Urban and Regional
Analysis Group

AN ALTERNATIVE ESTIMATOR FOR FIXED EFFECTS SPATIAL MODELS

Mary G. McGarvey
Mary Beth Walker

Working Paper No. 03-03
August 2003

Georgia State University

Andrew Young
School of Policy Studies

AN ALTERNATIVE ESTIMATOR FOR FIXED EFFECTS SPATIAL MODELS

Mary G. McGarvey
(email: mmcgarvey@unl.edu)
University of Nebraska-Lincoln

Mary Beth Walker
(email: mbwalker@gsu.edu)
Georgia State University

Key Words and Phrases: spatial correlation; fixed effect model; spatial correlation consistent covariance estimation; monte carlo evidence

JEL Classification: C33; C15

ABSTRACT

This paper examines inference in finite samples based on estimation using maximum likelihood and ordinary least-squares estimators in fixed effect models with spatially autocorrelated errors. The focus of the Monte Carlo experiments is on the effect of covariance misspecification on finite sample inference. We find that the maximum likelihood estimator of the conditional mean parameters is more efficient than OLS under most populations but the quasi-t statistics are most reliable using OLS with a spatial correlation consistent covariance estimator.

1. Introduction

It is well known that inference can be flawed when the data are characterized by spatial correlation. The increased use of spatial estimators in applications in urban and regional and environmental economics suggests that researchers are concerned about spatial correlation and recognize the need to incorporate it in estimation. By far, the most common method of handling spatially correlated errors is to use the first-order model attributed to Cliff and Ord (1973) and popularized by Anselin (1988).

In this paper, we evaluate an alternative estimation strategy for allowing for a spatially correlated error structure that does not require the specification of a particular spatial model. We suggest using a GMM estimator of the conditional mean parameters with a spatial correlation consistent (SCC) covariance estimator. This strategy is possible when the researcher has panel data available. Our primary interest is how inference is affected by the use of the alternative estimation strategy, although we also examine relative efficiency issues.

The behavior of the quasi-*t* statistics used to conduct inference depends on the covariance estimator used to construct the test statistic. To provide information on this, we compare finite sample performance of quasi-*t* statistics based on maximum likelihood estimation of a first-order model to quasi-t statistics based on OLS estimation using two SCC estimators. These SCC estimators are special cases of Driscoll and Kraay's (1998) covariance estimator. For comparison purposes, we also include the standard OLS covariance estimator and a heteroskedastic consistent estimator. We consider different patterns of spatial correlation in addition to the first-order spatial model.

We compare the ML and OLS coefficient estimators' efficiency through an examination of relative finite sample bias and variance. In general, we find that both estimators are unbiased but the maximum likelihood estimator tends to have a smaller variance. Even when the covariance structure incorporated in the likelihood function is incorrect, we find the maximum likelihood estimator for the conditional mean parameters is usually more efficient than OLS.[1]

Overall, our results indicate that least-squares estimators of conditional mean parameters combined with a robust covariance matrix estimator provide a reasonable alternative to maximum likelihood estimation of a first-order spatial model. In some cases where the likelihood function is misspecified, the least-squares strategy is far superior.

II. The Model and Estimators

The basic model that we consider is,

$$y_t = X_t\beta + u_t, \qquad t = 1, 2, ..., T \tag{1}$$

where y_t is an $(n\times 1)$ vector and X_t is an $(n\times k)$ matrix containing the t^{th} observation on the dependent and explanatory variables, and u_t represents the error vector for a panel of n units. In many applications, X_t will contain fixed effects dummy variables. Often it is reasonable to assume that the errors are uncorrelated over time, but correlated across units. Because our focus in this paper is spatial autocorrelation, we ignore any potential error correlation over time. In this case, the model's assumptions can be written as:

$$E(u_t \mid X_t) = 0,\ \ E(u_t u_s' \mid X_t) = \Omega_t,\ t = s \text{ and } E(u_t u_s' \mid X_t) = 0,\ t \neq s. \tag{2}$$

The asymptotic results are based on both n and T approaching infinity.

We consider two estimation strategies. First, with the additional assumptions that $plim\left(\frac{X'X}{nT}\right)$ and $plim\left(\frac{X'X}{nT}\right)$ are finite and non-singular, one can consistently estimate the conditional mean vector by OLS and base inference on the limiting distribution,

$$\sqrt{nT}\,(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, plim\left(\frac{X'X}{nT}\right)^{-1} plim\left(\sum_{t=1}^{T}\frac{X_t'\Omega_t X_t}{nT}\right) plim\left(\frac{X'X}{nT}\right)^{-1}), \tag{3}$$

[1] We examine the behavior of quasi-t statistics using the *estimated covariance*, as would be the practice in any applied work. It is the *true* covariances of the OLS and ML estimators that are used to assess relative efficiency.

where $\hat{\beta}_{OLS}$ is the OLS estimator and $X = (X_1', X_2',..., X_t')'$, the $(nT \times k)$ matrix of explanatory variables. With $\hat{u}_t$ defined as the vector of OLS residuals, one can use $\sum_{t=1}^{T} \frac{X_t' \hat{u}_t \hat{u}_t' X_t}{nT}$ as an estimator of the middle term in the covariance matrix and the resulting spatial correlation consistent (SCC) covariance estimator,

$$SCC1 = \left(\frac{X'X}{nT} \right)^{-1} \sum_{t=1}^{T} \frac{X_t' \hat{u}_t \hat{u}_t' X_t}{nT} \left(\frac{X'X}{nT} \right)^{-1}, \tag{4}$$

is a special case of that proposed by Driscoll and Kraay (1998).[2]

An alternative estimation scheme is to maintain equation (1) and specify the error as a first-order process. This can be written as:

$$u_t = \rho W u_t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2 I_n), \tag{5}$$

which implies that $E(u_t u_s') = \sigma_\varepsilon^2 (I_n - \rho W)^{-1} (I_n - \rho W)^{-1\prime}, t = s, E(u_t u_s') = 0, t \neq s.$ The diagonal elements of the nonstochastic weighting matrix *W* are zero and the off-diagonal elements, w_{ij} , are chosen to reflect the degree of dependence between the error of unit *i* and the error of unit *j*. With the weighting matrix chosen *a priori*, estimation has typically proceeded by assuming a normal distribution and maximizing the appropriate likelihood function.[3] Asymptotic inference is based on

$$\sqrt{nT}(\hat{\beta}_{MLE} - \beta) \xrightarrow{d} N(0, V_{MLE}), \tag{6}$$

where $\hat{\beta}_{MLE}$ is the maximum likelihood estimator and V_{MLE} is the relevant $(k \times k)$ submatrix of the inverse of the information matrix corresponding to the estimators of β. In the event that the errors do not follow a normal distribution, this estimator remains consistent as a quasi-maximum likelihood estimator. Asymptotic inference based on the MLE estimator is asymptotically efficient under the first-order spatial error structure of equation (5) but the covariance estimator is inconsistent under more general specifications such as equation (2).

The general stochastic specification of equation (2) allows for error correlation across units and an unknown pattern of heteroskedasticity for each time period as well as

[2] The SCC covariance estimator proposed by Driscoll and Kraay allows the model's errors to be correlated over time as well as across units.

over time. Although asymptotic inference based on the OLS covariance estimator in (4) is valid under this specification, we might expect that this level of generality would extract a cost in terms of finite sample inference, based on simulation evidence of other heteroskedastic consistent covariance estimators.

One way to restrict the stochastic specification and yet retain the general form of the contemporaneous error covariance within units is to impose a homoskedastic structure over time.[4] With this restriction, a consistent estimator of the middle matrix in equation (3) replaces Ω_t with $\hat{\Omega}$, where $\hat{\Omega} = \sum_{t=1} \hat{u}_t \hat{u}_t' / T$.[5] The resulting spatial correlation consistent (SCC) covariance estimator is,

$$SCC2 = \left(\frac{X'X}{nT}\right)^{-1} \sum_{t=1}^{T} \frac{X_t' \hat{\Omega} X_t}{nT} \left(\frac{X'X}{nT}\right)^{-1}, \qquad (7)$$

a more restrictive version of *SCC1*.

The *SCC* covariance matrices given in equations (4) or (7) do not appear frequently in the applied literature estimating fixed effect models with panel data. One explanation is that robust covariance estimators often result in incorrect finite sample inference. We address this concern by comparing the performance of OLS test statistics based on the SCC covariance estimators to those more commonly used in applied work.

III. Simulation and Results

The data and model we use for our simulations are very loosely based on those of Case, Rosen and Hines (1993) who estimate direct per capita expenditures by state and local governments as a function of state economic data and population demographics. Using a panel of 48 states over the years 1970 through 1985, Case, Rosen and Hines find

---

[3] See Kelejian and Prucha (1999) for an alternative GMM estimator for the first-order model.

[4] Note that this specification corresponds to the textbook Seemingly Unrelated Regression model with the coefficient vector restricted to be equal across units.

[5] Although $\hat{\Omega}$ will be singular in cases where $n > T$ and thus of no use as a weighting matrix for GLS, $X'(I_T \otimes \hat{\Omega})X / nT$ will be of rank k for well-behaved X and will converge to $plim[X'(I_T \otimes \Omega)X / nT]$ making inference based on OLS a feasible alternative.

strong evidence of spatial correlation. Our data set extends the sample period through 1995, so that there are now 26 times-series observations. Two of the weight matrices suggested in their work are adopted for our simulations.

Our simulations are based on the population model:

$$E(exp_{it} \mid X) = X_{it}\beta, \quad i = 1, 2, ..., n; \quad t = 1, 2, ..., T, \tag{8}$$

where β is set to zero and the matrix X_{it} includes fixed effects dummy variables, per capita real income and the proportion of the population that is black.

We simulate the behavior of the OLS quasi-t statistic, $\hat{t} = \hat{\beta}_1 / se(\hat{\beta}_1)$, using different estimates of the standard error. In addition to the SCC estimators, we also compute quasi-t statistics using covariance estimators often used in applications. These include the OLS covariance estimator under the assumption $u_t \sim iid(0, \sigma_u^2 I_n)$; this covariance estimator will be referred to as 'Standard OLS.' We include both White's (1980) usual heteroskedastic-consistent covariance estimator that allows the individual error variances to differ across units and possibly over time and a more restricted version of White's estimator that assumes constant error variances over time.

We compare these OLS quasi-t statistics to those based on maximum likelihood estimation of the first-order spatial correlation model. These estimation methods, covariance matrix estimators and the error structure assumed by each are summarized in Table 1.

The maximum likelihood estimator for the model given by equations (1) and (5) is computed using two weight matrices, one is based on comparisons of real per capita income, the other based on similar proportions of African-Americans in the state populations. The quasi-t statistic obtained using estimated standard errors from this estimator is denoted MLE.

We generate errors from two spatially correlated populations. Our first error structure is,

$$u_t = \rho W u_t + \varepsilon_t, \qquad \rho = 0, \pm.125, .250, .375, .500, .625, .750, .875, .950 \tag{9}$$

$$\varepsilon_t \sim N(0, I_n), \quad t = 1, 2, \ldots, T.$$

When $\rho = 0$, the error reduces to $u_{it} \sim iid\ N(0,1)$. In this case the usual OLS covariance matrix estimator is efficient. The other covariance estimators, as well as the ML covariance estimator, are consistent, but inefficient in the case of *iid* errors. Thus, under this error specification, all quasi-t statistics converge to standard normal distributions in the limit.

We also employ a different spatial correlation scheme that was suggested by Driscoll and Kraay:

$$u_t = \lambda v_t + \varepsilon_t, \quad (v_t, \varepsilon_t')' \sim iid\ N(0, \Sigma), \tag{10}$$

where λ is an $(n \times 1)$ vector of parameters, v_t is a scalar *iid* shock common to each unit, and the elements of ε_t are idiosyncratic errors uncorrelated with v_t, each other and over time. Following Driscoll and Kraay, we normalize the errors to have unit variances so that the correlation of u_i and u_j is $\lambda_i \lambda_j$. We generate samples from ten populations by drawing the value of λ from $\lambda_i \sim iid\ U(\ell, \ell + .10)$, where $\ell = 0, .1, .2, ..., .9$, so that the mean correlation of states' errors ranges from $E(\lambda_i)E(\lambda_j)$=.0025 to $E(\lambda_i)E(\lambda_j)$=.9025.

With the spatially correlated populations (9) and (10), the covariance estimator constructed under the incorrect assumption of *iid* errors (standard OLS), and the two White covariance estimators are inconsistent. Those constructed under the more general error specification assumptions (SCC1 and SCC2) are, of course, consistent but inefficient estimators of the true covariance matrix. Thus, under these populations, only the quasi-t statistics based on SCC1 and SCC2 converge to a standard normal random variable. In the case of the first-order spatial process, the maximum likelihood estimator of β, when based on the true weighting matrix, is asymptotically efficient, its estimated covariance matrix is consistent, and the MLE quasi-t statistic converges to a standard normal random variable. For the process in (10), the maximum likelihood estimator of β remains consistent, but its estimated covariance matrix is inconsistent.

Figure 1 shows the ratio of actual test size to a nominal size of .05 for test statistics for various values of ρ, using error specification (9). Because the two versions of the White and SCC estimators produce similar results, we include only White 1 and

SCC2 in our figures.[6] Ratios less than 1 indicate that the test statistic in the finite sample is rejecting too few times, on average, under the null, suggesting that the standard error estimates are, on average, larger than they should be. Ratios greater than 1 indicate over-rejection and suggest that standard errors are being under-estimated.

The graph shows that the standard OLS covariance estimator is very unreliable when errors are spatially correlated. Interestingly, test statistics based on standard OLS under-reject when ρ is negative and over-reject when ρ is positive. The magnitude of bias in the estimated standard error increases with the absolute value of ρ.

Test statistics based on the SCC covariance estimators over-reject the true null regardless of the value of ρ. The more restricted estimator, SCC2, produces test sizes that more closely conform to the nominal values than SCC1. This is reasonable because SCC2 imposes the correct restriction that the correlations do not change over time.

The White covariance estimators behave very similarly to the standard OLS estimator for non-zero values of ρ. Test statistics under-reject when ρ is negative and over-reject when ρ is positive with the absolute bias in the standard error estimate increasing with the degree of spatial correlation. With $\rho = 0$, standard OLS outperforms both White estimators while the White estimators produce results that conform more closely to nominal values than either SCC estimator.

Figure 2 presents results under the spatial correlation population described in (10). The ratio of actual to nominal size is graphed against $\bar{\lambda}$, the mean of the distribution of λ_i, so that the mean error correlation is $\bar{\lambda}^2$. When the mean correlation is only .0025 ($\bar{\lambda} = .05$), all of the test statistics slightly over-reject the true null except for standard OLS. As the degree of correlation increases, the over-rejection rates of test statistics based on standard OLS and White increase dramatically. In a spatially correlated population with mean correlation $>$ .30 ($\bar{\lambda} \geq .55$), test statistics based on standard OLS or White covariance estimators reject a true null over 50% of the time. Although the rejection rates for MLE test statistic are not as high as those based on standard OLS or White, they are still far greater than their nominal size and appear to vary with the degree

[6] Complete results are available from the authors on request.

of spatial correlation. The performance of the SCC estimator remains stable over all values of $\bar{\lambda}$: the test statistic rejects a true null about 6% of the time.

A third experiment confirms the conventional belief that estimates are sensitive to the weight matrix chosen. Although the sample was generated using a weight matrix based on similar proportions of blacks in state populations, the MLE's were obtained using weights based on similar levels of per capita income. Figure 3 shows that inference based on the MLE test statistics suffers from the same pattern of size distortions as inference based on OLS using the standard OLS or White covariance estimators: under-rejection in populations with negatively correlated errors and over-rejection in populations with positively correlated errors.

A natural concern in using the OLS strategy is the possibility of a substantial loss of efficiency in the estimates of conditional mean parameters. Applied researchers who suspect a first-order spatial correlation scheme in their data may believe that the increase in efficiency using the MLE estimator outweighs the possible inconsistency in the covariance matrix estimators if the weight matrix is incorrectly chosen. To provide some guidance on this issue, we compute bias, variance and mean square error for ML estimates of β_1 relative to OLS estimates of the same parameter for each of our populations.

Table 2 indicates that both the MLE and OLS estimators are approximately unbiased as is evident from comparing the relative variances of the two estimators to their relative mean squared errors. The relative variances indicate that the MLE is almost always a more efficient estimator of conditional mean parameters than OLS.[7] These results might entice the applied researcher into believing that because one's primary interest is in estimates of the conditional mean parameters, the use of maximum likelihood even under a misspecified first-order spatial correlation process is justified. Notice, though, that these relative variance ratios refer to the estimators' true variances. It is the *estimated* MLE variances and standard errors that are used to construct test statistics. Our comparison of rejection frequencies of OLS test statistics to MLE test

[7] This result continues to hold when the population errors are non-normal. Because our Monte Carlo results are essentially unchanged with non-normal errors we do not report these results.

statistics shows that OLS test statistics using an SCC covariance estimator perform nearly as well as a correctly-specified-MLE test statistics under a first-order spatial population and often perform much better under other spatially correlated populations or when the MLE weight matrix is incorrect.

IV. Conclusions

In this paper, we examined the effect of covariance misspecification on inference in fixed effects models with spatially correlated errors. We evaluated the finite sample performance of maximum likelihood and OLS estimation using Monte Carlo experiments under a variety of spatially autocorrelated populations. The Monte Carlo evidence indicated that for models with spatially correlated errors, robust covariance matrix estimators can be used very successfully. These estimators provide consistent estimates of the variances of the conditional mean estimates without imposing any functional or distributional assumptions on the form of the spatial correlation.

Although maximum likelihood produced relatively more efficient conditional mean estimators than OLS even under misspecified covariance structures, we found that finite sample inference was most reliable using OLS with an SCC covariance estimator. The estimated rejection frequencies using maximum likelihood are close to their nominal levels under most populations but are far from reliable under certain types of misspecification. OLS test statistics based on either the standard OLS covariance estimator or White's heteroskedastic consistent estimator performed very poorly in most populations. In populations with positive spatial correlation, the nominal sizes understated the true sizes associated with these test statistics and in populations with negative spatial correlation, the opposite occurred.

Based on these results we recommend that researchers consider using OLS to estimate the conditional mean parameters of panel data models with spatially correlated errors and base inference on an SCC covariance estimator. Over all the populations we considered, rejection frequencies associated with these OLS quasi-t statistics were slightly larger than their nominal sizes. This deviation was roughly constant across all populations, however, suggesting that inference could be improved easily using a simple size correction.

References

Anselin, Luc. (1988), ***Spatial Econometrics: Methods and Models***, (Dordrecht: Kluwer Academic).

Anselin, Luc, Anil K. Bera, Raymond Florax and Mann J. Yoon (1996), "Simple Diagnostic Tests for Spatial Dependence," ***Regional Science and Urban Economics***, 26, 77-104.

Baltagi, and Lee (2003), "Prediction in the Panel Data Model with Spatial Correlation," in New Advances in Spatial Econometrics, ed. by Luc Anselin and Raymond Florax, Springer-Verlag, forthcoming.

Besley, T. and A. Case, (1995), "Incumbent Behavior: Vote-seeking, Tax-setting, and Yardstick Competition," ***American Economic Review***, 85, 25-45.

Brueckner, Jan K. (1998), "Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls," ***Journal of Urban Economics***, 44, 438-467.

Case, Anne C. (1991), "Spatial Patterns in Household Demand," ***Econometrica***, 59, 953-966.

Case, Anne C., Harvey S. Rosen and James R. Hines, Jr. (1993), "Budget Spillovers and Fiscal Policy Interdependence," ***Journal of Public Economics***, 52, 285-307.

Chen, Xiaohong and Timothy G. Conley (2000), "A New Semiparametric Spatial Model for Panel Time Series," ***Journal of Econometrics***, forthcoming.

Cliff, A., and J. Ord, 1973, ***Spatial Autocorrelation***, (Pion, London).

Conley, Timothy G. (1999), "GMM estimation with cross sectional dependence," ***Journal of Econometrics***, 92, 1-45.

Driscoll, John C. and Aart C. Kraay (1998), "Consistent Covariance Matrix Estimation with Spatially Dependent Data," ***The Review of Economics and Statistics***, 80, 549-560.

Kelejian, Harry and Ingmar Prucha (1999), "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," ***International Economic Review***, 40, 509-533.

White, H. (1980),"Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," ***Econometrica,*** 48, 817-838.

Figure 1: Population- First-order Spatial Correlation

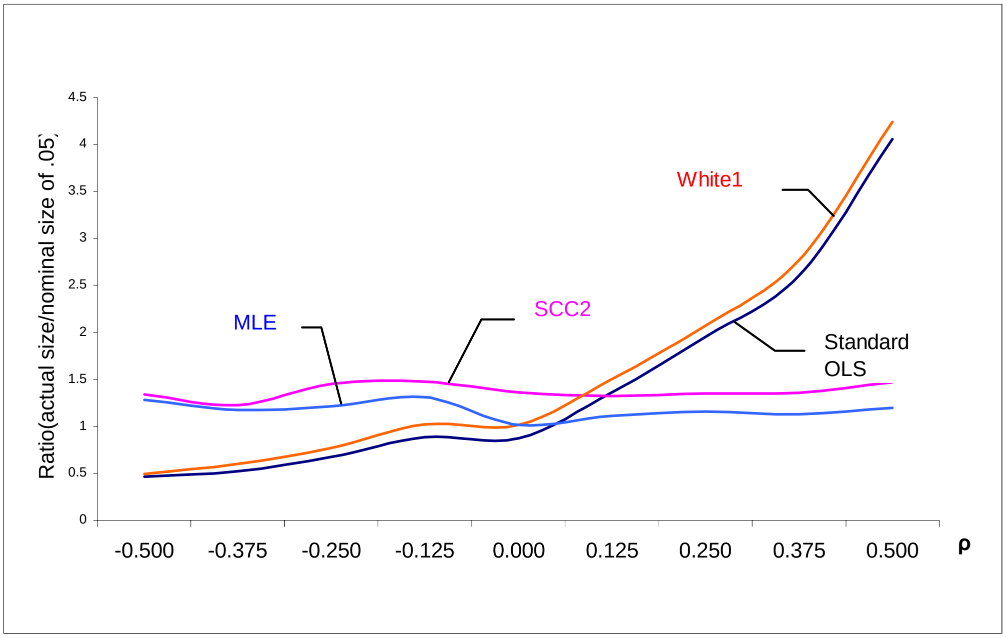

Figure 2: Population-General Spatial Model

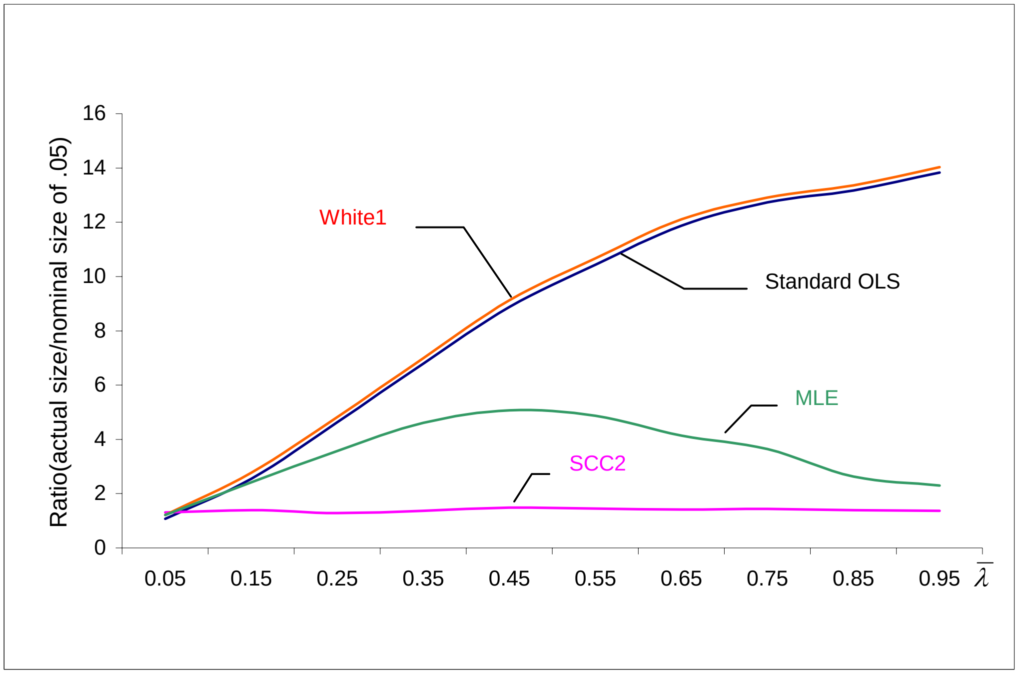

Figure 3: Population-First-order Spatial Correlation
MLE uses wrong weight matrix

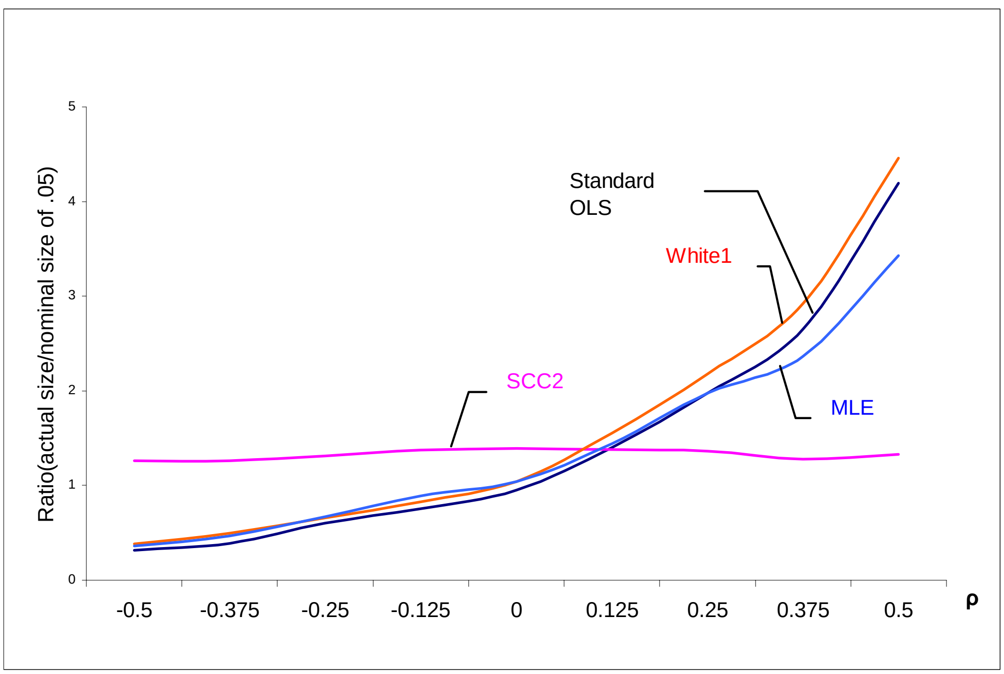

Table 1: Estimators and Maintained Hypotheses

$$\text{Model}: \quad y_t = X_t \beta + u_t$$

Estimation Method	**Covariance Estimator**	**Assumed Covariance Structure**
OLS	Iid	$E(u_t u_s') = \sigma^2 I_n, \ t = s$ $= 0, \quad t \neq s$
OLS	White1	$E(u_{ti} u_{sj}) = \sigma_{ti}^2, \ t = s, i = j$ $= 0, \quad t \neq s$ or $i \neq j$
OLS	White2	$E(u_{ti} u_{sj}) = \sigma_i^2, \ t = s, i = j$ $= 0, \quad t \neq s$ or $i \neq j$
OLS	SCC1	$E(u_t u_s') = \Omega_t, \ t = s$ $= 0, \quad t \neq s$
OLS	SCC2	$E(u_t u_s') = \Omega, \ t = s$ $= 0, \quad t \neq s$
MLE	MLE	$E(u_t u_s') = \sigma_\varepsilon^2 (I_n - \rho W)^{-1} (I_n - \rho W)^{-1\prime}$ $= 0, \quad t \neq s$

Table 2: Relative Bias, Variance, and MSE
MLE relative to OLS

Population	Rho	Bias	Variance	MSE
First-order Spatial Correlation	0.0000	1.0000	1.0000	1.0000
	0.1250	1.0000	0.9231	0.9230
	0.5000	0.8333	0.7333	0.7326
	0.8750	0.3462	0.0727	0.0728
	-0.1250	1.0000	1.0000	1.0000
	-0.5000	0.6000	0.8000	0.8000
	-0.8750	-0.2500	0.5000	0.5000
General Spatial Correlation	$\bar{\lambda}$ =0.15	1.0000	1.0000	1.0000
	$\bar{\lambda}$ =0.45	1.5000	0.6216	0.6216
	$\bar{\lambda}$ =0.75	0.6000	0.1705	0.1705
	$\bar{\lambda}$ =0.95	-3.0000	0.0217	0.0217
Wrong Weight Matrix	0.5000	1.2500	0.9375	0.9375