CONSISTENT COVARIANCE MATRIX ESTIMATION WITH SPATIALLY DEPENDENT PANEL DATA

John C. Driscoll and Aart C. Kraay*

Abstract—Many panel data sets encountered in macroeconomics, international economics, regional science, and finance are characterized by cross-sectional or "spatial" dependence. Standard techniques that fail to account for this dependence will result in inconsistently estimated standard errors. In this paper we present conditions under which a simple extension of common nonparametric covariance matrix estimation techniques yields standard error estimates that are robust to very general forms of spatial and temporal dependence as the time dimension becomes large. We illustrate the relevance of this approach using Monte Carlo simulations and a number of empirical examples.

I. Introduction

CROSS-SECTIONAL or "spatial" dependence is a problematic aspect of many panel data sets in which the cross-sectional units are not randomly sampled. For example, spatial correlations may be present in macroeconomics, regional science, or international economics applications in which the cross-sectional units are a nonrandom sample of states, countries, or industries observed over time, as these units are likely to be subject to both observable and unobservable common disturbances. Similarly, in finance applications, cross-sectional units may be shares or portfolios of assets that respond, perhaps heterogeneously, to aggregate market shocks. Although this spatial dependence generally will not interfere with consistent parameter estimation, standard techniques that fail to account for the presence of spatial correlations will yield inconsistent estimates of the standard errors of these parameters.

Corrections for spatial correlation are possible, but suffer from a number of limitations. Consider first standard panel data techniques which assume a fixed time dimension of size T and require the cross-sectional dimension N to become large for their asymptotic justification. In this case, parametric corrections for spatial correlation are possible only if one places strong restrictions on their form, since the number of spatial correlations increases at the rate N^2, while the number of observations grows at only rate N.[1] If the restrictions are misspecified, the properties of these proce-

dures are in general unknown. In view of this, it would appear to be desirable to implement nonparametric corrections for spatial dependence that are analogous to the standard nonparametric time-series corrections for serial dependence. However, this approach has two major drawbacks. First, in most applications there is no natural ordering in the cross-sectional dimension upon which to base the cross-sectional analogues of the necessary mixing conditions.[2] Second, even if such an ordering were available, mixing conditions that require the dependence between two observations that are "far apart" in the cross-sectional ordering to be small rule out canonical forms of cross-sectional dependence such as equal cross-sectional correlations.

In many applications, however, the time dimension of the panel is sufficiently large to justify reliance on asymptotics as $T \rightarrow \infty$, holding fixed the size of the cross-sectional dimension N. In this case, the problem of consistent covariance matrix estimation appears to be more tractable, as it is in principle possible to obtain consistent estimates of the $N \times N$ matrix of cross-sectional correlations by averaging over the time dimension. The estimated cross-sectional covariance matrix can then be used to construct standard errors which are robust to the presence of spatial correlation. The cross-sectional covariance matrix can be estimated either using parametric methods or using standard spectral density matrix estimation techniques of the sort popularized in econometrics applications by Newey and West (1987).[3] Unfortunately this approach also has two drawbacks. First, in finite samples in which the cross-sectional dimension N is large, these techniques may not be feasible since it will not be possible to obtain a nonsingular estimate of the $N \times N$ matrix of cross-sectional correlations using the NT available observations. Second, even when T is sufficiently large relative to N for the estimator to be feasible, its finite sample properties may be quite poor in situations where N and T are

[1] A simple and common strategy is to assume that the spatial correlations are equal for every pair of cross-sectional units so that the inclusion of time dummies eliminates the spatial dependence. Keane and Runkle (1990), Case (1991), and Elliot (1993) are examples of more complicated parametric structures. Anselin (1988) provides a survey of the extensive regional science literature on this problem primarily in the context of pure

cross-sectional regressions. Frees (1995) provides a discussion of testing for spatial correlation using large-N asymptotics.

[2] Conley (1994) circumvents this difficulty by assuming the existence of prior information on a measure of distance in the cross-sectional dimension. The paper of Froot (1989) represents an interesting combination of the parametric and nonparametric approaches. He imposes the parametric restriction that subgroups (industries) in the cross-sectional dimension of firms are independent, and then applies a panel analog of the White (1980) heteroskedasticity-consistent covariance matrix estimation techniques to nonparametrically estimate the sum of the diagonal blocks in the covariance matrix, relying on large-N asymptotics.

[3] An example of the first approach is a variant of the seemingly unrelated regressions (SUR) model of Zellner (1962) in which coefficients are restricted to be equal across equations; three-stage least squares (3SLS) in which a similar restriction is made is another example. Variants on the second approach have found some application in the finance literature. Lehmann (1990) applies the Newey and West (1987) technique in a study of residual risk, but does not provide conditions on the structure of the spatial correlations under which this estimator is consistent.

of comparable orders of magnitude, since the many elements of the cross-sectional covariance matrix will be poorly estimated.

In this paper we propose a simple modification of the standard nonparametric time series covariance matrix estimator which remedies the deficiencies of techniques which rely on large T asymptotics. In particular, we show that a simple transformation of the orthogonality conditions which identify the parameters of the model permits us to construct a covariance matrix estimator which is robust to very general forms of spatial and temporal dependence as the time dimension becomes large. The consistency result holds for any value of N, including the limiting case in which $N \to \infty$ at any rate relative to T. By relying on nonparametric techniques, we avoid the difficulties associated with misspecified parametric estimators. Moreover, since we do not place any restrictions on the limiting behavior of N, the size of the cross-sectional dimension in finite samples is no longer a constraint on feasibility, and we can be confident of the quality of the asymptotic approximation in finite samples in which N and T are of comparable size, or even if N is much larger than T, provided that T is sufficiently large.

The remainder of this paper proceeds as follows. In section II we discuss the intuitions behind our approach, and then use mixing random fields to characterize a broad class of spatial and temporal dependence to which our covariance matrix estimator is robust. We then formally present the consistency result, which is a very simple variant on standard heteroskedasticity and autocorrelation consistent covariance matrix estimation techniques such as those in Newey and West (1987) or Andrews (1991). The next two sections emphasize the practical implications of spatial dependence for consistent covariance matrix estimation. Section III provides Monte Carlo simulations that demonstrate the consequences of failure to correct for spatial dependence, and compare the small-sample properties of the spatial correlation consistent estimator proposed here with common alternatives. We find that the presence of even modest spatial dependence can impart large biases to ordinary least squares (OLS) standard errors when the size of the cross-sectional dimension is large. We also show that the finite-sample performance of the spatial correlation-consistent estimator is very similar to that of the standard Newey and West (1987) time-series estimator, regardless of the size of the cross-sectional dimension. Moreover, despite the fact that the spatial correlation consistent standard error estimator relies on large-T asymptotics, its finite-sample performance dominates that of common alternatives which do not take spatial dependence into account, even when the time dimension is quite short. Section IV presents three empirical examples in which correcting for spatial correlation yields estimated standard errors that differ substantially from conventional estimates. Section V concludes.

II. Consistent Covariance Matrix Estimation with Spatial Dependence

A. Intuitions

In order to fix ideas, consider the class of models combining cross-sectional and time series data that is identified by an $R \times 1$ vector of orthogonality conditions $E[h_{it}(\theta)] = 0$, where $h_{it}(\theta) = h(z_{it}, \theta)$, z_{it} is a vector of data, θ is a $K \times 1$ vector of parameters with $K \leq R$, and the time and unit subscripts vary over $t = 1, \ldots, T$ and $i = 1, \ldots, N$. Under the assumption that N is fixed, conventional time-series generalized method of moments (GMM) estimation of these models proceeds by stacking the R orthogonality conditions for each of the N observations to obtain an $NR \times 1$ vector of moment conditions $E[\tilde{h}_t(\theta)] = 0$, where $\tilde{h}_t(\theta) = [h_{1t}(\theta)', \ldots, h_{Nt}(\theta)']'$. Provided that $\tilde{h}_t(\theta)$ satisfies well-known regularity conditions,[4] it is straightforward to construct the GMM estimator of the parameter vector θ as

$$\hat{\theta}_T = \underset{\{\theta\}}{\operatorname{argmin}} \left[\frac{1}{T}\sum_{t=1}^{T} \tilde{h}_t(\theta)\right]' \hat{\tilde{S}}_T^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \tilde{h}_t(\theta)\right]$$

where $\hat{\tilde{S}}_T$ is a consistent estimator of the $NR \times NR$ matrix,

$$\tilde{S}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E[\tilde{h}_t(\theta)\tilde{h}_s(\theta)'].$$

It is well known that consistent estimation of the variance of the GMM estimator also requires a consistent estimate of this matrix, and Newey and West (1987) have shown that a positive semidefinite estimator of $\tilde{S}_T$ can be obtained using spectral density matrix estimation techniques.

In many data sets encountered in macroeconomics, international economics, or finance, the size of the cross-sectional dimension N may be very large relative to the time dimension T. For example, tests of purchasing power parity or of consumption risk sharing are commonly carried out with samples of more than 100 countries, using 20 or 30 annual observations. In such situations, two difficulties arise. First, if the size of the cross-sectional dimension is too large relative to the time series dimension, it will not be possible to estimate the $NR(NR + 1)/2$ distinct elements in the matrix $\tilde{S}_T$ using the NT available observations in a manner which yields a nonsingular estimate. In this case, it is necessary to place prior restrictions on the form of the spatial correlations in order to reduce the dimensionality of the problem. Second, even if these estimators are feasible, the quality of the asymptotic approximation that is used to justify them is suspect, as the asymptotic theory implies that the ratio N/T

---

[4] For example, standard results on the consistency and asymptotic normality of the GMM estimator require that $\tilde{h}_t(\theta)$ be an NR-dimensional mixing sequence with bounded $(4 + \delta)$th moments, and a measurable and continuously differentiable function of θ.

tends to zero while in the finite sample N is in fact much larger than T.

In light of these two difficulties, it would be desirable to have an estimator which will be feasible even when N is large relative to T, and which does not require the assumption that N is constant for its asymptotic justification. We achieve this by basing our estimation on the following simple transformation of the orthogonality conditions. If we define an $R \times 1$ vector of cross-sectional averages, $h_t(\theta) = (1/N)\Sigma_{i=1}^{N} h_{it}(\theta)$, we can identify the model using only the $R \times 1$ vector of cross-sectional averages of the moment conditions, that is, $E[h_t(\theta)] = 0$. In this case, consistent estimation of the variance of the GMM estimator requires a consistent estimator of the $R \times R$ matrix,

$$S_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E[h_t(\theta)h_s(\theta)'].$$

Since S_T has only $R(R+1)/2$ distinct elements, the size of the cross-sectional dimension is no longer a constraint on the feasibility of estimating this matrix, eliminating the first difficulty.[5] Moreover, since $h_t(\theta)$ is the sum of N elements normalized by N, it will have well-behaved moments for any value of N. This will allow us to make N any nondecreasing function of T in the proof of consistency of the usual covariance matrix estimator, hence addressing the second concern.

A specific example may help to clarify our approach. Consider a simple univariate linear model with cross-sectional but no time-series dependence, that is, $y_{it} = x_{it}\beta + \epsilon_{it}$ with $E[\epsilon_{it}] = 0$ and $E[x_{it}\epsilon_{it}x_{jt}\epsilon_{jt}] = \omega_{ij}$ for all i, j, and t, and which is identified by the assumption that $E[x_{it}\epsilon_{it}] = 0$. It is straightforward to see that the GMM estimator of β based on the cross-sectional average of the orthogonality conditions, $E[h_t(\beta)] = (1/N)\Sigma_{i=1}^{N}E[x_{it}(y_{it} - x_{it}\beta)] = 0$ is simply the OLS estimator applied to the pooled time-series cross-sectional data. Consistent estimation of the variance of this estimator requires a consistent estimate of the sum

$$S_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[x_{it}x_{jt}\epsilon_{it}\epsilon_{jt}] = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}$$

which can readily be constructed by replacing the unobserved disturbances with the estimated residuals.[6] In contrast, suppose that we were to base our estimates on the $N \times 1$ vector of orthogonality conditions, $E[\tilde{h}_t(\beta)] = 0$, where the ith element of $\tilde{h}_t(\beta)$ is $\tilde{h}_{it}(\beta) = x_{it}(y_{it} - x_{it}\beta)$. Consistent

estimation of the variance of this estimator requires a consistent estimate of the $N \times N$ matrix $\tilde{S}_T$, whose ijth element is $\tilde{S}_{Tij} = (1/T)\Sigma_{t=1}^{T}E[x_{it}x_{jt}\epsilon_{it}\epsilon_{jt}] = \omega_{ij}$. Since this procedure (which is the Zellner (1962) seemingly unrelated regressions (SUR) model with the slope coefficients constrained to be equal for each cross section) requires an estimate of each of the $N(N+1)/2$ distinct elements in $\tilde{S}_T$ rather than simply an estimate of the *sum* of these elements, it clearly will not be feasible in situations where the cross-sectional dimension is large relative to the time series. Even when it is feasible, it is reasonable to expect that the finite sample performance of this estimator may be quite poor since each of the cross-sectional covariances will be poorly estimated unless the size of the cross-sectional dimension is small relative to the length of the time series dimension.

In the following subsection, we formalize the intuitions presented here. First, we use mixing random fields to characterize a general class of spatial and temporal dependence for the random variables $h_{it}(\theta)$. Once we verify that the sequence of cross-sectional averages of these random variables satisfies the well-known conditions for consistent covariance matrix estimation (such as those in Newey and West (1987)), the consistency of the usual estimator of S_T is immediate.

B. Random Fields

Random fields, which are simply collections of doubly-indexed random variables, provide a natural framework for describing situations in which both the cross-sectional dimension as well as the time-series dimension of a data set are large.[7] More formally, let Z^2 denote the two-dimensional lattice of integers, $Z^2 = \{(i, t) | i = 1, 2, \ldots, N, \ldots, t = 1, 2, \ldots, T, \ldots\}$, and let $(\Omega, \mathscr{F}, P)$ denote the standard probability triple. A random field is defined as follows:

Definition: The set of random variables $\{\epsilon_z | z \in Z^2\}$ on $(\Omega, \mathscr{F}, P)$ is a random field.

In order to summarize the cross-sectional and time-series dependence between the elements of the random field, consider sets of the form $A_t = \{(i, v) | v \le t\}$. The σ-algebra generated by the collection of random variables whose indices lie in the set A_t, which we denote by $\mathscr{F}^t_{-\infty} \equiv \sigma(\epsilon_z | z \in A_t)$, has the usual interpretation as the information set available at time t. Furthermore, let $\mathscr{F}^\infty_{t+s} \equiv \sigma(\epsilon_z | z \in A^C_{t+s})$, where A^C_t denotes the complement of A_t. Using this notation, we can summarize the dependence between two σ-algebras using α-mixing coefficients defined in a manner

---

[5] We are assuming that the number of orthogonality conditions R is less than the number of cross-sectional units N. This rules out applications where the parameters are permitted to vary across cross-sectional units.

[6] Clearly, OLS standard errors and heteroskedasticity-consistent standard errors would result in inconsistent estimates since they are based on the assumption that the off-diagonal elements of the variance–covariance matrix of the errors are zero, that is, $\omega_{ij} = 0$ for $i \ne j$.

[7] Random field structures have been used extensively in the statistics literature. See Rosenblatt (1970) and Bulinskii (1988) for asymptotic results. Doukhan (1994) provides an extensive survey of mixing in random fields in other contexts. Some economic applications include Wooldridge and White (1988), Quah (1990, 1994), and Conley (1994).

analogous to the standard univariate α-mixing[8] coefficient, that is,

$$\alpha(s) \equiv \sup_{\langle t \rangle} \sup_{\langle F_1 \in \mathscr{F}_{-\infty}^t, F_2 \in \mathscr{F}_{t+s}^\infty \rangle} |P[F_1 \cap F_2] - P[F_1]P[F_2]|$$

We define a mixing random field as follows:

Definition: A random field is α-mixing of size $r/(r-1)$, $r > 1$, if for some $\lambda > r/(r-1)$, $\alpha(s) = O(s^{-\lambda})$.

This definition of mixing departs from the more standard α-mixing structures on random fields in that it treats the cross-sectional dependence differently than the time-series dependence. Most definitions of mixing restrict the dependence in both dimensions symmetrically, requiring the dependence between two observations to decline either as the distance in the cross-sectional ordering becomes large, or as the time separation becomes large. This restriction on the dependence across units is required to deliver $(NT)^{1/2}$ asymptotic normality for double sums over i and t of the ϵ_{it}, just as in the one-dimensional case restrictions on the temporal dependence are required to deliver $T^{1/2}$ asymptotic normality for appropriately normalized sums.[9]

The definition of mixing presented here, however, does not restrict the degree of cross-sectional dependence. Instead, we only require the dependence between ϵ_{it} and $\epsilon_{jt\text{-}s}$ to be small when s is large, for any value of i and j. The advantage of this is that it will not preclude canonical forms of cross-sectional dependence, such as factor structures in which cross-sectional units may be equicorrelated in a given time period, or grouped structures in which observations are correlated according to possibly unobservable group characteristics. This greater permissible cross-sectional dependence comes at the cost that it will not be possible to obtain $(NT)^{1/2}$ asymptotics for double sums over i and t of the ϵ_{it}. However, we do not require this as we are interested only in situations in which the time series dimension of the panel is large, and hence it is possible to rely exclusively on $T^{1/2}$ asymptotics for this double sum.

C. Results

As discussed above, the primary usefulness of this structure is that the sequence of cross-sectional averages of this random field, upon which we will base our covariance matrix estimator, is well behaved. This is formalized in the following result:

Result 1: Suppose that $h_{it}(\theta)$, $t = 1, \ldots, T$, $i = 1, \ldots, N(T)$, is an α-mixing random field of size $r/(r-1)$, $r > 1$. Then

$$h_t(\theta) = \frac{1}{N(T)} \sum_{i=1}^{N(T)} h_{it}(\theta)$$

is an α-mixing sequence of the same size as $h_{it}(\theta)$ for any $N(T)$.

Proof: See appendix.

Moreover, it is immediate that if $E[|h_{it}|^\delta] < D$, for finite constants δ and D, then $E[|h_t|^\delta] < D$ also. Given this result, we can now state the main result on the consistency of the covariance matrix estimator based on the sequence of cross-sectional averages of the moment conditions:

Result 2: Suppose that: (a) $h_{it}(\theta) = h(z_{it}, \theta)$ satisfies conditions (1) and (2) of Newey and West (1987, theorem 2) and condition (5) holds; (b) $h_{it}(\theta)$ is a ϕ-mixing random field of size $2r/(2r-1)$ or an α-mixing random field of size $2r/(r-1)$ as defined above, with $E[h_{it}(\theta)] = 0$. Then

$$\hat{S}_T - S_T = \hat{\Omega}_0 + \sum_{j=1}^{m(T)} w(j, m(T))[\hat{\Omega}_j + \hat{\Omega}_j'] - S_T \xrightarrow{p} 0$$

as $T \to \infty$ for any $N(T)$, where $w(j, m(T)) = 1 - j/[m(T) + 1]$, $\hat{\Omega}_j = T^{-1}\Sigma_{t=j+1}^T h_t(\hat{\theta}_T)h_{t-j}(\hat{\theta}_T)'$, $h_t(\hat{\theta}_T) = N(T)^{-1}\Sigma_{i=1}^{N(T)} h_{it}(\hat{\theta}_T)$, S_T is defined as in section IIA, $\sqrt{T}|\hat{\theta}_T - \theta|$ is bounded in probability and $m(T) = O(T^{1/4})$.

Proof: Follows immediately from result 1 and Newey and West (1987, theorem 2).

A few comments are in order. First, note that the covariance matrix estimator is precisely the standard Newey and West (1987) heteroskedasticity and serial correlation consistent covariance matrix estimator, applied to the sequence of cross-sectional averages of the $h_{it}(\theta)$.[10] As a result, it is computationally very convenient and can easily be constructed using standard econometric software packages.[11] Second, as noted, we emphasize that the random field assumption encompasses a broad class of spatial and temporal dependence that may arise in many empirical applications, and that this estimator requires no further prior knowledge of the exact form of the contemporaneous and lagged cross-unit correlations.

Finally, it is worth mentioning that very little additional structure must be placed on the mixing random field in order to ensure that the GMM estimator itself is consistent and

---

[8] It is straightforward to extend these definitions and the results which follow to ϕ-mixing random fields by defining analogous ϕ-mixing coefficients as follows: $\phi(s) \equiv \sup_{\langle t \rangle} \sup_{\langle F_1 \in \mathscr{F}_{-\infty}^t, F_2 \in \mathscr{F}_{t+s}^\infty \rangle} |P[F_1 | F_2] - P[F_1]|$.

[9] For such random fields $(NT)^{1/2}$ asymptotics typically require N and T to go to infinity at the same rate, suggesting that in finite-sample applications, the cross-sectional and time-series dimension must be roughly equal for asymptotic approximations to be plausible. For example, Quah (1990) has the restriction that $T = \kappa N$. We do not require this restriction, as $N(T)$ can be any nondecreasing function of T.

[10] Various improvements to the Newey and West (1987) specification have subsequently been explored, such as Andrews (1991) and Andrews and Monahan (1992). These refinements are equally applicable here.

[11] GAUSS and TSP procedures to implement these calculations are available from the authors upon request.

asymptotically normal as $T \to \infty$. As we show in the appendix, consistency of the GMM estimator follows immediately from the assumptions of result 2, and with the additional assumption that the mixing random field is asymptotically covariance stationary, asymptotic normality of the GMM estimator itself can also be established. This is important for two reasons. First, without asymptotic normality of the underlying GMM estimator, a consistent covariance matrix estimator would be of little use for inference. Second, it establishes sufficient conditions for the existence of an estimator θ_T that is $\sqrt{T}$-consistent, as assumed in result 2.

III. Monte Carlo Evidence

In this section we use Monte Carlo experiments to investigate the finite sample properties of the spatial correlation consistent covariance matrix estimator described above, and to compare these properties with those of a number of alternative estimators commonly encountered in practical applications. Our interest in these properties is twofold. First, we are interested in the consequences of failing to correct for spatial dependence when it is present. We find that the presence of even relatively moderate spatial dependence is sufficient to impart substantial biases into standard error estimators that do not take the spatial dependence into account. Second, we argued above that conventional GMM estimators based on the $NR \times 1$ vector of orthogonality conditions require an estimate of all the elements of an $NR \times NR$ matrix of cross-sectional correlations. Even when these estimators are feasible, the validity of the asymptotic approximation (which requires the ratio of N/T to tend to zero) is a priori suspect in finite samples when N and T are of comparable orders of magnitude. The Monte Carlo experiments confirm that this suspicion is well grounded: these estimators perform quite poorly and tend to produce spuriously small standard error estimates.

In the remainder of this section, we present a simple data generating process which exhibits spatial and serial dependence, and discuss a number of alternative covariance matrix estimators. We then present and discuss the results of the Monte Carlo experiments.

A. Data Generating Process

For simplicity, we consider the problem of consistently estimating the standard error of the estimator of the slope coefficient in a bivariate linear regression of the form

$$y_{it} = x_{it}\beta + \epsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T.$$

Without loss of generality, we set $\beta = 0$. We introduce contemporaneous and lagged cross-sectional dependence into the model by assuming that the disturbance term is generated by the following factor structure:

$$\epsilon_{it} = \lambda_i f_t + v_{it}$$

where

$$f_t = \rho f_{t-1} + u_{it}.$$

The forcing terms u_{it} and v_{it} are mean zero, mutually independent normal random variables, uncorrelated over time and across units. To focus attention on the effects of spatial dependence, we normalize the variances of the forcing terms to ensure that the regression disturbances ϵ_{it} are homoskedastic and have unit variance, that is, $E[u_{it}^2] = 1 - \rho^2$ and $E[v_{it}^2] = 1 - \lambda_i^2$, so that $E[f_t^2] = E[\epsilon_{it}^2] = 1.$[12]

Cross-sectional dependence in the disturbances arises due to the presence of the unobserved factor f_t, which is common to all cross-sectional units. Since the factor follows an autoregressive process of order 1, both contemporaneous and lagged spatial dependence is present. The extent of the dependence between two cross-sectional units i and j depends on the magnitude of the constant factor loadings λ_i and λ_j and the degree of persistence in the factor ρ. In particular, our normalizations imply that the spatial and temporal correlation in the residuals is of the following simple form:

$$\xi_{ijs} = \text{CORR}\,[\epsilon_{it}\epsilon_{jt-s}] = \begin{cases} 1, & i = j, s = 0 \\ \lambda_i\lambda_j\rho^s, & \text{otherwise.} \end{cases}$$

It is also straightforward to verify that under weak regularity conditions, this factor structure satisfies the conditions required for the consistency of our covariance matrix estimator.[13] Finally, without loss of generality we set the regressor $x_{it} = 1$ so that the problem reduces to one of estimating the (zero) mean of the dependent variable.[14]

To complete the description of the data generating process, we need to choose values for the factor loadings and the autoregressive parameter of the factors. In our Monte Carlo simulations, we will allow the autoregressive parameter to vary over values ranging from zero to 0.5, corresponding to the moderate degree of temporal dependence likely to be present in many applications.[15] Choosing values for the

---

[12] We make this assumption only because it leads to a convenient expression for the spatial correlations. The spatial correlation consistent estimator proposed here is of course also robust to the presence of conditional heteroskedasticity.

[13] The only important condition we require is that the factors loadings λ_i be uniformly bounded constants. Since the AR(1) structure of f_t implies that it is an α-mixing sequence (White (1984), example 3.4.3)), it is immediate that $\lambda_i f_t$ forms an α-mixing random field. It is then straightforward to verify the remaining conditions in result 2. Since the AR(1) structure is covariance stationary, we also can verify that the GMM estimator itself is consistent and asymptotically normal.

[14] This assumption is made purely for expositional clarity, since it implies that the spatial and temporal dependence in the orthogonality condition $E[h_{it}] = E[x_{it}\epsilon_{it}] = E[\epsilon_{it}] = 0$ is the same as that of the residuals. If there is spatial and temporal dependence in the regressors as well, this will change the pattern of spatial and temporal dependence in the orthogonality conditions. We also ran the Monte Carlos reported below assuming that the regressors were generated by the same factor structure as the residuals, with similar results.

[15] We are assuming that, in most applications, the time dimension is sufficiently large that an appropriate transformation or the inclusion of

factor loadings, which in turn characterize the spatial dependence in the data, is more difficult, since spatial dependence is inherently difficult to observe empirically. To reflect the fact that the form of spatial dependence is unlikely to be known in practice, we choose the factor loadings randomly, assuming that they are drawn from a uniform distribution with support $(0, b)$.[16] It is straightforward to verify that this implies that the contemporaneous cross-sectional correlations $\xi_{ij0} = \lambda_i \lambda_j$ are drawn from the distribution $-b^{-2}[\ln (\xi_{ij0}) - 2 \ln (b)]$ over the support $(0, b^2)$, with an average value of $E[\xi_{ij0}] = b^2/4$. We consider two cases: $b = 0.707$, so that the spatial correlations range from 0 to 0.5 with an average value of 0.125, and $b = 1$, so that the spatial correlations range from 0 to 1 with an average value of 0.25. The first case amounts to a very conservative assumption on the strength of the spatial dependence, while the second allows for stronger but still moderate cross-sectional correlations.[17]

B. Alternative Variance Estimators

We now turn to the problem of obtaining a consistent estimate of the variance of point estimators of the regression coefficient β. Clearly, OLS applied to the pooled time-series and cross-sectional data will yield a consistent estimator of β. However, the presence of spatial and temporal dependence in the data implies that the usual OLS standard errors will be inconsistent. In order to provide a benchmark that illustrates the consequences of failure to take into account this spatial dependence, we first compute the inconsistent OLS standard errors.

As an example of an estimator that takes into account the cross-sectional dependence in the data, we consider the Zellner (1962) seemingly unrelated regressions (SUR) estimator of β.[18] As long as there is no serial dependence in the data (i.e., $\rho = 0$), this estimator will yield consistent esti-

lagged variables is sufficient to reduce the serial dependence in the data to these moderate levels. However, there is nothing to prevent us from considering more persistent processes.

[16] To be precise, we first randomly draw values for the factor loadings, and then hold these parameters constant over the subsequent draws from the data generating process. We have also run a variety of Monte Carlo experiments in which we choose specific values for the factor loadings, with qualitatively similar results.

[17] In order to obtain a rough idea as to whether this is a reasonable assumption, consider the cross-country correlations in output and productivity fluctuations, which has received considerable attention in the literature on international equilibrium business cycles. Kraay and Ventura (1997) observe that the average cross-country contemporaneous correlation of GDP growth rates among OECD economies is 0.52, while that of Solow residuals is 0.35. This suggests that our assumptions on the strength of the cross-sectional dependence are rather conservative.

[18] That is, $\beta_{SUR} = (X'(I \otimes \hat{\Sigma}^{-1})X)^{-1}X'(I \otimes \hat{\Sigma}^{-1})Y$, and $V[\hat{\beta}_{SUR}] = (X'(I \otimes \hat{\Sigma}^{-1}) X)^{-1}$ where X and Y denote the $NT \times 1$ vector of observations stacked by cross-sectional unit for every time period and $\hat{\Sigma}$ is the $N \times N$ matrix of estimated contemporaneous cross-sectional covariances based on the OLS residuals. Note that this is simply the conventional SUR estimator imposing the restriction that β be equal across cross-sectional units. Modifications of SUR to admit first-order serial correlation in the disturbances have also been developed (see Judge et al. (1985, sect. 12.3)). However, we do not consider these as our main emphasis is on the difficulties caused by the need to obtain an estimate of all the elements of

TABLE 1.—FINITE SAMPLE PERFORMANCE OF ALTERNATIVE VARIANCE ESTIMATORS T = 25, N = 20

Average Value of Contemporaneous Spatial Correlation	Ratio of Estimated to True Standard Deviation			Coverage Rate of Nominal 95% Confidence Interval		
	$\rho = 0$	$\rho = .25$	$\rho = .50$	$\rho = 0$	$\rho = .25$	$\rho = .50$
0						
OLS	1.018	0.990	0.968	0.950	0.948	0.943
SUR	0.448	0.434	0.424	0.621	0.591	0.585
SCC	0.943	0.922	0.898	0.914	0.914	0.900
0.125						
OLS	0.522	0.385	0.435	0.704	0.566	0.590
SUR	0.402	0.310	0.290	0.569	0.455	0.405
SCC	0.942	0.815	0.772	0.916	0.859	0.819
0.250						
OLS	0.536	0.293	0.247	0.701	0.428	0.352
SUR	0.384	0.286	0.222	0.555	.0420	0.326
SCC	0.921	0.860	0.729	0.905	.0882	0.812

Notes: This table summarizes the finite-sample properties of alternative variance estimators for the model $y_{it} = \beta x_{it} + \epsilon_{it}$, for various values of spatial and temporal dependence. The regressor x_{it} is set to one, β is set to zero and the disturbances are generated according to the factor structure described in section III A. The left-hand side of the table summarizes the bias in the estimator of the variance of the estimator of β by reporting the average value over 1000 replications of the ratio of this variance estimator to the observed sample standard deviation of the estimator of β. The right-hand side of the table reports the fraction of times a nominal 95% confidence interval contains the true value of β. The alternative estimators are as discussed in section III B: ordinary least squares (OLS), seemingly-unrelated regressions restricting the slope coefficient to be equal across cross-sectional units (SUR), and the spatial correlation consistent standard errors proposed in this paper (SCC).

mates of β that are more efficient than the OLS point estimates. Moreover, the standard error estimators will also be consistent, again provided that there is no serial dependence in the data. This estimator is of interest for two reasons. First, it too is frequently applied in situations in which there is spatial dependence. Second, it is straightforward to verify that this estimator is an example of the class of GMM estimators based on the $N \times 1$ vector of orthogonality conditions discussed in section II. As such it is subject to the two generic critiques of these estimators raised earlier: (1) it will not be feasible in situations where the size of the cross-sectional dimension is too large relative to the time dimension, since it will not be possible to obtain a non-singular estimate of the cross-sectional covariance matrix, and (2) even if it is feasible, its finite-sample properties may be quite poor since it requires estimates of a large number of parameters in the cross-sectional covariance matrix.

Finally, we calculate the spatial correlation consistent estimator of the variance of the OLS estimator of β. Specifically, we use the estimated residuals from the pooled time-series cross-sectional OLS regression to construct a sequence of cross-sectional averages of the estimated orthogonality conditions, $h_t(\hat{\beta}) = N^{-1} \sum_{i=1}^{N} h_{it}(\hat{\beta}) = N^{-1} \sum_{i=1}^{N} x_{it}\hat{\epsilon}_{it}$, and we then apply the standard Newey and West (1987) estimator to this sequence.[19]

C. Results

The results of our Monte Carlo experiments are summarized in table 1 and figure 1. Table 1 considers the case of a

the $N \times N$ covariance matrix Σ, which will be present whether or not serial correlation is present.

[19] We use a Bartlett kernel and select the bandwidth $m(T)$ using the corresponding optimal bandwidth estimator suggested by Andrews (1991).

FIGURE 1.—COVERAGE RATES FOR NOMINAL 95% CONFIDENCE INTERVALS
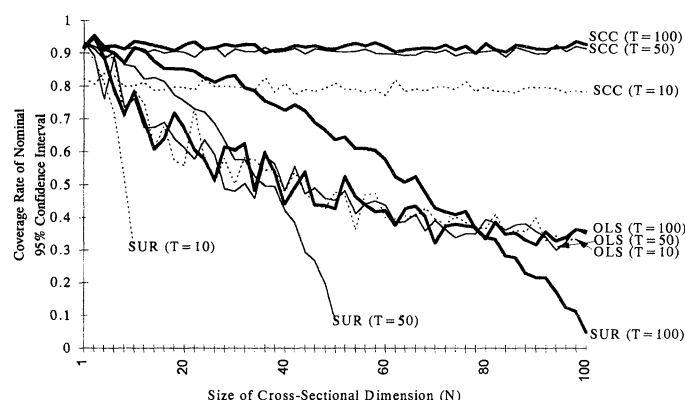
Notes: Figure 1 plots the fraction of times a nominal 95% confidence interval contains the true value of β over 1000 Monte Carlo replications, for the indicated N and T. The alternative estimators are as discussed in section IIIB: ordinary least squares (OLS), seemingly-unrelated regressions restricting the slope coefficient to be equal across cross-sectional units (SUR), and the spatial correlation consistent standard errors proposed in this paper (SCC).

moderate time-series dimension of size $T = 25$ and a moderate cross section of $N = 20$, and describes how the alternative standard error estimators discussed perform for a range of values of the spatial and temporal dependence. The left-hand side of table 1 presents the average value over 1000 replications of the various standard error estimators, expressed as a fraction of the sample standard deviation of the point estimator of β.[20] For consistent estimators such as SUR when $\rho = 0$ or the spatial correlation consistent estimator, this ratio should be close to 1, whereas values less than (greater than) 1 indicate downward (upward) finite sample bias. For inconsistent estimators such as OLS in the presence of spatial dependence or SUR when $\rho \neq 0$, we expect this ratio to differ from 1 in all samples. The right-hand side of table 1 summarizes the consequences for inference of these biases by reporting the fraction of times a nominal 95% confidence interval contains the true value of β.

Table 1 reveals that the OLS and SUR estimators exhibit substantial downward bias in the presence of spatial dependence. Even for moderate values of the spatial dependence, the OLS standard error estimator is biased downward by a factor of about 0.5, and the SUR estimator exhibits even greater downward bias even when it is specified correctly (when $\rho = 0$). This downward bias worsens as the temporal dependence increases. In contrast, the spatial correlation consistent standard error estimator displays relatively little finite sample bias, attaining between 92% and 94% of its true value in the case of no temporal dependence, and slightly worse as the temporal dependence increases. The poor performance of the OLS and SUR estimators is mirrored in the coverage rates reported in the right-hand side of table 1. The coverage rates associated with the spatial correlation consistent standard error estimator are between

[20] The simulations were run using the random number generator in Gauss 3.2. The autoregressive process for the factors was initialized with a draw from the distribution of $\bar{u}_{it}$.

80% and 90% when there is spatial dependence, whereas the OLS and SUR standard error estimators result in coverage rates ranging from about 35% to 60%.

However, the superior performance of the spatial correlation consistent standard error estimator comes at a price. As shown in the upper third of table 1, the spatial correlation consistent standard error estimator is dominated by the OLS estimator when there is no spatial dependence. In this case, the consistent OLS estimator displays only modest finite-sample bias, while the spatial correlation consistent estimator exhibits slightly greater downward bias. However, we shall see shortly that increasing the length of the time-series dimension does much to improve the performance of the spatial correlation consistent standard error estimator. In any case, given the much better performance of this estimator relative to alternatives when there is even modest spatial dependence in the data, this appears to be a small price to pay in return for the greater versatility of the spatial correlation consistent estimator.[21]

It is worth noting that, although the inconsistent OLS standard error estimator exhibits downwards bias in the presence of spatial dependence in our simulations, the sign of the bias is not a general result. Rather, it is a consequence of our data generating process, which introduces only positive off-diagonal elements into the cross-sectional correlation matrix. Since, loosely speaking, the OLS estimator will not take into account the contribution of these positive elements to the variance of the slope estimator, it will exhibit a downward bias.[22] If, on the other hand, the cross-sectional correlations take on both positive and negative values but average to zero, then corrections for spatial dependence will not result in variance estimates that are substantially different from uncorrected estimators.[23] However, this does not

[21] We also note this problem is not unique to our spatial correlation consistent estimator. For example, in his comprehensive survey of the alternative heteroskedasticity and autocorrelation consistent covariance matrix estimators, Andrews (1991) notes that correctly specified parametric estimators dominate nonparametric estimators in a time-series setting.

[22] The downward bias in the SUR estimator is not a consequence of our assumption of positive spatial dependence. Rather, it is due to the feasibility problems of estimators which require positive semidefinite estimates of the covariance matrix. As N becomes large for a fixed value of T, the estimated covariance matrix becomes "nearly" singular, and it is straightforward to show that this introduces downwards bias into standard error estimates.

[23] We also performed our simulations drawing the factor loadings from a uniform distribution with support $(-1, 1)$, so that the cross-sectional correlations averaged to zero. In this case, the performance of the OLS estimators was unaffected by the presence of spatial dependence, for the reasons given in the text. It is interesting to note that, as an artifact of the factor structure we use to introduce spatial dependence, including time dummies in the OLS regressions reported here would have a similar effect. This is because the transformation of taking deviations from period means would result in a similar factor structure, but with demeaned factor loadings. Since the demeaned factor loadings by construction average to zero, this yields similar results to drawing the factor loadings from the interval $(-1, 1)$. However, this too is not a general result. When there are arbitrary cross-sectional correlations in the residuals, taking deviations from period means will not reduce the average value of the off-diagonal elements of the covariance matrix of the residuals, and hence time dummies will not mitigate the problem of spatial dependence.

imply that corrections for spatial correlation are unimportant. In fact, in many empirical applications, it is a priori reasonable to assume that the cross-sectional correlations do not average to zero and are generally positive. Spatial correlations induced by unobserved common shocks will only average to zero if the effects of the shocks are purely redistributive. Moreover, we will see in the empirical examples that follow that the spatial correlation consistent standard errors are generally larger than the uncorrected estimates, indicating that the off-diagonal elements of the covariance matrix are in fact generally positive.

Thus far we have only considered how the performance of the various estimators depends on the presence of spatial and temporal dependence in the data. We now turn to the question of how the finite-sample properties of these estimators depend on the size of the cross-sectional and time-series dimensions. Figure 1 plots the coverage rates associated with nominal 95% confidence intervals constructed using the three estimators under consideration for time dimensions of $T = 10$, 50, and 100 and cross-sectional dimensions ranging from $N = 1$ to $N = 100$, and for the case of $\rho = 0.3$ and an average value of the contemporaneous spatial correlations of 0.125.[24] It is immediately apparent from figure 1 that, as suggested by the asymptotic theory which relies only on T becoming large, coverage rates based on the spatial correlation consistent standard errors do not depend on the size of the cross-sectional dimension. In contrast, the performance of the OLS and SUR estimators deteriorates rapidly as the size of the cross-sectional dimension increases.

Figure 1 also illustrates that for larger values of the time dimension, such as $T = 50$ and $T = 100$, the coverage rates based on the spatial correlation consistent estimator improve over those reported in table 1 (on average 0.908 and 0.920, respectively, versus 0.815 in the case of $T = 25$). Although these coverage rates still fall short of 95%, they unsurprisingly are comparable to the finite sample performance of standard time-series heteroskedasticity and autocorrelation-consistent standard error estimators. For example, Andrews (1991) reports coverage rates of 91.5% for the case of $T = 128$ and $\rho = 0.3$, using a slightly different data generating process and a quadratic-spectal kernel (table IV of that paper). Even for very short time dimensions such as $T = 10$, the spatial correlation consistent estimator manages respectable coverage rates of about 80%, which dominate the two alternatives for all but very small values of N. We conclude from this that in panel data sets where the time series is long enough that the practitioner would be comfortable reporting standard time-series heteroskedasticity and autocorrelation consistent standard errors, s/he should also be comfortable reporting spatial correlation consistent standard errors.

---

[24] For values of N and T where $T < (N + 1)/2$, SUR is infeasible, and coverage rates are not reported.

VI. Empirical Examples

The Monte Carlo evidence of the preceding section indicates that failing to take into account spatial dependence can have significant effects on estimated standard errors and, hence, on inference. In this section we illustrate the point further by comparing spatial correlation consistent standard errors with common alternatives in the context of three empirical examples: tests of purchasing power parity, tests of consumption risk sharing, and estimation of returns to scale. In particular, for each example we compute five sets of standard errors: (1) simple OLS/IV standard errors, (2) White (1980) heteroskedasticity consistent standard errors, which are robust to heteroskedasticity but not serial or spatial correlation, (3) Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors, which are robust to heteroskedasticity and within-unit serial correlation, but not to contemporaneous and lagged spatial correlation, (4) spatial correlation consistent standard errors as proposed here, with lag window $m(T) = 0$ so that they are robust to heteroskedasticity and contemporaneous spatial correlation alone, and (5) spatial correlation consistent standard errors without restricting the lag window to be zero so that they are robust to heteroskedasticity and serial and spatial correlation.[25]

In all three applications it is reasonable to expect that the data is characterized by heteroskedasticity and serial and spatial correlation. Thus, only the last estimator will yield consistent estimates, while the other three will be inconsistent and may be biased in any direction, depending on the nature of the serial and spatial dependence in the data. Nevertheless, it is useful to compare these alternative estimators because it provides a sense of the relative importance of corrections for heteroskedasticity and spatial and serial dependence in applied contexts. Throughout, we will emphasize two comparisons. Comparing the White (1980) standard errors in (2) with the spatial correlation consistent standard errors in (4) illustrates the importance of correcting for contemporaneous cross-unit correlation alone, while comparing (3) and (5) illustrates the importance of allowing for both contemporaneous and and lagged cross-sectional dependence, rather than only within-unit temporal dependence.

A. Purchasing Power Parity

A number of empirical tests of purchasing power parity (PPP) are based on regressions of the form

$$\Delta s_{it} = \alpha + \beta(\Delta p - \Delta p*)_{it} + \epsilon_{it}$$

---

[25] In this panel setting, we compute "conventional" Newey–West standard errors in (3) as follows. First, we construct the usual time-series estimator of $\hat{S}_{iT}$ for each cross-sectional unit. Under the assumption that the cross-sectional units are independent, it is straightforward to see that these can be averaged over the N cross sections to obtain a consistent estimator of $\hat{S}_T$. In both (3) and (5) we set the lag window $m(T) = 2$.

TABLE 2.—PURCHASING POWER PARITY REGRESSIONS

	Forward Regression			Reverse Regression		
	Constant Only	Time Effects	Country Effects	Constant Only	Time Effects	Country Effects
Coefficient	1.006	1.004	1.022	0.814	0.823	0.726
(1) OLS Standard Errors	0.010	0.010	0.013	0.008	0.008	0.009
Standard Errors Robust to:						
(2) Heteroskedasticity	0.054	0.055	0.067	0.051	0.053	0.058
(3) Heteroskedasticity and within-unit serial correlation	0.041	0.042	0.056	0.051	0.052	0.062
(4) Heteroskedasticity and contemporaneous spatial correlation	0.055	0.055	0.068	0.055	0.056	0.062
(5) Heteroskedasticity and contemporaneous and lagged spatial correlation (SCC errors)	0.035	0.035	0.049	0.054	0.053	0.063

Notes: The first three columns report results from estimating $\Delta s_{it} = \alpha + \beta(\Delta p - \Delta p^*)_{it} + \nu_{it}$, where s_{it} is country i's exchange rate relative to the U.S. and the independent variable is the inflation differential between the two countries. α is a set of deterministic variables which alternately contains a constant, time dummies and country dummies, as indicated in the three columns. The next three columns report results from reversing the dependent and independent variables. The first row reports pooled time-series cross-sectional OLS coefficient estimates; the remaining rows report alternative standard error estimates as discussed in the text, with the last row representing the SCC errors proposed in this paper. All regressions use annual data covering the period 1973 to 1993 and a sample of 107 countries.

where the dependent variable is the percentage change in the nominal exchange rate while the independent variable is the inflation differential with a base country, usually taken to be the United States.[26] A point estimate of β which does not differ significantly from one can be interpreted as evidence in favor of (relative) PPP.[27] We estimate several variants on this equation using annual data for a sample of 107 countries between 1973 and 1993.[28]

Table 2 presents OLS estimates of β, together with alternative estimates of its standard error. We note first that the simple OLS standard errors in (1) are much smaller than any of the alternatives. Comparing (2) and (4), however, suggests that in this example correcting for spatial correlation alone has little effect on estimated standard errors, with the spatial correlation-consistent standard error estimates differing from the White standard errors by at most 7% (0.062 versus 0.058) in the last column. In contrast, allowing for both contemporaneous and lagged spatial dependence appears to be more important. Comparing the Newey–West standard errors in (3) with the serial and spatial correlation consistent estimators reveals that the latter are as much as 17% smaller than the former (0.042 versus 0.035 in column 2).

B. Consumption Risk Sharing

Many empirical tests of consumption risk sharing are based on the intuition that, if markets are complete, individual consumption growth should be correlated only with aggregate shocks to income, since individuals are able to

[26] Here we follow the notation of Frankel and Rose (1996). For a survey of the empirical evidence on PPP, see Froot and Rogoff (1995).

[27] However, as discussed in Frankel and Rose (1996) and elsewhere, the alternative hypothesis does not have an economically meaningful interpretation. Tests for mean reversion in real exchange rates provide an alternative framework not subject to this criticism. See Frankel and Rose (1996), Wei and Parsley (1995), and O'Connell (1998) for examples. The paper by O'Connell is particularly relevant here as it considers the power of panel unit root tests in the presence of cross-sectional dependence.

[28] The exchange rate and inflation data are drawn from the International Monetary Fund's International Financial Statistics. Inflation is measured as the change in the logarithm of the consumer price index. The sample of countries is restricted to a sample of 107 countries for which complete data are available between 1973 and 1993.

TABLE 3.—CONSUMPTION RISK-SHARING REGRESSIONS

	53 Countries		OECD (22 countries)	
	1961–90	1971–90	1960–90	1971–90
Coefficient	1.001	1.058	0.773	0.836
(1) OLS Standard Errors	0.020	0.025	0.031	0.041
Standard Errors Robust to:				
(2) Heteroskedasticity	0.041	0.052	0.044	0.052
(3) Heteroskedasticity and within-unit serial correlation	0.046	0.055	0.052	0.057
(4) Heteroskedasticity and contemporaneous spatial correlation	0.037	0.044	0.048	0.070
(5) Heteroskedasticity and contemporaneous and lagged spatial correlation (SCC errors)	0.046	0.043	0.045	0.081

Notes: This table reports the results from estimating $\Delta \ln c_{jt} = \theta(t) + \beta \Delta x_{jt} + \epsilon_{jt}$, where the independent variable is growth in consumption of country j, $\theta(t)$ is a conformable vector of time dummies, and x_{jt} is the idiosyncratic component of GDP growth in country j, defined as the difference between its growth and growth in world average per capita GDP. The first row reports pooled time-series cross-sectional OLS coefficient estimates; the remaining rows report alternative standard error estimates as discussed in the text, with the last row representing the SCC errors proposed in this paper. All data are drawn from the Penn World Tables, Version 5.6.

diversify away all idiosyncratic income risk. Here we consider an example of such a test based on Lewis (1996), who estimates the following regression:

$$\Delta \ln c_{jt} = \theta(t) + \beta x_{jt} + \epsilon_{jt}$$

where c_{jt} is consumption in country j at time t, $\theta(t)$ is a vector of time dummies, and x_{jt} is an idiosyncratic shock to income, defined as the deviation of GDP growth in country j at time t from world average GDP growth. An estimate of β that does not differ significantly from zero can be interpreted as evidence in favor of consumption risk sharing. Table 3 presents the results of estimating this equation for samples of 22 and 53 countries over the period from 1960 to 1990, using OLS.[29] In this example, allowing for contemporaneous cross-sectional dependence has larger effects on esti-

[29] Following Lewis (1996), we take all countries in the Penn World Tables with data quality C+ or better, and restrict our attention to those countries with no missing observations between 1960 and 1990.

TABLE 4.—RETURNS TO SCALE IN U.S. MANUFACTURING INDUSTRIES

Estimation Method Instrument Set	3SLS			2SLS		
	I	II	III	I	II	III
Coefficient	1.120	1.143	1.048	1.312	1.497	1.334
(1) 2SLS/3SLS Standard Errors	0.057	0.056	0.063	0.104	0.095	0.115
Standard Errors Robust to:						
(2) Heteroskedasticity	0.096	0.088	0.125	0.100	0.093	0.137
(3) Heteroskedasticity and within-unit serial correlation	0.100	0.090	0.124	0.104	0.095	0.137
(4) Heteroskedasticity and contemporaneous spatial correlation	0.142	0.108	0.144	0.182	0.141	0.218
(5) Heteroskedasticity and contemporaneous and lagged spatial correlation (SCC errors)	0.185	0.091	0.114	0.195	0.102	0.124

Notes: This table reports results from estimating $\Delta v_{it} = \gamma^F \Delta x_{it} + \Delta a_{it}$, where the dependent variable is growth in value added for industry i and the independent variable is growth in a cost-weighted index of inputs. There are N = 20 industries and T = 32 years (1953–1984). The instrument sets are as follows: I—real growth in military expenditures, the political party of the president, and the world price of oil, plus one lag of each; II—only the contemporaneous values of I; and III—instrument set I augmented with four lagged values of monetary shocks, as described in more detail in Burnside (1996). The first three columns of results are based on three-stage-least-squares estimates. The last three columns are two-stage-least-squares estimates. The first row reports coefficient estimates, while the remaining rows report alternative standard error estimates as discussed in the text, with the last row representing the SCC standard errors proposed in this paper.

mated standard errors. For the 1971–1990 subsample, the spatial correlation consistent standard errors in (4) are 35% larger than the White standard errors in (2) for the OECD sample (0.070 versus 0.052), and 15% smaller in the full sample of 53 countries (0.044 versus 0.052). Comparing the estimated standard errors in (5) and (3) reveals that allowing for contemporaneous and lagged spatial dependence in addition to within-unit serial correlation has even greater effects on estimated standard errors. In the OECD sample over the period of 1971–1990, the spatial correlation consistent standard errors are 42% larger than the Newey–West standard errors (0.081 versus 0.057), while in the 53-country sample they are 22% smaller (0.043 versus 0.055).

C. Returns to Scale in U.S. Manufacturing Industries

The existence of increasing returns to scale has important implications for many branches of macroeconomics. In recent years an empirical literature has emerged which attempts to estimate the degree of returns to scale using regressions of the form

$$\Delta v_{it} = \gamma^F \Delta x_{it} + \Delta a_{it}$$

where Δv_{it} is growth in value added of industry i, Δx_{it} is growth in a cost-weighted index of inputs, and Δa_{it} is unobserved growth in productivity.[30] A point estimate of γ^F that is significantly greater than (less than) 1 can be interpreted as evidence of increasing (decreasing) returns. We estimate variants on this specification using a data set covering 20 manufacturing subsectors in the United States between 1953 and 1984, using various sets of instruments.[31]

Table 4 contains the results of these regressions. In the first panel we present point estimates obtained using three-stage least squares (3SLS), ordinary 3SLS standard errors,

and the four remaining alternatives (2)–(5) as before. Since 3SLS point estimates incorporate information in the (possibly poorly) estimated cross-sectional covariance matrix, we also report point estimates obtained using 2SLS and the corresponding standard errors.[32] Given that in this example the cross-sectional units are industries within a single country and are thus very likely to be subject to common shocks, it is not surprising that corrections for spatial correlation have large effects on estimated standard errors. As before, the importance of correcting for contemporaneous spatial dependence can be seen by comparing the heteroskedasticity consistent White standard errors in (2) with the heteroskedasticity and contemporaneous spatial correlation consistent standard errors in (4). In all cases the latter are larger than the former, and by as much as 82% in the case of column 4 of table 4 (0.182 versus 0.100). Lagged cross-sectional dependence also figures prominently, as shown by a comparison of the standard errors in (3) and (5), with spatial correlation consistent standard errors as much as 88% higher in the case of column 4 (0.195 versus 0.104).

V. Conclusions

Spatial and other forms of cross-sectional correlation are likely to be an important complicating factor in many empirical studies. Standard techniques which fail to take into account this spatial dependence will lead to inconsistent standard error estimates. Moreover, available techniques that in principle are robust to the presence of spatial correlation are often either infeasible or else have poor finite sample properties. In this paper, we have shown that a simple transformation of the orthogonality conditions which identify a broad class of panel data models permits consistent covariance matrix estimation in the presence of very general spatial dependence. Moreover, in contrast to existing techniques, the size of the cross-sectional dimension does

[30] Here we follow the notation of Burnside (1996). See Caballero and Lyons (1989, 1992), Hall (1988, 1990), Basu and Fernald (1995), and Basu (1996) for further examples.

[31] The data and instruments are described in more detail in Burnside (1996). We are grateful to Craig Burnside for kindly providing us with the data set and his programs. The data were originally constructed by Robert Hall, to whom thanks are also due.

[32] To be precise, the spatial correlation consistent standard error estimator is constructed as follows: First, we constructed a sequence of R-vectors of estimated orthogonality conditions by taking the cross-sectional averages of the products of the instruments and either the 2SLS or 3SLS residuals. We then applied the Newey–West estimator to this sequence of cross-sectional averages.

not constrain the feasibility of the estimator proposed here. Monte Carlo experiments demonstrate that the finite-sample properties of this estimator are quite good, and are superior to those of other commonly used techniques. We demonstrated the empirical relevance of corrections for spatial dependence through a number of examples in which spatial correlation consistent standard errors differ substantially from those obtained using conventional methods.

Finally, we note that this paper has relied exclusively on large T asymptotics to deliver consistent covariance matrix estimates in the presence of cross-unit correlations. However, when T is small or when there is only a single cross section, the problem of consistent nonparametric covariance matrix estimation appears to be much less tractable. The reason for this is that, unlike in the time dimension, there is no natural ordering in the cross-sectional dimension upon which to base mixing restrictions, and hence it is not possible to construct the pure cross-sectional analogs of nonparametric time-series covariance matrix estimators. Thus it would appear that consistent covariance matrix estimation in models of a single cross section with spatial correlations will have to continue to rely on some prior knowledge of the form of these spatial correlations.

REFERENCES

Amemiya, Takeshi, *Advanced Econometrics* (Cambridge, MA: Harvard University Press, 1985).
Andrews, Donald W. K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59 (May 1991), 817–858.
Andrews, Donald W. K., and J. Christopher Monahan, "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60 (July 1992), 953–966.
Anselin, Luc, *Spatial Econometrics* (Dordrecht: Kluwer, 1988).
Basu, Susanto, "Procyclical Productivity: Increasing Returns or Cyclical Utilization,?" *Quarterly Journal of Economics* 111, (Aug. 1996), 719–752.
Basu, Susanto, and J. G. Fernald, "Aggregate Productivity and the Productivity of Aggregates," NBER Working Paper 5382 (1995).
Bulinskii, A. V., "On Various Conditions for Mixing and Asymptotic Normality of Random Fields," *Soviet Math. Dokl.* 37:2 (1988) 443–448.
Burnside, Craig, "Production Function Regression, Returns to Scale, and Externalities," *Journal of Monetary Economics* 37 (Apr. 1996), 177–201.
Caballero, Ricardo, and Richard Lyons, "The Role of External Economies in U.S. Manufacturing," NBER Working Paper 3033 (1989).
——— "External Effects in U.S. Procyclical Productivity," *Journal of Monetary Economics* 29 (Apr. 1992), 209–226.
Case, Anne C., "Spatial Patterns in Household Demand," *Econometrica* 59 (July 1991), 953–965.
Conley, Timothy G., "Econometric Modelling of Cross Sectional Dependence," manuscript, University of Chicago (1994).
Doukhan, Paul, *Mixing: Properties and Examples* (New York: Springer, 1994).
Elliott, Graham, "Spatial Correlations and Cross-Country Regressions," manuscript, Harvard University (1993).
Frankel, Jeffrey A., and Andrew K. Rose, "A Panel Project on Purchasing Power Parity: Mean Reversion within and between Countries," *Journal of International Economics* 40 (Feb. 1996), 209–224.
Frees, Edward W., "Assessing Cross-Sectional Correlation in Panel Data," *Journal of Econometrics* 69 (Oct. 1995), 393–414.
Froot, Kenneth A., "Consistent Covariance Matrix Estimation with Cross-Sectional Dependence and Heteroskedasticity in Financial Data," *Journal of Financial and Quantitative Analysis* 24 (Sept. 1989), 333–355.

Froot, Kenneth A., and K. Rogoff, "Perspectives on PPP and Long-Run Real Exchange Rates," in Gene Grossman and Kenneth Rogoff (eds.), *The Handbook of International Economics*, vol. 3 (Amsterdam: North Holland, 1995).
Hall, Robert E., "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy* 96 (Oct. 1988), 921–947.
——— "Invariance Properties of Solow's Productivity Residual," in Peter Diamond (ed.), *Growth/Productivity/Employment* (Cambridge, MA: MIT Press), 1990, 71–112.
Hamilton, James D., *Time Series Analysis* (Princeton, NJ: Princeton University Press, 1994).
Hansen, Lars P., "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* 50 (July 1982), 1029–1054.
Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee, *The Theory and Practice of Econometrics* (New York. Wiley, 1985).
Keane, Michael, and David Runkle, "Testing the Rationality of Price Forecasts: Evidence from Panel Data," *American Economic Review* 80 (Sept. 1990), 715–735.
Kraay, Aart, and Jaume Ventura, "Trade and Fluctuations," manuscript, World Bank and MIT (1997).
Lehmann, Bruce, "Residual Risk Revisited," *Journal of Econometrics* 45(July–Aug. 1990), 71–97.
Lewis, Karen, "What Can Explain the Apparent Lack of International Consumption Risk Sharing,?" *Journal of Political Economy* 104 (Apr. 1996), 267–297.
McLeish, Donald L., "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability* 3 (1975), 826–836.
Newey, Whitney K., and Kenneth D. West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (May 1987), 703–708.
O'Connell, Paul, "The Overvaluation of Purchasing Power Parity," *Journal of International Economics* 44 (Feb. 1998), 1–19.
Quah, Danny, "International Patterns of Growth: Persistence in Cross-Country Disparities," manuscript, MIT, Department of Economics (1990).
——— "Exploiting Cross-Section Variation for Unit Root Inference in Dynamic Data," *Economics Letters* 44:1–2 (1994), 9–19.
Rosenblatt, M., "Central Limit Theorem for Stationary Processes," in *Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2 (1970), 51–561.
Wei, S. J., and D. Parsley, "Purchasing Power Disparity during the Floating Rate Period: Exchange Rate Volatility, Trade Barriers and Other Culprits," NBER Working Paper 5032 (1995).
White, Halbert, "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica* 44 (Apr. 1980), 817–838.
——— *Asymptotic Theory for Econometricians* (New York: Academic Press, 1984).
Wooldridge, Jeffrey M., and Halbert White, "Some Invariance Principles and Central Limit Theorems for Dependent Heterogenous Processes," *Econometric Theory* 4 (1988), 210–230.
Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias," *Journal of the American Statistical Association* 57 (1962), 348–368.

APPENDIX

Proof of Result 1

The proof is simply a matter of verifying that h_t satisfies the definition of univariate mixing. Define $B_t = \{v | v < t\} \in Z^1$ as the natural one-dimensional analog of A_t, and similarly $\mathscr{F}^t_{-\infty} \equiv \sigma(\epsilon_z | z \in B_t)$ and $\mathscr{F}^\infty_{t+s} \equiv \sigma(\epsilon_z | z \in B^C_{t+s})$. Define the mixing coefficients for the sequence ϵ_t as $\alpha^h(s) = \sup_{(t)} \sup_{(G_1 \in \mathscr{F}^t_{-\infty}, G_2 \in \mathscr{F}^\infty_{t+s})} |P[G_1 \cap G_2] - P[G_1]P[G_2]|$. Now we claim that $\mathscr{F}^t_{-\infty} \subseteq \mathscr{F}^t_{-\infty}$ and $\mathscr{F}^\infty_{t+s} \subseteq \mathscr{F}^\infty_{t+s}$. Given this claim, we have $\alpha^h(s) \leq \alpha(s)$ for all s, and hence $\alpha^h(s)$ converges at least as quickly as $\alpha(s)$. Thus the sequence h_t is mixing of the same size of h_{it}.

To verify the claim, note that $h_t: \Omega \rightarrow R^1$ is a Borel function of $\{h_{it} | i = 1, \ldots, N, \ldots\}$, and hence is $\sigma(h_{it} | i = 1, \ldots, N, \ldots)$-measurable, that is, $h_{it}^{-1}(\mathscr{B}) \subset \sigma(h_t | i = 1, \ldots, N, \ldots)$, where $\mathscr{B}$ is the σ-algebra generated by the Borel sets. Thus by definition $\sigma(h_t) = \sigma(h_t^{-1}(\beta)) \subset$

$\sigma(h_{it}| i = 1, \ldots, N, \ldots)$. Finally, note that $\mathscr{S}^t_{-\infty} = \mathbf{x}^t_{v=-\infty} \sigma(h_v)$ and $\mathscr{F}^t_{-\infty} = \mathbf{x}^t_{v=-\infty} \sigma(h_{iv}| i = 1, \ldots, N, \ldots)$, and so the claim is verified.

Consistency and Asymptotic Normality of the GMM Estimator

We first note that the assumptions of result 2 are sufficient to establish consistency of the GMM estimator. The proof consists of verifying standard conditions for the consistency of extremum estimators such as those in Amemiya (1985, theorem 4.1.1). Conditions A and B of Amemiya (1985, theorem 4.1.1) follow immediately from assumption (a) of our result 2. Condition C of this theorem requires the minimand in the GMM problem to satisfy a law of large numbers as $T \to \infty$. By result 1 in this paper, $h_t(\theta)$ is a univariate α-mixing sequence of size $2r/(r - 1) > r/(r - 1)$. Thus to apply the McLeish (1975) LLN (see White (1984, theorem 3.47) for α-mixing sequences of size $r/(r - 1)$, we need only verify that $h_t(\theta)$ has finite $(r + \delta)^{\text{th}}$ moments. This follows immediately from assumption (1), which imposes the stronger condition (required for consistency of the covariance matrix estimator) that $h_{it}(\theta)$ has finite $4(r + \delta)^{\text{th}}$ moments, and the fact that $h_t(\theta)$ is the sum of $N(T)$ random variables normalized by $N(T)$. Thus the GMM estimator is consistent as $T \to \infty$.

To establish asymptotic normality of the GMM estimator, we need only verify standard conditions such as those in Hamilton (1994, proposition 14.1) (which are in turn adapted from Hansen (1982)). Condition (a) of Hamilton (1994, proposition 14.1) follows from the consistency of the GMM estimator. To verify condition (b) of Hamilton (1994, proposition 14.1), we need to show that the sequence $h_t(\theta)$ satisfies a central limit theorem for dependent processes such as White (1984, theorem 5.19). This theorem requires $h_t(\theta)$ to be an α-mixing sequence of size $r/(r - 1)$ with finite $2r$th moments. Both these conditions have been verified. In addition,

this central limit theorem requires the assumption of asymptotic covariance stationarity, Var $[T^{-1/2}\Sigma^{a+T}_{t=a+1}h_t(\theta)]$ $V \to 0$ uniformly in a as $T \to \infty$ for V positive definite. If we make this assumption, together with the regularity condition on the partial derivatives of $h_{it}(\theta)$ in condition (c) of Hamilton (1994, proposition 14.1), asymptotic normality of the GMM estimator as $T \to \infty$ is immediate.

One remark on the covariance stationarity assumption required for asymptotic normality is in order. If the cross-sectional dependence is sufficiently weak and if $N(T)$ grows quickly enough relative to T, the condition that V be positive definite may not hold. To see this, note that

$$\text{Var}\left[T^{-1/2} \sum_{t=a+1}^{a+T} h_t(\theta)\right] = \frac{1}{T} \sum_{t=a+1}^{a+T} \sum_{s=a+1}^{a-T} \left(\frac{1}{N(T)^2} \sum_{i=1}^{N(T)} \sum_{j=1}^{N(T)} E[h_{it}(\theta)h_{js}(\theta)]\right).$$

If the cross-sectional dependence is weak in the sense that the inner double sum over the cross-sections of the cross-sectional covariances is of order less than $N(T)^2$, say, for example, $N(T)^{2-2\epsilon}$, and if $N(T)$ grows fast enough relative to T, then it is possible that $V = 0$ since the term in the parentheses vanishes as $N(T)$ becomes large. However, this point need not detain us since it simply amounts to an observation that if the cross-sectional dependence is not too strong, the GMM estimator converges in distribution at a rate faster than $T^{1/2}$. In particular, if we replace $h_t(\theta)$ with $g_t(\theta) = N(T)^\epsilon h_t(\theta)$ in the preceding arguments, asymptotic normality of the GMM estimator is immediate. Thus the estimator is asymptotically normal at least at rate $T^{1/2}$, and the weaker the cross-sectional dependence (i.e., the larger is ϵ), the faster is the rate of convergence. The proofs of consistency of the GMM estimator and consistency of the covariance matrix estimator are unaffected by this point.