Testing for random effects in panel models with spatially correlated disturbances

Helmut Herwartz *

November 2005

Abstract

In the empirical analysis of panel data the Breusch Pagan statistic has become a standard tool to infer on unobserved heterogeneity over the cross section. Put differently, the test statistic is central to discriminate between the pooled regression and the random effects model. Conditional versions of the test statistic have been provided to immunize inference on unobserved heterogeneity against random time effects or patterns of spatial error correlation. Panel data models with spatially correlated error terms are typically set out under the presumption of some known adjacency matrix parameterizing the correlation structure up to a scaling factor. This paper delivers a bootstrap scheme to generate critical values for the Breusch Pagan statistic allowing robust inference under misspecification of the adjacency matrix. Moreover, asymptotic results are derived for the case of a finite cross section and infinite time dimension. Finite sample simulations show that misspecification of spatial covariance features could lead to large size distortions which could be overcome by the robust bootstrap procedure.

Keywords: Breusch Pagan test, random effects model, wild bootstrap, spatial error correlation.
JEL Classification: C12, C33.
1 Introduction

The use of cross country or cross regional data and the application of panel data techniques is becoming more and more popular in macro- as well as spatial econometrics. Traditional panel data models, applied e.g. in microeconometric studies, presume the cross section dimension to be large, or even infinite, in comparison with the time dimension. Data dimensions in applied spatial or macroeconometric studies are often add odds with the latter presumption since the number of countries or regions is naturally bounded whereas the time dimension could be of considerable magnitude. Therefore, asymptotic properties of estimators or test statistics mostly derived for panel data with finite or infinite time series but infinite cross section dimension are somewhat at odds with key features of typical macroeconomic panel data.

Panel data models are often formalized under the restrictive assumption that underlying model disturbances are homoskedastic over time and the cross section and free of contemporaneous error correlation between the cross section members. Intuitively (economic) distance may be seen as some factor governing cross sectional dependence over a set of cross section members such as macroeconomies, regions, provinces etc. As a consequence spatial econometric models (Anselin, 1988) are attracting significant interest in both theoretical and applied econometrics. For a recent review of methodological progress in this field and a portfolio of applied contributions the reader may consult Anselin, Florax and Rey (2004). Given that iid assumptions are hardly realistic for macroeconomic or regional data conditional versions of prominent test statistics such as the Breusch Pagan statistic (BP test) testing the pooled regression model have been put forth. To name a few examples Baltagi, Chang and Qi (1992) provide a version of the BP test which retains its validity in presence (or absence) of random time effects. Herwartz (2005) proposes a resampling procedure immunizing the test statistic against general patterns of cross equation correlation in the spirit of seemingly unrelated regressions (Zellner, 1962) and, moreover, provides a BP test version which is robust against cross sectional heteroskedasticity of model disturbances. Baltagi, Song and Koh (2003) introduce a conditional version of the BP test for the purpose of testing the prevalence of individual effects under spatially correlated model disturbances.

In presence of time or cross sectional heteroskedasticity the asymptotic distribution of common test statistics derived under an iid assumption will depend on nuisance parameters. Moreover, in the framework of spatial modeling misspecification of the adjacency matrix is also likely to deteriorate pivotalness of the BP statistic such that inference based on critical values from the Gaussian or $\chi^2$-distribution will be invalid. Note that in presence of nuisance parameters an analytic first order asymptotic approximation of a test statistics’ limit distribution is often cumbersome if not impossible. Bootstrap approaches have been successfully applied to obtain critical values for nonpivotal test statistics. It is the purpose of this paper to provide a resampling scheme that immunizes the conditional BP test introduced by Baltagi et al. (2003) against misspecification of the spatial correlation matrix. The proposed method exploits a convenient feature of the so-called wild bootstrap being robust under heteroskedastic error distributions (Wu, 1986; Liu, 1988; Mammen, 1993) and cross sectional error correlation (Herwartz and Neumann, 2005).
The derivation of asymptotic results does not rely on normality of underlying error terms and allows the case of a finite cross sectional dimension. Implicitly the proposed testing device will be robust under heteroskedasticity in both directions of panel data.

The remainder of the paper is organized as follows: The panel data model with spatially correlated error terms is given in Section 2. Section 3 provides the conditional BP statistic and discusses its limit behavior in the fully and semiasymptotic case. The bootstrap approach to obtain critical values for the conditional BP test is outlined in Section 4. A simulation study, given in Section 5, investigates the finite sample performance of the resampling scheme and of inference (falsely) presuming an asymptotic Gaussian distribution of the test statistic. Conclusions are drawn in Section 6. An Appendix delivers the proofs of the asymptotic results stated in Section 3 and Section 4.

## 2 The spatial panel data model

Consider the following panel data model with random individual effects by observation and in vector notations, respectively:

\[ y_{it} = x_{it}^\prime \beta + e_{it}, \quad e_{it} = \mu_i + v_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \]  
\[ \Delta \quad y_t = X_t \beta + e_t, \quad e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})', \]  
\[ \Delta \quad y = X \beta + e, \quad e = (e_1', e_2', \ldots, e_T'). \]

In (1) \( \beta \) is a \( K \) dimensional parameter vector and, accordingly, \( x_{it} \) is a \( K \times 1 \) nonstochastic vector of explanatory variables with the first component equal to unity. Error terms \( v_{it} \) are assumed to be contemporaneously correlated with the following spatial structure:

\[ v_t = \rho W v_t + u_t, \quad u_t \sim iid(0, \sigma^2_u I_N), \]  
\[ -1 < \rho < 1, \]  
\[ \text{(4)} \]

where \( I_R \) is short for a \( R \) dimensional identity matrix, \( W \) is a known \( N \times N \) spatial weight matrix with zero diagonal elements the columns of which sum to unity and \( \rho \) is the spatial autocorrelation parameter (Anselin, 1988). Note that the uninformative error vectors \( u_t \) can be recovered from \( v_t \) as \( u_t = B v_t, \quad B = I_N - \rho W \). Instead of stating a specific parametric distribution for the error terms in \( u/\sigma_u, \quad u = (u_1', u_2', \ldots, u_T')', \) it is assumed that the matrix of fourth order moments exists and is finite, i.e.

\[ E[uu' \otimes uu'] = \Upsilon \sigma^4_u, \quad \Upsilon \text{ finite, nonsingular.} \]  
\[ \text{(5)} \]

The random individual effects \( \mu_i \sim (0, \sigma^2_\mu) \) are independent from the spatially correlated disturbances \( v_{it} \). Note that in case \( \sigma^2_\mu = 0 \) the pooled regression with an error covariance structure \( \text{Cov}[v] = I_T \otimes \sigma^2_u B^{-1}(B')^{-1}, \quad v = (v_1', v_2', \ldots, v_T)' \) is obtained as a special case of (1).

In this paper mostly the semiasymptotic case with \( T \to \infty \) and \( N \) finite will be considered. To guarantee consistency of the Ordinary Least Squares (OLS) or weighted least
squares (GLS) estimator of the model in (3) we make the following additional assumptions. Define a matrix \( \Xi_t \) collecting the observations on explanatory variables in time \( t \), i.e.

\[
\Xi_t = \text{diag} \left( x_{1t}, x_{2t}, \ldots, x_{Nt} \right) \in (K \times N)
\]

Note that the vector \( \Xi_t \) has covariance \( \Xi_t \tilde{\Omega} \Xi_t' = \Theta_t, \tilde{\Omega} = \sigma_u^2 B^{-1}(B')^{-1} + I_N \sigma^2_{\mu} \). To invoke a central limit theorem for the sum over heteroskedastic mean zero random variables a multivariate extension of the Lindeberg-Feller conditions is assumed to hold (see e.g. Greene, 2003):

(A2) \[
\lim_{T \to \infty} \bar{M} = M, \text{ where } \bar{M} = \frac{1}{T} \sum_{t=1}^{T} \Xi_t \tilde{\Omega} \Xi_t', \text{ } M \text{ finite, nonsingular.} \tag{6}
\]

(A3) \[
\lim_{T \to \infty} (TM)^{-1} \Theta_t = \lim_{T \to \infty} (T \bar{M})^{-1} \Theta_t = 0. \tag{7}
\]

As mentioned, in absence of individual specific effects, \( \sigma^2_{\mu} = 0 \), the model in (1) collapses to the pooled regression model which is efficiently estimated by means of the following GLS estimator

\[
\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \quad \Omega^{-1} = I_T \otimes (B'B).
\]

Under presence of individual effects, \( \sigma^2_{\mu} > 0 \), the latter estimator will be unbiased but inefficient in general. In this case efficient estimation requires a modified GLS approach based on a covariance matrix \( \Omega = I_T \otimes \tilde{\Omega} \), where in practice first step estimates of both variance parameters will be required. In light of the latter arguments testing the pair of hypotheses

\[
H_0 : \sigma^2_{\mu} = 0 \quad \text{vs.} \quad H_1 : \sigma^2_{\mu} > 0 \tag{9}
\]

is important to determine an efficient estimator for the model in (1). The unconditional version of the BP test is routinely provided by standard econometric software packages, as e.g. LIMDEP. According to the Lagrange Multiplier (LM) principle the BP statistic is determined under the null hypothesis, which facilitates its computation also in the present case where estimation under \( H_0 \) does not require a first step estimate of the variance parameters \( \sigma^2_u \) and \( \sigma^2_{\mu} \).

In its initial version the unconditional BP statistic (Breusch and Pagan, 1980) was derived as a two sided test thereby ignoring the one sided nature of the testing problem formalized in (9). A more natural one sided version of the test statistic goes back to Honda (1985) who also relaxed the assumption of underlying Gaussian disturbances and proves the asymptotic distribution to be Gaussian if both data dimensions \( N \) and \( T \) approach infinity. A standardized version of the one sided statistic improving its empirical size features has been introduced by Moulton and Rudolph (1989). Conditional versions of the BP statistic which allow contemporaneous error correlation induced by random time effects and spatial dependence have been introduced by Baltagi et al. (1992) and Baltagi
et al. (2003), respectively. The latter test and its distributional features in case of a finite cross section or misspecification of the adjacency matrix will be discussed in the next Section.

3 The test statistic and its components

In this section the stochastic features of the BP statistic allowing for the presence (or absence) of spatial error correlation will be addressed. Before providing the statistic in the form introduced by Baltagi et al. (2003) the adopted GLS approach is motivated and, for completeness, GLS estimation of the model parameters is briefly sketched. After the provision of the conditional BP statistic its asymptotic features are given thereby contrasting in particular the scenarios of an infinite vs. a finite cross section dimension.

3.1 A GLS perspective

Adopting a Maximum Likelihood (ML) approach Baltagi et al. (2003) assume iid normality of the underlying disturbances $u_{it}$ and consider the case where the time dimension $T$ is finite and $N$ approaches infinity. Owing to natural features of spatial modeling the underlying assumptions concerning the dimensions of the data lack econometric intuition. Intuitively a spatial analyst might regard the cross section dimension to be given, or small say, whereas in the time dimension the number of available sample points could be unbounded in principle. Moreover, viewing the matrix $W$ as a known $N \times N$ matrix is somewhat at odds with an asymptotic perspective formalizing the case $N \to \infty$. For the latter reasons the empirical properties of the conditional BP statistic in case of a finite cross section will be described in some detail in this Section.

In Baltagi et al. (2003) the allowance of a fixed time series dimension is due to the willingness to assume iid normality when setting out a ML approach. Since this assumption is rather strong the ML framework is replaced here by a GLS approach where error terms $u_{it}$ are iid by assumption. Instead of stating any specific distributional assumption the time series dimension is supposed to converge to infinity. Note that under normality of error terms GLS and ML estimation or inference coincide.

3.2 GLS estimation

To implement GLS estimation of the parameters $\rho$ and $\sigma_u^2$ under $H_0 : \sigma_u^2 = 0$ and conditional on some choice of $W$ the following steps are involved (Anselin 1988):

1. Estimate the pooled regression by means of OLS and obtain residuals $\hat{e}_{OLS} = y - Xb$, $b = (X'X)^{-1}X'y$.

2. Use OLS residuals to determine $\hat{\rho}$ such that the concentrated log-likelihood function

$$l_c = -\frac{NT}{2} \ln \left( \frac{1}{NT} \hat{e}_{OLS}' (I_N \otimes B') (I_N \otimes B) \hat{e}_{OLS} \right) + T \ln |I_N - \rho W|$$

is maximized.
3. Conditional on $\hat{\rho}$ the pooled regression model is estimated by means of the GLS estimator given in (8), from which in turn residual estimates and a maximizer of the concentrated likelihood in (10) can be determined.

4. The latter step is iterated until convergence of $\hat{\rho}$, which then is the EGLS estimator of the spatial correlation parameter. Conditional on this estimator the parameters $\beta$ and $\sigma^2_u$ and disturbances of the spatial model are easily estimated under the null hypothesis implying $e = v$, as

\[
\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \quad \Omega^{-1} = I_T \otimes (B'B), \tag{11}
\]

\[
\hat{e} = \hat{v} = y - X\hat{\beta} = v - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}v, \tag{12}
\]

\[
\hat{\sigma}^2_u = \hat{v}'(I_N \otimes B')(I_N \otimes B)\hat{v}/(NT). \tag{13}
\]

EGLS estimation proceeds from some given estimate of the correlation parameter $\rho$ (and thus of $B$). Consistency of $\hat{\rho}$ implies that asymptotically the EGLS based statistics will converge to their GLS counterparts. To derive the asymptotic behavior of the test statistics and its components under the null hypothesis it is also worthwhile pointing out that with $T \to \infty$ the variance estimator $\hat{\sigma}^2_u$ is consistent for the corresponding true quantity such that the asymptotic features of the statistic do not depend on uncertainty concerning the variance parameter. To facilitate the notation in the following GLS type statistics will be considered throughout. Since the semiasymptotic distribution of the BP statistic will depend on the uncertainty with regard to the underlying model disturbances the only respect where a distinction between estimated variables and their true counterparts enters the notation will distinguish between $e = v$ and $\hat{e} = \hat{v}$, respectively.

### 3.3 The conditional test statistic

Allowing for the potential of spatial error correlation the conditional version of the one sided BP statistic as derived by Baltagi et al. (2003) for the case $u_t \sim \text{iid} \mathcal{N}(0, \sigma^2_u I_N)$, $N \to \infty$ and $T$ fixed is:

\[
\Delta = \hat{\delta}/\sqrt{\text{Var}[\hat{\delta}]} \xrightarrow{d} \mathcal{N}(0, 1), \tag{15}
\]

where

\[
\hat{\delta} = \frac{1}{\sqrt{2\sigma^2_u}} \hat{v}'[J_T \otimes (B'B)^2]\hat{v} - \frac{T}{\sqrt{2}} \text{tr}(B'B) = \hat{\delta}_1 - \hat{\delta}_2, \tag{16}
\]

and $J_T$ is a $T \times T$ unit matrix. Estimated disturbances $\hat{v}$ are obtained from (E)GLS estimation as outlined in Section 3.2. Using $A^2 = AA$ to abbreviate a matrix product the following short hand notations are used in (14): $c = \text{tr}[(W'B + B'W)(B'B)^{-1}][W'B + B'W]$, $d = \text{tr}[W'B + B'W]$, $e = \text{tr}[(B'B)^2]$, $g = \text{tr}[(W'B + B'W)(B'B)^{-1}]$, $h = \text{tr}[B'B]$. Note that
as quoted from Baltagi et al. (2003) the convergence result in (15) holds for finite $T$ only under $u_t \sim \mathcal{N}(0, \sigma_u^2)$. From (14) it can be seen that for the case where, in addition, $T \to \infty$ the Gaussian limit is obtained from $\delta$ after rescaling it by some sequence that increases at rate $T \sqrt{N}$ since all quantities $c, d, e, g, h$ increase at order $O(N)$.

### 3.4 Decomposition of the test statistic

The discussion of the asymptotic properties of the test statistic in (14) or its core component

$$\hat{\delta}_1 = \frac{1}{\sqrt{2\sigma_u^2}} \hat{v}'[J_T \otimes (B'B)^2]\hat{v},$$

will proceed in two steps: In the first place $\hat{\delta}_1$ in (17) is further decomposed in terms involving the true underlying error terms $v$ instead of their GLS estimates. Then, secondly, the asymptotic properties of the leading term in this decomposition are derived in more detail.

Owing to the representations in (17) and (12) the following terms enter the core component of the test statistic under the null hypothesis

$$\hat{\delta}_1 = \frac{1}{\sqrt{2\sigma_u^2}} \hat{v}'[J_T \otimes (B'B)^2]\hat{v} = \frac{1}{\sqrt{2\sigma_u^2}} (L_1 - L_2 - L_3 + L_4),$$

where

\[
\begin{align*}
L_1 &= v'[J_T \otimes (B'B)^2]v \\
L_2 &= v'[J_T \otimes (B'B)^2]X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}v, \\
L_3 &= v'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'[J_T \otimes (B'B)^2]v, \\
L_4 &= v'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'[J_T \otimes (B'B)^2]X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}v.
\end{align*}
\]

The following proposition states the stochastic orders of $L_2$ to $L_4$.

**Proposition 1** Under the assumptions $A1$, $A2$ and $A3$, $T \to \infty$, $N$ finite and $H_0: \sigma^2 = 0$

$$L_i = O_p(T), \quad i = 2, 3, 4.$$ 

Proof: See Appendix.

An implication of the latter result for the distribution of the test statistic $\Delta$ in case with finite $N$ is that dividing the quantity $\hat{\delta}$ in (14) by an $O(T)$ sequence will not suffice to guard against the estimation error associated with $\hat{v}$ under $H_0$ since $L_2/T$ to $L_4/T$ will not vanish asymptotically. As a consequence the limit distribution in the semiasymptotic case will be affected by nuisance parameters.

The following proposition states that under the null hypothesis the term $\hat{\delta}/T$ in (14) and, thus, the LM statistic $\Delta$ itself obeys a bounded distribution in case with finite $N$. 

7
Proposition 2  Under the assumptions A1, A2 and A3, \( T \to \infty \), \( N \) finite and \( H_0 : \sigma^2 = 0 \)

\[
\frac{\hat{\delta}_1 - \delta_2}{T} = O_p(1) \quad (23)
\]

\[
d \to \frac{1}{\sqrt{2}} \sum_{i=1}^{N} \lambda_i (\tilde{\xi}_i^2 - 1) + C, \quad (24)
\]

where the \( \tilde{\xi}_i^2 \) are independent \( \chi^2(1) \) distributed random variables, \( \lambda_i, \ i = 1, \ldots, N \), are the eigenvalues of \( B'B \) and \( C = O_p(1) \).

Proof: See Appendix.

The distribution of \( C \) in (24) is a quadratic form in \( u \). The precise form of this distribution is, however, cumbersome to derive and will in general depend on the set of explanatory variables and, more crucial, on the presumed spatial correlation pattern. Additional remarks on this distribution are made in the Appendix.

Two observations are worthwhile to point out in light of the result in (24). Firstly, given that the first term is a sum of mean zero but heterogeneous random variables a CLT will apply, apart from a suitable standardization, to a rescaled statistic \( \Delta/(T \sqrt{N}) \) as \( N \to \infty \). The proof of the proposition will give some more comments on the fully asymptotic case. However, for a CLT to hold the degree of "cross sectional heterogeneity" of the eigenvalues \( \lambda_i \) has to be bounded in some way (see the Appendix for a formal condition). In turn, then, one may intuitively imagine that the actual speed of convergence to the Gaussian limit will be faster in cases with more homogenous eigenvalues of the \( B'B \) matrix. Secondly, from the formal proof given in the Appendix it will become also evident that the random variables \( \tilde{\xi}_i \) will fail to be independently \( \chi^2(1) \) distributed in case the underlying spatial matrix is misspecified. Moreover, the centering in (16) by means of \( \hat{\delta}_2 = T \text{tr}(B'B)/\sqrt{2} \) will likely fail. For the case of misspecification the random variable \( \Delta/(T \sqrt{N}) \), \( N, T \to \infty \), will still be \( O_p(1) \) but not pivotal anymore. Note for completeness that the common convergence result stated in (24) does not allow cross sectional or time heteroskedasticity of the disturbances \( u_{it} \).

In the light of the practical problems involved when Gaussian quantiles are used as critical values for the conditional BP statistic in (14) it is tempting to look for an alternative device to create critical values retaining its validity for finite \( N \), misspecification of \( W \) or (cross sectional or time) heteroskedasticity. In the following section a bootstrap approach will be outlined and shown to work in the semiasymptotic case.

4  A bootstrap version of the conditional Breusch Pagan test

Bootstrap methods are potential candidates to obtain correct critical values for nonpivotal test statistics. The particular issues raised for the BP statistic, finiteness of the cross sectional dimension or misspecification of the spatial correlation features summarized in
$W$, make the so-called wild bootstrap or external bootstrap a potential approach to test for random individual effects. The wild bootstrap has often been motivated to cope with heteroskedasticity of model disturbances (Wu, 1986; Liu, 1988; Mammen, 1993). Herwartz and Neumann (2005) prove the validity of the wild bootstrap for likelihood ratio type statistics when testing parametric restrictions in cross sectionally correlated systems of univariate error correction models. In a similar vein Herwartz (2005) adopts the wild bootstrap to immunize the BP statistic introduced by Honda (1985) against general patterns of cross sectional error correlation. Regarding the feasibility of the resampling scheme it is noteworthy that the external bootstrap does not require any parametric guess about the potential sources that may underly the failure of pivotalness. Resampling the one sided version of the BP statistic in (14) proceeds along the following steps:

1. Estimate the (pooled regression) model in (3) under $H_0$ by means of (E)GLS conditional on some guess about $W$ and a first step estimator $\hat{\rho}$. Obtain $\hat{\beta}$, estimated disturbances $\hat{v}_t$ and the test statistic $\Delta$.

2. For $g = 1, \ldots, G$, with $G$ sufficiently large,
   - draw vector valued bootstrap disturbances $v^*_t$ matching the second order moments to the corresponding powers of $\hat{v}_t$ as
     \[ v^*_t = \hat{v}_t \cdot \eta_t, \quad \eta_t \sim (0,1), \quad t = 1, \ldots, T, \] (25)
     where $\eta_t$ is a scalar random variable that is independent of the variables in the model;
   - generate bootstrap samples of the $N \times 1$ vectors $y_t$ as
     \[ y^*_t = X_t \hat{\beta} + v^*_t, \quad t = 1, \ldots, T; \]
   - obtain a bootstrap version $\Delta^*$ of $\Delta$ from $y^*$ and $X$ from GLS estimation adopting the same choices for $W$ and $\rho$ as for the initial statistic in Step 1.

3. Decision: Reject $H_0$ with significance level $\alpha$ if $\Delta$ exceeds the $(1 - \alpha)$-quantile of $\Delta^*$ ($\Delta > \Delta^*_{1-\alpha}$).

The following proposition asserts that the bootstrap based test has asymptotically the prescribed size:

**Proposition 3** Under the assumptions A1, A2 and A3, $T \to \infty$, $N$ finite and $H_0 : \sigma^2_\mu = 0$

\[ \sup_{-\infty < c < \infty} |P(\Delta^* \leq c) - P(\Delta \leq c)| \to 0. \]
Proof: See Appendix.

As argued in Herwartz and Neumann (2005) the wild bootstrap is suitable to mimic cross sectional features of the disturbances underlying some panel data model. Bootstrap disturbances match the covariance structure of the underlying true disturbances on average, if

\[
\frac{1}{T} \sum_{t=1}^{T} \text{Cov}(\eta_t \hat{v}_t) = \frac{1}{T} \sum_{t=1}^{T} \hat{v}_t \hat{v}'_t = \frac{1}{T} \sum_{t=1}^{T} v_t v'_t + o_p(1) = \frac{1}{T} \sum_{t=1}^{T} \text{Cov}(v_t) + o_p(1). \tag{26}
\]

A proof for the result in (26) applying to resampling the BP statistic is given in the Appendix. Given this result it is worthwhile pointing out that the bootstrap approach is also justified in case the true covariance undergoes some (moderate) time variation preserving the existence of a finite unconditional covariance matrix. Similarly since misspecification of \( W \) is not crucial for consistency of the (E)GLS estimator of \( \beta \) the bootstrap procedure will retain its validity under the null hypothesis in cases where a false a-priori choice of \( W \) enters the analysis. For the same reason resampling will also work under the null hypothesis in case some "arbitrary" spatial correlation parameter \( \rho \) is chosen for inference.

Several approaches to generate \( \eta_t \) are available from the literature and differ with respect to the low order moments of \( \hat{v}_t \) which are mimicked by the bootstrap design. For the simulations discussed in Section 5 \( \eta_t \) is generated according to the following representation (Mammen, 1993):

\[
\eta_t = d_{t1}/\sqrt{2} + (d_{t2}^2 - 1)/2, \quad d_t = (d_{t1}, d_{t2})', \quad d_t \sim \mathcal{N}(0, I_2). \tag{27}
\]

5 Simulation study

The empirical properties of both the first order asymptotic approximation and the bootstrap approach to determine critical values for the conditional BP statistic are investigated by means of a Monte Carlo study. Mainly two issues are of particular interest when comparing the two avenues of inference on individual effects sketched in this paper. Firstly, it will be of interest in how far inference implemented by means of Gaussian quantiles suffers from misspecification of the spatial correlation matrix or the spatial correlation parameter. Since in practice an analyst has to rely on some a-priori choice of the adjacency matrix there is always a risk of misspecification. In case \( W \) is falsely selected the test statistic \( \Delta \) in (14) will loose its pivotalness whereas the bootstrap procedure, in principle, mimics also the impact of nuisance parameters on the asymptotic distribution of the conditional BP test. In the second place inference in case of a finite cross section will be addressed. From the theoretical discussion only the bootstrap approach promises asymptotically \( (T \to \infty) \) valid significance levels whereas the first order asymptotic approximation is likely to be adversely affected the more heterogeneous are the eigenvalues of the matrix \( (B'B) \).
5.1 The simulation design

The data generating process used for the simulations is the homogeneous model

\[ y_{it} = 1 + x_{it,2} + e_{it}; \quad e_{it} = \mu_i + v_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N. \]  
\[ (28) \]

The exogenous variables \( x_{it,2} \) are generated once from a Gaussian distribution with mean zero and variance 9, \( x_{it,2} \sim N(0, 9) \), and then fixed over all realizations. The disturbances \( v_{it} \) are drawn throughout under a scenario of spatial error correlation formalized by means of the following adjacency matrix:

\[
W = I_{(N/5)} \otimes W^*_0, \quad W^*_0 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0.5 \\
0 & 0 & 0.5 & 0 & 0.5 \\
0 & 0.5 & 0 & 0 & 0
\end{pmatrix}.
\]

The value of the spatial correlation parameter is fixed to \( \rho = 0.8 \) throughout. Summarizing the latter features of the true underlying model the following second order properties of vector valued error terms are obtained:

\[ v_t = B^{-1}u_t, \quad B = (I_N - 0.8(I_{(N/5)} \otimes W^*_0)), \quad u_t \sim \text{iid} N(0, I_N). \]

The particular choice of the adjacency matrix \( W \) is rather arbitrary since it formalizes a link structure for the cross section members such that groups of five ”individuals” share the same network structure excluding any further cross group correlation. Since a main focus of the Monte Carlo experiment is to study the effects of misspecification of \( W \) the proposed structure has the advantage that misspecification of \( W \) may result as a variation of only a few parameters of \( W^*_0 \). For the case \( N = 5 \), however, the ”realistic” matrix \( W^* \) formalizes all cross sectional links.

Inference on individual effects will be investigated under three alternative scenarios regarding the formalization of the spatial error correlation. The analyst is presumed to have knowledge of the correct baseline matrix \( W^*_0 \) and, alternatively, cases of slight misspecification of correlation weights and of a false correlation pattern will be considered. For the latter two cases the presumed adjacency matrices \( W^*_i \) are

\[
W^*_1 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.6 & 0.4 \\
0 & 0 & 0.6 & 0 & 0.6 \\
0 & 0.4 & 0.4 & 0 & 0
\end{pmatrix} \quad \text{and} \quad W^*_2 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix}, \quad (29)
\]

respectively. Apart from misspecification of a presumed adjacency matrix the Monte Carlo analysis also sheds light on the empirical size feature obtained when inference is based on a false a-priori choice of \( \rho = 0.4 \).

To investigate size properties \( e_{it} \) in (28) is just equal to \( v_{it} \). For the estimation of empirical power properties error terms \( e_{it} \) also contain a normally distributed individual
effect, $\mu_i \sim N(0, \sigma_\mu^2), \sigma_\mu^2 = 0.01$. Note that to estimate the empirical power of alternative tests a relatively small variance of random individual effects has been chosen such that the detection of random effects is not trivial even in larger samples. The considered cross section and time dimensions are $N = 5, 10, 20, 50$ and $T = 5, 10, 20, 50, 100$, respectively. Each data generating process is simulated 5000 times and for the wild bootstrap the number of bootstrap replications is $G = 200$.

### 5.2 Monte Carlo results

#### 5.2.1 Inference under misspecification

Table 1 provides empirical size (left hand side panels) and power estimates (right hand side panels) for the competing (E)GLS based approaches to test the hypotheses $H_0 : \sigma_\mu^2 = 0$ vs. $H_1 : \sigma_\mu^2 > 0$. Size adjusted power estimates are not provided since, in particular, inference under a structural misspecification of the spatial correlation matrix by means of Gaussian critical values obtains huge size distortions. The nominal test level for all experiments is 5%. Qualitatively similar results are obtained when testing with a nominal level of 10% such that these results are not given here for space considerations. To facilitate the evaluation of empirical size properties estimates $\hat{\alpha}$ differing from $\alpha$ with 5% significance are indicated with bold entries. Significant size violations are diagnosed in case the empirical rejection frequencies obtained under the null hypothesis are not covered by a confidence band constructed around the nominal level, $\alpha = 0.05 \pm 1.96 \sqrt{\alpha(1-\alpha)/5000}$, obtaining as the respective lower and upper bound 0.044 and 0.056.

Insert Table 1 about here

Taking critical values for the BP statistic from the Gaussian distribution amounts to a rather conservative testing strategy if the structure of the adjacency matrix is misspecified. Inference on individual effects presuming $W^{(2)}$ as the relevant weight matrix yields empirical significance levels below 1% for all experiments with a cross section dimension of at least $N = 20$. The empirical size achieved by the bootstrap critical values is by far closer to the nominal significance level. For some experiments with $N = 10$ or $N = 20$ the empirical rejection frequencies under the null hypothesis cannot be distinguished from the nominal level by statistical means with 5% significance. Note that the bootstrap approach promises validity of inference under the null hypothesis in case the time dimension approaches infinity. For this reason it should not be too surprising that small but significant size distortions are reported for bootstrap inference if the cross sectional dimension gets larger in relation to the time series dimension, as it is the case e.g. for $N = 50$. Power properties of standard inference under structural misspecification of $W$ mirror directly the low rejection frequencies characterizing the test under the null hypothesis. Regarding experiments with $N = 50$ power estimates vary over a range from 0.12% ($T = 5$) to 40 % ($T = 100$) which is far below the corresponding results characterizing bootstrap inference. Over the latter scenarios generating critical values by means of the wild bootstrap obtains empirical power estimates between 9.32% ($T = 5$) and 79.3 % ($T = 100$). When comparing the power properties over alternative DGPs, however, it becomes clear that
the imposition of a wrong structural assumption is not costless when adopting a robust bootstrap procedure. Comparing the power estimates achieved by bootstrap inference over different 'degrees' of misspecification it is seen that empirical power estimates are substantially smaller in case a wrong structure of $W$ is presumed ($W^*_2$) in comparison with a false choice of spatial correlation weights ($W^*_1$). For instance, with $N = 20$ and $T = 50$ performing bootstrap inference under the alternative hypothesis by means of $W^*_{(0)}, W^*_{(1)}, W^*_{(2)}$ yields empirical power estimates of 64.28%, 60.02% and 'only' 27.36%, respectively. Note, however, that taking critical values from the Gaussian distribution is inferior in terms of both empirical test features: The empirical levels indicate that the test is invalid and 'power' estimates are considerably smaller than the bootstrap analoges.

As argued before the case using $W^*_{(1)}$ is regarded as a scenario of 'weak' misspecification of spatial features. For this reason one may expect the empirical test properties for this scenario to be closer to the results discussed for the case where the correct spatial structure is presumed in comparison with those cases where the structure of the weight matrix is subjected to specification error ($W^*_2$). The empirical results provided in Table 1 are throughout in line with the latter conjecture. At the first sight it appears that misspecification implied by adopting $W^*_{(1)}$ instead of $W^*_{(0)}$ improves the empirical size features of the test implemented by the use of Gaussian critical values. The latter observation, however, is likely more an artefact which can be addressed to the heterogeneity of cross sectional eigenvalues of $B'B$ on the one hand and misspecification of the adjacency matrix on the other hand. Whereas the first impact may result in a significantly oversized test procedure (see also the discussion in Section 5.2.2) the latter, in turn, could result in some undersizing of the test. Aggregating over both effects the empirical size properties reported in Table 1 for using Gaussian critical values of the BP statistic under misspecification of $W$ are intuitively understood.

Severe undersizing is also evident for the GLS based statistic derived under $\rho = 0.4$ irrespective of the presumed correct or false adjacency matrix. To provide a representative view at the corresponding simulation results the lower panel of Table 1 provides empirical size and power estimates for the case with $N = 20$. Again bootstrap based size estimates are mostly in line with the nominal test level. Slight but significant oversizing is only diagnosed when the time dimension is small ($T = 5, 10$) and the adjacency matrix is structurally misspecified ($W^*_2$). However, similar to the effects of a wrong presumption concerning the matrix $W$ a false parameter choice for $\rho$ involves considerable power losses. Using bootstrap critical values for test statistic implemented with the true adjacency matrix ($W^*_{(0)}$) for $\rho = 0.4$ and its EGLS counterpart results in empirical rejection frequencies under the alternative hypothesis of 62.4% and 90.5%, respectively.

5.2.2 Inference under finiteness of the cross section

Apart from the misspecification issue discussed above a further motivation for the bootstrap has been its validity under nonnormal disturbances in case of a finite cross section dimension $N$ and infinite time dimension $T$. Asymptotic normality of the test statistic in case $N \to \infty$ hinged upon the cross sectional heterogeneity of the eigenvalues of the matrix $(B'B)$. In the present case the relevant eigenvalues are implied by $W^*_{(1)}$ and the
correlation parameter $\rho = 0.8$. The actual eigenvalues turn out to be 0.04 0.04 1.96 1.96 3.24 each occurring with multiplicity $N/5$. The ratio of the largest vs. the smallest eigenvalue is 81 pointing to substantial implied "cross sectional" heterogeneity. In the light of this heterogeneity the poor speed of convergence to the Gaussian limit is intuitively understood. Employing a correct adjacency matrix standard EGLS inference for given $T = 10$ yields empirical significance levels varying between 4.48% ($N = 5$) and 8.88% ($N = 10$). Over scenarios with $T > N$ the overall magnitude of empirical size distortions appears to become smaller with an increasing cross sectional dimension. From the perspective of empirical practice the latter results are of direct relevance. When discussing estimation or inference results the analyst should always take the a-priori assumed cross sectional heterogeneity into account which is implicit in $W$. Generating critical values for the BP statistic under a correct choice of $W$ by means of the wild bootstrap turns out to deliver empirical size estimates that cannot be distinguished from the nominal test level for most of the performed Monte Carlo experiments. In particular, whenever $T > N$ the bootstrap approach obtains most accurate empirical size estimates. Moreover, the resampling procedure has satisfying power properties, since the rejection frequencies under the alternative hypothesis are rather close to the corresponding estimates obtained for implementations of the test using Gaussian quantiles as critical values.

6 Conclusions

Nuisance parameters affect inference on unobserved heterogeneity by means of the BP test in case of a finite cross section or misspecification of the spatial matrix $W$. In these cases the use of Gaussian quantiles as critical values may involve large size distortions. The recommended bootstrap scheme to generate critical values for the BP statistic accounts implicitly for the nuisance parameters owing to a variety of potential sources. The latter implicitly also includes patterns of cross sectional or time heteroskedasticity. Given its validity under finiteness of the cross sectional dimension the proposed resampling scheme is quite tempting for the practical implementation of spatial econometric models.

As a further aspect of robustness the bootstrap scheme to infer on unobserved heterogeneity in spatial panel models will provide valid critical values in cases where the presumed spatial correlation pattern undergoes some (moderate) time variation. Since ‘adjacency’ might intuitively be seen as some ‘fixed’ concept the latter property could appear to be of limited value for practical aspects. Note, however, that ‘economic distances’ are not necessarily invariant over time but might depend on changing latent factors. In this case common parametric approaches to estimation or inference bear the risk of providing spurious results or at least will lead to invalid significance levels.

As discussed the wild bootstrap allows robust inference on unobserved panel heterogeneity. The proof of its validity, however, basically departs from consistent estimation of the slope parameters underlying the panel model. Owing to the latter argument one may conjecture that the bootstrap approach also applies to other issues of inference that are of interest in panel data modeling, as, for instance, testing poolability and/or significance of slope parameters. In turn the asymptotic results will likely carry over if the assumption of
cross sectional homogeneity of slope parameters is relaxed. Thus, the proposed bootstrap scheme will also be useful for a variety of inference problems in heterogenous panels.

Finally, the ease of implementing the proposed resampling scheme is worthwhile to be pointed out. In this respect it is not only the simplicity of obtaining bootstrap disturbances just from iid quantities with mean zero and unit variance but also the fact that key parameters of a panel model, as e.g. the spatial correlation parameters, have not to be estimated anymore but might a-priori be set to some 'reasonable' level.

Appendix

Proof of Proposition 1:

To derive the stochastic orders of $L_2$ to $L_4$ in (20) to (22) note first that $L_4$ is a quadratic form in the vector of spatially correlated error terms $v$ and similarly $L_2 + L_3$ is also a quadratic form in $v$. Noting that $J_R = j_R j_R'$ where $j_R$ is a $R-$dimensional unit vector and $\Omega^{-1} = (I_T \otimes B')(I_T \otimes B)$ the latter quadratic forms allow the following representations in terms of the error vector $u = (I_T \otimes B)v$:

\[
\frac{1}{\sigma_u^2}(L_2 + L_3) = u' \{(j_T \otimes B)(j_T' \otimes B')X_{(1)}(X_{(1)}'X_{(1)})^{-1}X_{(1)}' \}
\]

\[
+ X_{(1)}(X_{(1)}'X_{(1)})^{-1}X_{(1)}'(j_T \otimes B)(j_T' \otimes B')\}u/\sigma_u^2
\]

\[
= u'(P_1' + P_1)u/\sigma_u^2
\]  

\[
\frac{1}{\sigma_u^2}L_4 = u' \{(X_{(1)}(X_{(1)}'X_{(1)})^{-1}X_{(1)}'(j_T \otimes B)(j_T' \otimes B')X_{(1)}(X_{(1)}'X_{(1)})^{-1}X_{(1)}')\}u/\sigma_u^2
\]

\[
= u'Qu/\sigma_u^2,
\]  

where the $NT \times K$ matrix $X_{(1)}$ collects the $N \times K$ matrices of transformed explanatory variables $X_{(1),t} = BX_t$ in stacked form. Since the random vectors $u/\sigma_u$ are standardized the first order moments of the quadratic forms in (30) and (31) are (see e.g. Schott 1997):

\[
E[(L_2 + L_3)/\sigma_u^2] = \text{tr}(P), \quad P = P_1 + P_1', \quad E[L_4/\sigma_u^2] = \text{tr}(Q),
\]

\[
\text{Var}[(L_2 + L_3)/\sigma_u^2] = \text{tr}((P \otimes P)\Upsilon), \quad \text{Var}[L_4/\sigma_u^2] = \text{tr}((Q \otimes Q)\Upsilon),
\]

where the $(NT)^2 \times (NT)^2$ matrix $\Upsilon$ basically comprises the fourth order moments of $u$ which are finite by A3, i.e.

\[
\sigma_u^4\Upsilon = E[uu' \otimes uu'].
\]  

In view of the latter arguments the quadratic forms in (30) and (31) will be of order $O_p(T)$ if the first and second order expectations given above are of orders $O(T)$ and $O(T^2)$, respectively. The following Lemmas state these orders:

Lemma 1: Under the assumptions A2 and A3, $T \to \infty$ and $N$ finite

\[
\text{tr}(P) = 2\text{tr}(P_1) = O(T), \quad \text{tr}(Q) = O(T).
\]
Proof of Lemma 1:
From the definition of $P_1$ and properties of the trace operator we obtain

$$\text{tr}(P_1) = \text{tr}[(X_{(2)}(j_T \otimes I_N)(j_T' \otimes I_N)X_{(2)})(X_{(1)}'X_{(1)})^{-1}],$$

with $X_{(2)}$ collecting the $N \times K$ matrices of twice transformed explanatory variables $X_{(2),t} = B'X_{(1),t} = B'BX_t$ in stacked form. In (33) the term $(j_T' \otimes I_N)X_{(2)}$ is a $N \times K$ dimensional matrix of cross section specific sums of transformed explanatory variables $x_{(2),t}$, i.e.

$$X = (j_T' \otimes I_N)X_{(2)} = \begin{pmatrix} \sum_{t=1}^T x_{(2),1t}' \\ \sum_{t=1}^T x_{(2),2t}' \\ \vdots \\ \sum_{t=1}^T x_{(2),Nt}' \end{pmatrix}.$$

Under the stationarity assumption made for the explanatory variables and finiteness of $B'B$ each of the latter sums will converge at rate $T$. Thus, for finite $N$ $\text{tr}[X'X(X_{(1)}'X_{(1)})^{-1}]$ is $O(T)$. In case the cross sectional dimension also tends to infinity $\text{tr}(P_1)$ will continue to converge at rate $T$ since the cross sectional dimension enters the limit behavior of both $X'X$ and $X_{(1)}'X_{(1)}$ with unit power. Similar arguments as for $\text{tr}(P_1)$ also apply to the matrix $Q$ governing the quadratic form $L_4$. Note that $\text{tr}(Q) = \text{tr}(P_1)$ is immediately obtained from reordering the components of $Q$. □

Lemma 2: Under the assumptions A1, A2 and A3, $T \to \infty$, and $N$ finite

$$\text{tr}((P \otimes P)\Upsilon) = O(T^2), \text{tr}(Q \otimes Q)\Upsilon = O(T^2).$$

Proof of Lemma 2:
Deriving the variances of the quadratic forms of interest would be greatly facilitated under the assumption of standardized Gaussian disturbances $u_t/\sigma_u$. In this case the following matrix of fourth order moments is obtained (Schott 1997):

$$\Upsilon = E[uu' \otimes uu']/\sigma_u^4 = I_{(NT)^2} + K_{NT,NT} + \text{vec}(I_{NT})\text{vec}(I_{NT})',$$ (34)

where $K_{R,R}$ is the $R$ dimensional square commutation matrix and vec(.) is an operator stacking the columns of a $(R \times S)$ matrix into an $RS$ dimensional column vector. Owing to the presumed moment structure of $u$ $\Upsilon$ will mostly contain zero entries. $NT$ elements of $\Upsilon$ will be equal to 3, the fourth order moment of the Gaussian distribution, and $3NT(NT-1)$ cross moments will give rise to a unit entry in $\Upsilon$. Given that the contribution of the fourth moments to the variance of the quadratic forms in (30) and (31) is of a smaller order in comparison with the contribution of cross moments the order stated in Lemma 2 continues to hold even under a non Gaussian distribution of $u_t/\sigma_u$ having finite fourth moments. The following proof adopts, however, the representation of $\Upsilon$ given in (34). From properties of the trace operator and the commutation matrix given in Lütkepohl
(1996), Section 4.1.1. line (15), Section 9.2.2. lines (5) and (6), and the structure of vec$(I_{NT})vec(I_{NT})'$ the following representation holds
\[
\text{tr}((P \otimes P)\mathcal{Y}) = \text{tr}[P \otimes P + (P \otimes P)K_{NT,NT} + (P \otimes P)\text{vec}(I_{NT})\text{vec}(I_{NT})'] \\
= (\text{tr}(P))^2 + 2\text{tr}(P'P).
\]
(35)
(36)
The first term on the right hand side of (36) is of order $O(T^2)$ and along similar lines followed for the proof of Lemma 1 we get
\[
\text{tr}(P'P) = \text{tr}(P_1P_1 + P_1P'_1 + P'_1P_1) \\
= 2\text{tr}(P_1P_1) + 2\text{tr}(P'_1P_1).
\]
The latter terms are seen to be, respectively:
\[
\text{tr}(P_1P_1) = \text{tr}[X(2)\prime(J_T \otimes I_N)X(2)(X_1(1)X_1(1))^{-1}X(2)\prime(J_T \otimes I_N)X(2)(X_1(1)X_1(1))^{-1}] \\
= \text{tr}[X'X(X_1(1)X_1(1))^{-1}X'X(X_1(1)X_1(1))^{-1}] \\
\text{tr}(P'_1P_1) = \text{tr}[X(1)(X_1(1)X_1(1))^{-1}X(1)\prime(j_T \otimes B)(j_T \otimes B')X(1)(X_1(1)X_1(1))^{-1}X_1'] \\
= T\text{tr}[X(3)'(j_T \otimes I_N)(j_T \otimes I_N)X(3)(X_1(1)X_1(1))^{-1}],
\]
(37)
(38)
where typical matrices stacked in $X(3)$ are $X(3)_{t,s} = BX(2)_{t,s} = B(B'B)_{t,s}$. Employing a similar argument as for proving Lemma 1 the expression $X(3)'(j_T \otimes I_N)(j_T \otimes I_N)X(3)$ in (38) is $O(T)$ for finite $N$. In sum, the components entering tr($(P \otimes P)\mathcal{Y}$) are of order $O(T^2)$. Along similar lines tr($(Q \otimes Q)\mathcal{Y}$) can also be shown to increase at rate $O(T^2)$. □

Given Lemma 1 and Lemma 2 Proposition 1 follows immediately.

**Proof of Proposition 2:**
Under the null hypothesis, $H_0 : \sigma^2 = 0$, the following results are obtained for the stochastic properties of $\delta/T$:
\[
\frac{\delta_1 - \delta_2}{T} = \frac{\delta_1}{T} + \frac{LA - (L2 + L3)}{T} \frac{1}{\sqrt{2}\sigma_u^2} - \frac{\delta_2}{T} \\
= \delta_1/T + O_p(1) - \text{tr}(B'B)/\sqrt{2}.
\]
(39)
For $\delta_1$ on the right hand side of (39) one obtains
\[
\delta_1 = \frac{1}{\sqrt{2}\sigma_u^2}v'[J_T \otimes (B'B)^2]\nu = \frac{1}{\sqrt{2}\sigma_u^2} \left\{ \sum_{t=1}^{T} v_t'v_t(B'B)^2v_t + 2 \sum_{t=1}^{T} \sum_{s=t+1}^{T} v_t'(B'B)^2v_s \right\} \\
= \frac{1}{\sqrt{2}\sigma_u^2} \left\{ \sum_{t=1}^{T} \nu_t^2 + 2 \sum_{t=1}^{T} \sum_{s=t+1}^{T} \nu_t^2 \nu_s \right\},
\]
(40)
where transformed error terms $v_t$ are defined as $v_t = B'Bv_t = B'Y_t$. By construction the error vectors $v_t/\sigma_u$ obey the covariance structure $\text{Cov}[v_t/\sigma_u] \sim (0, B'B)$. In turn, the latter matrix allows the decomposition
\[
B'B = \Gamma\Lambda\Gamma',
\]
(41)
with $\Lambda$ being a diagonal matrix having the eigenvalues $\lambda_i$, $i = 1, \ldots, N$, of $B'B$ along the diagonal. The respective eigenvectors provide the columns of the matrix $\Gamma$. From the decomposition in (41) one may define uninformative innovations $\xi_t$ as

$$\xi_t = \Lambda^{-1/2}v_t/\sigma_u, \quad \xi_t \sim \text{iid}(0, I_N),$$

and obtain the following result for $\delta_1/T$ from (40):

$$\delta_1/T = \frac{1}{\sqrt{2}} \left\{ \sum_{t=1}^{T} \xi_t' \Lambda \xi_t + 2 \sum_{t=1}^{T} \sum_{s=t+1}^{T} \xi_s' \Lambda \xi_s \right\} \frac{1}{T}$$

$$= \frac{1}{\sqrt{2}} \sum_{i=1}^{N} \lambda_i \xi_i^2, \quad \xi_i = \sum_{t=1}^{T} \xi_{it} / \sqrt{T}. \quad (42)$$

Summarizing the latter arguments we obtain the result for $\delta_1/T$ stated in Proposition 2 since $\text{tr}(B'B) = \sum_{i} \lambda_i$ and a CLT for mean zero iid random variables implies that the distribution of terms $\hat{\xi}_i = \xi_i / \sqrt{T}$ is Gaussian and, moreover, independent of other cross sectional units $\hat{\xi}_j, j \neq i$. For the latter reason each term entering the sum in (24) will exhibit a $\chi^2$ distribution with one degree of freedom multiplied by the eigenvalue $\lambda_i$. Moreover, $E[\lambda_i \hat{\xi}_i^2] = \lambda_i$ as $T \to \infty$. For the validity of a CLT when ”averaging” over a cross section of heterogeneous random variables it is crucial that the cross sectional entities $\lambda_i \hat{\xi}_i$ are not too different with regard to their second order features. A CLT will hold (Greene 2003) if the eigenvalues $\lambda_i$ are such that

$$\lim_{N \to \infty} \frac{\lambda_i^2 N}{\lambda_i^2 N} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^2, \text{ exists}$$

and, moreover,

$$\lim_{N \to \infty} \max_i \lambda_i / (N \sqrt{\lambda^2}) = 0.$$

Then, noting that $\text{Var}[\lambda_i \hat{\xi}_i^2] = 2 \lambda_i^2$ a CLT applies in the fully asymptotic case ($N, T \to \infty$), i.e.

$$\frac{\hat{\delta}_1 - \hat{\delta}_2}{T \sqrt{\lambda N \lambda^2}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (43)$$

Note that as $N, T \to \infty$ asymptotic equivalence implies $\text{Var}[\hat{\delta}] \overset{p}{\to} T \sqrt{N \lambda^2}$, where $\text{Var}[\hat{\delta}]$ is given in detail in (14).

**Proof of Proposition 3:**

The proof of the asymptotic validity of bootstrap approximation of the conditional BP statistic follows the same lines as provided in Herwartz (2005) for the unconditional BP statistic. Note that since explanatory variables in $\mathbf{x}_t$ are nonstochastic $\Delta$ is essentially a scalar function of an $N$-dimensional vector of normally distributed random variables, i.e.

$$\Delta = f(v),$$
where \( v = T^{-1/2}(v_1, v_2, \ldots, v_N)' \), \( v_i = \sum_{t} v_{it} \). Now consider the bootstrap approximation of error terms \( v_{it} \),

\[
v_{it}^* = \eta_t \hat{v}_{it} = \eta_t v_{it} - \eta_t x_{it}'(X_{(1)}'X_{(1)})^{-1}X_{(1)}'u.
\]

For the corresponding vector process \( \mathbf{v}^* = T^{-1/2}(v_1^*, v_2^*, \ldots, v_N^*)' \), \( v_i^* = \sum_{t} v_{it}^* \), we have asymptotic normality from the central limit theorem owing to the iid property of \( \eta_t \). For a typical element of \( E[\mathbf{v}^*\mathbf{v}^*'] \) we obtain for finite \( N \)

\[
\text{Cov} \left[ \left( \frac{1}{\sqrt{T}} v_i^* \right) \left( \frac{1}{\sqrt{T}} v_j^* \right) \right] = E \left[ \frac{1}{T} \sum_t \eta_t^2 v_{it}v_{jt} + \frac{1}{T} \sum_t \eta_t^2 O_p(T^{-1}) \right] = \text{Cov} \left[ \left( \frac{1}{\sqrt{T}} v_i \right) \left( \frac{1}{\sqrt{T}} v_j \right) \right] + o_p(1).
\]

Since \( \mathbf{v} \) and its bootstrap approximation \( \mathbf{v}^* \) share the same asymptotic distribution,

\[
f(v^*|\mathbf{X}) \overset{d}{\rightarrow} f(v|\mathbf{X}),
\]

proposition 3 follows from the continuous mapping theorem (Pollard 1984). \( \square \)

Note that the validity of the bootstrap may be proven along similar lines if the explanatory variables were stochastic and the derivation of the test statistic would proceed conditional on the set of explanatory variables \( \mathbf{X} = \{ x_{it} | i = 1, \ldots, N, \ t = 1, \ldots, T \} \). Moreover, from the proof of Proposition 3 it is clear that bootstrap inference might be also applied for other issues arising in the framework of spatial panel models whenever a particular statistic asymptotically depends only on \( \mathbf{v} \).

References


Lütkepohl, H., Handbook of Matrices, John Wiley and Sons, Chichester, 1996.


Table 1: Simulation results

<table>
<thead>
<tr>
<th>Size</th>
<th>EGLS based inference</th>
<th>GLS based inference, ( \rho = 0.4, N = 20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{(0)}^* \Delta</td>
<td>4.26 4.48 7.98 8.80 6.16</td>
<td>0.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>5.32 4.68 4.74 4.86 4.84</td>
<td>5.20 5.22 5.22 4.76 5.42</td>
</tr>
<tr>
<td>W_{(1)}^* \Delta</td>
<td>3.34 3.88 7.10 6.90 4.44</td>
<td>3.78 4.60 4.72 4.78 4.60</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>4.24 3.96 4.58 4.50 4.62</td>
<td>6.42 5.80 5.40 5.53 4.71</td>
</tr>
<tr>
<td>W_{(2)}^* \Delta</td>
<td>2.26 0.62 0.78 0.36 0.12</td>
<td>0.06 0.02 0.18 0.12 0.18</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>2.42 2.36 2.88 3.96 4.04</td>
<td>7.28 5.50 5.90 5.72 5.41</td>
</tr>
<tr>
<td>N = 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{(0)}^* \Delta</td>
<td>5.32 8.88 7.64 5.86 7.60</td>
<td>5.56 6.64 6.50 6.40 6.34</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>5.80 5.12 4.28 4.72 5.50</td>
<td>6.02 5.28 5.26 4.51 4.82</td>
</tr>
<tr>
<td>W_{(1)}^* \Delta</td>
<td>3.94 7.48 5.68 4.84 5.74</td>
<td>3.78 4.60 4.72 4.78 4.60</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>4.94 4.72 4.08 5.04 5.34</td>
<td>6.06 4.56 4.76 5.04 5.16</td>
</tr>
<tr>
<td>W_{(2)}^* \Delta</td>
<td>0.84 1.00 0.80 0.88 0.82</td>
<td>0.40 0.62 0.38 0.32 0.64</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>3.84 3.86 4.44 4.56 4.94</td>
<td>5.60 5.34 5.12 4.28 4.96</td>
</tr>
<tr>
<td>N = 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{(0)}^* \Delta</td>
<td>5.56 7.00 5.78 6.80 6.38</td>
<td>5.56 6.64 6.50 6.40 6.34</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>5.84 5.02 4.92 5.02 4.96</td>
<td>6.02 5.28 5.26 4.51 4.82</td>
</tr>
<tr>
<td>W_{(1)}^* \Delta</td>
<td>4.22 5.26 4.18 5.04 4.82</td>
<td>3.78 4.60 4.72 4.78 4.60</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>6.06 4.56 4.76 5.04 5.16</td>
<td>6.42 5.80 5.40 5.53 4.71</td>
</tr>
<tr>
<td>W_{(2)}^* \Delta</td>
<td>0.40 0.62 0.38 0.32 0.64</td>
<td>0.06 0.02 0.18 0.12 0.18</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>3.84 3.86 4.44 4.56 4.94</td>
<td>7.28 5.50 5.90 5.72 5.41</td>
</tr>
<tr>
<td>N = 50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{(0)}^* \Delta</td>
<td>5.56 6.64 6.50 6.40 6.34</td>
<td>0.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>6.02 5.28 5.26 4.51 4.82</td>
<td>5.20 5.22 5.22 4.76 5.42</td>
</tr>
<tr>
<td>W_{(1)}^* \Delta</td>
<td>3.78 4.60 4.72 4.78 4.60</td>
<td>4.56 5.38 5.20 4.56 5.22</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>6.42 5.80 5.40 5.53 4.71</td>
<td>5.46 5.38 5.20 4.56 5.22</td>
</tr>
<tr>
<td>W_{(2)}^* \Delta</td>
<td>0.10 0.22 0.20 0.12 0.36</td>
<td>7.04 6.04 5.46 4.68 4.82</td>
</tr>
<tr>
<td>\Delta^*</td>
<td>7.04 7.04 4.68 4.68 4.82</td>
<td>7.04 7.44 8.06 13.22 26.14</td>
</tr>
</tbody>
</table>

Size and power estimates of competing tests of \( H_0 : \sigma_{\mu}^2 = 0 \) under alternative selections of the spatial correlation matrix \( W \). \( \Delta \) and \( \Delta^* \) denote the one sided conditional Breusch Pagan statistic, and its bootstrap counterpart to generate critical values. The upper panels (bottom panel) show(s) results for EGLS (GLS) inference where the spatial correlation parameter estimated unrestrictedly (is set to \( \rho = 0.4 \)).