ways to improve the finite sample performance of these tests. Clearly, after β_0 and σ_v^2 being replaced by $\hat{\beta}(\delta_0)$ and $\hat{\sigma}_v(\delta_0)$ in the last four components of $S^*(\psi_0)$ given in (12.3), the concentrated AQS functions no longer have mean zero, although they do asymptotically.

Furthermore, the variance of the concentrated AQS functions may also be affected. Thus, re-adjustments on the mean and variance may help improving the finite sample performance of the AQS tests (see Baltagi and Yang 2013a,b).

Rewrite the numerator, $\hat{\sigma}_v^2(\delta)S_{\rm C}^*(\delta)$, of the concentrated AQS function in (12.4) as

$$S_{\rm C,N}^*(\delta) = \begin{cases} \Delta\hat{u}(\delta)'\Omega^{-1}\Delta Y_{-1} + \phi_1\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), \\ \Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \phi_2\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), \\ \Delta\hat{u}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \phi_3\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), \\ \Delta\hat{u}(\delta)'(C^{-1}\otimes\mathcal{A})\Delta\hat{u}(\delta) - \phi_4\Delta\hat{u}(\delta)'\Omega^{-1}\Delta\hat{u}(\delta), \end{cases} \quad (12.18)$$

where $\phi_1 = \frac{1}{N}\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1})$, $\phi_2 = \frac{1}{N}\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}\mathbf{W}_1)$, $\phi_3 = \frac{1}{N}\mathrm{tr}(\mathbf{C}^{-1}\mathbf{D}_{-1}\mathbf{W}_2)$ and $\phi_4 = \frac{1}{n}\mathrm{tr}(G_3)$.

The ideas are:
finding the mean of $S_{\rm C,N}^*(\delta_0)$ and recentering, and then finding the variance estimate of the recentered $S_{\rm C,N}^*(\delta_0)$ and restandardizing.

Letting $\Omega^{\frac{1}{2}}$ be the symmetric square root matrix of Ω, and $\Delta X^* = \Omega^{-\frac{1}{2}}\Delta X$, we have

$$\Omega^{-\frac{1}{2}}\Delta\hat{u}(\delta) = \mathbf{M}\Omega^{-\frac{1}{2}}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1}),$$

where $\mathbf{M} = I_N - \Delta X^*(\Delta X^{*\prime}\Delta X^*)^{-1}\Delta X^{*\prime}$ is a projection matrix.

Noting that $\mathbf{M}\Delta X^* = 0$, and that at the true δ_0,

$$\Omega_0^{-\frac{1}{2}}(\mathbf{B}_{10}\Delta Y - \mathbf{B}_{20}\Delta Y_{-1}) = \Delta X^*\beta_0 + \Omega_0^{-\frac{1}{2}}\mathbf{B}_{30}^{-1}\Delta v,$$

we obtain

$$S^*_{\mathrm{C,N}}(\delta_0) = \begin{cases} \Delta v'\mathbf{B}_{30}^{\prime -1}\mathbf{M}_0^*\Delta Y_{-1} + \phi_{10}\Delta v'\mathbf{M}_0^{**}\Delta v, \\ \Delta v'\mathbf{B}_{30}^{\prime -1}\mathbf{M}_0^*\mathbf{W}_1\Delta Y + \phi_{20}\Delta v'\mathbf{M}_0^{**}\Delta v, \\ \Delta v'\mathbf{B}_{30}^{\prime -1}\mathbf{M}_0^*\mathbf{W}_2\Delta Y_{-1} + \phi_{30}\Delta v'\mathbf{M}_0^{**}\Delta v, \\ \Delta v'\mathbf{M}_0^{**}(C\otimes G_{30})\mathbf{M}_0^{**}\Delta v - \phi_{40}\Delta v'\mathbf{M}_0^{**}\Delta v, \end{cases} \quad (12.19)$$

where $\mathbf{M}^* = \Omega^{-\frac{1}{2}}\mathbf{M}\Omega^{-\frac{1}{2}}$ and $\mathbf{M}^{**} = \mathbf{B}_3^{\prime -1}\mathbf{M}^*\mathbf{B}_3^{-1}$.

It is easy to show that $\mathrm{E}[S^*_{\mathrm{C,N}}(\delta_0)$ has elements:

$$\mu_{\rho_0} = \sigma^2_{v0}\mathrm{tr}[(\mathbf{B}'_{30}\mathbf{B}_{30})^{-1}\mathbf{M}^*_0(\phi_{10}\mathbf{C} - \mathbf{D}_{-10})],$$
$$\mu_{\lambda_{10}} = \sigma^2_{v0}\mathrm{tr}[(\mathbf{B}'_{30}\mathbf{B}_{30})^{-1}\mathbf{M}^*_0(\phi_{20}\mathbf{C} - \mathbf{W}_1\mathbf{D}_0)],$$
$$\mu_{\lambda_{20}} = \sigma^2_{v0}\mathrm{tr}[(\mathbf{B}'_{30}\mathbf{B}_{30})^{-1}\mathbf{M}^*_0(\phi_{30}\mathbf{C} - \mathbf{W}_2\mathbf{D}_{-10})],$$
$$\mu_{\lambda_{30}} = \sigma^2_{v0}\mathrm{tr}[\mathbf{M}^{**}_0(C \otimes G_{30} - \phi_{40}\mathbf{C})].$$

The recentered AQS function thus takes the form:

$$S^{\diamond}_{\mathrm{C,N}}(\delta) = S^*_{\mathrm{C,N}}(\delta) - (\mu_\rho, \mu_{\lambda_1}, \mu_{\lambda_2}, \mu_{\lambda_3})'. \tag{12.20}$$

Note: $\mathbf{M}^* = \Omega^{-1} - \Omega^{-1}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}$. Thus, only the calculation of Ω^{-1} is necessary and the calculations of $\Omega^{\frac{1}{2}}$ and $\Omega^{-\frac{1}{2}}$ are avoided.

To develop an OPMD estimate of the VC matrix of $S^{\diamond}_{\mathrm{C,N}}(\delta_0)$, similar to (12.5) we have,

$$S^{\diamond}_{\mathrm{C,N}}(\delta_0) = \begin{cases} \Delta v'\Psi_1\Delta \mathbf{y}_1 + \Delta v'\Pi_1 + \Delta v'\Phi_1\Delta v - \mu_{\rho_0}, \\ \Delta v'\Psi_2\Delta \mathbf{y}_1 + \Delta v'\Pi_2 + \Delta v'\Phi_2\Delta v - \mu_{\lambda_{10}}, \\ \Delta v'\Psi_3\Delta \mathbf{y}_1 + \Delta v'\Pi_3 + \Delta v'\Phi_3\Delta v - \mu_{\lambda_{20}}, \\ \Delta v'\Phi_4\Delta v - \mu_{\lambda_{30}}, \end{cases} \quad (12.21)$$

where $\Pi_1 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\boldsymbol{\eta}_{-1}$, $\Pi_2 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_1\boldsymbol{\eta}$, $\Pi_3 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_2\boldsymbol{\eta}_{-1}$;

$$\Phi_1 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbb{S}_{-1} + \phi_{10}\mathbf{M}_0^{**},\ \Phi_2 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_1\mathbb{S} + \phi_{20}\mathbf{M}_0^{**},$$

$$\Phi_3 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_2\mathbb{S}_{-1} + \phi_{30}\mathbf{M}_0^{**},\ \Phi_4 = \mathbf{M}_0^{**}(C \otimes G_{30})\mathbf{M}_0^{**} - \phi_{40}\mathbf{M}_0^{**},$$

$$\Psi_1 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbb{R}_{-1},\ \Psi_2 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_1\mathbb{R},\ \Psi_3 = \mathbf{B}_{30}'^{-1}\mathbf{M}_0^*\mathbf{W}_2\mathbb{R}_{-1}.$$

Similar to $\{\mathbf{g}_i\}$ defined based on (12.5), we define $\{\mathbf{g}_i^\circ\}$ based on (12.21).

Now, $\{\mathbf{g}_i^\diamond\}$ are functions of unknown parameters δ_0 and errors Δv.

Replacing δ_0 by $\tilde{\delta}$ and Δv by $\widetilde{\Delta} v$ in $\{\mathbf{g}_i^\diamond\}$ to give $\{\tilde{\mathbf{g}}_i^\diamond\}$, one obtains an OPMD estimate of $\Gamma^\diamond(\delta_0) = \mathrm{Var}[S^\diamond_{\mathrm{C,N}}(\delta_0)]$:

$$\widehat{\Gamma}^\diamond = \sum_{i=1}^n \tilde{\mathbf{g}}_i^\diamond \tilde{\mathbf{g}}_i^{\diamond\prime}. \tag{12.22}$$

Again, to develop the standardized AQS tests in a unified manner, recall $\delta = (\pi', \varphi')'$ and the null hypothesis specifies $\varphi = 0$.

Let $\Sigma^\diamond(\delta_0) = -\mathrm{E}[\frac{\partial}{\partial \delta} S^\diamond(\delta_0)]$. Partition $\Sigma^\diamond(\delta)$ and $\Gamma^\diamond(\delta)$ according to π and φ, and denote their submatrices by $\Sigma^\diamond_{ab}(\delta)$ and $\Gamma^\diamond_{ab}(\delta)$, $a = \pi, \varphi$, $b = \pi, \varphi$.

Let $S^\diamond(\delta) = (S^{\diamond\prime}_\pi(\delta), S^{\diamond\prime}_\varphi(\delta))'$ and $\mathbf{g}_i^\diamond = (\mathbf{g}^{\diamond\prime}_{i,\pi}, \mathbf{g}^{\diamond\prime}_{i,\varphi})'$.

Now, the construction of the test of $\varphi = 0$ depends on $S^*_\varphi(\tilde{\pi}, 0)$ and its variance, where $\tilde{\pi}$ is the null estimate of π.

Similar to (12.14), a Taylor expansion leads to the following asymptotic MD representation:

$$\frac{1}{\sqrt{N}}S^{\diamond}_{\varphi}(\tilde{\pi}, 0_k) = \frac{1}{\sqrt{N}}S^{\diamond}_{\varphi}(\pi_0, 0_k) - \frac{1}{\sqrt{N}}\Sigma^{\diamond}_{\varphi\pi}\Sigma^{\diamond -1}_{\pi\pi}S^{\diamond}_{\pi}(\pi_0, 0_k) + o_p(1)$$
$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{n}(\mathbf{g}^{\diamond}_{i,\varphi} - \Lambda^{\diamond}\mathbf{g}^{\diamond}_{i,\pi}) + o_p(1), \quad (12.23)$$

where $\Lambda^{\diamond} = \Sigma^{\diamond}_{\varphi\pi}\Sigma^{\diamond -1}_{\pi\pi}$. Therefore, the standardized AQS (SAQS) test statistic for testing $H_0 : \varphi = 0$ takes a similar form as the AQS test:

$$T^{\diamond}_{\mathtt{M}} = S^{\diamond\prime}_{\varphi}(\tilde{\pi}, 0_k)\Big\{\sum_{i=1}^{n}(\tilde{\mathbf{g}}^{\diamond}_{i,\varphi} - \tilde{\Lambda}^{\diamond}\tilde{\mathbf{g}}^{\diamond}_{i,\pi})(\tilde{\mathbf{g}}^{\diamond\prime}_{i,\varphi} - \tilde{\Lambda}^{\diamond}\tilde{\mathbf{g}}^{\diamond}_{i,\pi})'\Big\}^{-1}S^{\diamond}_{\varphi}(\tilde{\pi}, 0_k), \quad (12.24)$$

where $\mathtt{M}$ corresponds to $\mathtt{PD}$, $\mathtt{DPD}$, $\mathtt{DSPDr}$, etc., for testing the hypotheses $H_0^{\mathtt{PD}}$, $H_0^{\mathtt{DPD}}$, $H_0^{\mathtt{DSPDr}}$, etc., postulated in Sec. 12.1.

As for $T_{\mathtt{M},0}$ below Theorem 12.1, if IME holds asymptotically, i.e., $\Sigma^{\diamond} = \Gamma^{\diamond} + o(N)$, the test can be simplified to

$$T^{\diamond}_{\mathtt{M},0} = S^{\diamond\prime}(\tilde{\delta})\Big(\sum_{i=1}^{n}\tilde{\mathbf{g}}^{\diamond}_{i}\tilde{\mathbf{g}}^{\diamond\prime}_{i}\Big)^{-1}S^{\diamond}(\tilde{\delta}),$$

where $\tilde{\delta} = (\tilde{\pi}', 0_k')'$.

Furthermore, if null specifies $\delta = 0$, $T^{\diamond}_{\mathtt{PD}}$ reduces to $T^{\diamond}_{\mathtt{PD},0}$ and there is no need of (12.23).

Theorem 12.2

Under Assumptions A-D, if $\tilde{\pi}$ is $\sqrt{N}$-consistent, we have under $H_0^{\mathtt{M}}$,

$$T^{\diamond}_{\mathtt{M}} \xrightarrow{D} \chi^2_k, \text{ as } n \to \infty,$$

where $\mathtt{M}$ denotes a null model specified in Sec. 12.1.

- Monte Carlo results show that the SAQS tests can offer much improvements over the AQS tests when n is not large, particularly when spatial dependence is heavy.
- In each SAQS test, the null estimate $\tilde{\pi}$ can be obtained in the same way as that for the AQS test or solving a subset of equations obtained from $S^{\diamond}_{\mathtt{C},\mathtt{N}}(\delta)$, and $T^{\diamond}_{\mathtt{M}}$ is implemented similarly.

- All the conditional AQS tests discussed in Sec. 12.2 have their counterparts based on the standardized AQS function.
- Similar to the case of the regular AQS tests presented in Sec. 12.2, the standardized AQS tests can also be implemented in a unified manner based on the general expressions (12.20) or (12.21), the VC matrix estimate defined in (12.22), and $\tilde{\Sigma}^{\diamond} = -\frac{\partial}{\partial} S^{\diamond}_{\mathrm{C,N}}(\delta)|_{\delta=\tilde{\delta}_{\mathrm{M}}}$.
- Similar to the AQS tests T_{PD} and T_{DPD}, the two standardized AQS tests, $T^{\diamond}_{\mathrm{PD}}$ and $T^{\diamond}_{\mathrm{DPD}}$, are also robust against both nonnormality and unknown CH.
- Others are in general robust only against nonnormality as the corresponding AQS tests. Therefore, it is desirable to have AQS tests fully robust against unknown CH.

As indicated in the early section, when the null model involves both dynamic and spatial parameters, the AQS tests may not be robust against the unknown CH, and there is no simple way to further adjust the AQS function to make it CH-robust.

Li and Yang (2020b) introduce an alternative way of adjusting the conditional QS functions to give a set of CH-robust AQS functions:

$$S_{\mathrm{H}}^{*}(\psi)=\begin{cases}\frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta u(\theta),\\ \frac{1}{2\sigma_v^4}\Delta u(\theta)'\Omega^{-1}\Delta u(\theta)-\frac{N}{2\sigma_v^2},\\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\Delta Y_{-1}+\frac{1}{\sigma_v^2}\Delta u(\theta)'\mathbf{E}_{\rho}\Delta u(\theta),\\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y+\frac{1}{\sigma_v^2}\Delta u(\theta)'\mathbf{E}_{\lambda_1}\Delta u(\theta),\\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}+\frac{1}{\sigma_v^2}\Delta u(\theta)'\mathbf{E}_{\lambda_2}\Delta u(\theta),\\ \frac{1}{\sigma_v^2}\Delta u(\theta)'\left[C^{-1}\otimes(\mathcal{A}-\mathbf{E}_{\lambda_3})\right]\Delta u(\theta),\end{cases} \tag{12.25}$$

$(\mathbf{E}_{\rho},\mathbf{E}_{\lambda_1},\mathbf{E}_{\lambda_2})=\Omega^{-1}\mathbf{C}^{-1}(\mathbf{D}_{-1},\mathbf{W}_1\mathbf{D},\mathbf{W}_2\mathbf{D}_{-1})$, $\mathbf{E}_{\lambda_3}=B_3'\mathtt{diag}(G_3)[\mathtt{diag}(B_3^{-1})]^{-1}$.

Solving the estimating equations, $S^*_{\mathrm{H}}(\psi) = 0$, gives the CH-robust M-estimator $\hat{\psi}_{\mathrm{H}}$. As before, this can be done by first solving the equations for β and σ_v^2, given $\delta = (\rho, \lambda')'$, to give

$$\begin{aligned}
\hat{\beta}_{\mathrm{H}}(\delta) &= (\Delta X' \Omega^{-1} \Delta X)^{-1} \Delta X' \Omega^{-1} (\mathbf{B}_1 \Delta Y - \mathbf{B}_2 \Delta Y_{-1}), \\
\hat{\sigma}^2_{v,\mathrm{H}}(\delta) &= \tfrac{1}{N} \Delta \hat{u}(\delta)' \Omega^{-1} \Delta \hat{u}(\delta),
\end{aligned}$$

where $\Delta \hat{u}(\delta) = \Delta u(\hat{\beta}(\delta), \rho, \lambda_1, \lambda_2)$. Then, substituting $\hat{\beta}_{\mathrm{H}}(\delta)$ and $\hat{\sigma}^2_{v,\mathrm{H}}(\delta)$ back into the last four components of (12.25) gives the concentrated AQS functions:

$$S^{*c}_{\mathrm{H}}(\delta) = \begin{cases}
\frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \Omega^{-1} \Delta Y_{-1} + \frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \mathbf{E}_\rho \Delta \hat{u}(\delta), \\
\frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \Omega^{-1} \mathbf{W}_1 \Delta Y + \frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \mathbf{E}_{\lambda_1} \Delta \hat{u}(\delta), \\
\frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \Omega^{-1} \mathbf{W}_2 \Delta Y_{-1} + \frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \mathbf{E}_{\lambda_2} \Delta \hat{u}(\delta), \\
\frac{1}{\hat{\sigma}^2_{v,\mathrm{M}}(\delta)} \Delta \hat{u}(\delta)' \left[C^{-1} \otimes (\mathcal{A} - \mathbf{E}_{\lambda_3}) \right] \Delta \hat{u}(\delta).
\end{cases} \qquad (12.26)$$

Solving $S^{*c}_{\mathrm{H}}(\delta) = 0$ gives the CH-robust M-estimator $\hat{\delta}_{\mathrm{H}}$ of δ, and then the CH-robust M-estimators of β and σ_v^2: $\hat{\beta}_{\mathrm{H}} \equiv \hat{\beta}_{\mathrm{H}}(\hat{\delta}_{\mathrm{H}})$ and $\hat{\sigma}^2_{v,\mathrm{H}} \equiv \hat{\sigma}^2_{v,\mathrm{H}}(\hat{\delta}_{\mathrm{H}})$.

By the representations for ΔY and ΔY_{-1} used in Sec. 2.1 and using the relationship $\Delta u = \mathbf{B}_{30}^{-1}\Delta v$, the AQS function at ψ_0 can be written as

$$S_{\mathrm{H}}^{*}(\psi_0) = \begin{cases} \Pi_1'\Delta v, \\ \Delta v'\Phi_1\Delta v - \frac{n(T-1)}{2\sigma_{v0}^2}, \\ \Delta v'\Psi_1\Delta \mathbf{y}_1 + \Pi_2'\Delta v + \Delta v'\Phi_2\Delta v, \\ \Delta v'\Psi_2\Delta \mathbf{y}_1 + \Pi_3'\Delta v + \Delta v'\Phi_3\Delta v, \\ \Delta v'\Psi_3\Delta \mathbf{y}_1 + \Pi_4'\Delta v + \Delta v'\Phi_4\Delta v, \\ \Delta v'\Phi_5\Delta v, \end{cases} \tag{12.27}$$

where $\Pi_1 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\Delta X$, $\Pi_2 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\mathbf{W}_1\boldsymbol{\eta}$, $\Pi_4 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\mathbf{W}_2\boldsymbol{\eta}_{-1}$,

$\Phi_1 = \frac{1}{2\sigma_{v0}^4}\mathbf{C}^{-1}$, $\Phi_2 = \frac{1}{\sigma_{v0}^2}(\mathbb{C}_{b0}\mathbb{S}_{-1} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\rho 0}\mathbf{B}_{30}^{-1})$,

$\Phi_3 = \frac{1}{\sigma_{v0}^2}(\mathbb{C}_{b0}\mathbf{W}_1\mathbb{S} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\lambda_{10}}\mathbf{B}_{30}^{-1})$, $\Phi_4 = \frac{1}{\sigma_{v0}^2}(\mathbb{C}_{b0}\mathbf{W}_2\mathbb{S}_{-1} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\lambda_{20}}\mathbf{B}_{30}^{-1})$,

$\Phi_5 = \frac{1}{\sigma_{v0}^2}[C^{-1} \otimes \left(B_{30}^{-1\prime}(\mathcal{A}_0 - \mathbf{E}_{\lambda_{30}})B_{30}^{-1}\right)]$,

$\Psi_1 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\mathbb{R}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\mathbf{W}_1\mathbb{R}$, and $\Psi_3 = \frac{1}{\sigma_{v0}^2}\mathbb{C}_{b0}\mathbf{W}_2\mathbb{R}_{-1}$.

The similarity between (12.5) and (12.27) immediately leads to an MD representation for the CH-robust AQS function, i.e., $S^*_{\mathrm{H}}(\psi_0) = \sum_{i=1}^n \mathbf{g}_{\mathrm{H}i}$, referring to (12.6)-(12.11).

The vectors $S^*_{\mathrm{H}}(\psi)$ and $\mathbf{g}_{\mathrm{H}i}$, and the matrix $\Sigma^*_{\mathrm{H}}(\psi) = -\mathrm{E}[\frac{\partial}{\partial\psi'} S^*_{\mathrm{H}}(\psi_0)]$ are partitioned in the same way according to ϑ and φ.

A similar asymptotic MD representation, as in (12.14)and (12.23), holds for $S^{*\prime}_{\mathrm{H},\varphi}(\tilde{\vartheta}_{\mathrm{H}}, 0_k)$, where $\tilde{\vartheta}_{\mathrm{H}}$ is the constrained estimator under the null.

A CH-robust AQS test for testing the hypothesis $H_0 : \varphi = 0$ is thus,

$$T^{\dagger}_{\mathrm{M}} = S^{*\prime}_{\mathrm{H},\varphi}(\tilde{\vartheta}_{\mathrm{H}}, 0_k)\Big\{\sum_{i=1}^n (\tilde{\mathbf{g}}_{\mathrm{H}i,\varphi} - \tilde{\Lambda}^*_{\mathrm{H}}\tilde{\mathbf{g}}_{\mathrm{H}i,\vartheta})(\tilde{\mathbf{g}}_{\mathrm{H}i,\varphi} - \tilde{\Lambda}^*_{\mathrm{H}}\tilde{\mathbf{g}}_{\mathrm{H}i,\vartheta})'\Big\}^{-1} S^{*}_{\mathrm{H},\varphi}(\tilde{\vartheta}_{\mathrm{H}}, 0_k), \tag{12.28}$$

where M = PD, DPD, DSPD1, $\cdots$, DSPD5, and SPD, associated with the null hypotheses defined in Sec. 12.1, $\tilde{\Lambda}^*_{\mathrm{H}} = \tilde{\Sigma}^*_{\mathrm{H},\varphi\vartheta}\tilde{\Sigma}^{*-1}_{\mathrm{H},\vartheta\vartheta}$, and $\tilde{\mathbf{g}}_{\mathrm{H}i,\vartheta}$ and $\tilde{\mathbf{g}}_{\mathrm{H}i,\varphi}$ are the null estimates of $\mathbf{g}_{\mathrm{H}i,\vartheta}$ and $\mathbf{g}_{\mathrm{H}i,\varphi}$. We take $\tilde{\Sigma}^*_{\mathrm{H}} = -\frac{\partial}{\partial\psi} S^*_{\mathrm{H}}(\psi)|_{\psi=\tilde{\psi}_{\mathrm{H}}}$ with $\frac{\partial}{\partial\psi} S^*_{\mathrm{H}}(\psi)$ being given in Appendix B.

Theorem 12.3

Under Assumptions A, B, C and D, if $\tilde{\vartheta}_{\mathrm{H}}$ is $\sqrt{N}$-consistent, then under H_0^{M},*

$$T_{\mathrm{M}}^{\dagger} \xrightarrow{D} \chi_k^2, \text{ as } n \to \infty,$$

where M denotes a null model specified in Sec. 12.1.

- Working with the numerator of $S_{\mathrm{H}}^{*c}(\delta)$ given in (12.26), one may be able to obtain finite sample improved tests that are fully robust against unknown CH.
- However, this does not seem to be an easy task, as the existence of unknown CH renders the simple recentering method followed in Sec. 12.3 for the homoskedastic case unapplicable.
- This is seen from the results given in Li and Yang (2020b):

$$\mathrm{E}(\Delta Y_{-1}\Delta v') = -\sigma_{v0}^2 \mathbf{D}_{-10}\mathbf{B}_{30}^{-1}\mathbf{H} \text{ and } \mathrm{E}(\Delta Y \Delta v') = -\sigma_{v0}^2 \mathbf{D}_0 \mathbf{B}_{30}^{-1}\mathbf{H},$$

where $\mathbf{H} = I_{T-1} \otimes \mathcal{H}_n$ and $\mathcal{H}_n = \mathrm{diag}\{h_{ni}, i = 1, \ldots, n\}$.

Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed tests:

- T_{M}, the regular AQS test,
- $T_{\mathrm{M}}^{\diamond}$, the standardized AQS (SAQS) test,
- $T_{\mathrm{M}}^{\dagger}$, the CH-robust AQS test,

in terms of size and size-adjusted power of the tests.

The following data generating process (DGP) is followed:

$$y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + \beta_0 \iota_n + X_t \beta_1 + Z\gamma + \mu + u_t, \quad u_t = \lambda_3 W_3 u_t + v_t,$$

with certain parameter(s) being dropped corresponding to each specific test, for generating observations at the null.

The elements of X_t are generated as in Yang (2018a), and the elements of Z are randomly generated from *Bernoulli*(0.5).

The spatial weight matrices are generated according to `Rook` contiguity, `Queen` contiguity, or group interaction schemes: `Group-I` or `Group-II`.

The `Rook` and `Queen` schemes are standard.

For `Group-I`, we first generate $k = \sqrt{n}$ groups of sizes $n_g \sim U(.5\bar{n}, 1.5\bar{n})$, $g = 1, \cdots, k$ and $\bar{n} = n/k$, and then adjust n_g so that $\sum_{g=1}^{k} n_g = n$.

For `Group-II`, we first generate 6 groups of fixed sizes $(3, 5, 7, 9, 11, 15)$, and replicate these groups r times to give $n = r \times 50$.

See Lin and Lee (2010) and Yang (2018a) for details in generating these spatial layouts.

The cross-sectional heteroskedasticity (CH) is generated according to:

- CH-1: $h_{ni} \propto \frac{1}{T}\sum_{t=1}^{T}|\Delta X_{nt}|$;
- CH-2: $h_{ni} \propto n_g$ for ith unit in gth group of size n_g; and
- CH-3: $h_{ni} \propto n_g$ if $n_g \le \bar{n}_k$; and $\propto 1/n_g^2$ otherwise, where $\bar{n}_k$ is the average group size.
- CH-0: h_{ni}, i.e., the case of homoskedasticity.

Group-I gives strongest spatial interaction andCH-3 gives the most severe cross-sectional heteroskedasticity.

Under Group-II, variation in number of neighbors for each spatial unit stays constant as n increases; in all other spatial layouts, it vanishes as n increases although slower for Group-I (see Yang, 2010).

The values of $(\beta_0, \beta_1, \gamma, \sigma_\mu, \sigma_v)$ are set to $(5, 1, 1, 1, 1)$, $T = 3$ or 6, and $n = (50, 100, 200, 500)$.

Each set of Monte Carlo results is based on 5000 samples (for $T = 3$) or 2000 (for $T = 6$).

The fixed effects μ are generated according to $\frac{1}{T}\sum_{t=1}^{T} X_t + e$, where $e \sim (0, I_N)$.

The error (v_{it}) distributions can be

- (*i*) normal,
- (*ii*) normal mixture (10% $N(0, 4)$ and 90% $N(0, 1)$), or
- (*iii*) lognormal.

In both (*ii*) and (*iii*), the generated errors are standardized to have mean zero and variance σ_v^2.

For testing H_0^{PD}: $\delta = 0$:

- When n is not large, the AQS test T_{PD} and the CH-robust AQS (RAQS) test $T_{\mathrm{PD}}^{\dagger}$ can be severely oversized, whereas the standardized AQS (SAQS) test $T_{\mathrm{PD}}^{\diamond}$ can be slightly undersized.
- As n increases, the empirical sizes of $T_{\mathrm{PD}}^{\diamond}$ quickly approach to their nominal values corresponding to the χ_4^2 distribution.
- As T increases from 3 (Table 1a) to 6 (Table 1b), all tests improve significantly.
- As shown by Corollary 12.1 and Theorem 12.3, these tests are all robust against unknown CH. The results given in Table 1b confirm this. The results further reveal that the severity of CH has a much greater impact on the AQS and RAQS tests than on the SAQS test in finite sample performance.
- As all three tests are asymptotically valid, it is important to compare their finite sample performance in terms of the power of the tests. This has to be done with sizes being adjusted. The results in Table 1c show that the size-adjusted power is the highest for $T_{\mathrm{PD}}^{\diamond}$ and the lowest for $T_{\mathrm{PD}}^{\dagger}$, as expected.

For testing H_0^{DPD}: $\lambda = 0$:

- The results show an excellent performance of the SAQS test with its empirical sizes being very close to their nominal values even when $n = 50$.
- In contrast, the regular and robust AQS tests can have sever size distortions when n is not so large, which get smaller in a significantly slower speed than those of the SAQS test, in particular under CH.
- While all three tests are robust against unknown CH as shown by Corollary 12.2 and Theorem 12.3, their finite sample properties differ (from both reported and unreported results), with T_{DPD} and $T_{\text{DPD}}^{\dagger}$ being affected by the severity of CH much more than the SAQS test $T_{\text{DPD}}^{\diamond}$.
- When T increases from 3 to 6, the AQS and RAQS tests improve significantly. The SAQS test is in general slightly more powerful than the AQS and RAQS tests.
- The true value of ρ does not have a significant effect on both tests.

For testing other hypotheses, general observations hold.

A small set of results are given below, and more in Yang (2021b).

Table 1a Empirical Size of Tests of $H_0^{\mathrm{PD}}: \delta = 0$; `Group-I`, $T = 3$, `CH-0`

		T_{PD}			$T_{\mathrm{PD}}^{\diamond}$			$T_{\mathrm{PD}}^{\dagger}$		
n	dgp	10%	5%	1%	10%	5%	1%	10%	5%	1%
50	1	15.30	8.08	1.22	10.78	4.28	0.46	15.54	8.20	1.16
	2	22.80	14.18	3.90	8.46	3.20	0.22	25.76	16.60	5.18
	3	16.48	9.46	2.16	8.78	3.44	0.22	18.24	10.54	2.64
100	1	12.60	6.48	1.28	10.42	5.08	0.60	13.30	7.38	1.56
	2	17.22	10.34	3.16	9.78	4.16	0.72	19.68	12.16	3.94
	3	14.00	7.96	2.10	9.80	4.54	0.80	16.04	8.98	2.50
200	1	11.14	6.52	1.26	10.56	5.30	0.90	12.66	6.84	1.50
	2	14.08	7.74	1.90	9.22	4.18	0.58	16.34	9.18	2.80
	3	13.70	7.20	1.76	10.78	5.12	0.78	14.60	8.14	1.96
500	1	10.78	5.58	1.32	10.34	4.94	1.28	11.80	6.22	1.46
	2	12.66	6.96	1.44	10.38	5.16	0.80	13.30	7.16	2.00
	3	11.98	6.40	1.62	10.82	5.50	1.08	13.22	7.14	1.56

Note: for dpg, 1 = normal, 2 = normal mixture, 3 = lognormal

Table 1a, Cont'd, `CH-1`

		T_{PD}			$T^{\diamond}_{\text{PD}}$			$T^{\dagger}_{\text{PD}}$		
n	dgp	10%	5%	1%	10%	5%	1%	10%	5%	1%
50	1	19.40	11.52	2.96	9.86	3.82	0.30	19.44	11.92	2.90
	2	27.20	18.52	6.98	8.18	2.96	0.16	29.78	21.16	9.34
	3	22.42	13.94	4.64	8.26	3.28	0.22	24.26	15.10	4.94
100	1	15.70	9.12	2.46	10.12	4.52	0.52	15.46	9.42	2.30
	2	22.14	14.34	4.98	8.86	3.74	0.32	25.44	16.30	6.18
	3	18.52	11.10	3.50	9.66	4.28	0.64	20.50	13.00	4.30
200	1	14.00	7.54	1.72	11.08	5.34	0.76	14.10	7.78	2.14
	2	17.08	9.88	2.64	9.50	3.94	0.56	19.02	11.32	3.74
	3	14.72	8.24	2.10	9.86	4.56	0.90	15.78	8.94	2.38
500	1	11.44	5.94	1.42	10.54	5.24	1.16	12.30	6.52	1.50
	2	12.84	6.98	1.44	9.12	4.18	0.60	14.84	8.66	2.42
	3	11.32	6.04	1.20	9.74	4.70	0.74	13.48	7.20	1.70

Note: for dpg, 1 = normal, 2 = normal mixture, 3 = lognormal

Table 1a, Cont'd, CH−2

		T_{PD}			$T_{PD}^{\diamond}$			$T_{PD}^{\dagger}$		
n	dgp	10%	5%	1%	10%	5%	1%	10%	5%	1%
50	1	15.08	8.42	1.74	10.74	4.72	0.58	16.08	9.02	2.18
	2	21.58	13.04	4.10	8.58	3.40	0.20	23.90	15.06	5.36
	3	17.44	9.94	2.64	9.16	3.58	0.40	19.04	10.98	2.98
100	1	12.26	6.68	1.58	10.42	4.86	0.90	13.72	7.04	1.82
	2	17.52	9.96	2.98	9.52	4.18	0.52	20.08	12.68	4.06
	3	14.42	7.94	2.18	10.32	4.42	0.62	15.76	9.22	2.26
200	1	11.36	5.94	1.18	10.08	4.86	0.74	12.44	6.72	1.24
	2	14.48	8.98	2.26	10.12	4.78	0.74	15.78	9.48	2.58
	3	13.74	7.80	1.92	11.20	5.78	0.78	15.52	8.72	2.00
500	1	10.74	5.56	1.02	10.26	5.06	0.86	12.58	6.68	1.30
	2	11.34	5.82	1.40	9.60	4.66	0.86	13.52	7.66	1.94
	3	11.04	5.84	1.56	10.04	4.96	1.24	12.68	7.02	1.74

Note: for dpg, 1 = normal, 2 = normal mixture, 3 = lognormal

Table 1a, Cont'd, CH−3

n	dgp	T_{PD} 10%	T_{PD} 5%	T_{PD} 1%	$T^{\diamond}_{\text{PD}}$ 10%	$T^{\diamond}_{\text{PD}}$ 5%	$T^{\diamond}_{\text{PD}}$ 1%	$T^{\dagger}_{\text{PD}}$ 10%	$T^{\dagger}_{\text{PD}}$ 5%	$T^{\dagger}_{\text{PD}}$ 1%
50	1	23.94	14.86	4.80	7.74	2.64	0.18	26.34	16.88	5.80
	2	33.68	24.62	12.16	6.26	1.94	0.00	39.92	30.00	15.90
	3	28.44	19.60	8.26	7.06	2.18	0.02	32.94	23.00	10.14
100	1	22.80	14.76	5.32	9.64	3.78	0.10	26.88	18.24	7.14
	2	31.50	22.62	11.42	7.40	2.72	0.08	40.06	30.60	16.96
	3	26.26	18.16	7.78	7.90	2.84	0.20	33.40	24.02	11.40
200	1	15.44	9.04	2.72	10.42	4.56	0.72	17.18	10.80	3.50
	2	22.70	14.42	5.44	9.48	3.64	0.28	26.18	18.00	7.62
	3	19.08	11.78	3.88	10.28	4.82	0.54	21.08	13.38	4.70
500	1	13.48	7.48	1.92	10.96	5.26	0.94	14.90	8.34	2.34
	2	16.76	9.84	2.96	9.10	4.02	0.58	19.86	12.16	4.52
	3	14.32	8.20	2.26	10.14	4.52	0.86	17.74	11.04	3.44

Note: for dpg, 1 = normal, 2 = normal mixture, 3 = lognormal

To facilitate the practical applications of the proposed tests, we provide an empirical illustration using the dataset 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 the chapter is estimated using

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

Estimation and inference results given in Lecture 10 are repeated here for easy reference. 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 10.6b. 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

For testing H_0^{PD}: $\delta = 0$. Data: First 6 Years

- $T_{\mathrm{PD}}^0 = 41.3321$; p-value = 0.000000 (a naive version of T_{PD})
- $T_{\mathrm{PD}} = 40.2631$; p-value = 0.000000
- $T_{\mathrm{PD}}^{\diamond} = 24.8616$; p-value = 0.000054
- TAQS4 = 35.7826; p-value = 0.000000 (a naive version of $T_{\mathrm{PD}}^{\dagger}$)
- $T_{\mathrm{PD}}^{\dagger} = 32.0701$; p-value = 0.000002

For testing H_0^{PD}: $\delta = 0$. Data: Last 6 Years

- $T_{\mathrm{PD}}^0 = 37.4127$; p-value = 0.000000 (a naive version of T_{PD})
- $T_{\mathrm{PD}} = 36.9916$; p-value = 0.000000
- $T_{\mathrm{PD}}^{\diamond} = 26.8423$; p-value = 0.000021
- $T_{\mathrm{PD}}^{\dagger 0} = 20.0693$; p-value = 0.000484 (a naive version of $T_{\mathrm{PD}}^{\dagger}$)
- $T_{\mathrm{PD}}^{\dagger} = 17.7975$; p-value = 0.001352

For testing H_0^{PD}: $\delta = 0$. Full Data

- $T_{\mathrm{PD}}^0 = 46.1398$; p-value = 0.000000 (a naive version of T_{PD})
- $T_{\mathrm{PD}} = 45.4386$; p-value = 0.000000
- $T_{\mathrm{PD}}^{\diamond} = 32.1654$; p-value = 0.000002
- $T_{\mathrm{PD}}^{\dagger 0} = 29.0929$; p-value = 0.000007 (a naive version of $T_{\mathrm{PD}}^{\dagger}$)
- $T_{\mathrm{PD}}^{\dagger} = 28.4627$; p-value = 0.000010

All tests strongly reject the null hypothesis of H_0^{PD}: $\delta = 0$; the heteroskedasticity-robust test $T_{\mathrm{PD}}^{\dagger}$ seems more trustworthy.

For testing H_0^{DPD}: $\lambda = 0$. Data: first 6 years

- $T_{\mathrm{DPD}}^0 = 41.3321$; p-value = 0.000000
- $T_{\mathrm{DPD}} = 40.2631$; p-value = 0.000000
- $T_{\mathrm{DPD}}^{\diamond} = 24.8616$; p-value = 0.000054
- $T_{\mathrm{DPD}}^{\dagger 0} = 35.7826$; p-value = 0.000000
- $T_{\mathrm{DPD}}^{\dagger} = 32.0701$; p-value = 0.000002

For testing H_0^{DPD}: $\lambda = 0$. Data: last 6 years

- T^0_{DPD} = 37.4127; p-value = 0.000000
- T_{DPD} = 36.9916; p-value = 0.000000
- $T^{\diamond}_{\mathrm{DPD}}$ = 26.8423; p-value = 0.000021
- $T^{\dagger 0}_{\mathrm{DPD}}$ = 20.0693; p-value = 0.000484
- $T^{\dagger}_{\mathrm{DPD}}$ = 17.7975; p-value = 0.001352

For testing H_0^{DPD}: $\lambda = 0$. Full data

- T^0_{DPD} = 46.1398; p-value = 0.000000
- T_{DPD} = 45.4386; p-value = 0.000000
- $T^{\diamond}_{\mathrm{DPD}}$ = 32.1654; p-value = 0.000002
- $T^{\dagger 0}_{\mathrm{DPD}}$ = 29.0929; p-value = 0.000007
- $T^{\dagger}_{\mathrm{DPD}}$ = 28.4627; p-value = 0.000010

All tests strongly reject the null hypothesis of H_0^{DPD}: $\lambda = 0$; the heteroskedasticity-robust test $T^{\dagger}_{\mathrm{PD}}$ seems more trustworthy.

For testing H_0^{PD}: $\delta = 0$. Data: first 4 years

- T_{PD}^{0} = 25.6656; p-value = 0.000037
- T_{PD} = 15.0803; p-value = 0.004538
- $T_{\mathrm{PD}}^{\diamond}$ = 3.8901; p-value = 0.421083
- $T_{\mathrm{PD}}^{\dagger 0}$ = 28.7358; p-value = 0.000009
- $T_{\mathrm{PD}}^{\dagger}$ = 13.0713; p-value = 0.010933

For testing H_0^{PD}: $\delta = 0$. Data: last 6 years

- T_{PD}^{0} = 31.2495; p-value = 0.000003
- T_{PD} = 30.3429; p-value = 0.000004
- $T_{\mathrm{PD}}^{\diamond}$ = 22.9664; p-value = 0.000129
- $T_{\mathrm{PD}}^{\dagger 0}$ = 29.2902; p-value = 0.000007
- $T_{\mathrm{PD}}^{\dagger}$ = 27.5405; p-value = 0.000015

Tests of Hypotheses: Cigarette Demand Data

For testing H_0^{PD}: $\delta = 0$. Full data

- $T_{\mathrm{PD}}^{0} = 45.6237$; p-value = 0.000000
- $T_{\mathrm{PD}} = 45.3431$; p-value = 0.000000
- $T_{\mathrm{PD}}^{\diamond} = 21.8516$; p-value = 0.000215
- $T_{\mathrm{PD}}^{\dagger 0} = 24.4905$; p-value = 0.000064
- $T_{\mathrm{PD}}^{\dagger} = 18.6271$; p-value = 0.000930

For testing H_0^{DPD}: $\lambda = 0$. Data: first 4 years

- $T_{\mathrm{DPD}}^{0} = 4.2939$; p-value = 0.231423
- $T_{\mathrm{PD}} = 3.6993$; p-value = 0.295822
- $T_{\mathrm{PD}}^{\diamond} = 3.9825$; p-value = 0.408384
- $T_{\mathrm{PD}}^{\dagger} = 7.7190$; p-value = 0.102432
- $T_{\mathrm{PD}}^{\dagger} = 5.5975$; p-value = 0.102432

For testing H_0^{PD}: $\lambda = 0$. Data: last years

- T^0_{DPD} = 4.5023; p-value = 0.212088
- T_{PD} = 2.4453; p-value = 0.485260
- $T^{\diamond}_{\mathrm{PD}}$ = 3.1956; p-value = 0.525634
- $T^{\dagger}_{\mathrm{PD}}$ = 3.7728; p-value = 0.437633
- $T^{\dagger}_{\mathrm{PD}}$ = 2.1558; p-value = 0.437633

For testing H_0^{PD}: $\lambda = 0$. Data: last years

- T^0_{DPD} = 17.4455; pval1 = 0.000572
- T_{PD} = 11.5752; p-value = 0.008990
- $T^{\diamond}_{\mathrm{PD}}$ = 13.5485; p-value = 0.008885
- $T^{\dagger}_{\mathrm{PD}}$ = 18.3122; p-value = 0.001072
- $T^{\dagger}_{\mathrm{PD}}$ = 14.4653; p-value = 0.001072

- Baltagi, B., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal* 16, 103-134.
- Baltagi, B., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests of spatial dependence. *Regional Science and Urban Economics* 43, 725-739.
- Li, L. Y., Yang, Z. L. 2020b. Estimation of fixed effects spatial dynamic panel data models with small T and unknown heteroskedasticity. *Regional Science and Urban Economics*81, 103520.
- 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. `Supplement` to "Unified *M*-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics* 205, 423-447", `http://www.mysmu.edu/faculty/zlyang/`
- Yang, Z. L. (2021a). Joint tests for dynamic and spatial effects in short dynamic panel data models with fixed effects and heteroskedasticity. *Empirical Economics* 60, 51-92.
- Yang, Z. L. (2021b). `Supplementary Material` to "Joint tests for dynamic and spatial effects in short dynamic panel data models with fixed effects and heteroskedasticity. *Empirical Economics* 60, 51-92", `http://www.mysmu.edu/faculty/zlyang/`

Additional References

- Anselin, L., Bera, A. K., Florax, R., Yoon, M. J., 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26, 77-104.
- Anselin, L., 2001. Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference* 97, 113-139.
- Anselin L., Bera, A. K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: *Handbook of Applied Economic Statistics, Edited by Aman Ullah and David E. A. Giles*. New York: Marcel Dekker.
- Baltagi, B. H., Song, S. H., Koh W., 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123-150.
- Baltagi, B. H., Song, S. H., Jung, Koh W., 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5-51.
- Bera, A. K., Dogan, O., Taspinar, S., Leiluo Y., 2019. Robust LM tests for spatial dynamic panel data models. *Regional Science and Urban Economics* 76, 47-66.
- Debarsy, N., Ertur, C., 2010. Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics* 40, 453-70.
- Born, B., Breitung, J., 2011. Simple regression based tests for spatial dependence. *Econometrics Journal* 14, 330-342.

- Chudik, A., Pesaran, M. H., 2017. A bias-corrected method of moments approach to estimation of dynamic short-T panels. *CESifo Working Paper Series* No. 6688.
- Neyman, J., Scott, E. L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.
- Kelejian H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.
- Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Lee, L. F., Yu, J., 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.
- Li, L. Y., Yang, Z. L. 2020a. Spatial dynamic panel data models with correlated random effects. *Journal of Econometrics*221, 424-454.
- Liu, S. F., Yang, Z. L., 2015. Asymptotic distribution and finite sample bias correction of QML estimators for spatial error dependence model. *Econometrics* **3**, 376-411.
- 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.
- Jin, F., Lee, L. F., 2015. On the bootstrap for Moran's I test for spatial dependence. *Journal of Econometrics* 184, 295-314.

- Jin, F., Lee, L. F., 2018. Outer-product-of-gradients tests for spatial autoregressive models. *Regional Science and Urban Economics* 72, 35-57.
- Robinson, P.M., Rossi, F., 2014. Improved Lagrange multiplier tests in spatial autoregressions. *Econometrics Journal* 17, 139-164.
- Robinson, P.M., Rossi, F., 2015a. Refined tests for spatial correlation. *Econometric Theory* 31, 1249-1280.
- Robinson, P.M., Rossi, F., 2015b. Refinements in maximum likelihood inference on spatial autocorrelation in panel data. *Journal of Econometrics* **189**, 447-456.
- Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.
- Taspinar, S., Dogan, O., Bera, A. K., 2017. GMM gradient tests for spatial dynamic panel data models. *Regional Science and Urban Economics* 65, 65-88.
- Xu, Y. H., Yang, Z. L. (2020). Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Regional Science and Urban Economics* 81, 103488.
- Yang, Z. L., 2018c. Bootstrap LM tests for higher order spatial effects in spatial linear regression models. *Empirical Economics* 55, 35-68.

- Yang, Z. L., 2010. A robust LM test for spatial error components. *Regional Science and Urban Economics* 40, 299-310.
- Yang, Z. L., 2015. LM tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics* 185, 33-59.
- Yu, J., de Jong, R., Lee, L. F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.