Performance of the FGS3SLS Estimator in Small Samples: A Monte Carlo Study

Saket Sarraf, Kieran P. Donaghy, Geoffrey J.D. Hewings

REAL 13-T-02
May, 2013

Performance of the FGS3SLS Estimator in Small Samples: A Monte Carlo Study

Saket Sarraf
Affiliate Research Assistant Professor, Regional Economics Applications Laboratory, University of Illinois, Urbana, IL 61801, USA, and Principal, ps Collective, A-52, Galaxy Tower, Nr Hotel Grand Bhagwati, Bodakdev, Ahmedabad 380054, India
saket.sarraf@gmail.com

Kieran P. Donaghy
Department of City and Regional Planning, Cornell University, Ithaca, NY, 14853
kpd23@cornell.edu

Geoffrey Hewings
Regional Economics Applications Laboratory, University of Illinois, Urbana, IL 61801, USA
hewings@illinois.edu

Abstract: System of equations models with spatial lags in dependent variable and error terms can be estimated using the full information Feasible Generalized Spatial Three Stages Least Square (FGS3SLS) estimator proposed by Kelejian & Prucha (2004). The estimator is consistent and asymptotically normal, but its finite sample properties are not analytically determinable. In absence of very large samples as is the case in most applied work, it is difficult to interpret the results with confidence based on asymptotic results only. This paper evaluates the performance of the FGS3SLS estimator in finite samples and its sensitivity to varying degrees of spatial interaction and externalities using Monte Carlo simulations.

Key words: FGS3SLS, Spatial interaction models, Monte Carlo simulation, Finite sample properties

JEL Classification: C13, C21, C31, R14

1. Introduction

The modeling of spatial processes has attained a mainstream position in social sciences (Goodchild *et al.*, 2000). Anselin (2010) presents an historical analysis of how spatial econometrics has attained a mainstream status in applied econometrics and social science methodology. In the simplest cases, the variables of interest are spatially correlated with their neighbors and with other variables. As we move from one variable to a system of variables, modeling the spatial interactions becomes complex. The complexity further increases as the randomness becomes correlated spatially and across equations. Modeling the strength of spatial interactions and externalities requires the specification and estimation of spatial econometric models. However, the available estimators (Anselin, 1988; Case, 1991; Case *et al.*, 1993) lack methodological sophistication and computational simplicity to accurately estimate simultaneous systems with spatial autoregressive dependent variables and spatially interrelated cross sectional equations. They are often based on quasi-maximum likelihood procedures and might not have feasible solutions in medium to large samples. Further, they are designed for single equation frameworks (See Kelejian & Prucha, 1999 for an extensive discussion on this issue).

To estimate models for such processes, Kelejian & Prucha (2004) proposed the limited information Feasible Generalized Spatial Two Stage (FGS2SLS) and full information Feasible Generalized Three Stage Estimators (FGS3SLS). These estimators are based on generalized methods of moments using approximation of optimal instruments, and thus are computationally simple. Kelejian and Prucha show that the estimators are consistent and asymptotically normal. Some of the applied examples of this estimator include Ngeleza *et al.* (2006) to determine the geographical and institutional determinants of real income, Driffield (2006) for modeling spatial spillovers of foreign direct investment, Fishback *et al.* (2006) for modeling the impact of New Deal expenditures on mobility during the great depression. More recent applications includes applications in the fields of assessing regional growth (Gebremariam *et al.* 2012)…..

It is important to understand how this estimator behaves in applied studies given its relevance in estimating many of the complex spatial processes that have been largely ignored thus far. However, our understanding of this estimator is at best, rudimentary. The number of publications using this estimator is relatively few, and only its asymptotic properties have been established so far. In absence of very large samples, as is the case in much applied work, it is difficult to interpret the results with confidence based on asymptotic results only.

One alternative to employ in a situation such as this is to use finite sample approximations or asymptotic expansions. However, these approximations tend to be very complex, the results difficult to interpret and the computations very advanced. Some early work on this topic is summarized in Philips (1983) and Rothenberg (1984). In contrast, the method of Monte Carlo replaces the skills needed in asymptotic approximations by relying on computational power of computers. In this paper, the properties of the parameters of interest are studied through a series of stochastic simulations and their statistics are analyzed (Davidson & MacKinnon, 1993).

This paper investigates the performance of the FGS3SLS estimator for a system of simultaneous equations, with spatial autoregressive dependent variables and spatially autocorrelated error structures using Monte Carlo experiments. Performance is measured by its ability to estimate parameters of the model and sensitivity of the results to varying degree of spatial dependences, choice of spatial weight matrix, sample size and variance covariance matrices. The paper concludes by emphasizing the need for further studies on the subject to increase our understanding of the estimator's behavior in applied work.

The rest of the paper is organized as follows. Section 2 sets up the model used for the study. Section 3 briefly describes the estimator and section 4 describes the experimental design. The results of the simulation exercise are presented in section 5. Section 6 summarizes the main findings and concludes the study with direction for future research.

2. Model Structure

2.1 Formal Considerations

The performance of the FGS3SLS estimator was tested using a model specification closely resembling the structure of the model used in Sarraf (2012) to analyze the regional social dynamics and its impacts on land-use change. The model used here consists of a system of simultaneous equations with two endogenous variables, their spatial and temporal lags and two exogenous variables. The spatial lag of the dependent variable is treated as endogenous while the temporal lag is considered as predetermined, since the model is conditioned on past values of the dependent variable. The disturbances are assumed to be correlated across space and across different equations. This form of model allows the analyst (a) to capture spatial processes like diffusion across space, (b) to address problems of ecological fallacy or presence of some local conditions leading to spatially correlated error structures, and (c) to determine the correlation between two spatial processes. Further, the specification allows forecasting of the value of dependent variables conditional on its own past values, and other exogenous variables after accounting for the underlying spatial processes.

Let y_1 represent percent abandoned housing units in a census tract and y_2 represent net in-migration of households. Equation 1 states that percent abandoned units y_1 depend on: (1) the magnitude of net in-migration of households (y_2) and percent housing abandonment in neighboring tracts (Wy_1) in the current period; (2) the percentage of the housing abandoned in the previous period (x_1); (3) the distance from interstate (x_3); and (4) a random component (u_1). Simultaneously, the net in-migration of households is endogenous and depends on the percentage of units abandoned since higher housing abandonment tends to repel more households from the region. According to the equation (2), the magnitude of net in migration of households (y_2) in a tract depends on: (1) the percentage of housing abandonment (y_1) and net in migration of households in the neighboring tracts (Wy_2) in the current period; (2) lagged values of net in-migration (x_2); (3) the condition of infrastructure (x_4); and (4) a random component u_2. Thus, housing abandonment and net in migration of households are jointly determined. Note that x_3 is treated as fixed over time while x_4 is time dependent but still exogenous.

$$y_1 + \gamma_1 y_2 + \lambda_1 W_1 y_1 + \beta_1 x_1 + \beta_3 x_3 = u_1 \tag{1}$$

$$\gamma_2 y_1 + y_2 + \lambda_2 W_2 y_2 + \beta_2 x_2 + \beta_4 x_4 = u_2 \tag{2}$$

where, $y_i, (i = 1, 2)$ represents the endogenous variables we are interested to forecast. $W_i y_i$'s are the spatially lagged dependent variables with the spatial lag parameter λ_i. W_i is the row standardized weight matrix of known constants describing the neighborhood structure of observations. x_1 and x_2 are the temporally lagged values of dependent variable y_1 and y_2 respectively, x_3 and x_4 are the exogenously determined variables whose values either remain fixed throughout the simulation or are known *a priori*. u_1 and u_2 represent the stochastic component of the model whose behavior is elaborated below.

The disturbance vectors u_1 and u_2 in equations (1) and (2) are assumed to be correlated across space and across equations. The spatial geography over which the social dynamics are occurring is different from the administrative geography of census tracts. The aggregation of data at the tract level leads to correlation of disturbances across tracts. Further any randomness affecting housing abandonment and change in number of households may be correlated. Thus, the current specification allows for randomness that is also correlated across equations.

$$u_1 = \rho_1 W_3 u_1 + \varepsilon_1 \tag{3}$$

$$u_2 = \rho_2 W_4 u_2 + \varepsilon_2 \tag{4}$$

$$with, \quad \Sigma = Cov(\varepsilon_1, \varepsilon_2) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \tag{5}$$

Equations (3) and (4) characterize the correlation across space where $W_3 u_1$ and $W_4 u_2$ are average values of error terms in the neighboring locations, and ρ_1 and ρ_2 depict the degree of spatial correlation of the error terms. ε_1 and ε_2 are non-spatially correlated disturbances but are correlated across equations with the variance covariance matrix Σ (equation 5). This completes the specification of the hypothetical model.

2.2 Generalized form

For brevity, the model system represented in equations (1) to (5) can be rewritten in matrix notation as:

$$\begin{pmatrix} I_n + \lambda_1.W_1 & \gamma_1.I_n \\ \gamma_2.I_n & I_n + \lambda_2.W_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \beta_1.I_n & 0.I_n \\ 0.I_n & \beta_2.I_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \beta_3.I_n & 0.I_n \\ 0.I_n & \beta_4.I_n \end{pmatrix} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} (1-\rho_1 W_3)^{-1}.\varepsilon_1 \\ (1-\rho_2 W_4)^{-1}.\varepsilon_2 \end{pmatrix} \tag{6}$$

$$or, \quad B.Y + T_a.X_a + T_b.X_b = U \tag{7}$$

where, $Y = (y_1, y_2)$ is a vector of endogenous variables, $X_a = (x_1, x_2)$ is a vector of temporally lagged endogenous variable Y, $X_b = (x_3, x_4)$ is a vector of exogenous variables and I_n is an identity matrix of dimension n. B, T_a and T_b represent the coefficients associated with these variables in equation (6). $U = (u_1, u_2)$ represents the vector of disturbance terms. The estimator is described in Appendix C.

3. Monte Carlo Experiments

With the model structure in place, designing the Monte Carlo experiment consists of three additional parts, namely: defining the parameter settings; generating the spatio-temporal array of synthetic data for different variables consistent with the underlying spatial process; and designing the simulations to reduce errors due to randomization and analysis of alternative scenarios. Each of these steps is elaborated below.

3.1 Parameter settings

This section assigns values to the parameters used in the model specified in equations (1) through (5) including the values of all the coefficients, the variance covariance matrix of disturbance terms, the weight matrix and the spatial dependence parameters.

The parameters for the spatial lag of the dependent variable and for spatial autocorrelation in the error terms { λ_i , ρ_i } include all possible combinations from the set {-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8} in different experiments for each choice of Σ. For clarity of the exposition, we assume a common neighborhood structure $W(=W_1 = W_2 = W_3 = W_4)$, $\rho_1 = \rho_2$ and $\lambda_1 = \lambda_2$. It should be noted that in applications, this is not the case. Weight matrices for different variables will take different specifications depending on the nature of spatial processes that influence them (see for example, Cuaresma, 2010). However, there is no loss of generality by using the same weight matrix for different process for the Monte Carlo experiments.

We consider three samples sizes of 100, 250 and 500 observations each. For each sample size, two different weight matrices are considered. The specification of *W* closely follows the weight matrix described in Kelejian & Prucha (1999) and Das, Kelejian, & Prucha (2003). These matrices differ in the degree of sparseness. For the first specification, a hypothetical circular world is considered where each observation (y_i and u_i) is related to exactly one neighbor immediately before it and one neighbor immediately after it. Thus, the i^{th} row of *W* has non-zero entities only in *i-1* and *i+1* column, for each *i = 2, 3…. (n-1).* For the first row, the non-zero elements are in the 2^{nd} and n^{th} column while for the last row, the non-zero elements are in the *n-1*th and 1st column. Further, the matrix *W* is row standardized such that sum of elements in each row =1. This matrix is termed as “one ahead and one behind.” The second matrix is analogously defined as “three ahead and three behind” where each observation is related to exactly three other observations behind and ahead of it. Thus, the average number of neighbors in the first matrix is 2 while in the second matrix it is 6. Kelejian and Prucha report that the results from hypothetical weight matrices and "real world" weight matrices are similar. We conjecture similar outcomes in this case.

The parameter associated with the non-spatial components of the model is specified as below, representing both positive and negative association. The choice of these values has no or very little bearing on the research question.

$\gamma_1 = -0.3, \quad \gamma_2 = 0.7, \quad \beta_1 = 2.0, \quad \beta_2 = 2.5 \quad \beta_3 = 2.5, \quad \beta_4 = -2.0$

Similarly, two alternative forms of the variance covariance matrix Σ are used corresponding to an R^2 value of roughly 0.75 and 0.6 respectively, where R^2 is defined as the average squared correlation coefficient (Carter & Nagar, 1977) between y_i and the mean value of y_i as explained by the model in different experiments:

$$\Sigma_1 = \begin{pmatrix} 900 & 450 \\ 450 & 900 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} 3000 & -2500 \\ -2500 & 4000 \end{pmatrix}$$

In the first case, $\sigma_1 = \sigma_2$ and the correlation between error terms $corr(\varepsilon_1, \varepsilon_2) = \sigma_{12} / \sigma_1\sigma_2$ is 0.50. In the second case, $\sigma_1 \neq \sigma_2$ and the correlation between error terms is -0.72.

3.2 Generating Synthetic Data

A dataset that is a realization of the spatial processes under study is needed for the purpose of estimation. It should be a generated from interdependencies between variables, random components and the spatial interactions between them as specified in the model structure. For each scenario, a different dataset is generated influenced by the parameter settings, nature and strength of spatial dependence, variance-covariance structure and sample size. This procedure will ensure that the variation in results of different scenarios only reflects the changes in the scenarios rather than the randomness in the data generation process thus making comparison possible. The data generation process consists of two parts, namely generating the values of the disturbance terms and that for the regression variables.

3.3 Generating the values of disturbance terms

The process of generating spatially correlated random components starts with random draws from independently and identically distributed normal random variables $\nu_i, (i = 1, 2)$ with zero mean and unit variance. These values are then transformed to reduced form disturbances ε_i that are correlated across equations with zero mean and variance covariance matrix Σ using the following transformation:

$E = V\Sigma_*$ where, $E = (\varepsilon_1, \varepsilon_2)$, $V = (\nu_1, \nu_2)$

and Σ_* is the *m* x *m* lower triangular matrix such that $\Sigma_*^{'}\Sigma_* = \Sigma$. The disturbance terms, u_i, in the model are then obtained by using the transformation $u_i = (I - \rho W)^{-1}\varepsilon_i$ resulting in randomness that is correlated across space as well as across equations.

3.4 Generating the values of regression variables

The starting values for a large number of time series for the two exogenous variables $X_{b,t} = \left(x_3, x_4\right)$ are independently drawn from a normal distribution with zero mean and unit variance. x_3 is treated as fixed over time while x_4 is assumed to grow at a rate of one percent in every period. To avoid the sensitivity of results to exogenous variables, they are generated using the same set of random realizations in every experiment.

Values of Y_t are generated conditional on $X_{a,t}$ and $X_{b,t}$ using a reduced form autoregressive data generation process described as follows. Re-writing equation (7) with a time subscript, substituting $X_{a,t} = Y_{t-1}$ and taking expectations, we obtain:

$$B.Y_t + T_a.Y_{t-1} + T_b.X_{b,t} = 0 \quad (8)$$

$$Y_t = -(B)^{-1}.(T_a.Y_{t-1} + T_b.X_{b,t}) \quad (9)$$

True values for Y_t are generated using equation (9) for each period starting from initial values of $Y_{t=0}$ from normal random variables, and exogenously generated values for the variable $X_{b,t}$. This process is iterated several times to ensure that the pre-determined variable Y_{t-1} is generated using the same underlying spatial process as Y_t. The observed value of Y_t is subsequently obtained by perturbing its true values with disturbances $U=(u_1,u_2)$ whose values were generated in the previous section.

$$Y_{t,observed}=Y_t \pm U_t \quad (10)$$

The results from Monte Carlo simulations are at best random. In order to obtain sufficiently accurate results, a large number of repetitions is required. The errors due to the number of repetitions were reduced by use of antithetic variates. Thus, in equation (10) both positive and negative error terms are used to generate the observed values of *Y*.

3.5 Simulation design

Random samples are drawn from a specified distribution, and a set of data consistent with the model is generated. It is then used to estimate model parameters using the FGS2SLS and FGS3SLS estimator. This process is repeated several times. The estimates are then averaged to obtain the expected values of parameters of interest. The whole process is repeated for varying degrees of spatial dependences, sample size and the neighborhood structure to analyze the performance of FGS3SLS estimator under different conditions to analyze the sensitivity of the results to the data generation process. The complete code for the experiments is written in the statistical package R (R Development Core Team, 2005).

4. Results

Monte Carlo simulation using the above parameters and synthetically generated data is performed for all combinations of the weight matrix *W*, sample size *n* and the spatial lag parameter λ. 500 random samples of errors are generated for each set of *n*, ρ, the neighborhood matrix *W* and the covariance matrix Σ. Each vector of errors is used twice (as thetic and anti-thetic variates) resulting in 1000 repetitions for each experiment. This setup yields a total of 2 values for Σ, 2 for *W*, 9 for λ_i, 9 for ρ_i and 3 for *n* resulting in 972 experiments with 1000 repetitions for each experiment.

The performance of the Feasible Generalized Spatial Three Stage Least Square estimator (FGS3SLS) was found to be superior to the Feasible Generalized Spatial Two Stage Least Square estimator (FGS2SLS) which in turn was found to be superior to the ordinary two stage least square estimator under varying conditions. Table 1 demonstrates the gains from using FGS3SLS for the parameter settings described in this paper. The estimates from FGS3SLS have smaller bias and variance compared to the FGS2SLS estimator. The gains from using the former are greater when the spatial correlation in disturbance terms is high, the spatial lag parameter has a low absolute value, the sample size is small and the neighborhood structure is less dense.

<<insert table 1 here >>

Given the overall superiority of the FGS3SLS estimator under different conditions, we will only focus on the properties of FGS3SLS in the subsequent analysis. The simulations permit analysis of the impact of sample size, neighborhood density, variance-covariance structure of disturbances and the strength of spatial dependence on parameter estimates obtained using this estimator.

4.1 Impact of sample size on parameter estimates

In this section, we analyze the impact of sample size on parameter estimates using root mean square errors (RMSE) as a measure of performance for the FGS3SLS estimator. An attempt is made to isolate the interaction effects of sample size with neighborhood density (average number of neighbors), variance covariance matrix of error structures, degree of spatial dependences in endogenous variables and spatial autocorrelation in errors. For brevity of presentation, we choose one value of ρ and show the impact of varying sample size on RMSE of $\hat{\gamma}_2$ for different values of λ. Similarly, we choose one value of λ and show the impact of varying sample size on RMSE of $\hat{\gamma}_2$ for different values of ρ. The exercise is repeated for the two variance-covariance matrices (see figure 1).

<<insert figure 1 here>>

Increasing the sample size from 100 to 250 observations had a huge impact on the RMSE of a parameter estimates irrespective of other control variables like neighborhood density or the variance covariance matrix. However, the gains in increasing the sample size from 250 to 500 were marginal except at extreme values of spatial dependence parameters λ and ρ. A large sample size improves the performance much more when the spatial lag parameter of the dependent variable is small, the spatial autocorrelation in errors is high, the number of neighbors is large and the variance covariance structure of error are large.

4.2 Impact of the average number of neighbors specified in the weight matrix *W*

The choice of neighborhood structure as defined by *W* is often decided *a priori* using exploratory data analysis or is based on the goodness of fit criteria. This is because the data generation process is not known in practice and the theory behind selection of the weight matrix is weak (see Cuaresma, 2010).

According to the simulations, the impact of neighborhood density on RMSE of parameter estimates depends on the strength of spatial dependences (ρ, λ) as shown in figure 2. For all parameter estimates except that of ρ, increasing the average number of neighbors increased the RMSE noticeably for the following two combinations of spatial dependence parameters – (a) extreme negative values of ρ and high positive λ, and (b) small absolute values of λ and high positive ρ. However, for small absolute values of λ and extreme negative values of ρ, the RMSE actually decreased. The estimates of ρ conditional on *W* behaved slightly differently. Increasing the density marginally increased the RMSE for small ρ irrespective of λ but was drastically decreased for extreme negative values of ρ (except at high positive λ).

<< *insert figure 2 here* >>

The experiments with different number of average neighbors revealed that as the structure becomes denser, the bias in parameter estimates increases many times. The effect is more pronounced as the spatial autocorrelation in the dependent variable and error structure increases. An increase in the sample size consistently and greatly reduces the bias due to the increase in neighborhood density. Thus, in a large sample, the increase in bias due to a denser neighborhood structure is marginal. The result for estimates of γ_2 for different values of sample size and degree of spatial dependences are shown in figure 3. Estimates of other model parameters behaved in similar fashion.

<< *insert figure 3 here*>>

The simulation results suggest that the choice of neighborhood structure should not only involve goodness of fit criteria but also concern for increased bias in parameter estimates due to denser neighborhood structure.

4.3 Estimates of λ and ρ

The bias in the estimate of the spatial autocorrelation parameter in error terms ρ was analyzed under different sample sizes, variance-covariance structure and weight matrices conditional on different values of the spatial lag parameter λ. Similar analysis was conducted for the estimates of the spatial lag parameter λ conditional on ρ (figure 4). The estimator does not provide a direct way to calculate the variance of ρ and therefore, it was derived computationally. One point of caution is that the estimation of ρ requires an optimization procedure where the objective function may not be well defined and is susceptible to the choice of starting parameters. This was not found to be the case in our experiments as the results were stable with respect to the choice of starting parameters. However, it is a concern to be borne in mind while using the estimator.

Estimates of λ were very robust to varying degrees of spatial dependences over most of the (ρ, λ) space. As the neighborhood density increases, there is an increase in the bias and is mostly independent of the value of ρ on which it is conditioned. The estimator performs well at low and moderate degrees of spatial dependencies in endogenous variables except when there is a simultaneous presence of a very high spatial dependence in randomness. Surprisingly, higher bias in the parameter estimate of λ was accompanied by higher variances, signifying the poor performance of the estimator in such conditions.

The bias and variance of ρ was largely independent of the values of λ it was conditioned upon except at very high values of λ. The bias increased very rapidly when its true parameter value increased from -0.8 to +0.8. However, unlike λ, there was a clear bias-variance trade off in the estimates of ρ.

<<insert figure 4 here>>

5. Conclusion

In this paper, we analyzed the small sample properties of the limited information Feasible Generalized Spatial Two Stage Least Squares (FGS2SLS) and the full information Feasible Generalized Spatial Three Stage Least Squares (FGS3SLS) estimator for a system of simultaneous equations with spatial dependence in error terms and in the dependent variable. Given relatively few published applications of this estimator and lack of theoretical understanding about its behavior in small samples, this paper provides a starting point for analyzing the behavior of this estimator. A Monte Carlo framework was used to explore the impacts of sample size, neighborhood structure, variance co-variance matrix and varying degree of spatial dependence parameters on the estimators' performance.

The FGS3SLS estimator performed better than the FGS2SLS estimator in terms of smaller bias and lower variance. The gains of using the former are higher when the spatial correlation in the disturbance terms is high, the spatial lag parameter has a low absolute value, the sample size is small and the neighborhood structure is dense. Given the superiority of the FGS3SLS estimator over the FGS2SLS in the simulations described in this paper, the detailed study of the impacts of sample size, neighborhood structure, variance-covariance matrix and degree of spatial dependence on estimator's behavior that was made was limited to the FGS3SLS estimator.

The performance of the FGS3SLS estimator drastically improved when the sample size was increased from 100 to 250 observations. Increasing the sample size to 500 observations yielded only marginal gains. Gains with increasing sample size are more significant when the heterogeneity is high, the spatial lag parameter of the dependent variable is small, the spatial autocorrelation in errors is high, the number of neighbors is large and the variance covariance structure of error is large. The performance of the estimator was found to be sensitive to the value of the spatial dependence parameters. It deteriorated with low values of the spatial lag parameter in the dependent variable (λ) and at extremely high values of the spatial dependence in the error structure (ρ). Thus, the estimator pays a premium in terms of bias and variance when the spatial lag is small but has huge gains as the spatial lag increases. The estimator for λ performed well at low and moderate degrees of spatial dependencies in the endogenous variables except when there is a simultaneous presence of a very high spatial dependence in randomness. Spatial structures with higher average number of neighbors led to higher bias and variances in the estimates. The effect is more pronounced as the spatial autocorrelation in the dependent variable and error structure increases. In large samples, the increase in bias due to denser neighborhood structure is marginal. The results presented here are sensitive to the model specification, choice of the data generation process, distribution of the exogenous variable, etc. However, the results are useful as a comparative exercise to assess the relative changes in performance under different conditions and should not be taken as an absolute measure of performance.

Understanding the impacts of the sample size, varying degrees of spatial dependencies, neighborhood structure and the error structure on the forecasted value is essential. However, the importance of this work in analyzing the forecasts of spatial data and comparing with the results with true values was not addressed in this paper.

Additional research is needed in order to enhance the use of this estimator in applied work. It is computationally intensive and there is no software or standard code to implement this estimator. Efforts in this direction are very much warranted. A useful extension would be to analyze the impact of increasing model complexity and choice of instruments on the performance of the estimator. Further, the estimates of ρ are obtained from an optimization routine, where the objective function may have multiple optima. In such cases, the parameter estimate of ρ may be susceptible to the choice of starting values and various techniques may be needed to insure that a global optimum is reached. This makes the task more computationally demanding. Work is also needed to theoretically corroborate the findings of this paper in a generalized framework. Over the last five decades, we have learnt a great deal about the properties of the three stage least squares estimator in terms of impacts of misspecification, nonlinearity, multicollinearity, etc., many of which have been studied through Monte Carlo simulations. A parallel series of literature needs to be developed for the Feasible Generalized Three Stage Least Square estimator.

6. References

Anselin, L. (1988) *Spatial econometrics: Methods and models*, Boston, Kluwer Academic Publishers.

Anselin, L. (2010) Thirty years of spatial econometrics, *Papers in Regional Science*, 89, 3-25

Carter, R. A. L. & Nagar, A. L. (1997) Coefficients of correlation for simultaneous equation systems, *Journal of Econometrics*, 6, 39–50.

Case, A. (1991) Spatial patterns in household demand, *Econometrica*, 59, 953–966.

Case, A., Rosen, H. S. & Hines, J. R. (1993) Budget spillovers and fiscal policy independence: Evidence from the states, *Journal of Public Economics*, 52, 285–307.

Cliff, A. & Ord, J. (1973) *Spatial autocorrelation*, London, Pion.

Cliff, A. & Ord, J. (1981) *Spatial processes, models and applications*, London, Pion.
Cuaresma, J. C. & Feldkircher, M. (2010) Spatial filtering, model uncertainty and the speed of income convergence in europe. *Working paper 160, Oesterreichische National Bank*
Das, D., Kelejian, H. H. & Prucha, I. R. (2003) Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances, *Papers in Regional Science*, 82, 1-26.
Davidson, R. & MacKinnon, J. G. (1993) *Estimation and inference in econometrics*, New York, Oxford University Press.
Driffield, N. (2006) On the search for spillovers from foreign direct investment (FDI) with spatial dependency, *Regional Studies*, 40(1), 107-119.
Fishback, P. V., Horrace, W. C. & Kantor, S. (2006) The impact of New Deal expenditures on mobility during the great depression, *Explorations in Economic History*, 43(2), 179-222.
Goodchild, M., Anselin, L., Appelbaum, R. & Harthorn, B. (2000) Toward spatially integrated social science, *International Regional Science Review*, 23, 139–159.
Kelejian, H. H. & Prucha, I. R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40, 509-533.
Kelejian, H. H. & Prucha, I. R. (2004) Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics*, 118, 27-50.
Ngeleza, G. K., Florax, R. J. G. M. & Masters, W. A. (2006) *Geographic and institutional determinants of real income: A spatio-temporal simultaneous equation approach*, Working Papers 06-15, College of Agriculture, Department of Agricultural Economics, Purdue University.
Phillips, P. C. P. (1983) Exact small sample theory in the simultaneous equations model, in: Z. Griliches & M. Intriligato (eds.) *Handbook of Econometrics*, pp 441-516, Vol 1, New York, North Holland.
R Development Core Team. (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://www.R-project.org.
Rothenberg, T. J. (1984) Hypothesis testing in linear models when the error covariance matrix is nonscalar, *Econometrica*, 52, 827–842.
Sarraf, S. (2012) *Three essays on social dynamics and land-use change: Framework, model and estimator*, Unpublished doctoral dissertation, University of Illinois, Urbana Champaign.

λ_i	ρ_i	λ_1	γ_1	β_1	β_3	λ_1	γ_1	β_1	β_3
FGS3SLS		Bias				Variance			
-0.8	-0.8			-0.001	-0.006				0.001
-0.8	-0.4			-0.001	0.002				0.002
-0.8	0.0			-0.003	0.023				0.002
-0.8	0.2		0.001	-0.003	0.039				0.003
-0.8	0.6	-0.001	0.001	-0.008	0.106				0.009
-0.4	-0.8	-0.004	0.002	-0.014	0.249				0.033
-0.4	-0.4	-0.005	0.003	-0.016	0.301				0.047
-0.4	0.0	-0.006	0.004	-0.019	0.360				0.049
-0.4	0.2	-0.005	0.004	-0.019	0.380				0.043
-0.4	0.6	-0.009	0.007	-0.031	0.652			0.001	0.107
0.0	-0.8	-0.002	0.011	-0.009	0.328				0.108
0.0	-0.4	-0.001	0.012	-0.008	0.384				0.142
0.0	0.0	-0.001	0.010	-0.005	0.386				0.157
0.0	0.2	-0.003	0.010	-0.005	0.390				0.120
0.0	0.6	-0.019	0.013	-0.001	0.624	0.002			0.199
0.2	-0.8	-0.001	0.010	-0.006	0.232	0.001			0.152
0.2	-0.4	-0.001	0.007	-0.006	0.177	0.001			0.133
0.2	0.0	-0.001	0.006	-0.005	0.143	0.001			0.105
0.2	0.2	-0.005	0.005	-0.002	0.131	0.001			0.099
0.2	0.6	-0.029	0.004	0.012	0.098	0.004		0.001	0.089
0.6	-0.8	-0.005	0.003	-0.001	0.028	0.001			0.077
0.6	-0.4	-0.005	0.002		0.013				0.083
0.6	0.0	-0.003	0.001		-0.021				0.059
0.6	0.2	-0.005	0.001	0.002	-0.037				0.051
0.6	0.6	-0.013		0.010	-0.187	0.001			0.063
Gains over FGS2SLS		Reduciton in Absolute Bias				Reduction in Variance			
-0.8	-0.8			0.001	0.010				
-0.8	-0.4		0.001	0.002	0.027				
-0.8	0.0	0.001	0.001	0.003	0.042				0.001
-0.8	0.2	0.001		0.005	0.055				0.002
-0.8	0.6	0.002	0.002	0.010	0.118				0.010
-0.4	-0.8	0.007	0.003	0.018	0.250				0.053
-0.4	-0.4	0.007	0.003	0.019	0.264				0.068
-0.4	0.0	0.007	0.003	0.019	0.286				0.063
-0.4	0.2	0.009	0.004	0.019	0.299				0.075
-0.4	0.6	0.012	0.006	0.023	0.394				0.208
0.0	-0.8	0.002	0.014	-0.002	0.521				0.130
0.0	-0.4	0.003	0.014	-0.004	0.523				0.145
0.0	0.0	0.005	0.016	-0.004	0.582				0.126
0.0	0.2	0.008	0.015	-0.004	0.553				0.160
0.0	0.6	0.029	0.017	0.013	0.659				0.197
0.2	-0.8	-0.001	0.015	-0.005	0.536				0.201
0.2	-0.4	0.005	0.013	-0.005	0.489				0.129
0.2	0.0	0.014	0.011		0.513				0.095
0.2	0.2	0.017	0.010	0.008	0.508				0.101
0.2	0.6	0.049	0.011	0.031	0.643				0.116
0.6	-0.8	-0.003	0.005	0.004	0.258			0.001	0.027
0.6	-0.4		0.004	0.002	0.236	0.001			0.008
0.6	0.0	0.003	0.003		0.191				0.010
0.6	0.2	0.004	0.003		0.168				0.006
0.6	0.6	0.017	0.002	0.011	-0.057			0.001	0.002

Table 1: FGS3SLS Bias and Variances for n=250, Σ_2 *, W=6,* $\gamma_1 = -0.3$, $\beta_1 = 2.0$, $\beta_3 = 2.5$

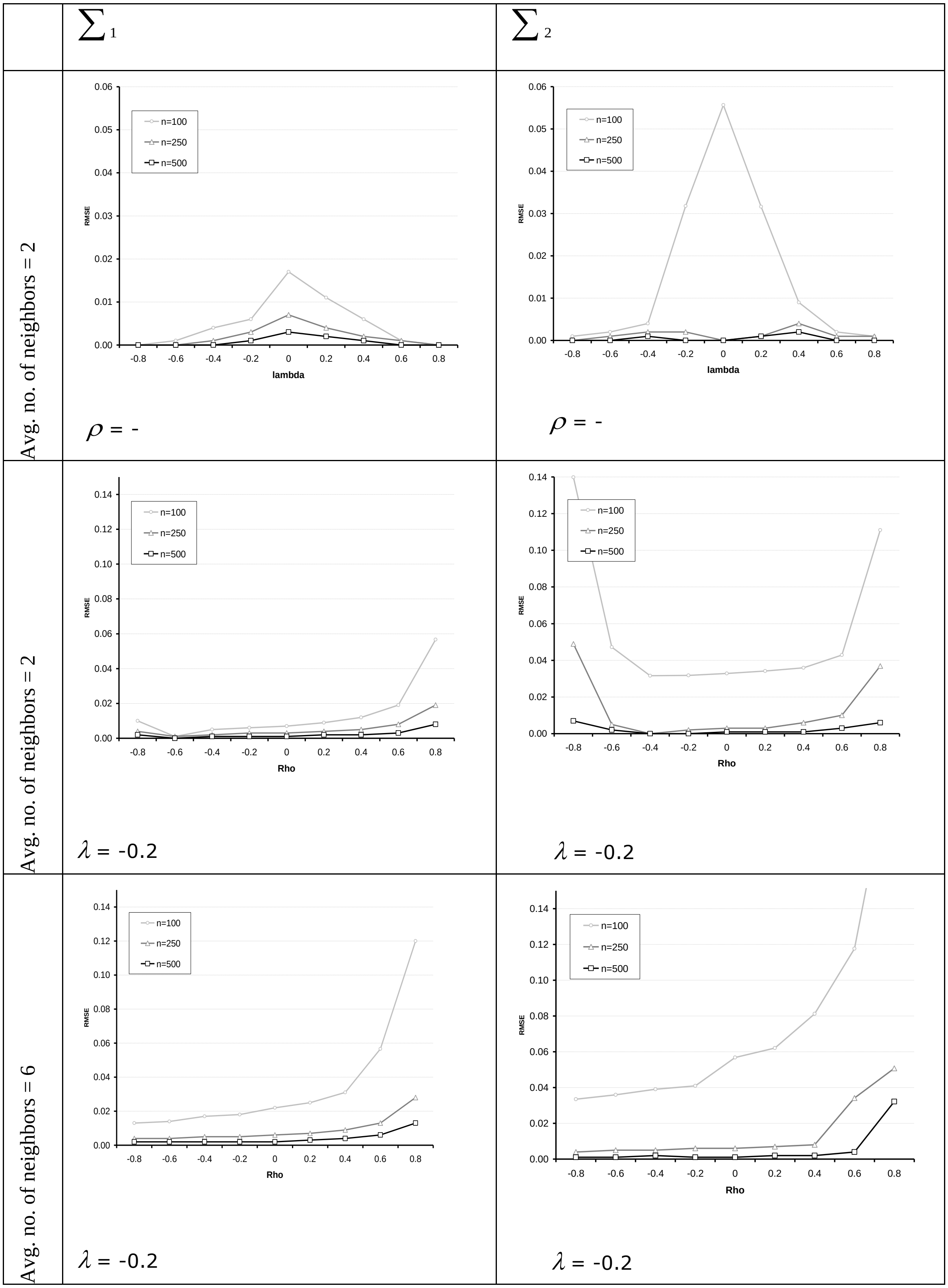


Figure 1: Impact of sample size on RMSE of γ_2

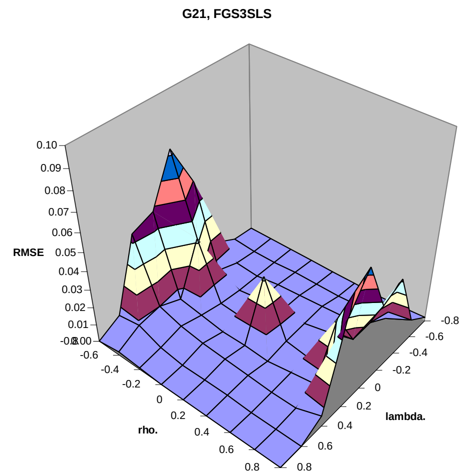
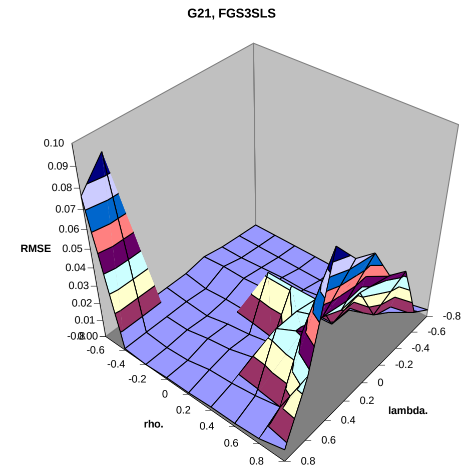
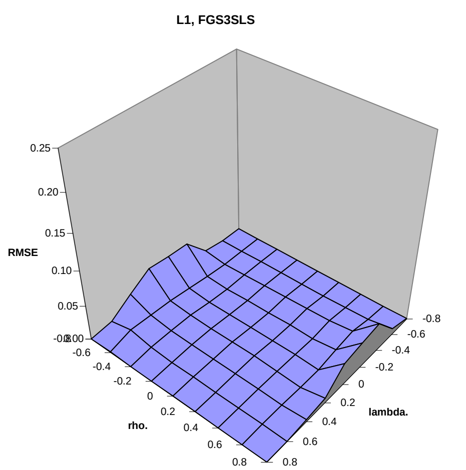
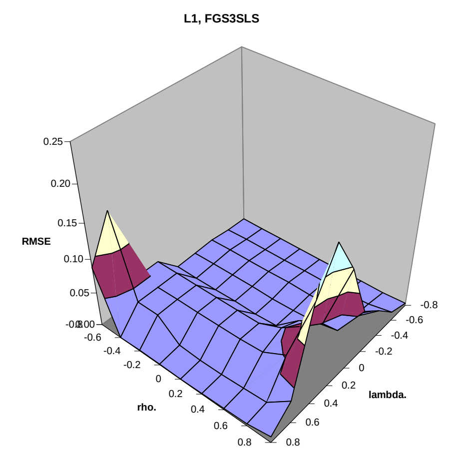
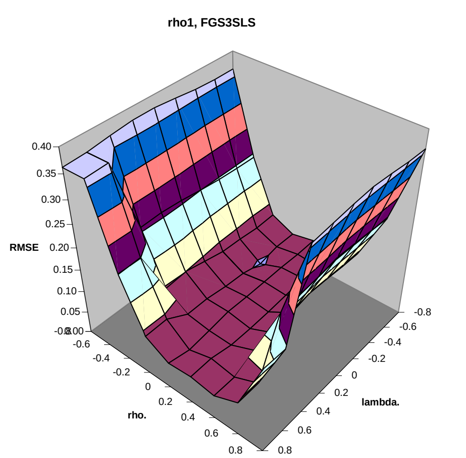
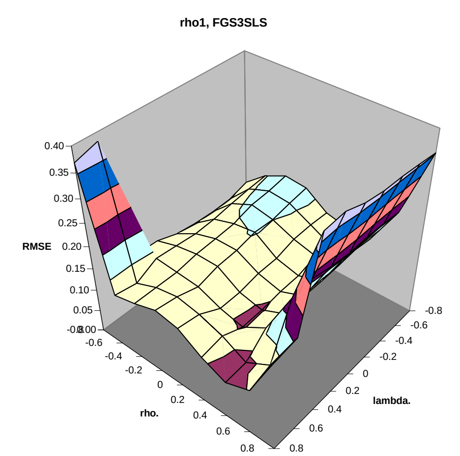
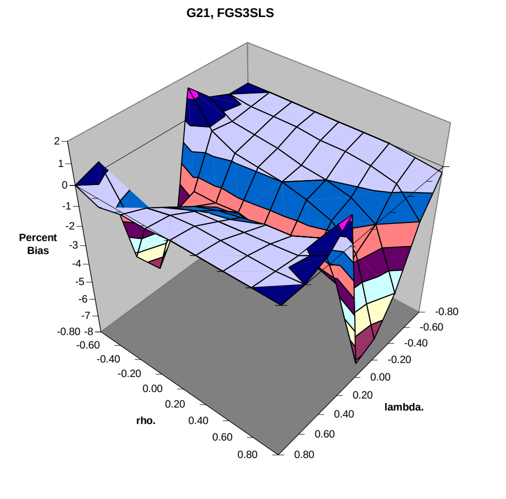
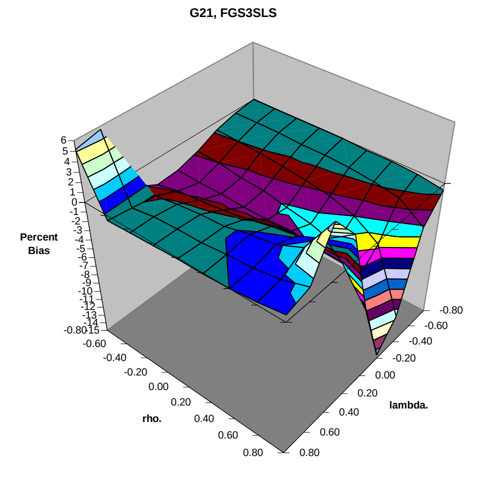
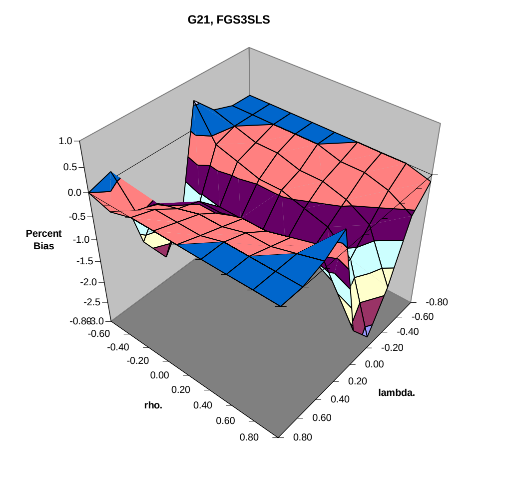
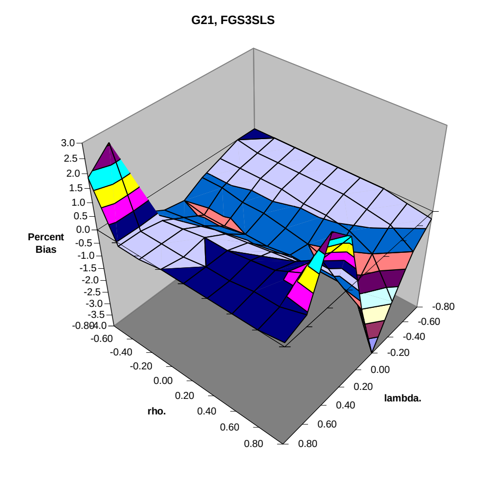
	Average number of neighbors=2	Average number of neighbors=6
γ_2		
λ		
ρ		

Figure 2: Impact of average number of neighbors on RMSE for Σ_2 and n = 250

<table>
<tr><td></td><td>Average number of neighbors=2</td><td>Average number of neighbors=6</td></tr>
<tr><td>N=

100</td><td>
</td><td>
</td></tr>
<tr><td>N=

250</td><td>
</td><td>
</td></tr>
<tr><td>N=

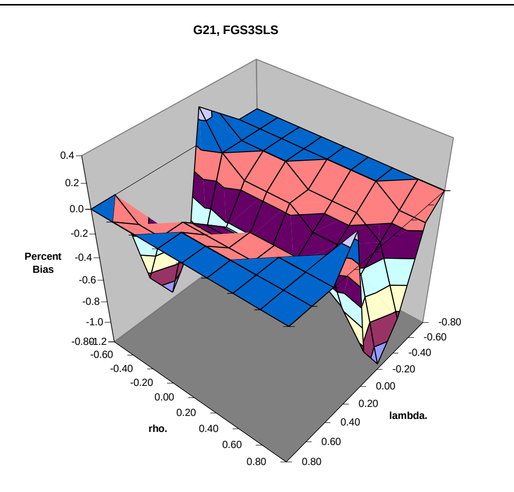
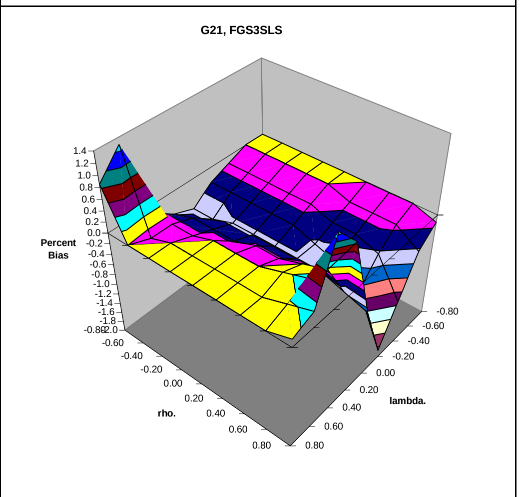
500</td><td>
</td><td>
</td></tr>
</table>

Figure 3: Impact of average no. of neighbors on percentage bias of $\gamma_2 (= 0.7)$ *for* Σ_1

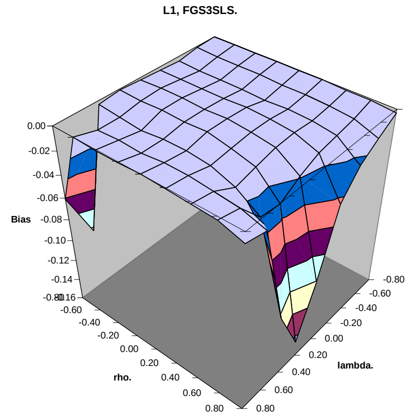
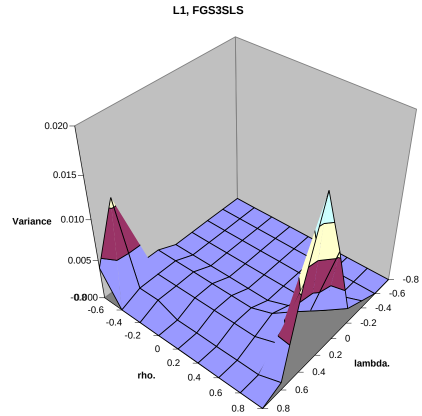
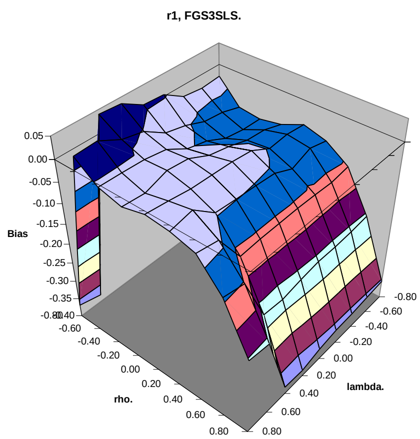
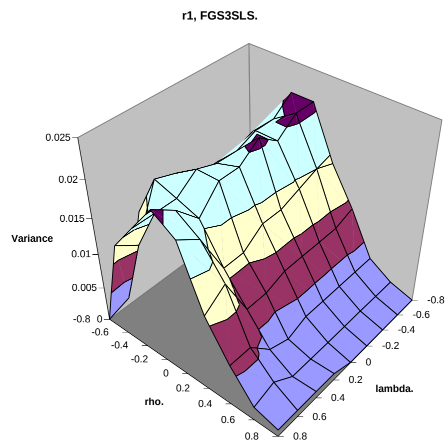
	Bias	Variance
λ		
ρ		

Figure 4: Estimates of λ *and* ρ *, at n=250 for* Σ_2 *, average number of neighbors = 6*