

Simultaneous Spatial and Functional Form Transformations

By

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

Ronald Barry
Associate Professor of Statistics
Department of Mathematical Sciences
University of Alaska
Fairbanks, Alaska 99775-6660
(907)-474-7226
FAX: (907)-474-5394
FFRPB@aurora.alaska.edu

V. Carlos Slawson, Jr.
Assistant Professor of Real Estate
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803
(225)-388-6238
FAX: (225)-388-6366
clawson@lsu.edu

C.F. Sirmans
Professor of Real Estate and Urban Economics
Director, Center for Real Estate and Urban Studies
368 Fairfield Rd, Rm 426, U-41RE
Storrs, CT 06269-2041
(860) 486-3227
FAX: (860) 486-0349
CF@sba.uconn.edu

**Forthcoming, *Advances in Spatial Econometrics*, edited by
Raymond Florax and Luc Anselin.**

Abstract:

Parsimonious regression models using locational data often yield non-normal, heteroskedastic, and spatially dependent residuals. This manuscript develops a model which simultaneously performs spatial and functional form transformations to mitigate these problems. We apply the model to 11,006 observations on housing prices in Baton Rouge. For these data, the model reduces the interquartile range of the errors in the untransformed variable's space by 38.38% relative to a simple model using the untransformed variable. In addition, the pattern of residuals improves dramatically and the generalized additive model exhibits interesting transformations. Finally, the manuscript documents the computational innovations which make it possible to perform the maximization of the likelihood in under 10 seconds.

Keywords: Splines, Spatial Statistics, Non-parametric regression, Jacobians, Generalized Additive Models, GIS, Spatial Autoregression, House prices.

Acknowledgements:

We would like to thank Paul Eilers and Brian Marx for their comments, as well as the LSU Statistics Department Seminar participants. In addition, Pace and Barry would like to thank the University of Alaska for its generous research support. Pace and Sirmans would like to thank the Center for Real Estate and Urban Studies, University of Connecticut for their support. Pace and Slawson would like to thank Louisiana State University and the Greater Baton Rouge Association of Realtors for their support.

Simultaneous Spatial and Functional Form Transformations

I. Introduction

Technological advances such as the global positioning system (GPS) and low-cost, high-quality geographic information systems (GIS) have led to an explosion in the volume of large data sets with locational coordinates for each observation. For example, the Census provides large amounts of data for over 250,000 locations in the US (block groups). Moreover, geographic information systems can often provide approximate locational coordinates for street addresses (geocoding). Given the volume of business information which contains a street address field, this allows the creation of extremely large spatial data sets. Such data, as well as other types of spatial data, often exhibit spatial dependence and thus require spatial statistical methods for efficient estimation, valid inference, and optimal prediction.

Several barriers exist to performing spatial statistics with large data sets. Spatial statistical methods require the computation of determinants or inverses of n by n matrices. Allowing for space does not necessarily cure all of the problems encountered in typical data. For example, simple models fitted to housing and other economic data often exhibit heteroskedasticity, visible problems of misspecification for extreme observations, and non-normality (*e.g.*, Goodman and Thibodeau (1995), Subramanian and Carson (1988), Belsley, Kuh and Welsch (1980)). Simultaneously attacking these problems along with spatial dependence for large data sets presents a challenge.

Functional form transformations provide one technique which can simultaneously ameliorate all of these problems. For example, better specification of the functional form could reduce spatial autocorrelation of errors given spatial clustering of similar observations. While not guaranteed, functional form transformations often simultaneously reduce heteroskedasticity and residual non-normality. Because of the potential interaction between the spatial transformation and the functional form transformation, it seems desirable to fit these simultaneously.

Accordingly, we wish to examine the following transformation of the dependent variable,

$$(I - \alpha D)Y(\theta)$$

where D represents a n by n spatial weight matrix, α represents the spatial autoregressive parameter, and $Y(\theta)$ represents the dependent variable transformation parameterized by a vector of o parameters, θ . Least squares would not work for this problem as it would reduce the sum-of-squared

errors by reducing the range of the transformed variable. As an extreme case OLS could choose θ to make $Y(\theta)$ almost constant for a sufficiently flexible form and a regression with an intercept term would yield almost no error. Hence, this problem requires maximum likelihood with a Jacobian for the spatial transformation and a Jacobian for the functional form transformation.

The above form of the problem involves transformation of the functional form of the dependent variable first and the spatial transformation second. This seems a more natural formulation than transformation of the functional form of $(I - \alpha D)Y$ since the functional form of the dependent variable often has an interesting subject matter interpretation. However, spatial transformation first followed by functional form transformation is feasible and may offer some advantages.

The Box-Cox transformation is the most frequently used in regression. Recently, Griffith, Paelinck, and van Gastel (1998) discussed the importance of transformations for spatial data and examined bivariate Box-Cox/Box-Tidwell transformations of the dependent and independent variable in a spatial autoregression. The use of a parameter for the dependent variable as well as a parameter for the independent variable provided substantial flexibility in the potential transformation. Note, the Box-Cox/Box-Tidwell approach has an additional overhead in spatial problems as one must compute the spatially lagged value of the new transformed variables at each iteration.

We take a different route in modeling the functional form of the variables in a spatial autoregression. Specifically, we use B -splines (De Boor (1978), Ramsay (1988)) which are piecewise polynomials with conditions enforced among the pieces. The knots specify where each local polynomial begins and ends and the degree specifies the amount of smoothness among the pieces. A spline of degree 0 has no smoothness, a spline of degree 1 is piecewise linear, a spline of degree 2 is piecewise quadratic, and so forth.

Relative to the common Box-Cox transformation, the B -spline transformations do not require strictly positive untransformed variables and can assume more complicated shapes (Box and Cox (1964)). The standard one-parameter Box-Cox transformation either has a concave or convex shape. The B -spline transformation can yield convex shapes over part of the domain and concave shapes over the rest of the domain. Moreover, B -splines can yield more severe transformations of the dependent variable than the Box-Cox transformation. Burbidge (1988) discusses the deficiencies of the Box-Cox transformation and the need for more severe transformations of the extreme values of the untransformed dependent variable. These beneficial features do have a price. Relative to

transformations such as the Box-Cox, splines may require substantially more degrees-of-freedom. This could create problems for small data sets or those with low amounts of signal-to-noise (*i.e.*, low R^2).

Computationally, there are three components to the log-likelihood for this problem. These include (1) a spatial Jacobian; (2) a functional form Jacobian; and (3) the log of the sum-of-squared errors term.

To address the spatial Jacobian part of the log-likelihood, we use the techniques proposed by Pace and Barry (1997a,b,c) to quickly compute the Jacobian of the spatial transformation ($\ln|I - \alpha D|$). This involves the computation of $\ln|I - \alpha D|$ across a grid of values of α . With sparse D , special techniques exist which make this computational tractable.

To address the functional form Jacobian part of the likelihood, we employ two additional techniques to greatly accelerate computational speed. First, we use an intermediate transformation of the dependent variable. Intermediate transformations are often used in nonparametric regression (regression with very flexible functional forms). By adopting a transformation which partially models the nonlinearity, it requires less flexibility (fewer degrees-of-freedom) to model the remaining nonlinearity. The goal of our particular intermediate transformation is to make the dependent variable's histogram approximately symmetric.

Second, given an approximately symmetric dependent variable, we can employ evenly spaced knots. Equally spaced knots result in more observations between the central knots and fewer observations between the extreme knots. This makes the spline transformation more flexible in the tails and less flexible in the center, a desirable result. Such evenly spaced knots have often been used with B -splines (Hastie and Tibshirani (1990, p. 24)). Evenly spaced knots lead to a very simple functional form Jacobian (Eilers and Marx (1996), Shikin and Plis (1995, p. 44)) suitable for rapid computation.

To address the log of sum-of-squared errors portion of the log-likelihood, we use the linearity of the B -spline and spatial transformations to write the overall sum-of-squared errors as a series of the sum-of-squared errors from regressions on the individual parts of the transformation. This allows us to recombine the sum-of-squared errors from a set of regressions rather than recompute the sum-of-squared errors fresh each iteration.

Cumulatively, these computational techniques accelerate the log-likelihood computations so that each iteration takes little time. Each estimate requires around 1,000 iterations. Yet, these could be

computed in less than 10 seconds on a 200 megahertz Pentium Pro computer, even though the data set had 11,006 observations.

We apply this to a housing data set from Baton Rouge, Louisiana. The Real Estate Research Institute at Louisiana State University estimates regressions periodically to form an index of real estate prices over time. Since each house does not sell each quarter, the regression controls for the differences in sample composition over time by using a variety of independent variables such as age, living area, other area, number of bathrooms, number of bedrooms, and date of sale. In addition, variants of these data have been used to examine prediction accuracy of regression models (*e.g.*, Knight, Sirmans and Turnbull (1994)).

In real estate, predictions of the price of unsold homes have been extensively used for tax assessments. In fact, the majority of the districts in the country (and many foreign countries) use some form of statistical analysis to predict the prices of unsold homes (Eckert (1990)). In addition, the secondary mortgage markets have begun exploring the use of statistical appraisal for determining the value of collateral for loans (Gelfand *et al.* (1998), Eckert and O'Connor (1992)). Note, both of these applications give rise to very large spatial data sets.

To handle these needs, we estimated a general model which includes the previously discussed transformations of the dependent variable, transformations of the independent variables, spatially lagged independent variables, time indicator, and miscellaneous variables. As an illustration of the efficacy of the proposed techniques, the general model reduced the interquartile range of the residuals by 38.38% relative to a simple model using the untransformed dependent variable. Moreover, the resulting dependent variable transformation greatly improved the pattern of the residuals.

Most estimates of the Box-Cox parameters yield a model somewhere between a linear and logarithmic transformation. The estimated dependent variable transformation also fell between a linear and a logarithmic transformation – it was close to a linear transformation for low-priced properties but approached a logarithmic transformation for the high-priced properties. In fact, it actually provided more damping than the logarithmic transformation for extremely high-priced properties. Finally, the estimated functional forms of the independent variables seemed plausible and of interest.

Section II develops the joint spatial and dependent variable transformation estimator while section III applies the estimator to the Baton Rouge data. Section IV concludes the paper.

II. Simultaneous Spatial and Variable Transformations

This overall section presents the estimator and the various techniques facilitating computation. Section A sets up the log-likelihood, section B discusses the application of splines to the problem, section C shows how to simplify the SSE, section D provides a computational simplification of the spatial Jacobian, section E gives a simple way of computing the functional form Jacobian, and section F extends the model to transformations of the independent variables.

A. A Transformed Dependent Variable with Spatial Autoregression

Suppose the transformed variable follows a spatial autoregressive process,

$$\begin{aligned} Y(\theta) &= X\beta + u \\ u &= \alpha Du + \varepsilon \end{aligned} \tag{1}$$

where $Y(\theta)$ denotes the transformed dependent variable n element vector which depends upon the o element vector of parameters θ . In addition, X denotes an n by p matrix of the independent variables, D denotes an n by n spatial weight matrix, α represents the autoregressive parameter ($1 > \alpha \geq 0$), β denotes the p element vector of regression parameters, u denotes the spatially autocorrelated error term, while ε denotes a normal *iid* error term.

The spatial weight matrix D has some special structure. First, it has zeros on the main diagonal which prevents an observation from predicting itself. Second, it is a non-negative matrix and positive entries in the j th column of the i th row means observation j directly affects observation i . We do not assume symmetry and so the converse does not necessarily hold. Third, we assume each observation is only directly affected by its m closest neighbors. This makes D very sparse (high proportion of zeros) which greatly aids computational performance. Fourth, D is row-stochastic and so each row sums to 1. This gives D a smoothing or linear filter interpretation (Davidson and MacKinnon (1993)). Intuitively, $DY(\theta)$ provides a construct similar to a lag in time series for $Y(\theta)$.

To estimate (1), we rewrite it as,

$$(I - \alpha D)Y(\theta) = X\beta + \varepsilon \tag{2}$$

For a known α and θ , one could proceed to apply OLS to (2). Unfortunately, estimating α and θ via OLS results in biased estimates.

To motivate the defect in using OLS to estimate the parameters in this situation, consider momentarily the very simple model $(\xi I)Y = X\beta + \varepsilon$ where ξ represents a scalar parameter. If we

employ OLS to estimate both ξ and β , the estimated value of the parameter ξ would equal 0 for any value of β . This would turn the dependent variable vector ξY into a vector of zeros that a model with an intercept would fit perfectly.

To prevent this form of extreme behavior, one must employ maximum likelihood which explicitly penalizes such pathological transformations using the Jacobian of the transformation. The Jacobian of the transformation measures the n -dimensional volume change caused by stretching or compressing any or all of the potential n dimensions. By premultiplying Y via the matrix ξI , we are performing a linear transformation. In this case we are compressing or stretching each of the n dimensions of Y by a factor ξ . Relative to a unit value for ξ , values of $\xi < 1$ correspond to more singular transformations. The Jacobian of the transformation is the determinant of the matrix of derivatives which in this instance is ξ^n ($|\xi I| = \xi^n$).¹ To make the example even simpler, we are dealing with a cube when n is 3. If we multiply each dimension of the cube by a factor of 2, we increase the volume of the cube by a factor of 8 (2^3). The need for the Jacobian is not specific to the normal maximum likelihood, but arises whenever making transformations with proper, continuous densities (Davidson and MacKinnon (1993, p. 489), Freund and Walpole (1980, p. 230-252)).

Assuming normality, the profile log-likelihood for this example equals a constant plus the log of the Jacobian less $(n/2) \log(SSE(\xi))$. Taking as a reference point the sum-of-squared error when $\xi = 1$ ($SSE(\xi = 1)$), then $SSE(\xi) = SSE(\xi = 1)\xi^2$. As an example, multiplying Y by a constant of $1/2$ would multiply SSE by a constant of $1/4$. Hence, the profile log-likelihood becomes $n \log(\xi) - (n/2) \log(SSE(\xi = 1)\xi^2)$. A simple expansion shows that the likelihood will be the same for any choice of ξ . Hence, the maximum likelihood choice for β does not depend upon ξ . Thus, one cannot affect the maximum likelihood estimate by a simple scaling of the dependent variable, a highly desirable result.²

In this simple case, the role of the Jacobian in maximum likelihood is clear. The Jacobian continues to play a similar role in more complicated transformations such as those arising from spatial transformations or from functional form transformations. Successive transformations result in Jacobians multiplied by each other in the multivariate density. Hence, for simultaneous

¹ Determinants measure the n dimensional volume of the geometric object defined by its rows (or equivalently columns). See Lay (1997, p. 199-204) for a nice discussion of this point.

² Davidson and MacKinnon (1993) provide an excellent introduction to transformations in the context of maximum likelihood.

transformations the log of the Jacobian of ABC would equal the sum of the logs of the individual Jacobians (e.g., $\ln(J_{ABC}) = \ln(J_A) + \ln(J_B) + \ln(J_C)$ where J denotes the relevant Jacobian).

Hence, the profile log-likelihood for estimation using a spatial and a functional form transformation equals,

$$L(\alpha, \theta) = C_{lik} + \ln(J(Y)_\alpha) + \ln(J(Y)_\theta) - \left(\frac{n}{2}\right) \ln(SSE(\alpha, \theta)) \quad (3)$$

where $J(Y)_\alpha$ and $J(Y)_\theta$ represent the Jacobians of the spatial and dependent variable transformations and C_{lik} represents an arbitrary constant.

Attacking the maximization of the above log-likelihood in the most straightforward way would likely result in very long execution times. We show methods for greatly accelerating the computation of each of these terms. We detail these computational accelerations in the sections below.

B. Linear Expansions of Non-Linear Functions

If one computed $Y(\theta)$ and subsequently $(I - \alpha D)Y(\theta)$ for every iteration of the maximization of the log-likelihood, this could greatly reduce the speed of the algorithm as $(I - \alpha D)$ is an n by n (albeit sparse) matrix. Hence, we first seek ways of avoiding this step. Fortunately, if we can expand $Y(\theta)$ linearly, we can avoid this set of computations. A number of ways of linearly expanding a function exist. We could use indicator variables, polynomials, or splines. For our computations we chose B -splines (De Boor (1978, 1999)).

In fact, splines generalize both indicator variables and polynomials. Indicator variables provide a locally constant fit to a function for their non-zero portions. B -splines of degree 0 yield indicator variables. The advantage of indicator variables or B -splines of degree 0 is their local fit. Their disadvantage is that locally constant approximations are not necessarily continuous or smooth.

Polynomials, however, exhibit both continuity and smoothness. Polynomials attempt to approximate a function globally and gyrations of the function over parts of the domain can cause the polynomial to poorly fit other parts of the domain. A polynomial equates to a high degree B -spline with few knots.

Specifically, B -splines are piecewise polynomials with conditions enforced among the pieces. The knots specify where each local polynomial begins and ends and the degree specifies the amount

of smoothness among the pieces. A spline of degree 0 has no smoothness, a spline of degree 1 is piecewise linear, a spline of degree 2 is piecewise quadratic, and so forth.

To provide some physical intuition, a spline was a thin strip of wood used in constructing ships. The spline attached to two points separated by less than its length would cause the spline to produce a curve. By introducing supports (ribs of the ship), the curve could be modified into many shapes. Hence, the spline knots act similar to the ship's ribs. Moreover, the flexibility of the strip of wood would determine the smoothness (affect the degree of the spline). The piecewise linear splines used here correspond to laying a string across the ribs of the ship.

Also, one can restrict B -splines to yield strictly monotonic transformations. One must have monotonic transformations of dependent variables for prediction in the original dependent variable space (Ramsay (1988)). Finally, B -splines can interpolate a given set of values (assuming satisfaction of the Schoenberg-Whitney conditions (De Boors (1999, p. 1.10))). The Schoenberg-Whitney conditions essentially require that each of the B -spline basis vectors have at least one non-zero value. Hence, given a set of values, some weighting of the associated B -spline basis vectors could return the same set of values.

To explain splines in detail is beyond the scope of this article. However, a specific example greatly aids in understanding some of their features. In example 1 we consider four values for the dependent variable Y of 1, 1.5, 2.25, and 3.0. Given knots of 1, 2, and 3 (with 1 and 3 being repeated), we used the SPCOL function in the Matlab Spline Toolbox 2.01 to produce the following matrix $B(Y)$ comprised of three basis vectors. The exact values of the basis vectors depend upon Y and hence we emphasize this by writing $B(Y)$.

Example 1			
Y	$B(Y)$		
1.00	1.00	0.00	0.00
1.50	0.50	0.50	0.00
2.25	0.00	0.75	0.25
3.00	0.00	0.00	1.00

In example 1, $B(Y)$ illustrates a couple of B -spline features. First, $B(Y)$, the collection of basis vectors, contains only non-negative numbers. Second, each row sums to one. Third, the basis vectors have zero elements for elements of Y sufficiently far away from the knots. If we compute $B(Y)\theta$ for $\theta_{\text{interpolate}} = [1 \ 2 \ 3]$, we find it yields Y exactly. For other θ such that $\theta_1 < \theta_2 < \theta_3$, the plots of $B(Y)\theta$ versus Y show a monotonically increasing piecewise linear relations. Figures 1a, 1b, 1c, and 1d show four such plots. In every case, the selected θ satisfied the monotonicity constraints. Figure 1a shows how the function $B(Y)\theta$ exactly replicated the original Y (interpolated). Figure 1b shows a slightly concave transformation while 1c shows a more severe concave transformation. Figure 1d shows a convex transformation. With more points, one could generate combinations of convex and concave transformations (over different domains).

Assuming satisfaction of the Schoenberg-Whitney conditions, with B -splines our transformed dependent variable becomes,

$$Y(\theta) = B(Y)\theta \quad (4)$$

where $B(Y)$ represents the n by o matrix containing the basis vectors and θ represents the o by 1 parameter vector. The number of basis vectors, o , depends upon the number of knots and the degree of smoothness required. As o rises, the transformed dependent variable $Y(\theta)$ can assume progressively more flexible forms.

Substituting (4) into (2) yields (5).

$$(I - \alpha D)Y(\theta) = (I - \alpha D)B(Y)\theta = [B(Y) \quad DB(Y)] \begin{bmatrix} \theta \\ -\alpha\theta \end{bmatrix} \quad (5)$$

Hence, one can linearly expand the joint spatial and dependent variable into the product of a n by $2o$ matrix and a $2o$ by 1 parameter vector.

C. SSE Simplifications

Let M represent the idempotent least squares matrix $I - X(X'X)^{-1}X'$. We can write the residuals from the regression of $(I - \alpha D)Y(\theta)$ on X as (6),

$$e = M \left([B(Y) \quad DB(Y)] \begin{bmatrix} \theta \\ -\alpha\theta \end{bmatrix} \right) = [MB(Y) \quad MDB(Y)] \begin{bmatrix} \theta \\ -\alpha\theta \end{bmatrix} = E\rho \quad (6)$$

where the n by $2o$ matrix E contains all the residuals from the individual regressions and the vector ρ represents the $2o$ element parameter vector. The linearity of the problem means the least squares residuals e on the overall transformed variable $(I - \alpha D)Y(\theta)$ are simply a linear combination of the least squares residuals from regressing each basis vector in $B(Y)$ and their spatial lags $DB(Y)$ on X . Hence, forming parameterized sum-of-squared errors yields (7).

$$SSE(\alpha, \theta) = e'e = \rho'(E'E)\rho \quad (7)$$

Note, the $2o$ by $2o$ error cross-product matrix $E'E$ is only computed once. Subsequent iterations of $\rho'(E'E)\rho$ involve only order of o^3 operations, a very small number which does not depend upon n , the number of observations or k , the number of regressors. Moreover, o is usually much less than k and strictly less than n . This reduction in the dimensionality of the sum-of-squared errors leads to an low dimensional profile likelihood (Meeker and Escobar (1995)).

D. Spatial Jacobian Simplifications

Historically, the spatial Jacobian, $\ln|I - \alpha D|$, constituted the main barrier to fast computation of spatial estimators (*e.g.*, Li (1995)). However, the use of a limited number of spatial neighbors lead to sparse matrices. Pace and Barry (1997b, c) show how various permutations of the rows and columns of such sparse matrices $(I - \alpha D)$ can vastly accelerate the computation of $\ln|I - \alpha D|$. Although computation of $\ln|I - \alpha D|$ is inexpensive for a particular value of α , one can further accelerate the computations by computing $\ln|I - \alpha D|$ for a large numbers of values of α (*e.g.*, 100) and interpolating intermediate values. Insofar as α has a limited range (for stochastic D) and the function $\ln|I - \alpha D|$ is quite smooth, the interpolation exhibits very low error.

Moreover, these computations are performed only when changing the weight matrix D . Hence, one can reuse the grid of values (and interpolated points) when fitting different models involving Y and X for a given D .

Pace and Barry have released a public domain Matlab-based package, “Spatial Toolbox 1.0” available at www.finance.lsu.edu/re which implements these spatial Jacobian simplifications and contains copies of the articles which describe the implementation details.

E. Functional Form Jacobian Simplifications

The functional form log-Jacobian has a particularly simple form for piecewise linear splines with evenly spaced knots,

$$\ln(J(Y)_\theta) = C + n_2 \ln(\theta_2 - \theta_1) + n_3 \ln(\theta_3 - \theta_2) + \dots + n_{o-1} \ln(\theta_{(o-1)} - \theta_{(o-2)}) \quad (8)$$

where $n_2, n_3, \dots, n_{(o-1)}$ represents the number of non-zero elements of all but the first and last basis vectors and the distance between knots determines the constant C (Eilers and Marx (1996), Shikin and Plis (1995, p. 44)). This very simple form lends itself to extremely rapid execution. Piecewise linear splines also facilitate enforcing strict monotonicity. Provided $(\theta_{j+1} - \theta_j) > 0$, $J(Y)_\theta > 0$.

Unfortunately, an even placement of knots may not work well in many cases. However, transforming the original variable Y may result in a variable $g(Y)$ where an even knot placement will work better. In which case, the log of the Jacobian involving an intermediate transformation can be partitioned into the original log-Jacobian and a log-Jacobian for the intermediate transformation.

$$\ln(J(g(Y))_\theta) = \sum_{i=1}^n \ln\left(\frac{dg(Y)}{dY_i}\right) + \ln(J(Y)_\theta) \quad (9)$$

The intermediate transformation $g(\cdot)$ does not depend upon the parameters α or θ and hence these do not affect its contribution to the functional form Jacobian. However, the intermediate transformation $g(\cdot)$ does help adjust the placement of knots and therefore has some effect upon the final fit. Parameterizing knot placement within a maximum likelihood framework could make it easier to assess its statistical consequences.

Even knot placement results in nested models in some cases. For example, if the most flexible model uses 12 knots, sub-models with six, four, three, and two knots correspond to parameter restrictions placed on the 12 knot model. Again, this aids the assessment of the statistical consequences of knot placement.

F. Extension to Functions of the Independent Variables

Naturally, one could include a spline expansion of the independent variables. In addition, one could include spatial lags of the independent variables. Let Z represent the untransformed independent variables. We could model X , the regressors as,

$$X = [B(Z) \quad DB(Z)] \quad (10)$$

where $B(Z)$ represents the spline expansion of each one of the columns of Z . Note, without deletion of one basis vector for each column of Z , X would be linearly dependent as the sum of the rows of all the basis vectors always equals 1 for B -splines. Hence, if each basis function expansion takes o vectors, $B(Z)$ will have dimension of $p(o-1)$. Adding the spatial lags doubles the variable count. The spline expansion of each one of the core independent variables Z allows one to create a generalized additive model (Hastie and Tibshirani (1990)). In addition, this particular model allows the spatially lagged variables to follow a different functional form.

$$(I - \alpha D)Y(\theta) = [B(Z) \quad DB(Z)]\beta + \varepsilon = \sum_{i=1}^p f(Z_i) + \sum_{i=1}^p Dh(Z_i) + \varepsilon \quad (11)$$

This very general specification subsumes the case of autocorrelated errors. This restriction would also make $f(\cdot) = h(\cdot)$. Imposing this restriction would substantially slow the speed of computing the estimates. However, the use of restricted least squares would still provide much more speed than a formulation which required computing $(X'X)$ each iteration. Moreover, this restriction will often be rejected by the data as n becomes large.

III. Baton Rouge Housing

This overall section presents the application of the techniques developed in the previous section to housing data from Baton Rouge. Section A discusses the data, section B gives details on the construction of the spatial weight matrix, section C provides timing and other information on the determinant computations, section D presents the general model, section E discusses the estimated dependent variable transformation, section F discusses the estimated independent variable transformations, section G shows how to conduct the inference in this model, section H discusses model performance in the untransformed variable space, and section I conducts an experiment to document the uniqueness of the estimates and computation times.

A. Data

We selected observations from the Baton Rouge Multiple Listing Service which (1) could be geocoded (given a location in latitude and longitude based upon the house's address); (2) had complete information on living area, total area, number of bedrooms, and number of full and half baths. In addition, we also discarded negative entries for these characteristics. In total, 11,006 observations survived these joint criteria.

B. Spatial Weight Matrix

To construct the spatial weight matrix D , compare the distance d_{ij} between every pair of observations i and j to d_i^m , the distance from observation i and its m th nearest neighbor. It seems reasonable to set to 0 the direct influence of distant observations upon a particular observation. Accordingly, assign a weight of $\frac{1}{m}$ to observations whenever d_{ij} is greater than 0 and is less than or equal to d_i^m as in (12),

$$0 < d_{ij} \leq d_i^m \leftrightarrow D_{ij} = \frac{1}{m}. \quad (12)$$

By construction D will be row-stochastic but not necessarily symmetric. For this particular problem, we set m equal to 4.

C. Determinant Computations

Following Pace and Barry (1997b) we computed $\ln|I - \alpha D|$ for $\alpha = 0.01, 0.02, \dots, 0.99$. The LU decomposition of $(I - \alpha D)$ results in the triangular matrices L and U , where the diagonal of U contains the pivots r_i . By construction, $(I - \alpha D)$ is strictly diagonally dominant and hence has bounded error sensitivity (Golub and Van Loan (1989)). The magnitude of the determinant is determined by the product of the pivots r_i or the log-determinant by the sum of $\ln(r_i)$.

Computation of the 100 determinants took 57.6 seconds on an 200 megahertz Pentium Pro computer. By employing some of the permutation algorithms discussed in Pace and Barry (1997b) or by employing some devices to exploit symmetry as in Pace and Barry (1997c) we could further accelerate these times.

Given the grid of log-determinant values, we employed linear interpolation to arrive at intermediate values.

D. Model

We fitted the following model to the data. Each of the functions $f(\cdot), h(\cdot)$ for the independent variable's living area, other area, and age comes from piecewise linear B -splines with knots at the minimum value, the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 99th quantiles, and the maximum value. Specifically, we used the Matlab Spline Toolbox (Version 1.1.3) function SPCOL to create the necessary basis vectors. Hence, applying SPCOL to a particular variable such as age would result in an n by 11 matrix whose columns contained the basis vectors. A particular linear combination of

these basis vectors would create the function $f(\cdot)$ while a different linear combination of the same basis vectors would create $h(\cdot)$. De Boor wrote the Spline Toolbox and the functions in it closely resemble those described in De Boor (1978).

For the discrete full bath and beds variables, these functions are formed from indicator variables at each of the values these discrete variables assume. In addition, we used single indicator variables to control for age missing values, for age greater than 150 years, for the presence of half-baths, and for the year of sale. For both the spline and the sets of indicator variables, we deleted one column to prevent linear dependence as the row-sum of B -splines equals 1 as does the sum of a complete set of indicator variables.

$$\begin{aligned}
(I - \alpha D)g(\text{Price}) = & f_1(\text{living area}) + f_2(\text{other area}) + f_3(\text{age}) \\
& + f_4(\text{full baths}) + f_5(\text{beds}) + \\
& + Dh_1(\text{living area}) + Dh_2(\text{other area}) + Dh_3(\text{age}) \\
& + Dh_4(\text{full baths}) + Dh_5(\text{beds}) + I_1(\text{age missing})\beta_1 \\
& + I_2(\text{age} > 150)\beta_2 + I_3(\text{half bath} > 0)\beta_3 + I_{1985-1992}\beta_{4-11} + \varepsilon
\end{aligned} \tag{13}$$

The full model involves 113 parameters. This very general model will hopefully span the true model. Moreover, the general model provides a way of investigating other potential problems and a starting point for subset selection. See Hendry, Pagan, and Sargan (1984) for more on the advantages of general to specific modeling.

E. Estimated Dependent Variable Transformations

As discussed in II.E, the use of an intermediate transformation $g(\cdot)$ makes it possible to modify the effects of equal knot placements. We selected the Box-Cox transformation $g(Y) = (Y^\varphi - 1) / \varphi$ with log-Jacobian $(1 - \varphi) \sum \ln(Y_i)$ for this step. We examined the transformation for a grid of φ and selected $\varphi=0.25$ based upon maximizing the normality of Y as measured by the studentized range. This induced approximate symmetry which made equal knot placement viable. We used 11 equally placed knots.

Based upon other work with transformations (*e.g.*, Burbidge (1988)) we expected most reasonable transformations would induce linearity for the bulk of the observations. The approximate normality of Y coupled with equal placement gave the desirable result of having a greater number of knots in the tails as opposed to the center of the density of Y . This gave the potential transformation more flexibility in the tails where the differences among transformations emerge.

Figure 2 shows Y , $\ln(Y)$, and $Y(\theta)$, the optimal piecewise linear spline transformation of Y , plotted against $\ln(Y)$. The optimal transformation $Y(\theta)$ acts similar to a linear transformation for low-priced houses and acts more like the logarithmic transformation for high-priced houses.

Figure 3 shows the effects of this optimal transformation. Figure 3c shows the extreme heteroskedasticity (positively related to price) created by not using any transformation. Note the untransformed dependent variable model systematically underpredicts the high-priced properties as well.

Figure 3d shows the extreme heteroskedasticity (negatively related to price) created by using the logarithmic transformation. Note the logarithmically transformed dependent variable model overpredicts low-priced properties as well.

Figure 3b shows the intermediate transformation (Box-Cox with $\lambda = 0.25$) created heteroskedasticity for both low and high-priced properties and also created problems of systematic over and under prediction at the extremes of the price density.

Figure 3a shows how the spline transformation cures the problem of heteroskedasticity. Moreover, inspection of the low and high-priced properties does not reveal a systematic pattern of under or over prediction. Figure 4a shows the histogram of standardized residuals from the spatial regression on the transformed dependent variable with a normal curve superimposed. Similarly, Figure 4b shows the histogram of standardized residuals from the spatial regression on the untransformed dependent variable with a normal curve superimposed. Relative to the untransformed dependent variable spatial regression, the errors from the spatial regression on the transformed variable show substantially less leptokurtosis.

Previous work, such as Knight, Sirmans and Turnbull (1994), avoided the problem of heteroskedasticity by truncating large portions of the sample based upon price.

F. Estimated Independent Variable Transformations

Figure 5 shows the optimal functions of the independent variables. Note, we did not enforce strict monotonicity with these optimal functions. Figure 5a depicts $f_1(\text{living area})$, which apart from a decreasing section for very small houses not often observed in the sample, shows a positive, concave relation between $Y(\theta)$ and living area. Miscoding of observations, such as leaving out a digit in the living area field, provides one possible explanation for this decreasing section. For

example, if there are average-priced houses with 0 reported living area, the model might actually show a rise in price as living area goes to 0.

As depicted by Figure 5b, age shows a decreasing relation up until about 40 years when it rises and declines again at 100 years. The Age variable confounds two phenomena. First, physical and hence economic depreciation rises with age. Second, age reflects the year of construction. If the year of construction proxies for features such as wood floors, high ceilings, or other desirable traits, one could see a non-monotonic relation between age and price. In addition, remodeling confuses the issue as the age of the improvements differs from the age of the original structure. Goodman and Thibodeau (1995) also found a non-monotonic relation between age and price. “Dwellings 20-40 years old appreciated slightly, while older dwellings depreciate.”

As depicted by Figure 5c, other area shows a very positive, concave relation between $Y(\theta)$ and other area. As depicted by Figure 5d, baths shows a positive, concave relation between $Y(\theta)$ and baths up until four baths. Subsequently, it declines slightly. Again not many houses have five baths or more.

One would not necessarily expect a monotonic relation between bedrooms and price. Holding other variables constant, more bedrooms means smaller bedrooms. Hence, bedrooms is a design value with some optimal value. As depicted by Figure 5e, this optimum is at three bedrooms, a plausible value. Finally, Figure 5f shows the relation between $Y(\theta)$ and year-of-sale. This shows the precipitous drop in housing prices in 1988 which has been documented by others (*e.g.*, Knight, Sirmans and Turnbull (1994)).

We also examined the optimal independent variable transformations for the original untransformed dependent variable (no spatial or dependent variable transformations). For the most part, these arrived at qualitatively similar independent variable transformations. Some differences appeared. For example, the optimal transformation for living area was slightly convex instead of concave, baths showed a more precipitous drop for houses with more than five bathrooms, and age showed a rise after 20 years (as opposed to around 35 years for the model with spatial and dependent variable transformations).

G. Inference

Given the fast computation of the log-likelihood, it seems reasonable to conduct inference via likelihood ratio tests. Table 1 presents these likelihood ratios for a wide variety of hypotheses. In all cases these were significant at well beyond the 1% level. Hence, both the spatial and the transformation parts of the model seem highly significant. The spatial autoregressive parameter, α , equaled 0.5820 and had a deviance ($-2\log(\text{LR})$) of 3936.62 with only one hypothesis. The transformation $Y(\theta)$ also proved quite significant with a deviance of 8114.82 with 10 hypotheses. Only 10 parameters vary independently due to the affine invariance of the regressand for linear regression. Note, deleting the transformation parameters equates to running a pure spatial model. For the pure spatial model α equaled 0.5099. Hence, rather than the transformation removing spatial autocorrelation through better specification, the model acted to transform the dependent variable to increase the use of the autocorrelation correction.

The individual variables were all significant with living area showing the greatest impact on the log-likelihood with a deviance of 3364.92. The general model dominated simpler models with fewer variables. Compared to running a regression with the untransformed dependent variable coupled with a simple set of independent variables ignoring space and transformations, the deviance was 14782.04 with 82 hypotheses.

The use of restricted least squares, which avoids recomputing $(X'X)$, further aids in the speed of computing these likelihood ratio tests.

Finally, we do not account for the statistical consequences created by the monotonicity constraint. However, one could easily use a Bayesian inequality estimator as in Geweke (1986) to show how the prior associated with the monotonicity constraint affects the posterior distributions of the parameters of interest. See Gilley and Pace (1995) for an application of this estimator to another house price data set.

H. Performance in the Original Dependent Variable Space

Part of the goal of fitting the general model was to improve upon prediction over simpler models in the original dependent variable space (Price). Given the Y and the strictly positive monotonic transformation $Y(\theta)$, we can take the prediction in the transformed space, $\hat{Y}(\theta)$ and with interpolation compute the prediction in the original space, \hat{Y} . Even if $\hat{Y}(\theta)$ comes from an unbiased

estimator of $Y(\theta)$, \hat{Y} does not unbiasedly estimate Y . To control for this bias, we allowed for it using the smearing estimator of Duan (1983).

We computed the predictions for a variety of models in the original dependent variable space. The performances of these models in the original dependent variable space appear in Table 2. We began with Model 5, a simple model in price space without transformation or spatial modeling of the independent or dependent variables. One could consider Model 5 as the standard model without using any transformations. The results from Model 5 closely match others in the literature. For example, Knight, Sirmans and Turnbull (1994) examined the relation between list and transactions prices for the Baton Rouge data to investigate buyer search behavior. Their model uses a very similar specification and has a R^2 of 0.72. The R^2 for Model 5 was a very similar 0.7299. This provides a benchmark for the subsequent models.

The residuals are asymmetric in Model 5 so while the mean error equals 0 by construction, the median error equals -530.14 dollars and the 25th and 75th quartiles are $-10,660.61$ and $9,707.98$ dollars. Given the average price of the houses in the sample is $\$75,597$, this does not represent particularly good performance. Model 4, which includes spatial independent variables and transformed independent variables, improves considerably on Model 5. It shows more symmetric errors and dominates Model 5 for every order statistic. Similarly, Model 3 adds transformation of Y , and also improves on Model 4 for most order statistics. Model 2 does not use transformations of Y but does add spatially lagged Y . It shows a large reduction relative to previous models for all but the minimum and 1st quantiles of the empirical error density.

Model 1, the general model, displays considerable improvements over the previous models, except for the 95th quantile to the maximum of the empirical error density where the spatial model without dependent variable transformations (Model 2) displays lower error. Relative to the simple Model 5, Model 1 has a 38.38% lower interquartile range of the empirical error density. In addition, relative to Model 4, the next best performing model, it shows a 8.6% reduction in the interquartile range of the empirical error density. Hence, the improvements in the transformed space carry back to the untransformed space.

I. Timing and Uniqueness

Local maxima are the bane of complicated maximum likelihood models. To examine this problem in the context of this problem, we estimated the model 250 times with different random

starting points. We picked α randomly from $[0,1)$. We picked θ_i from $[0,1]$ with the restriction that $\theta_{i-1} > \theta_i$ to generate strictly positive monotonic starting points.

It took 493 iterations at minimum and 1642 iterations at maximum to find the optimum. On average it took less than 10 seconds to arrive at the maximum likelihood estimates (given previous computation of $E'E$ and $\ln|I - \alpha D|$) using a computer with a 200Mhz Pentium Pro processor. All of the 250 estimates converged to the same log-likelihood value with a maximum error of 0.08 from the iteration which took the longest to converge.

IV. Conclusion

Locational data may suffer from both spatial dependence and a host of other problems such as heteroskedasticity, visible evidence of misspecification for extreme values of the dependent variable, and non-normality. Functional form transformations of the dependent variable often jointly mitigate these problems. Moreover, the transformation to reduce spatial dependence and the transformation of the functional form of the dependent variable can interact. For example, a reduction in the degree of functional form misspecification can also reduce the degree of spatial autocorrelation in the residuals. Alternatively, the functional form transformation may make the spatial transformation more effective. In fact, the latter occurred for the Baton Rouge data as the spatial autoregressive parameter rose from 0.5099 when using the untransformed variable to 0.5820 when using the transformed variable.

Application of the joint spatial and functional form transformations to the Baton Rouge data provided a number of gains relative to simpler models. First, the pattern of residuals in the transformed space improved dramatically. For example, unlike the residuals from simpler models, the general model's residuals seemed evenly divided by sign for all predicted values. Second, the magnitude of the sample residuals dropped dramatically even in the untransformed variable's space. Specifically, the interquartile range of the residuals from the general model using all the transformations when taken back into the untransformed variable's space fell by 38.38% relative to the residuals on a simple model with the untransformed variable. Third, the general model provided interesting insights into the functional form of the dependent and independent variables. The estimated functional form for the dependent variable followed an approximately linear transformation for low-priced properties, an approximately logarithmic transformation for high-

priced properties, and a somewhat more severe than logarithmic transformation for the very highest-priced properties.

The computation of the model employs several innovations. First, it relies upon the sparse matrix techniques proposed by Pace and Barry (1997a,b, c) to compute 100 log-determinants of the 11,006 by 11,006 spatial transformation matrix in 57.6 seconds using a 200 megahertz Pentium Pro computer. Interpolation of this grid of log-determinants provides the spatial log-Jacobian which greatly accelerates maximum likelihood maximization. Second, it uses an intermediate transformation to allow the use of evenly-spaced knots which have a particularly simple log-Jacobian for the functional form. Third, it expresses the overall sum-of-squared error as a linear combination of the sum-of-squared errors on individual parts of the transformations. Consequently, the actual maximization of the log-likelihood for the joint transformation takes less than 10 seconds on average (given prior computation of the spatial log-Jacobian and the individual sum-of-squared error computations). This part of the maximization of the log-likelihood does not directly depend upon the number of observations or the total number of regressors. The optimum appears unique as 250 iterations with different starting points returned the same log-likelihood value.

The computational speed of this model has at least two implications. First, inference can proceed by relatively straightforward likelihood ratio tests. The use of restricted least squares, which avoids recomputing $(X'X)$, further aids in the speed of computing the likelihood ratios. Second, the model becomes useful for exploratory work with large spatial data sets, an area which currently suffers from a lack of tools. By simultaneously fitting a generalized additive model and controlling for spatial dependence, it potentially provides a good first view of locational data. Such views can suggest simpler parametric specifications and the need for other adjustments such as reweighting. Naturally, the model could accommodate reweighting with an additional Jacobian for the weights.

While we primarily worked with economic data with this model, we suspect it could have applications to other fields. As the volume of spatial data continues to rise, methods which simultaneously and quickly adapt to the problems which arise in large data sets should come into more common use.

Bibliography

- Belsley, David. A., Edwin Kuh, and Roy. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*. New York: John Wiley, 1980.
- Box, G.E.P., and D.R. Cox. "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B* 26 (1964), 211-243.
- Burbidge, John B., Lonnie Magee and A. Leslie Robb. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83 (1988), 123-127.
- Cheney, Ward and David Kincaid, *Numerical Mathematics and Computing*, 2nd Edition, Brooks-Cole: Pacific Grove, CA, 1985.
- Davidson, Russell, and James MacKinnon. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.
- De Boor, Carl. *A Practical Guide to Splines*. Berlin: Springer, 1978.
- De Boor, Carl. *Matlab Spline Toolbox User Guide Version 2.0.1*. Mathworks: Natick, MA, 1999.
- Duan, N. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78 (1983), 605-610.
- Eckert, J. *Property Appraisals and Assessment Administration*. Chicago: International Association of Assessing Officers, 1990.
- Eckert, Joseph K., and Patrick M. O'Connor. "Computer-Assisted Review Assurance (CARA): A California Case Study." *Property Tax Journal* 11 (1992), 59-80.
- Eilers, Paul H. C., and Brian D. Marx. "Flexible Smoothing with B-Splines and Penalties." *Statistical Science* 11 (1996), 89-121.
- Freund, John, and Ronald Walpole, *Mathematical Statistics*, Third Edition, Prentice-Hall, 1980.
- Gelfand, Alan E., Sujit K. Ghosh, John R. Knight and C.F. Sirmans. "Spatio-Temporal Modeling of Residential Sales Data." *Journal of Business and Economic Statistics* 16 (1998), p. 312-21.
- Geweke, John. "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics*, Vol. 1, January 1986, p. 127-141.
- Gilley, O.W., and R. Kelley Pace, "Improving Hedonic Estimation with an Inequality Restricted Estimator," *Review of Economics and Statistics* 77 (1995), 609-621.
- Golub, G. H., and C. F. Van Loan. *Matrix Computations, second edition*. John Hopkins, 1989.
- Goodman, Allen C., and Thomas G. Thibodeau. "Age-Related Heteroskedasticity in Hedonic House Price Equations." *Journal of Housing Research* 6 (1995), 25-42.
- Griffith, Daniel, Jean Paelinck, and Reinaud van Gastel. "The Box-Cox Transformation: Computational and Interpretation Features of the Parameters," In D. Griffith, C. Amrhein, and J-M. Huriot. *Econometric Advances in Spatial Modelling and Methodology*. Amsterdam: Kluwer, 1998, 46-56.

- Hastie, T.J., and Robert Tibshirani. *Generalized Additive Models*. London: Chapman and Hall, 1990.
- Hendry, David, Adrian Pagan and Denis Sargan. "Dynamic Specification." In Z. Griliches, and M. Intriligator., Eds., *Handbook of Economics*. Amsterdam: North Holland, 1984.
- Knight, John, C.F. Sirmans, and Geoffrey Turnbull. "List Price Signaling and Buyer Behavior in the Housing Market." *Journal of Real Estate Finance and Economics* 9 (1994), 177-192.
- Lay, David. *Linear Algebra and its Applications*. Reading, MA: Addison-Wesley, 1997.
- Li, Bin. "Implementing Spatial Statistics on Parallel Computers." In S. Arlinghaus., Eds., *Practical Handbook of Spatial Statistics*,. Boca Raton: CRC Press, 1995, pp. 107-148.
- Meeker, W. Q., and L.A. Escobar. "Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimates." *The American Statistician* 49 (1995), 48-53.
- Pace, R. Kelley, and Ronald Barry. "Sparse Spatial Autoregressions." *Statistics and Probability Letters* 33 (1997), 291-297.
- Pace, R. Kelley, and Ronald Barry. "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable." *Geographical Analysis* 29 (1997), 232-247.
- Pace, R. Kelley, and Ronald Barry. "Fast CARs." *Journal of Statistical Computation and Simulation* 59 (1997), 123-147.
- Ramsay, J. O. "Monotone Regression Splines in Action." *Statistical Science* 3 (1988), 425-441.
- Shikin, Eugene, and Alexander Plis. *Handbook on Splines for the User*, Boca Raton, CRC Press, 1995.
- Subramanian, Shankar, and Richard T. Carson. "Robust Regression in the Presence of Heteroskedasticity." In : G.F. Rhodes and T.B. Fomby., Eds., *Advances in Econometrics*. Greenwich, Conn: JAI Press, 1988, pp. 85-138.

Table 1 — Likelihood Ratio Tests

Models	Log-likelihood	Deviance	Degrees of Freedom	Critical Values 1%
Unrestricted Model	-154849.65			
Model Sans Beds Indicators	-154905.70	112.10	14	29.1
Model Sans Bath Indicators	-154936.91	174.52	12	26.2
Model Sans Age Spline	-154979.21	259.12	20	37.6
Model Sans Other Area Spline	-155131.04	562.78	20	37.6
Model Sans Time	-155317.68	936.06	11	24.7
Model Sans Living Area Spline	-156532.11	3364.92	20	37.6
Model Sans Lagged Dependent Variables ($\alpha=0$)	-156817.96	3936.62	1	6.6
Model Sans Spatial Lagged Independent Variables	-155095.00	490.70	56	83.5
Model Sans Transformation ($\theta=0$)	-158907.06	8114.82	10	23.2
Log Dependent Variable Model with Spatial Lagged Independent Variables	-160313.48	10927.66	12	26.2
Linear Dependent Variable Model with Spatial Lagged Independent Variables	-160206.97	10714.64	12	26.2
Log Dependent Variable Model Sans Spatial Lagged Independent Variables	-162032.58	14365.87	82	114.7
Linear Dependent Variable Model Sans Spatial Lagged Independent Variables	-162240.67	14782.04	82	114.7

Table 2 — Sample Error Statistics Across Models For Prediction of the Untransformed Dependent Variable

	Model 1	Model 2	Model 3	Model 4	Model 5
Spatial <i>Y</i>	1	1	0	0	0
Spatial <i>X</i>	1	1	1	1	0
Transformed <i>Y</i>	1	0	1	0	0
Transformed <i>X</i>	1	1	1	1	0
Min	-173303.03	-228289.63	-220671.59	-241016.53	-252491.46
1 st	-35807.31	-45655.02	-43785.12	-50025.25	-58528.35
5 th	-20261.98	-23054.30	-25135.14	-28153.28	-33423.99
10 th	-14270.14	-15912.10	-17853.04	-19654.08	-23087.08
25 th	-6387.17	-6809.01	-8684.07	-9123.23	-10660.61
50 th	42.30	348.76	-340.64	-15.06	-530.14
75 th	6164.72	6927.99	7989.47	8762.24	9707.98
90 th	13924.82	14010.39	18207.61	18189.98	21588.44
95 th	21122.11	20214.30	27702.55	26686.41	32908.49
99 th	52523.81	48008.43	63033.51	54432.72	73978.17
Max	328574.03	276177.59	409496.21	341369.79	389299.09
Interquartile Range	12,551.89	13,737.00	16,673.54	17,885.47	20,368.59

Figure 1a

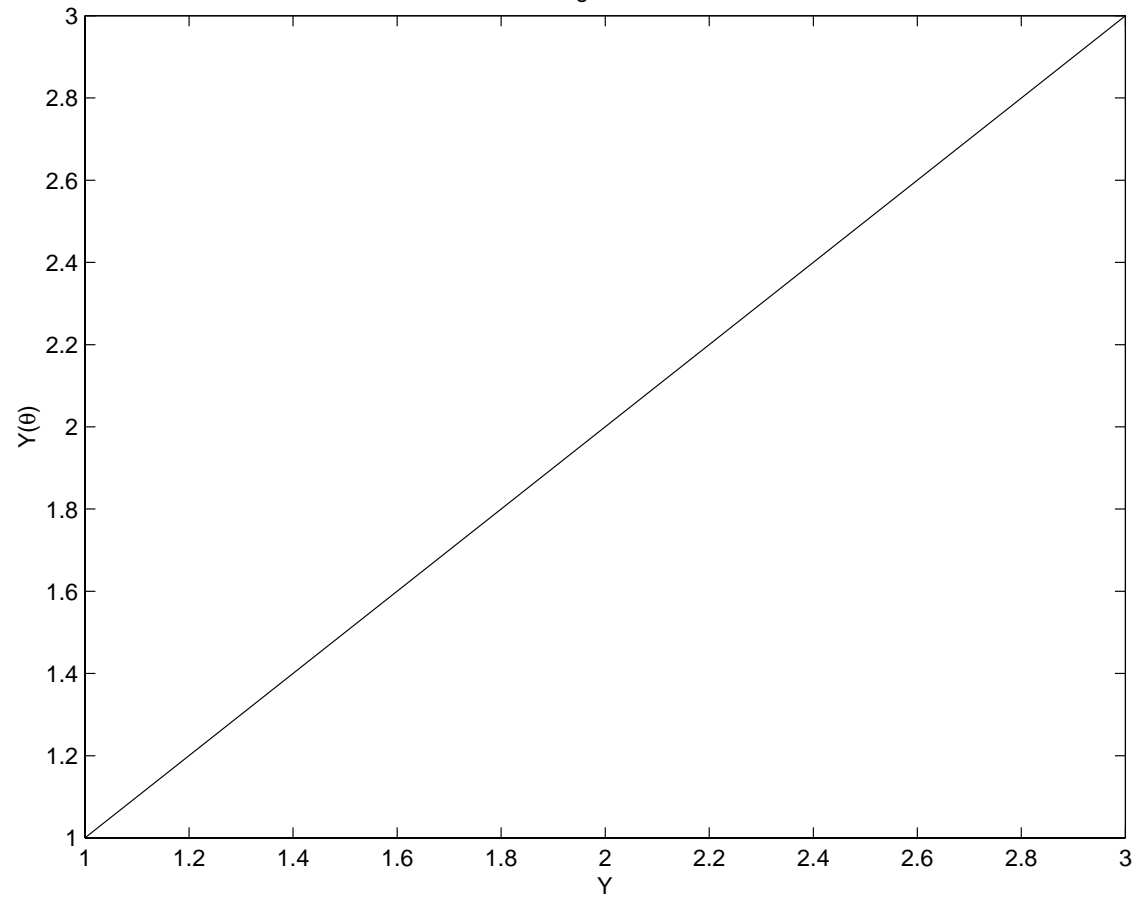
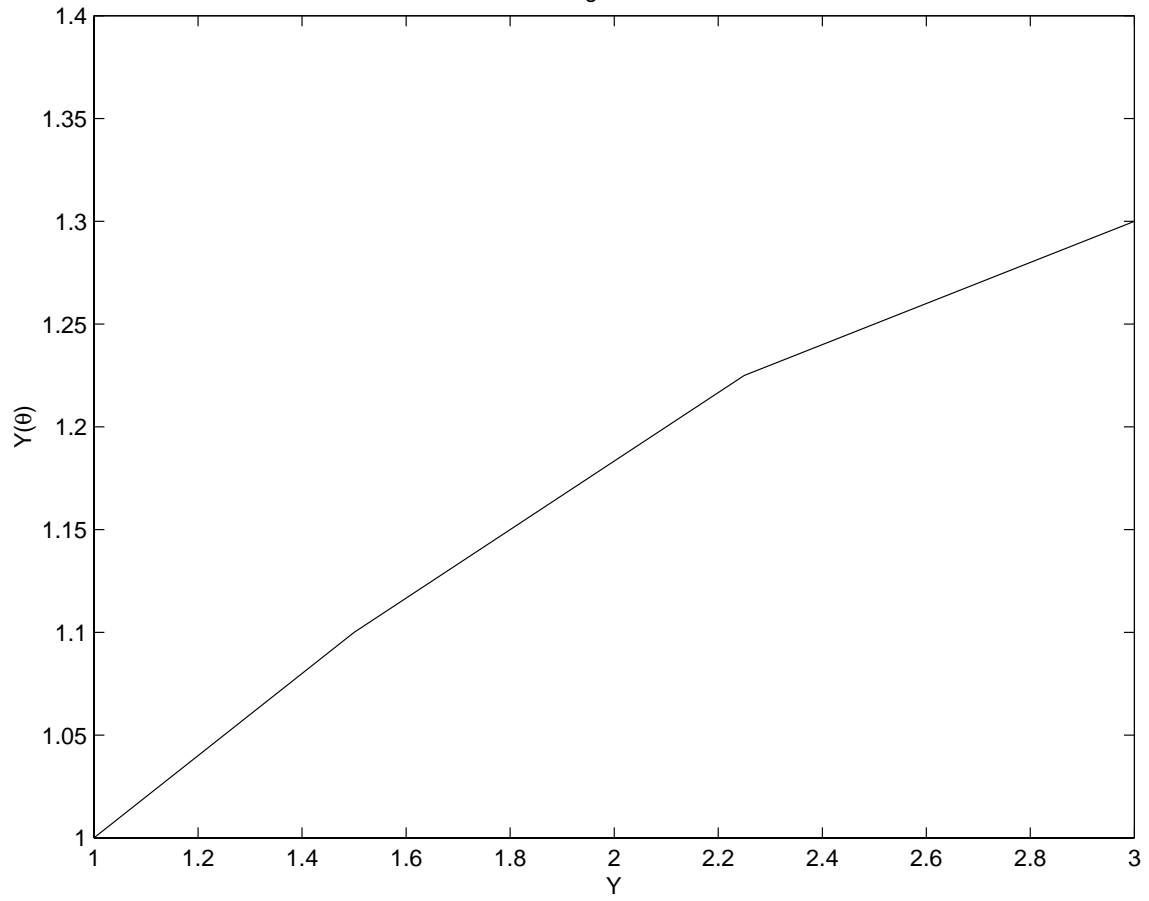
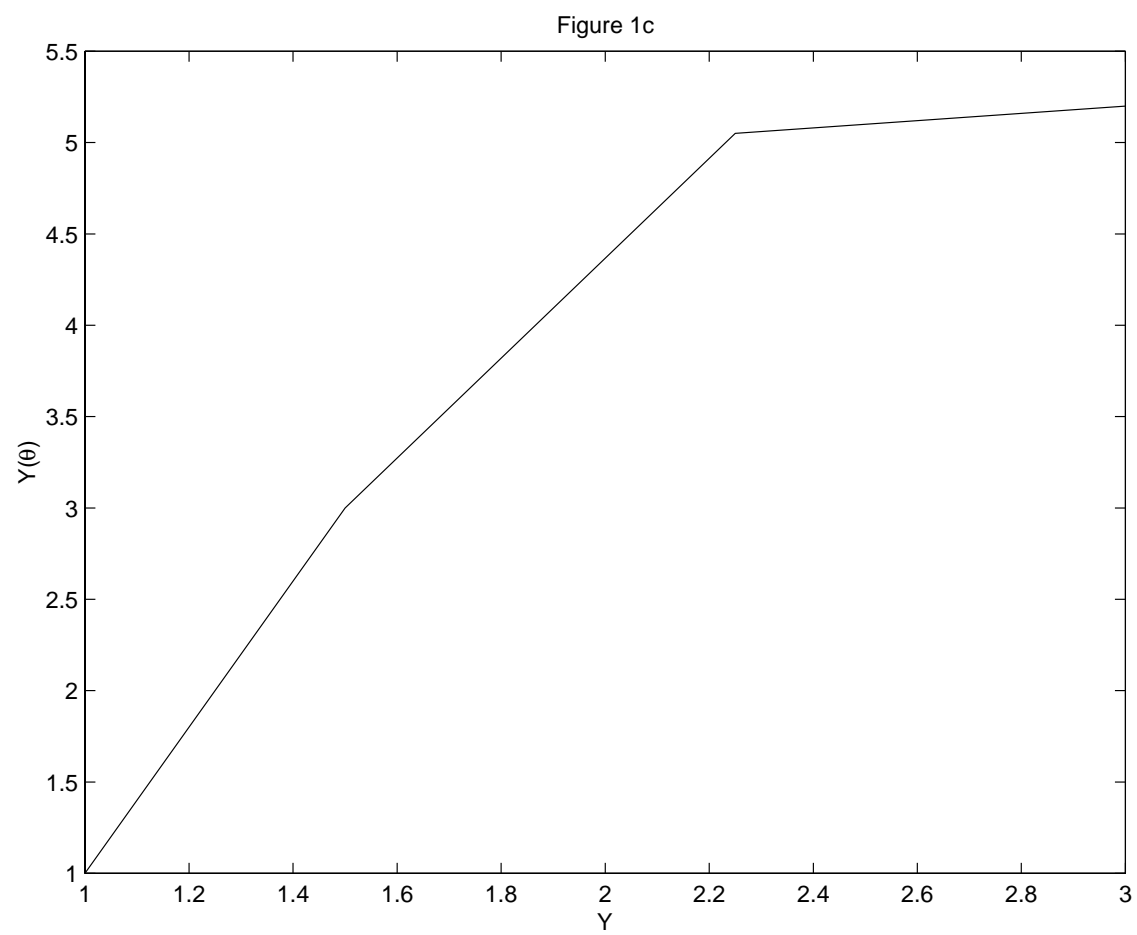
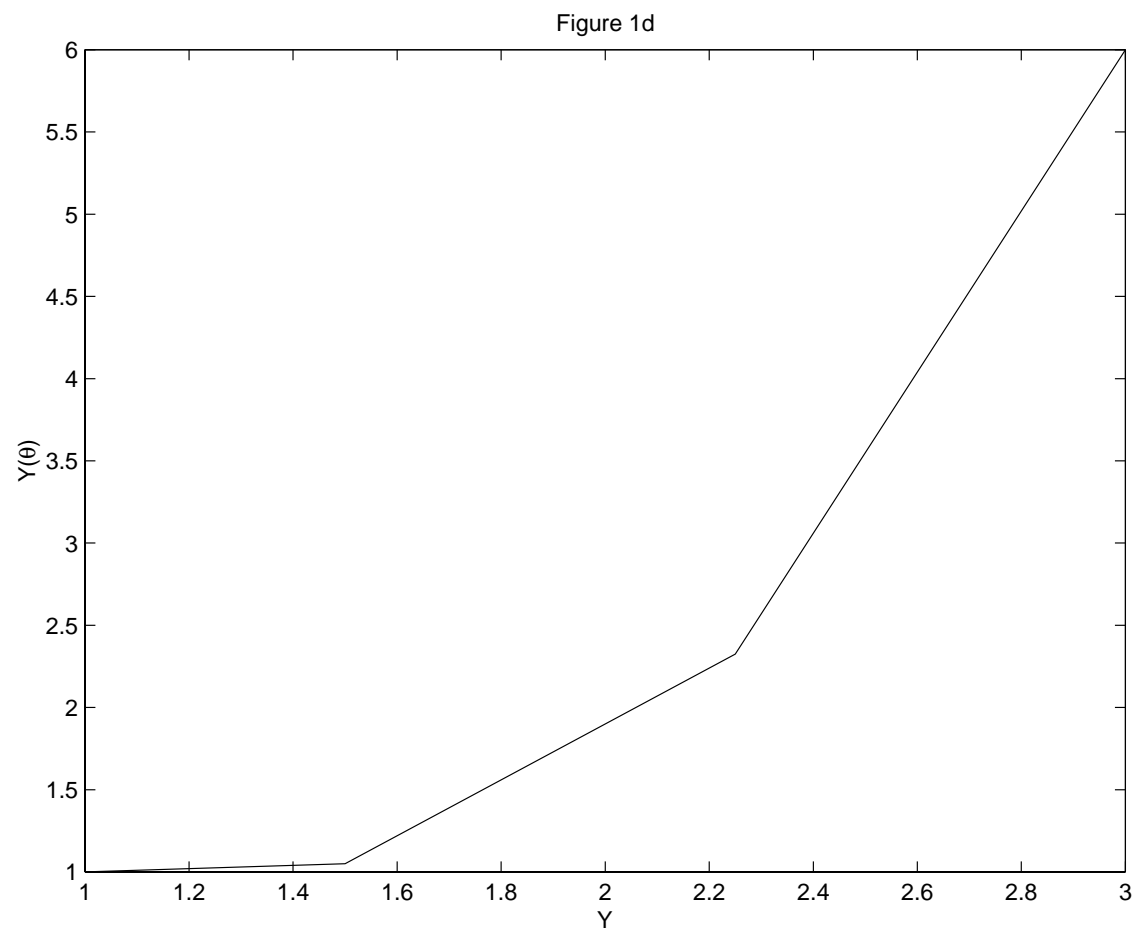


Figure 1b







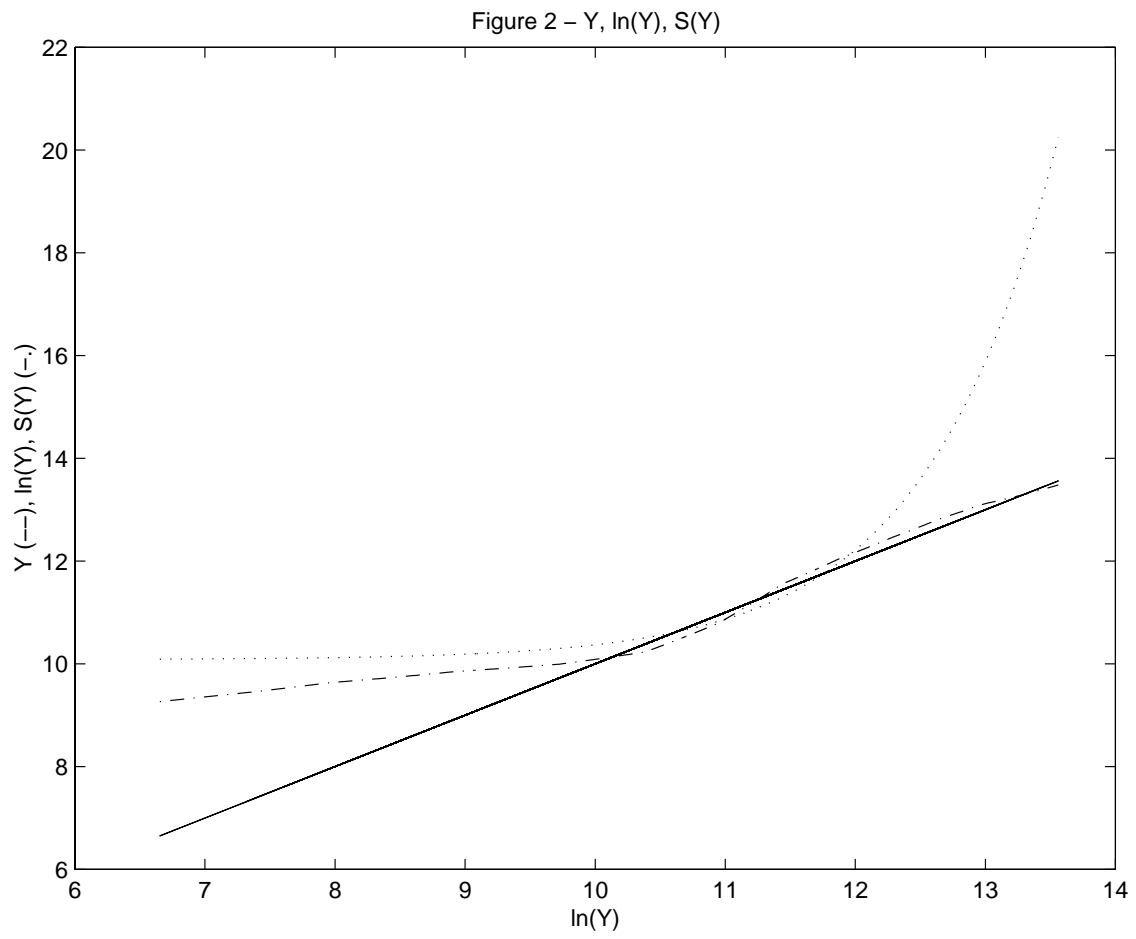


Figure 3a – Predictions v S(Y)

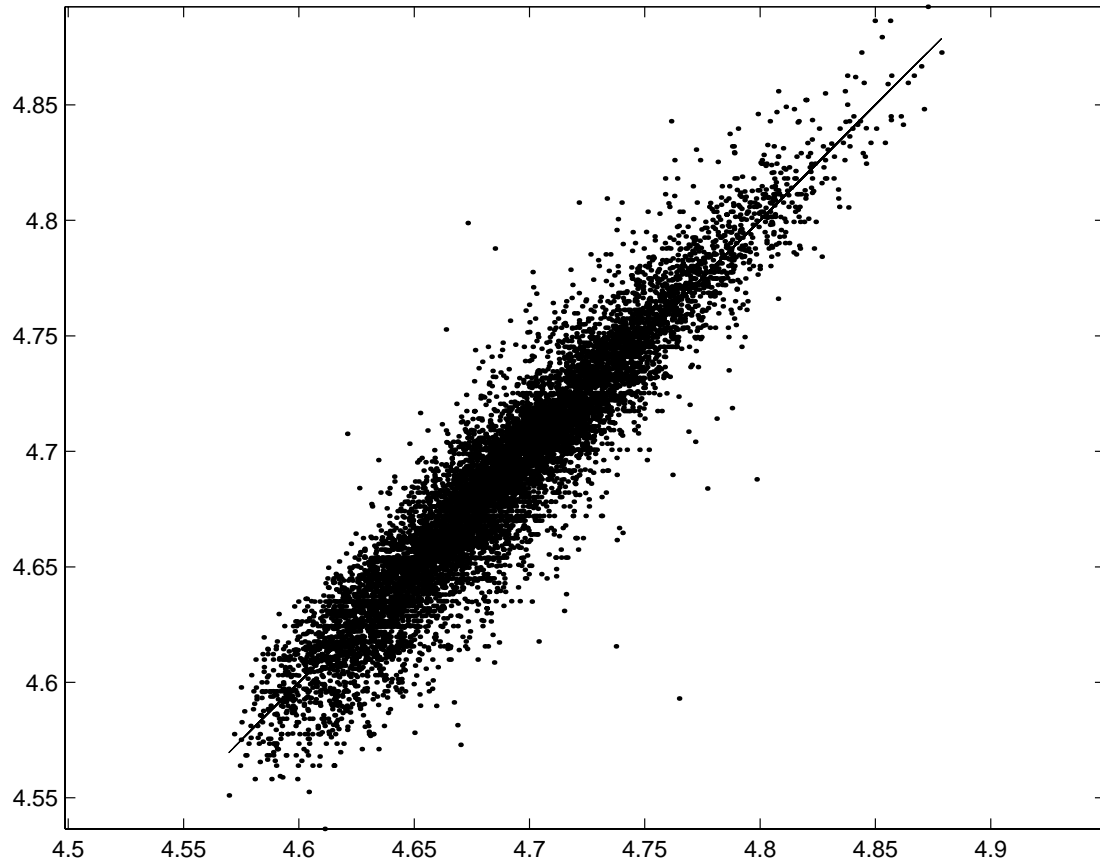
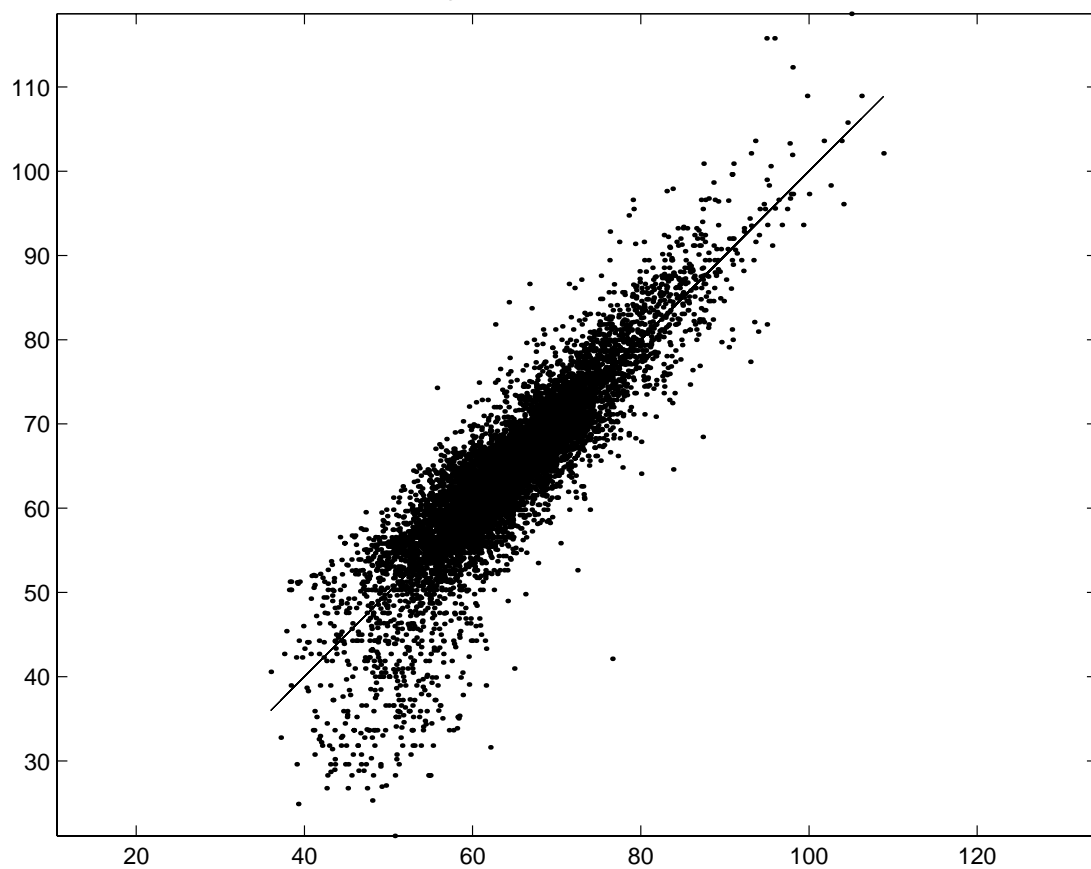


Figure 3b – Predictions v $Y^{1/4}$



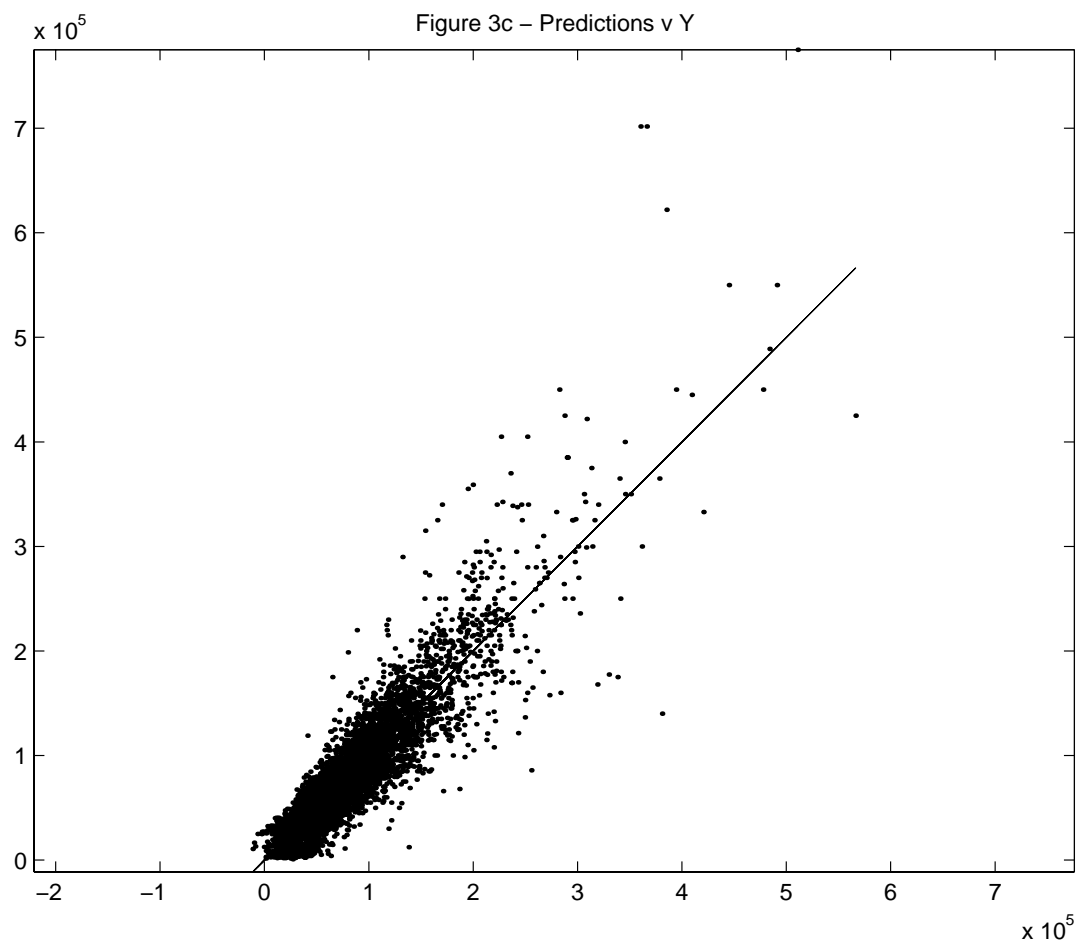


Figure 3d – Predictions v $\ln(Y)$

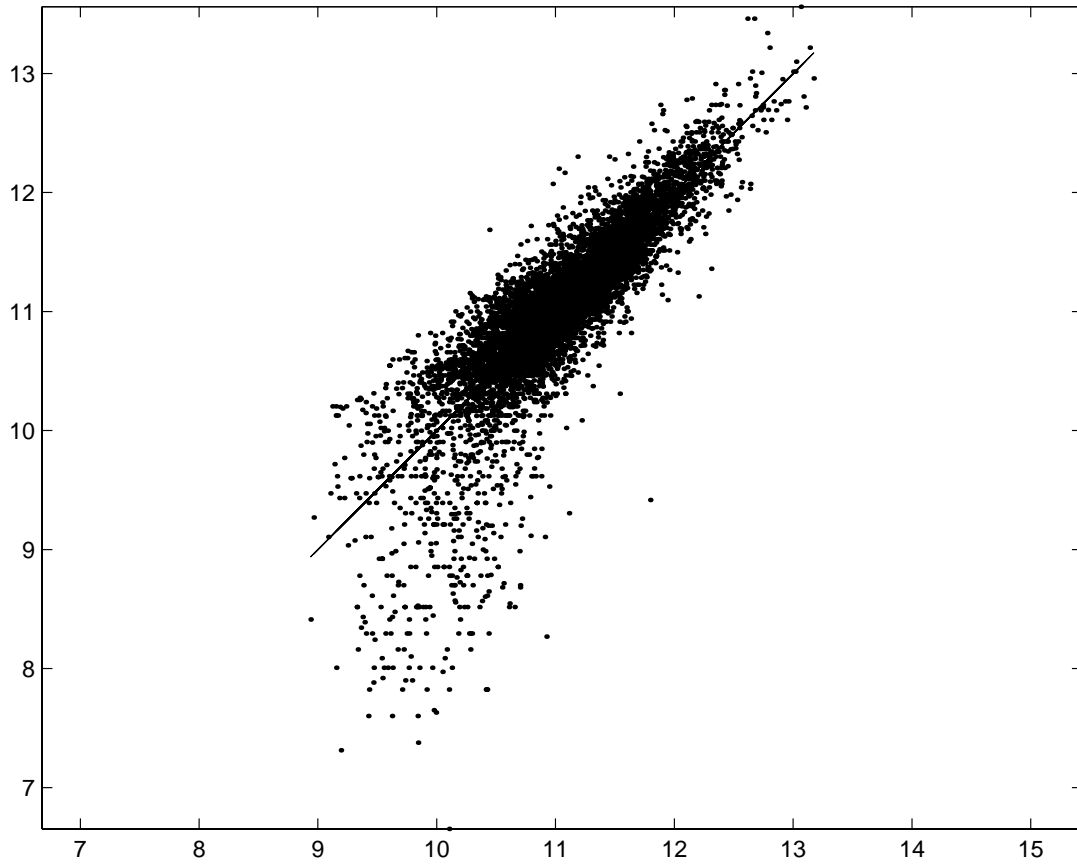


Figure 4a – Histogram of Spatial Regression Errors on Transformed Y

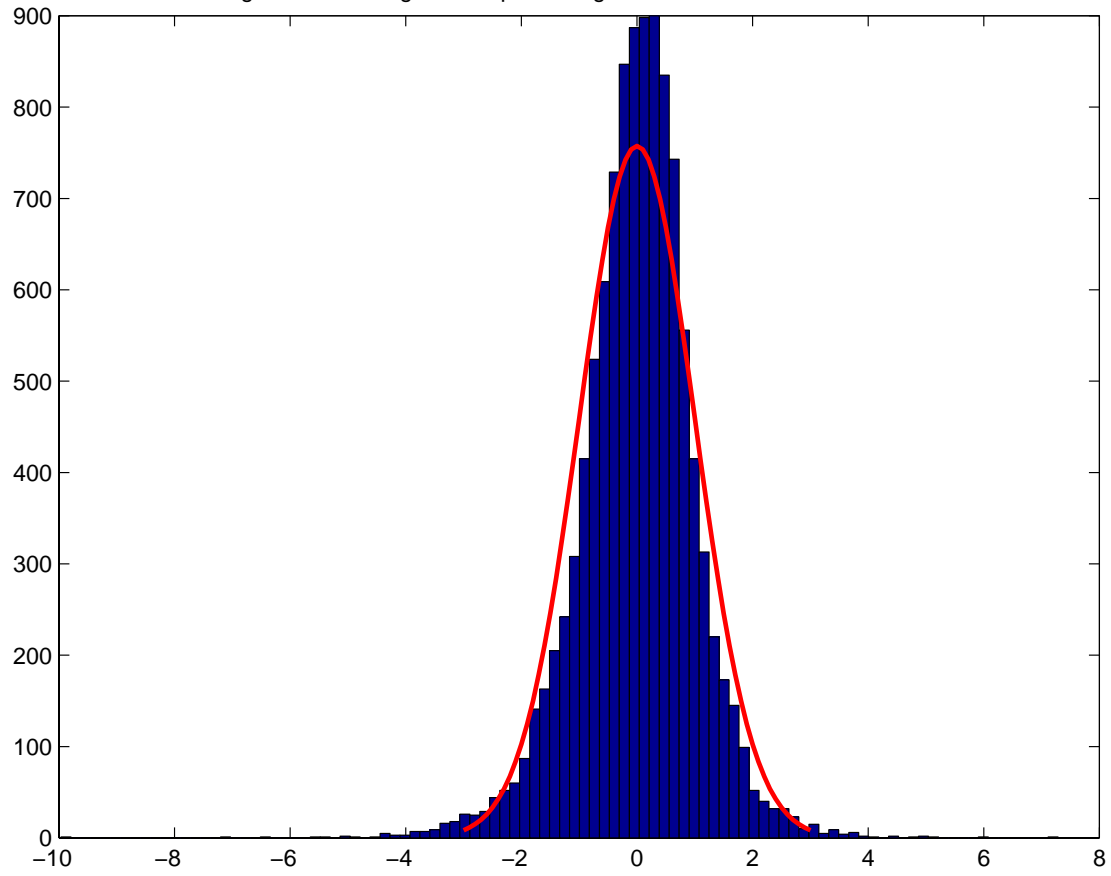


Figure 4b – Histogram of Spatial Regression Errors on Untransformed Y

