A Family of Geographically Weighted Regression Models

James P. LeSage
Department of Economics
University of Toledo
2801 W. Bancroft St. Toledo, Ohio 43606
e-mail: jpl@spatial-econometrics.com

November 19, 2001

Abstract

A Bayesian treatment of locally linear regression methods introduced in McMillen (1996) and labeled geographically weighted regressions (GWR) in Brunsdon, Fotheringham and Charlton (1996) is set forth in this paper. GWR uses distance-decay-weighted sub-samples of the data to produce locally linear estimates for every point in space. While the use of locally linear regression represents a true contribution in the area of spatial econometrics, it also presents problems. It is argued that a Bayesian treatment can resolve these problems and has a great many advantages over ordinary least-squares estimation used by the GWR method.

1 Introduction

A Bayesian approach to locally linear regression methods introduced in McMillen (1996) and labeled geographically weighted regressions (GWR) in Brunsdon, Fotheringham and Charlton (1996) is set forth in this paper. The main contribution of the GWR methodology is use of distance weighted sub-samples of the data to produce locally linear regression estimates for every point in space. Each set of parameter estimates is based on a distance-weighted sub-sample of "neighboring observations", which has a great deal of intuitive appeal in spatial econometrics. While this approach has a definite appeal, it also presents some problems. The Bayesian method introduced here can resolve some difficulties that arise in GWR models when the sample observations contain outliers or non-constant variance.

The distance-based weights used in GWR for data at observation i take the form of a vector W_i which can be determined based on a vector of distances d_i between observation i and all other observations in the sample. Note that the symbol W is used in this text to denote the spatial weight matrix in spatial autoregressive models, but here the symbol W_i is used to represent distance-based weights for observation i, consistent with other literature on GWR models. This distance vector along with a distance decay parameter are used to construct a weighting function that places relatively more weight on sample observations from neighboring observations in the spatial data sample.

A host of alternative approach have been suggested for constructing the weight function. One approach suggested by Brunsdon et al (1996) is:

$$W_i = \sqrt{\exp(-d_i/\theta)} \tag{1}$$

The parameter θ is a decay or "bandwidth" parameter. Changing the bandwidth results in a different exponential decay profile, which in turn produces estimates that vary more or less rapidly over space. Another weighting scheme is the tri-cube function proposed by McMillen and McDonald (1998):

$$W_i = (1 - (d_i/q_i)^3)^3 \quad \mathrm{I}(d_i < q_i) \tag{2}$$

Where q_i represents the distance of the qth nearest neighbor to observation i and I() is an indicator function that equals one when the condition is true and zero otherwise. Still another approach is to rely on a Gaussian function ϕ:

$$W_i = \phi(d_i/\sigma\theta) \tag{3}$$

Where ϕ denotes the standard normal density and σ represents the standard deviation of the distance vector d_i.

The notation used here may be confusing since we usually rely on subscripted variables to denote scalar elements of a vector. Here, the subscripted variable d_i represents a vector of distances between observation i and all other sample data observations.

A single value of the bandwidth parameter θ is determined using a cross-validation procedure often used in locally linear regression methods. A score function taking the form:

$$\sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(\theta)]^2 \tag{4}$$

is minimized with respect to θ, where $\hat{y}_{\neq i}(\theta)$ denotes the fitted value of y_i with the observations for point i omitted from the calibration process. Note that for the case of the tri-cube weighting function, we would compute an integer q (the number of nearest neighbors) using cross-validation. We focus on the exponential and Gaussian weighting methods for simplicity, ignoring the tri-cube weights.

The non-parametric GWR model relies on a sequence of locally linear regressions to produce estimates for every point in space using a sub-sample of data information from nearby observations. Let y denote an nx1 vector of dependent variable observations collected at n points in space, X an nxk matrix of explanatory variables, and ε an nx1 vector of normally distributed, constant variance disturbances. Letting W_i represent an nxn diagonal matrix containing the vector d_i of distance-based weights for observation i that reflect the distance between observation i and all other observations, we can write the GWR model as:

$$W_i y = W_i X \beta_i + \varepsilon_i \tag{5}$$

The subscript i on β_i indicates that this kx1 parameter vector is associated with observation i. The GWR model produces n such vectors of parameter estimates, one for each observation. These estimates are produced using:

$$\hat{\beta}_i = (X' W_i^2 X)^{-1}(X' W_i^2 y) \tag{6}$$

The GWR estimates for β_i are conditional on the parameter θ we select. That is, changing θ will produce a different set of GWR estimates. Our Bayesian approach relies on the same cross-validation estimate of θ, but

adjusts the weights for outliers or aberrant observations. An area for future work would be devising a method to determine the bandwidth as part of the estimation problem, resulting in a posterior distribution that could be used to draw inferences regarding how sensitive the GWR estimates are to alternative values of this parameter. Posterior Bayesian estimates from this type of model would not be conditional on the value of the bandwidth, as this parameter would be "integrated out" during estimation.

One problem with GWR estimates is that valid inferences cannot be drawn for the regression parameters using traditional least squares approaches. To see this, consider that locally linear estimates use the same sample data observations (with different weights) to produce a sequence of estimates for all points in space. Given the conditional nature of the GWR on the bandwidth estimate and the lack of independence between estimates for each location, regression-based measures of dispersion for the estimates are incorrect.

Another problem is that the presence of aberrant observations due to spatial enclave effects or shifts in regime can exert undue influence on locally linear estimates. Consider that all nearby observations in a sub-sequence of the series of locally linear estimates may be "contaminated" by an outlier at a single point in space. The Bayesian approach introduced here solves this problem using robust estimates that are insensitive to aberrant observations. These observations are automatically detected and downweighted to lessen their influence on the estimates.

A third problem is that the locally linear estimates based on a distance weighted sub-sample of observations may suffer from "weak data" problems. The effective number of observations used to produce estimates for some points in space may be very small. This problem can be solved with the Bayesian approach by incorporating subjective prior information. We introduce some explicit parameter smoothing relationships in the Bayesian model that can be used to impose restrictions on the spatial nature of parameter variation. Stochastic restrictions based on subjective prior information represent a traditional Bayesian approach for overcoming "weak data" problems.

The Bayesian formulation can be implemented with or without the relationship for smoothing parameters over space, and we illustrate both uses in different applied settings. The Bayesian model subsumes the GWR method as part of a much broader class of spatial econometric models. For example, the Bayesian GWR can be implemented with a variety of parameter smoothing relationships. One relationship results in a locally linear variant of the spatial expansion method introduced by Casetti (1972,1992). Another

parameter smoothing relation is based on a monocentric city model where parameters vary systematically with distance from the center of the city, and still others are based on distance decay or contiguity relationships.

Section 2 sets forth the GWR and Bayesian GWR (BGWR) methods. Section 3 discusses the Markov Chain, Monte Carlo estimation method used to implement the BGWR, and Section 4 provides three examples that compare the GWR and BGWR methods.

2 The GWR and Bayesian GWR models

The Bayesian approach, which we label BGWR is best described using matrix expressions shown in (7) and (8). First, note that (7) is the same as the GWR relationship, but the addition of (8) provides an explicit statement of the parameter smoothing that takes place across space. Parameter smoothing in (8) relies on a locally linear combination of neighboring areas, where neighbors are defined in terms of the GWR distance weighting function that decays over space. Other parameter smoothing relationships will be introduced later.

$$W_i y = W_i X \beta_i + \varepsilon_i \tag{7}$$

$$\beta_i = \begin{pmatrix} w_{i1} \otimes I_k & \ldots & w_{in} \otimes I_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + u_i \tag{8}$$

The terms w_{ij} in (8) represent normalized distance-based weights so the row-vector $(w_{i1}, \ldots, w_{in})$ sums to unity, and we set $w_{ii} = 0$. That is, $w_{ij} = \exp(-d_{ij}/\theta)/\sum_{j=1}^{n} \exp(-d_{ij}\theta)$.

To complete our model specification, we add distributions for the terms ε_i and u_i:

$$\varepsilon_i \sim N[0, \sigma^2 V_i], \qquad V_i = \mathrm{diag}(v_1, v_2, \ldots, v_n) \tag{9}$$

$$u_i \sim N[0, \sigma^2 \delta^2 (X' W_i^2 X)^{-1})] \tag{10}$$

The $V_i = \mathrm{diag}(v_1, v_2, \ldots, v_n)$, represent a set of n variance scaling parameters (to be estimated) that allow for non-constant variance as we move across space. Of course, the idea of estimating n terms $v_j, j = 1, \ldots, n$ at each observation i for a total of n^2 parameters (and nk regression parameters β_i) with only n sample data observations may seem truly problematical! The way around this is to assign a prior distribution for the n^2 terms

$V_i, i = 1, \ldots, n$ that depends on a single *hyperparameter*. The V_i parameters are assumed to be i.i.d. $\chi^2(r)$ distributed, where r is a hyperparameter that controls the amount of dispersion in the V_i estimates across observations. This allows us to introduce a single hyperparameter r to the estimation problem and receive in return n^2 parameter estimates.

This type of prior has been used by Lindley (1971) for cell variances in an analysis of variance problem, Geweke (1993) in modeling heteroscedasticity and outliers and LeSage (1997) in a spatial autoregressive modeling context. The specifics regarding the prior assigned to the V_i terms can be motivated by considering that the mean of prior equals unity, and the prior variance is $2/r$. This implies that as r becomes very large, the prior imposes homoscedasticity on the BGWR model and the disturbance variance becomes $\sigma^2 I_n$ for all observations i.

The distribution for the stochastic parameter u_i in the parameter smoothing relationship is normal with mean zero and a variance based on Zellner's (1971) $g-$prior. This prior variance is proportional to the parameter variance-covariance matrix, $\sigma^2(X'W_i^2X)^{-1}$ with δ^2 acting as the scale factor. The use of this prior specification allows individual parameters β_i to vary by different amounts depending on their magnitude.

The parameter δ^2 acts as a scale factor to impose tight or loose adherence to the parameter smoothing specification. Consider a case where δ was very small, then the smoothing restriction would force β_i to look like a distance-weighted linear combination of other β_i from neighboring observations. On the other hand, as $\delta \rightarrow \infty$ (and $V_i = I_n$) we produce the GWR estimates. To see this, we rewrite the BGWR model in a more compact form:

$$\begin{aligned} \tilde{y}_i &= \tilde{X}_i\beta_i + \varepsilon_i \\ \beta_i &= J_i\gamma + u_i \end{aligned} \tag{11}$$

Where the definitions of the matrix expressions are:

$$\begin{aligned} \tilde{y}_i &= W_i y \\ \tilde{X}_i &= W_i X \\ J_i &= \begin{pmatrix} w_{i1} \otimes I_k & \ldots & w_{in} \otimes I_k \end{pmatrix} \\ \gamma &= \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \end{aligned}$$

As indicated earlier, the notation is somewhat confusing in that $\tilde{y}_i$ denotes an $n-$vector, not a scalar magnitude. Similarly, ε_i is an $n-$vector and $\tilde{X}_i$ is an n by k matrix. Note that (11) can be written in the form of a Theil-Goldberger (1961) estimation problem as shown in (12).

$$\begin{pmatrix} \tilde{y}_i \\ J_i\gamma \end{pmatrix} = \begin{pmatrix} \tilde{X}_i \\ -I_k \end{pmatrix} \beta_i + \begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \tag{12}$$

Assuming $V_i = I_n$, the estimates β_i take the form:

$$\begin{aligned} \hat{\beta}_i &= R(\tilde{X}_i'\tilde{y}_i + \tilde{X}_i'\tilde{X}_i J_i\gamma/\delta^2) \\ R &= (\tilde{X}_i'\tilde{X}_i + \tilde{X}_i'\tilde{X}_i/\delta^2)^{-1} \end{aligned}$$

As δ approaches ∞, the terms associated with the Theil-Goldberger "stochastic restriction", $\tilde{X}_i'\tilde{X}_i J_i\gamma/\delta^2$ and $\tilde{X}_i'\tilde{X}_i/\delta^2$ become zero, and we have the GWR estimates:

$$\hat{\beta}_i = (\tilde{X}_i'\tilde{X}_i)^{-1}(\tilde{X}_i'\tilde{y}_i) \tag{13}$$

In practice, we can use a diffuse prior for δ which allows the amount of parameter smoothing to be estimated from sample data information, rather than by subjective prior information. Details concerning estimation of the parameters in the BGWR model are taken up in the next section. Before turning to these issues, we consider some alternative spatial parameter smoothing relationships that might be used in lieu of (8) in the BGWR model.

One alternative smoothing specification would be the "monocentric city smoothing" set forth in (14). This relation assumes that the data observations have been ordered by distance from the center of the spatial sample.

$$\begin{aligned} \beta_i &= \beta_{i-1} + u_i \\ u_i &\sim N[0, \sigma^2\delta^2(X'W_i^2X)^{-1}] \end{aligned} \tag{14}$$

Given that the observations are ordered by distance from the center, the smoothing relation indicates that β_i should be similar to the coefficient β_{i-1} from a neighboring concentric ring. Note that we rely on the same GWR distance-weighted data sub-samples, created by transforming the data using: $W_i y, W_i X$. This means that the estimates still have a "locally linear" interpretation as in the GWR. We rely on the same distributional assumption

for the term u_i from the BGWR which allows us to estimate the parameters from this model by making minor changes to the approach used for the BGWR based on the smoothing relation in (8).

Another alternative is a "spatial expansion smoothing" based on the ideas introduced by Casetti (1972). This is shown in (15), where Z_{xi}, Z_{yi} denote latitude-longitude coordinates associated with observation i.

$$\begin{aligned} \beta_i &= \begin{pmatrix} Z_{xi} \otimes I_k & Z_{yi} \otimes I_k \end{pmatrix} \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix} + u_i \\ u_i &\sim N[0, \sigma^2 \delta^2 (X' W_i^2 X)^{-1})] \end{aligned} \tag{15}$$

This parameter smoothing relation creates a locally linear combination based on the latitude-longitude coordinates of each observation. As in the case of the monocentric city specification, we retain the same assumptions regarding the stochastic term u_i, making this model simple to estimate with only minor changes to the basic BGWR methodology.

Finally, we could adopt a "contiguity smoothing" relationship based on a first-order spatial contiguity matrix as shown in (16). The terms c_{ij} represent the ith row of a row-standardized first-order contiguity matrix. This creates a parameter smoothing relationship that averages over the parameters from observations that neighbor observation i.

$$\begin{aligned} \beta_i &= \begin{pmatrix} c_{i1} \otimes I_k & \dots & c_{in} \otimes I_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + u_i \\ u_i &\sim N[0, \sigma^2 \delta^2 (X' W_i^2 X)^{-1})] \end{aligned} \tag{16}$$

These approaches to specifying a geographically weighted regression model suggest that researchers need to think about which type of spatial parameter smoothing relationship is most appropriate for their application. Additionally, where the nature of the problem does not clearly favor one approach over another, statistical tests of alternative models based on different smoothing relations might be carried out. Posterior probabilities can be constructed that will shed light on which smoothing relationship is most consistent with the sample data. This subject is taken up in Section 3.1 and illustrations are provided in Section 4.

3 Estimation of the BGWR model

A recent methodology known as Markov Chain Monte Carlo is based on the idea that rather than compute a probability density, say $p(\theta|y)$, we would be just as happy to have a large random sample from $p(\theta|y)$ as to know the precise form of the density. Intuitively, if the sample were large enough, we could approximate the form of the probability density using kernel density estimators or histograms. In addition, we could compute accurate measures of central tendency and dispersion for the density, using the mean and standard deviation of the large sample. This insight leads to the question of how to efficiently simulate a large number of random samples from $p(\theta|y)$.

Metropolis, et al. (1953) demonstrated that one could construct a Markov chain stochastic process for $(\theta_t, t \geq 0)$ that unfolds over time such that: 1) it has the same state space (set of possible values) as θ, 2) it is easy to simulate, and 3) the equilibrium or stationary distribution which we use to draw samples is $p(\theta|y)$ after the Markov chain has been run for a long enough time. Given this result, we can construct and run a Markov chain for a very large number of iterations to produce a sample of $(\theta_t, t = 1, \ldots)$ from the posterior distribution and use simple descriptive statistics to examine any features of the posterior in which we are interested.

This approach, known as Markov Chain Monte Carlo, (MCMC) or Gibbs sampling has greatly reduced the computational problems that previously plagued application of the Bayesian methodology. Gelfand and Smith (1990), as well as a host of others, have popularized this methodology by demonstrating its use in a wide variety of statistical applications where intractable posterior distributions previously hindered Bayesian analysis. A simple introduction to the method can be found in Casella and George (1992) and an expository article dealing specifically with the normal linear model is Gelfand, Hills, Racine-Poon and Smith (1990). Two recent books that deal in detail with all facets of these methods are: Gelman, Carlin, Stern and Rubin (1995) and Gilks, Richardson and Spiegelhalter (1996).

We rely on Gibbs sampling to produce estimates for the BGWR model, which represent the multivariate posterior probability density for all of the parameters in our model. This approach is particularly attractive in this application because the conditional densities are simple and easy to obtain. LeSage (1997) demonstrates this approach for Bayesian estimation of spatial autoregressive models, which represents a more complicated case.

To implement the Gibbs sampler we need to derive and draw samples from the conditional posterior distributions for each group of parameters,

β_i, σ, δ, and V_i in the model. Let $P(\beta_i|\sigma, \delta, V_i, \gamma)$ denote the conditional density of β_i, where γ represents the values of other β_j for observations $j \neq i$. Using similar notation for the the other conditional densities, the Gibbs sampling process can be viewed as follows:

1. start with arbitrary values for the parameters $\beta_i^0, \sigma_i^0, \delta^0, V_i^0, \gamma^0$
2. for each observation $i = 1, \ldots, n$,
   (a) sample a value, β_i^1 from $P(\beta_i|\delta^0, \sigma_i^0, V_i^0, \gamma^0)$
   (b) sample a value, σ_i^1 from $P(\sigma_i|\delta^0, V_i^0, \beta_i^1, \gamma^0)$
   (c) sample a value, V_i^1 from $P(V_i|\delta^0, \beta_i^1, \sigma_i^1, \gamma^0)$
3. use the sampled values $\beta_i^1, i = 1, \ldots, n$ from each of the n draws above to update γ^0 to γ^1.
4. sample a value, δ^1 from $P(\delta|\sigma_i^1, \beta_i^1 V_i^1, \gamma^1)$
5. go to step 1 using $\beta_i^1, \sigma_i^1, \delta^1, V_i^1, \gamma^1$ in place of the arbitrary starting values.

Steps 2 to 4 outlined above represents a single pass through the sampler, and we make a large number of passes to collect a sample of parameter values from which we construct our posterior distributions. Note that this is computationally intensive as it requires a loop over all observations for each draw. In one of our examples we implement a simpler version of the Gibbs sampler that can be used to produce robust estimates when no parameter smoothing relationship is in the model. This sampling routine involves a single loop over each of the n observations that carries out all draws, as shown below.

1. start with arbitrary values for the parameters $\beta_i^0, \sigma_i^0, V_i^0$
2. for each observation $i = 1, \ldots, n$, sample all draws using a sequence over:
3. Step 1: sample a value, β_i^1 from $P(\beta_i|\sigma_i^0, V_i^0)$
4. Step 2: sample a value, σ^1 from $P(\sigma_i|V_i^0, \beta_i^1)$
5. Step 3: sample a value, V_i^1 from $P(V_i|\beta_i^1, \sigma_i^1)$

6. go to Step 1 using $\beta_i^1, \sigma_i^1, V_i^1$ in place of the arbitrary starting values. Continue returning to Step 1 until all draws have been obtained.

7. Move to observation $i = i + 1$ and obtain all draws for this next observation.

8. When we reach observation n, we have sampled all draws for each observation.

This approach samples all draws for each observation, requiring a single pass through the N observation sample. The computational burden associated with the first sampler arises from the need to update the parameters in γ for all observations before moving to the next draw. This is because these values are used in the distance and contiguity smoothing relationships.

The second sampler takes around 10 seconds to produce 1,000 draws for each observation, irrespective of the sample size. Sample size is irrelevant because we exclude distance weighted observations that have negligible weights. This reduces the size of the matrices that need be computed during sampling to a fairly constant size that does not depend on the number of observations. In contrast, the first sampler takes around 2 seconds per draw for even moderate sample sizes of 100 observations, and computational time increases dramatically with the number of observations.

For the case of the monocentric city prior we could rely on the GWR estimate for the first observation and proceed to carry out draws for the remaining observations using the second sampler presented above. The draw for observation 2 would rely on the posterior mean computed from the draws for observation 1. Note that we need the posterior from observation 1 to define the parameter smoothing prior for observation 2. Assuming the observations are ordered by distance from a central observation, this would achieve our goal of stochastically restricting observations from nearby concentric rings to be similar. Observation 2 would be similar to 1, 3 would be similar to 2, and so on.

Another computationally efficient way to implement these models with a parameter smoothing relationship would be to use the GWR estimates as elements in γ. This would allow us to use the second sampler that makes multiple draws for each observation, requiring only one pass over the observations. A drawback to this approach is that the parameter smoothing relationship doesn't evolve as part of the estimation process. It is stochastically restricted to the fixed GWR estimates.

We rely on the compact statement of the BGWR model in (11) to facilitate presentation of the conditional distributions that we rely on during

the sampling. The conditional posterior distribution of β_i given σ_i, δ, γ and V_i is a multivariate normal shown in (17).

$$p(\beta_i|\ldots) \propto N(\hat{\beta}_i, \sigma_i^2 R) \tag{17}$$

Where:

$$\begin{aligned} \hat{\beta}_i &= R(\tilde{X}_i' V_i^{-1} \tilde{y}_i + \tilde{X}_i' \tilde{X}_i J_i \gamma/\delta^2) \\ R &= (\tilde{X}_i' V_i^{-1} \tilde{X}_i + \tilde{X}_i' \tilde{X}_i/\delta^2)^{-1} \end{aligned} \tag{18}$$

This result follows from the assumed variance-covariance structures for ε_i, u_i and the Theil-Goldberger (1961) representation shown in (12). The conditional posterior distribution for σ is a $\chi^2(m)$ distribution shown in (19), where m denotes the number of observations with non-negligible weights.

$$\begin{aligned} p(\sigma_i|\ldots) &\propto \sigma_i^{-(m+1)} \exp\{-\frac{1}{2\sigma_i^2}(\varepsilon_i' V_i^{-1} \varepsilon_i)\} \\ \varepsilon_i &= \tilde{y}_i - \tilde{X}_i \beta_i \end{aligned} \tag{19}$$

The conditional posterior distribution for V_i is shown in (20), which indicates that we draw an m-vector based on a $\chi^2(r+1)$ distribution.

$$p\{[(e_i^2/\sigma_i^2) + r]/V_i \mid \ldots\} \propto \chi^2(r+1) \tag{20}$$

To see the role of the parameter v_{ij}, consider two cases. First, suppose (e_j^2/σ_i^2) is small (say zero), because the GWR distance-based weights work well to relate y and X for observation j. In this case, observation j is not an outlier. Assume that we use a small value of the hyperparameter r, say $r = 5$, which means our prior belief is that heterogeneity exits. The conditional posterior will have a mean and mode of:

$$\begin{aligned} \text{mean}(v_{ij}) &= (\sigma_i^{-2} e_j^2 + r)/(r+1) = r/(r+1) = (5/6) \\ \text{mode}(v_{ij}) &= (\sigma_i^{-2} e_j^2 + r)/(r-1) = r/(r-1) = (5/4) \end{aligned} \tag{21}$$

Where the results in (21) follow from the fact that the mean of the prior distribution for V_{ij} is $r/(r-2)$ and the mode of the prior equals $r/(r+2)$.

In the case shown in (21), the impact of $v_{ij} \approx 1$ in the model is negligible, and the typical distance-based weighting scheme would dominate.

For the case of exponential weights, a weight, $w_{ij} = \exp(-d_i)/\theta v_{ij}$ would be accorded to observation j. Note that a prior belief in homogeneity that assigns a large value of $r = 20$, would produce a similar weighting outcome. The conditional posterior mean of $r/(r+1) = 20/21$, is approximately unity, as is the mode of $(r+1)/r = 20/19$.

Second, consider the case where (e_j^2/σ_i^2) is large (say 20), because the GWR distance-based weights do not work well to relate y and X for observation j. Here, we have the case of an outlier for observation j. Using the same small value of the hyperparameter $r = 5$, the conditional posterior will have a mean and mode of:

$$\begin{aligned} \text{mean}(v_{ij}) &= (20+r)/(r+1) = (25/6) \\ \text{mode}(v_{ij}) &= (20+r)/(r-1) = (25/4) \end{aligned} \tag{22}$$

For this aberrant observation case, the role of $v_{ij} \approx 5$ will be to downweight the distance associated with this observation. The distance-based weight, $w_{ij} = \exp(-d_i)/\theta v_{ij}$ would be deflated by a factor of approximately 5 for this aberrant observation. It is important to note that, a prior belief of homogeneity (expressed by a large value of $r = 20$) in this case would produce a conditional posterior mean of $(20+r)/(r+1) = (40/21)$. Downweighting of the distance-based weights would be only by a factor of 2, rather than 5 found for the smaller value of r.

It should be clear that as r becomes very large, say 50 or 100, the posterior mean and mode will be close to unity irrespective of the fit measured by e_j^2/σ_i^2 This replicates the distance-based weighting scheme used in the non-Bayesian GWR model.

A graphical illustration of how this works in practice can be seen in Figure 1. The figure depicts the adjusted distance-based weights, $W_i V_i^{-1}$ alongside the GWR weights W_i for observations 31 to 36 in the Anselin (1988) Columbus neighborhood crime data set. In Section 4.1 we motivate that observation #34 represents an outlier.

Beginning with observation 31, the aberrant observation #34 is downweighted when estimates are produced for observations 31 to 36 (excluding observation #34 itself). A symbol 'o' has been placed on the BGWR weight in the figure to help distinguish observation 34. This downweighting of the distance-based weight for observation #34 occurs during estimation of β_i for observations 31 to 36, all of which are near #34 in terms of the GWR distance measure. It will be seen that this alternative weighting produces a divergence in the BGWR estimates and those from GWR for observations neighboring on #34.

Finally, the conditional distribution for δ is a $\chi^2(nk)$ distribution based on (23).

$$p(\delta|\ldots) \propto \delta^{-nk}\exp\{-\sum_{i=1}^{n}(\beta_i - J_i\gamma)'(\tilde{X}_i'\tilde{X}_i)^{-1}(\beta_i - J_i\gamma)/2\sigma_i^2\delta^2\} \qquad (23)$$

Now consider the modifications needed to the conditional distributions to implement the alternative spatial smoothing relationships set forth in Section 3. Because the same assumptions were used for the disturbances ε_i and u_i, we need only alter the conditional distributions for β_i and δ. First, consider the case of the monocentric city smoothing relationship. The conditional distribution for β_i is multivariate normal with mean $\hat{\beta}_i$ and variance-covariance $\sigma^2 R$ as shown in (24).

$$\begin{aligned} \hat{\beta}_i &= R(\tilde{X}_i'V_i^{-1}\tilde{y}_i + \tilde{X}_i'\tilde{X}_i\beta_{i-1}/\delta^2) \\ R &= (\tilde{X}_i'V_i^{-1}\tilde{X}_i + \tilde{X}_i'\tilde{X}_i/\delta^2)^{-1} \end{aligned} \qquad (24)$$

The conditional distribution for δ is a $\chi^2(nk)$ based on the expression in (25).

$$p(\delta|\ldots) \propto \delta^{-nk}\exp\{-\sum_{i=1}^{n}(\beta_i - \beta_{i-1})'(\tilde{X}'\tilde{X})^{-1}(\beta_i - \beta_{i-1})/\sigma_i^2\delta^2\} \qquad (25)$$

For the case of the spatial expansion and contiguity smoothing relationships, we can maintain the conditional expressions for β_i and δ from the case of the basic BGWR, and simply modify the definition of J, to be consistent with these smoothing relations.

3.1 Informative priors

Implementing the BGWR model with very large values for δ will essentially eliminate the parameter smoothing relationship from the model. The BGWR estimates will then collapse to the GWR estimates (in the case of a large value for the hyperparameter r that leads to $V_i = I_n$), and this represents a very computationally intensive way to obtain GWR estimates. If there is a desire to obtain robust BGWR estimates without imposing a parameter smoothing relationship in the model, the second sampling scheme presented in Section 3 can do this in a more computationally efficient manner.

The parameter smoothing relationships are useful in cases where the sample data is weak or objective prior information suggests spatial parameter smoothing that follows a particular specification. Alternatives exist for placing an informative prior on the parameter δ. One is to rely on a Gamma(a, b) prior distribution which has a mean of a/b and variance of a/b^2. Given this prior, we could eliminate the conditional density for δ and replace it with a random draw from the Gamma(a, b) distribution during sampling.

Another approach to the parameter δ is to assign an improper prior value using say, $\delta = 1$. Setting δ may be problematical because the scale is unknown and depends on the inherent variability in the GWR estimates. Consider that $\delta = 1$ will assign a prior variance for the parameters in the smoothing relationship based on the variance-covariance matrix of the GWR estimates. This may represent a tight or loose imposition of the parameter smoothing relationship, depending on the amount of variability in the GWR estimates. If the estimates vary widely over space, this choice of δ may not produce estimates that conform very tightly to the parameter smoothing relationship. In general we can say that smaller values of δ reflect a tighter imposition of the spatial parameter smoothing relationship and larger values reflect a looser imposition, but this is unhelpful in particular modeling situations.

A practical approach to setting values for δ would be to generate an estimate based on a diffuse prior for δ and examine the posterior mean for this parameter. Setting values of δ smaller than the posterior mean from the diffuse implementation should produce a prior that imposes the parameter smoothing relationship more tightly. One might use magnitudes for δ that scale down the diffuse δ estimate by 0.5, 0.25 and 0.1 to examine the impact of the parameter smoothing relationship on the BGWR estimates.

Posterior probabilities can be used as a guide for comparing alternative parameter smoothing relationships and various values for δ. These can be calculated using the log posterior for every observation divided by the sum of the log posterior over all models at each observation. Expression (26) shows the log posterior for a single observation of our BGWR model. Posterior probabilities based on these quantities provide an indication of which parameter smoothing relationship fits the sample data best as we range over observations.

$$\text{Log}P_i = \sum_{j=1}^{n} W_{ij}\{\log\phi([y_j - X_i\beta_i]/\sigma_i v_{ij}) - \log\sigma_i v_{ij}\} \tag{26}$$

Keep in mind that these posterior probabilities reflect a measure of fit to the sample data, as is clear from (26). In applications where robust estimates are desired, it is not clear that choice of models should be made using measures of fit. Robust estimates require a trade-off between fit and insensitivity to aberrant observations.

A similar Gamma prior for the hyperparameter r can be used, where values $a = 8, b = 2$ would indicate small values of r around 4. This should provide fairly robust estimates if there is spatial heterogeneity. In the absence of heterogeneity, the resulting V_i estimates will be near unity so the BGWR distance weights will be similar to those from GWR, even with a small value of r. We can also set an improper prior value for this hyperparameter, say $r = 4$ Additionally, a $\chi^2(c, d)$ natural conjugate prior for the parameter σ could be used in place of the diffuse prior set forth here. This would affect the conditional distribution used during Gibbs sampling in only a minor way.

Some other alternatives offer additional flexibility when implementing the BGWR model. For example, one can restrict specific parameters to exhibit no variation over the spatial sample observations. This might be useful if we wish to restrict the constant term over space. Or, it may be that the constant term is the only parameter that we allow to vary over space.

These alternatives can be implemented by adjusting the prior variances in the parameter smoothing relationship:

$$var - cov(\beta_i) = \sigma^2 \delta^2 (\tilde{X}_i' \tilde{X}_i)^{-1} \tag{27}$$

For example, assuming the constant term is in the first column of the matrix $\tilde{X}_i$, setting the first row and column elements of $(\tilde{X}_i' \tilde{X}_i)^{-1}$ to zero would restrict the intercept term to remain constant over all observations.

4 Examples

Section 4.1 provides two comparisons of the GWR and BGWR estimates without reliance on a parameter smoothing relationship. These illustrations demonstrate the sensitivity of GWR estimates to aberrant observations and show how outliers are downweighted by the V_i terms in the BGWR model.

An illustration that compares the GWR to the BGWR based on monocentric, distance and contiguity smoothing relations is provided in Section 4.2, along with the posterior probabilities for these alternative spatial smoothing approaches.

4.1 A comparison of GWR and BGWR

As an initial illustration of the problems created by outliers in GWR estimation, a generated data set containing 100 observations was used. A regression variable y was generated using coefficients that vary over a regular grid according to the quadrant in which the observation falls. Coefficients of 1 and -1 were used for two explanatory variables. A switch from 1 to -1 in the coefficients occurs at observation 50, which is the type of spatial variation in relationships that the GWR model was devised to detect.

After producing GWR estimates based on this data set, we create a single outlier at observation 60 by multiplying the explanatory variables by 10. Another set of GWR estimates along with BGWR model estimates were produced using this outlier contaminated data set. If the BGWR model is producing robust estimates, we would expect to see estimates that are similar to those from the GWR model based on the data set with no outlier.

The results from this experiment are shown in Figure 2 where the adverse impact of the single outlier at observation 60 is clear. GWR estimates from the data set with no outlier captured the shift in relationship at observation 50 with a great deal of precision, as did the robust BGWR estimates based on the data set containing the outlier. In contrast, the GWR estimates based on the data set with a single outlier do not capture the abrupt shift in the relationship over space. It would be difficult to infer the abrupt shift in regime at the appropriate point in space based on these GWR estimates.

In addition to adversely impacting the coefficient trajectories over space, the single outlier also affects the $t-$statistics that would be used to draw inferences regarding shifts in regime as we move over space. Figure 3 shows $t-$statistics from the GWR model based on both data sets as well as the BGWR $t-$statistics for the data set containing the outlier. Here again, we see that the BGWR estimates are close to those from the GWR model based on no outliers. A closer examination of the $t-$statistic from the GWR model in the case of the outlier data set indicated that the estimate of the noise variance, σ^2 which enter into calculation of the $t-$statistics was the source of the problem.

As an applied illustration of the Bayesian GWR model we used a spatial data set from Anselin (1988) on neighborhood crime in Columbus, Ohio. A model was estimated using neighborhood crime incidents as the dependent variable, household income and house values along with a constant term as explanatory variables, that is:

$$\text{Crime}_i = \beta_{1i} + \beta_{2i}(\text{Household Income})_i + \beta_{3i}(\text{House Value})_i + \varepsilon_i \quad (28)$$

Estimates from a GWR model are compared to those from a BGWR model based on $r = 4$ representing a heteroscedastic prior, and a Gaussian weighting approach. For this sample of 49 observations and 3 explanatory variables, it took around 250 seconds to produce 1,250 draws, and 120 seconds for 550 draws on an Apple 266 Mhz. G3 Powerbook. The posterior means of the parameter estimates were virtually identical for the sample of 550 and 1,250 draws, suggesting no problems with convergence of the Gibbs sampler.

Figure 4 shows the comparison of GWR and BGWR estimates from the heteroscedastic version of the model. We see definite evidence of a departure between the GWR and BGWR estimates. The large V_i estimates presented in Figure 5 point to non-constant variance as we move over the spatial sample.

An interesting question is — are these differences significant in a statistical sense? We can answer this question using the 1,000 draws produced by the Gibbs sampler to compute a two standard deviation band around the BGWR estimates. If the GWR estimates fall within this confidence interval, we would conclude the estimates are not significantly different. Figure 6 shows the GWR estimates and the confidence bands for the BGWR estimates. The actual BGWR estimates were omitted from the graph for clarity. We see that the GWR estimates are near the two standard deviation confidence intervals for sample observations in the range from 20 to 44, which implies we might draw different inferences from the GWR and BGWR estimates.

Another way to visualize the impact of non-constant variance over space is to examine a map of the absolute differences between the GWR and BGWR estimates. Neighborhoods surrounding areas with large V_i values should exhibit differences in the GWR and BGWR estimates. A change in the noise variance for a single observation tends to produce different trajectories for the estimates in all surrounding neighborhoods because the GWR relies on a sequence of sub-samples of the data.

Figures 7 and 8 show maps of the absolute differences between the GWR and BGWR coefficient estimates for household income and housing values in the 49 Columbus neighborhoods. Darker areas reflect larger differences between the GWR and BGWR estimates.

In the case of the income coefficient shown in Figure 7, we see a pattern where the absolute differences between the GWR and BGWR estimates are largest around neighborhoods bordering on observations 2 in the west, 16 and 27 in the north, 20 and 24 near the center and observation 34 in the south. Note that large V_i estimates for these observations shown in

Figure 5 produced large differences between GWR and BGWR estimates for surrounding neighborhoods, not just the observations containing large V_i values. A similar pattern exists in Figure 8 showing absolute differences between the GWR and BGWR estimates for housing values.

The mean of the V_i estimates averaged over all observations in the spatial sample can be used as a diagnostic measure to detect aberrant observations. These V_i values reflect observations that consistently produced large residuals during estimation of each β_i parameter. The average V_i draws in Figure 5 indicate that observations 2, 16 and 27, 20 and 24 as well as observation 34 were consistently downweighted during estimation of the β_i for all 49 observations. This is desirable if we wish to keep these aberrant observations from contaminating the estimates produced for neighbors.

Ultimately, the role of the parameters V_i in the BGWR model and the prior assigned to these parameters reflect our prior knowledge that distance alone may not be reliable as the basis for spatial relationships between variables. If distance-based weights are used in the presence of aberrant observations, inferences will be contaminated for whole neighborhoods and regions in our analysis. Incorporating this prior knowledge turns out to be relatively simple in the Bayesian framework, and it appears to effectively robustify estimates against the presence of spatial outliers.

4.2 Alternative spatial smoothing relations

To illustrate alternative parameter smoothing relationships we use a data set consisting of employment, payroll earnings and the number of establishments in all fifty zip (postal) codes from Cuyahoga county Ohio during the first quarter of 1989. The data set was created by aggregating establishment level data used by the State of Ohio for unemployment insurance purposes. It represents employment for workers covered by the state unemployment insurance program. The regression model used was:

$$\ln(E_i/F_i) = \beta_{0i} + \beta_{1i}\ln(P_i/E_i) + \beta_{2i}\ln(F_i) + \varepsilon_i \tag{29}$$

Where E_i is employment in zip code i, P_i represents payroll earnings and F_i denotes the number of establishments. The relationship indicates that employment per firm is a function of earnings per worker and the number of firms in the zip code area. For presentation purposes we sorted the sample of 50 observations by the dependent variable from low to high, so observation #1 represents the zip code district with the smallest level of employment per firm.

Three alternative parameter smoothing relationships were used, the monocentric city prior centered on the central business district, the distance decay prior and the contiguity prior. We would expect the monocentric city prior to work well in this application. An initial set of estimates based on a diffuse prior for δ are discussed below and would typically be generated to calibrate the tightness of alternative settings for the prior on the parameter smoothing relations.

A Gaussian distance weighting method was used, but estimates based on the exponential weighting method were quite similar. All three BGWR models were based on a hyperparameter $r = 4$ reflecting a heteroscedastic prior.

A graph of the three sets of estimates is shown in Figure 9, where it should be kept in mind that the observations are sorted by employment per firm from low to high. This helps when interpreting variation in the estimates over the observations.

The first thing to note is the relatively unstable GWR estimates for the constant term and earnings per worker when compared to the BGWR estimates. Evidence of parameter smoothing is clearly present. Bayesian methods attempt to introduce a small amount of bias in an effort to produce a substantial increase in precision. This seems a reasonable trade-off if it allows clearer inferences. The diffuse prior for the smoothing relationships produced estimates for δ^2 equal to 138 for the monocentric city prior, 142 and 113 for the distance and contiguity priors. These large values indicate that the sample data are inconsistent with these parameter smoothing relationships, so their use would likely introduce some bias in the estimates. From the plot of the coefficients it is clear that no systematic bias is introduced, rather we see evidence of smoothing that impacts only volatile GWR estimates that take rapid jumps from one observation to the next.

Note that the GWR and BGWR estimates for the coefficients on the number of firms are remarkably similar. There are two factors at work to create a divergence between the GWR and BGWR estimates. One is the introduction of v_i parameters to capture non-constant variance over space and the other is the parameter smoothing relationship. The GWR coefficient on the firm variable is apparently insensitive to any non-constant variance in this data set. In addition, the BGWR estimates are not affected by the parameter smoothing relationships we introduced. An explanation for this is that a least-squares estimate for this coefficient produced a $t-$statistic of 1.5, significant at only the 15% level. Since our parameter smoothing prior relies on the variance-covariance matrix from least-squares (adjusted by the distance weights), it is likely that the parameter smoothing relationships

are imposed very loosely for this coefficient. Of course, this will result in estimates equivalent to the GWR estimates.

A final point is that all three parameter smoothing relations produced relatively similar estimates. The monocentric city prior was most divergent with the distance and contiguity priors very similar. We would expect this since the latter priors rely on the entire sample of estimates whereas the monocentric city prior relies only on the estimate from a neighboring observation.

The times required for 550 draws with these models were: 320 seconds for the monocentric city prior, 324 seconds for the distance-based prior, and 331 seconds for the contiguity prior.

Turning attention to the question of which parameter smoothing relation is most consistent with the sample data, a graph of the posterior probabilities for each of the three models is shown in the top panel of Figure 10. It seems quite clear that the monocentric smoothing relation is most consistent with the data as it receives slightly higher posterior probability values for all observations. There is however no dominating evidence in favor of a single model, since the other two models receive substantial posterior probability weight over all observations, summing to over 60 percent.

For purposes of inference, a single set of parameters can be generated using these posterior probabilities to weight the three sets of parameters. This represents a Bayesian solution to the model specification issue (see Leamer, 1983). In this application, the parameters averaged using the posterior probabilities would look very similar to those in Figure 9, since the weights are roughly equal and the coefficients are very similar.

Figure 10 also shows a graph of the estimated v_i parameters from all three versions of the BGWR model. These are nearly identical and point to observations at the beginning and end of the sample as regions of non-constant variance as well as observations around 17, 20, 35, 38 and 44 as perhaps outliers. Because the observations are sorted from small to large, the large v_i estimates at the beginning and end of the sample indicate our model is not working well for these extremes in firm size. It is interesting to note that outlying GWR estimates by comparison with the smoothed BGWR estimates correlate highly with observations where the v_i estimates are large. As we saw in the generated data example, the GWR model tends to "chase" after the outliers, and we see evidence of this here as well.

A final question is — how sensitive are these inferences regarding the three models to the diffuse prior used for the parameter δ? To test alternative smoothing priors in an attempt to find a single best model we impose the priors in a relatively tight fashion. In the face of a very strict implemen-

tation of the smoothing relationship, the posterior probabilities will tend to concentrate on the model that is most consistent with the data. To illustrate this, we constructed another set of estimates and posterior probabilities based on scaling δ to 0.1 times the estimate of δ from the diffuse prior. This should reflect a fairly tight imposition of the prior for the parameter smoothing relationships.

The posterior probabilities and estimates from these three models were very similar to those from the diffuse prior implementation. This suggests that even with this tighter imposition of the prior, all three parameter smoothing relationships are relatively compatible with the sample data. No smoothing relationship obtains a distinctive advantage over the others.

We need to keep the trade-off between bias and efficiency in mind when implementing tight versions of the parameter smoothing relationships. For this application, the fact that both diffuse and tight implementation of the parameter smoothing relationships produced similar estimates indicates our inferences would be robust with respect to relatively large changes in the smoothing priors.

5 Conclusions

We have demonstrated that GWR models can be subsumed as a special case of a broader set of Bayesian models. This was accomplished by adding a parameter smoothing relationship to the GWR model that stochastically restricts the estimates based on spatial relationships.

In addition to replicating the GWR estimates, the Bayesian model presented here can produce estimates based on parameter smoothing specifications that rely on distance, contiguity relationships, monocentric distance from a central point, or the latitude-longitude locations proposed by Casetti (1972).

The Bayesian GWR model also solves some problems that arise when the GWR model encounters non-constant variance over space or outliers. Given the locally linear nature of the GWR estimates, aberrant observations tend to contaminate entire sub-sequences of the estimates. The BGWR model robustifies against these observations by automatically detecting and downweighting their influence on the estimates. A further advantage of this approach is that a diagnostic plot can be used to identity observations associated with regions of non-constant variance or spatial outliers.

If the goal of locally linear estimation is to make inferences regarding spatial variation in the relationship, contamination from outliers may lead

to an erroneous conclusion that the relationship is changing. In fact the relationship may be stable but subject to the influence of a single outlying observation. In contrast, the BGWR estimates indicate changes in the parameters of the relationship as we move over space that abstract from aberrant observations. From the standpoint of inference, we can be relatively certain that changing BGWR estimates truly reflect a change in the underlying relationship as we move through space. In contrast, the GWR estimates are more difficult to interpret, since changes in the estimates may reflect spatial changes in the relationship, or the presence of an aberrant observation.

A final issue that plagues the GWR is that conventional measures of dispersion may not be valid because the assumption of independence is not realistic given the reuse of sample observations. Bayesian estimates produced using the Gibbs sampler overcome these problems using measures of dispersion based on the posterior distributions derived from the Gibbs sampler that are not affected by a lack of sample independence.

6 References

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

Brunsdon, C., A. S. Fotheringham, and M.E. Charlton. (1996) "Geographically weighted regression: A method for exploring spatial nonstationarity", *Geographical Analysis*, Volume 28, pp. 281-298.

Casella, G. and E.I. George. (1992) "Explaining the Gibbs Sampler", *American Statistician*, Volume 46, pp. 167-174.

Casetti, E. (1972) "Generating Models by the Expansion Method: Applications to Geographic Research", *Geographical Analysis*, December, Volume 4, pp. 81-91.

Casetti, E. (1992) "Bayesian Regression and the Expansion Method", *Geographical Analysis*, January, Volume 24, pp. 58-74.

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

Gelfand, Alan E., Susan E. Hills, Amy Racine-Poon and Adrian F.M. Smith. (1990) "Illustration of Bayesian Inference in Normal Data

Models Using Gibbs Sampling", *Journal of the American Statistical Association*, 85, pp. 972-985.

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

Geweke, John. (1993) "Bayesian Treatment of the Independent Student-t Linear Model", *Journal of Applied Econometrics*, Vol. 8, s19-s40.

Gilks, W.R., S. Richardson and D.J. Spiegelhalter. (1996) *Markov Chain Monte Carlo in Practice*, (London: Chapman & Hall).

Leamer, Edward E. (1983) "Model Choice and Specification Analysis", in *Handbook of Econometrics, Volume 1*, Zvi Griliches and Michael D. Intriligator, eds. (North-Holland: Amsterdam).

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

Lindley, David V. (1971) "The Estimation of many Parameters," in V.P. Godambe and D.A. Sprott (eds.), *Foundations of Statistical Inference.* Toronto: Holt, Reinhart, and Winston.

McMillen, Daniel P. (1996) "One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics*, Vol. 40, pp. 100-124.

McMillen, Daniel P. and John F. McDonald. (1997) "A Nonparametric Analysis of Employment Density in a Polycentric City," *Journal of Regional Science*, Vol. 37, pp. 591-612.

McMillen, Daniel P. and John F. McDonald. (1998) "Locally weighted maximum likelihood estimation: Monte Carlo evidence and an application", Paper presented at the Regional Science Association International meetings, Santa Fe, NM.

Metroplis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. (1953) "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, Vol. 21, pp. 1087-1092.

Theil, Henri and Arthur S. Goldberger. (1961) "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, Vol. 2, pp. 65-78.

Zellner, Arnold. (1971) *An Introduction to Bayesian Inference in Econometrics*. (New York: John Wiley & Sons.)

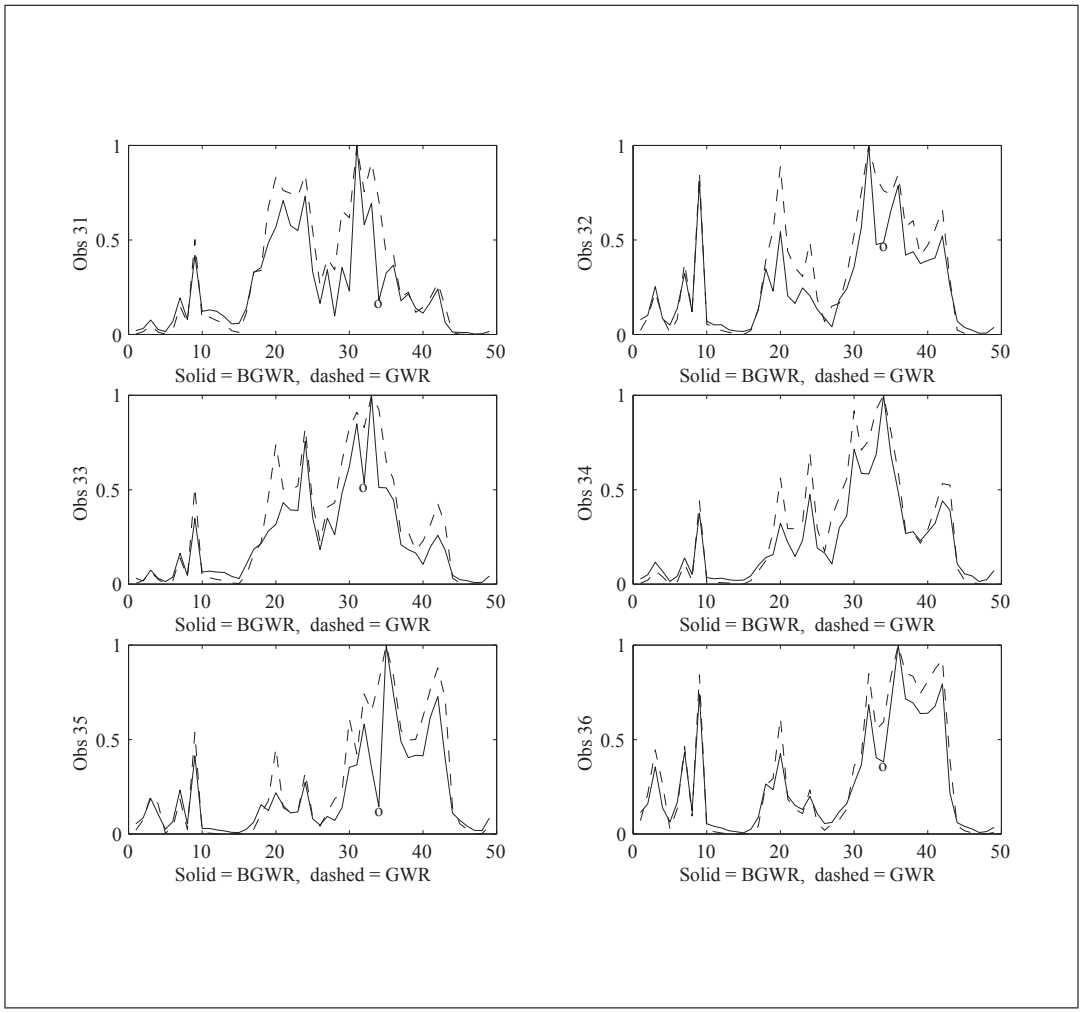

Figure 1: Distance-based weights adjusted by V_i

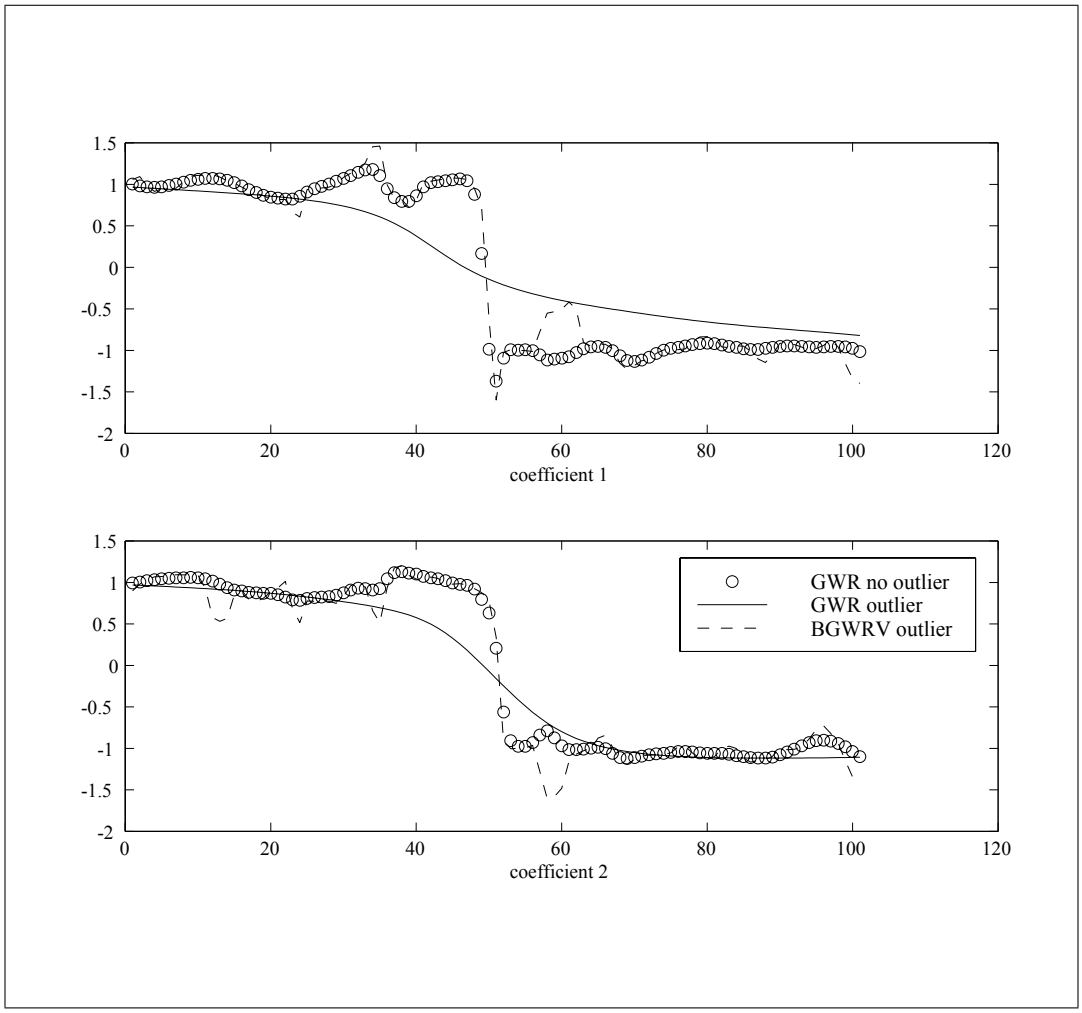

Figure 2: β_i estimates for GWR and BGWRV with an outlier

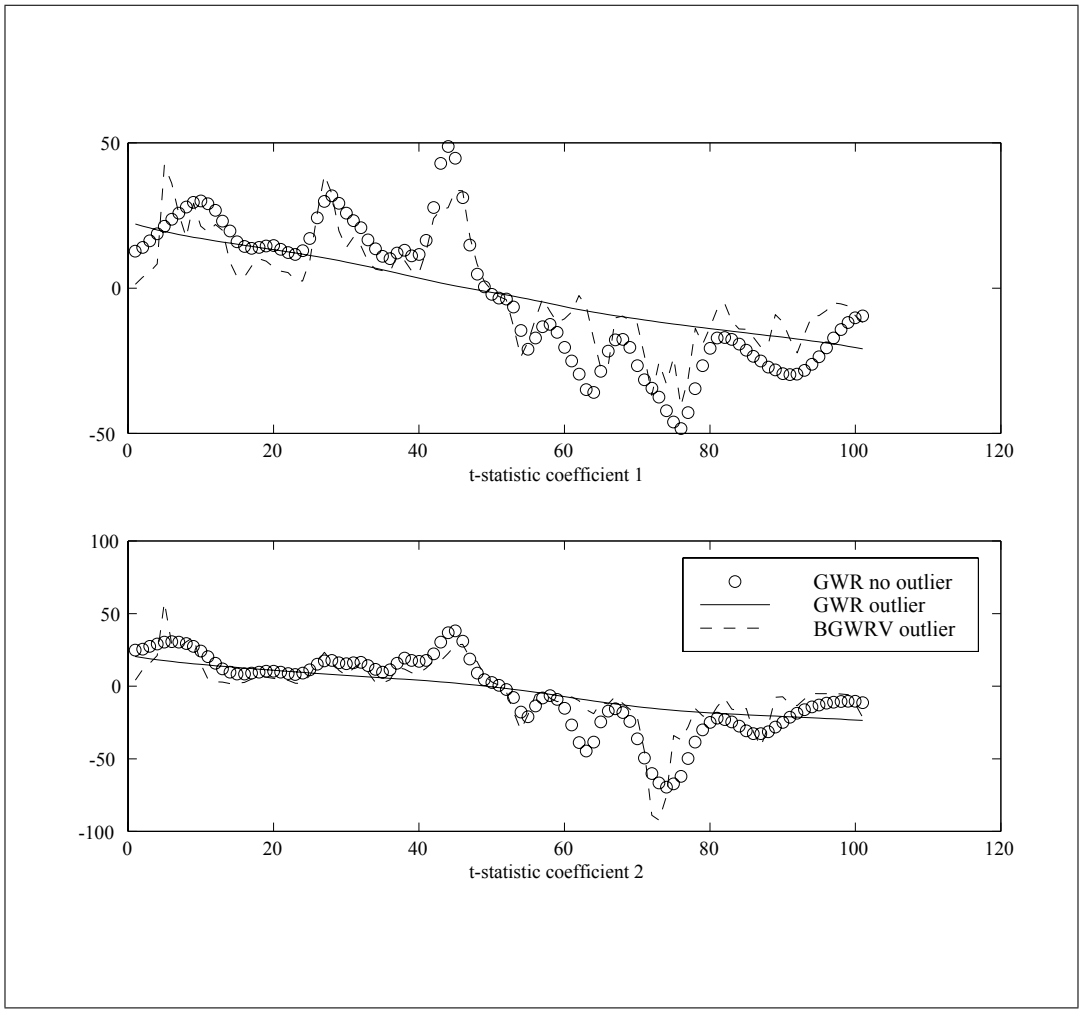

Figure 3: $t-$statistics for the GWR and BGWRV with an outlier

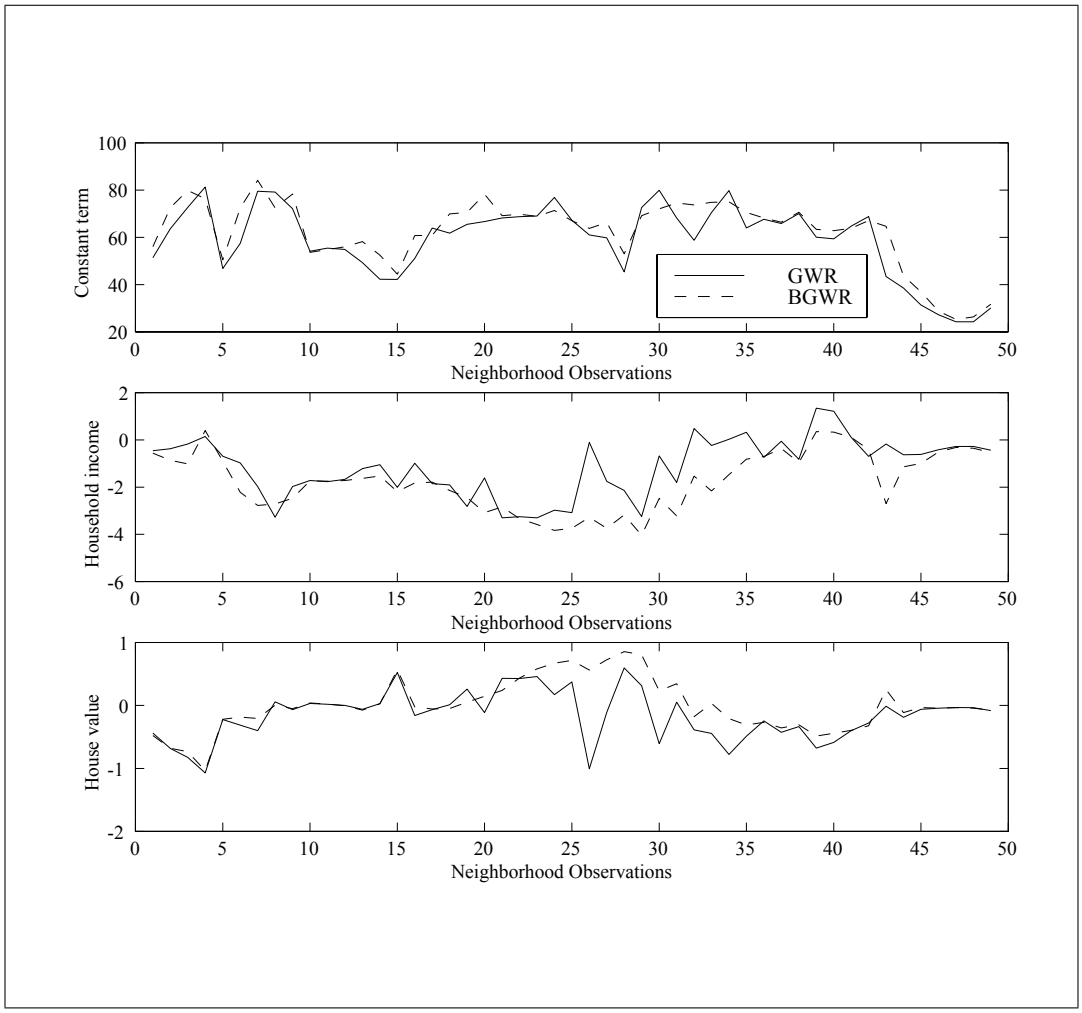

Figure 4: GWR versus BGWR estimates for Columbus data set

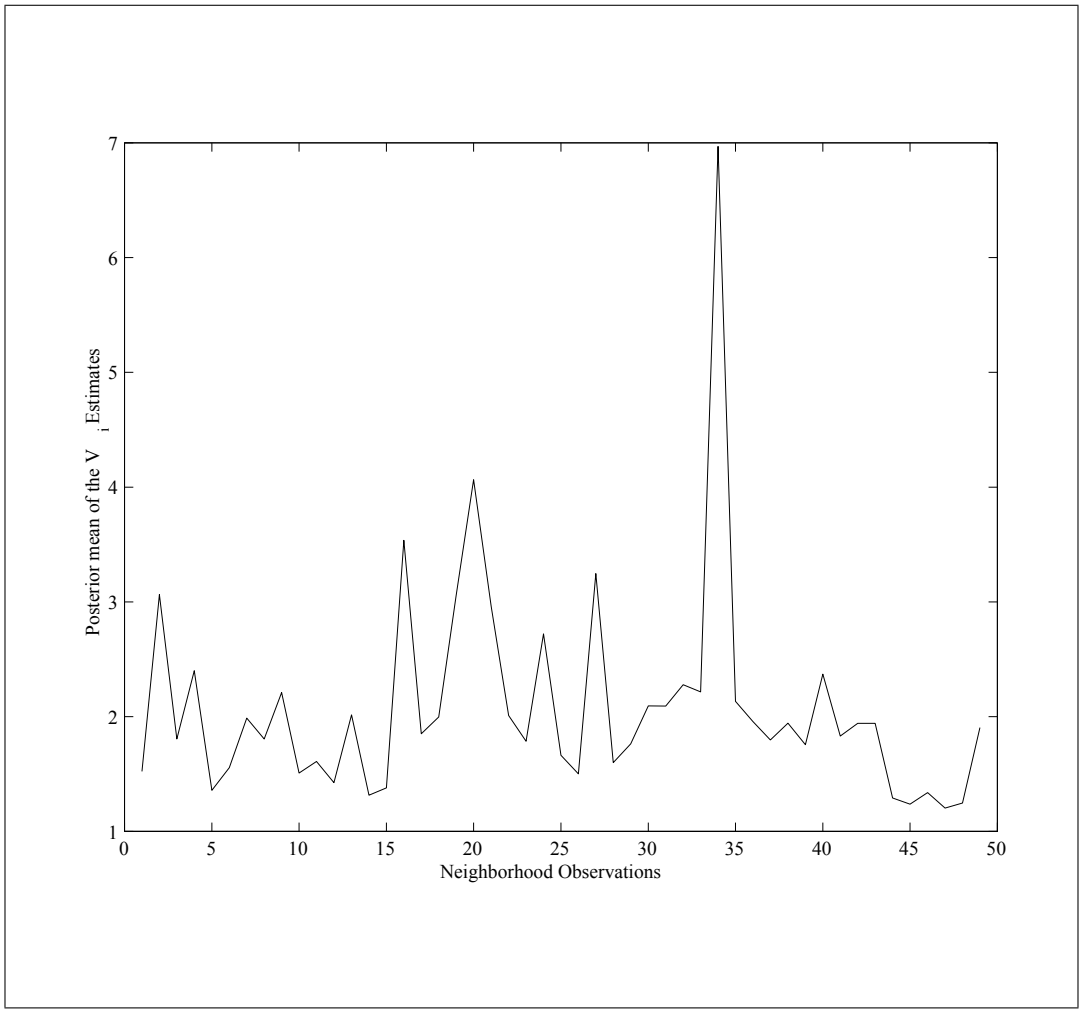

Figure 5: Average V_i estimates over all draws and observations

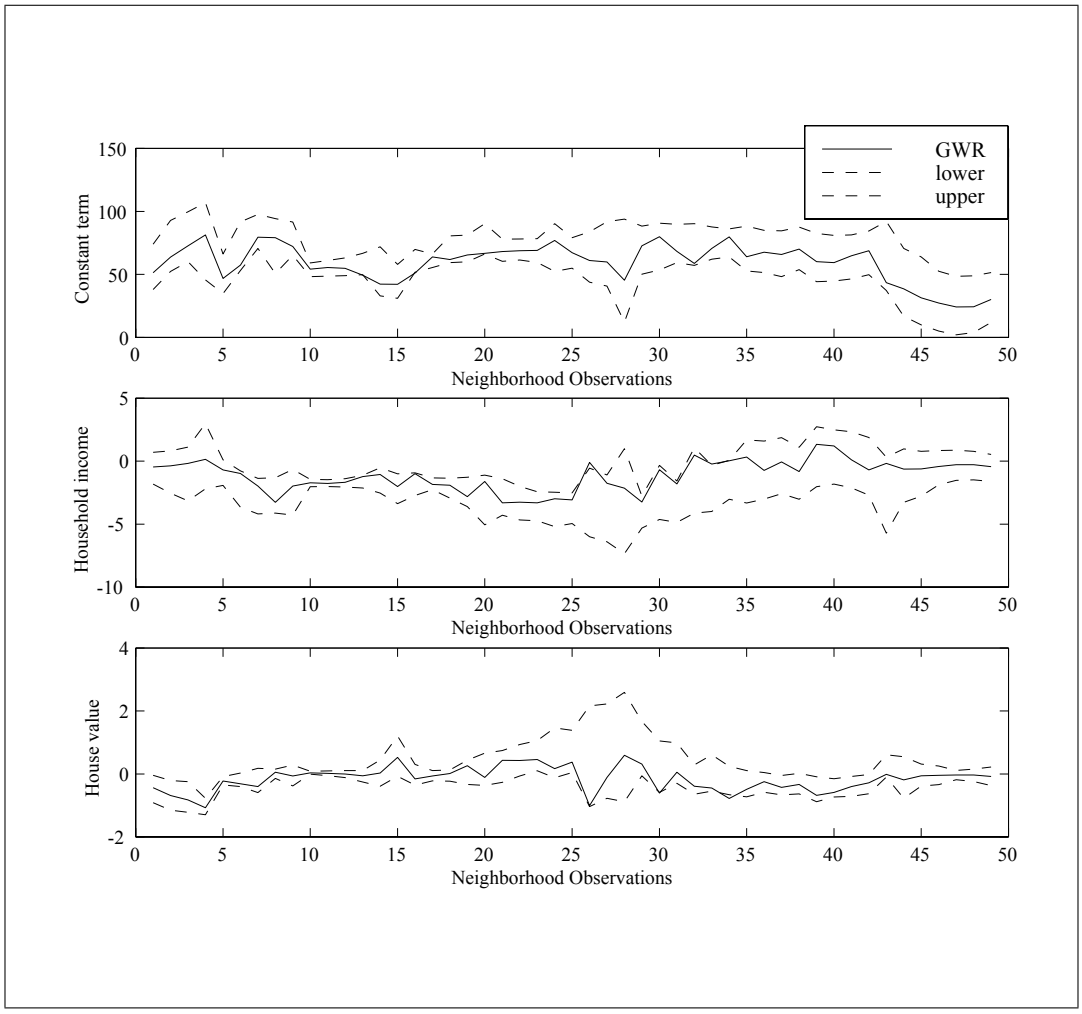

Figure 6: GWR versus BGWR confidence intervals

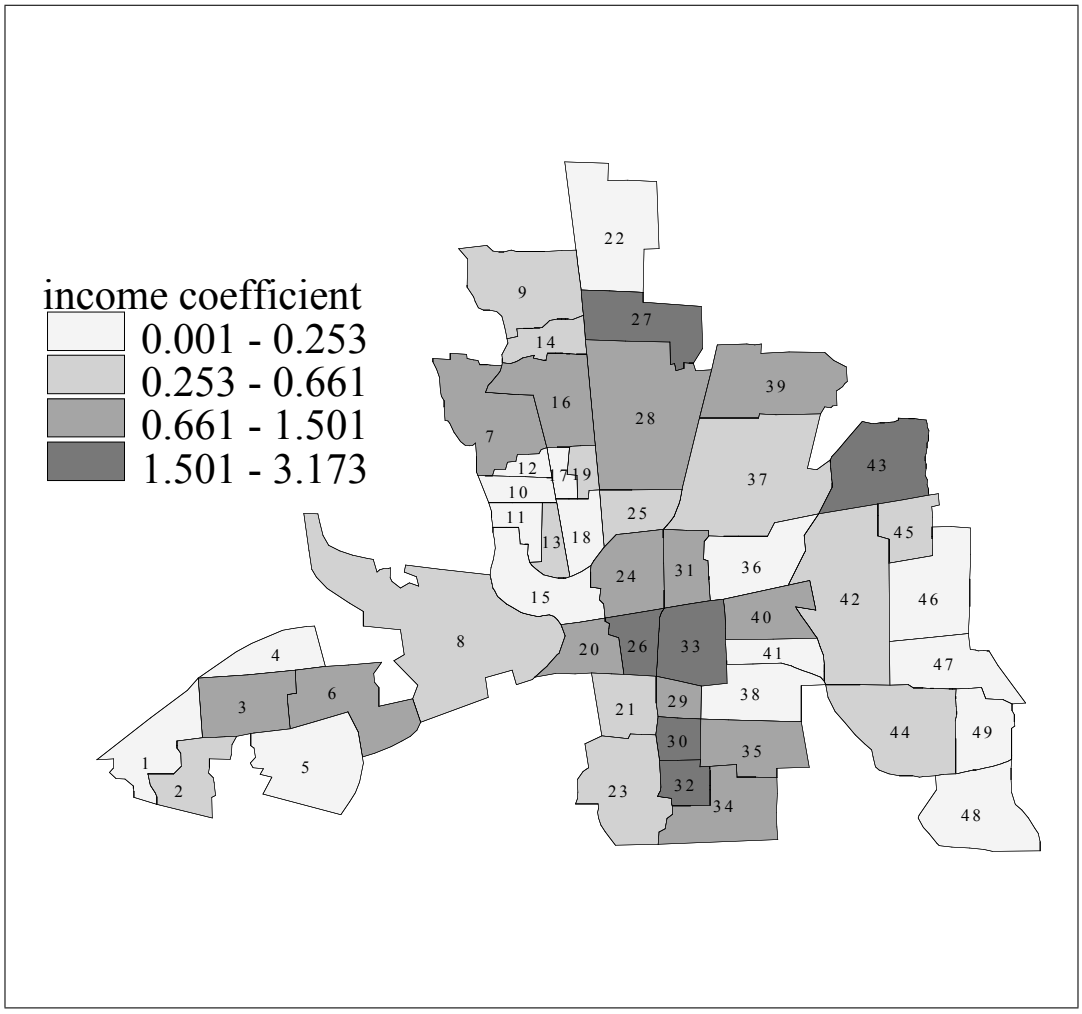

Figure 7: Absolute differences between GWR and BGWR household income estimates

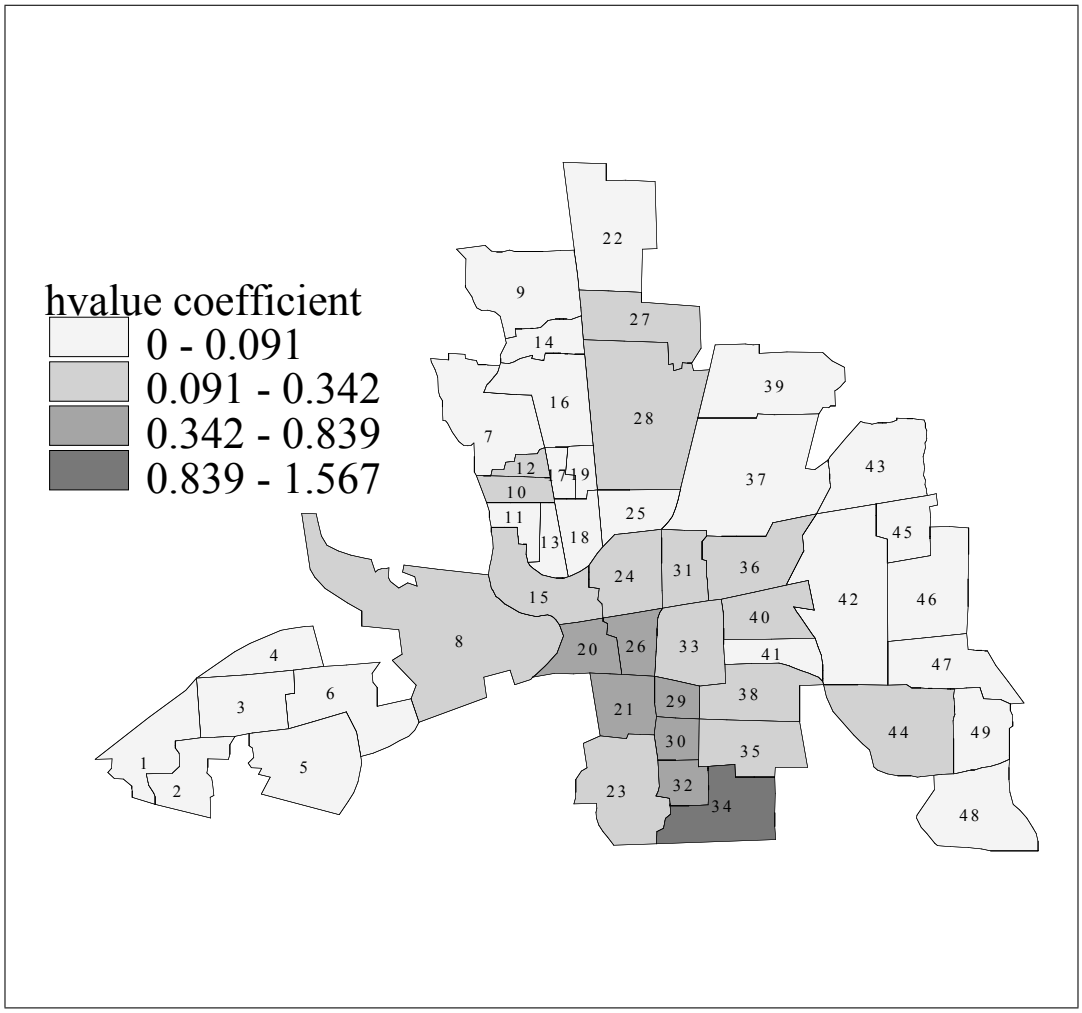

Figure 8: Absolute differences between GWR and BGWR house value estimates

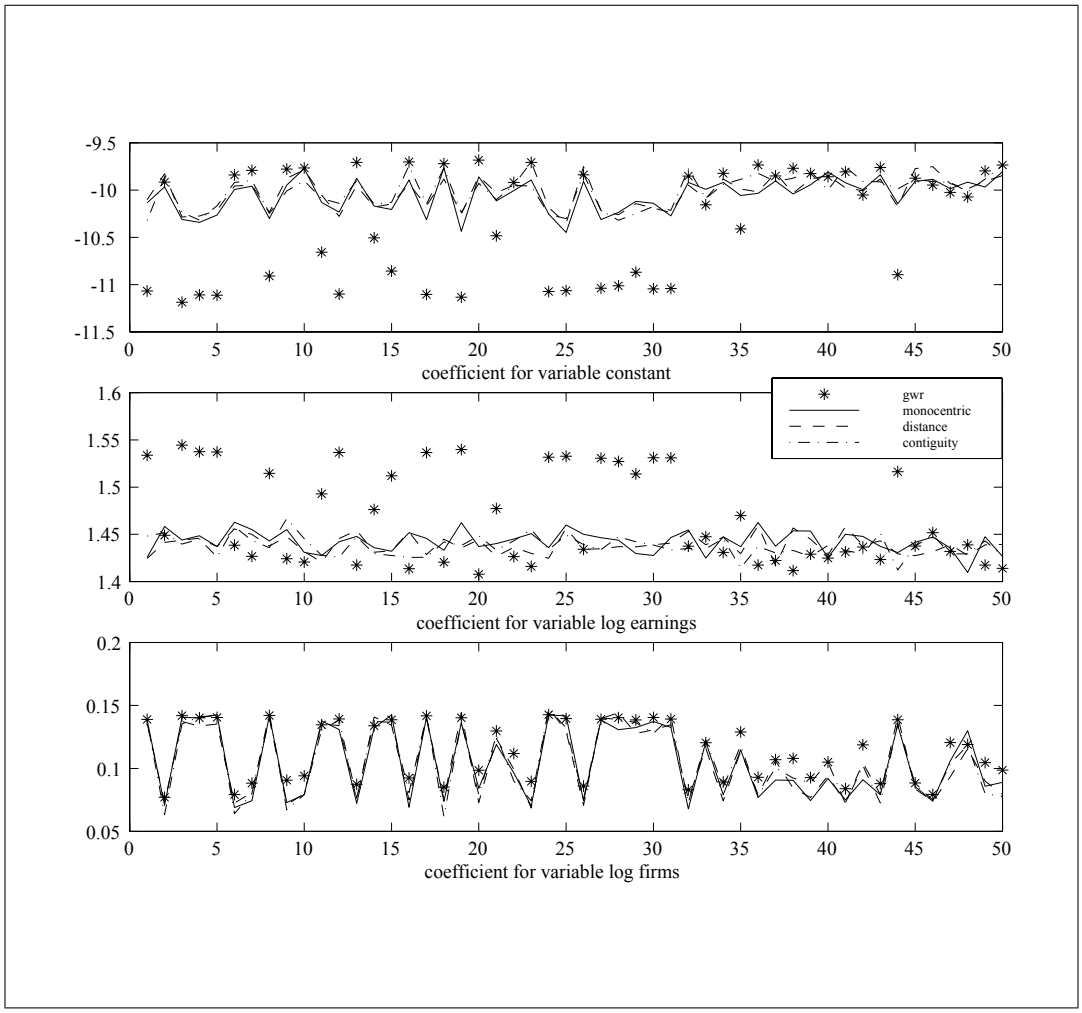

Figure 9: Ohio GWR versus BGWR estimates

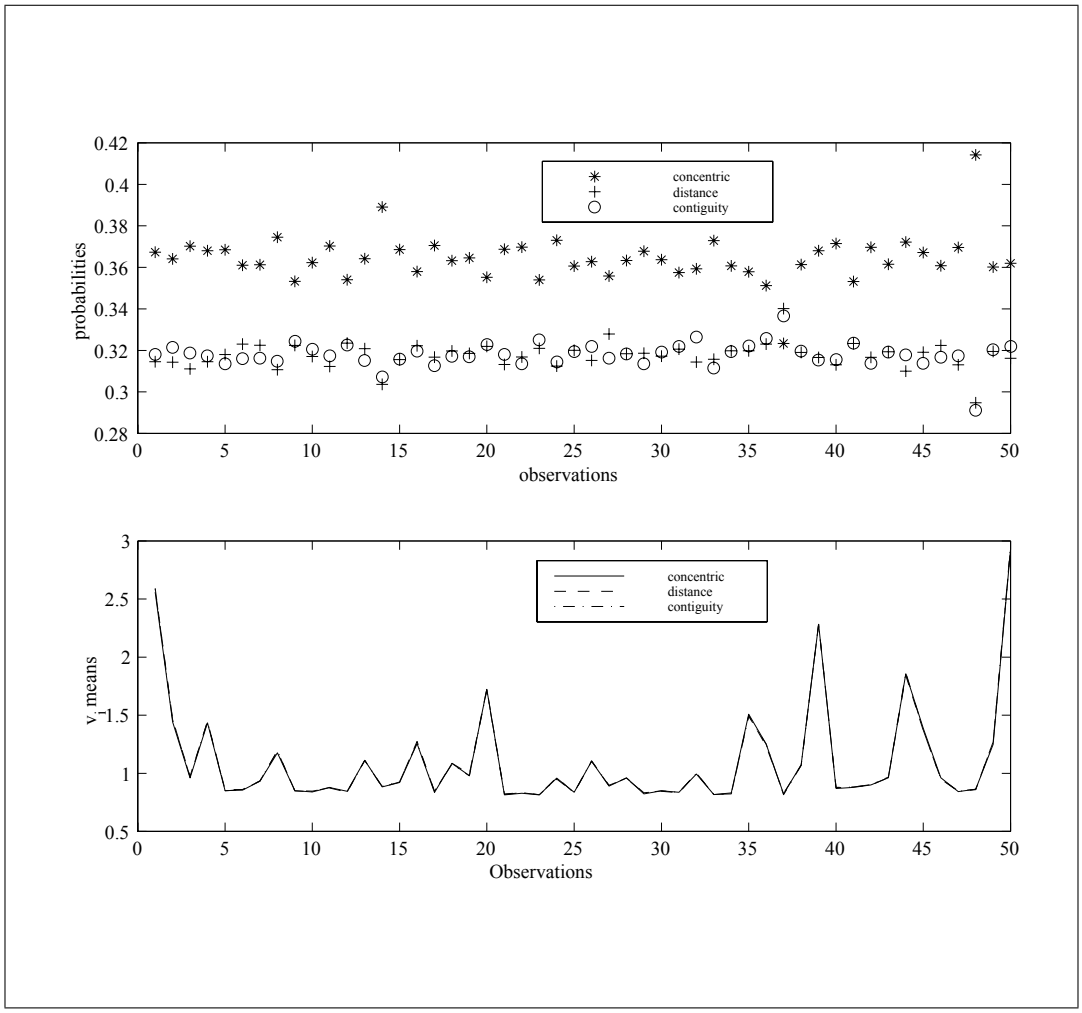

Figure 10: Posterior probabilities and v_i estimates

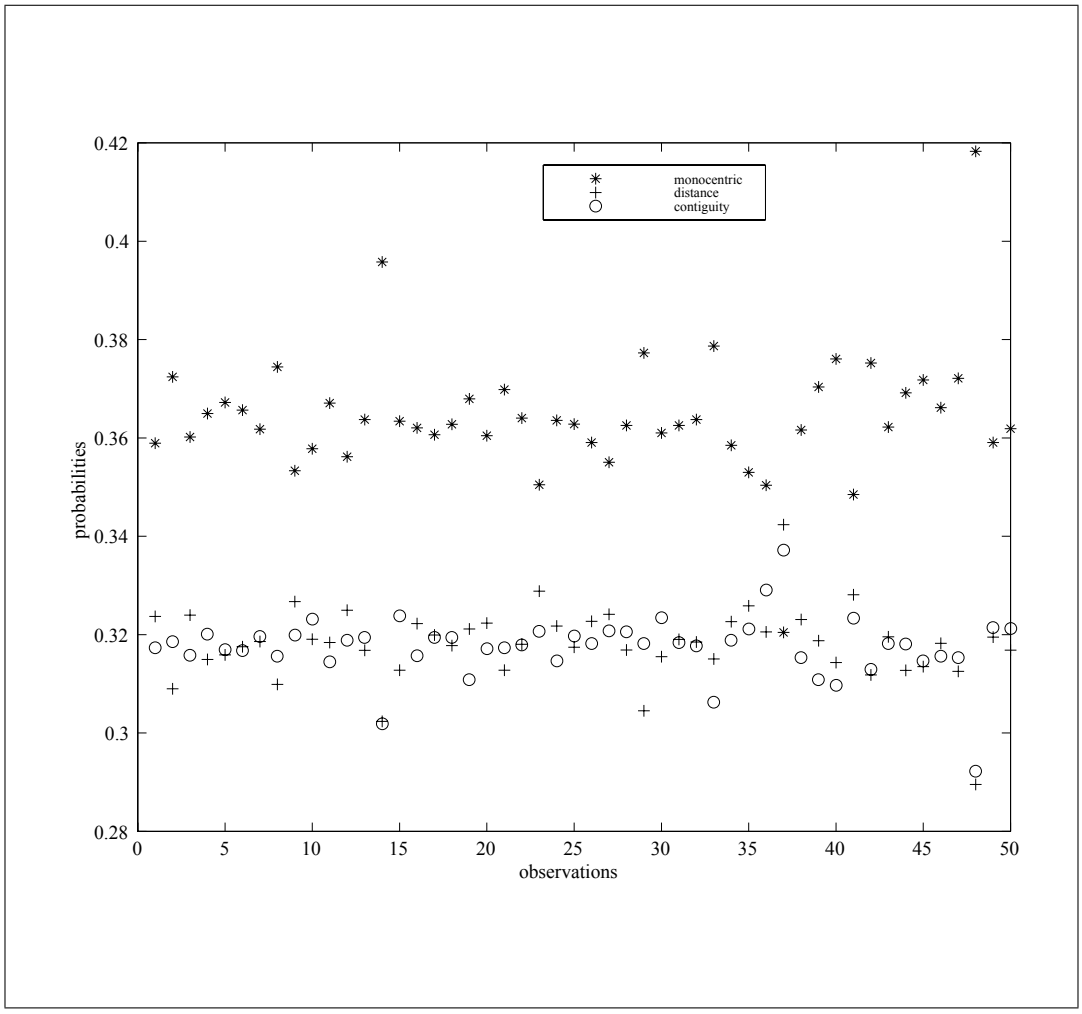

Figure 11: Estimates based on a tight imposition of the prior