Using Matrix Exponentials to Explore Spatial Structure in Regression Relationships

James P. LeSage[1]
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com

R. Kelley Pace
LREC Endowed Chair of Real Estate
Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
OFF: (225)-388-6256, FAX: (225)-334-1227
kelley@pace.am, www.spatial-statistics.com

October 11, 2000

[1]The authors would like to thank the Ohio Urban Universities Program for funding development of the ES202 data used in this study.

Abstract

Spatial regression relationships typically specify the structure of spatial relationships among observations using a fixed n by n weight matrix, where n denotes the number of observations. For computational convenience, this weight matrix is usually assumed to represent fixed sample information, making inferences conditional on the particular spatial structure embodied in the weight matrix used in the model. In an economic context, where the spatial structure can arise from externalities or spillovers, the magnitude and extent of influence from one observation or spatial location on other observations at nearby locations may be a subject of interest and inference. To address the dual goals of improving inference and understanding spatial structure, we set forth a flexible specification for the spatial weights that allows hyperparameters to control the number of neighboring entities as well as decay of influence over space. This specification is combined with the matrix exponential spatial specification (MESS) introduced in Pace and LeSage (2000) leading to a Bayesian model that allows posterior inferences regarding the magnitude and extent of spatial influence. We illustrate the method in an application that examines the role of geographical proximity in the economic growth of 219 urban zip-code areas in northeast Ohio. Inferences regarding the magnitude and extent of spatial influence in this setting may be of interest because theories of economic growth have increasingly stressed technological spillovers as a source of growth (see Romer, 1994; Grossman and Helpman, 1994).

KEYWORDS: spatial statistics, spatial autoregression, nearest neighbor, Bayesian Markov Chain Monte Carlo estimation, spatial econometrics.

1 Introduction

Spatial autoregressive models typically specify the spatial dependence among observations using a n by n weight matrix, where n denotes the number of observations. Traditional specifications treat the weight matrix as fixed sample information for computational convenience and to avoid problems such as multimodality in the likelihood. This results in inferences that are conditional on the specific form of weight matrix used to specify the model. In many applications interest centers upon the spatial structure itself. For example, economic constructs such as spillovers or externalities imply that an economic unit at one point in space exerts an influence upon other economic units. Both the magnitude of influence and its decay with distance are important. To address this need for a model of spatial structure we set forth a flexible specification for the spatial weights that allows hyperparameters to control the number of neighboring entities as well as decay of influence over space. This specification combined with the matrix exponential spatial specification (MESS) introduced in Pace and LeSage (2000) leads to a Bayesian model that allows posterior inferences regarding the magnitude and extent of spatial influence.

We argue that this spatial specification has computational advantages over more traditional specifications and allows simultaneous inference regarding the usual parameters in the model as well as the nature and extent of the spatial relationships involved. Pace and LeSage (2000) introduce an approach to maximum likelihood estimation of spatial regression models based on the matrix exponential covariance specification introduced by Chiu, Leonard, and Tsui (1996). They show that matrix exponentials can spatially transform the dependent and/or independent variables so that the determinant of the matrix exponential transformation identically equals 1. This eliminates the computationally troublesome log-determinant term from the log-likelihood and reduces maximum likelihood estimation to minimizing a quadratic form subject to a polynomial constraint. Further, they show that the resulting minimization problem has a unique, closed-form interior solution. Thus, maximum likelihood for the MESS model reduces to nonlinearly constrained least squares.

The MESS model introduced in Pace and LeSage (2000) relies on a flexible specification of the spatial weight structure involved in the model that allows hyperparameters to control various aspects of the spatial weight specification. In a maximum likelihood setting, this complicates the optimization problem and Pace and LeSage (2000) examine a profile of the likelihood surface with respect to estimates based on varying values of the hyperparameters and use likelihood ratio tests to draw inferences.

This paper introduces a Bayesian variant of the MESS model. This variant can produce estimates identical to those from the maximum likelihood MESS model, if implemented with a diffuse prior. Estimation of the model proceeds via Markov Chain Monte Carlo (MCMC), making it more computationally demanding than maximum likelihood estimates, but one that provides a number of other advantages.

First, the MCMC method can handle the problem of local optima, that often arise in spatial modeling. In the presence of local optima and other erratic behavior of the likelihood, conventional likelihood methods may provide misleading inference whereas the ability of Bayesian MCMC methods to directly sample from the posterior can avoid some of these problems. Second, Bayesian methods can more gracefully address the problem of nuisance parameters (Wolpert (2000, p. 771-772)). Third, Bayesian models usually are advantageous in situations involving parameters that take on discrete values, as is the case with the hyperparameter controlling the number of nearest neighbors in the MESS model.

Finally, and most importantly, Bayesian models based upon MCMC provide greater modeling flexibility in many circumstances than alternative approaches. For example, MCMC methods can accommodate missing or censored data samples.[1] Censoring often arises in spatial modeling problems because governmental agencies employ censoring for confidentiality purposes. In addition, Bayesian methods can formally incorporate relevant prior information. This flexibility also allows relatively simple modifications to produce robust estimates that address spatially idiosyncratic behavior sometimes displayed by aberrant observations or spatial enclaves.

While MCMC provides a number of important advantages, a naive implementation for traditional spatial autoregressive models would involve calculating the log-determinant of an nxn matrix on each pass through the sampler (see LeSage, 2000). The use of MESS, where the log-determinant term vanishes, greatly accelerates the MCMC computations. Another desirable aspect of the MESS model is the availability of the likelihood which greatly facilitates Bayesian inference. This is in contrast to other specifications that have been introduced to overcome computational problems in spatial models such as the generalized-moments (GM) estimation techniques

[1] The Bayesian MESS model could easily be extended to the case of limited or censored dependent variables. Following LeSage (2000), we need only add one additional conditional distribution to the MCMC sampler that produces a latent variable reflecting the limited or truncated dependent variable distribution. This results in a model where posterior inference can proceed as in the case of continuous dependent variables.

proposed by Kelejian and Prucha (1998,1999).

In the Bayesian MESS model introduced here, MCMC estimation methods produce posterior distributions for the hyperparameters used in the weight specification. This easily implementable approach provides an elegant solution to replace the profile likelihood grid search over hyperparameter values. Another advantage to the Bayesian MESS model is that prior information regarding the regression parameters, spatial weight hyperparameters, spatial dependence parameter and disturbance variance can be introduced. Spatial modeling often involves large samples, so prior information regarding regression parameters will probably not have much influence on the estimation outcome. However, priors placed on the spatial weight specification parameters, the spatial dependence parameter, or the robustness parameter may exert an important influence.

Section 2 sets forth relevant aspects of the MESS model from Pace and LeSage (2000) and section 3 introduces the Bayesian version of the model along with estimation via Markov Chain Monte Carlo methods. Computational considerations are also discussed in this section. In section 4 we illustrate the method in an application that examines the role of geographical proximity in the economic growth of 219 urban zip-code areas in northeast Ohio. Inferences regarding the magnitude and extent of spatial influence in this setting may be of interest because theories of economic growth have increasingly stressed technological spillovers as a source of growth (see Romer, 1994; Grossman and Helpman, 1994). We use this illustrative application to demonstrate inference regarding the spatial structure of employment growth relations in urban zip-code areas that may arise from knowledge spillovers.

2 The MESS model

The computational difficulties associated with estimation of spatial econometric regression models discussed in Anselin (1988) are well-known. Direct estimation via maximum likelihood for spatial models requires computation of a determinant involving an nxn matrix. Brute force implementation of maximum likelihood methods become prohibitively expensive for very large data sets that are the focus of recent spatial statistical applications.

Advances for the particular case of nearest neighbor spatial dependence described in Pace and Zou (2000) provide a closed-form solution that produces maximum likelihood estimates. Another alternative that also leads to closed-form maximum likelihood estimates is presented in Pace and LeSage (2000). They introduce the matrix exponential spatial specification (MESS) that relies on a matrix exponential spatial transformation of the dependent variable. They show that this leads to a situation where the troublesome log determinant term vanishes from the log likelihood function.

The MESS model holds an important advantage over other attempts to eliminate the computational burden of estimating spatial models, such as the GM estimation technique proposed by Kelejian and Prucha (1998,1999). The availability of the likelihood in the MESS model allows both classical and Bayesian inference. This is in contrast to the GM methods that require adopting another inferential paradigm.

Consider estimation of models where the dependent variable y undergoes a linear transformation Sy as in (1).

$$Sy = X\beta + \varepsilon \tag{1}$$

The vector y contains the n observations on the dependent variable, X represents the nxk matrix of observations on the independent variables, S is a positive definite nxn matrix, and the n-element vector ε is distributed $N(0, \sigma^2 I_n)$. The log-likelihood for the MESS model in (1) is,

$$L = C + \ln|S| - (n/2)\ln(y'S'MSy) \tag{2}$$

where C represents a scalar constant and both $M = I - H$ and $H = X(X'X)^{-1}X'$ are idempotent matrices. The term $|S|$ is the Jacobian of the transformation from y to Sy. Pace and LeSage (2000) use the matrix exponential defined in (3) to model S as:

$$S = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \tag{3}$$

where W represents an nxn non-negative matrix with zeros on the diagonal and α represents a scalar real parameter. While a number of ways exist to specify W, a common specification sets $W_{ij} > 0$ for observations $j = 1 \ldots n$ sufficiently close (as measured by some metric) to observation i. By construction, $W_{ii} = 0$ to preclude an observation from directly predicting itself. If $W_{ij} > 0$ for the nearest neighbors of observation i, $W_{ij}^2 > 0$ contains neighbors to these nearest neighbors for observation i. Similar relations hold for higher powers of W which identify higher-order neighbors. Thus the matrix exponential S, associated with matrix W, can be interpreted as assigning rapidly declining weights for observations involving higher-order neighboring relationships. That is, observations reflecting higher-order neighbors (neighbors of neighbors) receive less weight than lower-order neighbors.

Pace and LeSage (2000) rely on a property of the matrix exponential, $|e^{\alpha W}| = e^{\text{trace}(\alpha W)}$ to simplify the MESS log-likelihood. Since $\text{trace}(W) = 0$ and by extension $|e^{\alpha W}| = e^{\text{trace}(\alpha W)} = e^0 = 1$, the log-likelihood takes the form: $L = C - (n/2)\ln(y'S'MSy)$. This produces a situation where maximizing the log-likelihood is equivalent to minimizing $(y'S'MSy)$ with respect to S. Note that S always appears in expressions involving pre-multiplication of y or X, eliminating the need for computation involving high-order operation counts. Computing Sy or SX involves low-order matrix-vector product computations that require little time for sparse weight matrices such as W.

The MESS model in (1) can be extended to allow for spatial dependence in the explanatory variables matrix X as shown in (4).

$$X = [\iota \ U \ WU \ldots W^{q-1}U] \tag{4}$$

Where U represents a matrix of observations on p non-constant independent variables and q is an integer large enough for convergence in the Taylor series expansion.

In this case, X approximately spans SU and thus the MESS model based on (4) nests a spatial autoregression in the errors. Hence, a set of linear restrictions on the parameters associated with the columns of X could yield the error autoregression. This allows the MESS model to effectively accommodate different structures for the spatial lags of Y and U (Anselin (1988), p. 225-230).

Additional flexibility can be introduced by specifying a spatial weight that includes a decay parameter ρ that lies between 0 and 1, along with a variable number of nearest neighbor spatial weight matrices N_i, where the subscript i indexes the ith nearest neighbor. The weight structure specification is shown in (5), where m denotes the maximum number of neighbors considered.

$$W = \sum_{i=1}^{m} \left(\frac{\rho^i N_i}{\sum_{i=1}^{m} \rho^i} \right) \tag{5}$$

In (5), ρ^i weights the relative effect of the ith individual neighbor matrix, so that S depends on the parameters ρ as well as m in both its construction and the metric used. By construction, each row in N sums to 1 and has zeros on the diagonal. To see the role of the spatial decay hyperparameter ρ, consider that a value of $\rho = 0.87$ implies a decay profile where the 6th nearest neighbor exerts less than 1/2 the influence of the nearest neighbor. We might think of this value of ρ as having a "half-life" of six neighbors.

On the other hand, a value of $\rho = 0.95$ has a half-life between 14 and 15 neighbors.

The flexibility arising from this type of weight specification adds to the burden of estimation requiring that we draw an inference on the parameters ρ and m. Together these hyperparameters determine the nature of the spatial weight structure. To the extent that the weight structure specification in (5) is flexible enough to adequately approximate more traditional weight matrices based on contiguity, the model introduced here can replicate results from models that assume the matrix W is fixed and known. However, all inferences regarding β and σ drawn from a model based on a fixed matrix W are conditional on the particular W matrix employed. The model we introduce here produces inferences regarding β and σ that are conditional only on a family of spatial weight transformations that we denote Sy, where $S = e^{\alpha W}$, with the matrices W taking the form in (5). Of course, this raises the issue of inference regarding these hyperparameters, and we show that the Bayesian MESS model introduced here can produce a posterior distribution for the joint distribution of the parameters α, ρ and m as well as the other model parameters of interest, β and σ.

3 The Bayesian MESS model

A Bayesian approach to the MESS model would include specification of prior distributions for the parameters in the model, $\alpha, \beta, \sigma, \rho$ and m. Prior information regarding the parameters β and σ is unlikely to exert much influence on the posterior distribution of these parameter estimates in the case of very large samples that are often the focus of spatial modeling. However, the parameters α, ρ and m are likely to exert an influence even in large samples, because they determine the structure of the spatial transformation, Sy in the model.

The Bayesian MESS is presented in (6), where the priors are enumerated.

$$
\begin{aligned}
Sy &= X\beta + \varepsilon \\
S &= e^{\alpha W} \\
W &= \sum_{i=1}^{m} \left(\rho^i N_i / \sum_{i=1}^{m} \rho^i \right) \\
\varepsilon &\sim N(0, \sigma^2 V) \quad V = \mathrm{diag}(v_1, \ldots, v_n) \\
\beta &\sim N(c, T) \\
r/v_i &\sim \mathrm{ID}\chi^2(r)
\end{aligned}
$$

$$
\begin{aligned}
1/\sigma^2 &\sim \Gamma(d, \nu) \\
\alpha &\sim U[-\infty, 0], \quad \text{or} \quad N(a, B) \\
\rho &\sim U[0, 1] \\
m &\sim U^D[1, m_{\max}]
\end{aligned} \tag{6}
$$

We rely on a normal-gamma prior for the parameters β, σ using a prior mean c and variance T for the normal prior on β and prior parameters d, ν in the gamma prior on σ. In the case of very large samples involving upwards of 10,000 observations, the normal-gamma priors for β, σ should exert relatively little influence.

In contrast, the prior assigned to the parameter α associated with spatial dependence should exert an impact on the estimation outcomes even in large samples because of the important role of the spatial structure in the model. The prior assigned for α can be a relatively non-informative uniform prior that allows for the case of no spatial effects when $\alpha = 0$, or an informative prior based on a normal distribution centered on a with prior variance B as indicated in (6).

The relative variance terms $(v_1, v_2, \ldots, v_n)$, are assumed fixed but unknown parameters that need to be estimated. The prior distribution for the v_i terms takes the form of an independent $\chi^2(r)/r$ distribution. Recall that the χ^2 distribution is a single parameter distribution, where we have represented this parameter as r. This allows us to estimate the additional n parameters v_i in the model by adding the single parameter r to our estimation procedure.

This type of prior has been used Geweke (1993) in modeling heteroscedasticity and outliers in the context of linear regression. The specifics regarding the prior assigned to the v_i terms can be motivated by considering that the mean equals unity and the variance of the prior is $2/r$. This implies that as r becomes very large, the terms v_i will all approach unity, resulting in $V = I_n$, the traditional assumption of constant variance across space. The role of $V \neq I_n$ is to accommodate outliers and observations containing large variances by down-weighting these observations. Even in large samples, this prior will exert an impact on the estimation outcome.

A relatively non-informative approach was taken for the hyperparameters ρ and m where we rely on a uniform prior distribution for ρ and a discrete uniform distribution for m, the number of nearest neighbors. The term $m_{\max}$ denotes a maximum number of nearest neighbors to be considered in the spatial weight structure, and U^D denotes the discrete uniform distribution that imposes an integer restriction on values taken by m. Note that practitioners may often have prior knowledge regarding the number of neighboring observations that are important in specific problems, or the extent to which spatial influence decays over neighboring units. Informative priors could be developed and used here as well, but in problems where interest centers on inference regarding the spatial structure relatively non-informative priors would be used for these hyperparameters.

Given these distributional assumptions, it follows that the prior densities for $\beta, \sigma^2, \alpha, \rho, m, v_i$ are given up to constants of proportionality by (7), (where we rely on a uniform prior for α).

$$\begin{aligned}
\pi(\beta) &\propto \exp[-\frac{1}{2}(\beta - c)' T^{-1}(\beta - c)] \\
\pi(\sigma^2) &\propto (\sigma^2)^{-(d+1)} \exp\left(-\frac{\nu}{\sigma^2}\right) \\
\pi(\rho) &\propto 1 \\
\pi(\alpha) &\propto 1 \\
\pi(m) &\propto 1 \\
\pi(v_i) &\propto v_i^{-(\frac{r}{2}+1)} \exp\left(-\frac{r}{2v_i}\right)
\end{aligned} \tag{7}$$

3.1 Estimation of the model

Given the prior densities from section 3, the Bayesian identity

$$p(\beta, \sigma^2, V, \rho, \alpha, m) \cdot p(y) = p(y|\beta, \sigma^2, V, \rho, \alpha, m) \cdot \pi(\beta, \sigma^2, V, \rho, \alpha, m) \tag{8}$$

together with the assumed prior independence of the parameters allows us to establish the posterior joint density for the parameters, $p(\beta, \sigma^2, V, \rho, \alpha, m)$. This posterior is not amenable to analysis, a problem that has often plagued Bayesian methods in the past. We can however derive the posterior distribution for the parameters in our model using a methodology known as Markov Chain Monte Carlo (MCMC).

MCMC is based on the idea that rather than compute the posterior density of our parameters based on the expression in (8), we would be just as happy to have a large random sample from the posterior of our parameters, which we designate $p(\theta|D)$, using θ to denote the parameters and D the sample data. If the sample from $p(\theta|D)$ were large enough, we could approximate the form of the probability density using kernel density estimators or histograms, eliminating the need to know the precise analytical form of the density.

We rely on Metropolis-Hastings to compute the posterior distributions for the parameters α, ρ and m in the MESS model. A normal distribution is used as the proposal density for α and rejection sampling can be used to constrain α to the range $(-\infty, 0)$ which imposes positive spatial autocorrelation. This prior restriction is often used in spatial modeling because negative spatial correlation is difficult to motivate. A uniform proposal distribution for ρ over the interval $(0, 1)$ was used along with a discrete uniform for m over the interval $(1, m_{\max})$.

The parameters β, V and σ in the MESS model can be estimated using draws from the conditional distributions of these parameters, a process known as Gibbs sampling. Assume a parameter vector $\theta = (\theta_1, \theta_2)$, a prior $p(\theta)$, and likelihood $l(\theta|y, X)$, that produces a posterior distribution $p(\theta|D) = c \cdot p(\theta) l(\theta|y, X)$, with c a normalizing constant. The case we encounter in the MESS model is one where the posterior distribution over all parameters is difficult to work with. However, if we partition our parameters into two sets θ_1, θ_2 and had initial estimates for θ_1, we could estimate θ_2 conditional on θ_1 using $p(\theta_2|D, \hat{\theta}_1)$. As a concrete example, in our case an estimate of β conditional on $\sigma, V, \alpha, \rho, m$ is very easy to derive and compute. Denote the estimate, $\hat{\theta}_2$ derived by using the posterior mean or mode of $p(\theta_2|D, \hat{\theta}_1)$, and consider that we are now able to construct a new estimate of θ_1 based on the conditional distribution $p(\theta_1|D, \hat{\theta}_2)$. Note that for our problem it is also easy to compute estimates of σ and V conditional on the other parameters in the MESS model. This new estimate for θ_1 can be used to construct another value for θ_2, and so on.

Summarizing, we will rely on Metropolis sampling for the parameters α, ρ and m within a sequence of Gibbs sampling steps to obtain β, σ and V, a procedure that is often labeled "Metropolis within Gibbs sampling" (Gelman, Carlin, Stern and Rubin, 1995).

3.2 The conditional distributions for β, σ and V

To implement our Metropolis within Gibbs sampling approach to estimation we need the conditional distributions for β, σ and V which are presented here.

For the case of the parameter vector β conditional on the other parameters in the model, $\alpha, \sigma, V, \rho, m$ we find that:

$$\begin{aligned} p(\beta|\alpha, \sigma, V, \rho, m) &\sim N(\bar{b}, \sigma^2 \bar{B}) \\ \bar{b} &= (X'V^{-1}X + T^{-1})^{-1}(X'V^{-1}Sy + \sigma^2 T^{-1}c) \end{aligned}$$

$$\bar{B} = (X'V^{-1}X + T^{-1})^{-1} \tag{9}$$

Note that given the parameters V, α, ρ, σ and m, the vector Sy and $X'V^{-1}X$ can be treated as known, making this conditional distribution easy to compute and sample from. This is often the case in MCMC estimation, which makes the method attractive.

The conditional distribution of σ is shown in (10), (see Gelman, Carlin, Stern and Rubin, 1995).

$$p(\sigma^2|\beta, \alpha, V, \rho, m) \propto (\sigma^2)^{-(\frac{n}{2}+d+1)}\exp\left[-e'V^{-1}e + \frac{2\nu}{2\sigma^2}\right] \tag{10}$$

where $e = Sy - X\beta$, which is proportional to an inverse gamma distribution with parameters $(n/2) + d$ and $e'V^{-1}e + 2\nu$.

Geweke (1993) shows that the conditional distribution of V given the other parameters is proportional to a chi-square density with $r + 1$ degrees of freedom. Specifically, we can express the conditional posterior of each v_i as:

$$\frac{e_i^2 + r}{v_i}|(\beta, \alpha, \sigma^2, v_{-i}, \rho, m) \sim \chi^2(r+1) \tag{11}$$

where $v_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ for each i.

As noted above, the conditional distributions for α, ρ and m take unknown distributional forms that require Metropolis-Hastings sampling. By way of summary, the MCMC estimation scheme involves starting with arbitrary initial values for the parameters which we denote $\beta^0, \sigma^0, V^0, \alpha^0, \rho^0, m^0$. We then sample sequentially from the following set of conditional distributions for the parameters in our model.

1. $p(\beta|\sigma^0, V^0, \alpha^0, \rho^0, m^0)$, which is a multinormal distribution with mean and variance defined in (9). This updated value for the parameter vector β we label β^1.

2. $p(\sigma|\beta^1, V^0, \alpha^0, \rho^0, m^0)$, which is chi-squared distributed $n+2d$ degrees of freedom as shown in (10). Note that we rely on the updated value of the parameter vector $\beta = \beta^1$ when evaluating this conditional density. We label the updated parameter $\sigma = \sigma^1$ and note that we will continue to employ the updated values of previously sampled parameters when evaluating the next conditional densities in the sequence.

3. $p(v_i|\beta^1, \sigma^1, v_{-i}, \alpha^0, \rho^0, m^0)$ which can be obtained from the chi-squared distribution shown in (11). Note that this draw can be accomplished as a vector, providing greater speed.

4. $p(\alpha|\beta^1, \sigma^1, V^1, \rho^0, m^0)$, which we sample using a Metropolis step with a normal proposal density, along with rejection sampling to constrain α to the interval $(-\infty, 0)$. We can rely on the likelihood to evaluate the candidate value of α except in the case of a normal $N(a, B)$ prior on α. Here where we rely on: $-e'e/(2\sigma^2) - 0.5\{(\alpha - a)^2/B\sigma^2\}$, with $e = Sy - X\beta$.

5. $p(\rho|\beta^1, \sigma^1, V^1, \alpha^1, m^0)$, which we sample using a Metropolis step based on a uniform distribution that constrains ρ to the interval (0,1). Here again, we rely on the likelihood to evaluate the candidate value of ρ. As in the case of the parameter α it would be easy to implement a normal or some alternative prior distributional form for this hyperparameter.

6. $p(m|\beta^1, \sigma^1, V^1, \alpha^1, \rho^1)$, which we sample using a Metropolis step based on a discrete uniform distribution that constrains m to be an integer from the interval $(1, m_{\text{max}})$. As in the case of α and ρ, we rely on the likelihood to evaluate the candidate value of m.

We now return to step 1) employing the updated parameter values in place of the initial values $\beta^0, \sigma^0, V^0, \alpha^0, \rho^0, m^0$. On each pass through the sequence we collect the parameter draws which are used to construct a joint posterior distribution for the parameters in our model. Gelfand and Smith (1990) demonstrate that MCMC sampling from the sequence of complete conditional distributions for all parameters in the model produces a set of estimates that converge in the limit to the true (joint) posterior distribution of the parameters. That is, despite the use of conditional distributions in our sampling scheme, a large sample of the draws can be used to produce valid posterior inferences regarding the mean and moments of the parameters.

3.3 Computational considerations

Use of the likelihood when evaluating candidate values of α, ρ and m in the MCMC sampling scheme requires that we form the matrix exponential $S = e^{\alpha W}$, which in turns requires computation of $W = \sum_{i=1}^{m}(\rho^i N_i / \sum_{i=1}^{m} \rho^i)$, based on the current values for the other two parameters. For example, in the case of update $\alpha = \alpha^1$, we use $\rho = \rho^0$ and $m = m^0$ to find W. The nearest neighbor matrices N_i can be computed outside the sampling loop to save time, but the remaining calculations can still be computationally demanding if the number of observations in the problem is large.

Further aggravating this problem is the need to evaluate both the existing value of the parameters α, ρ and m, given the updated values for β, σ and V

as well as the candidate values. In all, we need to form the matrix product Sy, along with the matrix W six times on each pass through the sampling loop.

To enhance the speed of the sampler, we compute the part of Sy that depends only on ρ and m, for a grid of values over these two parameters prior to beginning the sampler. During evaluation of the conditionals and the Metropolis-Hastings steps, a simple table look-up recovers the stored component of Sy and applies the remaining calculations needed to fully form Sy.

The ranges for these grids can be specified by the user, with a trade-off between selecting a large grid that ensures coverage of the region of posterior support and a narrow grid that requires less time. In a typical spatial problem, the ranges might be $0.5 \leq \rho \leq 1$, and $4 < m < 30$. If the grid range is too small, the posterior distributions for these parameters should take the form of a truncated distribution, indicating inadequate coverage of the region of support.

Simpler models than that presented in (6) could be considered. For example either ρ or m, or both ρ and m could be fixed a priori. This would enhance the speed of the sampler because eliminating one of the two hyperparameters from the model reduces the computational time needed by almost one-third since it eliminates two of the six computationally intensive steps involving formation of Sy.

The use of nearest neighbors also accelerates computation. Suppose the indices to the neighboring observations lie in the nxm matrix G where $G_{1...m}$ equal individual index vectors. For a vector g the values of $N_1 g$ equal $g(G_1)$ and so forth. Using index arithmetic in place of matrix multiplication can greatly reduce computation time as indexing into a matrix is one of the faster digital operations.

Before turning to an illustrative example, we present some timing results for applications of both maximum likelihood and MCMC estimation of the MESS model, as well as more conventional spatial econometrics models. A generated data experiment involving a data set containing 3,107 US counties from Pace and Barry (1997) was used to perform the timing comparisons. There are 4 explanatory variables in the model and estimates were produced using a host of Bayesian variants of the spatial econometric models described in Anselin (1988). These Bayesian versions of the spatial regression models are presented in LeSage (1997) along with a description of estimation via MCMC sampling. The traditional spatial models used along with labels we rely on in presenting timing results are enumerated below.

SAR: $y = \rho W y + X\beta + \varepsilon$

SEM: $y = X\beta + u, \quad u = \lambda W u + \varepsilon$

SAC: $y = \rho W_1 y + X\beta + u, \quad u = \lambda W_2 u + \varepsilon$

These timing results reflect the use of a Monte Carlo estimator presented in Barry and Pace (1999) that approximates the log determinant over a grid of values for the spatial dependence parameter ρ in these models (see also Barry and Pace, 1997). Other computational enhancements were used, including sparse matrix algorithms from MATLAB along with special table look-up of the log determinant values during sampling evaluation of the conditional distributions. The timing results we report represent state-of-the-art algorithms implemented in MATLAB functions contained in the public domain *Spatial Econometrics Toolbox* available at http://www.spatial-econometrics.com. All computations were carried out on a 650 Mhz. Pentium III laptop running Windows 2000 and MATLAB version 5.3.

A number of variants on the Bayesian MESS model are also presented to provide some feel for the computational effort involved in determining various parameters in the model. For comparability with the SAR, SEM and SAC models, no spatial lags of the explanatory variables were used. That is, the model took the form: $Sy = X\beta + \varepsilon$. The various MESS models used in the experiments are enumerated below from simplest to most complex.

MESS1 - a model with both ρ and m fixed, and no v_i parameters.

MESS2 - a model with ρ fixed, m estimated and no v_i parameters.

MESS3 - a model with m fixed, ρ estimated and no v_i parameters.

MESS4 - a model with both ρ and m estimated and no v_i parameters.

MESS5 - a model with both ρ and m estimated as well as estimates for the v_i parameters.

All MCMC estimates involved 1,250 draws, which were adequate to produce converged estimates. The Bayesian SAR, SEM, and SAC models all invoked the heteroscedastic prior estimating a vector of v_i parameters. The SAR and SEM models relied on a first-order contiguity matrix, and the SAC model used a first-order matrix for W_1 and a second-order matrix for W_2. For the maximum likelihood MESS and Bayesian MESS1 models, the hyperparameter for the number of neighbors was set arbitrarily at 5. The other hyperparameter, ρ in these models was fixed at 0.9. In the case of the Bayesian MESS2 through MESS5 models, the grid search over ρ extended from 0.5 to 1, in 0.01 increments and from 4 to 10 neighbors in integer increments. These were thought to be representative of values practitioners might use in applications.

Table 1 shows total time taken as well as times required for component aspects involved in solving the estimation problem. The total time reported is greater than the sum of the component timings shown in the table because this involved computing measures of fit for the model such as $R-$squared, predicted values, residuals as well as means and standard deviations of the estimates based on the posterior distribution of draws. From the results reported in Table 1, the computational efficiency of the maximum likelihood MESS model is quite clear. It required only 1.16 seconds to solve this problem, with 0.85 seconds required to compute the five nearest neighbors and 0.28 seconds to compute a numerical hessian used to produce a variance-covariance measure of dispersion for the estimates.

The Bayesian models MESS1 through MESS4 hold a clear MCMC speed advantage over MCMC estimates for the conventional SAR, SEM, and SAC spatial econometric models. Keep in mind that the traditional models assume the spatial weight matrix W is fixed and known, making inferences conditional on this aspect of the problem. In contrast, the MESS2 through MESS5 models produce estimates for a flexible weight matrix specification as well as the other parameters in the model.

The times reported for the MESS5 model illustrate the computational costs of estimating the v_i variance scaling parameters that accommodate non-constant variance over space. Nonetheless, the Bayesian MESS models can be applied to spatial samples involving all US counties and produce estimates in two to six minutes time on a personal computer.

A large sample of 35,702 home sales from Lucas County, Ohio was used to produce another comparison of computational speed. These results are presented in Table 2 in the same format at that used in Table 1. This experiment used 1,250 draws and the same settings and grid ranges as in the previous experiment. This should provide a feel for the computational impact of changing the number of observations, ceteris paribus. A standardized first-order contiguity matrix was used for the SAR and SEM models that was very sparse, containing 214,168 non-zero elements from a total of 1,274,632,804 elements.

The number of observations in this experiment is 11.5 times that of the first experiment reported in Table 1. The average difference between the time required for the Bayesian SAR and SEM models here and in Ta-

Table 1: Timing results (in seconds) for 3,107 observations

Bayesian Traditional Models			
Model	total time	time for log(det) over a grid of θ values	sampling time
Bayes SAR	321.4	3.6	299.3
Bayes SEM	309.9	3.6	287.8
Bayes SAC	498.7	3.6	478.4
Maximum Likelihood Models			
Model	total time	time for neighbors calculation	time for hessian
Max Lik MESS	1.16	0.85	0.28
Bayesian MESS models			
Model	total time	time for neighbors plus hyperparameters grid	sampling time
Bayes MESS1	45.2	0.88	43.9
Bayes MESS2	80.4	1.16	79.2
Bayes MESS3	81.9	2.58	79.2
Bayes MESS4	126.4	18.2	108.1
Bayes MESS5	392.5	17.7	374.5

Table 2: Timing results (in seconds) for 35,702 observations

Bayesian Traditional Models			
Model	total time	time for log(det) over a grid of θ values	sampling time
Bayes SAR	3545.1	41.8	3503.0
Bayes SEM	3491.1	41.7	3449.1
Maximum Likelihood Models			
	total time	time for neighbors calculation	time for hessian
Max Lik SAR	555.2	—	8.3
Max Lik MESS	15.2	10.9	3.7
Bayesian MESS models			
	total time	time for neighbors plus hyperparameters grid	sampling time
Bayes MESS1	569.5	10.7	557.9
Bayes MESS2	1031.0	40.1	988.7
Bayes MESS3	1020.6	34.5	985.4
Bayes MESS4	1600.5	207.0	1381.5
Bayes MESS5	4959.2	206.2	4749.4

ble 1 was 11.14, so the time required for these two models scale linearly in the number of observations. Similarly, the average time required for the Bayesian MESS models here relative to Table 1 is 12.6, so again the models scale approximately linearly in the number of observations.

An interesting result with regard to the Bayesian MESS1 model is that estimation via MCMC sampling took 569.5 seconds compared to 555.2 seconds for maximum likelihood estimation of the traditional SAR model. Both of these models take the spatial weight matrix as given, as does the Bayesian SAR model implemented using MCMC methods. Here we see a speed improvement of over 6 times for the Bayesian MESS1 model relative to the Bayesian SAR model. This improvement in computation times is attributable to the elimination of the log determinant term in the likelihood function and the conditional distributions involved in the MESS models.[2]

One caveat regarding these timing results is that they represent the use of MATLAB, an interpreted matrix programming language rather than a compiled language such as FORTRAN or C. Timing results based on a FORTRAN implementation of maximum likelihood MESS estimation suggest a six-fold improvement in timing results would arise from using a compiled language in place of MATLAB.

In summary, the Bayesian MESS model introduced here is computationally feasible for very large spatial problems and MCMC estimation methods are around six times as fast as Bayesian variants of the more traditional spatial econometric models implemented via MCMC sampling.

4 Illustrations of the Bayesian MESS model

We provide illustrations of the Bayesian MESS model in section 4.1 using a generated model with only 49 observations taken from Anselin (1988). Use of a generated example where the true model and parameters are known allows us to illustrate the ability of the model to find the true spatial weight structure used in generating the model.

One point we wish to illustrate here is that both maximum likelihood and Bayesian MESS models can be used to estimate relationships generated using the more traditional spatial regression models. Inferences regarding spatial dependence, disturbance noise variance and the regression parameters produced by the MESS models will be the same as those drawn from

[2]Some of the improvement comes from the use of nearest neighbors in the MESS model along with index calculations. The index calculations double the speed relative to sparse matrix calculations.

estimates based on the more traditional models. Because of the computational speed advantages associated with the MESS model, this seems a highly desirable situation.

Section 4.2 provides an illustration where inferences regarding the magnitude and extent of spatial influence may be of interest. The magnitude and extent of spatial knowledge spillovers in a sample of 219 northeast Ohio urban zip-code areas is used to illustrate inference regarding the spatial weight structure of the model.

4.1 A generated data example

A traditional spatial autoregressive (SAR) model: $y = \theta W y + X\beta + \varepsilon$ was used to generate the vector y based on 49 spatial observations from Columbus neighborhoods presented in Anselin (1988). The spatial weight matrix, $W = \sum_{i=1}^{m} \rho^i N_i / \sum_{i=1}^{m} \rho^i$, was based on $m = 5$ nearest neighbors and distance decay determined by $\rho = 0.9$. The two explanatory variables from Anselin's data set (in studentized form) along with a constant term and θW were used to generate a vector $y = (I_n - \theta W)^{-1} X\beta + (I_n - \theta W)^{-1}\varepsilon$. The parameters β and the noise variance, σ_ε^2 were set to unity and the spatial correlation coefficient θ was set to 0.65.

This generated data was used to estimate parameters based on a maximum likelihood SAR model, maximum likelihood MESS and Bayesian MESS models. Of course, traditional implementation of the SAR model would likely rely on a first-order contiguity matrix treated as fixed sample information, which we label W_1. Maximum likelihood MESS would attempt to determine values for the hyperparameters ρ, m using a profile likelihood grid search over these values. Bayesian MESS would produce posterior estimates for the hyperparameters as part of the MCMC estimation as in the models labeled MESS4 and MESS5 in the previous section. Of course, it would be possible to rely on MCMC estimation and the simpler models labeled MESS1 to MESS3 in the previous section, but the computational requirements for this small sample are minimal.

We illustrate the difference in estimates and inferences that arise from using these three approaches. Note that two variants of the SAR model were estimated, one based on a first-order contiguity matrix, W_1 and another based on the true W matrix used to generate the model. In practice of course, one would not know the true form of the W matrix. One point to note is that the first-order contiguity matrix for this data set contains an average number of neighbors equal to 4.73 with a standard deviation of 1.96. Of the total 49x49=2,401 elements there are 232 non-zero entries. We might expect that the differences between SAR models based on W_1 and the true W containing five nearest neighbors and a small amount of distance decay should be small.

The maximum likelihood MESS model searched over a grid of ρ values from 0.01 to 1 in 0.01 increments and neighbors m ranging from 1 to 10. Estimates were produced based on the values of ρ and m that maximized the profile log likelihood function. The Bayesian MESS model was run to produce 5500 draws with the first 500 discarded to allow the MCMC chain to converge to a steady state.[3] Diffuse priors were used for β and σ and two variants of the model were estimated: one that included the parameters V and another that did not. The latter Bayesian model assumes that $\varepsilon \sim N(0, \sigma^2 I_n)$, which is consistent with the assumption made by the maximum likelihood SAR and MESS models.

The estimation results are presented in Table 3. Measures of precision for the parameter estimates are not reported in the table because all coefficients were significant at the 0.01 level. In the table we see that the SAR model based on the true spatial weight matrix W performed better than the model based on W_1, as we would expect. (True values used to generate the data are reported in the first column next to the parameter labels). Both the profile likelihood approach and the posterior distribution from the Bayesian MESS models identified the correct number of neighbors used to generate the data. The Bayesian MESS models produced posterior estimates for ρ based on the mean of the draws equal to 0.91 and 0.89 compared to the true value of 0.90, whereas the profile likelihood search resulted in an estimate of $\rho = 1.0$. Nonetheless, the MESS models produced very similar β estimates as well as estimates for the spatial dependence parameter in this model, α. The estimate of σ^2 from one Bayesian MESS model was close to the true value of unity, while the other Bayesian model produced an estimate closer to that from the maximum likelihood SAR model based on the true W matrix.

The profile likelihood approach identified the correct number of neighbors used to generate the data and points to a value of $\rho = 1$, versus the true value of 0.9. For contrast, the posterior distribution of ρ is presented in Figure 1 where a kernel density estimate based on the 5,000 draws retained from the MCMC sampler was used. This posterior was skewed, having a mean of 0.9171, a median of 0.9393 and a mode of 0.9793. This partially explains the difference between the maximum likelihood estimate of unity

[3]This is actually an excessive number of draws, since the estimates were the same to one or two decimal places as those from a sample of 1250 draws with the first 250 discarded.

Table 3: A comparison of models from experiment 1

	SAR W_1	SAR W	ML MESS	MESS4	MESS5
Variable					
$\beta_0 = 1\dagger$	1.3144	1.1328	1.1848	1.1967	1.1690
$\beta_1 = 1$	1.1994	0.9852	1.0444	1.0607	1.0071
$\beta_2 = 1$	1.0110	1.0015	1.0144	1.0102	0.9861
$\sigma^2 = 1$	1.4781	0.7886	0.8616	0.9558	0.7819
$\theta = 0.65$	0.5148	0.6372			
α			-0.8879	-0.8871	-0.9197
R^2	0.8464	0.9181	0.9160	0.9141	0.9134
$m = 5$			5	5.0466	5.0720
$\rho = 0.90$			1.0	0.9171	0.8982

† true values used to generate the data.

and the Bayesian estimate reported in Table 3. It should be clear that the posterior distributions for the hyperparameters ρ and m provide a convenient summary that allows the user to rely on mean, median or modes in cases where the resulting distributions are skewed.

In most applications a simple histogram of the draws would suffice, and we illustrate this for the posterior distribution of the draws for m in Figure 2. In addition to posterior point estimates for the hyperparameters based on measures of central tendency, measures of dispersion can be easily calculated using the standard deviation of the MCMC draws, or quantiles calculated from a count of the draws within upper and lower limits.

Note that the parameter α in the MESS model plays the role of θ in the traditional SAR, SEM, and SAC models capturing the extent of spatial dependence. Inferences about spatial dependence are based on a test of the magnitude of α versus zero. Figure 3 shows the posterior distribution of α from the MESS4 model, which should make it clear that this estimate would lead to an inference of spatial dependence, that is, $\alpha \neq 0$.

As an illustration of the ability of the MESS model to find the correct model specification, we produced estimates for models based on a first-order contiguity matrix used to generate the data in this experiment as well as models based on the two through six nearest neighbors. Note that use of spatial weight matrices based on nearest neighbors represents a misspecification since the first-order contiguity matrix was used to generate the dependent

Table 4: Specification search example involving six models

# of neighbors	ML MESS Log Likelihood	MCMC MESS Posterior Probability
Correct W matrix	-75.7670	0.9539
2 neighbors	-85.0179	0.0001
3 neighbors	-80.2273	0.0094
4 neighbors	-81.9299	0.0017
5 neighbors	-79.3733	0.0247
6 neighbors	-80.3274	0.0102

variable vector y. No hyperparameters were used in this experiment, so the specification: $W_i = e^{\alpha N_i}, i = 1, \ldots, 6$, was used, where N_i denotes a nearest neighbor weight matrix based on i neighbors.

The question of interest here is whether the MESS models can distinguish the first-order contiguity matrix used to generate the data from the nearest neighbor matrices. Posterior probabilities for these six models are shown in Table 4 for the Bayesian MESS and the log likelihood function values are shown for the maximum likelihood MESS model.[4] From the table we see that the MESS models correctly identified the model associated with the true weight matrix. Almost all of the posterior probability weight was placed on this model, indicating that the flexibility associated with a specification that allows varying the number of neighbors did not lead the model to pick an inferior spatial weight structure when confronted with the true structure.

Relatively diffuse priors along with a prior reflecting a belief in constant variance across space were used in the experiments above to illustrate that the Bayesian MESS model can replicate maximum likelihood estimates. This is however a computationally expensive approach to producing MESS estimates. A practical motivation for the Bayesian model would be cases involving outliers or non-constant variance across space. To illustrate the Bayesian approach to non-constant variance over space we compare six models based on alternative values for the hyperparameter r that specifies our prior on heterogeneity versus homogeneity in the disturbance variances. These tests are carried out using two data sets, one with homoscedastic and another

[4]Posterior probabilities can be computed using a sample of marginal log likelihood function values generated during the MCMC sampling as described in Kass and Raftery, 1995.

Table 5: Homogeneity test results for two data sets

r value	Homoscedastic data Posterior Probabilities	Heteroscedastic data Posterior Probabilities
50	0.9435	0.0001
20	0.0554	0.0039
10	0.0011	0.2347
7	0.0001	0.6200
4	0.0000	0.1413
1	0.0000	0.0000

with heteroscedastic disturbances. Non-constant variances were created by scaling up the noise variance for the last 20 observations during generation of the y vector. This might occur in practice if a neighborhood in space reflects more inherent noise in the regression relationship being examined. The last 20 observations might represent one region of the spatial sample.

We test a sequence of declining values for r with large values reflecting a prior belief in homogeneity and smaller values indicating heterogeneity. Posterior probabilities for these alternative values of r are shown in Table 5 for both sets of generated data. For the case of constant variances, the posterior model probabilities correctly point to a model based on large r values of 50. In the case of heteroscedastic disturbances, the models based on r values of 10, 7 and 4 receive high posterior probability weights, reflecting the non-constant variance.

In addition to correctly identifying the existence of heterogeneity in the disturbance variances, a plot of the posterior means of the v_i estimates can be a useful diagnostic regarding the nature and extent of the heterogeneity. Figure 4 shows a plot of these estimates for the Bayesian MESS model as well as the Bayesian SAR model. From the figure we see that the pattern of inflated variances over the last 20 observations is correctly identified by the v_i estimates from both models.

4.2 An application

As an example of a study with potential spatial aspects, Glaeser, Kallal, Scheinkman, and Shleifer (1992) (GKSS hereafter) examine employment and earnings growth of large industries in 170 U.S. cities between 1956 and 1987.

They find that local competition and urban diversity are positively associated with industry-level employment and earnings growth while regional specialization is not significantly associated with employment or earnings growth. Their motivation for examining the role of these factors is that theories of economic growth have increasingly stressed technological spillovers as a source of growth (see Romer, 1994; Grossman and Helpman, 1994). Geographical proximity suggests that knowledge spillovers take place in urban areas where transmission of ideas is rapid and convenient because of the high density of individuals, occupations and industries.

Discussions of spillovers naturally suggest a spatial variant of the GKSS model where the number of neighbors and the rate of decay of influence over space might be a focus of inference. To examine this, we estimate a GKSS style model using employment and earnings growth data over the period from the first quarter of 1989 to the first quarter of 1998 in 219 zip-code areas in northeastern Ohio.[5] These zip-code areas included Akron, Canton, Cleveland, and Youngstown-Warren in the following way. Because our interest centers on earnings growth in urban and surrounding areas, we relied on the U.S. Postal Service designations of zip codes associated with the four cities: Akron, Canton, Cleveland, and Youngstown-Warren. We also included suburban areas using a mapping program to determine zip-code areas that shared borders with the urban zip-code areas. Geographers refer to these neighboring areas as "first-order contiguous". Also included in the sample were areas defined by zip-codes that were neighbors to the suburban areas, i.e., have shared borders with the suburban areas. These would represent "second-order contiguous" areas with respect to the urban areas. Figure 5 shows a map of the 219 contiguous zip-code areas resulting from this configuration.

The dependent variable in the model is the growth rate of aggregate zip-code area employment or payroll earnings over the period 1989 first quarter

[5]This study draws on a preliminary set of estimates for quarterly employment, payroll and establishments at the zip-code area level. This new source of labor market information was constructed by a network of university researchers coordinated by the Ohio Urban Universities Program. In cooperation with the Ohio Bureau of Employment Services, establishment level information (ES202 data) on quarterly employment and payroll reported by firms for unemployment insurance purposes has been developed to form a database that contains employment, nominal payroll and the number of establishments in each of Ohio's 1,008 zip code areas. The raw ES202 data information represents an administrative database that is not immediately useful for the type of analysis in this study. The initial ES202 information was statistically transformed to facilitate the analysis we undertake in the sequel. For a detailed report on the statistical methods and transformations employed, see LeSage (1999).

to 1998 first quarter, computed as $\log(y_T) - \log(y_0)$. Following GKSS a set of explanatory variables was constructed to reflect:

1. **diversity** of geographically neighboring industries as this may promote innovation and growth if knowledge transfers come from outside the core industry. GKSS attribute these ideas to Jacobs (1969).

2. **specialization** or geographical concentrations of industries which might promote growth if technology transfers take place through spying and interfirm movement of skilled labor. Marshall (1890), Arrow (1962) and Romer (1986) are all credited by GKSS with making cases for this type of technology externalities.

3. **competition** which would favor Jacobs case for diversity, whereas a lack of competition might be consistent with specialization or geographical concentration suggested by Marshall, Arrow and Romer.

4. **initial conditions** which might be an important determinant of employment and earnings growth rates. Geographic areas with large initial endowments of employment (payroll) should exhibit lower growth rates than areas with lower initial levels of employment (payroll).

Although GKSS motivate their model and relationships in terms of geographically neighboring industries and firms, geographical concentration and spatial spillovers or technology transfers, their data sample involves a cross-section of 170 US cities and the estimation methodology was ordinary least-squares. Here we use a zip-code area spatial data sample that allows for a finer spatial scale where the geographical influence of neighboring areas on employment and earnings growth might be detected. The Bayesian MESS model in (12) was estimated here to take account of competition, diversity and specialization in neighboring zip-code areas as well as the area itself.

$$Sy = \begin{bmatrix} X & WX \end{bmatrix} \beta + \varepsilon \tag{12}$$

In addition, the model estimates the extent of influence over space using the parameterized version of the spatial weight matrix from (5) allowing us to draw inferences about this issue.

The variable measuring competition was constructed by dividing the number of establishments per employee for the largest 2-digit industry in each zip-code area by the number of establishments per employee for this industry in the state of Ohio. Values of this variable greater than one would indicate a more competitive environment for the largest industry in the zip-code area than in the state as a whole. Smaller values indicate less competition. This variable is similar to that used by GKSS to measure competition and was based on initial period employment and establishments.

As a measure of specialization we followed GKSS and used a ratio of the proportion of total area employment contained in the largest 2-digit industry for each zip-code area divided by the proportion of total state employment held by this industry at the statewide level. Values greater than one would indicate a larger degree of specialization in this industry than the average industrial specialization in the state. Again, initial period values were used.

Diversity was measured as in GKSS using employment in the largest five 2-digit industries during 1989Q1 in each zip-code area as a share of total zip-code area employment. Larger magnitudes for this variable reflect less industrial diversity in the zip-code area.

To control for initial conditions the log-level of payroll (employment) during the initial period 1989Q1 was entered in the earnings (employment) growth equation. A negative sign for the coefficient on this variable would indicate convergence in growth between the zip-code areas over time. In addition, establishment growth rates (calculated as $(\log(y_T) - \log(y_0))$ were used as a control variable.

Estimates based on 10,500 draws (with the first 500 used as "burn-in" draws for the MCMC sampler) are shown in Table 6. In addition to Bayesian MESS estimates based on an assumption of normal disturbances, we compare these results to estimates from a non-spatial least-squares model.

The parameter α in the model indicates significant spatial dependence for both MESS models, consistent with the importance of neighboring zip-code areas. The posterior distributions for α from both employment and earnings relations are shown in Figure 6, where we see clear evidence of spatial dependence. Evidence regarding the posterior mean, median and mode as well as the variance for the spatial dependence parameter in both relations is remarkably similar, as we might expect because of the reasonably high correlation between employment and earnings.

There are numerous motivations for drawing inferences regarding the spatial structure of the relationship between growth and variables that describe the environment in which firms operate. An important part of the technology transfer argument is that knowledge spillovers result from employment density. Anas, Arnott and Small (1998) provide a cogent case for the impact of differing spatial scales on employment density. They present a graphical depiction of employment density in Los Angeles County using three different degrees of spatial averaging. The density appears homoge-

Table 6: A comparison of earnings growth models

	least-squares		Bayesian MESS	
Variable	Coeff.	p-level	Coeff.	p-level†
constant term	1.7765	0.0000	-0.4910	0.3482
log(1989 level)	-0.0699	0.0001	-0.0886	0.0000
establishment growth rate	0.4629	0.0000	0.4181	0.0000
diversity	-0.9069	0.0057	-0.9039	0.0021
competition	0.0866	0.0725	0.0699	0.0748
specialization	0.1065	0.2017	0.1198	0.0749
W (log(1989 level))			0.0960	0.0106
W (est growth rate)			-0.0377	0.4064
W (diversity)			2.6342	0.0457
W (competition)			-0.1696	0.1747
W (specialization)			-0.1037	0.3454
Hyperparameters			mean	std.
α			-0.4838	0.1913
ρ			0.8994	0.0859
# neighbors			13.0271	3.6813
adjusted R^2	0.3296		0.4026	
σ_ε^2	0.1446		0.1348	

† Bayesian p-levels, see Gelman, Carlin, Stern and Rubin (1995)

Table 7: A comparison of employment growth models

	least-squares		Bayesian MESS	
Variable	Coeff.	p-level	Coeff.	p-level†
constant term	0.6086	0.0009	-0.3421	0.3331
log(1989 level)	-0.0411	0.0084	-0.0472	0.0058
establishment growth rate	0.4452	0.0000	0.4208	0.0000
diversity	-0.8874	0.0004	-0.8183	0.0009
competition	0.1215	0.0010	0.1217	0.0010
specialization	0.0229	0.7201	0.0142	0.4132
W (log(1989 level))			0.0542	0.0824
W (est growth rate)			-0.0917	0.2556
W (diversity)			1.7025	0.1015
W (competition)			-0.1969	0.0934
W (specialization)			0.0687	0.3609
Hyperparameters			mean	std.
α			-0.4814	0.2056
ρ			0.9143	0.0759
# neighbors			13.7074	3.6374
adjusted R^2	0.4216		0.4675	
σ^2_ε	0.0851		0.0821	
† Bayesian p-levels, see Gelman, Carlin, Stern and Rubin (1995)				

nous at a course scale, but contains a great deal of intricate structure when viewed at finer scales. The traditional practice that fixes spatial weights a priori may reflect a coarse spatial scale that obscures differences in urban, suburban and outlying employment density and their impact on the economic environment in which firms operate.

A second motivation that the extent of spatial influence is important when considering technology transfer comes from the sociological literature on social networks. Granovetter (1974) using a sample of 282 male professional and technical workers in the Boston area finds that 57% of current jobs were found through personal contacts or referrals. Corcoran et al. find similar results using a larger sample from the 1978 wave of the PSID. Wellman (1996) finds that 38% of all active contacts take place between persons that live less than 1 mile apart, and 64% involve agents living within 5 miles of each other. Combining these two findings, we have a case that the spatial structure may need to reflect the appropriate distance and physical proximity to capture social interactions of the type that lead to spillovers.

Third, Jaffe et al. (1993) point out that citation of patents can be used to document knowledge flows. Patent citations reference previous ideas underlying development of the patent. In their study (Jaffe et al., 1993) matched company citations of university patents by states and MSAs finding strong evidence that citations of university patents were geographically localized near the academic institutions. This seems to reinforce the second motivation regarding distance and social interaction.

Fourth, if diversity, competition and specialization are important factors determining knowledge spillovers, firms my attempt to relocate within a given metropolitan region to take advantage of different propensities for spillover in various geographic areas within the region. Again, an a priori setting of the spatial weights may not reflect an appropriate scale to capture these movements within the region.

To illustrate how one might address these issues, we draw inferences on the spatial extent of influence from neighbors using the hyperparameters ρ and m from the model. The parameter ρ reflects distance decay and m indicates the number of neighbors, which together determine the spatial weight structure in the Bayesian MESS model. The means and standard deviations reported for ρ in Table 6 need to be considered in light of the skewed asymmetric posterior distributions shown in Figure 7. The mean and median ρ from the earnings equation are: 0.8994 and 0.9217, respectively, while the mean and median ρ values for the employment relation are: 0.9143 and 0.9361. Using a value of 0.91 between the mean and median for the earnings relation, the half-life would be 9 neighbors. That is, the 9th nearest neighbor would exert an influence half that of the nearest neighbor. Similar conclusions would hold for the employment relation where $\rho = 0.92$ represents a value midway between the mean and median posterior estimates.

This information in conjunction with posterior information for the parameter m reflecting the number of neighbors allows us to draw inferences about the spatial extent of influence arising from the economic environmental variables reflecting competition, diversity and specialization. The posterior means for the neighbor parameters in the earnings and employment relations were 13.0271 and 13.7074 and the medians were 13 and 14. This should be interpreted in conjunction with the posterior information for the hyperparameter ρ, that suggests a half-life of nine nearest zip-code areas. There are 46 zip-code areas in Akron (including areas that are first and second order contiguous to the central urban areas), 38 in Canton, 76 in Cleveland and 54 in Youngstown-Warren, so we might interpret 13 or 14 neighbors as representing around one-third of the smaller cities, one fourth of Youngstown-Warren and one sixth of Cleveland. That is, the scope of influence from the economic environment operating in neighboring zip-code areas is fairly small, and this influence decays to half by the time we extend out to nine neighboring zip-code areas. Another way to put this into perspective is to note that the average number of first-order contiguous neighbors (borders touching) for the sample of zip-code areas was 6.12 and the standard deviation was 1.40, suggesting that spillover influences of any magnitude extend slightly beyond first-order contiguous zip-code areas.

This inference suggests that spillover influences may be constrained to relatively small areas consistent with the social interaction literature discussed above. Intuitively, many of the same factors that work to make distance and physical proximity an important determinant of social interaction may also work to constrain firm interaction. From a policy perspective, these results are interesting because economic development officials frequently argue that spillover benefits arise from development projects that impact the larger geographic region in which they operate.

Turning attention to the least-squares estimates we find they are in agreement with those from GKSS for the three variables competition, diversity and specialization. Local competition is significant and positively associated with earnings and employment growth. The least-squares coefficient for specialization is also consistent with results from GKSS in that it is not significantly associated with earnings or employment growth. Diversity is negative and significant for both the earnings and employment growth relations. Note that larger values of the diversity variable reflect less diversity, hence the negative coefficient indicating that more diversity in a zip-code area promotes growth.

Note however that the MESS estimates for the spatial lags are not consistent with results from GKSS for the diversity variable, indicating that taking the spatial influence of neighboring zip-code areas into account may affect our inference about this variable. One might expect some differences between our spatial results and those of GKSS because: 1) our data represents a finer spatial scale, 2) our dependent variable represents aggregate zip-code level employment and earnings growth unlike the industry-level variables used by GKSS, and 3) the change in spatial scale from cities to contiguous zip-code areas results in measures of diversity, specialization and competition that also exhibit a finer spatial scale.

Finally, the estimates for the coefficient reflecting the log of initial employment and payroll in the two relations is negative and significant in least-squares and MESS, indicating zip-code level convergence in both earnings and employment growth over time, a result consistent with GKSS. The initial levels in neighboring areas is positive and weakly significant in the earnings equation but not in the case of employment. This suggests that high earnings levels in neighboring areas would lead to higher earnings growth, which seems intuitively plausible. Establishment growth exerts a positive and significant impact on employment growth, but not on earnings growth, again a plausible result.

5 Conclusion

We argue that the magnitude and extent of spatial influence exerted by variables in spatial regression relationships may represent a subject of interest. A flexible specification based on hyperparameters for the number of neighboring entities and decay of influence over space was combined with the matrix exponential spatial specification (MESS) introduced in Pace and LeSage (2000). The resulting Bayesian model allows posterior inferences regarding the magnitude and extent of spatial influence as well as traditional inferences about spatial dependence and the role of explanatory variables in the spatial regression relationship.

The reliance of our model on a flexible specification for the spatial weight structure produces a situation where the Bayesian approach to estimation introduced here holds some advantages over maximum likelihood approaches. For example, posterior inferences regarding the parameters of the weight structure can be obtained from the MCMC estimation method used to implement the Bayesian MESS model. Another advantage to the Bayesian approach is that non-constant variance over space and spatial outliers can be handled in an eloquent fashion. Finally, the Bayesian variant of the MESS model introduced here could be extended to the case of limited or censored dependent variable models using the methods introduced in LeSage (2000).

Experiments carried out in this study suggest that accurate posterior inferences regarding parameters that specify the flexible spatial weight structure are possible using the MCMC estimation methods. Unlike conventional approaches to spatial econometric models, the approach taken here does not assume the spatial weight matrix is part of the sample data information. We provide some evidence that the flexible weight structure in the MESS model can replicate results from more traditional spatial regression models, even in cases were the estimated model was different from the data generating process. This suggests that the MESS model may be flexible enough to replicate more traditional spatial econometric regression models, while avoiding many of the computational problems that plague those models. However, inferences regarding β and σ as well as spatial dependence drawn from a traditional model based on a fixed matrix W are conditional on the particular W matrix employed in the model. The model we introduce produces inferences regarding β and σ as well as spatial dependence that are conditional only on a family of spatial weight transformations that we denote Sy, where $S = e^{\alpha W}$, with the matrices W taking a flexible form. The Bayesian approach introduced here produces a posterior distribution for the joint distribution of the parameters in the spatial weight structure as well as the other model parameters.

The primary disadvantage of the Bayesian approach is the need to rely on MCMC estimation procedures which are computationally intense by comparison with the maximum likelihood approach to the MESS model. Nonetheless, MCMC estimation of the Bayesian MESS model exhibited a two- to six-fold increase in computational speed over MCMC estimation of Bayesian versions of traditional spatial econometric models introduced in LeSage (1997). Moreover, as demonstrated in section 3.3, the Bayesian MESS technique can handle applications with many observations (i.e., 35,702 in Table 2).

References

Anas, Alex, Richard Arnott and Kenneth A. Small, (1998) "Urban Spatial Structure," *Journal of Economic Literature*, Vol. 36, no. 3, pp. 1426-1464.

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

Arrow, Kenneth J. (1962) "The Economic Implications of Learning by Doing," *Review of Economic Studies*, pp. 155-73.

Barry, Ronald, and R. Kelley Pace, (1997) "Kriging with Large Data Sets Using Sparse Matrix Techniques," *Communications in Statistics: Computation and Simulation*, Volume 26, Number 2, pp. 619-629.

Barry, Ronald, and R. Kelley Pace, (1999) "A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices," *Linear Algebra and its Applications*, Volume 289, Number 1-3, pp. 41-54.

Chiu, Tom Y.M., Tom Leonard, and Kam-Wah Tsui, (1996) "The Matrix-Logarithmic Covariance Model," *Journal of the American Statistical Association*, 91, pp. 198-210.

Corcoran, Mary, Linda Datcher and Greg Duncan, 1980. "Information and Influence Networks in Labor Markets," in *Five Thousand American Families: Patterns of Economic Progress*, George Duncan and James Morgan (eds.), vol. 7, p. 1-37, (Ann Arbor, MI: Institute for Social Research).

Gelfand, Alan E., and A.F.M Smith. (1990) "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Vol. 85, pp. 398-409.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995) *Bayesian Data Analysis*, (Chapman & Hall: London).

Geweke, John. (1993) "Bayesian Treatment of the Independent Student t Linear Model," *Journal of Applied Econometrics*, Vol. 8, pp. 19-40.

Glaeser, Edward L, Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer, (1992) "Growth in Cities," *Journal of Political Economy*, Vol. 100, number 61, pp. 1127-1152.

Granovetter, Mark S., 1974. *Getting a Job: A Study of Contracts and Careers,* (Cambridge, MA: Harvard University Press).

Grossman, Gene M. and Elhanan Helpman, (1994) "Endogenous Innovation in the Theory of Growth," *The Journal of Economic Perspectives*, Vol. 8, no. 1, pp. 23-44.

Jacobs, Jane, (1969) *The Economy of Cities*, (New York: Vintage).

Jaffe, A., M. Trajtenberg and R. Henderson, (1993) "Geographic localization of knowledge spillovers as evidenced by patent citations," *Quarterly Journal of Economics*, Vol. 108, pp. 577-598.

Kass, R.E., and A. E. Raftery (1995) "Bayes Factors," *Journal of the American Statistical Association*, Volume 90, pp. 773-795.

Kelejian, H., and I.R. Prucha, (1998) "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate and Finance Economics*, Volume 17, Number 1, pp. 99-121.

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

LeSage, James P. (1997) "Bayesian Estimation of Spatial Autoregressive Models," *International Regional Science Review*, Volume 20, number 1&2, pp. 113-129.

LeSage, James P., 1999. "A Report on Statistical Cleaning of the Ohio UUP ES202 data," Urban Affairs Center, University of Toledo.

LeSage, James P., (2000) "Bayesian Estimation of Limited Dependent variable Spatial Autoregressive Models," *Geographical Analysis*, Volume 32, number 1, pp. 19-35.

Marshall, Alfred. 1890. *Principles of Economics*, (London: Macmillan).

Pace, R. Kelley, and Dongya Zou, (2000) "Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence," *Geographical Analysis*, Volume 32, Number 2, pp. 154-172.

Pace, R. Kelley. and J. P. LeSage, (2000), "Closed-Form Maximum Likelihood Estimates for Spatial Problems," unpublished manuscript available at http://www-spatial-statistics.com.

Romer, Paul M. (1986) "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, Vol. 94, pp. 1002-37.

Romer, Paul M., (1994) "The Origins of Endogenous Growth," *The Journal of Economic Perspectives*, Vol. 8, no. 1, pp. 3-22.

Wellman, Barry, (1996) “Are Personal Communities Local? A Dumptarian Reconsideration,” *Social Networks*, Vol. 18, pp. 347-354.

Wolpert, Robert, (2000) “On the Probability of Observing Misleading Statistical Evidence: A Comment,” *Journal of the American Statistical Association*, Vol. 95, pp. 771-772.

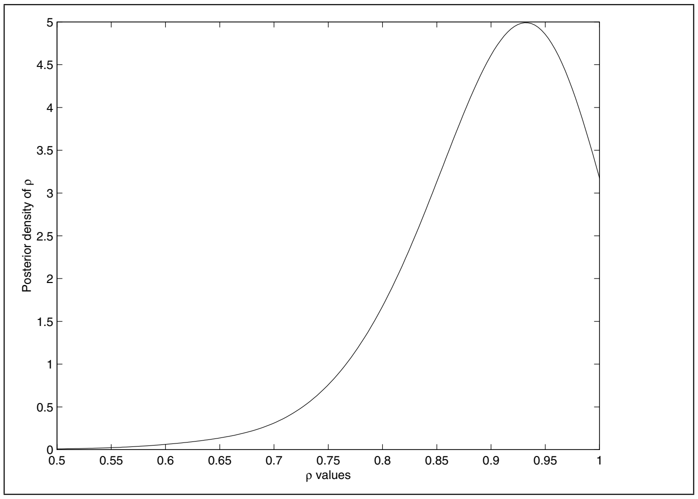

Figure 1: Posterior density for ρ hyperparameter

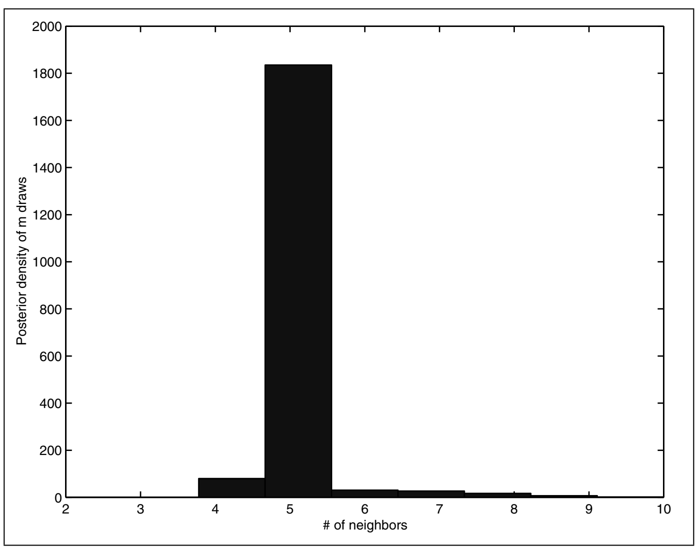

Figure 2: Histogram of the draws for m

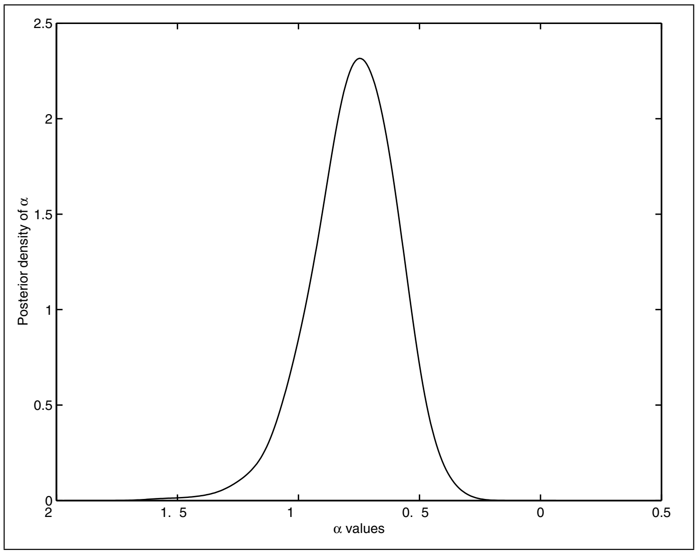

Figure 3: Posterior distribution of α parameter

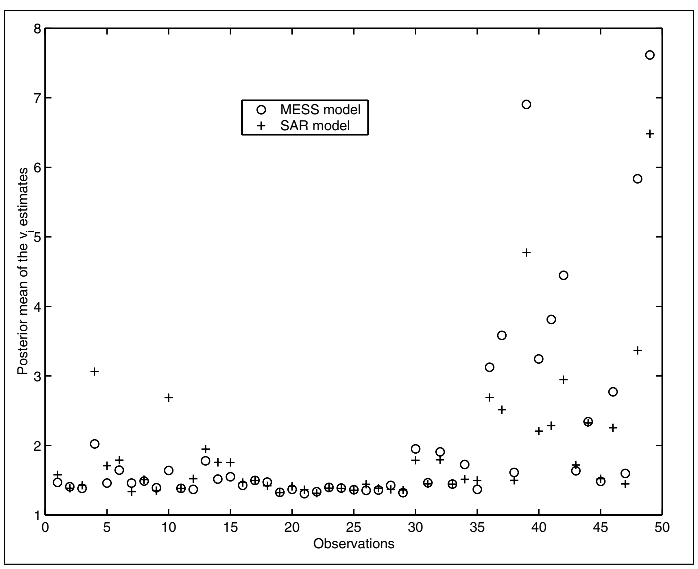

Figure 4: Posterior means of the v_i estimates for a heteroscedastic model

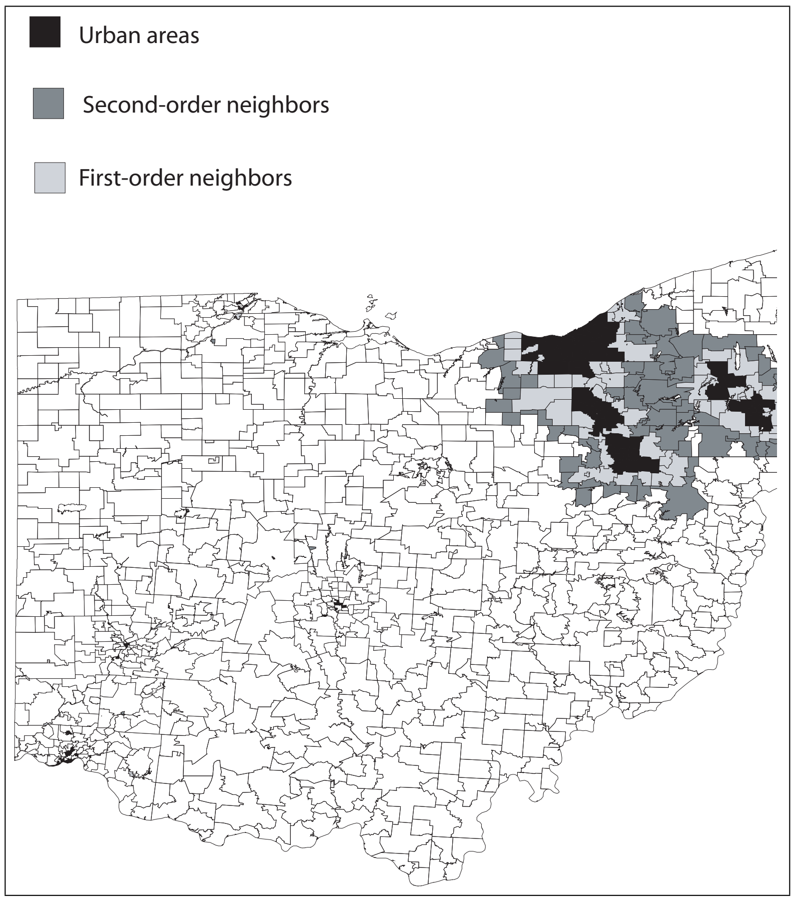

Figure 5: Zip-code areas in northeast Ohio used for employment and earnings growth models

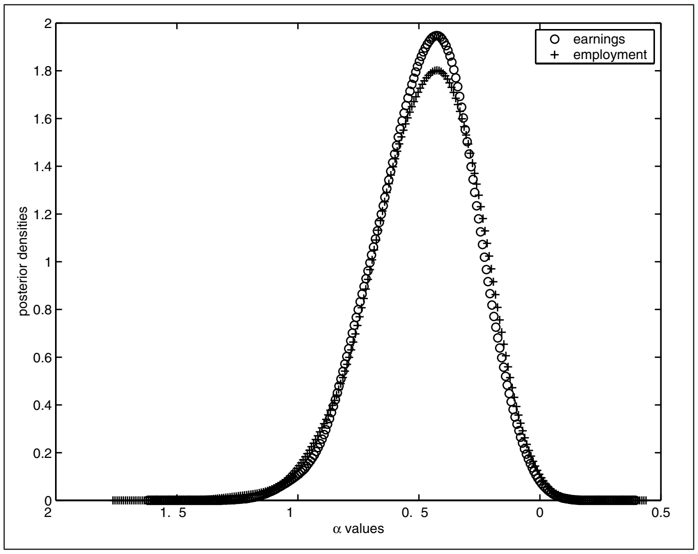

Figure 6: Posterior distribution for α in both employment and earnings growth models

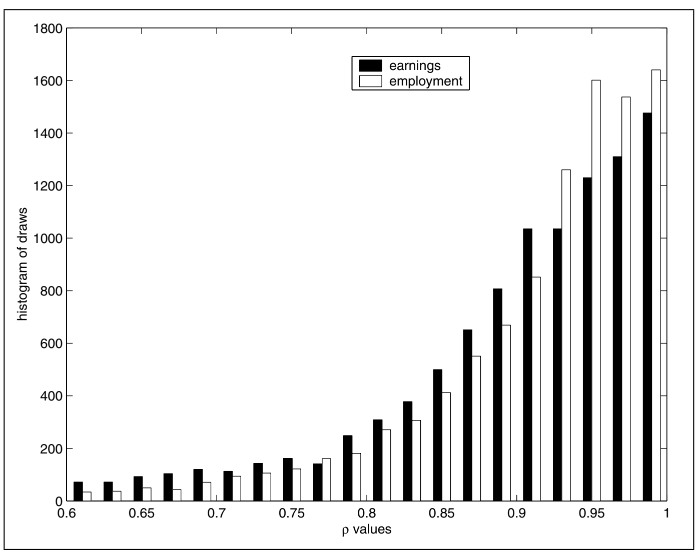

Figure 7: Posterior distribution for ρ in both employment and earnings growth models