

Copyright
by
Xiaokun Wang
2007

**The Dissertation Committee for Xiaokun Wang certifies that this is the
approved version of the following dissertation:**

**CAPTURING PATTERNS OF SPATIAL AND TEMPORAL
AUTOCORRELATION IN ORDERED RESPONSE DATA:
A CASE STUDY OF LAND USE AND AIR QUALITY
CHANGES IN AUSTIN, TEXAS**

Committee:

Kara Kockelman, Supervisor

Kelley A. Crews-Meyer

Stephen Donald

Steven T. Waller

Charles M. Walton

Howard M. Liljestrand

**CAPTURING PATTERNS OF SPATIAL AND TEMPORAL
AUTOCORRELATION IN ORDERED RESPONSE DATA:
A CASE STUDY OF LAND USE AND AIR QUALITY
CHANGES IN AUSTIN, TEXAS**

by

Xiaokun Wang, B.S.;M.S.

Dissertation

Presented to the Faculty of the Graduate School of
the University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2007

For Isabella

And Wenhan

ACKNOWLEDGEMENTS

I would like to thank Dr. Kara Kockelman. Without her help and support through the last four years, this dissertation would not have been possible. She has been a wonderful mentor and a great role model. I am deeply indebted for her guidance in both academic and personal lives. I also would like to thank the participation and advice of members of the preliminary and final dissertation committees: Drs. Kelley Crews-Meyer, Stephen Donald, Steven T. Waller, Charles M. Walton, Paul Wilson and Howard M. Liljestrang.

I am grateful to Dr. Kockelman's assistant, Annette Perrone, for her encouragement and immeasurable help with administrative and logistical matters. And I would like to extend a general note of thanks to all the members of U.T. Austin's transportation program, for their support during my years as a PhD student. A special thanks goes to Zheng Li and Jianming Ma, for their friendship and help through the past four years.

I also want to thank my parents, for always giving me unconditional love. And I would like to thank my new daughter, Isabella, who was born just a month before this dissertation was completed. This "twin birth" of sorts has given me enormous courage and strength during the toughest nine months of my life.

Lastly, and most importantly, I certainly must acknowledge the love of my life: my husband, Wenhan Hu. Many changes have occurred in my life over the past five years. Throughout it all, he has been at my side, warming me with his understanding, encouragement and support. I thank him, for his undying love.

**CAPTURING PATTERNS OF SPATIAL AND TEMPORAL
AUTOCORRELATION IN ORDERED RESPONSE DATA:
A CASE STUDY OF LAND USE AND AIR QUALITY
CHANGES IN AUSTIN, TEXAS**

Publication No. _____

Xiaokun Wang, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Kara Kockelman

Many databases involve ordered discrete responses in a temporal and spatial context, including, for example, land development intensity levels, vehicle ownership, and pavement conditions. An appreciation of such behaviors requires rigorous statistical methods, recognizing spatial effects and dynamic processes.

This dissertation develops a dynamic spatial ordered probit (DSOP) model in order to capture patterns of spatial and temporal autocorrelation in ordered categorical response data. This model is estimated in a Bayesian framework using Gibbs sampling and data augmentation, in order to generate all autocorrelated latent variables. The specifications, methodologies, and applications undertaken

here advance the field of spatial econometrics while enhancing our understanding of land use and air quality changes.

The proposed DSOP model incorporates spatial effects in an ordered probit model by allowing for inter-regional spatial interactions and heteroskedasticity, along with random effects across regions (where “region” describes any cluster of observational units). The model assumes an autoregressive, AR(1), process across latent response values, thereby recognizing time-series dynamics in panel data sets.

The model code and estimation approach is first tested on simulated data sets, in order to reproduce known parameter values and provide insights into estimation performance. Root mean squared errors (RMSE) are used to evaluate the accuracy of estimates, and the deviance information criterion (DIC) is used for model comparisons. It is found that the DSOP model yields much more accurate estimates than standard, non-spatial techniques. As for model selection, even considering the penalty for using more parameters, the DSOP model is clearly preferred to standard OP, dynamic OP and spatial OP models.

The model and methods are then used to analyze both land use and air quality (ozone) dynamics in Austin, Texas. In analyzing Austin’s land use intensity patterns over a 4-point panel, the observational units are 300 m × 300 m grid cells derived from satellite images (at 30 m resolution). The sample contains 2,771 such grid cells, spread among 57 clusters (zip code regions), covering about 10% of the overall study area. In this analysis, temporal and spatial autocorrelation effects are found to be significantly positive. In addition, increases in travel times to the region’s central business district (CBD) are estimated to substantially reduce land development intensity.

The observational units for the ozone variation analysis are $4 \text{ km} \times 4 \text{ km}$ grid cells, and all 132 observations falling in the study area are used. While variations in ozone concentration levels are found to exhibit strong patterns of temporal autocorrelation, they appear strikingly random in a spatial context (after controlling for local land cover, transportation, and temperature conditions). While transportation and land cover conditions appear to influence ozone levels, their effects are not as instantaneous, nor as practically significant as the impact of temperature.

The proposed and tested DSOP model is felt to be a significant contribution to the field of spatial econometrics, where binary applications (for discrete response data) have been seen as the cutting edge. The Bayesian framework and Gibbs sampling techniques used here permit such complexity, in world of two-dimensional autocorrelation.

TABLE OF CONTENTS

Chapter 1. Introduction and Motivation	1
1.1 Overview and Motivation	1
1.2 Models of Land Development Intensity	3
1.3 Issues In Land Development Data: Data Availability And Definition Of Observational Units	6
1.4 Ozone Concentration Models	8
1.5 Study Objectives	10
1.6 Organization.....	11
1.7 Chapter Summary	12
Chapter 2. Literature Review	13
2.1 Spatial Econometrics	13
2.2 Models of Discrete Responses.....	19
2.3 Spatial Econometrics for Models of Discrete Response.....	20
2.4 The Bayesian Approach to Model Estimation.....	23
2.4.1 Bayesian Theory	24
2.4.2 Gibbs Sampling	25
2.4.3 Data Augmentation.....	26
2.5 Bayesian Approach for Dynamic Spatial Ordered Categorical Models ...	26
2.6 Chapter Summary	27
Chapter 3. Specification of a Dynamic Spatial Ordered Probit (DSOP) Model, for Bayesian analysis	28
3.1 Introduction.....	28
3.2 Model Specification.....	28
3.2.1 Standard Ordered Probit (OP) Model	28
3.2.2 Spatial Ordered Probit (SOP) Model.....	29
3.2.3 Dynamics to the Spatial Ordered Probit Model.....	32
3.2.4 Specification of the Dynamic Spatial Ordered Probit (DSOP) Model.....	34
3.3 Parameter Estimation via MCMC Simulation.....	36
3.3.1 Conditional Distribution	36
3.3.2 A Summary of Conditional Posterior Distributions	37
3.3.3 Prior Distributions for All Parameters	39
3.3.4 Full Conditional Posterior Distributions.....	42
3.3.4.1 Conditional Posterior Distribution of β	42

3.3.4.2 Conditional Posterior Distribution of θ	44
3.3.4.3 Conditional Posterior Distribution of λ	45
3.3.4.4 Conditional Posterior Distribution of ρ	45
3.3.4.5 Conditional Posterior Distribution of σ^2	46
3.3.4.6 Conditional Posterior Distribution of V	47
3.3.4.7 Conditional Posterior Distribution of γ	49
3.3.4.8 Conditional Posterior Distribution of U_0	50
3.3.4.9 Conditional Posterior Distribution of U	51
3.3.5 Summary of Parameter Distributions	53
3.4 MCMC Sampling.....	55
3.5 Discussion of Special Cases	58
3.5.1 Case of No Temporal Autocorrelation.....	58
3.5.2 Homoskedastic Case.....	58
3.5.3 Single Individual in Each Region.....	59
3.6 Chapter Summary	59
Chapter 4. Model Validation and ComparisonS	61
4.1 Introduction.....	61
4.2 Simulated Dataset.....	61
4.3 Model Estimation and Validation.....	64
4.4 Model Comparisons.....	69
4.5 Chapter Summary	76
Chapter 5. Data Description	77
5.1 Introduction.....	77
5.2 Land Development Intensity.....	77
5.2.1 Land Development Intensity Level.....	78
5.2.1.1 Land Cover Information Derived from Satellite Images.....	78
5.2.1.2 Uncertainty and Aggregation of Land Cover Information ...	80
5.2.1.3 Categorizing Land Development Intensity Levels	82
5.2.2 Census Data	84
5.2.3 Transportation Access.....	88
5.2.3.1 Road Networks in Different Years	89
5.2.3.2 Calculating “Travel Cost”.....	91
5.2.3.3 Facilities and Important Locations	94
5.2.3.4 Calculating Travel Time for Each Location	95
5.2.4 School Access.....	96
5.2.5 Topographic Information.....	97
5.2.6 Defining Regions and Sample Selection	99
5.2.7 Summary of Land Development Data	102
5.3 Ozone Concentration Levels.....	104

5.3.1 1999 Ozone Concentration Levels.....	104
5.3.2 Temperature Distribution.....	107
5.3.3 Transportation and Land Cover Information	109
5.3.4 Summary of Ozone Model Data	110
5.4 Chapter Summary	111
Chapter 6. Results of data Analysis	112
6.1 Introduction.....	112
6.2 Estimation of Land Development Intensity Levels	112
6.2.1 Estimation Results	112
6.2.2 Model Comparisons.....	122
6.2.3 Marginal Effects	124
6.2.4 Development Intensity Model Prediction.....	128
6.3 Estimation of Ozone Concentration Levels.....	131
6.3.1 Estimation Results	131
6.3.2 Marginal Effects	138
6.3.3 Ozone Model Prediction	140
6.4 Chapter Summary	141
Chapter 7. Conclusions	143
7.1 Summary.....	143
7.2 Limitations and Extensions.....	146
7.3 Concluding Remarks.....	150
References.....	153
Appendix A. Derivation of Hyperparameters for β's Conditional Posterior Distribution	164
Appendix B. Derivation of Hyperparameters for U_{ikt}'s Conditional Posterior Distribution.....	166
Appendix C. Matlab Code for the core component of The DSOP Model	167
Vita	173

LIST OF TABLES

Table 3.1 Conditional Distributions of DSOP’s Parameters	54
Table 4.1 Estimation Results using Simulated Data (Averages from 50 Samples)	67
Table 4.2 Summary Statistics for One Sample	69
Table 4.3 Estimation Results using One Sample and Different Specifications....	71
Table 4.4 Prediction Rates using Different OP Model Specifications	73
Table 5.1 Estimation Results for Road Existence (Binary Probit)	90
Table 5.2 Data Description for Land Development Intensity Level Analysis....	102
Table 5.3 Summary Statistics for Land Development Intensity Analysis	103
Table 5.4 Frequency of Ozone Concentration Level through a Day	107
Table 5.5 Frequency of Temperature through a Day	108
Table 5.6 Data Description for Ozone Concentration Level Analysis	111
Table 6.1 Estimation Results for Model of Land Development Intensity Levels	115
Table 6.2 Correlations between Parameter Estimates	123
Table 6.3 Goodness of Fit and Prediction Rates using Different OP Model Specifications.....	124
Table 6.4 Average Marginal Effects of Covariates on Development Intensity Levels over All Observations	127
Table 6.5 Marginal Effects of Covariates on Development Intensity Levels (Using One Observation).....	128
Table 6.6 Comparison of Base Year and Predicted Land Development Intensity Levels.....	131
Table 6.7 Estimation Results for Model of Ozone Concentration Levels	134
Table 6.8 Marginal Effects of Covariates on Ozone Concentration Levels over All Observations	138

LIST OF FIGURES

Figure 3.1 Flowchart for the MCMC Simulation	57
Figure 4.1 Location of Regions in Simulated Dataset	62
Figure 4.2 Assumed Variances of Individual Specific Effects across Regions....	63
Figure 4.3 Variances of Individual Specific Errors with 50 Samples	68
Figure 4.4 Histogram of Dependent Variable Values.....	70
Figure 4.5 Estimation Convergence Pattern for β_1	70
Figure 4.6 Variances of Individual Specific Errors with One Sample	71
Figure 4.7 Posterior Distributions for Parameters of Primary Interest.....	75
Figure 5.1 Original Land Cover Information for Year 1983	80
Figure 5.2 Computed Land Intensity Levels across Different Years	84
Figure 5.3 Neighborhood Populations as Shown at the level of 300m Grid Cells (where neighborhood is a 3km-radius circle)	86
Figure 5.4 Neighborhood Workers as Shown at the level of 300m Grid Cells (where neighborhood is a 3km-radius circle)	87
Figure 5.5 Average Median Household Income across Neighborhoods (where neighborhood is a 3km-radius circle)	88
Figure 5.6 Estimates of Austin Road Network in Different Years.....	91
Figure 5.7 Procedure for Deriving Estimates of Travel Time	93
Figure 5.8 Locations of Key Sites and Highways in Different Years	95
Figure 5.9 Number of Schools in Each Cell's Neighborhood	96
Figure 5.10 Elevation Distribution	98
Figure 5.11 Slope Distribution.....	99
Figure 5.12 Selected Sample for Development Intensity Analysis	101
Figure 5.13 Ozone Concentration Values and Corresponding Levels (4 to 5pm on Monday, September 13, 1999).....	106
Figure 6.1 Convergence Patterns of Development Intensity Level Estimation..	114
Figure 6.2 Distribution of Region-Specific Error Term Estimates (θ) for Land Development Intensity Levels	117
Figure 6.3 Distribution of the Variances of Individual Specific Error Term Estimates (ν) for Land Development Intensity Levels.....	118
Figure 6.4 Posterior Distributions of Land Development Intensity Level Model Parameters.....	121
Figure 6.5 Most Likely Development Intensity Levels Prediction and Uncertainty (following an assumed doubling of population).....	130

Figure 6.6 Convergence Patterns of Ozone Concentration Level Parameter Estimation	132
Figure 6.7 Distribution of Regional-Specific Error Term Estimates (θ) for Ozone Concentration Levels	135
Figure 6.8 Posterior Distributions of Ozone Concentration Level Model Parameters.....	137
Figure 6.9 Prediction and Comparison of Ozone Concentration Levels	142

NOTATION

In this dissertation, bold characters indicate vectors or matrixes, while regular characters are for scalars. The following summarizes the notation used in this dissertation, in order of appearance.

i : Index for regions, $i \in 1, 2, \dots, M$

k : Index for individuals, $k \in 1, 2, \dots, n_i$, and $\sum_{i=1}^M n_i = N$

t : Index for time periods, $t \in 1, 2, \dots, T$

s : Index for ranks or levels, $s \in 1, 2, \dots, S$

r : Index for iterations, $r \in 1, 2, \dots, R$

y_{ikt} : Observed dependent variable for individual k in region i at time t

\mathbf{y} : Vector of y_{ikt} , stacked first by region, then by time

\mathbf{X}_{ikt} : Vector of explanatory variables for individual k in region i at time t

\mathbf{X} : Matrix of explanatory variables, stacked first by region, then by time

U_{ikt} : Latent variable for individual k in region i at time t

\mathbf{U} : Vector of latent variables, stacked first by region, then by time

$\mathbf{U}^\lambda = \mathbf{U} - \lambda \mathbf{U}_{-1}$: Vector of temporally differentiated latent variables

U_{ik0} : Initial value of latent variable for individual k in region i

\mathbf{U}_0 : Vector of initial values

$\boldsymbol{\beta}$: Vector of unknown parameters corresponding to \mathbf{X}

λ : Temporal autocorrelation coefficient

ρ : Spatial autocorrelation coefficient

\mathbf{W} : The row-standardized weight matrix with elements w_{ij}

ς_{\min} : Minimum eigenvalue of \mathbf{W}

ς_{\max} : Maximum eigenvalue of \mathbf{W}

$\theta_i = \rho \sum_{j=1}^M w_{ij} \theta_j + u_i$: regional effect for region i

u_i : Regional effect that is not caused by specified spatial factors

σ^2 : Variance of $u_i, \forall i \in 1, 2, \dots, M$

\mathbf{u} : Vector of u_i ,

\mathbf{I}_M : Identity matrix of dimension M

$\boldsymbol{\theta} = (\mathbf{I}_M - \rho \mathbf{W})^{-1} \mathbf{u}$: Vector of regional effects.

$\mathbf{B}_\rho = \mathbf{I}_M - \rho \mathbf{W}$: Matrix representing regional dependencies

ε_{ik} : Individual-specific error term for individual k in region i

$\boldsymbol{\varepsilon}_i$: Vector of ε_{ik}

$\boldsymbol{\varepsilon}$: Vector of $\boldsymbol{\varepsilon}_i$

ν_i : Variance for $\varepsilon_{ik}, \forall k \in 1, 2, \dots, n_i$

$$\mathbf{V} = \begin{bmatrix} \nu_1 \mathbf{I}_{n_1} & & & \mathbf{0} \\ & \nu_2 \mathbf{I}_{n_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \nu_M \mathbf{I}_{n_M} \end{bmatrix}$$
 : Variance matrix for $\boldsymbol{\varepsilon}_i, \forall i \in 1, 2, \dots, T$

$\boldsymbol{\Omega} = \mathbf{I}_T \otimes \mathbf{V}$: Variance matrix for $\boldsymbol{\varepsilon}$

\mathbf{I}_T : Identity matrix of dimension T

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_{n_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{I}_{n_M} \end{bmatrix}$$
 : $N \times M$ matrix with each \mathbf{I}_{n_i} being a $n_i \times 1$ vector of ones.

$\boldsymbol{\Delta} = \mathbf{I}_T \otimes \mathbf{L}$: $NT \times M$ matrix

$\mathbf{e} = \mathbf{U}^\lambda - \boldsymbol{\Delta} \boldsymbol{\theta} - \mathbf{X} \boldsymbol{\beta}$: Vector of estimated individual-specific errors

$\xi_{ik} = \theta_i + \varepsilon_{ik}$: Total error for individual k in region i

$\boldsymbol{\gamma}$: Vector of threshold parameters (for ordered probit specification)

γ_s : Element of $\boldsymbol{\gamma}$, $\gamma_0 = -\infty$, $\gamma_S = +\infty$, and $\gamma_1 < \gamma_2 < \dots < \gamma_{S-1}$
 $\delta(A)$: Indicator function (1 if A is true and 0 otherwise)
 $\pi(*)$: Prior distribution of the random parameter $*$
 Θ_* : Set of arguments that the random parameter $*$ conditioned on
 $p(*|\Theta_*)$: Conditional posterior distribution of the random parameter $*$.
 \mathbf{c} : Vector of $\boldsymbol{\beta}$'s prior means
 $\mathbf{H} = h\mathbf{I}_Q$: $\boldsymbol{\beta}$'s prior variance matrix
 \mathbf{I}_Q : Identity matrix of dimension Q
 λ_0 : Hyperparameter for λ 's prior distribution (mean)
 D : Hyperparameter for λ 's prior distribution (variance)
 α : Hyperparameter for σ^2 's prior distribution (shape parameter)
 τ : Hyperparameter for σ^2 's prior distribution (scale parameter)
 ϖ : Hyperparameter for ν_i 's prior distribution
 \mathbf{q} : Vector of $\boldsymbol{\gamma}$'s prior means, with elements γ_{s0} .
 \mathbf{G} : Diagonal matrix of $\boldsymbol{\gamma}$'s prior variance, with diagonal elements g_s
 a_0 : Hyperparameter for U_{ik0} 's prior distribution (mean),
 $\forall i \in 1, 2, \dots, M, \forall k \in 1, 2, \dots, n_i$
 d_0 : Hyperparameter for U_{ik0} 's prior distribution (variance),
 $\forall i \in 1, 2, \dots, M, \forall k \in 1, 2, \dots, n_i$
 \mathbf{A} : Hyperparameter for $\boldsymbol{\beta}$'s posterior distribution (inverse of its variance matrix)
 \mathbf{b} : Hyperparameter for $\boldsymbol{\beta}$'s posterior distribution ($\boldsymbol{\beta}$'s mean is $\mathbf{A}^{-1}\mathbf{b}$)
 A_λ : Hyperparameter for λ 's posterior distribution (inverse of its variance)
 b_λ : Hyperparameter for λ 's posterior distribution (λ 's mean is $A_\lambda^{-1}b_\lambda$)
 \mathbf{A}_θ : Hyperparameter for $\boldsymbol{\theta}$'s posterior distribution (inverse of its variance matrix)

- \mathbf{b}_θ : Hyperparameter for $\boldsymbol{\theta}$'s posterior distribution ($\boldsymbol{\theta}$'s mean is $\mathbf{A}_\theta^{-1}\mathbf{b}_\theta$)
- γ_s^{inf} : Hyperparameter for γ_s 's posterior density, left threshold for truncation
- γ_s^{sup} : Hyperparameter for γ_s 's posterior density, right threshold for truncation
- A_{U_0} : Hyperparameter for U_{ik_0} 's posterior distribution (inverse of its variance)
- b_{U_0} : Hyperparameter for U_{ik_0} 's posterior distribution (U_{ik_0} 's mean is $A_{U_0}^{-1}b_{U_0}$)
- a_{ikt} : Hyperparameter for U_{ikt} 's posterior distribution (mean),
 $\forall i \in 1, 2, \dots, M, \forall k \in 1, 2, \dots, n_i, \forall t \in 1, 2, \dots, T$
- b_{ikt} : Hyperparameter for U_{ikt} 's posterior distribution (variance),
 $\forall i \in 1, 2, \dots, M, \forall k \in 1, 2, \dots, n_i, \forall t \in 1, 2, \dots, T-1$

CHAPTER 1. INTRODUCTION AND MOTIVATION

1.1 OVERVIEW AND MOTIVATION

In transportation-related studies, the variables of interest often are discrete in nature and involve temporal and spatial relationships. For example, travel mode choices, trade flow distributions, vehicle ownership and pavement surface deterioration levels can all be measured (and/or coded) as discrete responses, dependent on various influential factors. These discrete responses share a common feature: they all exhibit some degree of temporal and spatial dependence or autocorrelation. For example, in two slices of a panel survey of households, the count of vehicles owned by the same household will be highly correlated. This phenomenon is normally defined as temporal dependency or autocorrelation. Meanwhile, even after controlling for household attributes, auto ownership levels are expected to exhibit positive correlations in the spatial context. To some extent, such correlation patterns can be explained by uncertainty or proximity because, in reality, there are always influential factors that cannot be controlled (e.g. pedestrian friendliness of all neighborhoods). The sign and magnitude of such uncertainties tend to vary rather gradually over space. Of course in a spatial context, in contrast to time-series data, such dependencies are *two* dimensional – which adds complexity. Like temporal relationships, correlation tends to diminish with increases in distance between any two households/observed units.

In studies of social behaviors and human activities, many choices or attributes (e.g., religious beliefs, presidential election outcomes, and levels of crime) involve discrete responses in a temporal and spatial context. The widespread nature of such phenomena and a need for understanding these behaviors compel the quest for rigorous statistical methods for analysis of such data. No currently existing statistical methods fully meet this need.

As part of a long-term research framework aimed at establishing an integrated land use-transport and environmental model, this dissertation selects two distinct datasets to illustrate the temporal and spatial dependencies that exist. The first features dynamics in land development intensity levels under the influence of geology, demographics, transportation conditions and other, socio-economic factors. The second tracks ozone concentration levels and their dependence on local land use, traffic intensity and air temperature.

Both of these analyzes rely on Austin, Texas data sets. Of course, one reason for choosing this area is the convenience of data acquisition. More importantly, the traffic and air quality conditions in the study area make it an interesting case study: thanks to rapid population growth and economic expansion, the area has experienced some dramatic changes during the last two decades. As will be shown in more detail in Chapter 5, during this time period, the region's land development has both sprawled over space and escalated in intensity. One direct result of this development is congestion. The Texas Transportation Institute's urban mobility report (Schrank and Lomax, 2005) indicates that Austin ranks number 1 among all 30 medium-sized U.S. cities in its annual study.

Meanwhile, the area's air quality has been deteriorating. According to the U.S. Environmental Protection Agency (EPA, 2003), the three-county area (including Williamson, Travis, and Hays Counties) are very close to being designated as non-attainment for national ambient air quality standards.¹ The situation may get worse if no actions are taken. Findings from this dissertation may help remedy the current situation.

In the first model, land development intensity is defined based on how much land is covered by manmade materials, which are characterized by higher reflectance

¹ EPA designates an area as non-attainment if it has violated, or has contributed to violations of the national 8-hour ozone standard over a three-year period. (EPA, 2006)

levels and other visual clues provided via satellite images. (In fact, the application of satellite images is another major motivation for this study, as will be discussed in Section 1.3.) These “intensity levels” are indexed as integers, and their order is key. It should also be noted that the ozone concentration information used in this study is derived from the continuous values generated by the CAMx projection model (CAPCO et al., 2004). The actual observed data for ozone concentration in the study area is, unfortunately, unavailable.² Therefore, though the study can still provide some reasonable insights into the area’s air quality problem, data quality issues should be expected.

The land development and ozone datasets are used as examples, to illustrate the specification, estimation and application of dynamic spatial ordered probit models. The following two sections describe the temporal, spatial and discrete nature of such data as well as limitations of current studies in this area, highlighting the need for more rigorous statistical analysis methods and the potential of satellite images as an innovative data source.

1.2 MODELS OF LAND DEVELOPMENT INTENSITY

Modeling land development intensity levels illustrates the temporal, spatial and discrete nature of land development data. For urban areas, the evolution of land development intensity is a topic of interest to traffic demand modelers, policy makers, and land developers. Such changes influence regional economies and environmental conditions. For non-urban areas, analyzing the dynamics of land development intensity is also important: For example, undeveloped land around the world, including some precious lands like the Amazon rainforest, are being converted for agriculture and other human uses. Such changes can significantly

² In practice, the number of measuring sites is very limited. As will be discussed in Chapter 7, the study area has only two such sites in 1999 and seven in 2007.

contribute to climate change, desertification, resource depletion and loss of habitats and species.

In this dissertation, land development intensity is derived based on land cover information in the each 300 m x 300 m neighborhood. However, as discussed in Chapter 5, “land cover information” in reality is often intertwined with land use information. Therefore, before beginning a formal discussion of land development intensity, definitions of “land use” and “land cover” terms are provided, and their differences discussed.

“Land use” normally refers to what the land is actually used for. For example, residential, commercial, transportation or industrial uses connote uses like homes versus businesses, roads versus offices, and manufacturing plants versus airports. In contrast, “land cover” emphasizes how the land “looks,” rather than how it is being used. For example, whether land is covered by water, brush, forests, cement and soil is key information for a variety of environmental studies, including flood control and deforestation. Land use and land cover information is important for studies of land development. For this reason, the two terms are often used in tandem, as “land use/land cover”.

In a review of existing land use/land cover models, Parker et al. (2003) concluded that no single approach yet “dominates this nascent field.” However, some approaches are more common than others and already have been used in the integrated models of land use and transportation. Examples include de la Barra’s (1989) TRANUS, Waddell’s (2002) UrbanSim and the Hunt and Abraham’s (2003) PECAS model. The land development modules within these packages rely on standard, MNL specifications to handle the discrete nature of land use type. TRANUS and PECAS are both aggregate models with zones (typically census tracts or traffic analysis zones) as their observational units. UrbanSim can use grid cells or (in version 4) user-defined zones as its observational units.

None of these models recognizes spatial or temporal autocorrelations in a statistically rigorous manner. TRANUS and PECAS simply recognize spatial effects by controlling for results from a spatial input-output model. And UrbanSim controls for regional, neighborhood and cell characteristics to diminish spatial effects. The cellular automata-based SLEUTH model (Candau et al., 2000) has received some attention. It consists of Clarke et al.'s (1996, 1997) Urban Growth Model (UGM) and Deltatron Land Use/Land Cover Model. Though SLEUTH simulates land cover change recognizing the temporal and spatial context of each cell, it can only share information across immediate cells, so that more dispersed interactions and correlations are largely ignored. In addition, SLEUTH is not (yet) designed to flexibly accommodate the effects of many influential and related human factors (such as land prices, employment and population density.) Finally, it relies on rule-driven algorithms, rather than more behavioral or statistical models.

In fact, many studies that acknowledge the presence of spatial effects have tried to remove all spatial correlation – by either controlling for a variety of neighborhood attributes or using strategic sampling (to provide a dispersed sample, with minimal interactions). Some also attempt to recognize temporal dependencies by controlling for variables from previous periods. For example, Nelson and Hellerstein (1997) sampled selectively and created exogenous variables based on neighboring units' land cover data in order to study the deforestation effects of roadways via a multinomial logit model. Wear and Bolstad (1998) controlled for prior land uses in the neighborhood of each data cell in their study of southern Appalachian landscapes, which involved binary response data. Munroe et al. (2001) attempted to filter out spatial correlations through sampling and then removed the residual spatial dependence through a “trend surface” approach (Cliff and Ord, 1981). As with all other existing models dealing with discrete response

data in a temporal and/or spatial context, the applicability of these methods is still limited because of the neglect of spatial effects (even intentionally) and data dynamics. These other works are discussed in depth in Chapter 2.

1.3 ISSUES IN LAND DEVELOPMENT DATA: DATA AVAILABILITY AND DEFINITION OF OBSERVATIONAL UNITS

In order to better understand land development, practitioners have been looking for more accessible and disaggregate datasets. Municipal land use data sets emphasize parcel geometry, for purposes of tax assessment, utility provision, and so on. Such datasets can provide precise information on how the land is actually “used.” However, using parcels as spatial units may not be detailed enough for certain types of analyses, such as those focusing on vegetative and other species, crime occurrence, and so forth. In addition, “parcel” is not a very desirable unit for statistical studies when temporal relationships must be considered. Large, undeveloped parcels may sub-divide and/or experience partial redevelopment, which can make these hard to treat in a panel fashion.

Because of the ways such data are collected, reliable and timely parcel-based data is also hard to access. Current land use information is often only updated and released every five to ten years, if that. For some rapidly developed areas, this lack of updating frequency leads to the loss of very important information. More and more, practitioners, researchers, and the public are looking for alternative data resources that provide highly detailed, accessible, and low-cost information. And technological advances are paving the way.

Developments in remote sensing via satellite provide such an opportunity. Generally, remote sensing works on the principle of “the inverse problem” (Aster et al, 2004), which means transforming data set information into model

parameters. The spatial units detected by satellite images can be very small. These often are selected at a 30m x 30m resolution, but can be scaled down to 1ft x 1 ft. In addition, development in satellite image acquisition and classification techniques means that such data may be accessed more easily and at a lower cost. Moreover, the derivation of multiple-year data also allows direct incorporation of temporal and spatial correlation into a model's specification.

Another advantage of satellite data is that it offers much more precise information on vegetation, which can be critical to air quality, due to biogenic sources of (and sinks for) various chemicals of interest. (More details on this are provided in Section 1.4.) All these advances and aspirations suggest that using satellite data may ultimately be the optimal choice for integrated land use-transport-environment (ILUTE) models, which are highly valued in many regions in order to demonstrate compliance with air quality-related planning standards.

The acquisition of satellite data, however, is a complex process, full of potential errors and omissions. Sophistication and expertise are needed to extract information from multiple spectra, correct this information radiometrically and atmospherically, and finally, interpret it³. While a detailed discussion of these issues is not the focus of this dissertation, data quality issues associated with the use of remote sensing data should be noted. For example, Foody (2002), Townshend (1992), and Wilkinson (1996) describe various types of uncertainty inherent in satellite data and discuss data-quality limitations. Ideally, such uncertainty should be quantified. Typically, however, information on data quality is wholly missing (see, e.g., Johnston and Timlin, 2000). The presence of measurement errors in the data examined here is simply the nature of the beast. (Even carefully collected household surveys on travel behavior and the like

³ Some classic works that provide technical details for such data processing include those by Campbell (2002), Jensen (1996 and 2007), Lillesand et al. (2003), and Richards and Jia (2006).

contain mistakes, as humans forget where they have been, what they have purchased and so on.) In reality, very few – if any – data sets are truly flawless. In part for this reason, the model specification and estimation methods presented here can be viewed as the primary contribution of this dissertation.

1.4 OZONE CONCENTRATION MODELS

As a gas in the stratosphere that protects Earth from harmful ultraviolet rays, the ozone layer shields living things. However, in the troposphere, ozone is a powerful oxidizer and can damage lung tissue. Under the National Ambient Air Quality Standards (NAAQS), all Metropolitan Statistical Areas (MSAs) in the United States are required to develop strategies for attaining the standards and accommodate future growth. Thus, planners and policy makers must understand the spatial distribution of air pollutants, like ozone. Most predictive studies for ozone concentrations emphasize the photochemical process. Though such an approach is more behavioral in nature than purely statistical modeling, it can prove much more challenging to specify and does not offer information on the nature of unexplained variations in ozone concentrations.

Ozone concentration is usually expressed as a continuous value. For example, the California one-hour ozone standard is set at 0.09 parts per million (ppm) and the eight-hour average ozone standard is 0.070 ppm (BAAQMD, 2005). However, in order to illustrate these levels clearly for policy makers and the public, especially for the purpose of air quality forecasting, these values may be categorized into ordered categories (indexed, for example, as low, medium and high concentrations). (See, for example, Athanasiadis et al., 2007.)

Of course, many factors can and do influence ozone concentration levels through complex chemical and physical processes. Travel choices and land development patterns impact concentrations. For example, Niemeier et al. (2006) found that for most regions in the Northern Hemisphere, road traffic intensity is closely

associated with local ozone concentrations. They surmised that, if traffic-related emissions per capita in south Asia hit U.S. levels, that continent's surface ozone concentrations would increase by 50 to 100%. Wang et al. (2005) concluded that transportation sources are the main contributor to ozone concentrations, averaging roughly twice the effect of industrial emissions. Friedman et al. (2001) studied changes in commuting behaviors during the 1996 Summer Olympic Games in Atlanta and noted how decreased traffic densities were associated with a prolonged reduction in ozone pollution.

Land coverage development and intensity are also important determinants. And, of course, even if the land is not developed for human use, its features need to be classified for calculation of biogenic emissions. These are naturally occurring emissions from vegetation, which can be a strong function of tree type. For example, live oak trees are high emitters of isoprene, a highly reactive, volatile organic compound (VOC) that is a precursor to ozone. In areas such as eastern Texas, where this species is common, biogenic emissions of VOCs dominate the area's emissions inventory (Wiedinmyer, 1999). Another reason for requiring such land coverage information is the calculation of dry deposition rates. Dry deposition refers to the accumulation of particles and gases as they come into contact with soil, water or vegetation on the earth's surfaces. Allen (2002) suggests that during ozone season in Texas, dry deposition is the most important physical removal mechanism for air pollutants. Dry deposition rates for specific pollutants are typically computed according to land cover type. McDonald-Buller et al. (2001) investigated the sensitivity of dry deposition and ozone mixing ratios as a function of land cover classification and noted the importance of establishing accurate, internally consistent land cover data for air quality modeling. Thus, changes to both developed and undeveloped land cover type can significantly alter the magnitude spatial distribution of ozone.

Of course, many other factors also play a role. For example, Guldmann and Kim (2001) suggest that, in addition to land development and transportation characteristics, pollution measurements, meteorological factors and socioeconomic data can and do influence ozone concentrations. Loibl et al. (1994) show how relative altitude and time of day are influential. Pont and Fontan (2001) suggest that though local reduction in traffic is important, advection⁴ of ozone is also critical to its concentration.

Obviously, ignoring any of these relevant factors introduces uncertainty in model estimation and prediction. Such variables, if unobserved, can generate both temporal and spatial autocorrelations in model error terms. For example, meteorological factors (such as local wind speeds, rainfall, relative humidity, and temperature), precursors of ozone, and pollution control policies all exhibit positive temporal and spatial dependencies (see, for example, Lin, 2007, and Hancock, 1994). Therefore, it is reasonable to incorporate temporally and spatially lagged term and neighborhood effects in model specification.

In summary, ozone concentration levels are related to numerous factors. Among them, transportation conditions and land use/land cover information are critical. A statistically rigorous analysis of ozone concentration can be achieved with an ordered discrete choice model with a temporal lagged item and spatial autocorrelation in error terms.

1.5 STUDY OBJECTIVES

The first objective of this study is to develop a model that is appropriate for describing the temporal and spatial relationships that exist in ordered categorical data. Related issues also will be explored, indicating model estimation techniques, model validation and model comparisons (with simplified, less behaviorally

⁴ Advection refers to the transport of something from one region to another. Ozone's advection is predominantly horizontal, following weather system patterns.

reasonable models). Such model specifications and estimation techniques may be viewed as breakthroughs in the area of spatial econometrics. This dissertation's results can be extended to a wide range of topics, as long as the dependent variables are ordered discrete values and may involve temporal and spatial dependencies across observations.

A second objective is to develop a framework for interpreting, processing and applying remotely sensed (satellite imagery) data in land use and air quality analyses. This framework may offer researchers and planners much inspiration, by illuminating the potential of satellite databases.

A third objective is an understanding of how different factors affect land development decisions, especially the role of transportation conditions.

A final objective is a stronger understanding of the spatial patterns and dynamics of ozone concentrations for the environmental impact module in the integrated model. These results may facilitate policy making and the model specifications can be used in an integrated model of land use, transportation, and air quality.

1.6 ORGANIZATION

The rest of the dissertation is organized as follows. Chapter 2 conducts an extensive review of existing studies on related topics. Chapter 3 explains the intuition of model specification, incorporating a temporally lagged latent variable and a spatially autocorrelated regional⁵ effect. Chapter 3 also illustrates how to use Gibbs Sampling to estimate the unknown parameters in such models via a Bayesian framework. Chapter 4 validates model code performances by testing with a simulated dataset.

⁵ In this dissertation, "region" is used to indicate a cluster of observational units. It can be interpreted as a sub-area of the study zone, a neighborhood, or a socially defined group.

Chapter 5 describes the empirical datasets used in this study - one for the dynamics of land development intensity and the other for ozone concentration levels. Both datasets come from Austin's urban area, but their spatial scales and observational units differ. Chapter 5 explains details about the utilization of satellite data and its integration with other data sources. In Chapter 6, the model is applied to the Austin datasets, and the effects of different factors are discussed based on the estimation results. The estimates also are applied to predict land development intensity levels and ozone concentration levels in two hypothetical scenarios. Chapter 7 summarizes this dissertation's contributions and findings, while illuminating study limitations and opportunities for extension of the work.

1.7 CHAPTER SUMMARY

This chapter introduced key motivations for and objectives of this dissertation research, which are as follows: (1) establishing a statistically rigorous model for analyzing discrete response data in a spatial and temporal setting, (2) mining information present in satellite imagery, and (3) understanding the dynamics of land development and ozone concentrations in Austin, Texas. Chapter 2 provides further background while placing more emphasis on methodology. Subsequent chapters show how these objectives are realized, how the model is specified and tested, and how the data are assembled and analyzed, how the model compares to simpler specifications and approaches, and how the results lend themselves to useful interpretations for application in a broad range of topic areas.

CHAPTER 2. LITERATURE REVIEW

While Chapter 1 offers a discussion of some land development change and air quality modeling literature, this chapter emphasizes the methodological side. This chapter starts from a discussion of fundamental aspects. The exploration is then deepened into specific topics, which are more and more closely related to the innovative method that will be proposed in the study. The following sections introduce common notions in spatial econometrics and models of discrete responses, and discuss existing spatial models of discrete responses. Methods for specifying and estimating such models are compared, with emphasis on the Bayesian methods used in this study. The chapter ends with a discussion of the Bayesian approach for ordered, discrete responses in a dynamic and spatial setting.

2.1 SPATIAL ECONOMETRICS

Econometrics is a statistically rigorous method for mining useful information and exploring relationships of interest that are embedded in behavioral data. As an emerging subfield, spatial econometrics holds considerable interest. The need for spatial econometrics exists in fields like sociology, agriculture, ecology, environment and city planning. (See, e.g., recent reviews in Anselin [1999], and Anselin and Florax [2002].) In transportation and urban studies, recognition of spatial relationships is almost inevitable. Normally, land owners make development decisions based on their knowledge and prediction of neighboring land development (See, e.g., Waddell, 2002, and Candau et al., 2000). As a result, land development is often clustered. For example, one can expect that a parcel of land is more likely to be intensely developed if its neighborhood offers intensely developed land.

Such spatial autocorrelation also exists in many naturally evolving activities. For example, ozone concentration levels are influenced by land cover, weather and many other factors. Perhaps most importantly, they are influenced by the

concentrations of chemicals and emissions in neighboring areas. As a result, ozone concentration patterns tend to be smooth (see, e.g., Guttorp et al., 2006), and their analysis needs to be examined in both a temporal and spatial context (Gao and Niemeier, 2005).

As long as there is spatial interaction, there is a need for spatial econometrics (Anselin, 1988). In some cases, people need to understand the operation of a spatial process in order to estimate the magnitude and trend of the neighboring units' influence. In such circumstances, explicit spatial effects are required. Even in circumstances where analysts are not so interested in spatial relationships, they may require spatial econometric methods in order to obtain reasonable and robust model estimates. Otherwise, their model specification may be statistically and functionally problematic. For example, if positive spatial correlation exists but is ignored in a linear model, the estimation is biased and precision generally is overestimated (Magnussen, 1992, and Briggs, 2006). Intuitively, the coefficients tend to be biased high because areas with higher event magnitudes will have a greater impact on model estimates. Precision is exaggerated (i.e., lower standard errors are reported and estimates are more likely to be found "statistically significant") because the actual number of independent observations is lower than assumed (Anselin, 1988).

Of course, in order to account for spatial effects in data, one first needs to have an indicator of "distance" between observations. A natural choice is an indicator variable for general location (e.g., which city or neighborhood a zone belongs to) (Wangen and Biom, 2001). However, as heterogeneity or cluster divisions increase, the number of indicator variables may increase to the extent that the method is no longer computationally feasible. A more effective measurement tends to be relative distance between observations. As Tobler's first law of geography claims, everything is related to everything else, but near things are

more related than distant things (Tobler, 1970). Using this law, it is not difficult to construct a spatial weight matrix to index the relative location of all pairs of observational units. The value of each element can be either an index of contiguity or a monotonically decreasing function of impedance (Anselin et al., 2006, O’Sullivan and Unwin, 2003). Depending on the topic of interest, this impedance can be Euclidean distance, travel time, number of links in a chain of economic interactions or even rock hardness. This spatial weight matrix is also called a spatial lag operator, analogous to the lag operator in time series data (Anselin, 1999), which essentially produces a weighted average of the neighbors’ values (of error terms, explanatory variables or responses, for example).

This dissertation uses an index of (queen) contiguity to construct the two weight matrices used throughout. One advantage of such matrices is that most elements are zero, allowing utilization of sparse matrix algorithms. This numerical efficiency and the resulting, shorter calculation times are important in practice, especially for studies with large samples and/or complex specifications. More importantly, the convenience of using a contiguity matrix is not obtained at the cost of significant loss in spatial information, as compared to a more detailed (thus more memory-demanding and computationally intensive) distance-decay matrix. This is because, in practice, the weight matrix is typically row-standardized (as discussed in Chapter 3). In this case, using a distance decay matrix no longer provides more meaningful interpretations.⁶ Furthermore, when the magnitudes of distance differ considerably, normalizing a distance decay matrix may push many elements down to almost zero values, creating a matrix numerically similar to a contiguity matrix.

⁶ As Anselin (1988, P. 514) points out, “distance decay has a meaningful economic interpretation; and scaling the rows so that the weights sum to one may result in a loss of that interpretation.”

Spatial econometrics certainly shares some similarities with time series data analysis. However, in contrast to the relatively mature development of time-series data analysis, spatial econometrics is in its infancy. The reason for this is that the correlation relationships are two dimensional. (With standard longitudinal/time series data; there is a single direction of effect.) In two-dimensional space, one can not rely on a single, serial correlation parameter. Methodological issues are not the only reason hindering the development of spatial econometrics: computational issues also play a key role. Spatial data can be large-scale and memory intensive (Nelson and Geoghegan, 2001), and computational demands increase exponentially with data set size. Therefore, efficient estimation methods generally must accompany model specification in order to enjoy operational success. Over the past two decades, researchers have invested much effort with these considerations in mind. Existing spatial econometric studies can be generally categorized into three classes, each offering a distinct notion for dealing with the hypothesized spatial relationship.

The first class of specification is geographically weighted regression (GWR), as formulated by Fotheringham et al. (2002). To some extent, GWR can be considered a transition from spatial statistics to spatial econometrics. Essentially, GWR extends the idea of local spatial statistics to model estimates: Local spatial statistics can be calculated for each individual based on neighborhood information by applying a matrix of location weights (W) to the standard formulae for the statistic of interest. Similarly, GWR generates a separate regression equation for each observational unit (Mennis, 2006, and Briggs, 2006). The model estimates thus vary from location to location. The GWR model specification can be illustrated via the following simple regression example: using (z_1, z_2) to denote the coordinates of one data point's position, the model can be written as follows:

$$y(z_1, z_2) = \beta_0(z_1, z_2) + \beta_1(z_1, z_2)x_1 + \varepsilon(z_1, z_2) \quad (2.1)$$

In other words, now the model parameters are location-specific, rather than constant across all observations. This model can be estimated based on a predetermined weighting scheme which basically assigns data closer to (z_1, z_2) a higher weight than data further away. This method was first applied in ecology and biology, where it is reasonable to believe that spatial processes vary with location. For example, Zhang et al. (2004) used GWR to model spatial variation of tree diameter and height relationships, and Shi et al. (2006) applied GWR to model deer distributions. In analyzing human activities, Malczewski and Poetz (2005) used GWR to find local variations in the relationships between the risk of residential burglary and neighborhood socioeconomics in London, Ontario. An example of GWR in transportation is Zhao and Park's (2004) study of annual average daily traffic (by zone). As Paez (2005) explains, GWR is an appealing approach for exploring spatial non-stationarity. When the interest lies in spatial autocorrelation, rather than behavioral variation over space, the GWR method is no longer an appropriate choice.

The second class of spatial econometrics models is called spatial filtering. Work mentioned previously, including studies by Nelson and Hellerstein (1997), Wear and Bolstad (1998), and Munroe et al. (2001) (in Chapter 1), belongs to this class. Some other studies use spatial statistics as explanatory variables for similar reasons (i.e., to remove spatial correlations.) These statistics are normally derived from exogenous, location-specific data. For example, Kockelman (1997) included land use mix, accessibility, and neighborhood entropy statistics in her study of household travel behaviors. However, issues remain in such approaches because spatial correlation generally cannot be perfectly removed through sampling, filtering, or controlling for neighborhood attributes. Neighborhood information and/or any spatial statistics derived from it can be endogenous in many cases. In

other words, explanatory and dependent variables are often interdependent, and such variables typically are not available to analysts in predicting future behaviors. Moreover, application of such models (e.g., traffic and land use forecasting) generally requires modeling of all data points, so correlation exists in the applications.

The third class of model is direct incorporation of spatial effects in model specification. Though this method has also uses spatially lagged explanatory and/or response variables, as in the spatial filter models, many of the variables are treated as endogenous. Anselin (1999) notes that this method can correct for “structural instabilities” (including heteroskedasticity and variable model coefficients) occurring as a result of spatial heterogeneity in the data. Anselin (1988, 1999) suggests that there are three main methods for incorporating such spatial effects: use of spatial stochastic processes, a direct representation of correlations, and a non-parametric framework. The relatively common spatial autoregressive (SAR) and spatial moving average (SMA) specifications are examples of spatial stochastic processes. (Anselin, 1988, Anselin and Bera, 1998, and Anselin, 2003 provide extensive technical discussions on these two processes.)

In general, SAR and SMA are used for dependent variables and error terms, respectively. The former case is called “spatial lag”, while the latter is often called “spatial error”. By using these two specifications, it is assumed that the spatial process follows a recursive pattern. Such methods are rather used regularly by researchers, thanks to their flexibility and applicability. For example, Besner (2002) used SAR to analyze housing prices in the Montreal Urban Community real estate market, and Miyamoto et al. (2004) incorporated SAR processes in both dependent variable and error term specifications for their location choice analysis. Frazier and Kockelman (2005) use a SAR process to understand urban land cover change in Austin.

A frequently encountered issue in spatial stochastic processes is the lack of stationary covariance terms, which can violate asymptotic property assumptions established based on the central limit theorem and law of large numbers (Cressie, 1993). By using a “direct representation,” the covariance matrix among unobservable components can be directly expressed as an inverse function of distances. This method is most suitable when some prior knowledge (for the functional form) of the spatial interaction pattern exists. This also meets stationarity requirements (Anselin, 1999). Such models have been used primarily for analysis of housing prices, including works by Dubin (1988, 1992), Olmo (1995) and Basu and Thibodeau (1998). Wang and Kockelman (2006c) also applied such methods, for their analysis of land cover change in Austin. Though this specification is intuitive in some cases, formulation of the inverse function is often restricted, because flexible expressions may suffer from estimation and identification problems.

Anselin’s (1988, 1999) third suggestion for direct incorporation of spatial effects involves non-parametric methods. While these do not require an explicit spatial process or functional form for the distance decay, they do require a long panel of data, with a time dimension (T) that is much greater than the cross-sectional dimension (N) (see Fiebig, 1999). Thus, they are seldom found in practice.

2.2 MODELS OF DISCRETE RESPONSES

Models of discrete response are an important sub-area of econometrics. They are used to model discrete choices among sets of alternatives, rather than a continuous response (Greene, 2000). Such models play an important role in scientific studies, both social and natural. The specification of discrete response models tends to require specific assumptions on the error term distribution. Two commonly used specifications are probit and logit models. The most basic form is a binary response, where the value of the dependent variable is either 0 or 1,

indicating no or yes. Of course, in many circumstances, the number of alternatives is more than two. When the data is multinomial and unordered, a common model specification is established based on the utility maximization theory introduced by McFadden (1994). In this framework, the alternative offering the maximum utility (a latent variable) is chosen. If the data are ordered, the model specification is more similar to binary choice settings, but with a set of thresholds to distinguishing different level of response (alternatives) and requiring estimation. Land development decisions often offer a nice example of unordered, multinomial discrete choice, but land development intensity and ozone concentration, when categorized by level, offer an example for ordered (discrete) response data. Such nonlinear models for limited dependent variables generally require more complicated estimation techniques than do continuous data. When spatial and temporal effects also exist, the modeling process becomes even more complicated than the spatial econometric and time series data models used for continuous response data.

2.3 SPATIAL ECONOMETRICS FOR MODELS OF DISCRETE RESPONSE

Similar to the standard probit and logit models, existing studies using spatial econometrics for discrete response analysis can be divided into two categories based on error term assumptions: those assuming normally distributed error terms and those assuming GEV distributed error terms. Also, as in standard spatial econometrics, methods for dealing with spatial effects in discrete choice models can be categorized into the three types discussed earlier, in Section 2.1. The first method is GWR, where spatial variation in behavioral parameters is of strong interest. In an analysis on suburban subcenters and employment density, McMillen and McDonald (1998) propose the idea of applying standard logit or probit methods to distance weighted sub-samples of the data in place of least

squares, which is essentially using GWR to deal with discrete responses. LeSage (1999) provided code for producing binary logit and probit GWR estimates, using crime data. Atkinson et al. (2003) also used a GWR binary logit model to explore relationships between the presence (or absence) of riverbank erosion and geomorphological controls. Vanasse (2005) incorporated GWR in a binary logit model to study spatial variation in the management and outcomes of acute coronary syndrome.

The second method, spatial filtering, has been applied more broadly. It saves much specification and estimation effort. In addition to several land use/land cover models (e.g., Nelson and Hellerstein, 1997, Wear and Bolstad, 1998, and Munroe et al., 2001), there are also many other works that use this method. For example, an early study by Boots and Kanaroglou (1988) introduced a measure of spatial structure and used it as an explanatory variable when considering spatial effects in intra-metropolitan migration in Toronto. Dugundji and Walker (2004) controlled for spatial network independencies in their mixed logit model when studying mode choice behavior. Coughlin et al. (2003) incorporated global and regional spatial effects into an analysis of state lotteries.

The third method incorporates spatial effects directly in a discrete choice model setting and is the focus of this study. This method can be further divided into two approaches.

The first considers spatial autocorrelation across choices or alternatives, as often discussed for location choice models. This approach extends the commonly used GEV model by allowing correlated alternative-specific error terms in a mixed logit framework. For example, Miyamoto et al. (2004) assumed that location choice follows an SAR process, and used the weight matrix as a multiplier on dependent variables. Bhat and Guo (2004) used a contiguity matrix on their latent dependent variables to represent alternative-zone correlation patterns.

The second approach considers spatial autocorrelation across observational units (or individuals), the topic of this dissertation. Currently, studies recognizing such spatial autocorrelation are limited to binary choice settings. To some extent, Wang and Kockelman (2006c)'s work on estimating urban land cover evolution seems an exception, because multiple choices are studied in a mixed logit framework. However, rather than permitting a more flexible SAR process, Wang and Kockelman used a direct representation method and assumed a specific distance-decay function for inter-observational correlations, making the spatial correlation pattern across observations rather arbitrary. All other existing spatial probit and logit work is binary in nature. Anselin (2005) reviewed such spatial probit models and notes that, McMillen (1995) first used the EM algorithm to estimate a probit model with an SAR process. Vijverberg et al. (1999, 2000) specified probit models with both spatial errors and spatial lags, and then estimated these models by using recursive importance sampling (RIS) to approximate the n-dimensional log-likelihood. Beron and Vijverberg (2004) used a similar method to examine the bias caused by ignoring spatial relationships in a probit model. LeSage (2000) specified a model with a spatially correlated error term and used Gibbs sampling for estimation. Smith and LeSage (2002) extended this study by incorporating a regional effect and used Bayesian techniques to analyze the 1996 presidential election results. Similar studies include Kakamu and Wago's (2006) Bayesian estimation of a spatial probit model for panel data to analyze the business cycle in Japan. Another estimation approach is the generalized method of moments, or GMM. Pinkse and Slade (1998) first used GMM to estimate a probit model with spatial error components. Pinkse et al. (2005) refined that study by incorporating a dynamic structure for dependent variables and applying a one-step GMM. There is also spatial logit model that incorporates spatial autocorrelation across observations. For example, Klier and

McMillen (2007) used GMM to estimate a spatial logit model for analyzing the clustering of auto supplier plants in the U.S.

It has been argued that the most important issue for estimation of any spatial discrete response data model is efficiency. The use of GMM is limited because it requires orthogonality conditions (as discussed in works like Klier and McMillen, 2007, Pinkse and Slade, 1998, and Pinkse et al., 2005), and standard errors must be derived. For this reason, it presently is applied only to binary response models; it has not yet been extended to multiple-response models. All the other estimation methods can be called simulation estimators. As Anselin (2005) concluded, all current simulation estimators are slow, but Gibbs sampler is relatively less slow. In other words, among all three general methods discussed above, the most promising one for a model of multiple discrete response (ordered and unordered) with spatial effects (both autocorrelation and heteroskedasticity) is one using a Gibbs sampler in a Bayesian framework. The next section gives a brief review of Bayesian notions.

2.4 THE BAYESIAN APPROACH TO MODEL ESTIMATION

In contrast to frequentist methods (i.e., classical statistical analysis), the Bayesian approach is rather straightforward in both model estimation and results interpretation. The primary motivation for using a Bayesian approach is its rather direct interpretation of parameter estimates and probabilities. A Bayesian approach yields estimates of parameter *distributions* (rather than relying on asymptotics for normality). These distributions effectively define intervals that can be “regarded as having a high probability of containing the unknown quantit(ies) of interest” (Gelman et al., 2004). In contrast, frequentist methods focus on producing point estimates and rather standard confidence intervals, and resulting probabilities that are strictly interpreted as “long run (asymptotic) relative frequenc(ies)” (Koop et al., 2007).

In practice, an important advantage of a Bayesian framework is its flexibility, allowing it to deal with complex estimation problems more easily. In fact, this is the main reason for this dissertation’s choice of a Bayesian framework – in addition to wanting to develop new methods of model estimation for the transportation sciences (where frequentist methods are the norm). In general, Bayesian estimation via Markov chain Monte Carlo (MCMC) simulation relies on a set of conditional distributions to deduce each parameter’s marginal distribution. In this way, models with many parameters and complicated multiple-layered probability specifications can be decomposed into a set of simpler sub-problems. By contrast, with frequentist methods, the models have to deal directly with any complicated model specification and any statistical problems arising from it. Of course, another well-understood advantage of using a Bayesian approach is that by having priors, one can make use of established intuition and experience to balance new information found in sample data. Unfortunately, such priors are generally rare in the practice of transportation engineering, so this benefit is often not realized. The Bayesian approach is regularly used without informative priors. (For example, Wallerman et al. (2004) used it to analyze remotely sensed forestry data, and Hamilton et al. (2005) used it to estimate expansion and migration rates for Swiss populations.)

2.4.1 Bayesian Theory

Essentially, Bayesian approaches rest on the basic property of conditional probability known as Bayes’ rule:

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) = \frac{\pi(\boldsymbol{\theta}|\mathbf{X})\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X})}{\pi(\mathbf{y}|\mathbf{X})} \propto \pi(\boldsymbol{\theta}|\mathbf{X})\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}) \quad (2.2)$$

where \mathbf{y} is a vector of dependent variables and \mathbf{X} is a matrix explanatory information. Together they compose the observed data. $\boldsymbol{\theta}$ is a vector of unknown parameters. If explanatory variables are irrelevant to the parameters (as in most

cases), $\pi(\boldsymbol{\theta}|\mathbf{X}) = \pi(\boldsymbol{\theta})$. This is known as the prior, or prior distribution of the random parameters $\boldsymbol{\theta}$. One can incorporate intuition and/or experience in this prior distribution. $\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X})$ is the likelihood function of \mathbf{y} given \mathbf{X} and $\boldsymbol{\theta}$. Thus, Bayesian methods model information from two sources: one's beliefs and sample data. Together, they lead to updated information on $\boldsymbol{\theta}$, producing a posterior distribution of $\boldsymbol{\theta}$, which is denoted as $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$.

The most commonly used method for estimating unknown parameters is MCMC simulation technique, and the Gibbs sampler is one of the most popular⁷. The next section first introduces MCMC and then the Gibbs sampler.

2.4.2 Gibbs Sampling

As Gelman et al. (2004) summarize, MCMC simulation is based on drawing parameter values from approximate distributions and then correcting these draws to better approximate the target posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$. The draws are sampled sequentially, and the distribution of the sampled draws ignores the initial (“burn in”) sample values.

The Gibbs sampler is a particular MCMC algorithm, also called an alternating conditional sampler, because in each iteration it simulates components conditional on the values of all other parameters. In other words, at iteration t , each $\boldsymbol{\theta}_j^t$ is sampled from its conditional distribution given all other components of $\boldsymbol{\theta}$:

$\pi(\boldsymbol{\theta}_j^t | \boldsymbol{\theta}_{-j}^{t-1}, \mathbf{y})$, where $\boldsymbol{\theta}_{-j}$ represents all components of $\boldsymbol{\theta}$, except for $\boldsymbol{\theta}_j$.

⁷ The Metropolis-Hastings algorithm is also an important (and the standard) member of MCMC simulation techniques. It is especially useful when models are not conditionally conjugate. (Readers may see Chapter 3 for a detailed explanation of the term “conjugate”.) However, as Gelman et al. (2004) suggest, Gibbs sampling is the simplest of the MCMC algorithms, and should be chosen first whenever possible. This dissertation uses a Gibbs sampler throughout.

2.4.3 Data Augmentation

Data augmentation where one adds auxiliary variable, is an efficient way to simplify or accelerate convergence when using the Gibbs sampler (see, e.g., Tanner and Wong, 1987, Chib, 1992, Albert and Chib, 1993, and Gelman et al., 2004). In cases of non-binary discrete response the latent variables must be recovered, and so one must resort to data augmentation. Chapter 3 explains in more detail how this data augmentation technique is used here.

2.5 BAYESIAN APPROACH FOR DYNAMIC SPATIAL ORDERED CATEGORICAL MODELS

The Bayesian approach for (stationary, non-spatial) discrete choice models was introduced by Albert and Chib (1993). LeSage (2000) first extended Albert and Chib's approach to models involving spatial dependencies. Later work by Smith and LeSage (2002) further extended the model by incorporating an error specification that allows both spatial dependencies and general spatial heteroscedasticity. However, all such studies deal only with binary data. As previously discussed, many data sets offer multiple categories. No existing studies tackle such patterns in a spatial context. While Albert and Chib (1993) briefly mentioned possible extensions from binary data to ordered categorical data, they did not offer any methodological details. Several years later, Johnson and Albert (1999) suggested a detailed Bayesian framework for modeling ordinal data, and Cowles (1996) presented a method for accelerating MCMC convergence for models like the ordered probit. Girard and Parent (2001) even extended Albert and Chib's study (1993) to temporally autocorrelated ordered categorical data, but there is nothing spatial in these studies.

A related and interesting topic is the use of Bayesian techniques to analyze unordered discrete response data. Recently, both Scott (2004) and Fruhwirth-Schnatter and Waldl (2004) used data augmentation techniques for recovering

latent utilities in a multinomial logit set-up. However, due to very complex specification and estimation demands made via Bayesian simulation, no studies have attempted to accommodate dynamic and/or spatial features.

To summarize, existing studies either deal with binary data, spatial relationships, or dynamic patterns. None incorporates all these conditions simultaneously. This dissertation is inspired by such studies but adds sophistication while combining space and time for ordered categorical data. It goes beyond a simple extension or combination of these works. The contribution of these prior studies is discussed in more detail in the next chapter, through illustrations of model specification and estimation techniques.

2.6 CHAPTER SUMMARY

This chapter reviewed a variety of literature related to the specification and estimation of the proposed DSOP model. Section 2.1 introduced general methods and common notions used in spatial econometrics, while Section 2.2 briefly introduced models of discrete response. Section 2.3 then summarized existing methods for specifying and estimating models involving spatial effects and discrete responses.

After a comparison of these existing methods, it seems likely that the Bayesian framework with a Gibbs sampler is the best method for coping with the specification and estimation issues inherent in this study. Thus, Section 2.4 further discussed Bayesian methods, and Section 2.5 discussed Bayesian approaches for models of ordered categorical data in a dynamic and spatial setting.

CHAPTER 3. SPECIFICATION OF A DYNAMIC SPATIAL ORDERED PROBIT (DSOP) MODEL, FOR BAYESIAN ANALYSIS

3.1 INTRODUCTION

This chapter describes the specification of a dynamic spatial ordered probit model, and methods for its estimation. The intuition behind this proposed model specification is that, first, spatial effects are incorporated into a standard ordered probit model, and then dynamic features are included. The model is to be estimated in a Bayesian framework using Gibbs sampling. Prior distributions of all parameters and variables of interest are explored here, and their posterior distributions are derived. The MCMC simulation process is then summarized. Several special (and simple) cases of the models are discussed briefly as well.

3.2 MODEL SPECIFICATION

3.2.1 Standard Ordered Probit (OP) Model

A standard ordered probit model has been used widely for estimating discrete responses of an ordinal nature (Greene, 2000). The model is built upon a latent regression that is expressed as follows:

$$U_i = X_i' \beta + \xi_i \tag{3.1}$$

where i indexes observations, ($i = 1, \dots, N$), and U_i is a latent (unobserved) response variable for individual i . X_i is a $Q \times 1$ vector of explanatory variables, and β is the set of corresponding parameters. ξ_i stands for unobservable factors for observation i and (for a standard ordered probit model) is assumed to follow an iid standard normal distribution.

The observed response variable, y , for the i^{th} observation is as follows:

$$y_i = s \text{ if } \gamma_{s-1} < U_i < \gamma_s, \quad s = 1, \dots, S$$

That is, the observed variable is a censored form of the latent variable, and its possible outcomes are integers ranging from 1 to S . The latent variable U_i is allowed to vary between unknown boundaries $\gamma_0 < \gamma_1 < \dots < \gamma_{S-1} < \gamma_S$, where γ_0 is $-\infty$ and γ_S is $+\infty$. If constants are to be included in the explanatory variables, γ_1 also is normalized to equal zero. The probabilities for these S outcomes are as follows:

$$\begin{aligned} \Pr(y_i = 1 | \mathbf{X}_i) &= \Phi(\gamma_1 - \mathbf{X}_i' \boldsymbol{\beta}) - \Phi(\gamma_0 - \mathbf{X}_i' \boldsymbol{\beta}) \\ \Pr(y_i = 2 | \mathbf{X}_i) &= \Phi(\gamma_2 - \mathbf{X}_i' \boldsymbol{\beta}) - \Phi(\gamma_1 - \mathbf{X}_i' \boldsymbol{\beta}) \\ &\vdots \\ \Pr(y_i = S | \mathbf{X}_i) &= \Phi(\gamma_S - \mathbf{X}_i' \boldsymbol{\beta}) - \Phi(\gamma_{S-1} - \mathbf{X}_i' \boldsymbol{\beta}) \end{aligned} \quad (3.2)$$

where $\Phi(\bullet)$ is the Cumulative Distribution Function (CDF) for standard normal distribution. If ξ_i has non-unit variance, the CDF in Equation (3.2) must be modified as follows:

$$\Phi\left(\frac{\gamma_s - \mathbf{X}_i' \boldsymbol{\beta}}{\nu}\right) \quad (3.3)$$

3.2.2 Spatial Ordered Probit (SOP) Model

In many studies, individuals are surveyed from a region containing several sub-regions or neighborhoods. A certain number of observations is collected from each of these sub-regions. In such cases, the effects of different regions need to be considered, so that the latent variable is now in the form of

$$U_{ik} = \mathbf{X}_{ik}' \boldsymbol{\beta} + \xi_{ik}, \text{ with } \xi_{ik} = \theta_i + \varepsilon_{ik} \quad (3.4)$$

where i now indexes regions (instead of individuals) ($i = 1, \dots, M$) and k indexes individuals inside each region (i.e., $k = 1, \dots, n_i$). In other words, there are M regions, each containing n_i observations, so that the total number of observations is $\sum_{i=1}^M n_i = N$.

The main difference between Equations (3.4) and (3.1) is that the unobserved factor ξ_{ik} is now composed of two parts: a “regional effect” θ_i and an individual effect ε_{ik} . This θ_i captures all unobserved, common features for observations within region i . To some extent, this specification is very close to a random effect in panel data, only here the “common factor” is cross-sectional rather than temporal. As discussed in Chapters 1 and 2, these regional effects should exhibit spatial autocorrelation: individuals in region i are likely to be more similar to those in neighboring regions than those in more distant locations. Therefore, a spatial autoregressive process can be formulated here, where

$$\theta_i = \rho \sum_{j=1}^M w_{ij} \theta_j + u_i, \quad i = 1, \dots, M \quad (3.5)$$

and weight w_{ij} can be derived based on contiguity and/or distance. In addition,

the weight matrix is row-standardized⁸ so that $w_{ii} = 0$ and $\sum_{j=1}^M w_{ij} = 1$. The

magnitude of overall neighborhood influence is thus reflected by ρ , also called the spatial coefficient. u_i aims to capture any regional effects that are not spatially distributed, and is assumed to be iid normally distributed, with zero

⁸ The row-standardized approach is chosen because in this way the “ \mathcal{W} ” term becomes essentially a weighted average of observations at neighboring locations” (Anselin and Hudak, 1992). This leads to a more meaningful interpretation of ρ .

mean and common variance σ^2 . Stacking all regions, then, the vector of regional effects can be formulated as

$$\boldsymbol{\theta} = \rho \mathbf{W} \boldsymbol{\theta} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_M) \quad (3.6)$$

Here, \mathbf{W} is the exogenous weight matrix with elements w_{ij} and \mathbf{I}_M is an identity matrix with rank M . Let $\mathbf{B}_\rho = \mathbf{I}_M - \rho \mathbf{W}$, where the subscript ρ means that \mathbf{B}_ρ depends only on the unknown parameter ρ . Now, the vector of regional effects can be expressed as

$$\boldsymbol{\theta} = \mathbf{B}_\rho^{-1} \mathbf{u} \quad (3.7)$$

In other words, the distribution of $\boldsymbol{\theta}$ depends on two unknown parameters: ρ and σ^2 . It has a multivariate normal distribution:

$$\boldsymbol{\theta} | (\rho, \sigma^2) \sim N\left[0, \sigma^2 (\mathbf{B}'_\rho \mathbf{B}_\rho)^{-1}\right] \quad (3.8)$$

The intuition behind this “regional effect” can be explained as follows: In many cases, individuals in a region⁹ share common features, yet these features differ from region to region. One source of such differentiations is policy variations by regions. For example, parcels subject to the same zoning constraints may share common features, but differ across zone boundaries. Animals enjoying the same habitat share experiences, thanks to vegetation and micro climates. Their settings shift across wide rivers, mountain ranges, or high-capacity freeways. Multiple regions may exist based on these physical boundaries. In short, there are reasons to believe that observations across space are influenced by “local effects”, which may exhibit spatial autoregressive patterns as a function of proximity. The use of such regional effects to capture certain spatial dependencies also enhances

⁹ As used here, “region” means a cluster of observational units, within the same neighborhood or socially defined group (such as members of the same household or employees in the same firm).

computational efficiency: normally, the number of regions is much lower than the total number of observations, allowing use of a \mathbf{B}_ρ of relatively low rank. Thanks to a lower dimension, the inversion of \mathbf{B}_ρ and calculation of its eigenvalues, are much less memory-intensive. Of course, both of these computations are necessary for parameter estimation.

This “regional effect” offers an opportunity to make each individual a region, i.e., $n_i = 1, \forall i \in M$, (so $M = N$). This allows all individuals to be spatially auto-correlated without imposing regional boundaries. While increasing computational burdens, this is definitely feasible with a reasonable sample size.

The final item requiring specification is the individual effect, ε_{ik} . It is computationally simplest to assume an iid distribution for ε_{ik} . And, within each region, it is behaviorally reasonable to make such assumptions (i.e., all ε_{ik} follow a normal distribution with zero mean and variance ν_i . Across regions, it seems reasonable to expect heteroscedasticity. Stacking all observations and denoting $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$, one has

$$\mathbf{V} = \begin{pmatrix} \nu_1 \mathbf{I}_{n1} & & \\ & \ddots & \\ & & \nu_M \mathbf{I}_{nM} \end{pmatrix} \quad (3.9)$$

which is an $N \times N$ matrix with non-zero elements only along its diagonal.

3.2.3 Dynamics to the Spatial Ordered Probit Model

As discussed in Chapter 2, there are four ways to incorporate both spatial and temporal (sequential) dependencies (Anselin, 1999). In this study, it is assumed that a time-space recursive formulation is proper, which means that the current value depends on the previous period’s value (at the same location, and thus

affected by neighboring locations), along with various contemporaneous factors. Furthermore, after controlling for all these temporally lagged and contemporaneous variables, the residuals remain spatially autocorrelated:

$$U_{ikt} = \lambda U_{ikt-1} + \mathbf{X}_{ikt}' \boldsymbol{\beta} + \theta_{it} + \varepsilon_{ikt}, \quad t = 1, \dots, T \quad (3.10)$$

where t indexes time periods and λ is the temporal autocorrelation coefficient to be estimated. The absolute value of this λ must be less than one in order to guarantee temporal stationarity (Hamilton, 1994). Each individual is now observed T times (the dataset is a balanced panel), and the total number of observations is NT . θ_{it} is assumed to iid distributed over t and so is ε_{ikt} . In other words, after controlling for lagged dependent variables (U_{ikt-1}), the error terms are sequentially uncorrelated and identically distributed. Though a more flexible framework is, of course, to allow θ_{it} and ε_{ikt} to exhibit sequentially dependencies or at least heteroscedasticity, it is reasonable enough to believe that after one controls for lagged latent dependencies (both spatial and temporal), the remaining error terms may be temporally constant, i.e.,

$$\theta_{it} \equiv \theta_t \text{ or } \boldsymbol{\theta}_t = \boldsymbol{\theta}, \text{ for all } t = 1, \dots, T \quad (3.11)$$

and

$$\varepsilon_{ikt} \equiv \varepsilon_{ik} \text{ or } \boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}, \text{ for all } t = 1, \dots, T \quad (3.12)$$

A standard problem in considering such dynamic features is the treatment of initial conditions (Wooldridge, 2002). U_{ik0} is called the initial latent variable and is an unknown quantity, generally assumed to be normally distributed. Section 3.3.3 discusses the distributional assumptions of U_{ik0} in more details.

3.2.4 Specification of the Dynamic Spatial Ordered Probit (DSOP)

Model

Equations (3.10) through (3.12) specify a dynamic spatial ordered probit model. The two empirical examples used in this dissertation (land development intensity levels and ozone concentration categories) both fit this specification. For example, land development decisions strongly depend on pre-existing and existing conditions, as well as owner/developer expectations of future conditions (such as local and regional congestion, population, and school access). These expectations can be approximated using contemporaneous measures of access and land use intensity, after which some spatial correlation in unobserved factors is likely to remain.

It is similar with ozone concentration levels: changes are temporally continuous so inclusion of lagged variables is wise. The impact of some factors, such as temperature, may be instantaneous, so their contemporaneous values should be used. Advection and other unobserved factors may cause spatial dependence, so spatially autocorrelated effects (regional/clustered or observational in nature) should be considered. Certainly, recognition of such temporal dependencies and spatial autocorrelation (of nuisance terms) is behaviorally more convincing and statistically more rigorous than simply controlling for contemporaneous factors and ignoring other, underlying spatial dependencies.

The model specification can be expressed in vector form as follows: for each $t \in T$, observations can be stacked by region, then by individuals. The resulting utility vector is expressed as:

$$U_t = \lambda U_{t-1} + X_t \beta + L \theta + \varepsilon \quad (3.13)$$

where $\mathbf{U}_t = \begin{bmatrix} \mathbf{U}_{1t} \\ \vdots \\ \mathbf{U}_{it} \\ \vdots \\ \mathbf{U}_{Mt} \end{bmatrix}$, with each $\mathbf{U}_{it} = \begin{bmatrix} U_{i1t} \\ \vdots \\ U_{ikt} \\ \vdots \\ U_{in_t t} \end{bmatrix}$. It is similar with \mathbf{X}_t , only \mathbf{X}_t is an

$N \times Q$ matrix (instead of an $N \times 1$ vector). Here, $\mathbf{L} = \begin{bmatrix} \mathbf{l}_{n_1} & & \\ & \ddots & \\ & & \mathbf{l}_{n_M} \end{bmatrix}$, with each

$\mathbf{l}_{n_i} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_i}$ being a $n_i \times 1$ vector of 1's.

If observations over all time periods are stacked, the model can be written as

$$\mathbf{U}^\lambda = \mathbf{X}\boldsymbol{\beta} + \Delta\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (3.14)$$

where \mathbf{U}^λ is the vector of differences between adjacent time periods:

$$\mathbf{U}^\lambda = (\mathbf{U}_1^\lambda, \mathbf{U}_2^\lambda, \dots, \mathbf{U}_T^\lambda)', \text{ with each } \mathbf{U}_t^\lambda = \mathbf{U}_t - \lambda\mathbf{U}_{t-1}.$$

$$\Delta = \mathbf{l}_T \otimes \mathbf{L} \quad (3.15)$$

where \mathbf{l}_T is a $T \times 1$ vector of 1's.

Here, \mathbf{X} is an $NT \times Q$ matrix, and $\boldsymbol{\varepsilon}$ is an $NT \times 1$ vector with variance matrix

$$\boldsymbol{\Omega} = \mathbf{I}_T \otimes \mathbf{V} \quad (3.16)$$

The likelihood function is thus

$$\Pr(\mathbf{y}|\mathbf{U}, \boldsymbol{\gamma}) = \prod_{t=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \sum_{s=1}^S \delta(y_{ikt} = s) \cdot \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s) \quad (3.17)$$

where $\delta(A)$ is an indicator function equaling 1 when event A is true (and 0 otherwise). Now it is clear that the parameters of interest are $(\beta, \lambda, \rho, V, \sigma^2, \gamma)$, together with unobserved (“nuisance”) variables θ and U . One way to estimate these is via MCMC sampling under a Bayesian framework, as discussed below.

3.3 PARAMETER ESTIMATION VIA MCMC SIMULATION

Estimation of the dynamic spatial ordered probit model is achieved via MCMC methods, as described below. Sections 3.3.1 and 3.3.2 discuss the theoretical background for using a set of conditional distributions to approximate the joint posterior distributions, and Section 3.3.3 explains how prior distributions are chosen. Section 3.3.4 then describes how appropriate conditional posterior distributions for all parameters are established.

3.3.1 Conditional Distribution

As discussed in Chapter 2, MCMC simulation can be used in model estimation by sampling sequentially from the parameters’ complete set of conditional distributions. Gelfand and Smith (1990) showed that MCMC sampling leads to consistent estimates of the true joint posterior distribution of all parameters (including “nuisance parameters”, such as V, σ^2 and θ). Using Bayes’ basic rule, the following formulation always holds true:

$$\begin{aligned}
 p(\beta, \lambda, \rho, V, \sigma^2, \gamma, \theta, U, U_0 | y) & \propto p(y) \\
 & = p(y | \beta, \lambda, \rho, V, \sigma^2, \gamma, \theta, U, U_0) \pi(\beta, \lambda, \rho, V, \sigma^2, \gamma, \theta, U, U_0)
 \end{aligned} \tag{3.18}$$

Here, U_0 is a vector for all individuals’ utility in the initial period, $p(\bullet)$ indicates posterior densities, and $\pi(\bullet)$ stands for prior distribution assumptions. Assuming

certain forms of independent priors, the posterior joint density

$p(\boldsymbol{\beta}, \lambda, \rho, V, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\theta}, U, U_0 | \mathbf{y})$ will exhibit the following proportionality:

$$p(\boldsymbol{\beta}, \lambda, \rho, V, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\theta}, U, U_0 | \mathbf{y}) \propto p(\mathbf{y} | U, \boldsymbol{\gamma}) \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \pi(\boldsymbol{\theta} | \rho, \sigma^2) \pi(\boldsymbol{\gamma}) \pi(U_0) \pi(\boldsymbol{\beta}) \pi(\rho) \pi(\sigma^2) \pi(\lambda) \pi(V) \quad (3.19)$$

3.3.2 A Summary of Conditional Posterior Distributions

Considering two random events A and B , and observed values y , Bayes' Rule implies that the following always holds:

$$p(A | B, y) = \frac{p(A, B | y)}{p(B | y)} \propto p(A, B | y) \quad (3.20)$$

In other words, the conditional distribution $p(A | B, y)$ is proportional to the conditional posterior distribution of (A, B) , and only terms involving A need to be extracted. Therefore, from Equation (3.19), the conditional distributions can be derived as follows, for each parameter and variable of interest. The $\boldsymbol{\Theta}_*$ in these formulations represents the set of conditional arguments for the conditional distribution of $*$. It includes all arguments except $*$. (For example, $\boldsymbol{\Theta}_\beta$ stands for $(\lambda, \rho, V, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\theta}, U, U_0, \mathbf{y})$.)

$$p(\boldsymbol{\beta} | \boldsymbol{\Theta}_\beta) \propto \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \pi(\boldsymbol{\beta}) \quad (3.21)$$

$$p(\boldsymbol{\theta} | \boldsymbol{\Theta}_\theta) \propto \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \pi(\boldsymbol{\theta} | \rho, \sigma^2) \quad (3.22)$$

$$p(\lambda | \boldsymbol{\Theta}_\lambda) \propto \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \pi(\lambda) \quad (3.23)$$

$$p(\rho | \boldsymbol{\Theta}_\rho) \propto \pi(\boldsymbol{\theta} | \rho, \sigma^2) \pi(\rho) \quad (3.24)$$

$$p(\sigma^2 | \Theta_{\sigma^2}) \propto \pi(\theta | \rho, \sigma^2) \pi(\sigma^2) \quad (3.25)$$

$$p(V | \Theta_V) \propto \pi(U | U_0, \beta, \theta, \lambda, V) \pi(V) = \pi(U | U_0, \beta, \theta, \lambda, V) \prod_{i=1}^M \pi(v_i) \quad (3.26)$$

$$p(\gamma | \Theta_\gamma) \propto p(y | U, \gamma) \pi(\gamma) \quad (3.27)$$

$$p(U_0 | \Theta_{U_0}) \propto \pi(U | U_0, \beta, \theta, \lambda, V) \pi(U_0) \quad (3.28)$$

$$p(U | \Theta_U) \propto p(y | U, \gamma) \pi(U | U_0, \beta, \theta, \lambda, V) \quad (3.29)$$

These formulations (Equations 3.21 to 3.29) all involve three factors:

$\pi(U | U_0, \beta, \theta, \lambda, V)$, $p(y | U, \gamma)$ and $\pi(\theta | \rho, \sigma^2)$. The following paragraphs discuss these three items in more detail:

From Equation (3.13), it can be observed that for all $t \neq 0, t = 1, \dots, T$,

$U_t | U_{\neq t}, \beta, \theta, \lambda, V \square N(\lambda U_{t-1} + X_t \beta + L \theta, V)$, so the conditional prior distribution can be expressed as

$$\begin{aligned} \pi(U_t | U_{\neq t}, \beta, \theta, \lambda, V) &= |V|^{-1/2} \exp \left\{ -\frac{1}{2} (U_t - \lambda U_{t-1} - L \theta - X_t \beta)' V^{-1} (U_t - \lambda U_{t-1} - L \theta - X_t \beta) \right\} \\ &= |V|^{-1/2} \exp \left\{ -\frac{1}{2} (U_t^\lambda - L \theta - X_t \beta)' V^{-1} (U_t^\lambda - L \theta - X_t \beta) \right\} \end{aligned} \quad (3.30)$$

Therefore, for $U = (U_1, U_2, \dots, U_T)'$

$$\begin{aligned} \pi(U | U_0, \beta, \theta, \lambda, V) &= \prod_{t=1}^T |V|^{-1/2} \exp \left\{ -\frac{1}{2} (U_t^\lambda - L \theta - X_t \beta)' V^{-1} (U_t^\lambda - L \theta - X_t \beta) \right\} \\ &= |\Omega|^{-1/2} \exp \left\{ -\frac{1}{2} (U^\lambda - \Delta \theta - X \beta)' \Omega^{-1} (U^\lambda - \Delta \theta - X \beta) \right\} \end{aligned} \quad (3.31)$$

Alternatively, this can be expressed as

$$\pi(U|U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, \mathbf{V}) = \prod_{i=1}^T \prod_{k=1}^M \prod_{k=1}^{n_i} \left\{ \nu_i^{-1/2} \exp \left[\frac{-1}{2\nu_i} \left(U_{ikt} - \lambda U_{ikt-1} - \mathbf{X}_{ikt}' \boldsymbol{\beta} - \theta_i \right)^2 \right] \right\} \quad (3.32)$$

$p(\mathbf{y}|U, \boldsymbol{\gamma})$ is already given by Equation (3.17) and $\boldsymbol{\theta} | (\rho, \sigma^2)$ is given by Equation (3.8), so that

$$\pi(\boldsymbol{\theta} | \rho, \sigma^2) = \sigma^{-M/2} |\mathbf{B}_\rho| \exp \left(\frac{-1}{2\sigma^2} \boldsymbol{\theta}' \mathbf{B}_\rho' \mathbf{B}_\rho \boldsymbol{\theta} \right) \quad (3.33)$$

3.3.3 Prior Distributions for All Parameters

Two types of prior distributions are commonly used in Bayesian statistics. Diffuse (also called non-informative or flat) priors reflect the notion of “letting the data speak for themselves.” The definition of “diffuse” has been argued over years, and various rules for generating diffuse priors have been proposed. In this dissertation, a prior is considered “diffuse” as long as it allows the data to dominate the location and form of the posterior distributions.

Another commonly used prior distribution is conjugate prior. Conjugate priors are designed so that posterior distributions fall within the same family of distributions as the priors, thus facilitating the derivation of posterior distributions. For conjugate priors, prior information can be viewed as “fictitious sample information in that it can be combined with the sample in exactly the same way that additional sample information would be combined.” (Koop et al., 2006, pp.23) Without prior information, “diffuse” priors are empirically very reasonable. The prior distributions for needed parameters in this dissertation are set to be diffuse and conjugate, wherever possible, and only conjugate if a diffuse prior of any possible form leads to a non-standard and non-derivable posterior distribution.

Gelman et al. (1995) pointed out that the use of conjugate priors means that the resulting posteriors are weighted averages of standard maximum likelihood

estimators and prior mean values. And Geweke (1993) and LeSage (1999) provided standard prior distributions for linear models. In this study, most of the priors take the forms provided by Smith and LeSage (2002), while others are similar to work by Girard and Parent (2001).

For example, the parameter set β is assumed to enjoy a normal conjugate prior

$$\beta \square N(\mathbf{c}, \mathbf{H}) \quad (3.34)$$

where $\mathbf{H} = h\mathbf{I}_Q$. If \mathbf{c} is small enough and h goes to infinity, this prior becomes diffuse. Of course, if one has valid reasons for specifying other values of \mathbf{c} and h , it can be very helpful, especially with small sample sizes. Estimation may be improved through experience and intuition, which can impact selection of priors.

It is similar with the threshold parameters, where it is assumed that

$$\boldsymbol{\gamma} \square N(\mathbf{q}, \mathbf{G}) \square \mathcal{D}(\gamma_1 < \gamma_2 < \dots < \gamma_{S-1}) \quad (3.35)$$

where \mathbf{q} is a $S \times 1$ vector, with elements γ_{s0} . \mathbf{G} is a diagonal matrix, with elements g_s on its diagonal and zeros elsewhere. In this way, the threshold parameters also follow a normal conjugate prior, only now with one more constraint to ensure that all probabilities derived from these thresholds are positive. So as \mathbf{q} approaches zero and g_s approaches infinity, this also becomes a diffuse prior.

The variance of regional effects σ^2 and the variance of individual effects ν_i are assumed to be conjugate inverse-gamma priors:

$$1/\sigma^2 \square \Gamma(\alpha, \tau) \quad (3.36)$$

More specifically, σ^2 is given a diffuse prior by setting parameters $\alpha = \tau = 0$. All v_i are assumed to follow an inverse chi-square distribution with hyperparameter ϖ , which is a special case of the inverse gamma:

$$r/v_i \square \chi^2(\varpi) \tag{3.37}$$

Here, the spatial autocorrelation coefficient ρ is given a uniform prior that is diffuse. As Sun et al. (1999) prove, the lower and upper bounds for ρ are determined by the inverse of eigenvalues from weight matrix W . Let ς_{\min} and ς_{\max} denote the minimum and maximum eigenvalues; then

$$\rho \square U[\varsigma_{\min}^{-1}, \varsigma_{\max}^{-1}] \tag{3.38}$$

In other words,

$$\pi(\rho) \propto 1 \tag{3.39}$$

λ is specified to have a normal distribution but limited to the range $(-1, 1)$ in order to ensure stationarity, as discussed in Section 3.2.3:

$$\lambda \square N(\lambda_0, D) \square \mathcal{B}(|\lambda| < 1) \tag{3.40}$$

Selection of an initial value for the latent variable U is termed the “initial conditional problem.” Many have discussed this complicated issue (e.g., Vishniac, 1993; Wooldrige, 2005; and Barlevy and Nagaraja, 2006), and there are two ways to specify the initial condition. One is to give non-stochastic values to these initial utility values. This assumption, however, is very strong, implying that initial values are independent of all other conditions (such as the unobserved regional effects and heteroskedastic individual effects, both of which should exist from the very beginning, in the “initial state”). Another, more flexible approach is to assign

a distribution for initial values. Here U_0 is assumed to be normally distributed, in order to be compatible with U 's distribution in other periods. It has the following prior:

$$U_0 \square N(a_0 \mathbf{I}_N, d_0 \mathbf{I}_N) \quad (3.41)$$

where \mathbf{I}_N is a $N \times 1$ vector with all elements equal to 1 and \mathbf{I}_N is an N -dimension identity matrix. Therefore, this distribution approximates a diffuse prior when a_0 is bounded and d_0 goes to infinity.

3.3.4 Full Conditional Posterior Distributions

Sections 3.3.2 and 3.3.3 introduce calculations for conditional posterior distributions and list the parameters' prior distributions. This section shows in detail how each parameter's conditional posterior distribution can be mathematically derived. Each of the following sub-sections focuses on one parameter (or one variable of interest) and discusses its posterior distribution, hyperparameters, and links to MLE estimators. If the parameter has a non-standard posterior distribution, the numerical method for generating random numbers from that distribution is also briefly explained.

3.3.4.1 Conditional Posterior Distribution of β

From Equations (3.21) and (3.34), it can be derived that

$$\begin{aligned} p(\beta | \Theta_\beta) &\propto \pi(U | U_0, \beta, \theta, \lambda, V) \square \pi(\beta) \\ &\propto \exp\left\{-\frac{1}{2}(\beta - c)' H^{-1}(\beta - c)\right\} \square \\ &\quad \exp\left\{-\frac{1}{2}(U^\lambda - \Delta\theta - X\beta)' \Omega^{-1}(U^\lambda - \Delta\theta - X\beta)\right\} \end{aligned} \quad (3.42)$$

As many previous studies show (e.g., Gelman et al., 1995; and Smith and LeSage, 2002) and as described in Appendix A, this form can be simplified to

$$p(\boldsymbol{\beta}|\boldsymbol{\Theta}_\beta) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{A}^{-1}\mathbf{b})' \mathbf{A}(\boldsymbol{\beta} - \mathbf{A}^{-1}\mathbf{b})\right] \quad (3.43)$$

$$\text{where } \mathbf{A} = \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} + \mathbf{H}^{-1} \quad (3.44)$$

$$\text{and } \mathbf{b} = \mathbf{X}'\boldsymbol{\Omega}^{-1}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta}) + \mathbf{H}^{-1}\mathbf{c}. \quad (3.45)$$

These equations indicate that the posterior mean vector for $\boldsymbol{\beta}$ is $\mathbf{A}^{-1}\mathbf{b}$ and the variance-covariance matrix is \mathbf{A}^{-1} . In fact, as Gelman et al. (1995) show, such a posterior distribution is a weighted average of $\boldsymbol{\beta}$'s prior distribution and sample data information and the weights are the *inverse* of the variance-covariance matrices or associated “uncertainty” levels. Using maximum likelihood estimation methods, the estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{MLE} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta}) \quad (3.46)$$

Here, the prior mean of $\boldsymbol{\beta}$ is assumed to be \mathbf{c} and its prior variance is assumed to be \mathbf{H} . It is not difficult to show that the posterior mean can then be written as follows:

$$\begin{aligned} E(\boldsymbol{\beta}|\boldsymbol{\Theta}_\beta) &= \mathbf{A}^{-1}\mathbf{b} \\ &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} + \mathbf{H}^{-1})^{-1} [\mathbf{X}'\boldsymbol{\Omega}^{-1}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta}) + \mathbf{H}^{-1}\mathbf{c}] \\ &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} + \mathbf{H}^{-1})^{-1} [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{H}^{-1}\mathbf{c}] \end{aligned} \quad (3.47)$$

In Equation (3.47), as sample size and information quality increase, the variance $\boldsymbol{\Omega}$ should decrease, which allows $\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}$ to dominate, giving $\hat{\boldsymbol{\beta}}_{MLE}$ more weight.

3.3.4.2 Conditional Posterior Distribution of θ

Some manipulation of Equations (3.22) and (3.33) can show that

$$\begin{aligned}
p(\theta|\Theta_\theta) &\propto \pi(U|U_0, \beta, \theta, \lambda, V) \pi(\theta|\rho, \sigma^2) \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{U}^\lambda - \Delta\theta - \mathbf{X}\beta)' \Omega^{-1}(\mathbf{U}^\lambda - \Delta\theta - \mathbf{X}\beta)\right\} \exp\left(\frac{-1}{2\sigma^2} \theta' \mathbf{B}'_\rho \mathbf{B}_\rho \theta\right) \\
&= \exp\left\{-\frac{1}{2}\left[\theta' (\sigma^{-2} \mathbf{B}'_\rho \mathbf{B}_\rho) \theta + \theta' \Delta' \Omega^{-1} \Delta \theta - 2(\mathbf{U}^\lambda - \mathbf{X}\beta) \Omega^{-1} \Delta \theta + \mathbf{C}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta' (\sigma^{-2} \mathbf{B}'_\rho \mathbf{B}_\rho + \Delta' \Omega^{-1} \Delta) \theta - 2(\mathbf{U}^\lambda - \mathbf{X}\beta) \Omega^{-1} \Delta \theta\right]\right\} \quad (3.48)
\end{aligned}$$

where \mathbf{C} stands for the constant term, which does not involve θ . Similar to the derivation of the conditional posterior distribution for β , it can be shown that

$$p(\theta|\Theta_\theta) \propto \exp\left[-\frac{1}{2}(\theta - \mathbf{A}_\theta^{-1} \mathbf{b}_\theta)' \mathbf{A}_\theta (\theta - \mathbf{A}_\theta^{-1} \mathbf{b}_\theta)\right] \quad (3.49)$$

$$\text{where } \mathbf{A}_\theta = \sigma^{-2} \mathbf{B}'_\rho \mathbf{B}_\rho + \Delta' \Omega^{-1} \Delta \quad (3.50)$$

$$\text{and } \mathbf{b}_\theta = \Delta' \Omega^{-1} (\mathbf{U}^\lambda - \mathbf{X}\beta). \quad (3.51)$$

These equations indicate that the mean vector for θ is $\mathbf{A}_\theta^{-1} \mathbf{b}_\theta$ and the variance-covariance matrix is \mathbf{A}_θ^{-1} . It should be noticed here, however, \mathbf{A}_θ depends on \mathbf{B}_ρ , which depends on ρ . That is, each random draw involves a matrix inversion. This computation demands much memory, especially when the number of regions (M) is large. Therefore, an appropriate sampling approach is very important. There are two alternative ways to calculate this matrix inverse. One is to compute the inverse directly; the other way, as Smith and LeSage (2002) suggest (when M is larger), is to sample from univariate normal distributions for each θ_i conditional on all other elements of θ (excluding the i^{th} element).

3.3.4.3 Conditional Posterior Distribution of λ

From Equations (3.23), (3.32), and (3.40), one can obtain the full form of λ 's conditional posterior distribution, written as follows:

$$\begin{aligned}
 p(\lambda|\boldsymbol{\Theta}_\lambda) &\propto \pi(\mathbf{U}|U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, \mathbf{V}) \square \pi(\lambda) \\
 &\propto \exp\left\{-\frac{1}{2} \sum_{t=1}^T (\mathbf{U}_t - \lambda \mathbf{U}_{t-1} - \mathbf{L}\boldsymbol{\theta} - \mathbf{X}_t \boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{U}_t - \lambda \mathbf{U}_{t-1} - \mathbf{L}\boldsymbol{\theta} - \mathbf{X}_t \boldsymbol{\beta})\right\} \square \\
 &\quad \exp\left\{-\frac{1}{2} (\lambda - \lambda_0)' \mathbf{D}^{-1} (\lambda - \lambda_0)\right\} \square \mathcal{D}(|\lambda| < 1)
 \end{aligned} \tag{3.52}$$

This is another conjugate distribution; so, similar to $\boldsymbol{\beta}$'s conditional posterior distribution:

$$p(\lambda|\boldsymbol{\Theta}_\lambda) \propto \exp\left[-\frac{1}{2} (\lambda - A_\lambda^{-1} b_\lambda)' A_\lambda (\lambda - A_\lambda^{-1} b_\lambda)\right] \square \mathcal{D}(|\lambda| < 1) \tag{3.53}$$

$$\text{where } A_\lambda = \sum_{t=1}^T \mathbf{U}'_{t-1} \mathbf{V}^{-1} \mathbf{U}_{t-1} + \mathbf{D}^{-1} \tag{3.54}$$

$$\text{and } b_\lambda = \sum_{t=1}^T \mathbf{U}'_{t-1} \mathbf{V}^{-1} (\mathbf{U}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{L}\boldsymbol{\theta}) + \mathbf{D}^{-1} \lambda_0. \tag{3.55}$$

One evident difference between this distribution of λ and the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is that this is a truncated normal. In each draw, the value of λ needs to be limited to $(-1, 1)$.

3.3.4.4 Conditional Posterior Distribution of ρ

Equations (3.24), (3.33) and (3.39) lead to the following formulation for ρ 's conditional posterior distribution:

$$p(\rho|\boldsymbol{\Theta}_\rho) \propto \pi(\boldsymbol{\theta}|\rho, \sigma^2) \square \pi(\rho)$$

$$\propto |\mathbf{B}_\rho| \exp\left(\frac{-1}{2\sigma^2} \boldsymbol{\theta}' \mathbf{B}'_\rho \mathbf{B}_\rho \boldsymbol{\theta}\right) \quad (3.56)$$

and $\rho \in [\zeta_{\min}^{-1}, \zeta_{\max}^{-1}]$. As Smith and LeSage (2002) point out, this expression cannot be simplified into a standard distribution. They further suggest that one may use univariate numerical integration to obtain this posterior density, as described below:

First, a range of ρ values between $[\zeta_{\min}^{-1}, \zeta_{\max}^{-1}]$ is generated from a uniform distribution. Before MCMC sampling, a vector of determinant values for \mathbf{B}_ρ corresponding to this range of ρ values can be constructed. Thus, during the iterative sampling process, only the second item ($\exp\left(\frac{-1}{2\sigma^2} \boldsymbol{\theta}' \mathbf{B}'_\rho \mathbf{B}_\rho \boldsymbol{\theta}\right)$) needs to be updated for each draw. Equation (3.56) is then numerically integrated (via a sum of point-area estimates) over the range of ρ values. The normalizing constant is obtained, given the condition that ρ is limited to the interval $[\zeta_{\min}^{-1}, \zeta_{\max}^{-1}]$, and this renders Equation (3.56)'s proportionality an equality. After this approximation for ρ 's CDF is acquired, one can randomly draw the ρ value from its inversion. As Smith and LeSage (2002) have suggested, the advantage of this approach (over a standard Metropolis-Hastings approach) is that it is more efficient: each pass through the sampler produces a draw for ρ .

3.3.4.5 Conditional Posterior Distribution of σ^2

From Equations (3.25), (3.33) and (3.36), the following distribution for σ^2 can be obtained:

$$p(\sigma^2 | \boldsymbol{\theta}_{\sigma^2}) \propto \pi(\boldsymbol{\theta} | \rho, \sigma^2) \pi(\sigma^2)$$

$$\begin{aligned}
& \propto (\sigma^2)^{-M/2} \exp\left(\frac{-1}{2\sigma^2} \boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta}\right) (\sigma^2)^{-(\alpha+1)} \exp\left(\frac{\tau}{\sigma^2}\right) \\
& = (\sigma^2)^{-(M/2+\alpha+1)} \exp\left(-\frac{\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau}{2\sigma^2}\right)
\end{aligned} \tag{3.57}$$

This is an inverse gamma distribution with shape parameter $-M/2 + \alpha$ and scale parameter $(\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)/2$.

Letting $\kappa = (\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)/\sigma^2$, so that $\sigma^2 = (\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)/\kappa$, and following the work of Geweke (1993), Equation (3.57) can be expressed as follows:

$$\begin{aligned}
p(\sigma^2 | \boldsymbol{\Theta}_{\sigma^2}) & \propto \left((\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)/\kappa \right)^{-(M/2+\alpha+1)} \exp\left(-\frac{\kappa}{2}\right) \left| \frac{d\sigma^2}{d\kappa} \right| \\
& = \left((\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)/\kappa \right)^{-(M/2+\alpha+1)} \exp\left(-\frac{\kappa}{2}\right) \left| \frac{(\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau)}{\kappa^2} \right| \\
& \propto \kappa^{(M/2+\alpha-1)} \exp\left(-\frac{\kappa}{2}\right)
\end{aligned} \tag{3.58}$$

This density is proportional to a chi-square density with $M + 2\alpha$ degrees of freedom (DOF). Alternatively, the conditional posterior of σ^2 can be expressed as

$$\frac{\boldsymbol{\theta}' \mathbf{B}'_{\rho} \mathbf{B}_{\rho} \boldsymbol{\theta} + 2\tau}{\sigma^2} \Big| \boldsymbol{\Theta}_{\sigma^2} \propto \chi^2(M + 2\alpha) \tag{3.59}$$

3.3.4.6 Conditional Posterior Distribution of V

From Equations (3.26) and (3.37), it can be shown that

$$p(V | \boldsymbol{\Theta}_V) \propto \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \prod_{i=1}^M \pi(v_i)$$

$$\propto |\mathbf{\Omega}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta})' \mathbf{\Omega}^{-1}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta})\right\} \prod_{i=1}^M \pi(v_i) \quad (3.60)$$

By letting $\mathbf{e} = \mathbf{U}^\lambda - \Delta\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta}$, the distribution of V can also be derived term by term for each i :

$$\begin{aligned} p(v_i | \boldsymbol{\theta}_{v_i}) &\propto |\mathbf{\Omega}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{e}' \mathbf{\Omega}^{-1} \mathbf{e}\right) \pi(v_i) \\ &= \prod_{j=1}^M \left[v_j^{-n_j T/2} \exp\left(-\sum_{t=1}^T \sum_{j=1}^M \frac{\mathbf{e}_{jt}' \mathbf{e}_{jt}}{2v_j}\right) v_j^{-\left(\frac{\varpi}{2}+1\right)} \exp\left(-\frac{\varpi}{2v_j}\right) \right] \\ &\propto v_i^{-n_i T/2} \exp\left(-\sum_{t=1}^T \frac{\mathbf{e}_{it}' \mathbf{e}_{it}}{2v_i}\right) v_i^{-\left(\frac{\varpi}{2}+1\right)} \exp\left(-\frac{\varpi}{2v_i}\right) \quad (3.61) \\ &= v_i^{-\left(\frac{\varpi+n_i T}{2}+1\right)} \exp\left(-\frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{2v_i}\right) \end{aligned}$$

Similar to the derivation of σ^2 's posterior distribution, letting $\kappa_i = \frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{v_i}$,

then this can be shown as the following chi-square distribution:

$$\begin{aligned} p(v_i | \boldsymbol{\theta}_{v_i}) &= \left(\frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{\kappa_i} \right)^{-\left(\frac{\varpi+n_i T}{2}+1\right)} \exp\left(-\frac{\kappa_i}{2}\right) \frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{\kappa_i^2} \\ &= \kappa_i^{-\left(\frac{\varpi+n_i T}{2}+1\right)} \exp\left(-\frac{\kappa_i}{2}\right) \quad (3.62) \end{aligned}$$

Therefore, it follows a chi-square density with $r + n_i T$ degrees of freedom, i.e.,

$$\frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{\nu_i} \Bigg| \boldsymbol{\Theta}_{\nu_i} \square \chi^2(\varpi + n_i T) \quad (3.63)$$

Similar to $\boldsymbol{\beta}$, Smith and LeSage (2002) show that the posterior mean of ν_i is a weighted average of the maximum likelihood estimator $\hat{\nu}_i$ and the prior mean, μ_i , which equals $\varpi/(\varpi - 2)$. In the dynamic model, the weights can be calculated using a method very similar to that suggested by Smith and LeSage (2002). These weights are $n_i T$ and $\varpi - 2$, respectively.

As expected, this means that more weight is given to the sample information as sample size n_i or the panel length, T , increases. ϖ needs to be larger than 2, but also needs to be kept small if one wants to use a diffuse prior. Here the hyperparameter ϖ is assumed to be 4.

3.3.4.7 Conditional Posterior Distribution of $\boldsymbol{\gamma}$

Equations (3.17), (3.27), and (3.35) lead to the following formulation for the conditional posterior distribution of $\boldsymbol{\gamma}$:

$$p(\boldsymbol{\gamma} | \boldsymbol{\Theta}_{\boldsymbol{\gamma}}) \propto p(\mathbf{y} | \mathbf{U}, \boldsymbol{\gamma}) \square \pi(\boldsymbol{\gamma})$$

$$\propto \left[\prod_{t=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \prod_{s=1}^S \delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s) \right] \square$$

$$N(\mathbf{q}, \mathbf{G}) \square \mathcal{D}(\gamma_1 < \gamma_2 < \dots < \gamma_{S-1}) \quad (3.64)$$

This equation can be considered term by term for $s = 1, \dots, S - 1$, by only extracting terms that involve γ_s :

$$\begin{aligned}
p(\gamma_s | \boldsymbol{\Theta}_{r_s}) &\propto \prod_{t=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \delta(U_{ikt} < \gamma_s | y_{ikt} = s) \prod_{t=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \delta(U_{ikt} > \gamma_s | y_{ikt} = s+1) \prod \\
&\quad N(\gamma_{s0}, \mathbf{g}_s) \prod \delta(\gamma_{s-1} < \gamma_s < \gamma_{s+1}) \\
&= \delta(\gamma_s^{\text{inf}} < \gamma_s < \gamma_s^{\text{sup}}) \prod \exp\left\{-\frac{1}{2\mathbf{g}_s}(\gamma_s - \gamma_{s0})^2\right\} \quad (3.65)
\end{aligned}$$

$$\text{With } \gamma_s^{\text{inf}} = \max\{\max\{U_{ikt} : y_{ikt} = s\}; \gamma_{s-1}\} \quad (3.66)$$

$$\text{and } \gamma_s^{\text{sup}} = \min\{\min\{U_{ikt} : y_{ikt} = s+1\}; \gamma_{s+1}\}, \text{ for } \forall i \in M, k \in n_i, t \in T. \quad (3.67)$$

Similar to the derivation for λ , this is a truncated normal distribution. The normalizing constant can be found using a univariate normal distribution, with the given lower and upper bounds. The major difference is, however, that these lower and upper bounds are interdependent, which may make the final posterior distribution multimodal.

3.3.4.8 Conditional Posterior Distribution of U_0

Substituting Equations (3.32) and (3.41) into Equation (3.28), one can get the following formulation:

$$\begin{aligned}
p(U_0 | \boldsymbol{\Theta}_{U_0}) &\propto \pi(\mathbf{U} | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, \mathbf{V}) \prod \pi(U_0) \\
&\propto \prod_{i=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \left\{ \exp\left[-\frac{1}{2\nu_i}(U_{ikt} - \lambda U_{ikt-1} - \theta_i - \mathbf{X}_{ikt}\boldsymbol{\beta})^2\right] \prod \exp\left[-\frac{1}{2d_0}(U_{ik0} - a_0)^2\right] \right\} \\
&\propto \prod_{i=1}^M \prod_{k=1}^{n_i} \exp\left[-\frac{(U_{ik1} - \lambda U_{ik0} - \theta_i - \mathbf{X}_{ik1}\boldsymbol{\beta})^2}{2\nu_i} - \frac{(U_{ik0} - a_0)^2}{2d_0}\right] \quad (3.68)
\end{aligned}$$

Deriving U_0 term by term for each i and k , by extracting only items involving U_{ik0} , Equation (3.68) reduces to:

$$p(U_{ik0} | \boldsymbol{\theta}_{U_{ik0}}) \propto \exp \left[-\frac{(U_{ik1} - \lambda U_{ik0} - \theta_i - \mathbf{X}_{ik1} \boldsymbol{\beta})^2}{2v_i} - \frac{(U_{ik0} - a_0)^2}{2d_0} \right] \quad (3.69)$$

This is a univariate normal distribution. Thus, similar to the posterior distribution calculations for $\boldsymbol{\beta}$, the distribution is as follows:

$$U_{ik0} | \boldsymbol{\theta}_{U_{ik0}} \square N(A_{U0}^{-1} b_{U0}, A_{U0}^{-1}) \quad (3.70)$$

$$\text{where } A_{U0} = \lambda^2 v_i^{-1} + d_0^{-1} \quad (3.71)$$

$$\text{and } b_{U0} = \lambda v_i^{-1} (U_{ik1} - \theta_i - \mathbf{X}_{ik1} \boldsymbol{\beta}) + d_0^{-1} a_0. \quad (3.72)$$

3.3.4.9 Conditional Posterior Distribution of U

For latent variables other than the initial status, Equations (3.17), (3.29), and (3.32) lead to the following formulation:

$$p(U | \boldsymbol{\theta}_U) \propto p(\mathbf{y} | U, \boldsymbol{\gamma}) \square \pi(U | U_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \lambda, V) \quad (3.73)$$

$$\propto \prod_{t=1}^T \prod_{i=1}^M \prod_{k=1}^{n_i} \left\{ \left[\sum_{s=1}^S \delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s) \right] \square \left[v_i^{-1/2} \square \exp \left[-\frac{1}{2v_i} (U_{ikt} - \lambda U_{ikt-1} - \theta_i - \mathbf{X}_{ikt} \boldsymbol{\beta})^2 \right] \right] \right\}$$

U_{ikt} appears in the formulation for both periods t and $t+1$. Therefore, for any i, k, t observation, by extracting only items involving U_{ikt} , the posterior distribution for U_{ikt} can be expressed as follows:

$$p(U_{ikt} | \boldsymbol{\theta}_{U_{ikt}}) \propto \left\{ \sum_{s=1}^S [\delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s)] \right\} \square v_i^{-1} \square \exp \left\{ -\frac{1}{2v_i} \left[(U_{ikt} - \lambda U_{ikt-1} - \theta_i - \mathbf{X}_{ikt} \boldsymbol{\beta})^2 + (U_{ikt+1} - \lambda U_{ikt} - \theta_i - \mathbf{X}_{ikt+1} \boldsymbol{\beta})^2 \right] \right\} \quad (3.74)$$

This is a truncated normal distribution. The first expression in Equation (3.74),

$$\sum_{s=1}^S [\delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s)]$$

indicates that if $y_{ikt} = s$, the distribution is truncated on the left by γ_{s-1} and on the right by γ_s . The last item in Equation (3.74),

$$\exp \left\{ -\frac{1}{2v_i} \left[(U_{ikt} - \lambda U_{ikt-1} - \theta_i - \mathbf{X}_{ikt} \boldsymbol{\beta})^2 + (U_{ikt+1} - \lambda U_{ikt} - \theta_i - \mathbf{X}_{ikt+1} \boldsymbol{\beta})^2 \right] \right\}$$

suggests that the un-truncated part is a normal distribution. This part has mean a_{ikt} and variance b_{ikt} . (Readers may wish to see Appendix B for more details on this.)

$$\text{Here, } a_{ikt} = \left[\lambda U_{ikt+1} + \lambda U_{ikt-1} + (1-\lambda)\theta_i + (\mathbf{X}_{ikt} - \lambda \mathbf{X}_{ikt+1}) \boldsymbol{\beta} \right] / (1 + \lambda^2) \quad (3.75)$$

$$\text{and } b_{ikt} = v_i / (1 + \lambda^2). \quad (3.76)$$

Therefore, for each i, k and each $t = 1, \dots, T-1$,

$$U_{ikt} | \boldsymbol{\theta}_{U_{ikt}} \square N(a_{ikt}, b_{ikt}) \square \sum_{s=1}^S [\delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s)] \quad (3.77)$$

A special case is that, when $t = T$, U_{ikT} only appears in the exponential term with U_{ikT-1} . That is,

$$p(U_{ikT} | \boldsymbol{\theta}_{U_{ikT}}) \propto \left\{ \sum_{s=1}^S [\delta(y_{ikT} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikT} < \gamma_s)] \right\} v_i^{-1/2} \square \exp \left[-\frac{1}{2v_i} (U_{ikT} - \lambda U_{ikT-1} - \theta_i - \mathbf{X}_{ikT} \boldsymbol{\beta})^2 \right] \quad (3.78)$$

This also is a truncated normal distribution. The (un-truncated) normal distribution has a mean $a_{ikT} = \lambda U_{ikT-1} + \theta_i + \mathbf{X}_{ikT} \boldsymbol{\beta}$ and variance v_i ; and, if $y_{ikT} = s$, the distribution is truncated on the left by γ_{s-1} and on the right by γ_s .

3.3.5 Summary of Parameter Distributions

Section 3.3.3 discussed how each parameter's prior distribution was selected. Section 3.3.4 illustrated how the conditional posterior distributions were mathematically developed. Table 3.1 effectively summarizes the content of these two sections by listing all prior and posterior distributions of the parameters and variables. As the table suggests, most of these distributions follow standard distributions and can be conveniently generated using routines built in commercial mathematical analysis packages (such as Matlab and Gauss). The spatial coefficient ρ , however, follows a non-standard posterior distribution and has to be generated using numerical methods. The threshold parameter $\boldsymbol{\gamma}$ follows a multidimensional truncated normal distribution and the truncations co-vary. Therefore, the marginal distribution of each element in $\boldsymbol{\gamma}$ is also expected to look non-standard.

Table 3.1 Conditional Distributions of DSOP's Parameters

	Prior Distributions	Conditional Posterior Distributions	Hyperparameters
β	$N(\mathbf{c}, \mathbf{H})$	$\beta \Theta_\beta \square N(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1})$	$\mathbf{A} = \mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X} + \mathbf{H}^{-1}$ $\mathbf{b} = \mathbf{X}'\mathbf{\Omega}^{-1}(\mathbf{U}^\lambda - \Delta\boldsymbol{\theta}) + \mathbf{H}^{-1}\mathbf{c}$
$\boldsymbol{\theta}$	$N\left[\mathbf{0}, \sigma^2 (\mathbf{B}'_\rho \mathbf{B}_\rho)^{-1}\right]$	$\boldsymbol{\theta} \Theta_\theta \square N(\mathbf{A}_\theta^{-1}\mathbf{b}_\theta, \mathbf{A}_\theta^{-1})$	$\mathbf{A}_\theta = \sigma^{-2} \mathbf{B}'_\rho \mathbf{B}_\rho + \Delta' \mathbf{\Omega}^{-1} \Delta$ $\mathbf{b}_\theta = \Delta' \mathbf{\Omega}^{-1} (\mathbf{U}^\lambda - \mathbf{X}\boldsymbol{\beta})$
λ	$N(\lambda_0, D) \square \mathcal{D}(\lambda < 1)$	$\lambda \Theta_\lambda \square N(\mathbf{A}_\lambda^{-1}\mathbf{b}_\lambda, \mathbf{A}_\lambda) \square \mathcal{D}(\lambda < 1)$	$\mathbf{A}_\lambda = \sum_{t=1}^T \mathbf{U}_{t-1}' \mathbf{V}^{-1} \mathbf{U}_{t-1} + D^{-1}$ $\mathbf{b}_\lambda = \sum_{t=1}^T \mathbf{U}_{t-1}' \mathbf{V}^{-1} (\mathbf{U}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{L}\boldsymbol{\theta}) + D^{-1} \lambda_0$
ρ	$U\left[\zeta_{\min}^{-1}, \zeta_{\max}^{-1}\right]$	$p(\rho \Theta_\rho) \propto \mathbf{B}_\rho \square \exp\left(\frac{-1}{2\sigma^2} \boldsymbol{\theta}' \mathbf{B}'_\rho \mathbf{B}_\rho \boldsymbol{\theta}\right)$ (Non-standard distribution)	
σ^2	$\Gamma(\alpha, \tau)^{-1}$	$\frac{\boldsymbol{\theta}' \mathbf{B}'_\rho \mathbf{B}_\rho \boldsymbol{\theta} + 2\tau}{\sigma^2} \mid \Theta_{\sigma^2} \square \chi^2(M + 2\alpha)$	
ν_i	$\chi^2(\varpi)^{-1}$	$\frac{\sum_{t=1}^T \mathbf{e}_{it}' \mathbf{e}_{it} + \varpi}{\nu_i} \mid \Theta_{\nu_i} \square \chi^2(\varpi + n_i T)$	
$\boldsymbol{\gamma}$	$N(\mathbf{q}, \mathbf{G}) \square \mathcal{D}(\gamma_1 < \gamma_2 < \dots < \gamma_{S-1})$	$\gamma_s \Theta_{\gamma_s} \square \delta(\gamma_s^{\text{inf}} < \gamma_s < \gamma_s^{\text{sup}}) \square N(\gamma_{s0}, \mathbf{g}_s)$	$\gamma_s^{\text{inf}} = \max\{\max\{U_{ikt} : y_{ikt} = s\}; \gamma_{s-1}\}$ $\gamma_s^{\text{sup}} = \min\{\min\{U_{ikt} : y_{ikt} = s+1\}; \gamma_{s+1}\}$
\mathbf{U}_0	$N(\mathbf{a}_0 \mathbf{I}_N, d_0 \mathbf{I}_N)$	$U_{ik0} \Theta_{U_{ik0}} \square N(\mathbf{A}_{U_0}^{-1} \mathbf{b}_{U_0}, \mathbf{A}_{U_0}^{-1})$	$\mathbf{A}_{U_0} = \lambda^2 \nu_i^{-1} + d_0^{-1}$ $\mathbf{b}_{U_0} = \lambda \nu_i^{-1} (\mathbf{U}_{ik1} - \theta_i - \mathbf{X}_{ik1} \boldsymbol{\beta}) + d_0^{-1} \mathbf{a}_0$
$\mathbf{U}_t, t \neq 0$	$N(\lambda \mathbf{U}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{L}\boldsymbol{\theta}, \mathbf{V})$	$U_{ikt} \Theta_{U_{ikt}} \square N(\mathbf{a}_{ikt}, \mathbf{b}_{ikt}) \square \sum_{s=1}^S \delta(y_{ikt} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikt} < \gamma_s)$	$a_{ikt} = (\lambda U_{ikt+1} + \lambda U_{ikt-1} + (1-\lambda)\theta_i + (\mathbf{X}_{ikt} - \lambda \mathbf{X}_{ikt+1}) \boldsymbol{\beta}) / (1 + \lambda^2)$ $b_{ikt} = \nu_i / (1 + \lambda^2)$
		$U_{ikT} \Theta_{U_{ikT}} \square N(\mathbf{a}_{ikT}, \nu_i) \square \sum_{s=1}^S \delta(y_{ikT} = s) \square \mathcal{D}(\gamma_{s-1} < U_{ikT} < \gamma_s)$	$a_{ikT} = \lambda U_{ikT-1} + \theta_i + \mathbf{X}_{ikT} \boldsymbol{\beta}$

3.4 MCMC SAMPLING

The MCMC sampling process begins with an initial parameter set $(\boldsymbol{\beta}^0, \lambda^0, \rho^0, \sigma^0, V^0, \boldsymbol{\gamma}^0, U^0)$, where superscripts indicate the current number of draws, or iteration step (for value updating.) All parameters or variables of interest are sampled sequentially, from the following conditional distributions.¹⁰

Step 1. $p(\boldsymbol{\beta} | \lambda^0, \boldsymbol{\theta}^0, \rho^0, \sigma^0, V^0, \boldsymbol{\gamma}^0, \mathbf{y}, U^0)$. This is a multivariate normal distribution of dimension Q , which leads to a weighted average of the maximum likelihood estimator and prior values. With all initial values and Equation (3.43), the value for parameter $\boldsymbol{\beta}$ can be updated to $\boldsymbol{\beta}^1$.

Step 2. $p(\boldsymbol{\theta} | \boldsymbol{\beta}^1, \lambda^0, \rho^0, \sigma^0, V^0, \boldsymbol{\gamma}^0, \mathbf{y}, U^0)$. This is a multivariate normal distribution of dimension M . With updated value $\boldsymbol{\beta}^1$ and all other, initial values, Equation (3.49) can be used to update the value of $\boldsymbol{\theta}$ to get $\boldsymbol{\theta}^1$.

Step 3. $p(\lambda | \boldsymbol{\beta}^1, \boldsymbol{\theta}^1, \rho^0, \sigma^0, V^0, \boldsymbol{\gamma}^0, \mathbf{y}, U^0)$. This is a truncated univariate normal distribution. Equation (3.53) can be used to update its value to λ^1 .

Step 4. $p(\rho | \boldsymbol{\beta}^1, \boldsymbol{\theta}^1, \lambda^1, \sigma^0, V^0, \boldsymbol{\gamma}^0, \mathbf{y}, U^0)$. Since this is not a standard distribution, the method of univariate numerical integration is used to find the normalizing constant, and then random numbers are drawn from the numerically approximated distribution.

¹⁰ This process is coded in Matlab (Mathworks, 2006). For generating (pseudo) random multivariate normal and chi-square vectors, routines from LeSage (1999)'s spatial econometric toolbox were used. The generation of random multivariate normal vectors involves Cholesky decomposition, which decomposes the variance-covariance matrix.

Step 5. $p(\sigma|\beta^1, \theta^1, \lambda^1, \rho^1, V^0, \gamma^0, \mathbf{y}, U^0)$. As Equation (3.59) indicates, this variable can be generated using a chi square distribution with $(M + 2\alpha)$ degrees of freedom. Using the updated values from the above steps and other initial values, σ can be updated to σ^1 .

Step 6. $p(v_i|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, v_{-i}, \gamma^0, \mathbf{y}, U^0)$. Each v_i follows a chi square distribution with $r + n_i T$ degrees of freedom. Equation (3.63) can be used to update these values.

Step 7. $p(\gamma^1|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, V^1, \mathbf{y}, U^0)$. This is a set of S univariate truncated normal distributions. For each γ_s ($s = 1, 2, \dots, S$), the truncated normal distribution can be normalized thanks to information on the lower and upper bounds.

Step 8. $p(U|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, V^1, \gamma^1, \mathbf{y})$. The updating process for latent variables U contains three steps:

Step 8.1. $p(U_0|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, V^1, \gamma^1, \mathbf{y})$. First, the initial utility values are updated, using the a set of N univariate normal distributions, based on Equation (3.70).

Step 8.2. $p(U_t|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, V^1, \gamma^1, \mathbf{y})$, $1 < t < T$. These are truncated normal distributions. Totally $N(T-1)$ values need to be updated, based on Equation (3.77).

Step 8.3. $p(U_T|\beta^1, \theta^1, \lambda^1, \rho^1, \sigma^1, V^1, \gamma^1, \mathbf{y})$. Equation (3.78) can be used to update values for latent variables at time period T . Totally N truncated normal distributions are used to update the U_T values.

After all these steps are completed, one needs to go back to Step 1 and use updated parameter values to replace the initial values. The whole process is

carried out iteratively by always using the most recent values of the parameters and variables, until the desired number of draws is achieved. The flowchart for sampling the parameters of interest is shown as Figure 3.1.

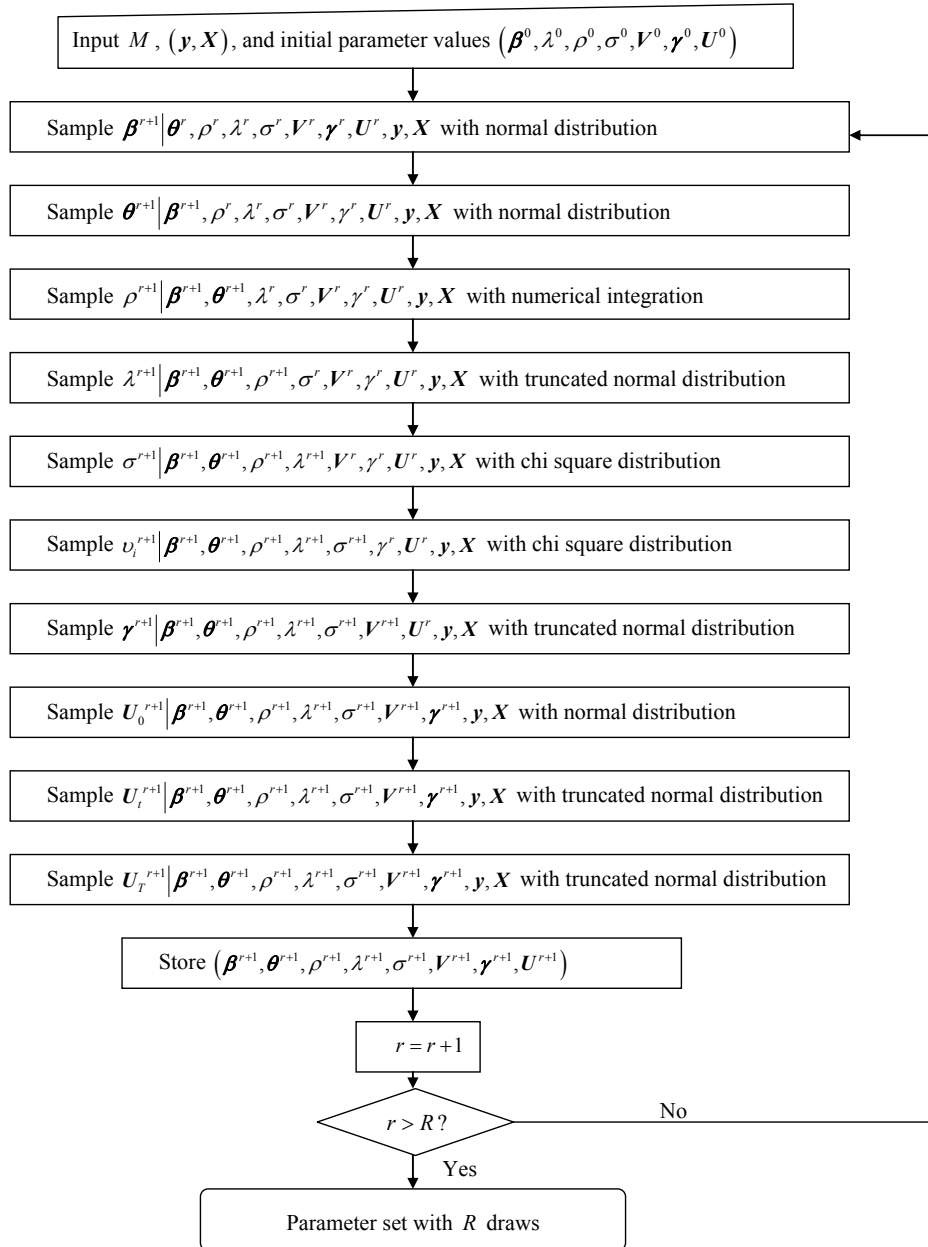


Figure 3.1 Flowchart for the MCMC Simulation

3.5 DISCUSSION OF SPECIAL CASES

Sections 3.2 through 3.4 provide details of the DSOP model's specification and estimation process. As previously mentioned, this model recognizes regional effects, spatial heterogeneity, spatial autocorrelation, and temporal autocorrelation in a latent setting with ordered categorical responses. This general framework can reduce to several simpler forms, which may be used for cases of special interest. The first special case discussed below is when the dataset exhibits no temporal autocorrelation (i.e., individuals' current responses do not rely on their previous status). The second case is that all individuals are homoskedastic. The third case is for when spatial dependencies directly occur among individuals themselves (not through regional clusters).

3.5.1 Case of No Temporal Autocorrelation

When the dataset is just cross-sectional, i.e., with no temporal autocorrelation, T equals 1 everywhere, which transforms several vectors into scalars. The subscript t and the term $-\lambda U_{t-1}$ will be erased from all formulations. In addition, discussions on λ and the special cases for U_{ikt} when $t = 0$ and $t = T$ are now unnecessary. One way to use these formulations is to simply substitute the values for λ and T as 0 and 1.

3.5.2 Homoskedastic Case

A special (and simpler) case of the model specification is homoskedastic, when all individuals' error terms have equal variances. In other words, $\nu_i \equiv \nu$.

In this setting, conditional posterior distribution for ν_i can be reduced to a

simpler form, where $\frac{T\mathbf{e}'\mathbf{e}+r}{\nu}$ follows a chi-square density with $r + NT$ degrees

of freedom. It can be observed that now if the sample size is larger, $\frac{T\mathbf{e}'\mathbf{e}+r}{\nu}$ approaches a $\chi^2(NT)$ distribution, just as a homoskedastic Bayesian linear model. The distribution for U_{ikt} , ($t = 0, \dots, T$) then reduces to a simpler version by changing ν_i to ν , and estimation time falls.

3.5.3 Single Individual in Each Region

As previously discussed, if necessary the number of regions can be set equal to the number of observations by making each region contain only one individual. In other words, when $n_i \equiv 1$, $M = N$. One can choose to deal with this situation simply by setting n_i to one. Normally, however, such model specifications require homoskedasticity across individual error terms; otherwise, there are too many variances to estimate, greatly complicating estimation. For this reason, as mentioned in section 3.5.2, ν_i is changed a constant ν term. Moreover, the index k is removed in all formulations, as n_i equals one. This converts some vectors into scalars. Substituting these changes into all formulations, estimates with “individualized” regional effects can be obtained.

3.6 CHAPTER SUMMARY

This chapter described the specification of the dynamic spatial ordered probit (DSOP) model, with which spatial heterogeneity, spatial autocorrelation and temporal autocorrelation are all recognized. The chapter also explained why the land use intensity and ozone-concentration data sets, can be reasonably analyzed using such a model specification. The chapter then described how the model can be estimated within a Bayesian framework. Prior and posterior distributions for all parameters and variables of interest were discussed in detail, and the overall

MCMC sampling process was summarized. Finally, special, simplifying cases of the model were discussed briefly.

CHAPTER 4. MODEL VALIDATION AND COMPARISONS

4.1 INTRODUCTION

Before application to empirical data (on land development intensity and ozone levels), the dynamic spatial ordered probit (DSOP) model developed in Chapter 3 is tested using a simulated dataset. Because such self-generated data have known parameter values and controlled interactions, they are more reliable for evaluating performance of the model specification and the proposed estimation techniques. This section describes how this experimental data set was generated and how the DSOP model was estimated. The estimation results are compared to true values, and the influence of parameter magnitudes on estimation consistency is evaluated. Finally, performance of this DSOP model is compared to an ordinary ordered probit (OP) model, a dynamic (but spatially independent) ordered probit (DOP) model, and a spatial (but cross-sectional) ordered probit (SOP) model estimated using the Bayesian approach. These comparisons help illuminate the superiority of the DSOP model.

4.2 SIMULATED DATASET

In the simulated dataset, there are 30 regions, each containing 10 individuals observed over 8 time periods. Each individual has a response level of 1, 2, or 3. That is, $M = 30$ and $n_i = 10, \forall i$, (so that $N = 300$), $T = 8$ and $S = 3$. There are $300 \times 8 = 2400$ observed responses in total and 3 possible levels. Figure 4.1 shows the location of these 30 regions. The weight matrix is generated based on (queen) contiguity. For example, region 10 is considered contiguous with regions 3, 4, 5, 9, 11, 15, 16 and 17. It is then row-standardized.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

Figure 4.1 Location of Regions in Simulated Dataset
(Region 10 and its contiguous neighbors shown in grey)

The region-specific effect is generated using the following formulations:

$$\boldsymbol{\theta} = (\mathbf{I}_M - \rho \mathbf{W}) \mathbf{u} \quad (4.1)$$

$$\mathbf{u} \square N(0, \sigma^2 \mathbf{I}_M) \quad (4.2)$$

where the spatial autocorrelation coefficient ρ is set to be 0.1, 0.6, 0.7 and 0.9 in different experiments. The variance σ^2 is equal to 1 so that \mathbf{u} for each region follows an iid standard normal distribution.

The individual-specific variables are normally distributed independently and heteroskedastic over the regions. Assumed values of variance v_1 through v_{30} are shown in Figure 4.2:

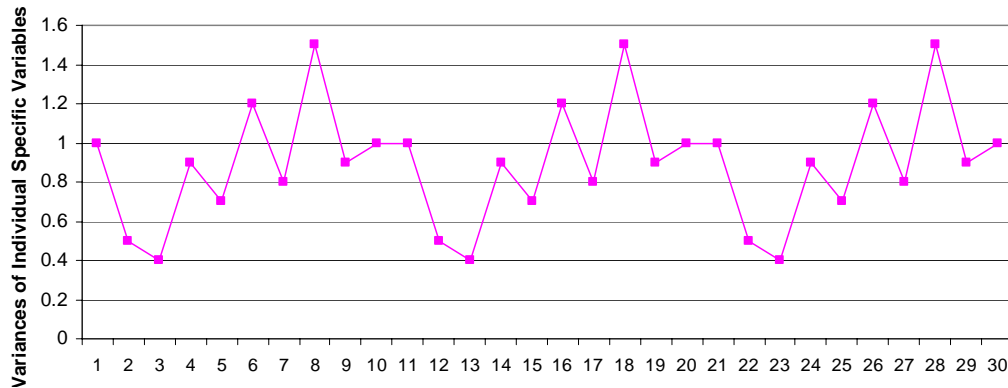


Figure 4.2 Assumed Variances of Individual Specific Effects across Regions

The specific value of each variance is set arbitrarily, between 0.4 and 1.5. This range of magnitudes helps ensure that the uncertainties caused by individual specific errors are important but do not overwhelm/dominate latent variables' effects, which is felt to be the most common case in reality. The variance for region 1 is fixed at its true value in estimation, which is necessary for identification.¹¹

The explanatory variables include the lagged utility (unobserved dependent variable) U_{t-1} and four other observed values. The temporal autocorrelation coefficient λ (i.e., the parameter for the lagged dependent variable) is set to equal 0.1, 0.5 and 0.9 in different experiments. The four variables are generated using a standard uniform distribution (bounded between 0 and 1). Their corresponding parameters (slope coefficients) are arbitrarily set as -1.7, 2, 1 and 0.5, respectively.

There are $S=3$ ordered categories, with thresholds $\gamma_1=0$ and $\gamma_2=2.1$. To summarize, the dataset is generated using the following model assumptions:

$$U = 0.5U_{t-1} - 1.7x_1 + 2x_2 + x_3 + 0.5x_4 + \theta + \varepsilon \quad (4.3)$$

¹¹ This can be inferred from Figure 21.4 in Greene (2002): in an OP model, parameters and variances can be scaled simultaneously (so that the normal curve becomes flatter or sharper), with probabilities remaining constant. In other words, it is necessary to normalize at least one of the parameters or variances for the purpose of identification.

where $y = 1$ if $U \leq 0$, $y = 2$ if $0 < U \leq 2.1$, and $y = 3$ if $U > 2.1$

and θ is multivariate normal vector of region-specific effects with zero mean and variance matrix $(I_M - \rho W)$, where ρ is set to equal 0.1, 0.6, 0.7 and 0.9, across separate experiments. As noted above, ε is a normally distributed individual-specific error term with zero mean and variable variance (heteroskedastic) across regions. (The variances of these error terms range from 0.4 to 1.5. (Figure 4.2))

4.3 MODEL ESTIMATION AND VALIDATION

The simulated data samples were analyzed using the DOSP model. The resulting estimates are compared here to their true values, in order to appreciate the model's estimation ability. However, in addition to the identification problem mentioned above, the use of small simulated data samples involves other potential problems. These problems need to be carefully handled before a robust model evaluation can be achieved.

The first problem lies in the simulated sample data itself: in the process of random number generation, extreme values can appear. To address this, researchers often use a high number of draws (to try to avoid the influence of extreme values). Here, however, the simulated sample data also are randomly generated. (As noted in Section 4.2, x was generated from a standard uniform distribution, and u and ε were generated using a standard normal distribution.) Unlike the number of draws used for estimation, the sample size here cannot be too large because a linear increase in sample size leads to an quadratic increase in computational burden. With 2,400 data points, the influence of extreme values is almost inevitable and "bad" samples are very likely to be generated. For example, the individual effect error term can become so large that it masks the contribution of explanatory variables and regional effects, leading to the conclusion that spatial autocorrelation or the influence of certain variables is insignificant.

Another example is that the values of explanatory variables and error terms may happen to be large for all data points, leading to a set of high latent dependent variable values, which means that few cells get labeled as Level 0. With such skewed data, the estimation may yield unreliable results. In order to neutralize this effect, for each parameter set, the data was re-generated 50 times, producing 600 samples (50 replicates \times 4 ρ values \times 3 λ values = 600). The averages of their estimated means and standard deviations are discussed below.

A second problem is estimation convergence. With a Bayesian approach, proof of convergence is a complicated issue. In this study, the estimation is assumed to converge when sampled parameter distributions appear to stabilize. Ideally, the number of draws (R) should be set as high as possible, but computational time and memory requirements also need to be taken into account. Especially when 600 samples (each containing 2,400 data points with complicated interactions) are to be analyzed, computational efficiency is an important consideration. Several R values were examined first here, for a small number of samples. Their estimation performances and computational intensities were evaluated, and the final selection was $R = 2,000$ since, beyond this number of draws, model the results no longer noticeably improved. Furthermore, after 1,000 runs, the distributions of all parameters appear stable. Therefore, the first 1,000 runs were omitted (burn-in) and the mean and standard deviation are both calculated based on the final 1,000 draws.

As described above, 600 simulated data samples were generated and their parameters then estimated with diffuse priors. The averages of all parameter estimates' means are shown in Table 4.1. Table 4.1 also uses root mean squared errors (RMSE)¹² to describe estimation accuracy for each parameter set. As can

¹² RMSE is the square root of mean squared error (MSE), which is (an estimate of) the expected value of the squared "error" (i.e., the difference between estimated and true values). This indicator

be observed, all RMSEs lie below 1. Considering the magnitudes of the parameter values, the estimation results are quite close to true values.

Some interesting tendencies are apparent. As the temporal autocorrelation coefficient (λ) increases, the magnitudes of coefficients and variances for both individual and regional specific effects tend to exhibit higher bias (as shown in Table 4.1 and Figure 4.3). One reason for this phenomenon is that, as λ increases, the influence of temporally lagged, latent response values rises, adding uncertainty to the right-hand side of the model. In the estimation process, this uncertainty will be partially ascribed to the error terms, which leads to larger estimates of σ^2 and $v_i, \forall i \in M$. As mentioned in Section 4.2, this process will in turn produce higher β estimates (to accommodate the increase in scale).

An increase in the spatial autocorrelation coefficient (ρ) also leads to greater bias. As mentioned in Chapter 2, when positive spatial correlation exists but is not fully recognized, the coefficients tend to be more biased because areas with higher response magnitudes will have a greater impact on model estimates.

is often used in assessing a forecasting model's predictive power (Greene, 2002). It also can be used to evaluate estimation accuracy when true parameter values are known. A larger RMSE value indicates an increase in variations that the model does not account for.

Table 4.1 Estimation Results using Simulated Data (Averages from 50 Samples)

Parameter	True Value	Average of Means from Estimated Parameter Distributions											
	λ	0.1	0.1	0.1	0.1	0.5	0.5	0.5	0.5	0.9	0.9	0.9	0.9
	ρ	0.1	0.6	0.7	0.9	0.1	0.6	0.7	0.9	0.1	0.6	0.7	0.9
β_1	-1.7	-1.701	-1.715	-1.712	-1.888	-1.727	-1.776	-1.801	-1.887	-1.841	-1.833	-1.881	-1.822
β_2	2.0	1.965	1.984	2.019	2.125	2.046	2.049	2.097	2.278	2.191	2.140	2.105	2.154
β_3	1.0	0.965	0.972	1.004	1.012	1.012	1.030	1.033	1.129	1.046	1.090	1.072	1.086
β_4	0.5	0.519	0.519	0.543	0.551	0.542	0.518	0.554	0.646	0.561	0.545	0.539	0.647
λ	–	0.097	0.101	0.099	0.097	0.492	0.507	0.514	0.511	0.919	0.921	0.909	0.863
ρ	–	0.048	0.452	0.572	0.845	0.039	0.494	0.623	0.863	-0.001	0.498	0.616	0.855
σ^2	1.0	1.054	1.091	1.117	1.832	1.158	1.217	1.302	1.768	1.290	1.232	1.307	1.498
γ_1	0.0	-0.223	-0.133	-0.112	0.006	0.094	-0.333	-0.358	-0.351	-0.202	-0.330	-0.177	-0.138
γ_2	2.1	1.818	1.933	2.009	2.276	2.190	1.803	1.834	2.075	2.090	1.956	2.130	2.342
Average RMSE		0.371	0.278	0.231	0.883	0.225	0.517	0.566	0.930	0.445	0.492	0.429	0.630

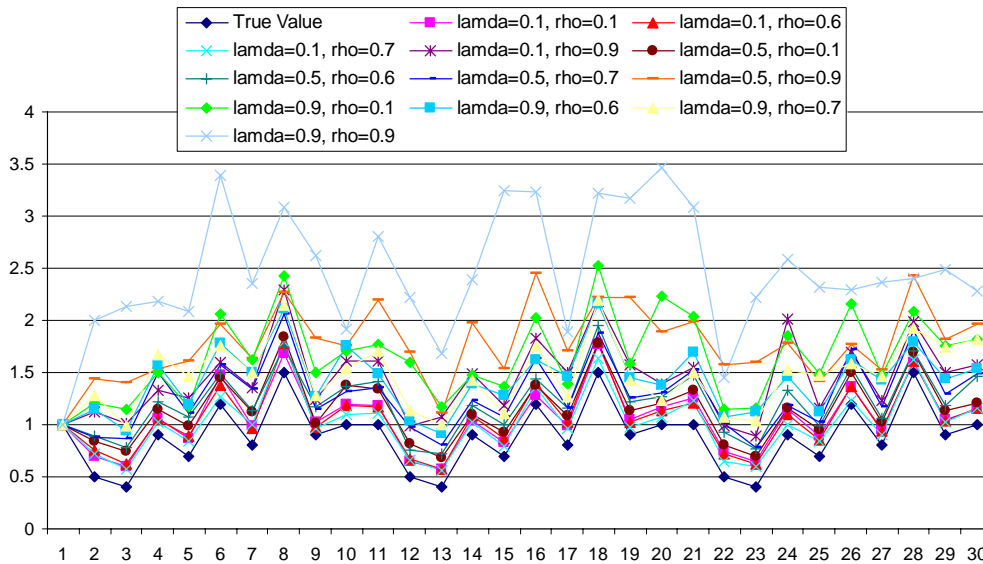


Figure 4.3 Variances of Individual Specific Errors with 50 Samples

In fact, this consistency problem is very common for nonlinear panel data models and dynamic models (see, e.g., Neyman and Scott, 1948) and has been studied for many years. A larger sample (larger N) and longer panel (larger T) may reduce this bias (Arellano and Hahn, 2005). Researchers have also proposed various approaches to reduce bias and achieve consistency with smaller N and T values (see, for example, Alvarez and Arellano, 2003, and Bester and Hansen, 2007). An efficient bias-reduction technique for the DSOP model makes an interesting topic for future study, but is not the focus of this dissertation and is not discussed at length here. In fact, such overestimation appears to be slight here: Most of the biases in slope parameters lie below 10%. Bias in estimates of the variances of individual specific errors (v_i) are higher. However, as can be observed in Figure 4.3, with the exception of the extreme case (where both λ and ρ are 0.9), biases in all other cases lie well below 100% and their relative magnitudes appear close to the true pattern.

In summary, the DSOP model performs well with the simulated data. It satisfactorily detects the temporal and spatial interaction effects as well as the influence of different variables.

4.4 MODEL COMPARISONS

To further validate the DSOP model, its performance is compared to those of simpler models (all estimated via a Bayesian approach), using a sample that provides a balanced mix of the three levels of the 50 samples.¹³ These simpler models include a standard ordered probit (OP) model; a dynamic ordered probit (DOP) model, which still allows for spatial heterogeneity but not spatial autocorrelation; and a spatial ordered probit (SOP) model, which incorporates all spatial effects but does not consider the temporal dependency. Data statistics for this sample are shown in Table 4.2, and the histogram of y values (Figure 4.4) indicates that enough observations exist for each level.

Table 4.2 Summary Statistics for One Sample

Variable	Mean	Standard Deviation	Minimum	Maximum
x_1	0.4978	0.2873	6.579E-04	9.995E-01
x_2	0.4994	0.2893	5.842E-04	9.990E-01
x_3	0.4936	0.2901	4.174E-05	9.999E-01
x_4	0.4877	0.2895	3.049E-05	9.998E-01

¹³ In fact, model comparisons based on other samples also were performed. It was found that, as long as the samples appeared reasonably well balanced (across response levels) and estimators converged, all samples yielded quite similar results. In order to illustrate convergence patterns and results more clearly, only one sample's results are presented in detail here.

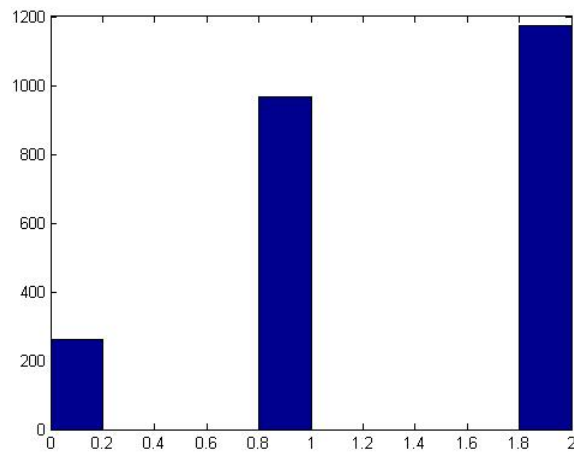


Figure 4.4 Histogram of Dependent Variable Values

As before, these models are run with 2,000 draws of which the first 1,000 draws are omitted (burn-in sample). As an example, Figure 4.5 shows the estimation convergence pattern for β_1 . Estimates of other parameters follow a similar pattern. The figure suggests that after the first 1,000 draws, the estimation becomes stable and may be convergent.

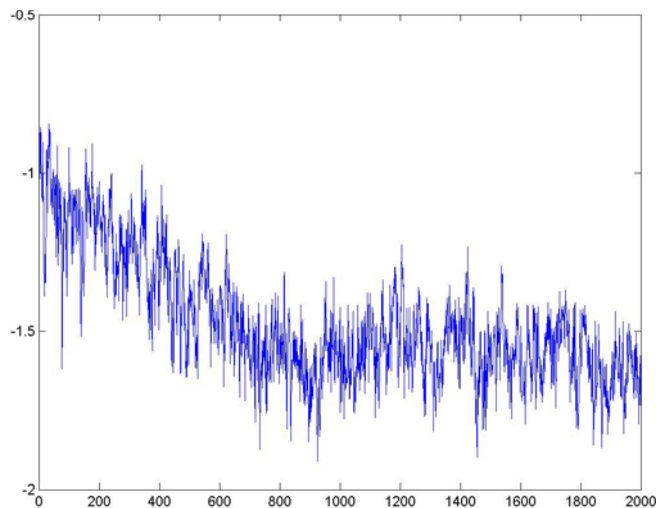


Figure 4.5 Estimation Convergence Pattern for β_1

Table 4.3 shows the estimation results for this sample. In addition, Figure 4.6 depicts estimates of v_i ($\forall i \in M$), using the DSOP model, where lower and higher

bounds are defined as 1st percentile and 99th percentile values. Mean estimates lie quite close to true values. Considering that only 80 observations are effectively used to estimate each v_i , the standard deviations are understandably large.

Table 4.3 Estimation Results using One Sample and Different Specifications

Para..	True Value	OP		DOP		SOP		DSOP	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
β_1	-1.7	-0.807	0.079	-1.581	0.104	-1.621	0.117	-1.608	0.119
β_2	2.0	1.727	0.078	2.201	0.112	2.150	0.128	2.166	0.128
β_3	1.0	0.999	0.079	1.043	0.107	1.014	0.099	1.000	0.097
β_4	0.5	0.634	0.076	0.502	0.089	0.461	0.092	0.469	0.098
λ	0.1	---	---	0.131	0.023	---	---	0.110	0.021
ρ	0.7	---	---	---	---	0.769	0.098	0.751	0.098
σ^2	1.0	---	---	2.182	0.742	1.323	0.426	1.116	0.342
γ_2	0.0	-0.203	0.014	-0.372	0.038	-0.226	0.084	0.081	0.057
γ_3	2.1	1.264	0.009	1.811	0.029	1.980	0.032	2.261	0.013
RMSE		1.769		1.472		0.463		0.293	
DIC		4360.4		3098.0		3106.1		3070.7	

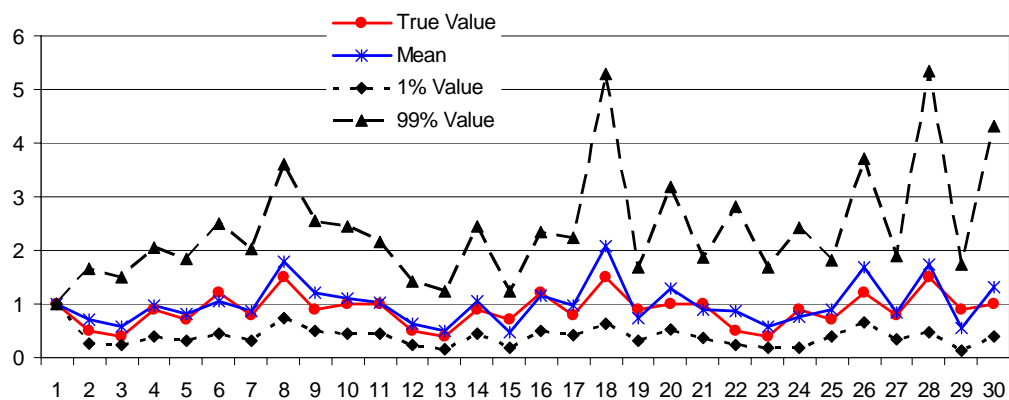


Figure 4.6 Variances of Individual Specific Errors with One Sample

Table 4.3 also shows RMSE and deviance information criteria (DIC)¹⁴ values for each specification. As before, RMSE indicates estimator accuracy. The DIC is an indicator of model fit. Both suggest that the DSOP model more accurately estimates the underlying parameters, with high statistical significance and fit of the sample data. In contrast, because of the inability to detect λ and ρ , the OP model's estimates are highly unsatisfactory. As shown in Table 4.3, it returns the appropriate signs and relative magnitudes for β parameters, but estimates deviate from true values quite a bit. The performances of the DOP and SOP models lie in-between. Though inferior to the DSOP model results, they are rated better than the OP model. RMSE measures suggest that the SOP model yields much more accurate estimates than the DOP model, which is quite understandable given the fact that λ is only 0.1 and ρ is 0.7 in this one particular sample. In other words, ignoring the temporal autocorrelation (i.e., restricting a 0.1 parameter to equal 0) should typically have less of an impact than a situation where one ignores a spatial autocorrelation term of 0.7. Interestingly, the DIC fit measure, suggests that the DOP model is very slightly preferred to the SOP model. The DOP model's smaller DIC value implies that, while the DOP model is not as able to produce accurate parameter estimates, it still fits sample data better than the SOP model, because it still accounts for spatial heterogeneity.

Table 4.4 illustrates predictive accuracy using the four methods. The standard OP model only correctly predicts dependent values for 47.0% of the 2,400 observations. The DOP model increases this percentage to 60.8%. The SOP

¹⁴The deviance information criterion (DIC) is a generalization of the Akaike information criterion (AIC) and Bayesian information criterion (BIC). It is particularly useful for Bayesian model comparison and selection (see Gelman et al., 2004, and Spiegelhalter et al., 2002). However, one limitation of the standard DIC is that it is only valid when posterior distributions are approximately multivariate normal. For models involving extremely asymmetric or bimodal posterior distributions (which happens for the DSOP model), some modified DIC need to be used instead. This dissertation uses the DIC calculation method for mixture models proposed by Celeux et al. (2006).

model's prediction rate is quite close to that of the DSOP model: 66.4%. Such a percentage is fairly satisfactory, given the presence of three response levels and considerable randomness in the sample dataset (σ^2 and U_i in the simulated data have similar magnitudes as all slope parameters, causing regional-specific and individual-specific errors to have a similar level of influence on latent utility values).

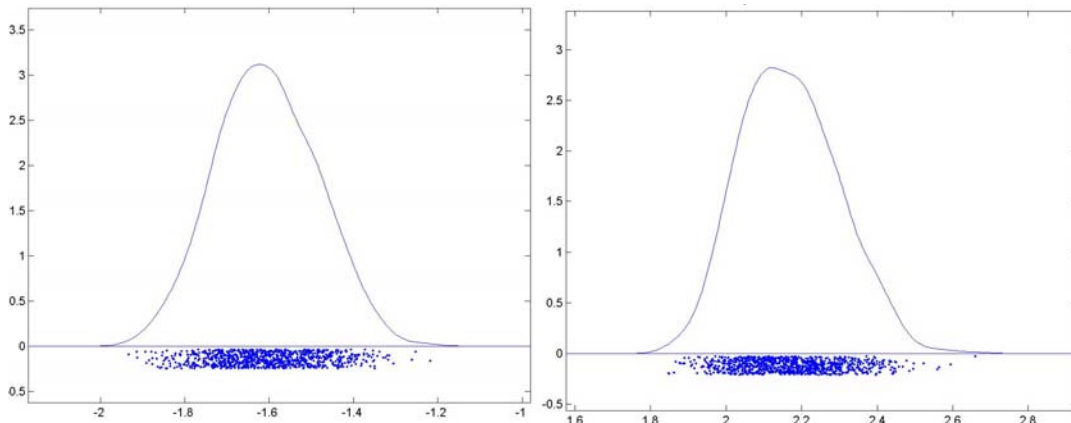
Table 4.4 Prediction Rates using Different OP Model Specifications

		Actual			Total	% Cases Correctly Predicted(%)	
		1	2	3			
y Value							
Predicted	OP	1	55	145	100	300	47.0
		2	106	372	372	850	
		3	100	448	702	1250	
	DOP	1	124	154	97	375	60.8
		2	116	511	253	880	
		3	21	300	824	1145	
	SOP	1	121	133	20	274	65.1
		2	124	536	249	909	
		3	16	296	905	1217	
	DSOP	1	121	112	17	250	66.4
		2	119	583	268	970	
		3	21	270	889	1180	
Total			261	965	1174	2400	

Such comparisons, of prediction rates, RMSE and DIC values, suggest that the DSOP model is superior to all the simpler models, as anticipated. It is followed by the SOP model, indicating the importance of recognizing the spatial autocorrelation in the dataset. Recognizing temporal dependency also significantly improves model performance, relative to a standard OP model. In this example study, this improvement is not as evident as recognizing the spatial autocorrelation, but this is partially due to the small true value of the temporal coefficient, λ . The OP model, though easy to specify and estimate, does not

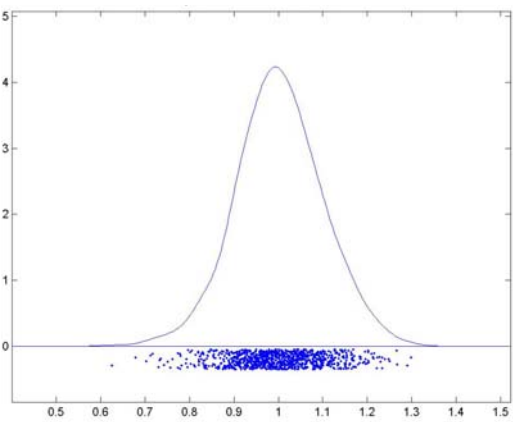
adequately utilize the observed information, thus returning inaccurate parameter estimates and response predictions.

Figure 4.7 illustrates the posterior distributions produced by the DSOP model for all parameters of primary interest. As summarized in Chapter 3's Section 3.3.5, β and λ present normal distributions, and σ^2 presents a chi-square distribution. Posterior distributions for ρ and γ are non-standard, and γ appears to be multimodal.

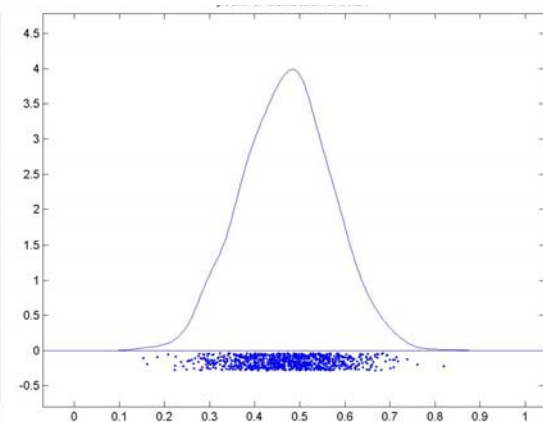


(a) Posterior Distribution for β_1

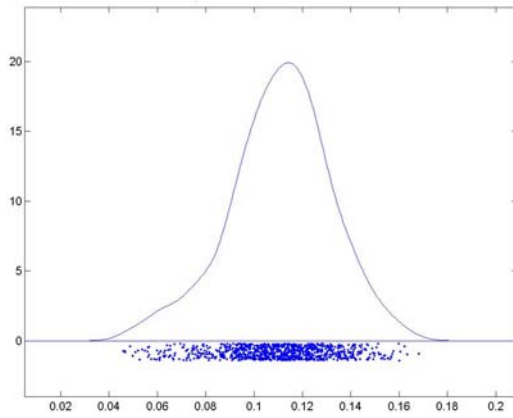
(b) Posterior Distribution for β_2



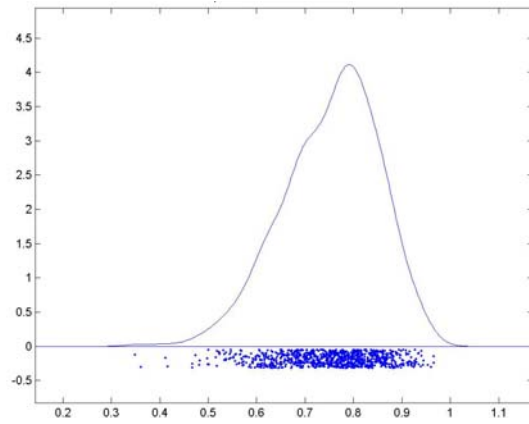
(c) Posterior Distribution for β_3



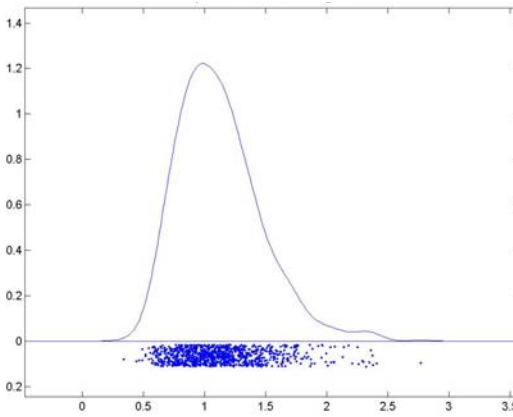
(d) Posterior Distribution for β_4



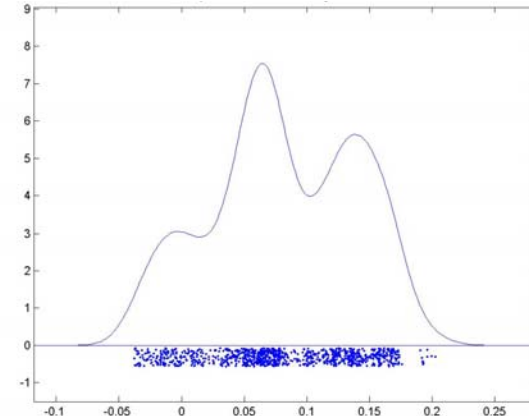
(e) Posterior Distribution for λ



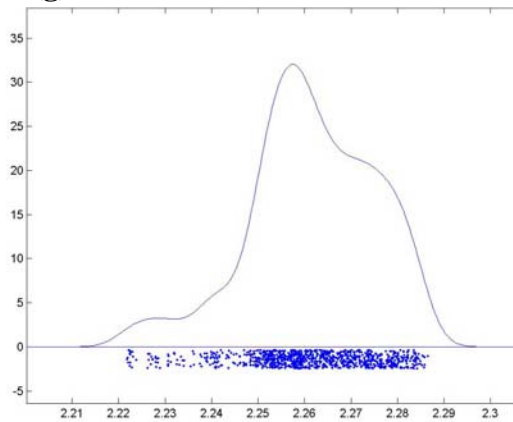
(f) Posterior Distribution for ρ



(g) Posterior Distribution for σ^2



(h) Posterior Distribution for γ_1



(i) Posterior Distribution for γ_2

Figure 4.7 Posterior Distributions for Parameters of Primary Interest

4.5 CHAPTER SUMMARY

This chapter described simulation of a test dataset, and then examined the DSOP model's performance using the simulated data. Estimator accuracy, model goodness of fit and prediction rates were compared across standard OP, DOP, SOP and DSOP models. As expected, model estimation, validation and comparison all lead to the conclusion that the DSOP model performs better than the other, simpler models. Finally, posterior distributions for parameters of primary interest were depicted, confirming the distribution formulations, as discussed in Chapter 3.

CHAPTER 5. DATA DESCRIPTION

5.1 INTRODUCTION

This chapter describes the two datasets used as empirical examples for the application of the DSOP model: land development intensity levels and ozone concentration levels. Section 5.2 explains how data for land development intensity levels was collected, processed and integrated, and Section 5.3 introduces the ozone concentration data. Assumptions made during the data assembly process are discussed in detail.

5.2 LAND DEVELOPMENT INTENSITY

The data used for land development dynamics comes from multiple data sources, including satellite images, the Census of Population, City of Austin school district and employment data, as well as transportation and geographic data from the Capital Area Council of Governments (CAPCOG). The land cover information is used as the dependent variable, and all others serve as explanatory variables. These include total neighborhood population, number of workers living in the neighborhood, average household income and number of schools in the neighborhood, travel time to the nearest major highway, travel time to the region's CBD, travel time to major (top 15) employers, travel time to the nearest airfield, average ground slope, and average elevation (of each 300m×300m grid cell).

The following sections describe how these different datasets were collected, processed and combined over space and time.

5.2.1 Land Development Intensity Level

The dependent variable, the land development intensity level, is derived from satellite images with 30m resolution. The following sections explain how the land cover information is obtained and classified based on light reflectance rates discerned from satellite images, how and why the 30m resolution grid cells are aggregated into 300m×300m grid cells, and how the original 9, unordered land use classes are categorized into the 4 development intensity levels.

5.2.1.1 Land Cover Information Derived from Satellite Images

The satellite images used for deriving land cover information come from Landsat 4, 5 and 7 systems and cover the urban area of Austin, Texas. LandSat 4, 5 and 7 were launched in 1982, 1984 and 1999, respectively. They all have an identical orbit with a cycle of 16 days. These satellites are able to take snapshots for every American city with a 30 m × 30 m resolution. These imaging systems collect reflectance of seven spectral bands. When information from all these bands is combined, reasonable land cover information can be derived.

There are two basic approaches for deriving such land cover data: supervised and unsupervised. Both approaches require knowledge of actual land cover information. The basic distinction is that a supervised technique uses this information from the beginning, thus guiding the classification process in a more rigorous, mathematical fashion.¹⁵

¹⁵Supervised classification uses the pre-collected actual land cover information as training data. With a large enough sample that well represents examples of all land cover classes, supervised classification allows the computer to match other pixels based on statistics in the same class. Generally, a decent number of training pixels should be provided for each class, although the precise number varies with algorithms. Unsupervised techniques, in contrast, first classify grid cells based on similarities in their band information, forming a series of clusters (groups with similar band statistics). These clusters are then linked to intended land cover classifications via very simple comparison of clusters to actual land cover types. In most cases these “clusters” and the intended classification scheme do not align (for example, classification *A* is linked to clusters 1,

The image processing work that produced the land cover information used in this dissertation was performed by students supervised by Dr. Barbara Parmenter at The University of Texas at Austin in 2002. Bands 1 - 5 and 7 were used as inputs for the classification algorithms. A hybrid supervised/unsupervised classification was performed. First, supervised classification was carried out using maximum likelihood decision rules with training data based on visual interpretation of USGS topographic maps and digital orthoimagery quarter quadrangles (DOQQs). An ISODATA¹⁶ clustering algorithm then was used for post-classification sorting of over-classified classes to reduce inter-class confusion. Though more years of satellite images probably were available, due to the computational intensity of classification work, cloud cover variations, and other, seasonal effects in datasets, only four years of satellite images were classified by Dr. Parmenter's team. These were for 1983, 1991, 1997, and 2000. Unfortunately, the time gaps between these years are unequal; ideally this variation in step size should be treated accordingly. Limited by time, this dissertation leaves this issue for future work

For each year, the study region covers a 48.5 km x 55.8 km area, containing around 3 million 30 m x 30 m pixels. Each of these pixels was classified as one of the nine land-cover types: water, barren, forest/woodland, shrubland, herbaceous natural/semi-natural, herbaceous planted/cultivated, fallow, residential, or commercial/industrial/transportation. As an illustration, Figure 5.1 shows the derived land cover types for the year 1983.

2, and 4, but cluster 4 also includes classification *B*). In such cases, certain classes will be "over-classified" (class *A* in this example), requiring several rounds of cluster "busting" or combining. 16 ISODATA stands for Iterative Self-Organizing Data Analysis Technique. It is an unsupervised classification approach. It is essentially a clustering technique based on minimum distances (of band values). Jensen (1996) provides more technical details about this approach.

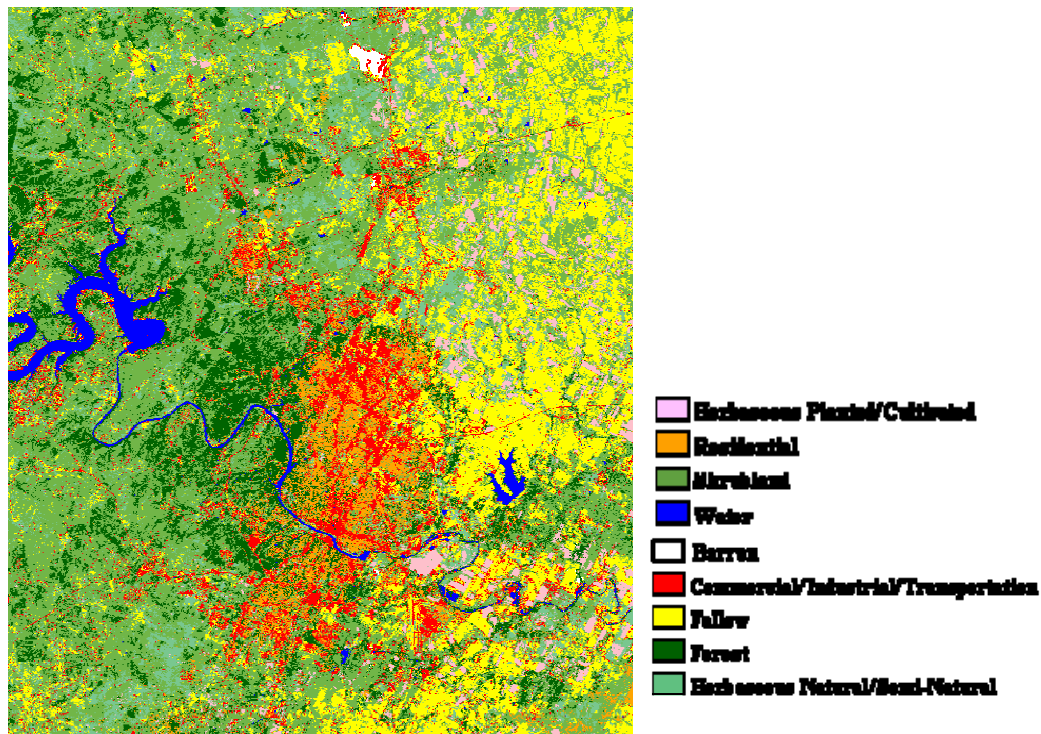


Figure 5.1 Original Land Cover Information for Year 1983

5.2.1.2 Uncertainty and Aggregation of Land Cover Information

Though the classified land cover information is felt to be of high quality (i.e., to present true land cover types [see Frazier, 2005]), numerous factors still influence its accuracy. Some of these factors can be traced back to the original images. For example, the image distortion caused by the satellite’s motion (relative to the Earth) and variations in atmospheric conditions (such as humidity and shadow effects) play a role. However, the most influential factor is still felt to be the classification process.

It is important to note that pixels were classified based on their spectral qualities (i.e., reflected light) rather than actual information on how humans “use” the land (e.g. tax appraisal records). Therefore, the terms residential and commercial/industrial/transportation are probably best interpreted as lands covered by different intensities of light-reflecting man-made materials. A visual

comparison of the classified data and DOQQs shows that the grid cells that are classified as commercial, industrial or transportation are those that are largely covered by cement or asphalt. Residential land is more likely to be area covered largely by cement but dotted with some vegetation.

Of course, the land cover information is based on parameters calibrated using the training data. As with any extrapolation/prediction, there is always added uncertainty when the calibrated rules or parameters are used for grid cells other than those with precisely known training data.

Further more, even with a 30 meter resolution (0.22 acre), one pixel can be composed of several land cover types. Thus, things can become confused when indexing each pixel as one specific type. Instead, classifying grid cells via some typology that can indicate mixtures of different land covers may be more reasonable.

One intuitive approach to moderate the above mentioned data imperfections involves aggregating observations in a neighborhood. In this way, some random classification errors can be cancelled. Therefore, the original dataset provided by Dr. Parmenter was aggregated using a square window that covers 100 grid cells. In other words, the new dataset now has a resolution of 300m × 300m. The study area now contains 29,946 of these larger grid cells, and remains a large sample with fairly small units. Another advantage of using larger grid cells is reductions in computational burden: the sample size is reduced by a factor of 100. For spatial studies, which track inter-observational interactions, this means that computational load is reduced by 10,000. Then, instead of giving each grid cell a specific land cover type, the new classification scheme is derived based on the proportions of barren land, water, vegetation and man-made materials. The following section describes in more detail how these classes were determined.

5.2.1.3 Categorizing Land Development Intensity Levels

As previously mentioned, each 300m cell's land development intensity level was determined based on the mixture of different land cover types. Of course, such definitions of "high intensity" or "low intensity" are rather flexible: the following definitions can be easily modified to adapt to different settings and user needs.

The nine land cover types were assigned different weights to indicate development intensity. Grid cells indexed as commercial/industrial/transportation are largely covered by cement or asphalt, indicating intense development activity, and therefore given a weight of 2. Residential cells were given a weight of 1.5. Grid cells classified as vegetation (shrubland, herbs, fallow, and forest) were given a weight of 0.5. Finally, if the surface is coded as barren or water, its weight is 0.¹⁷

For the aggregated, 300m neighborhoods, a simple average of these 100 weight indices was computed overall to produce a single value for development intensity. This intensity was then categorized into four levels: averages below 0.5 were ranked as Level 1 (which can be interpreted as almost no development containing mostly vegetation, barren land or water). Between 0.5 and 0.8, the neighborhood was categorized as Level 2, (i.e., slightly developed with around 40% land covered by man-made materials). Between 0.8 and 1.2, the class is Level 3, meaning that this area has medium development intensity with approximately 60% developed area. Above 1.2, the category is Level 4 (i.e., intensely developed, with at least 60% of its area covered by man-made materials).

Equations (5.1) and (5.2) summarize this definition process:

¹⁷ Though the weight scheme is flexible, this dissertation approximates these weights based on percentages of man-made materials covering the land (using visual comparison to the DOQQs). The proportion of these percentages for the four types is around 4:3:1:0, so the weights are assigned accordingly.

$$\begin{aligned}
INT = & 2 \times FRXN(\textit{Commercial} / \textit{Industrial} / \textit{Transportation}) \\
& + 1.5 \times FRXN(\textit{Residential}) \\
& + 0.5 \times FRXN(\textit{Herbaceous Planted} / \textit{Cultivated} \\
& \quad + \textit{Shrubland} + \textit{Forest} + \textit{Fallow} \\
& \quad + \textit{Herbaceous Natural} / \textit{Semi Natural})
\end{aligned} \tag{5.1}$$

where INT is the intensity index, $FRXN(\bullet)$ means the fraction of appropriate land cover type in the 300m x 300m neighborhood.

$$\begin{aligned}
y = 1 & \textit{ if } INT < 0.5 \\
y = 2 & \textit{ if } 0.5 \leq INT < 0.8 \\
y = 3 & \textit{ if } 0.8 \leq INT < 1.2 \\
y = 4 & \textit{ if } INT > 1.2
\end{aligned} \tag{5.2}$$

where y is the land development intensity level, the dependent variable used in this study. In addition to a fairly meaningful interpretation, these cut-offs also ensure a reasonably balanced mix of different development intensity levels. Figure 5.2 shows the derived land development intensity level for the study area in different model years.

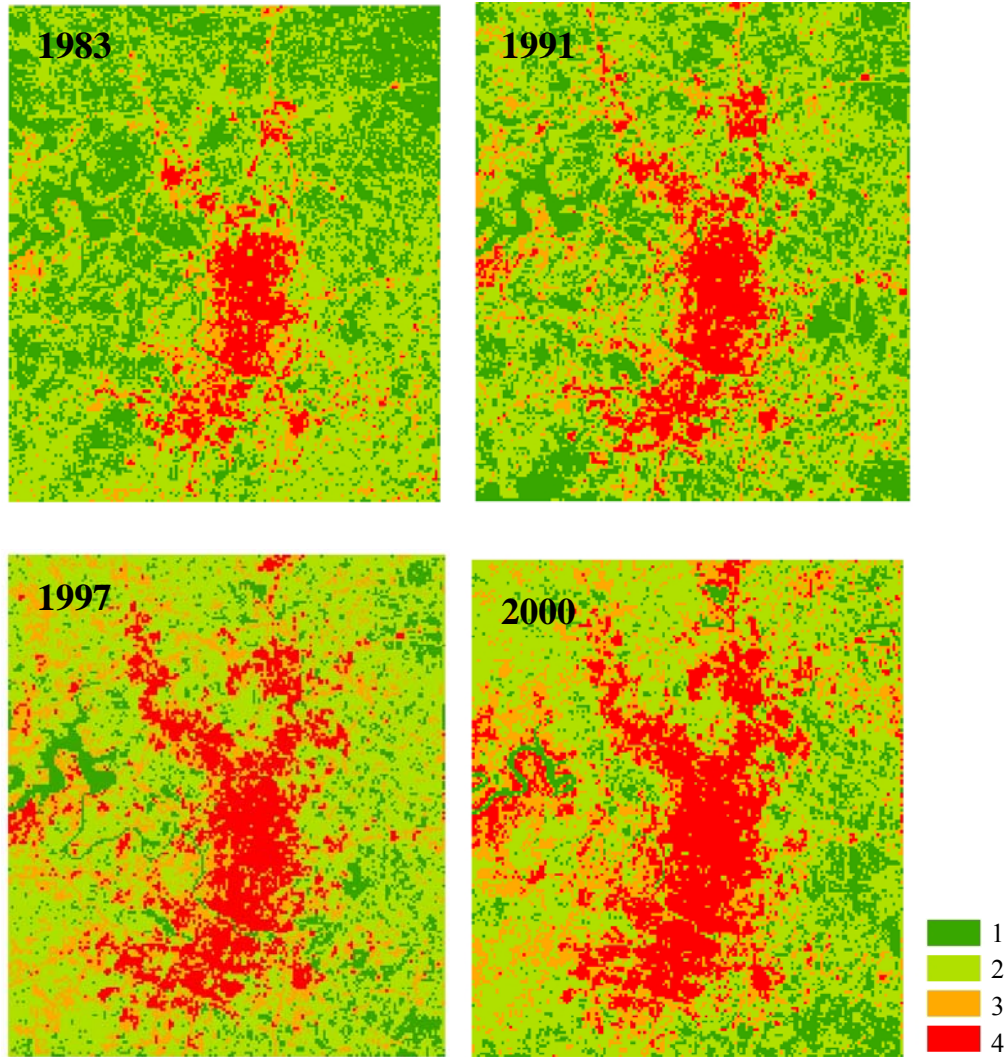


Figure 5.2 Computed Land Intensity Levels across Different Years

5.2.2 Census Data

Three variables used in this study of land use intensity come from the U.S. Census of Population for years 1990 and 2000: total population, number of workers and median household income. To be used as explanatory variables, these datasets need to be organized in the same frames (spatial and temporal) as the land cover data. However, the census data and the land cover data have

different spatial units and time points. For example, the smallest spatial unit for income in the census data is a block group, and the three land-cover data sets cover non-census years (1983, 1991, and 1997). In order to align these two datasets, the census data had to be spatially reorganized and off-year census data had to be extrapolated. The derived Census data used in this dissertation was processed primarily by Christopher Frazier (2004) for his Master's thesis.

Frazier (2004) used TransCAD's "overlay" function (Caliper Corporation 2004) to allot census data to each grid cell based on how much each block group lay within each 300m cell: for population and workers, the variables are derived using an area-weighted summation. For median household income, the variable was derived using population-weighted average of the Census values.

For temporal extrapolation, Frazier (2004) assumed an exponential form for workers and households. He calibrated the exponential model at the regional level and then rectified this uniform growth pattern by factors that indicate each cell's deviation from the "average" behavior. For median household income, a correction for inflation also was made (based on the Consumer Price Index).

It is expected that the development intensity level of a specific location depends not only on its own characteristics, but also on features of its neighborhood. For example, land owners are more likely to develop sites based on their expectation of the population of nearby areas. Therefore, after allotting population, workers and household income information to each grid cell, this dissertation calculates such variables for the neighborhood of each 300m grid cell. Neighborhood here is defined as a circle with 3km radius. This calculation is carried out in ArcMap (ESRI, 2005) using the "focal sum" function (and "focal mean" for income) after the vector map is rasterized. Figures 5.3 through 5.5 show the spatial distribution of these summed values (or averaged values, in the case of household income)

As can be observed, population in the study area is highest around downtown. Over the years, population has increased and expanded spatially. In general, the number of workers follows a very similar pattern. Even after accounting for inflation, the increases in household income are quite noticeable. The western study area exhibits higher household incomes, with peak values stay in the Westlake area.

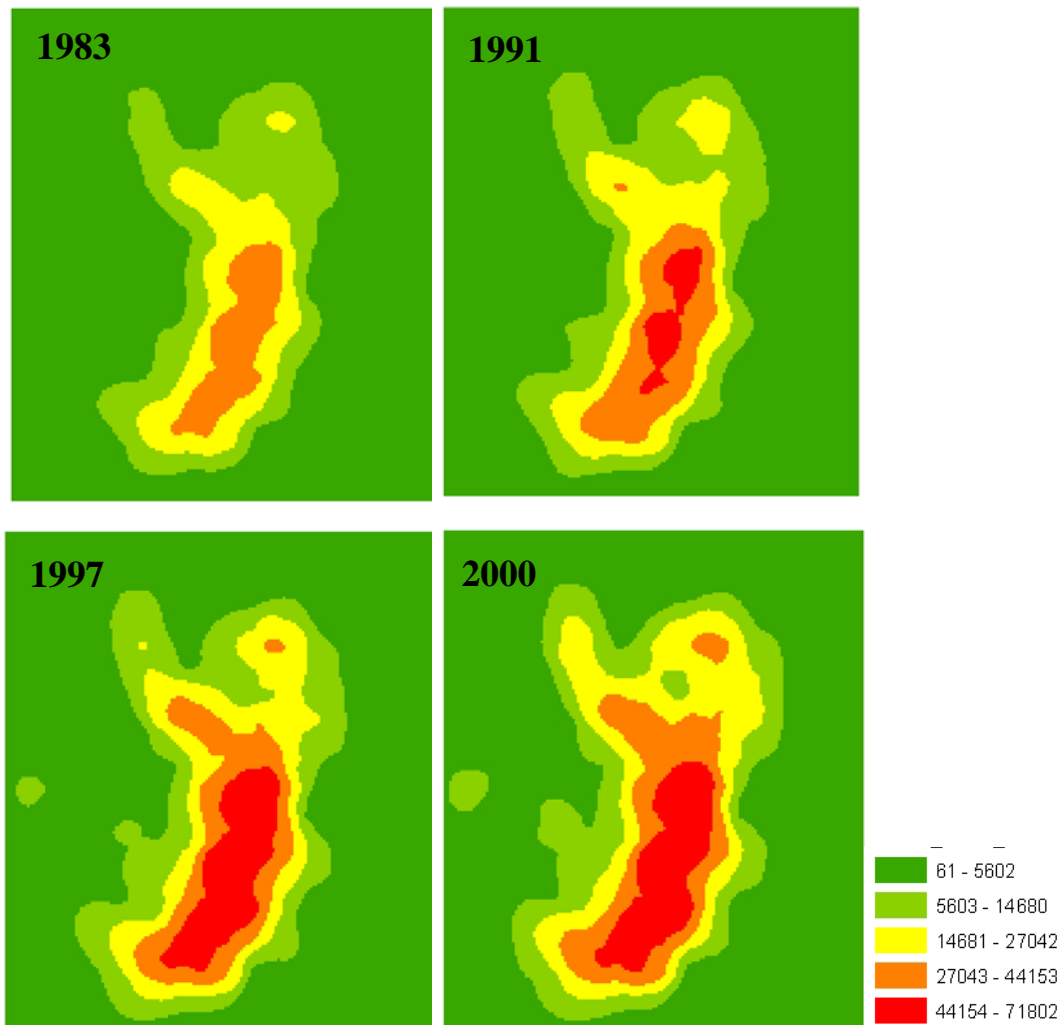


Figure 5.3 Neighborhood Populations as Shown at the level of 300m Grid Cells (where neighborhood is a 3km-radius circle)

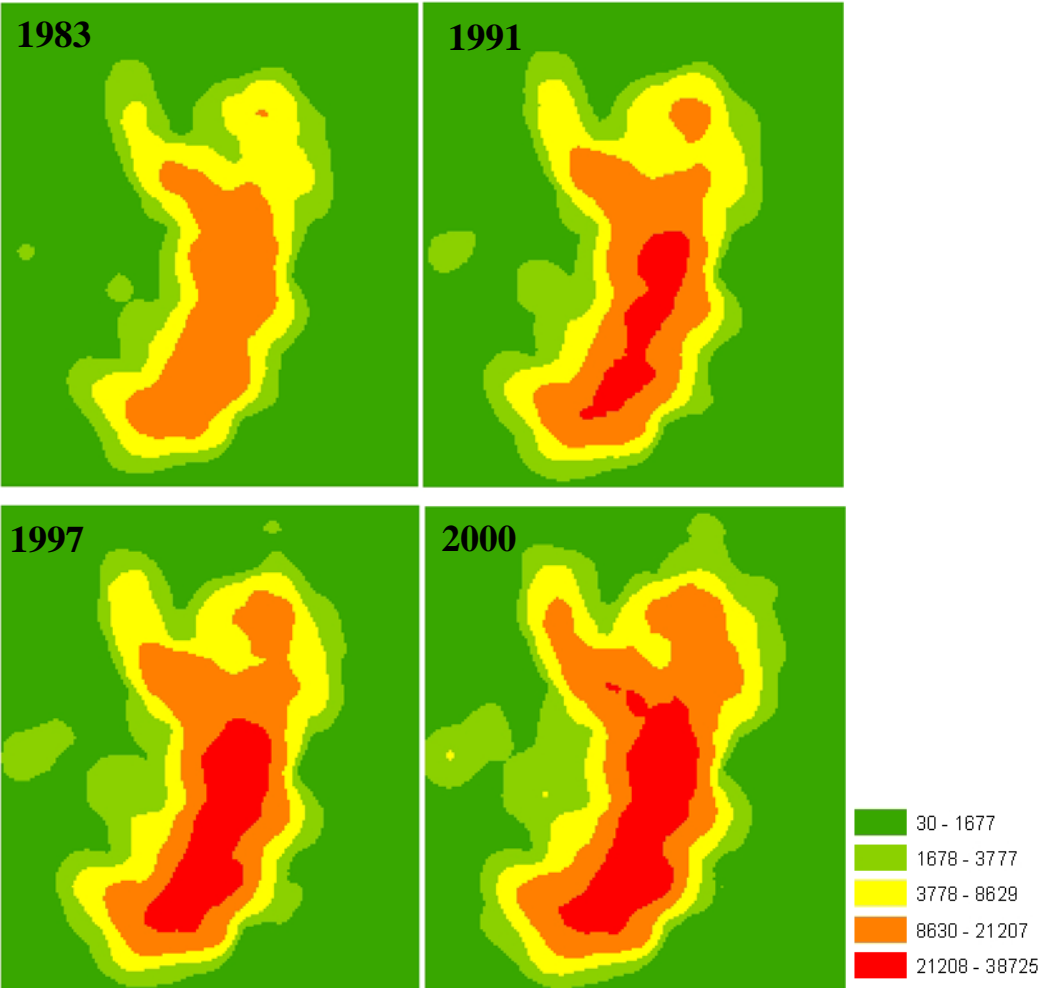


Figure 5.4 Neighborhood Workers as Shown at the level of 300m Grid Cells (where neighborhood is a 3km-radius circle)

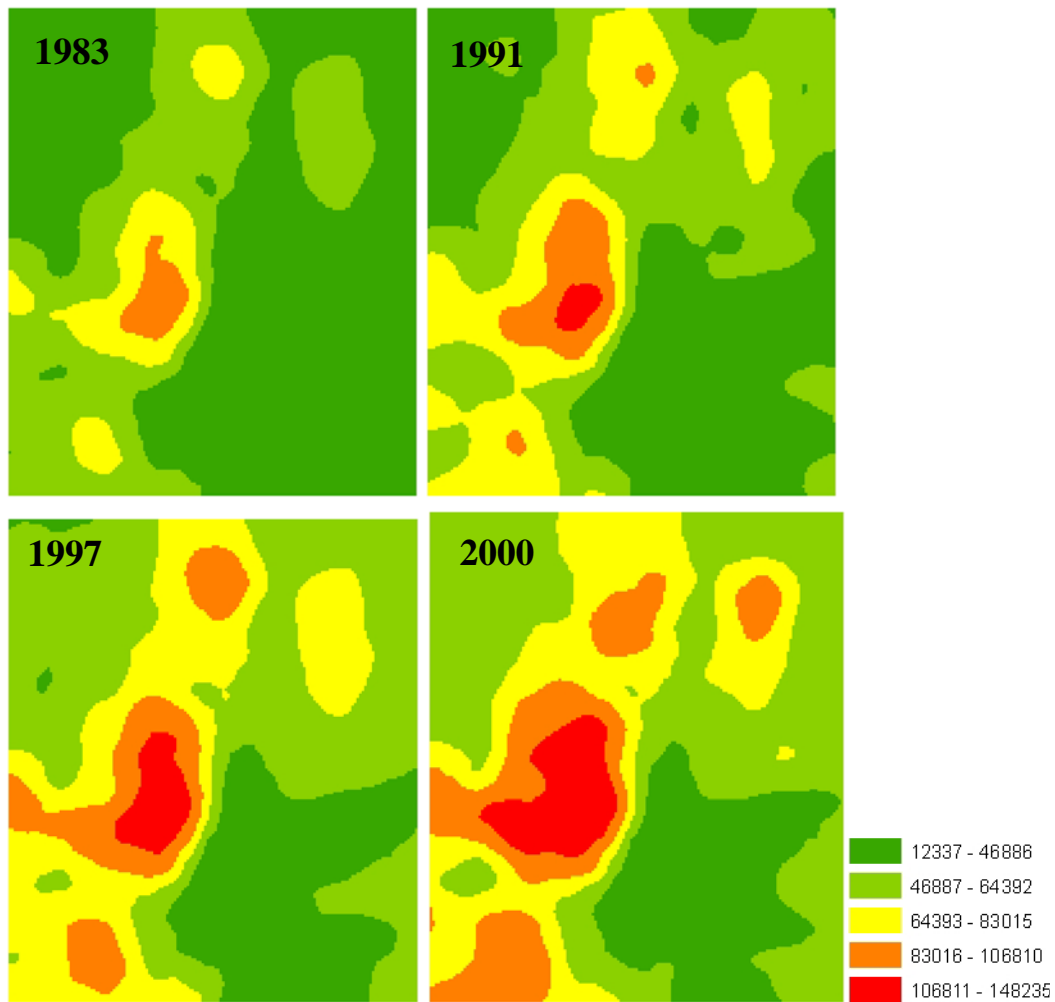


Figure 5.5 Average Median Household Income across Neighborhoods
 (where neighborhood is a 3km-radius circle)

5.2.3 Transportation Access

Development intensities can be significantly influenced by accessibility. One indicator of access is travel time to major highways, airports, and job sites. The following sections discuss how road networks and locations of major facilities and employers in different years were derived, and how travel times were calculated.

5.2.3.1 Road Networks in Different Years

When one is interested in travel times, proper road networks can be critical. Though Austin street maps can be traced back 100 years, only year 2002's map is digitally available (CAPCOG, 2005). It is possible to "vectorize" old street maps, with current techniques; this requires rather intense work and the results are subject to great uncertainty. Moreover, such methods remain impossible without hard-copy maps. This dissertation proposes a simple approach to derive road networks in different years based on land cover information. Though the method is not yet very sophisticated, the results appear quite reasonable. Moreover, this method provides a prototype for further study on how to use satellite data as a supplement to traditional data sources.

Limited by data availability, it is first assumed that the street condition in the study area did not change between years 2000 and 2002. That is, the 2002 street map represents the 2000 network. In this way, we are able to link land cover information and road existence using 2000 satellite data and the assumed 2000 network.

A road map is normally in the form of "vectors", i.e., links connected by nodes. Those vectors are first rasterized and thus transferred to grid cells with Value 1 indicating that there is road crossing the cell and 0 otherwise. Such a 0-1 situation can be analyzed using a standard binary probit model: the dependent variable is the roads existence and explanatory variables include a constant, and local fractions of commercial/industrial/transportation land and residential land. Table 5.1 provides the estimates of this simple probit model, as based on the 300m grid cells.

Table 5.1 Estimation Results for Road Existence (Binary Probit)

Variable	Mean	Standard Deviation
Constant	-0.444	-40.54
Fraction of commercial/industrial/transportation land (in 300m cell)	1.622	31.82
Fraction of residential land (in 300m cell)	2.632	63.57

Following this calibration, the model’s parameters are used to “predict” road existence in each grid cell based on the fractions of different land use/land cover types in other years (1983, 1991 and 1997).

The next step was to obtain a street map for each year by “trimming” links from the 2000 map, based on the above prediction. In ArcMap, the “join layer” (ESRI, 2005) function matched each year’s grid cell information to the 2000 road network. For each year, links with a predicted value of 1 were kept and others were deleted, resulting a somewhat reduced road network (since the road system was only assumed to grow, from 1983 through 2000/2002).

As Frazier (2002) suggests, the alignments/locations of major roads in the study area, including U.S. Highway 290, U.S. Highway 79, U.S. Highway 183, State Highway 71, Interstate 35, Loop 1, and Loop 360 did not change from 1983 to 2000. Thus, as a final refinement, the “cut off” road network of each year is combined with/overlaid on major roads. Replicated sections were removed.

Figure 5.6 shows the derived road networks in different years.

The resulting road networks remain vectors (i.e., represented by links and nodes, instead of tiny points). However, after processing, using such vectors to calculate travel distances and/or travel times is no longer a superior option (as compared to using rasterized data), since some sections of road may be incorrectly cut off. As

can be observed in Figure 5.6, the resulting road networks contain many small clusters of fragmented roads. Though the loss of these sections may be negligible in terms of total length, they can be critical to network connectivity.



Figure 5.6 Estimates of Austin Road Network in Different Years

5.2.3.2 Calculating “Travel Cost”

Thus, when estimating travel times and distances, the vector layers were first rasterized to a “weighted cost” raster layer. The cell size of this raster layer is set

to be as small as 10m x 10m, so that it can better approximate the vector layer. If there is a road, the cell is given Value 1, and if there is “No Data” (i.e., no road exists), the cell is given Value 5. This 5:1 cost ratio implies that the resulting total cost can be used in estimating travel times simply by dividing total cost by a common factor. In this dissertation, after conversion, the “total cost” was interpreted as travel times with travel speed assumptions of 8 mile/hr off-road and 40 mile/hr on road. Of course, if more information on road classification, capacity and/or congestion levels is available, the cost of each 10m cell can be more finely classified, to better represent different speeds under different conditions.

The first advantage of this method is that instead of ignoring off-road distances or assuming that a location is inaccessible, this method reasonably accounts for the impedance of “off-road” travel by giving it a “cost” that is five times on road cost. Second, because travel time (or cost) on roads is much less than time off road, the shortest path calculation is attracted to roads whenever possible. This implies that if two sections of roads are disconnected, but the gap is small, the shortest path will still go through these two sections instead of taking a more circuitous route (or simply reporting a cell to be “inaccessible”).

Though this overall method for deriving travel time is somewhat coarse (neglecting link type and issues of congestion, for example), it seems quite helpful when the road network expands noticeably over some years. After all, the “existence” of a road may be more important its travel time.

This multi-step process for deriving estimates of travel times is summarized by the flowchart shown in Figure 5.7.

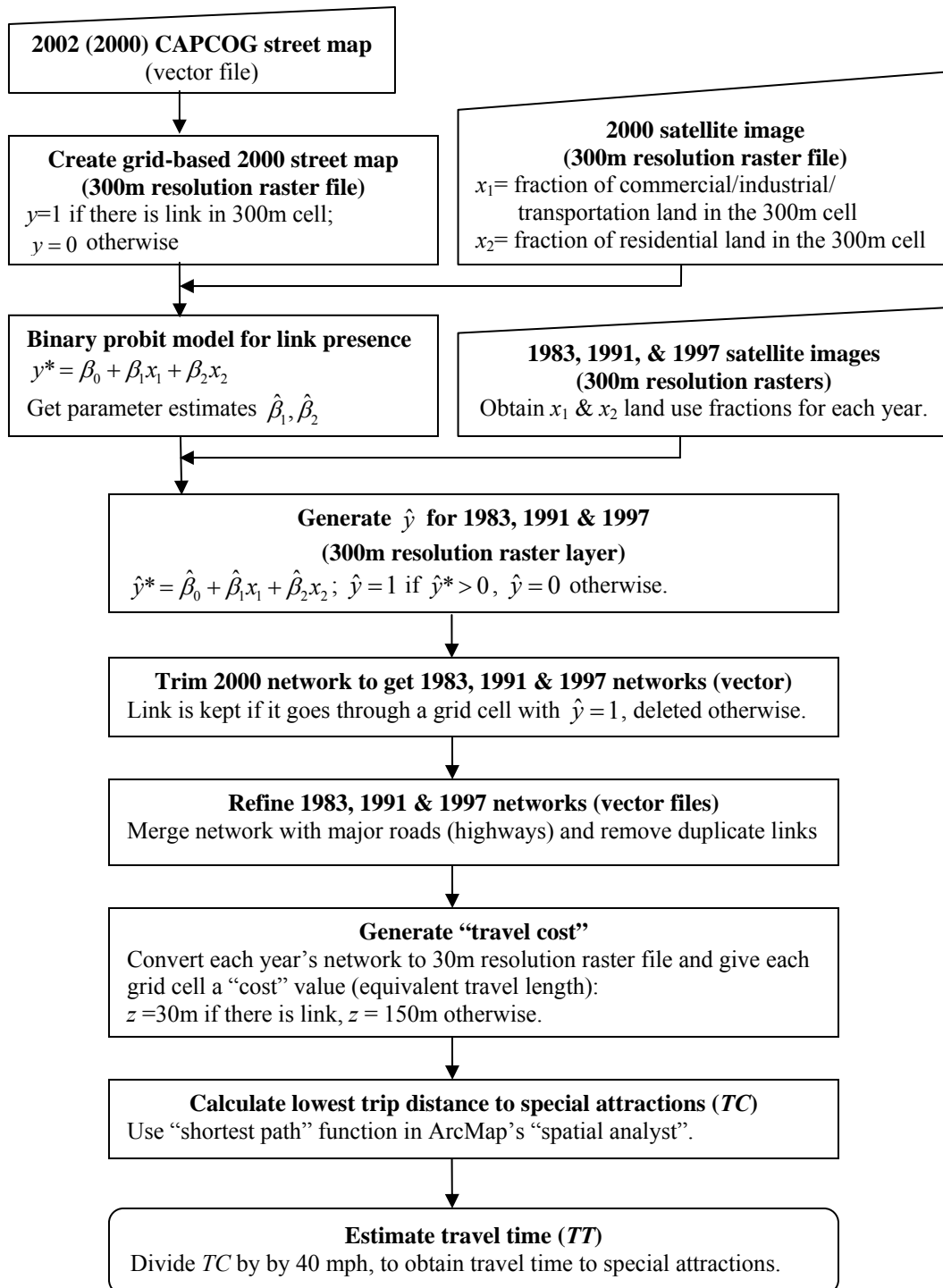


Figure 5.7 Procedure for Deriving Estimates of Travel Time

5.2.3.3 Facilities and Important Locations

Key facilities and locations considered in this dataset include major roads, all airfields, the central business district (CBD), and Austin Top 15 job sites. An airfield is helpful for long-distance travel. The CBD and sites of major employers attract many activities and trips. A major road is a key facility for travel of many types. The distances to all these sites and facilities are considered indicators of each location's attractiveness and accessibility.

As introduced in the previous section, the locations of Austin's major roads have not changed since 1983. The GIS map of these roads was obtained from CAPCOG's website (CAPCOG, 2006) along with airfield information (for each model year). The CBD is defined as a 2.4km x 3.3km rectangular area with its center located at the State Capitol Building and its long edge parallel to Interstate 35. Austin's major employer information was provided by the City of Austin (2006), but only for years 2000 and 2002, and only employers with more than 500 employees were geo-coded. There also is a 1997 Top 50 employers' map available, though not geo-coded. To make the measurements consistent, this dissertation located the top 15 employers in each model year. For years 1983 and 1991, the information was derived by tracing back the histories of the top 50 employers in 1997.

Figure 5.8 shows the locations of such attractions, including major roads, airfields, and top employers' headquarters.

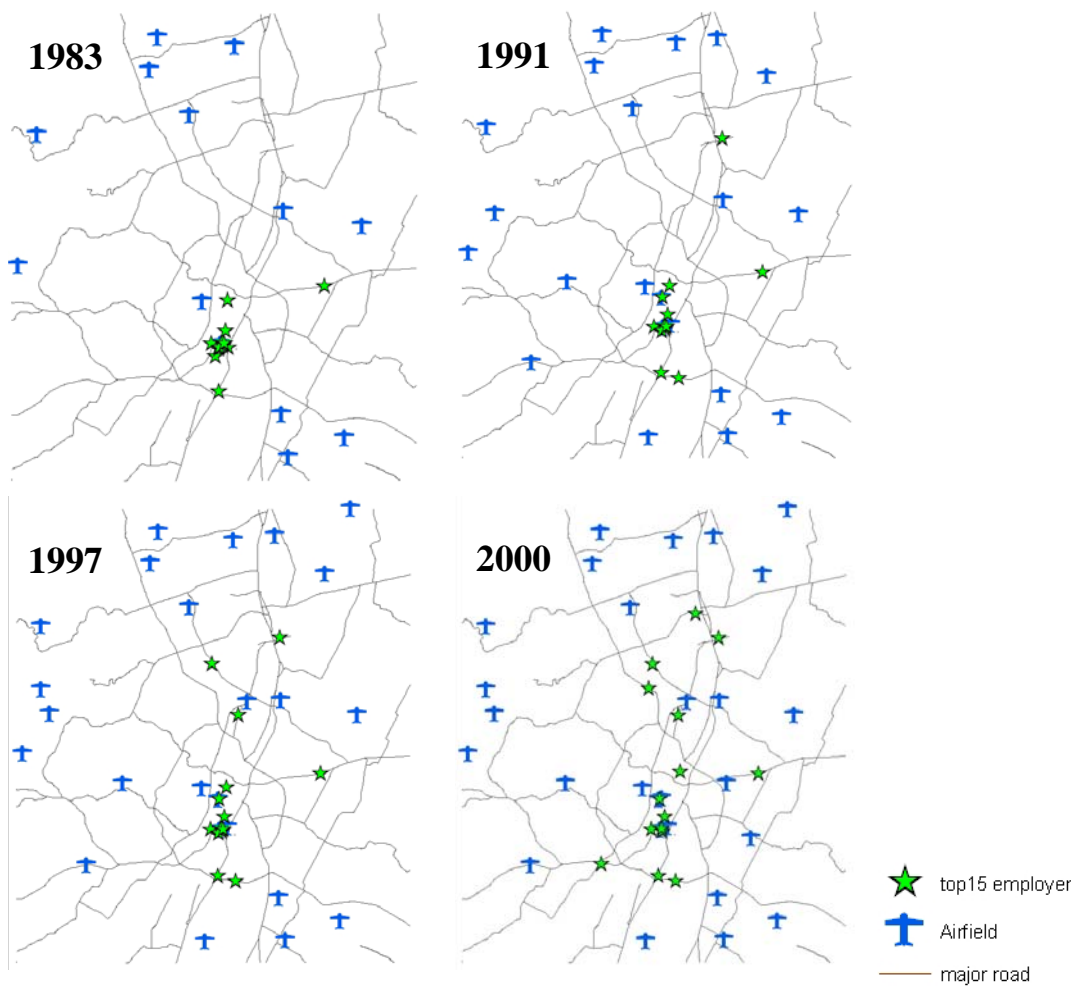


Figure 5.8 Locations of Key Sites and Highways in Different Years

5.2.3.4 Calculating Travel Time for Each Location

After travel cost to each 300m cell was defined as described in section 5.2.3.2 and the major facilities were located, travel times were calculated using the ArcMap’s “shortest path” routine. Each grid cell was given values indicating their travel time to the nearest major road, the nearest airfield, the CBD and the nearest top employer.

5.2.4 School Access

A location's accessibility to schools can be important to its development. Here, the number of Kindergarten through 12th grade (K-12) schools in the 3 km-radius neighborhood is calculated using ArcMap's "viewshed" function, based on the 2001 school information provided by the City of Austin (2007). Of course, over time new schools emerge, particularly in peripheral regions (as populations have grown). However, due to the lack of such, earlier school-siting information, this study assumes that the number of schools in any neighborhood has remained constant over the four model years. Figure 5.9 shows the spatial distribution of this variable.



Figure 5.9 Number of Schools in Each Cell's Neighborhood

5.2.5 Topographic Information

Topographic conditions can play an important role in site development. In this dissertation, two factors are considered: elevation and slope. These two variables may influence the intensity of land development because of their relationship to view lots, flood risk, and development costs.

The elevation and slope information are derived from CAPCOG's 10-foot resolution contour line map (CAPCOG, 2006). Elevations were calculated using ArcMap's spatial interpolation function, which generates a raster layer with values equaling the contour line values at locations where the lines actually appear. For locations between the lines, the function interpolates values based on neighboring contour lines using an "inverse distance weighted" algorithm (ESRI, 2006). The resulting elevations in the study area range from 186 to 1292 feet (i.e., 62 to 430 meters) above mean sea level. However, due to errors in the original contour line data (some contour line sections in high-elevation area have zero values), the interpolation returns some unrealistic values, clearly observable in Figure 5.10.

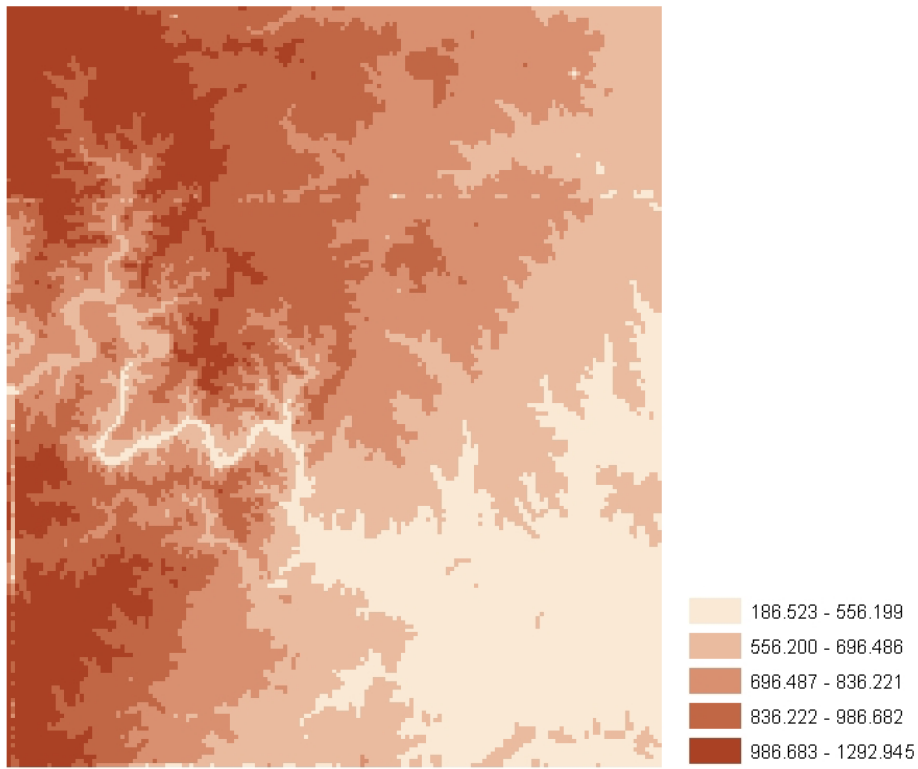


Figure 5.10 Elevation Distribution

Having obtained the elevation layer, slopes can be estimated using ArcMap’s “surface analysis” function. Because of the noise in elevation data (see Figure 5.10), the resulting slopes also contain some unrealistic values. Fortunately, the locations of these imperfections are easily detected, in the maps, and can be avoided when selecting the data sample.

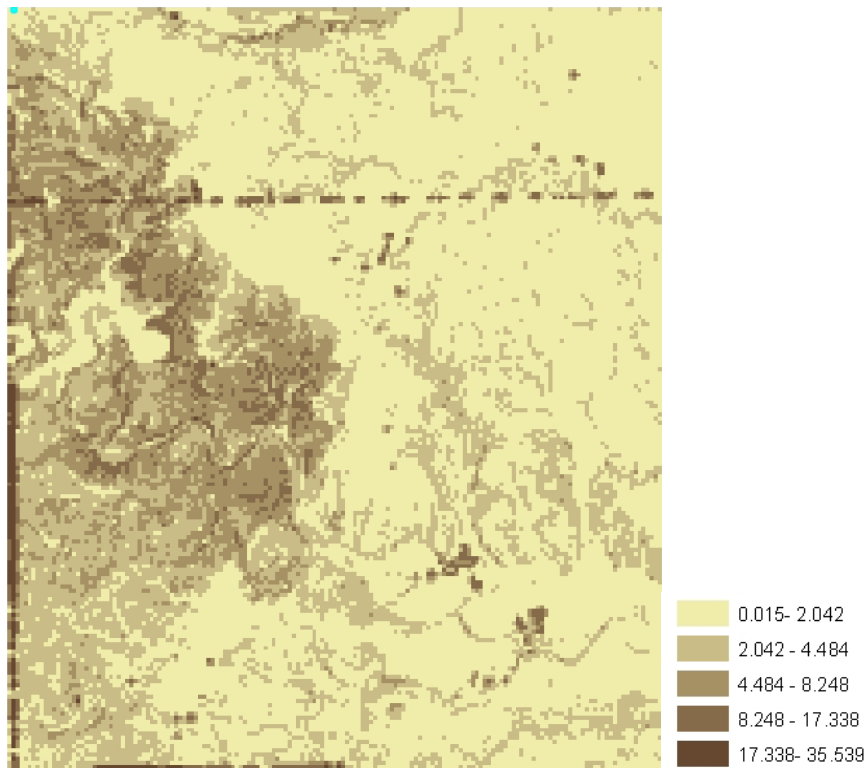


Figure 5.11 Slope Distribution

5.2.6 Defining Regions and Sample Selection

As discussed in Chapters 2 and 3, a reasonable rule for defining these is that observations in the same region share common features, and that these features differ across region boundaries. In ecological and environmental studies, boundaries may derive from natural spatial partitions, such as rivers and mountain ranges, with observations in the same region sharing vegetation and micro climate. For human activities, boundaries are more likely to be administrative units across which policies can change, such as zoning and school district scores.

In Austin’s urban area, zoning is based on neighborhood planning areas (NPAs). Changes in zoning constraints often occur across these boundaries. From this perspective, this NPA may be an ideal unit to be used as a region. However, only

information for the 98 NPAs in Travis County is readily available, so such information is not available for about 30% of the study area.

After careful consideration, regions in this study were defined based on zip codes, with several advantages. First, in the study area, a zip code area often overlaps with an NPA, or is a union of 2 to 4 NPAs. Second, some region boundaries also overlap with natural features like Austin's river. Furthermore, such a division produces 57 regions, a number large enough to offer interesting regional variation while small enough to moderate computational burdens.

After defining these regions, the next step was to select observations (grid cells) in each region. Of course, one can use all 29,946 300m grid cells as observations. However, there are good reasons for selecting only a subset of observations. First, the "boundary" of a region may be somewhat ambiguous and the differences between regions may be slight. If all grid cells are used, cells that are located in two different regions yet lie in close proximity may be more similar than grid cells that are far away from each other yet belong to the same region. The second reason is computational: 29,946 grid cells create a very large pool of observations with heavy computational burden for parameter estimation. A 10% sampling rate ($\sum n_i = 2,995$) is expected to return satisfactory estimation results with significantly reduced computation time and was used here.

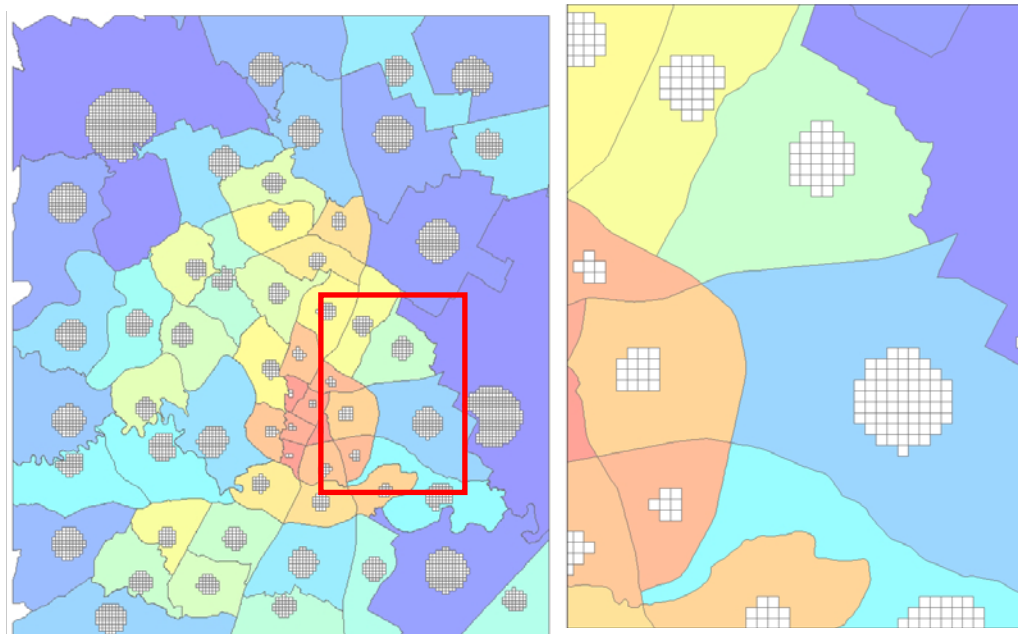
In order to ensure that observations in the same region are more alike than those in different ones, samples were selected around regional (zip code area) centroids. In this way, observations in the same region are spatially clustered (all contiguous) and thus expected to be more similar to each other than to observations in other regions.

Second, in order represent the entire study area, samples should distribute as evenly as possible across space. If an equal number of 300m cell observations is selected in each region, smaller regions will get more weight (than they "deserve")

in the sample. In order to spatially balance the selection, the number of observations in each region was set proportional to the region area.

Finally, 224 sampled cells were removed. These included those extending into a neighboring region (which only occurred in very narrow regions), those along edges of the study area (whose neighborhood information could not be obtained), and those falling into the areas (as discussed in Section 5.2.5) where elevation and slope values are unrealistic.

A total of 2,771 observations resulted from this processing. These observations are distributed across the 57 regions (zip code areas), and the number of observations per region ranges from 2 to 333. Figure 5.12 shows these regions and the sampled cells.



(a) Distribution of Selected Sample (b) Detail of the Highlighted Area

Figure 5.12 Selected Sample for Development Intensity Analysis

5.2.7 Summary of Land Development Data

In summary, variables for the development intensity DSOP analysis include elevation, slope, number of (neighborhood) schools, neighborhood population, and workers, average household income, travel time to top employers, travel time to the CBD, travel time to the nearest airfield, and travel time to the nearest major road. Table 5.2 summarizes definitions of all these variables, and Table 5.3 summarizes their statistics (over the 2,771 observations in each year). Table 5.3's statistics confirm the trends illustrated in Figures 5.1 through 5.7: over the years, development intensity levels, population, number of workers, and average household income have all increased. Average travel times (without considering congestion) to major facilities and employers have decreased because of new road in peripheral zones.

Table 5.2 Data Description for Land Development Intensity Level Analysis

Variable	Description
INTLV	Development intensity level
ELEVTN	Average elevation of the 300m grid cell (km)
SLOPE	Average slope of the 300m grid cell (%)
NSCHOOL	Number of K-12 schools in the neighborhood
POP	Population (thousand) in the neighborhood
WORKER	Number of workers (thousand) living in the neighborhood
INC	Average household income (thousand dollars) in the neighborhood
EMPTT	Travel time to nearest major (top 15) employer (hours)
CBDTT	Travel time to CBD (hours)
AIRTT	Travel time to nearest airfield (hours)
RDTT	Travel time to nearest highway (hours)

Table 5.3 Summary Statistics for Land Development Intensity Analysis

	Variable	Minimum	Maximum	Mean	Std. Deviation
Constant through Years	ELEVTN	0.136	0.390	0.251	0.061
	SLOPE	0.034	17.328	2.699	2.196
	NSCHOOL	0.000	7.000	1.208	1.377
1983	INTLV	0.000	3.000	0.826	0.774
	POP	0.225	37.531	4.632	7.298
	WORKER	0.121	19.997	2.408	3.918
	INC	17.330	88.941	45.368	15.109
	EMPTT	0.004	1.115	0.453	0.223
	CBDTT	0.000	0.358	0.154	0.070
	AIRTT	0.005	0.784	0.345	0.157
	RDTT	0.002	0.498	0.111	0.093
1991	INTLV	0.000	3.000	0.948	0.874
	POP	0.203	51.310	6.860	10.424
	WORKER	0.121	27.633	3.624	5.652
	INC	20.540	105.412	53.844	17.766
	EMPTT	0.004	0.733	0.298	0.149
	CBDTT	0.000	0.339	0.148	0.068
	AIRTT	0.004	0.630	0.259	0.120
	RDTT	0.002	0.430	0.092	0.082
1997	INTLV	0.000	3.000	1.300	0.827
	POP	0.389	64.873	8.007	12.615
	WORKER	0.211	35.220	4.240	6.900
	INC	23.332	119.738	61.077	20.341
	EMPTT	0.001	0.313	0.112	0.060
	CBDTT	0.000	0.308	0.142	0.065
	AIRTT	0.004	0.628	0.227	0.116
	RDTT	0.002	0.385	0.086	0.074
2000	INTLV	0.000	3.000	1.359	0.929
	POP	0.478	64.629	9.131	13.153
	WORKER	0.238	36.238	4.836	7.278
	INC	15.869	125.094	65.024	22.635
	EMPTT	0.001	0.182	0.070	0.037
	CBDTT	0.000	0.266	0.126	0.057
	AIRTT	0.005	0.437	0.154	0.070
	RDTT	0.002	0.251	0.054	0.044

5.3 OZONE CONCENTRATION LEVELS

As discussed in Chapter 1, various factors can influence ozone concentration levels. Ozone concentrations are forecast based on photochemical and deposition processes. One example is the widely used Comprehensive Air Quality Model with Extensions (CAMx). The CAMx model requires many inputs, including several meteorological condition indicators, land cover information, and dry deposition algorithms. (One may see Environ (2007) for a detailed introduction.)

This dissertation provides an alternative method for ozone concentration estimation, one that does not require an understanding of the details of the physical processes involved. This analysis seeks statistical relationships between various factors and ozone concentration levels. As an empirical example, it also illustrates application of the DSOP model in cases where each region contains only one individual ($n_i = 1 \ \forall i$).

The following sections describe variables used in this analysis of ozone concentration levels, including temperature, street length and fractions of different land cover types. Section 5.3.1 describes how ozone concentration levels were derived from the original data; Section 5.3.2 contains the temperature information; and Section 5.3.3 discusses how transportation and land cover information was incorporated and how these variables were interacted with time-of-day indicators.

5.3.1 1999 Ozone Concentration Levels

Ozone concentration levels were derived from continuous values originally prepared for an EPA project, and provided by Dr. Elena McDonald-Buller at the University of Texas at Austin (CAPCO et al., 2004). Using the CAMx model, many emissions inventories and a variety of behavioral assumptions, the project

researchers developed hourly ozone concentration estimates for a high-ozone episode: September 13-20, 1999.

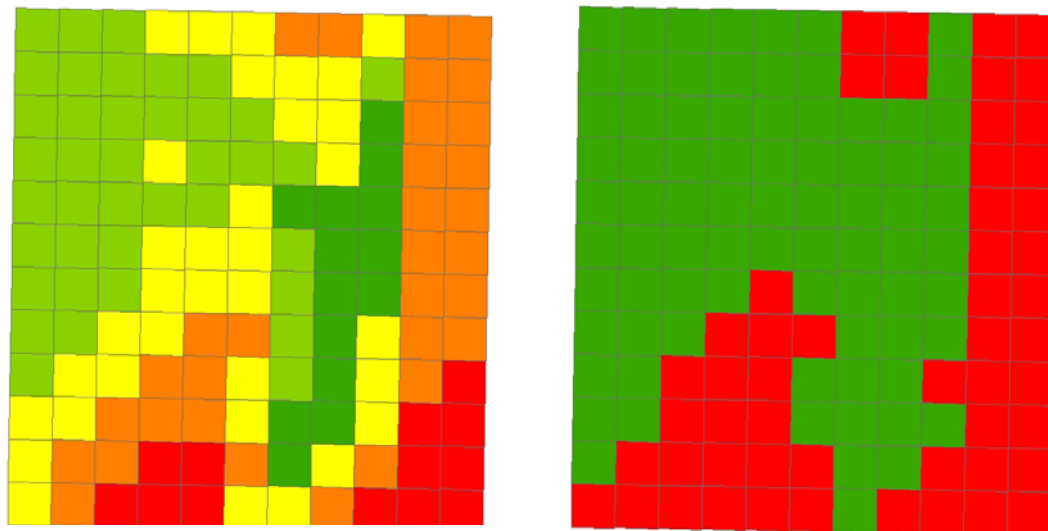
In the CAPCO (Capital Area Planning Council) study, there are three levels of spatial resolution and the finest is 4km. This 4km resolution area covers a 360 km x 432 km area (i.e., 90 x 108 grid cells) and includes all major urban centers within southern Texas and the Texas Gulf Coast.

In this dissertation, hourly ozone concentration data over one day (September 13, 1999) was selected. In addition, in order to make use of the transportation and land cover information already derived, only grid cells falling into the study area for development intensity analysis was used, resulting in a 44 km x 48 km area containing 132 such 4km x 4km grid cells. Thus, the resulting dataset is a 132 (*N*) × 24 (*T*) panel with values indicating ozone concentrations in parts per million (ppm). Each of these grid cells is treated as a region (and also as an individual).

The rule for defining ozone concentration levels is similar to the rule defining development intensity levels: the rule needs to be flexible and adaptable to the user's needs and every category needs to contain enough observations so that each is well represented. Here, the values were categorized into 5 groups: values below 0.035 are assigned Level 1, values between 0.035 and 0.04 are Level 2, those between 0.04 and 0.45 are Level 3, those between 0.045 and 0.05 are Level 4, and those above 0.05 are categorized as Level 5.¹⁸

Figure 5.13 illustrates the continuous ozone concentration values and their corresponding levels using data between 4 and 5pm on September 13, 1999 as an example.

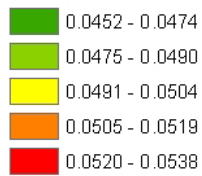
¹⁸ An area is designated as “non-attainment” when an 8-hour ozone concentration average exceeds 0.08 ppm. Ideally, this might constitute a key response level; however, the sample data does not contain any observations with such a high concentration.



(a) Ozone Concentration Values (ppm)

(b) Coded Concentration Levels

Legend (a)



Legend (b)



Figure 5.13 Ozone Concentration Values and Corresponding Levels (4 to 5pm on Monday, September 13, 1999)

Table 5.4 shows the changing trend of ozone concentration levels during the 24 hours: the levels are higher during daytime, especially in the afternoon, and lowest at night and in the early morning.

Table 5.4 Frequency of Ozone Concentration Level through a Day

Hour	Number of Grid Cells with Different Ozone Concentration Levels				
	1	2	3	4	5
0	2	130	0	0	0
1	33	99	0	0	0
2	55	77	0	0	0
3	68	64	0	0	0
4	93	39	0	0	0
5	80	52	0	0	0
6	88	44	0	0	0
7	115	17	0	0	0
8	119	13	0	0	0
9	64	65	3	0	0
10	0	46	86	0	0
11	0	0	32	100	0
12	0	0	0	103	29
13	0	0	1	110	21
14	0	0	1	88	43
15	0	0	0	64	68
16	0	0	0	79	53
17	0	0	4	58	70
18	0	0	0	17	115
19	0	0	0	9	123
20	0	0	6	58	68
21	0	7	66	36	23
22	1	43	87	1	0
23	7	87	38	0	0
Total	725	783	324	723	613

5.3.2 Temperature Distribution

Austin’s neighborhoods’ temperature information comes from the same EPA project datasets, provided by Dr. McDonald-Buller. This information was derived

using a revised MM5¹⁹ model (CAPCO et al., 2004). Table 5.5 illustrates the distribution and changes in temperatures over the 132 cells and 24 hours.

Table 5.5 Frequency of Temperature through a Day

Hour	Number of Grid Cells at a Given Temperature													
	°C	15*	16	17	18	19	20	21	22	23	24	25	26	27
	°F	59.0	60.8	62.6	64.4	66.2	68.0	69.8	71.6	73.4	75.2	77.0	78.8	80.6
0	0	0	0	72	60	0	0	0	0	0	0	0	0	0
1	0	0	1	111	20	0	0	0	0	0	0	0	0	0
2	0	0	6	126	0	0	0	0	0	0	0	0	0	0
3	0	3	121	8	0	0	0	0	0	0	0	0	0	0
4	0	60	72	0	0	0	0	0	0	0	0	0	0	0
5	0	111	21	0	0	0	0	0	0	0	0	0	0	0
6	51	81	0	0	0	0	0	0	0	0	0	0	0	0
7	7	109	16	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	94	38	0	0	0	0	0	0	0	0	0
9	0	0	0	0	4	128	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	132	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	108	24	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	132	0	0	0
13	0	0	0	0	0	0	0	0	0	0	132	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	107	25	0
15	0	0	0	0	0	0	0	0	0	0	0	132	0	0
16	0	0	0	0	0	0	0	0	0	0	0	132	0	0
17	0	0	0	0	0	0	0	0	0	0	16	116	0	0
18	0	0	0	0	0	0	0	0	0	47	85	0	0	0
19	0	0	0	0	0	16	116	0	0	0	0	0	0	0
20	0	0	0	0	2	130	0	0	0	0	0	0	0	0
21	0	0	0	74	58	0	0	0	0	0	0	0	0	0
22	0	0	0	132	0	0	0	0	0	0	0	0	0	0
23	0	56	76	0	0	0	0	0	0	0	0	0	0	0
Total	58	420	313	617	182	274	116	132	108	71	365	487	25	0

¹⁹ The MM5 or PSU/NCAR mesoscale model was developed by the National Center for Atmospheric Research at Pennsylvania State University to simulate/predict mesoscale atmospheric circulation.

5.3.3 Transportation and Land Cover Information

As introduced in Chapter 1, local traffic levels and land cover types may influence local ozone concentrations. Ideally, traffic counts and VMT by hour by cell would be available for use. These variables were not readily available (by time of day or for network links that are not coded or used in the region's TDM's network assignment routine), so the total length of street centerlines is used as a proxy for local VMT levels. (In reality, multi-lane facilities typically have higher traffic levels and so on.)

Land cover type influences ozone concentration because it contributes to both ozone generation (biogenic or anthropogenic) and deposition. Residential, commercial, transportation and industrial land (i.e., "developed" lands) may be categorized together, since they mainly contribute to anthropogenic emissions and their land cover materials may offer similar dry-deposition rates. Treed areas, brush, and agricultural land all contribute biogenic emissions and are expected to have similar dry deposition rates, so they may be aggregated as "vegetation." Barren land and water, though having quite different dry deposition rates, only account for a small proportion of the land in the study area, and so have been grouped together, as "undeveloped land", in order to avoid possible multi-collinearity issues.

The land cover information comes from the year-2000 satellite data provided by Dr. Parmenter (with 30 meter resolution). The month and day of this satellite data is September 3, very close to the model day, September 13. Hence, seasonality differences may be ignored. Based on this satellite data, fractions of the three aggregate land cover types described above were calculated. Furthermore, to account for variations in human activities and the effect of daylight (which can be influential to both ozone generation and deposition) across different times of day, these transportation and land cover fractions were interacted with several time-of-

day indicators. Total street length was multiplied by an indicator for peak travel hours (i.e., 7:00 to 10:00 and 16:00 to 19:00) and non-peak hours (any other time of day); developed land was interacted with working hours (i.e., 8:00 to 17:00) and non-work hours; and, because plant activity is strongly influenced by the presence of daylight, vegetated lands were interacted with a “day time” indicator (6:00 through 18:00) an night-time indicator (18:00 through 6:00). The fraction of undeveloped land was used as the base case.

5.3.4 Summary of Ozone Model Data

The dataset used for the ozone concentration model contains 132 individuals (or regions, in this case) over 24 hours, providing a total of 3,168 data points. Explanatory variables include temperature, street lengths interacted with indicators for peak/non-peak hours, percentages of developed land interacted with indicators for work/non-work hours, and percentages of vegetated land interacted with indicators for day/night time conditions. Table 5.6 summarizes definitions and statistics of all these variables. The mean and standard deviation of the ozone concentration levels imply that the dependent variable values are well balanced (i.e., each level has adequate observations, also shown by Table 5.4). The large standard deviations of all explanatory variables (as compared to the mean values) indicate substantial data variability.

Table 5.6 Data Description for Ozone Concentration Level Analysis

Variable	Description	Min	Max	Mean	Std. Dev.
OZONE	Ozone concentration level	1.00	5.00	2.91	1.47
TEMP	Temperature (Centigrade)	15.35	27.05	21.01	3.73
PEAKTRAF	Total length of street centerline (km) × indicator for peak travel hour	0.00	208.93	19.99	37.76
NONPTRAF	Total length of street centerline (km) × indicator for non-peak hour	0.00	208.93	39.99	45.32
WKDEV	Percentage of developed land (%) × indicator for work hours	0.00	93.43	14.61	22.66
NWKDEV	Percentage of developed land (%) × indicator for non-work hours	0.00	93.43	20.46	24.48
DTVEG	Percentage of vegetation (%) × indicator for day time	0.00	98.59	33.81	35.69
NTVEG	Percentage of vegetation (%) × indicator for night time	0.00	98.59	28.61	35.02
UNDEV	Percentage of under land (%; hold as base case)	0.00	30.92	2.51	5.52

5.4 CHAPTER SUMMARY

This chapter introduced the two distinct datasets used for model application, based on development intensity levels and ozone concentration levels over space and time. Sections 5.2 through 5.4 described in detail how all variables were derived, why they are expected to help explain land development and air quality dynamics, what these variables (statistically) look like, and what the potential data quality problems are.

In the next chapter, these two datasets will be analyzed using the DSOP model, thereby quantifying the influence of the various explanatory variables while illuminating the nature and level of spatial autocorrelation.

CHAPTER 6. RESULTS OF DATA ANALYSIS

6.1 INTRODUCTION

This chapter applies the DSOP model to both datasets: land development intensity levels and ozone concentration levels. As noted in Chapter 3, explanatory variables for both analyses include temporally lagged latent variables and various contemporaneous variables.

The following sections discuss estimation of both models and results. Estimates of parameter expected values and the statistical and practical significance of each variable are discussed. Finally, model estimates are used to predict response variables' values under hypothetical scenarios. The predictions can be visualized via a “most likely” result and an “uncertainty index.”

6.2 ESTIMATION OF LAND DEVELOPMENT INTENSITY LEVELS

In this section, land development intensity levels are analyzed using the DSOP model. First, the number of burn-in samples is determined. Estimate means, standard deviations, posterior distributions, and their marginal effects are then calculated and discussed. The performance of the DSOP model is compared to those from simpler models; and, finally, one-step predictions are generated and evaluated in terms of their relative uncertainty.

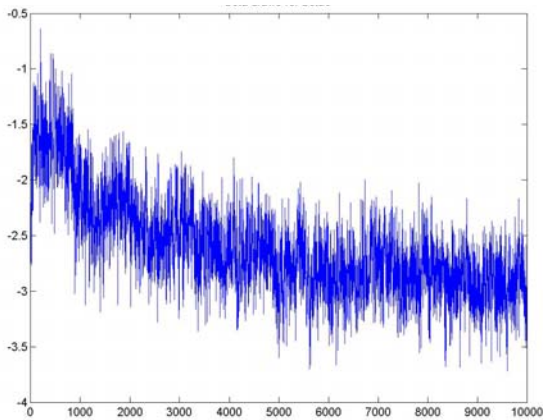
6.2.1 Estimation Results

As discussed in Chapter 3, one important issue in using a Bayesian approach is deciding when or whether the estimation converges. Rigorous proof of convergence is a complicated topic, so here “convergence” is based on the trace of variable estimates. If after a certain number of iterations, parameter estimates

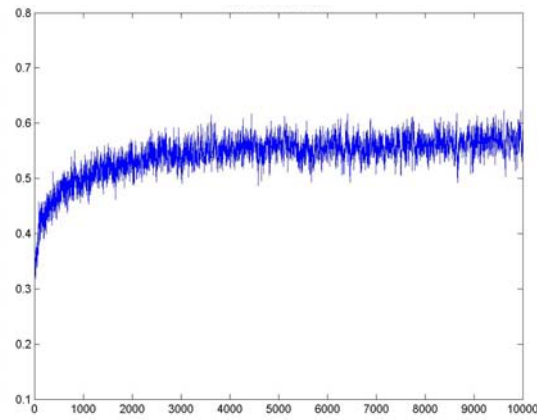
become stable, the estimation is designated as having converged. Results of iterations before this turning point are omitted and all inferences are drawn based on the converged iterations.

Figure 6.1 shows several typical estimation traces (convergence patterns) for parameters in the development intensity model. These patterns are representative, and the traces of other parameter estimations are all similar to them.

The model begins with diffuse priors and iterates 10,000 times. As observed in Figure 6.1, different parameters start “converging” after different numbers of runs. However, after 6000 runs, all traces appear stable, indicating an overall model convergence. Hence, the first 6000 runs were omitted (as a “burn-in” sample), and the model uses the latter 4000 runs to estimate parameter means and standard deviations, as shown in Table 6.1.



(a) Trace of β_{AIRT} AIRT



(b) Trace of λ

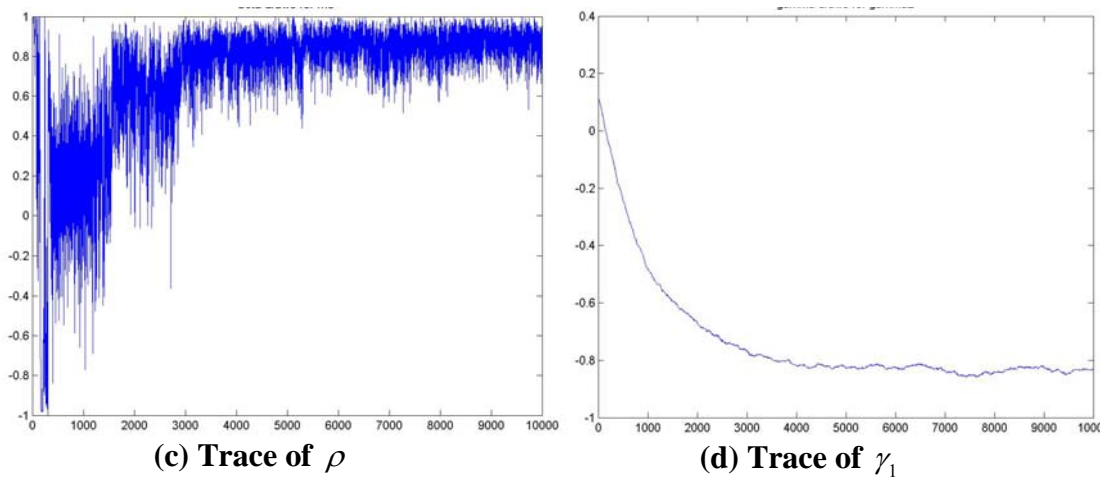


Figure 6.1 Convergence Patterns of Development Intensity Level Estimation

According to the results, neighborhood population and worker counts do not have statistically significant impacts on land development intensity level. Average household income, by contrast, appears to generally boost the development intensity. Distances to major employers, CBD area, and the nearest airfield all have statistically and practically significant effects on land development: the farther the cells lie from these attractions, the less likely they are to develop intensely. Interestingly, distance to highway is estimated to have a negative marginal effect on intensity, implying that (in the study area) development is more likely to occur at locations far from major roads. Considering that distances to the CBD and major employers already have been controlled for, this result can be interpreted as such: after access to work and the region's core is determined, developers tend to choose locations some distance away from the highway (and its noise, pollutants and safety issues). The result also suggests that locations with more neighborhood schools are more likely to be intensely developed while elevation is not a statistically influential factor, locations with steeper slopes are less attractive to land development.

Table 6.1 Estimation Results for Model of Land Development Intensity Levels

Variable	Mean	Std. Dev.	t-stat.
POP	-0.024	0.036	-0.668
WORKER	0.089	0.067	1.327
INC	0.019	0.002	9.143
EMPTT	-0.232	0.130	-1.778
CBDDT	-4.365	0.851	-5.126
AIRTT	-2.867	0.248	-11.550
RDDT	2.309	0.385	6.001
NSCHOOL	0.039	0.017	2.305
ELEVXN	-0.239	0.696	-0.343
SLOPE	-0.034	0.010	-3.394
λ	0.561	0.019	30.005
ρ	0.857	0.074	11.612
σ^2	0.871	0.222	3.931
γ_1	-0.834	0.011	-77.231
γ_2	2.235	0.031	71.393
γ_3	4.361	0.034	130.167

Unlike slope coefficients in a standard linear model, the beta estimates for this model involving latent variables cannot be used directly to determine the magnitude of their influence. In addition, as Greene (2005) explains, parameter signs in a model of ordered categorical response only indicate changes in likelihood of the two extreme outcomes ($y = 1$ and 4). Therefore, detailed discussions on the overall effects of such factors are provided in Section 6.2.2.

Another important estimation result is the practical and statistical significance of both the temporal autocorrelation coefficient (λ) and the spatial autocorrelation coefficient (ρ). These suggest that prior-period information has a very important influence on the (current) latent variable's value (mean $\lambda = 0.561$) and that, even after controlling for various neighborhood characteristics, residuals remain

strongly and positively correlated across space (mean $\rho=0.857$). These results support the notion that land development decisions depend heavily on neighboring conditions, and that spatial relationships should be reflected in model specification.

As a further confirmation, the mean values of regional specific error (θ_i) estimates (and their statistical significance) are shown in Figure 6.2. A clustering pattern (where similar values tend to co-locate, rather than lie randomly distributed across space) is clearly visible in this figure, so the spatial autocorrelation of these regional-specific error terms was tested using Moran's I (Moran, 1950), in ArcMap. It should be noted that the weight matrix used in ArcMap is based on the inverse of distance, not the contiguity approach used in this dissertation. And, of course, methodologically, Moran's I is quite different from a Bayesian approach. Therefore, any similarity with ρ may be limited to signs and general statistical significance.

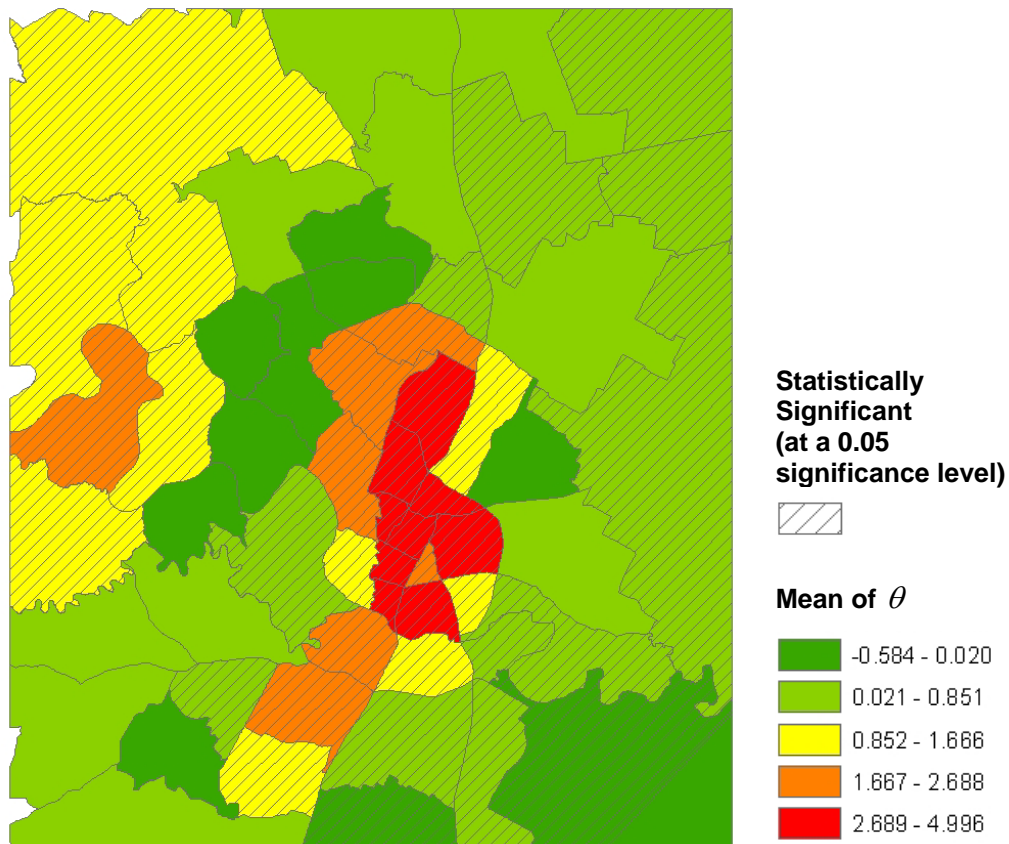


Figure 6.2 Distribution of Region-Specific Error Term Estimates (θ) for Land Development Intensity Levels

As expected, the Moran's I test results in ArcMap indicate clustering (i.e., positive spatial autocorrelation) of the θ values, across space. (The Moran's I value is very high: 0.56 with Z score of 6.7).

Figure 6.3 shows the estimation results for variances of these individual specific errors (v_i). Except for downtown regions, where only a few grid cell observations exist per region, all variance estimates are statistically significant. "City edges" (i.e., areas between Austin's central, highly developed area and the outer, less

developed areas) tend to have larger variances because these are where new developments are most likely to occur.

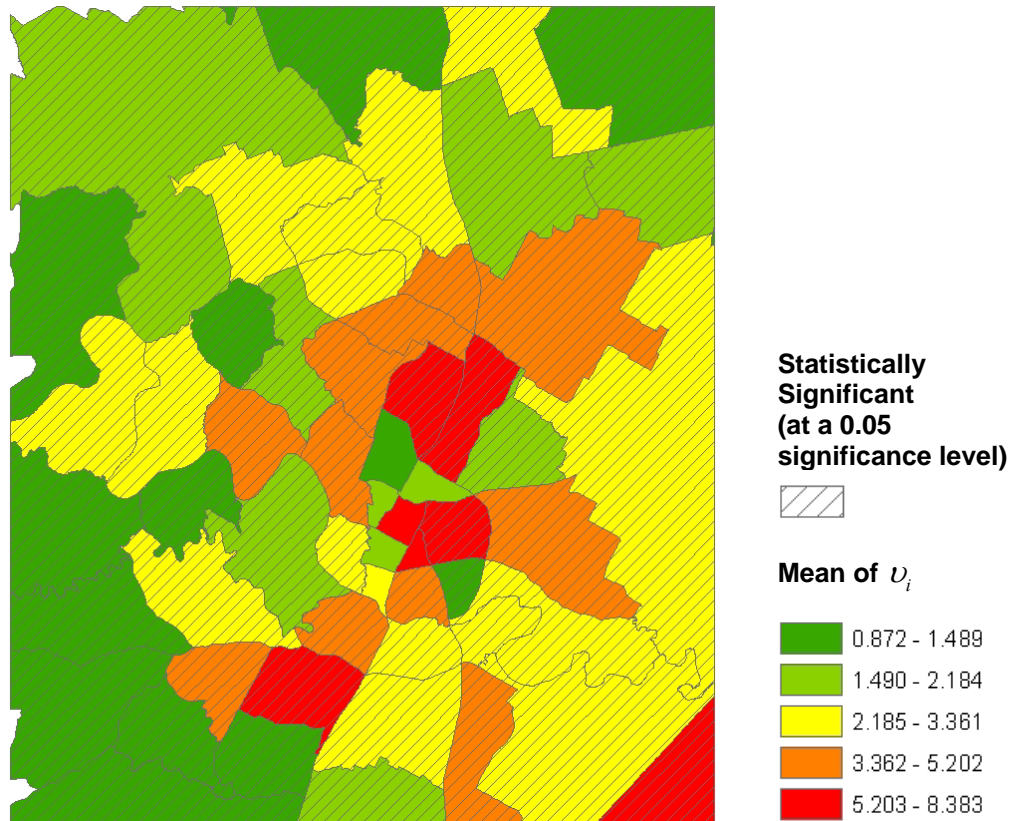
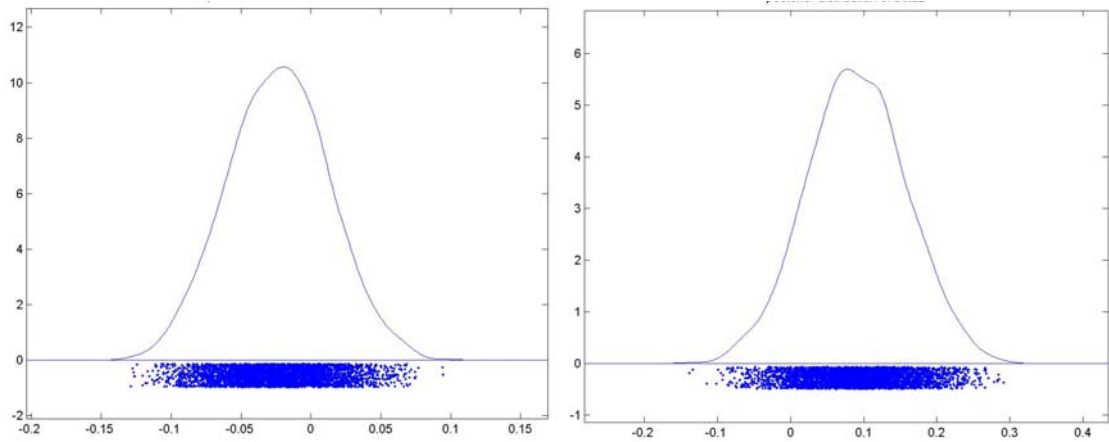


Figure 6.3 Distribution of the Variances of Individual Specific Error Term Estimates (v) for Land Development Intensity Levels

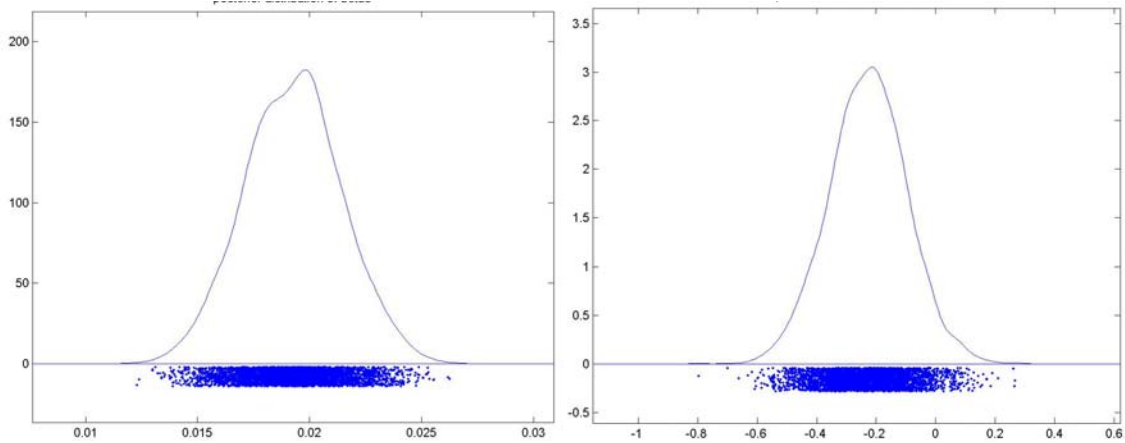
Figure 6.4 shows the posterior distributions of all parameters, based on the final 4000 runs. As also summarized in Table 3.1, all exogenous control variables are specified to follow normal posterior distributions. As expected λ has a truncated normal distribution, ρ has a non-standard distribution, and σ^2 follows a Chi square distribution. The posterior distributions of threshold parameters γ are very interesting. In Chapter 3, they are shown to follow a normal distribution mixed with a multivariate uniform distribution. According to graphs (n) through (p) in Figure 6.4, the resulting distributions are multimodal. Additionally, as expected,

the shapes of γ s present some similarity, suggesting that their values are co-dependent. This is to be expected: based on the γ posterior distribution (as shown in Table 3.1), it is clear that γ_s 's left threshold depends on γ_{s-1} and its right threshold depends on γ_{s+1} . This dependency can also be explained intuitively: the gap between γ_s and γ_{s+1} determines the probability of $y = s$. In order to maintain a generally constant gap, the values of γ_s and γ_{s+1} must move together.



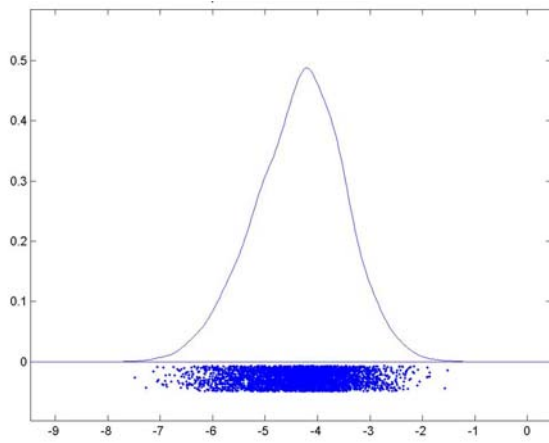
(a) Posterior Distribution of β_{POP}

(b) Posterior Distribution of β_{WORKER}

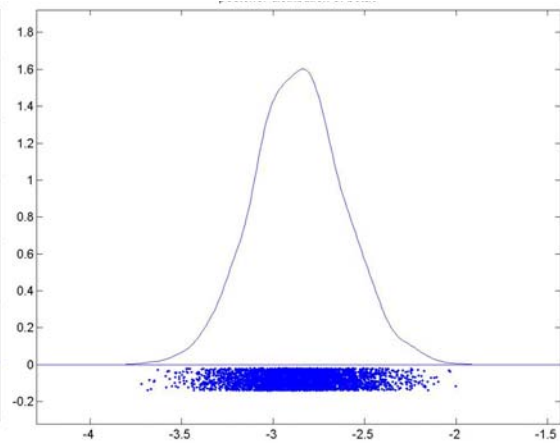


(c) Posterior Distribution of β_{INC}

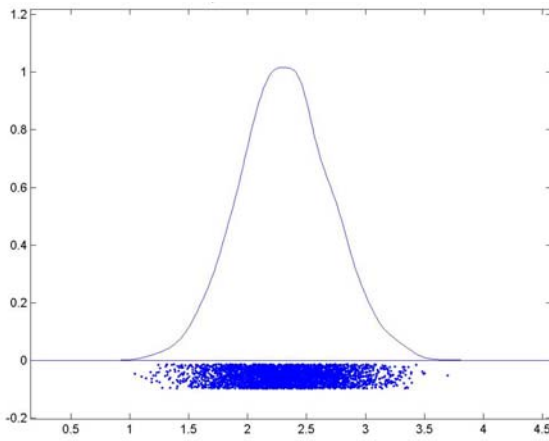
(d) Posterior Distribution of β_{EMPTT}



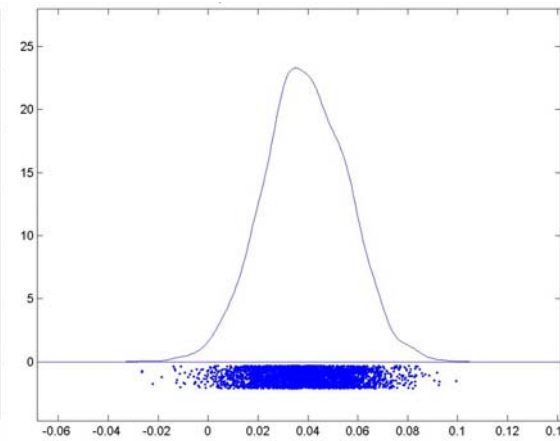
(e) Posterior Distribution of β_{CBDT}



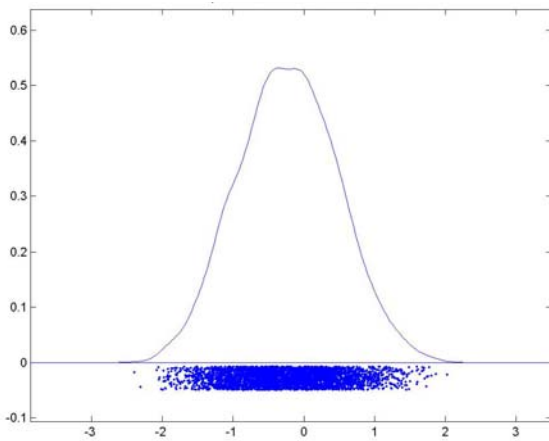
(f) Posterior Distribution of β_{AIRT}



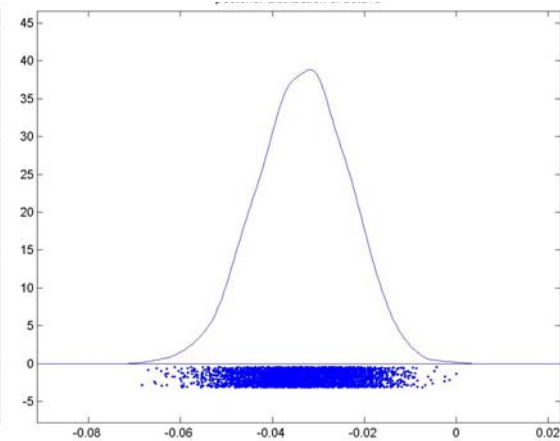
(g) Posterior Distribution of β_{RDT}



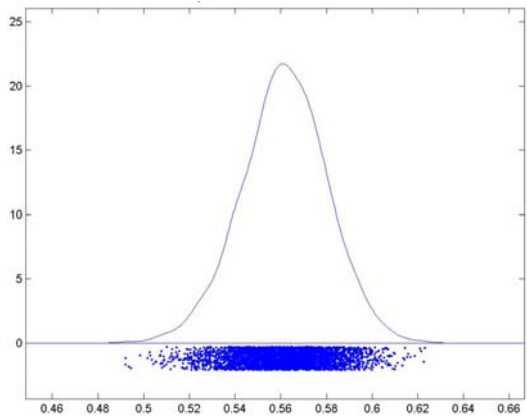
(h) Posterior Distribution of $\beta_{NSCHOOL}$



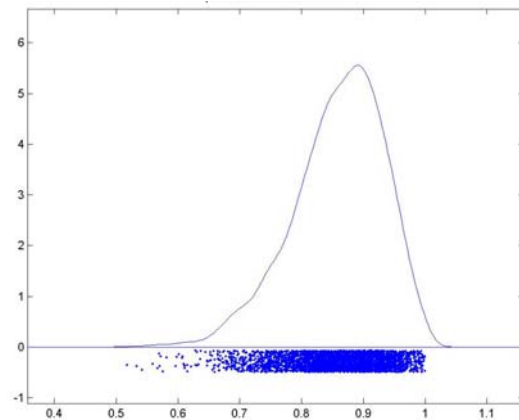
(i) Posterior Distribution of β_{ELEVXN}



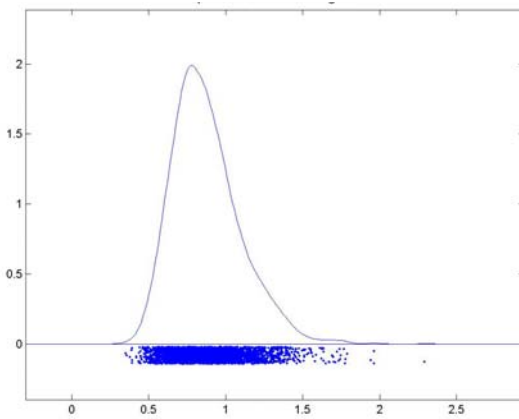
(j) Posterior Distribution of β_{SLOPE}



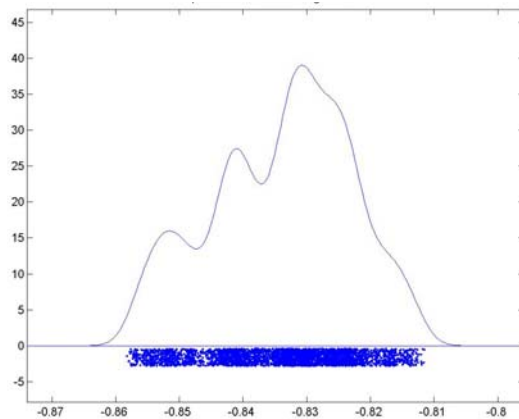
(k) Posterior Distribution of λ



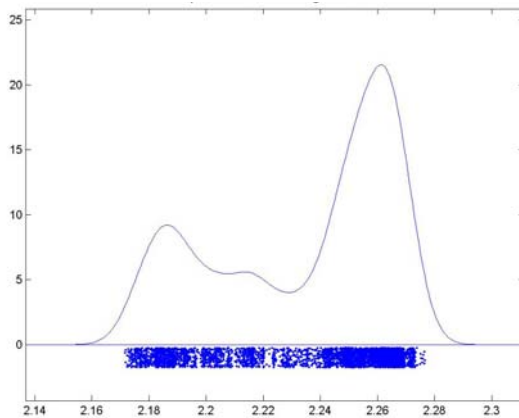
(l) Posterior Distribution of ρ



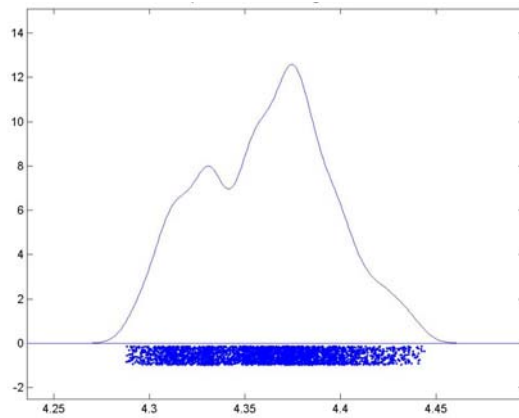
(m) Posterior Distribution of σ^2



(n) Posterior Distribution of γ_1



(o) Posterior Distribution of γ_2



(p) Posterior Distribution of γ_3

Figure 6.4 Posterior Distributions of Land Development Intensity Level Model Parameters

Table 6.2 shows the sample correlations between parameters, further illuminating the strong relationships across γ values. These results also indicate that the slope parameters for the population and number of workers variables are highly correlated (corr. = -0.93), suggesting that analysts may do well to remove one of these two highly correlated variables from the model specification. This problem is discussed in greater detail in Chapter 7.

6.2.2 Model Comparisons

Similar to Section 4.4, the performance of the DSOP model is compared to a standard ordered probit (OP) model, a dynamic ordered probit (DOP) model, and a spatial ordered probit (SOP) model. All these models were run using 10,000 draws, with the first 6,000 draws omitted (as the burn-in sample).

Table 6.3 summarizes the DIC values and prediction rates for each of these four specifications. In terms of model fit, it seems that no model is significantly better than the others. However, the DSOP model still outperforms all others, even after being penalized for using more parameters. Interestingly, while the OP model is unable to detect any spatial and temporal relationships in the dataset, its DIC values suggest it still may be preferred to the SOP and DOP models, thanks to its simpler/more parsimonious model specification. Of course, in terms of prediction rates the standard OP model only correctly predicts dependent values for 42.5% of the observations, and the SOP model increases this percentage to 47.8%. The DOP and DSOP models have quite close prediction rates: 48.7 and 48.8%. These results imply that, while considering dynamic and spatial patterns complicates the model specification and estimation process, the higher predictive accuracy may be worth the effort.

Table 6.2 Correlations between Parameter Estimates

	POP	WORKER	INC	EMPTT	CBDTT	AIRTT	RDTT	NSCHOOL	ELEVXN	SLOPE	λ	ρ	σ^2	γ_1	γ_2	γ_3
POP	1.000	-0.993	0.193	0.028	0.070	0.112	-0.130	-0.096	-0.061	-0.015	-0.084	-0.055	0.028	-0.008	-0.029	-0.020
WORKER	-0.993	1.000	-0.174	-0.009	-0.057	-0.114	0.126	0.091	0.043	0.017	0.043	0.039	-0.030	0.024	0.035	0.034
INC	0.193	-0.174	1.000	0.146	-0.077	0.325	-0.078	0.034	-0.330	-0.114	-0.320	-0.040	0.115	0.014	-0.075	0.031
EMPTT	0.028	-0.009	0.146	1.000	-0.068	-0.365	-0.137	-0.002	0.042	0.002	-0.050	-0.011	0.019	0.205	0.116	0.174
CBDTT	0.070	-0.057	-0.077	-0.068	1.000	-0.155	-0.434	-0.040	-0.256	-0.026	0.191	-0.017	0.004	0.108	0.111	0.144
AIRTT	0.112	-0.114	0.325	-0.365	-0.155	1.000	-0.288	0.032	-0.054	-0.117	0.070	-0.059	0.017	0.069	0.030	0.044
RDTT	-0.130	0.126	-0.078	-0.137	-0.434	-0.288	1.000	0.019	-0.019	0.066	-0.095	0.052	0.011	-0.173	-0.132	-0.168
NSCHOOL	-0.096	0.091	0.034	-0.002	-0.040	0.032	0.019	1.000	-0.059	-0.036	-0.059	-0.006	-0.016	-0.016	0.007	-0.030
ELEVXN	-0.061	0.043	-0.330	0.042	-0.256	-0.054	-0.019	-0.059	1.000	0.022	0.046	-0.107	-0.034	0.141	0.133	0.068
SLOPE	-0.015	0.017	-0.114	0.002	-0.026	-0.117	0.066	-0.036	0.022	1.000	0.092	-0.007	-0.008	-0.051	-0.028	-0.058
λ	-0.084	0.043	-0.320	-0.050	0.191	0.070	-0.095	-0.059	0.046	0.092	1.000	0.000	-0.066	-0.031	0.026	-0.016
ρ	-0.055	0.039	-0.040	-0.011	-0.017	-0.059	0.052	-0.006	-0.107	-0.007	0.000	1.000	-0.261	-0.037	-0.044	0.010
σ^2	0.028	-0.030	0.115	0.019	0.004	0.017	0.011	-0.016	-0.034	-0.008	-0.066	-0.261	1.000	0.025	-0.029	0.056
γ_1	-0.008	0.024	0.014	0.205	0.108	0.069	-0.173	-0.016	0.141	-0.051	-0.031	-0.037	0.025	1.000	0.789	0.823
γ_2	-0.029	0.035	-0.075	0.116	0.111	0.030	-0.132	0.007	0.133	-0.028	0.026	-0.044	-0.029	0.789	1.000	0.544
γ_3	-0.020	0.034	0.031	0.174	0.144	0.044	-0.168	-0.030	0.068	-0.058	-0.016	0.010	0.056	0.823	0.544	1.000

Note: Shaded values all exceed 0.5.

Table 6.3 Goodness of Fit and Prediction Rates using Different OP Model Specifications

Models	DIC	Predicted y Value	Actual y Value				% Cases Correctly Predicted (%)
			1	2	3	4	
DSOP	22587.9	1	1106	1417	179	47	48.8
		2	1354	3281	767	238	
		3	165	780	435	237	
		4	41	188	258	591	
DOP	23080.3	1	1120	1379	171	40	48.7
		2	1310	3237	750	236	
		3	208	841	479	279	
		4	28	209	239	558	
SOP	23091.3	1	1080	1379	187	49	47.8
		2	1294	3261	778	252	
		3	235	783	417	268	
		4	57	243	257	544	
OP	22800.0	1	992	1606	258	57	42.5
		2	1307	2913	770	324	
		3	273	822	371	299	
		4	94	325	240	433	

6.2.3 Marginal Effects

Based on Chapter 3's model specification, the marginal effects of explanatory variables X on the probabilities of each outcome level can be defined as follows:

$$\frac{\partial P(y_{ikt} = s)}{\partial x_{iktq}} = \left(\phi \left(\frac{\mu_{s-1} \lambda U_{ikt-1} - X_{ikt} \beta - \theta_i}{\nu_i} \right) - \phi \left(\frac{\mu_s - \lambda U_{ikt-1} - X_{ikt} \beta - \theta_i}{\nu_i} \right) \right) \cdot \beta_q \quad (6.1)$$

This marginal effect indicates the effect that a one-unit change in explanatory variable x_{iktq} has on the probability of different discrete outcomes, s . As mentioned in Section 6.2.1, the marginal effects on intermediate probabilities are not obvious at first glance, since a shift in the distribution can cause the

probability of intermediate response types to fall or rise, depending on the positioning of the average response (see, e.g., Wang and Kockelman, 2005).

As Equation (6.1) suggests, one variable's marginal effect is related not only to its own coefficient, but also to the values of all other coefficients. In addition, it can be inferred from Equation (6.1) that for each run, each observation and each period, there is a unique set of marginal effect values. With MLE, the marginal effects are often calculated using the final parameter estimates and average variable values. In this dissertation, the marginal effects are calculated separately for every observation in each period, and for every iteration. The results are then averaged to acquire a single, average response estimate, for every variable. Results obtained in this way are more reasonable and contain more information than single-equation results. This is an advantage of using the Bayesian approach: any derived statistics can be calculated on the heels of estimation.

Table 6.4 shows the magnitudes of "one unit" of different variables relative to their standard deviations, along with estimates of final (average) marginal effects. As shown, when the neighborhoods' average household income increases by \$1,000, the sample population's average probability of intense development increases by a mere 0.26% and the probability that it remains undeveloped falls by 0.523%. In other words, considering the growth rate of household income (approximately \$1,300/year [Figure 5.3]), its effect on land development is practically negligible.

When travel time to the nearest top employer increases by 10 minutes (0.17 hour) in all zones, the population's average probability of remaining undeveloped is estimated to increase by about 1.1%, and Level 2 through Level 4 probabilities are estimated to fall by 0.1%, 0.4%, and 0.6%, respectively. Again, though statistically significant, this factor seems to have a negligible practical effect.

In contrast, travel time to the region's CBD has an impressive effect: A 10 minute increase is linked to a 10% decrease in the probability of Level 4 development across the sample. (The probabilities of Levels 2 and 3 also fall, while the probability of finding undeveloped land rises by 20%.) Obviously, in the study area, development tends to cluster around the CBD area, so distance to the CBD is a key variable. As also indicated by other, existing works (e.g, Kockelman, 1997, and Zhou and Kockelman, 2007), this distance-to-CBD variable seems to be more predictive than other measures of access.

Travel time to an airfield appears to have a moderate impact on land development: A 10 minute increase is associated with a decrease in developments Levels 2, 3, and 4 by 1.4%, 5.1%, and 6.4%, respectively.

Travel time to highways is predicted to have the reverse impact: once the travel time to major employers, CBD and airfields are given, a 10 minute increase in travel time to a highway is associated with 4.1% more Level 2 development (very likely to be residential, commercial, or industrial, dotted with vegetation) and 5.2% more Level 3 development (densely developed residential, commercial, or industrial land).

The number of schools is also practically insignificant: one more school in the 28 km² (3 km radius) neighboring area is associated with outcome probability variations of less than 1%.

Finally, a 10-degree increase in slope is estimated to be associated with 1.0%, 3.6%, and 4.5% decreases in the probability of Level 2 through Level 4 development, respectively. Considering that the average slope in the samples is only 2.7 degrees, the impact of slope is rather insignificant in practice.

**Table 6.4 Average Marginal Effects of Covariates on Development Intensity
Levels over All Observations**

Variable	Ratio to Std. Dev.	Marginal Effect ($\times 10^{-2}$) (Change in Response Probability)			
		Level 1	Level 2	Level 3	Level 4
POP	0.092	0.652	-0.070	-0.257	-0.324
WORKER	0.168	-2.417	0.261	0.955	1.201
INC*	0.053	-0.523	0.057	0.207	0.259
EMPTT*	8.529	6.309	-0.682	-2.492	-3.135
CBDTT*	15.39	118.6	-12.87	-46.92	-58.80
AIRTT*	8.634	77.86	-8.467	-30.84	-38.55
RDTT*	13.65	-62.77	6.804	24.83	31.13
NSCHOOL*	0.726	-1.048	0.114	0.415	0.519
ELEV TN	16.39	6.429	-0.713	-2.515	-3.202
SLOPE*	0.455	0.912	-0.099	-0.362	-0.451

Notes: * indicates that the variable is statistically significant at the 0.05 level. Change in variable is one unit (e.g., 1 hour in case of travel times (TT)). “Ratio to Std. Dev.” is the ratio of one unit (e.g., 1 hour) to the standard deviation observed in the data set, for each variable.

In addition to Table 6.4, Table 6.5 is provided, to illustrate the marginal effects of such covariates when evaluating the responses of a single, “average” data point. Since spatial/neighborhood effects are important here, the “average observation” refers to an actual observation with values close to the sample mean values. (This differs from the usual, standard OP estimates, where the [hypothetical] unit of study enjoys exactly the mean or median values of all explanatory variables.) Values of explanatory variables for this observation are shown in Table 6.5, and compared to the sample mean values. As can be observed, for statistically significant variables, the marginal effects on Level 1 development for this “average observation” are close to the averages over all observations (Table 6.4). For Levels 2, 3 and 4, however, it seems that the marginal effects are generally higher for Level 2, and decrease with the development level. This is quite reasonable for this single point because its current intensity level is 1 and the

probability of its developing into Levels 3 or 4 is expected to remain low, despite the marginal changes in control variables' values.

**Table 6.5 Marginal Effects of Covariates on Development Intensity Levels
(Using One Observation)**

Variable	Sample Mean	Observ. Value	Marginal Effect ($\times 10^{-2}$) (Change in Response Probability)			
			Level 1	Level 2	Level 3	Level 4
INTLV	1.108	0.000	---	---	---	---
POP	7.158	6.971	0.477	-0.254	-0.156	-0.067
WORKER	3.777	3.795	-2.618	1.340	0.868	0.410
INC*	56.33	69.33	-0.793	0.398	0.264	0.131
EMPTT*	0.233	0.220	13.237	-6.678	-4.397	-2.162
CBDTT*	0.143	0.152	154.6	-77.56	-51.49	-25.54
AIRTT*	0.246	0.173	112.2	-56.41	-37.35	-18.45
RDTT*	0.086	0.056	-87.85	44.05	29.27	14.53
NSCHOOL*	1.208	1.000	1.265	-0.637	-0.420	-0.208
ELEVTVN	0.251	0.306	-1.543	0.775	0.513	0.255
SLOPE*	2.699	2.767	-0.256	0.181	0.074	0.001

Notes: * indicates that the variable is statistically significant at the 0.05 level. Change in variable is one unit (e.g., 1 hour in case of travel times (TT)).

In summary, most of the contemporaneous variables are practically insignificant. This suggests that when developers make decisions, past land conditions (represented by the lagged, latent dependent variables) are a more important consideration than current conditions. However, current access levels and transportation conditions, (especially travel time to the CBD area) are highly influential, and one might expect that access is also a key concern in developers' minds.

6.2.4 Development Intensity Model Prediction

One important application of this model is prediction of development intensities in the near future. This dissertation predicts future intensity levels for the 2,771 grid cells in the selected sample areas. In the prediction scenario, population

doubles from year 2000. To reflect possible congestion, all travel times (to major employers, CBD, nearest airfield, and nearest highway) were increased by 30%.

Similar to the calculation of marginal effects via a Bayesian approach, predictions can be achieved alongside model estimation. Over the final 4000 runs, (estimates of the) latent dependent variable values for the year 2000, estimated parameter and error term values, and control-variables for future scenario are used to generate future latent dependent variables. These latent variables then are compared to the threshold parameter values in each run, and development intensity levels for each location are calculated. Thus, for each of the 4000 draws and for each cell, there is a predicted development intensity level. The most common (frequently appearing) land development intensity levels in these 4000 runs for each sampled cell are shown in Figure 6.5 (a). As expected, more intensely developed land appears around the downtown area.

Of course, this single “most likely” pattern will not occur with a high likelihood. There is great flexibility and uncertainty in the future of these 2,771 grid cells. To help planners appreciate (and visualize) such uncertainty, an entropy statistic is used (see, e.g., Wang and Kockelman 2006, McKay 1995 and Kotz and Johnston 1982). That is, the uncertainty associated with the set of 4 potential land covers in cell i is specified as follows:

$$uncertainty_i = \frac{-1}{\ln(4)} \sum_{s=1}^4 P_{is} \ln(P_{is}) \quad (6.2)$$

This formulation generates a value between 0 and 1 for each cell. The higher the value, the more uncertain the prediction for that cell. When all four future land development intensity levels have equal probabilities ($P_{is} = 0.25 \forall s$), uncertainty entropy equals 1, indicating maximum uncertainty. When the same land intensity level emerges in all 4000 simulations, this uncertainty value is 0. As illustrated in

Figure 6.5 (b), higher uncertainty appears around the intermediate areas of the study area, or the central-city's edge. At these locations, the potential for variation is relatively large, resulting in a higher degree of uncertainty.

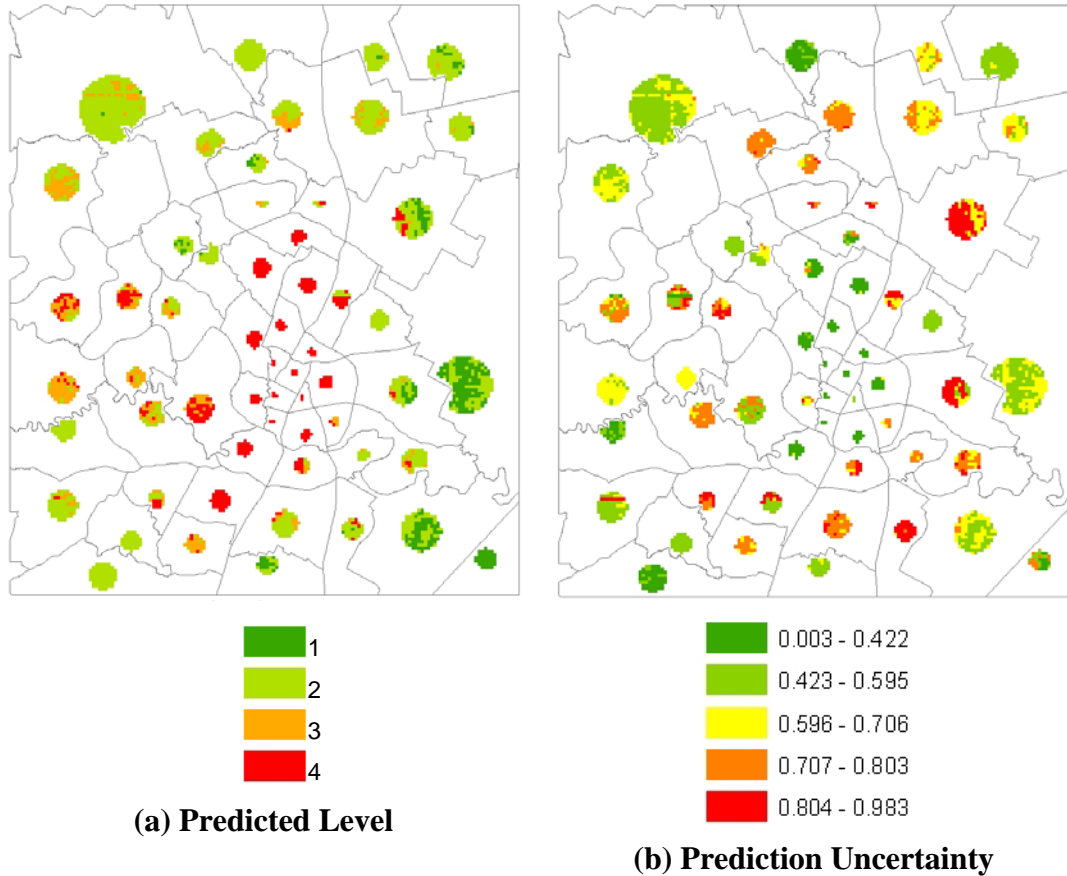


Figure 6.5 Most Likely Development Intensity Levels Prediction and Uncertainty (following an assumed doubling of population)

Table 6.6 compares these predictions to the year 2000 situation while there is strong similarity between land intensity levels in the two scenarios, some regression is apparent (i.e., changes from higher intensity to lower intensity levels). This is to be expected since some locations are presently more developed than the model would expect. Moreover, some locations may lose their attraction due to increases in travel time.

Table 6.6 Comparison of Base Year and Predicted Land Development Intensity Levels

		Most Likely Intensity Levels in Future Scenario				Total
		1	2	3	4	
Base Year Intensity Levels (Year 2000)	1	374	103	0	0	477
	2	11	1280	22	5	1318
	3	0	166	344	27	537
	4	0	2	71	366	439
Total		385	1551	437	398	2771

Model predictions, together with the previous discussion of parameter estimates and marginal effects, illustrate some interesting characteristics of Austin’s land development dynamics. The next section discusses analysis of a related but very different data set (ozone concentration levels), which follows a highly similar estimation and evaluation procedure²⁰.

6.3 ESTIMATION OF OZONE CONCENTRATION LEVELS

The estimation and analysis process for ozone concentration levels is similar to that for the land development intensity levels. As discussed earlier, burn-in samples are removed, based on the traces of estimates. Remaining values are summarized as means and standard deviations, together with their posterior distributions. Marginal effects of all variables are calculated and discussed, and predictions are carried out for a hypothetical scenario.

6.3.1 Estimation Results

As discussed in Section 3.4.3, since each region in the ozone dataset contains only one “individual.” Thus, individual-specific effects are assumed homoskedastic across regions. In addition, for the reasons discussed in Section

²⁰ Model comparisons (with simpler OP, DOP and SOP specifications) are omitted, along with calculation of single-point marginal effects.

4.3.1, all variances of individual-specific error terms are set to equal 1.0, so these errors all follow standard normal distributions.

Like the model for land development intensity, the model for ozone concentration level is initiated with diffuse priors. The total number of iterations is 8,000. As Figure 6.6 suggests, after 4000 runs all traces become stable, indicating convergence. Therefore, the first 4000 runs are omitted, and all inferences are drawn from results in iterations 4001 to 8000.

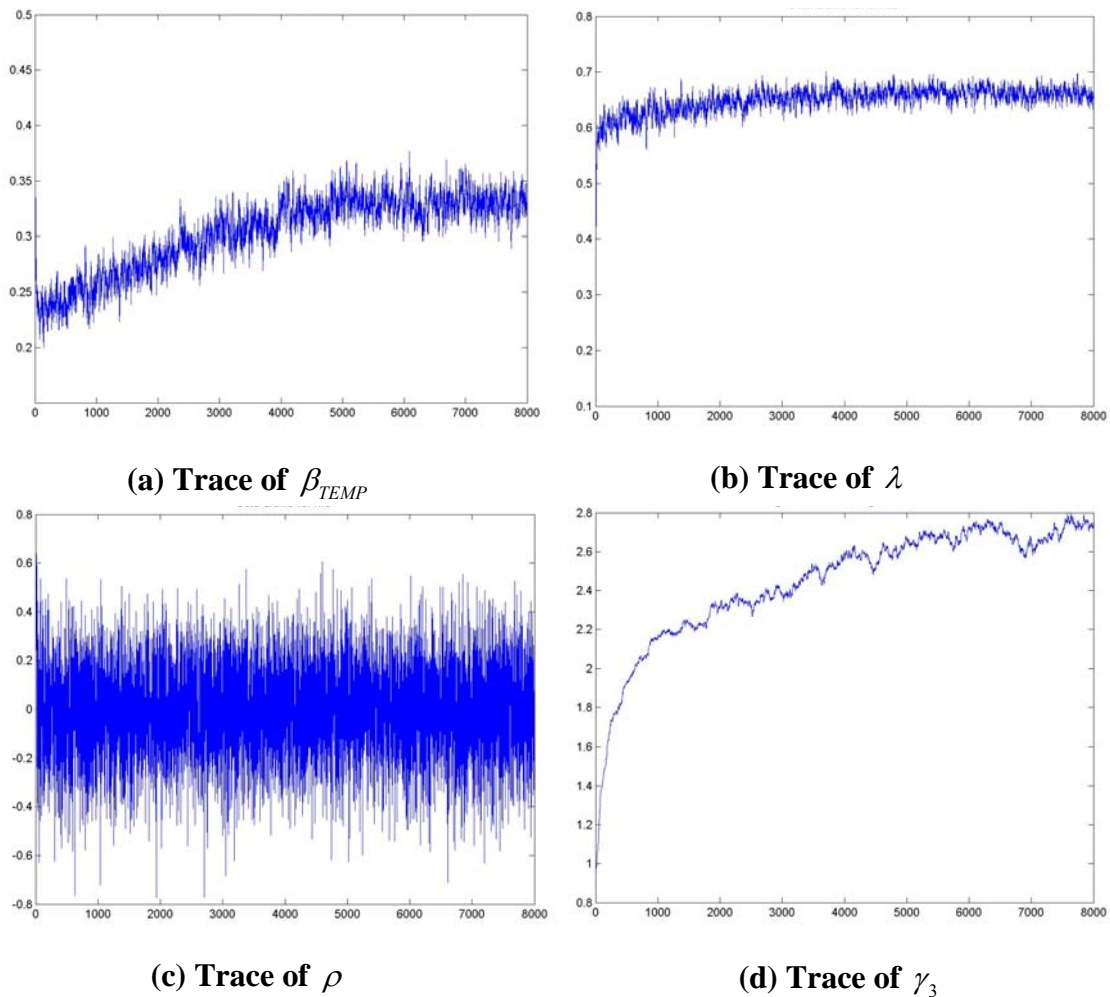


Figure 6.6 Convergence Patterns of Ozone Concentration Level Parameter Estimation

Table 6.7 shows parameter estimates based on the final 4000 runs. The estimation suggests that temperature has a statistically significant (and positive) effect on ozone concentration levels, as expected.

Interestingly, during peak travel hours, the total length of streets in the area (in this case equal to road density, since area is constant across grid cells) has no statistically significant effect. However, during non-peak hours, higher road density is associated with higher levels. Though this finding is consistent with Loibl et al. (1994)'s conclusion that time of day is influential, this phenomenon is somewhat counterintuitive, and may be explained by a delay in the photochemical process for ozone generation and deposition: the process may require several hours to develop.

Two other factors to consider are the fraction of developed land and vegetation. The fraction of developed land has the same effect during work and non-work hours. Vegetated land also has nearly the same effect day and night. These results indicate that, while land cover has a significant role, its effect is not instantaneous, possibly due to the time needed for the photochemical process. More details on these effects are provided in Section 6.3.2.

The estimation also shows that λ has a fairly high value and is statistically significant, indicating that the latent dependant variable of the previous period plays an important role. The ρ value is close to zero and slightly negative (on average) and is statistically insignificant, implying that the control variables adequately explain and spatial clustering in ozone concentration levels. The insignificance of this ρ value seems to contradict conclusions of studies by Lin (2007) and Hancock (1994), as previously discussed.

Table 6.7 Estimation Results for Model of Ozone Concentration Levels

Variable	Mean	Std. Dev.	t-stat.
TEMP	3.164E-01	1.410E-02	22.44
PEAKTRAF	1.300E-03	1.900E-03	0.68
NONPTRAF	4.900E-03	1.900E-03	2.58
WKDEV	-7.390E-02	5.200E-03	-14.21
NWKDEV	-7.360E-02	4.800E-03	-15.33
DTVEG	-6.020E-02	3.100E-03	-19.42
NTVEG	-5.910E-02	2.700E-03	-21.89
λ	6.583E-01	1.230E-02	53.52
ρ	-2.700E-03	1.874E-01	-0.01
σ^2	9.550E-02	2.480E-02	3.85
γ_1	-1.219E+00	8.710E-02	-13.99
γ_2	9.792E-01	5.800E-02	16.88
γ_3	2.462E+00	7.690E-02	32.02
γ_4	4.770E+00	1.040E-01	45.86

As in Section 6.2.1, the values of estimated regional-specific errors (θ_i), and their statistical significance (t-statistic greater than 1.64) are shown, in Figure 6.7. As indicated by the low ρ value returned by the model, the θ values seem to be randomly distributed across space (This result is supported by a Moran's I of -0.05, with a Z score of just 0.4).

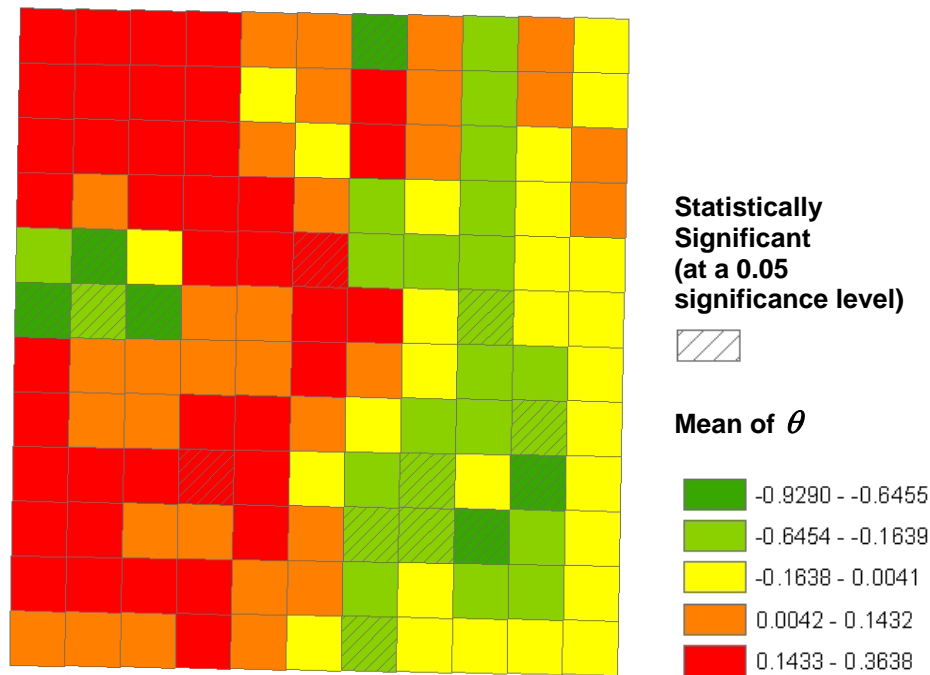
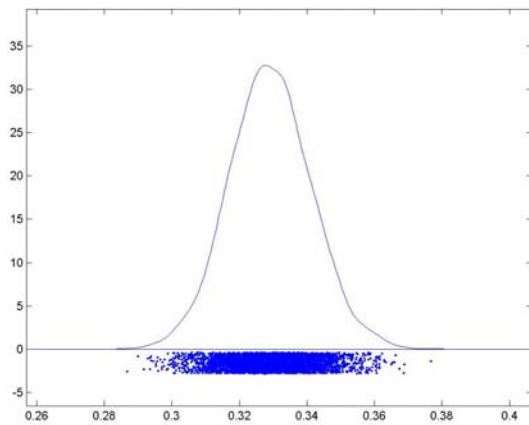
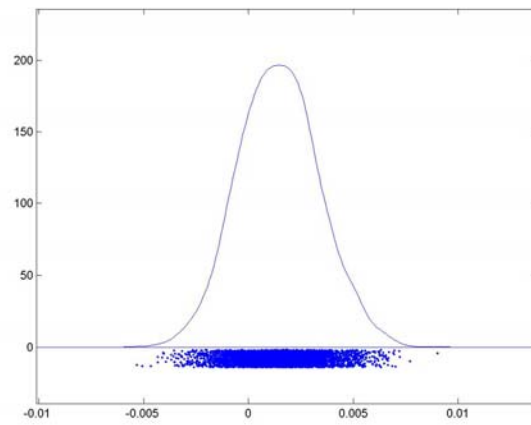


Figure 6.7 Distribution of Regional-Specific Error Term Estimates (θ) for Ozone Concentration Levels

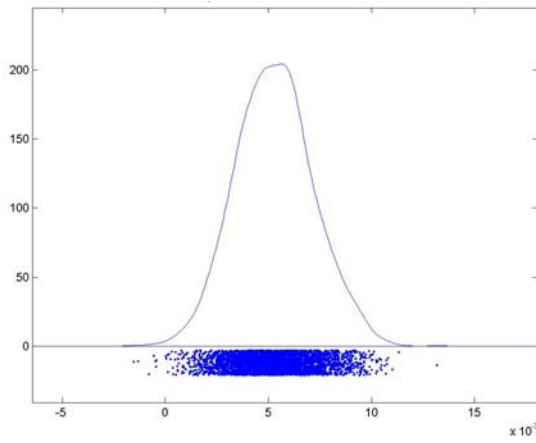
Posterior distributions of all parameters are shown in Figure 6.8. While the distributions for threshold parameters are multimodal as in the case of the development intensity results, the overall intervals are fairly narrow, offering statistically significant estimates.



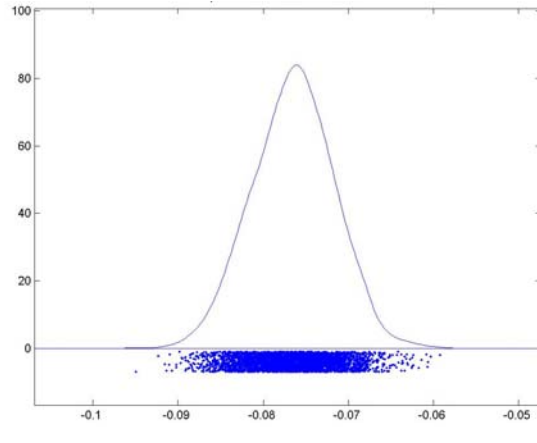
(a) Posterior Distribution of β_{TEMP}



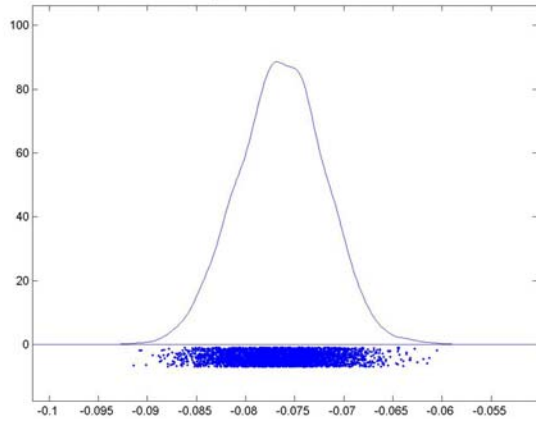
(b) Posterior Distribution of $\beta_{PEAKTRAF}$



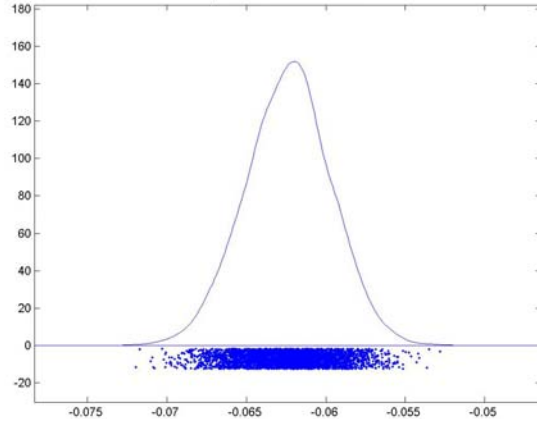
(c) Posterior Distribution of $\beta_{NONPTRAF}$



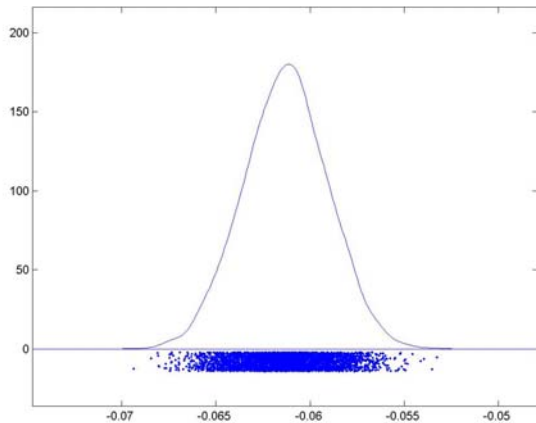
(d) Posterior Distribution of β_{WKDEV}



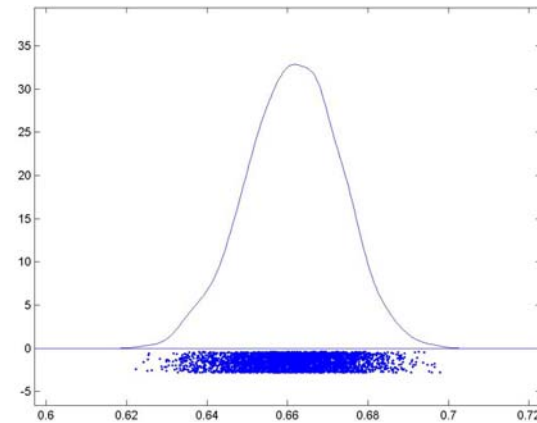
(e) Posterior Distribution of β_{NWKDEV}



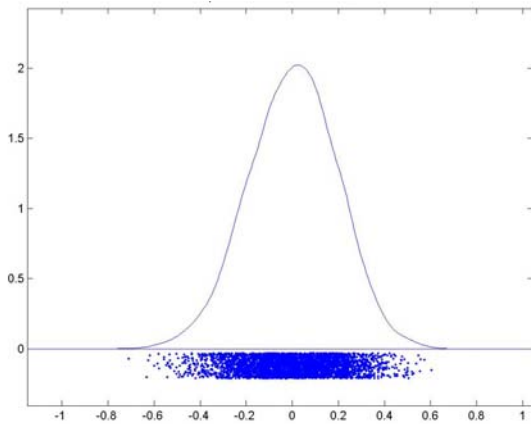
(f) Posterior Distribution of β_{DTVEG}



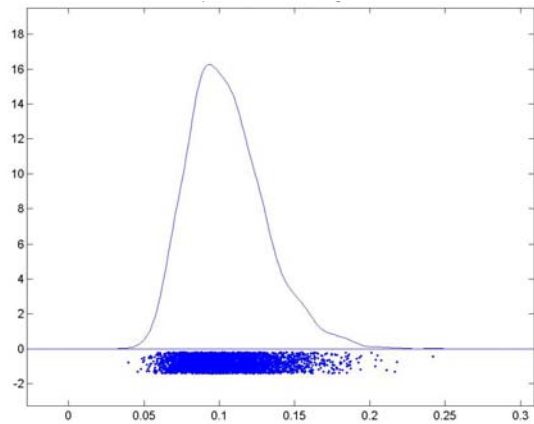
(g) Posterior Distribution of β_{NTVEG}



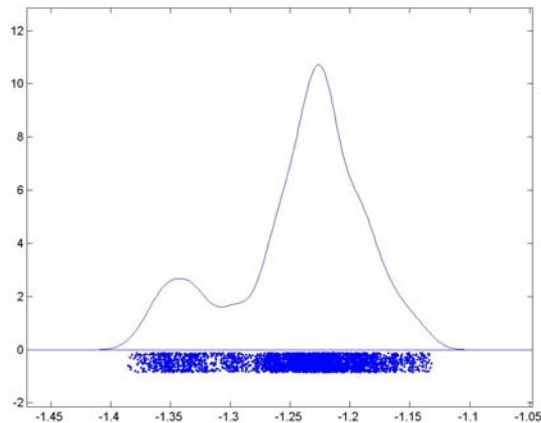
(h) Posterior Distribution of λ



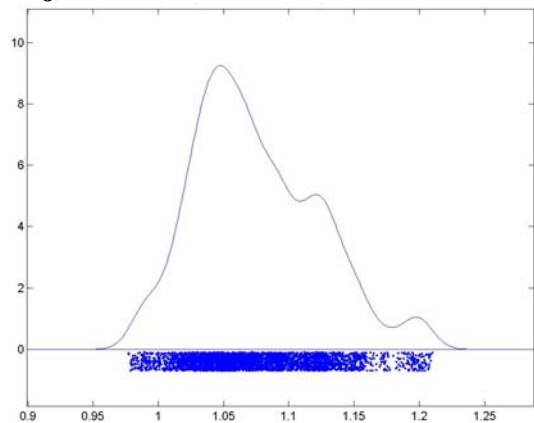
(i) Posterior Distribution of ρ



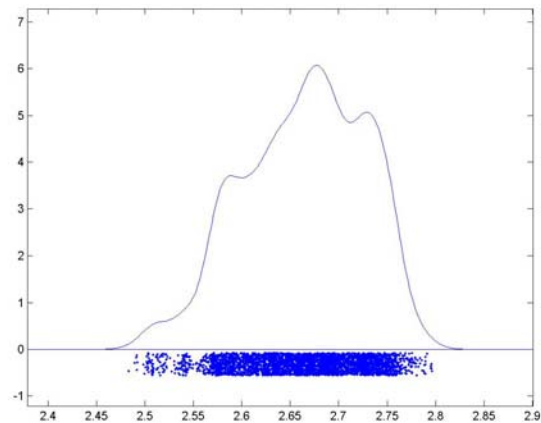
(j) Posterior Distribution of σ^2



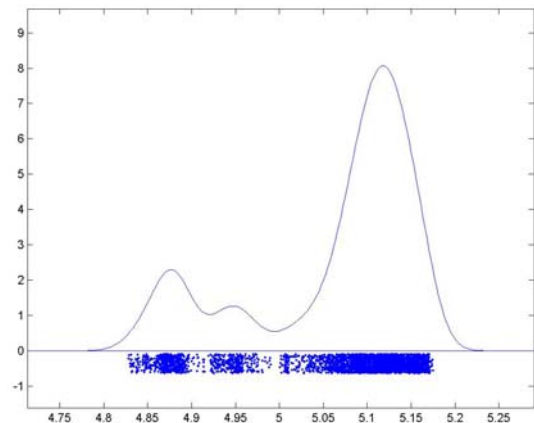
(k) Posterior Distribution of γ_1



(l) Posterior Distribution of γ_2



(m) Posterior Distribution of γ_3



(n) Posterior Distribution of γ_4

Figure 6.8 Posterior Distributions of Ozone Concentration Level Model Parameters

6.3.2 Marginal Effects

As with the results of the development intensity analysis, marginal effects were calculated for each observation in each time period. The average values of these results are summarized in Table 6.8.

One interesting result is the switch in signs of effects across ozone classes, presenting a “jumpy” pattern: Levels 1 and 4 share a consistent direction of change that opposes all others. This example highlights the fact that marginal effects for intermediate levels cannot be inferred directly from parameter signs when multiple observational units are involved.

Table 6.8 Marginal Effects of Covariates on Ozone Concentration Levels over All Observations

Variable	Ratio to Std. Dev.	Marginal Effect (10^{-2}) (Change in Response Probability)				
		Level 1	Level 2	Level 3	Level 4	Level 5
TEMP	0.026	-3.806	1.175	0.562	-1.653	3.723
PEAKTRAF	0.022	-0.016	0.005	0.002	-0.007	0.016
NONPTRAF	0.044	-0.061	0.019	0.009	-0.026	0.059
WDEV	0.041	0.886	-0.274	-0.131	0.385	-0.867
NWDEV	0.028	0.885	-0.273	-0.131	0.384	-0.865
DLVEG	0.029	0.722	-0.223	-0.107	0.314	-0.707
NTVEG	0.268	0.710	-0.219	-0.105	0.308	-0.694

Notes: “Ratio to Std. Dev.” is the ratio of one unit (e.g., 1 °C) to the standard deviation observed in the data set, for each variable.

Table 6.8 indicates how temperature may be expected to influence ozone concentration levels. By increasing temperatures one degree centigrade, the probabilities of Levels 1 and 4 are expected to fall by 3.8% and 1.7%, respectively; and probabilities of Levels 2, 3 and 4 are estimated to increase by 1.2%, 0.6%, and 3.7%, respectively. Considering that temperature can change by more than 10 degrees in a day, its effect is quite impressive.

The effect of street length, during both peak and non-peak hours, is negligible: even when lengths are increased by 20 km (roughly the current average), the corresponding change in different levels' probabilities is less than 1.3%. This result suggests that traffic local intensities (as approximated using road density) may not influence ozone concentrations in Austin. This conclusion is somewhat counterintuitive, and different from most existing studies (e.g., Niemeier et al., 2006, Wang et al., 2005, and Friedman et al., 2001). It could be due to winds shifting, emissions, and downstream ozone levels – as well as the importance of non-road emissions sources. Alternatively, it could be that without considering capacity or number of lanes, the street length variable cannot very well approximate traffic intensities. (For example, the presence of many short and narrow roads may suggest high total length but proxy for residential neighborhoods and relatively low traffic volumes.)

The fraction of developed land has a significant (negative) influence. If this fraction increases by just 1%, the two extreme probabilities are estimated to change by around 0.9% (when computed using a sample average). The fraction of vegetated land has a similar effect: an increase of 1% suggests a 0.7% increase in Level 1 concentrations and a 0.7% decrease in Level 5 concentrations. While developed and vegetated lands may be expected to contribute more to ozone generation, they also may assist ozone deposition. Thus, their net effect, when compared to barren land and water, may be to decrease ozone concentration levels. Though more insightful reasons for explaining the effects of land cover cannot be given here due to insufficient understanding of the photochemical process, the statistical relationships provided by the model estimation are helpful enough for planners to associate land cover with air quality.

6.3.3 Ozone Model Prediction

In this prediction scenario, cell/region temperatures are set to those at 0:00 to 1:00 on September 13. The fraction of developed land in each grid cell is assumed to be 1.2 times that of its current value, and vegetated land is 0.8 times that of its current value. The “previous period” is 23:00 to 24:00 on September 13, so we are predicting just 1 hour forward in time.

As with development intensity, the predictions are carried out during the model estimation. The only difference is that now the total number of alternatives is 5, instead of 4, so calculation of the uncertainty index needs to be modified accordingly.

Graphs (a) and (b) in Figure 6.9 show the most likely predicted ozone concentration levels and each cell’s uncertainty index. Ozone concentration levels generated by CAMx for 23:00-24:00 on September 13 and 0:00-1:00 on September 14 are also shown, as graphs (c) and (d), for comparison.

The prediction suggests almost no effect of land cover, which seems to contradict conclusions from several studies discussed in Chapter 1 (e.g., Wiedinmyer, 1999, Allen, 2002, and McDonald-Buller et al., 2001). Part of the reason for this finding is that the effects of developed land and land with vegetation are very similar. Therefore, increases in one offset reductions in the other. Since the land cover change has a negligible effect, and temporal dependencies (of the prior latent dependant variables) is strong, one expects a pattern lying somewhere between patterns shown in Figure 6.9’s graphs (c) and (d) which is clearly the case here. To some extent, this comparison validates the model.

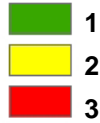
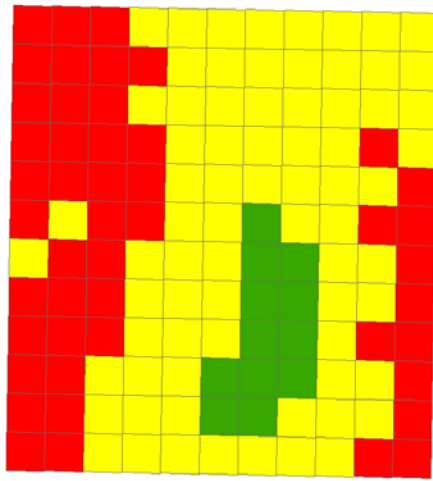
Graph (b) shows that higher uncertainties are associated with higher levels of ozone, but even the highest uncertainty is only around 0.75. A closer look at the data shows that the uncertainty is mainly caused by confusion or ambiguity

between Levels 2 and 3. Given the expectation that the predicted pattern should lie between cases (c) and (d), which are dominated by Levels 2 and 3, such confusion is quite understandable.

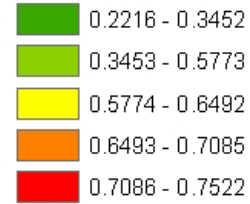
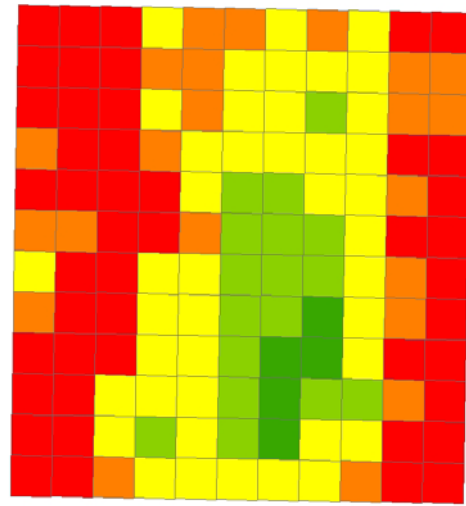
6.4 CHAPTER SUMMARY

This chapter presents the results of models for levels of Austin's land development and ozone concentrations, using the DSOP model in a Bayesian framework. Estimates of the parameters' posterior distributions are consistent with the formulations listed in Table 3.1. Parameter estimates, their marginal effects and model predictions disclose some interesting findings, which may help researchers and planners better understand Austin's land development and air quality dynamics.

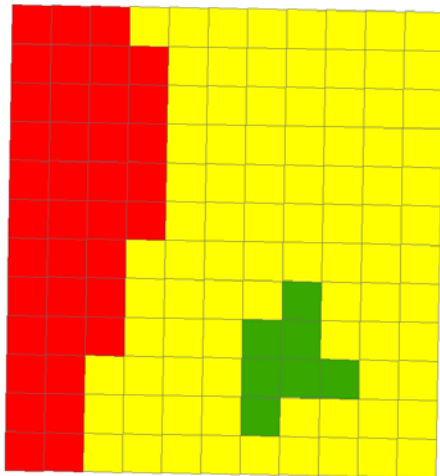
The next chapter summarizes findings from this chapter, while offering a brief review of the DSOP model's specification and estimation, and discussing limitations and potential extensions of this dissertation.



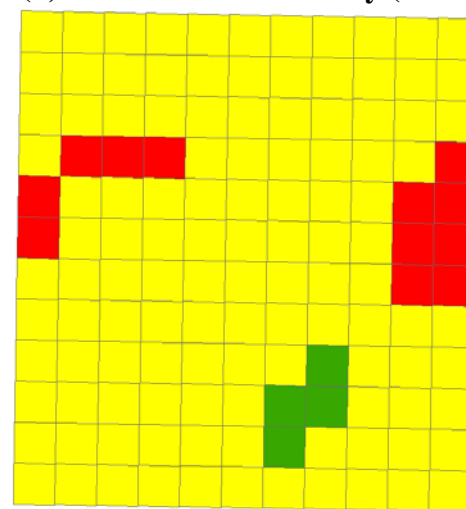
(a) Most Likely Ozone Concentration Levels



(b) Prediction Uncertainty (Entropy)



(c) Ozone Concentration Levels on September, 13, 1999 (11pm to midnight)



(d) Ozone Concentration Levels on September, 14, 1999 (midnight to 1am)

Figure 6.9 Prediction and Comparison of Ozone Concentration Levels

(Note: No Levels 4 and 5 at these points in time.)

CHAPTER 7. CONCLUSIONS

7.1 SUMMARY

Many transportation and land use studies involve latent (unobserved) variables exhibiting underlying spatial interactions and temporal dependency patterns. Examples include destination and location choices, crash counts (over a network), and pavement deterioration levels. These examples all present two common features. First, the variables of interest are indicators or censored versions of unobserved variables. Second, they all exhibit certain degrees of temporal and spatial autocorrelation. Such phenomena also exist in other fields, like ecology, biology and anthropology. To capture these temporal and spatial patterns and accurately estimate the impacts of potentially influential factors, a rigorous statistical method for analyzing such data is needed. The dynamic spatial ordered probit (DSOP) model established in this dissertation meets this need.

The DSOP model analyzes ordered response data based on latent variables exhibiting and spatial dependencies as well as individual heterogeneity. This dissertation makes three major contributions to the methodological development of spatial econometrics. First, the model incorporates spatial effects by allowing for both regional spatial interactions and heteroskedasticity across observations from different regions. Second, the model allows for an AR(1) process via the latent, lagged dependent variable, thus recognizing dynamic features. Third, when compared to existing spatial discrete choice models, the DSOP model is the first to emerge from an ordered probit model, where multiple levels of ranked categorical data can be analyzed.

The models developed here were estimated in a Bayesian framework using MCMC sampling and data augmentation techniques (to generate the autocorrelated latent variables). The estimation process approximates the

parameter set's joint probability using a set of conditional distributions. To achieve this, proper prior distributions for parameters and nuisance terms (latent dependent variables and variances) were assumed and their posterior distributions then derived. These posterior distributions include common distributions (like the truncated normal and chi square), mixture distributions (combining a normal and multivariate uniform), and nonstandard distributions (offering no closed-form expressions for hyperparameters). Matlab code was developed to draw from these distributions.

As noted in Chapter 2, one advantage of a Bayesian approach to model estimation is that the program code can be conveniently adapted to other model specifications, if they include some conditional distributions in the same form. For example, the drawing method derived and programmed here can be rather easily modified for other latent variable models, including count data and continuous responses with temporal dependencies and spatial autocorrelation.

The DSOP model specification and estimation methods were first validated (in Chapter 4) using 70 simulated datasets, for each of the 12 parameter sets. The results produced estimates that are quite close to true values. The comparison with a standard ordered probit model highlighted the accuracy of the DSOP model, while recognizing report temporal and spatial autocorrelation patterns.

The DSOP model then was applied to two Austin, Texas datasets, one for urban land development and the other for air quality. The first dataset involves development intensity levels derived from satellite images with controls for socioeconomic and topographic information (derived from several other sources). The data integration process itself is another contribution of this dissertation, as it suggests how to align remotely sensed data with various traditional data sources across both temporal and spatial coordinates. The second dataset involves ozone concentration levels, with controls for transportation access and infrastructure,

land cover conditions, and temperature. This long panel dataset ($T=24$) illustrates how the DSOP model can be used with datasets in which spatial interaction occurs across all individuals ($n_i = 1, \forall$ region i).

Analysis of these empirical databases discloses several important findings. In the case of land development intensity levels, the temporal autocorrelation coefficient is highly practically and statistically significant. This implies that when developers make decisions, the existing land conditions (represented by temporally lagged latent dependent values), is the determining consideration. Other control variables have much smaller marginal effects, suggesting that an AR(1)-type approach with spatial lags can be key to land development prediction. Estimates of transportation conditions, especially travel time to the CBD area, do seem to have a significant impact on land development choices, highlighting the important role of access. Even after controlling for neighborhood characteristics and lagged latent response levels, estimation residuals are high in this model, and positively correlated across space. This statistical result confirms the common intuition that land development tends to cluster rather than randomly distributed in space. In addition, as discussed in Chapter 2, the recognition of this effect avoids problems such as biased estimates and overestimated precision due to omission of spatial effects.

As expected, the model's application to the ozone concentration data revealed a highly continuous process: levels during one period are mainly determined by the values from the previous period, with mean temporal lag coefficient as high as 0.66. Temperature is another very influential factor while transportation and land cover were not very helpful. In addition, their mild effects are not instantaneous. The coefficient on the spatial error matrix was close to zero in this case, further implying that the temporal lagged utility and temperature adequately explain the changing ozone levels predicted by the CAMx model.

7.2 LIMITATIONS AND EXTENSIONS

This dissertation research sought to thoughtfully incorporate temporal and spatial relationships in models of ordered categorical response data. Due to limitations on time and data, more general model specifications and several potential applications were not realized. Several extensions of this work exist, as described below.

The most radical modification and meaningful extension of this work is a specification that accommodates unordered categorical response data. This extension may involve random utility maximization theory. Unlike the case of ordered data, such analysis typically requires one equation for each alternative. In other words, for each observation in each period, there will be a set of equations. This may appear similar to a seemingly unrelated regression (SUR) model of latent variables. It is expected that a dynamic spatial model for unordered categorical data will be much more complicated than the DSOP model, but its application will be broader. Fruhwirth-Schnatter and Wagner (2006) and Scott (2004) provide algorithms for calculating the multinomial logit model (MNL) with data augmentation techniques in a Bayesian framework. Wang and Kockelman (2006a) proposed a 3-step method that is methodologically similar to MCMC sampling²¹, for estimating SUR models with spatial and temporal dependencies. These studies may be incorporated within the existing DSOP model code, providing a way to analyze unordered categorical response data with temporal and spatial dependencies.

Another area for enhancement recognizes that the specifications calibrated here relied on contiguity matrices, rather than more complex measures of inter-

²¹ Wang and Kockelman (2006a) specified conditional maximum likelihood estimators (MLE) for different sets of parameters. These conditional distributions are iterated until convergence. From this perspective, the estimation procedure can be viewed as a mixture of traditional MLE and MCMC sampling, though their method maintains a frequentist perspective.

observational distances. As discussed in Chapter 2, following row-standardization a contiguity matrix is not inferior to a distance-decay matrix. Contiguity matrices still permit the autocorrelation to permeate the database (as long as all regions eventually touch via a series of neighbors), while allowing analysts the speed and convenience of sparse matrix algorithms. However, for cases where the weight matrix is not row-standardized (for example, when non-standardized values ensure stationarity of the spatial process, or the original economic/behavioral meaning of inter-element relationships needs to be kept), distance-decay matrices will clearly differ. Future study might well address the differences between these methods, including differences in their computational efficiencies, influence on the model goodness of fit, and results interpretation.

Another issue relates to variable gaps in land use data sets. The four time data years are 1983, 1991, 1997, and 2000, with gaps of 8, 6, and 3 years, respectively. Intuitively, when the gap is longer, the temporal dependencies should be weaker. A more appropriate model specification would control for gaps variations in some way. This research ignores the difference in time gaps and uses a single temporal coefficient (λ) across all periods. One possible extension of this study is to capitalize on time series data analysis methods for variable gap lengths, to try and avoid such issues. For example, one approach may be to express the temporal coefficient as an exponential function of the time gap.

Another extension relates to long-term forecasting. The applications described in Chapter 6 predict one period forward. It will be useful to explore multi-period forecasts and their associated uncertainty. It may be useful to treat a large time gap as one long period, with adjustments to the temporal coefficient, λ . Alternatively, one can do one-period-forward forecasting repeatedly, until the desired year's projection is achieved. Such forecasting can be carried out multiple times, and the associated uncertainty (across all potential, discrete responses) can

be evaluated using entropy statistics. Another important application relates to extending the data set's cross-sectional scope: how can latent utilities for out-of-sample observations be derived and what the associated uncertainty will be.

In spatial analysis, a common issue is the modifiable areal unit problem (MAUP). This dissertation determines individual and region units based on existing zone structures and computational considerations. This process, though flexible and adaptable to user needs, is somewhat arbitrary and lacks theoretical substantiation. In fact, estimation with more aggregate observational units tends to indicate lower spatial autocorrelation. Future research may examine how the choice of spatial unit (e.g., 30 m cells vs. 300 m cells) influences estimation results. This may prove particularly useful in applications of satellite information, where data quality and results are sensitive to resolution levels. Such research may be able to suggest optimal observational units, so that spatial information can be maximally explored and estimation results best explain reality.

As mentioned in Chapters 3 through 6, diagnosing convergence in a Bayesian setting is a complicated issue. This dissertation uses traces of parameter estimates to ascertain convergence. Ideally, existing convergence criteria (e.g., Gelman and Rubin 1992, and Cowles and Carlin, 1996) should be compared and the most suitable one used to better ensure appropriate burn-in period.

The DSOP model specification includes a temporal lag of the latent dependent variable and a spatial lag of the regional-specific error terms. A more general model specification could include temporal and spatial lags of both dependent variables and error terms. Such an extension would prove interesting and useful, though much more complicated. Potential difficulties include finding a proper formulation to combine spatial and temporal autocorrelations, ensuring parameter identifiability, and interpreting results.

The prediction scenario tested here are highly hypothetical. Alternatively, a prediction may be carried out for more realistic scenarios (For example, in the development intensity model, using year 2005's actual explanatory variable values would facilitate "prediction" of year 2005's intensity levels). The 2005 "prediction" can be compared to the observed situation, serving as another method of model validation.

The ozone dataset includes a total of 132 grid cells with 4km space resolution, and the estimation suggests no spatial autocorrelation. However, this $n=132$ sample size may be too small to discern clustering and other patterns of spatial autocorrelation. More importantly, both dependent and explanatory variables in the ozone dataset are derived from CAMx model predictions, rather than actual ozone measurements across Austin. In the future, predictions for a larger-scale area should be tried, and if possible, observed ozone concentration data should be collected and used, though such sites are generally few²². The results from using real data should be compared to those from the CAMx model to further validate this statistical method.

Data availability problems are mentioned several times in this dissertation. Of course, the main intent of this study is to offer new methods to assist analysis and policy making. To better assist in the use of the models' empirical results, more policy-related factors (such as land development constraints, congestion tolls, and vehicle inspection/maintenance plans) should be acquired and incorporated into the model. Moreover, as indicated in Chapter 6, there appears to be too-high correlation between population and number of workers variables in the model of development intensity levels. To some extent, this problem compromises the

²² In practice, sampling sites are very scarce. For example, the study area has only two such sites on September 13, 1999 (and seven in June 2007) (TCEQ, 2007). For San Francisco Bay Area basin, the number is 23 in 1999 and 22 in 2007 (CARB, 2007).

validity of the estimation results, at least on these variables' associated slope coefficients. Certainly, one of these two variables should probably be left out of the final specification.

Data quality, especially when based on remotely sensed, reconfigured, and classified data is used, is another key issue to be explored. This study uses land cover information originally at 30 m resolution, then aggregated and re-interpreted at 300 m resolution. Errors in the original data, as well as errors incurred from the aggregation and categorization process, should all be assessed. Implications of this process also should be discussed and justified.

Another extension of this study is to improve the Matlab code. As discussed in Chapter 1, spatial data is often large-scale and memory intensive (Nelson and Geoghegan, 2001). The computation time with the current code is acceptable, but not ideal: for the development intensity analysis ($N=2771$, $M=57$, and $T=8$), an Intel 1.66GHz CPU required about 3.5 hours; for the ozone concentration level analysis ($N=M=132$ and $T=24$), the same computer required about 2 hours. Better data structure and functions can be utilized to enhance the speed. User-friendly interfaces also can be added.

7.3 CONCLUDING REMARKS

Response variables of interest often appear in the form of ordered categorical data, and these data may exhibit temporal and spatial dependencies. Many such examples can be found in transportation-related studies. This dissertation's dynamic spatial ordered probit (DSOP) model captures patterns of spatial and temporal autocorrelations for ordered response data. The model is estimated using MCMC sampling in a Bayesian framework. This brand new DSOP model appears to successfully capture temporal and spatial patterns in distinct datasets while

quantifying effects of various explanatory variables. It offers a valuable framework for rigorously analyzing relationships in such complicated situations. This study also renders some general insights into the pragmatic advantages of a Bayesian framework over a frequentist method²³. For this type of work, the Bayesian approach appears more straightforward and much easier to apply than maximum (simulated) likelihood estimation (MSLE). Especially for models involving complicated statistical distributions and multi-layered specifications (as with the DSOP model), the advantage of a Bayesian framework is evident. By using “conditional” distributions, the Bayesian approach decomposes the joint estimation of many variables into much simpler, sequential simulations. In contrast, maximum (simulated) likelihood estimation (MSLE) must tackle an intractable likelihood function (and its gradients and possibly its Hessian matrix, with respect to the parameter set) (see, e.g., Wang and Kockelman [2006c]). With a Bayesian framework, a slight change in model specification only requires modifying a part of the simulation procedure. With MSLE, on the other hand, the model estimation method may need to be completely overhauled. However, the Bayesian approach also has its limitations. For example, in this study, because the estimation involves simulating latent variables and one (multivariate) posterior distribution (for threshold terms) is multimodal, marginal effects and the model’s goodness of fit need to be calculated simultaneously with the simulation. Otherwise, if an indicator (such as the deviance information criterion) needs to be obtained afterwards, the model must be completely rerun, which can be rather time consuming.

²³ Of course, much has been written (e.g., Geweke, 1993; Gelman et al., 2004; and Koop et al. 2007) about the differences in classical and Bayesian statistical viewpoints. Much of the discussion is somewhat “philosophical” in nature, and “superiority” has never been conclusively determined (Gelman et al., 2004).

Finally, this dissertation demonstrates how to use satellite data meaningfully, in land development and air quality analysis. This provides inspiration for tapping the potential of satellite databases. As more frequent and accurate satellite images become available, this evolving data source will be used for far more extensive topics, such as global climate changes, loss of Amazon rainforest, Africa's desertification, human migration, and even real-time traffic condition forecasting. It is important that transportation researchers and others begin to unleash the potential of these data sets, by recognizing the spatial relationships that exist and by exploiting their presence.

REFERENCES

- Albert, J. H. and Chib, S. (1993) "Bayesian analysis binary and polychotomous response data." *Journal of the American Statistical Association* 88: 669-679.
- Allen, D.T., Durrenberger, C. and TNRCC Technical Analysis Division (2002) "Accelerated science evaluation of ozone formation in the Houston Galveston area: Photochemical air quality modeling." Technical paper. Accessed May 10, 2007:
http://www.utexas.edu/research/ceer/texaqsarchive/pdfs/Modeling02_17_02.PDF.
- Alvarez, J. and Arellano, M. (2003) "The time series and cross-section asymptotics of dynamic panel data estimators." *Econometrica* 71(4): 1121-1159.
- Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Press.
- Anselin, L. (1999) *Spatial Econometrics*. Working paper. Accessed July 10, 2005:
http://www.csiss.org/learning_resources/content/papers/baltchap.pdf.
- Anselin, L. (2001) "Issues in spatial probit models." Workshop on Qualitative Dependent Variable Estimation and Spatial Effects, College of ACES, University of Illinois, April 20, 2001.
- Anselin, L. (2003) "Spatial externalities, spatial multipliers and spatial econometrics." *International Regional Science Review* 26(2): 153-166.
- Anselin, L. and Bera, A.(1998). "Spatial dependence in linear regression models with an introduction to spatial econometrics." In Ullah, A. and Giles, D. E.A. (Eds.), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.
- Anselin, L. and Hudak, S. (1992) "Spatial econometrics in practice: A review of software options." *Regional Science and Urban Economics* 22(3): 509-536.
- Anselin, L., Florax, R., and Rey, S. (2004) (Eds.), *Advances in Spatial Econometrics. Methodology, Tools and Applications*. Berlin: Springer-Verlag.
- Anselin, L., Syabri, I. and Kho, Y. (2006) "GeoDa: an introduction to spatial data analysis." *Geographical Analysis* 38(1): 5-22.

- Arellano, M. and Hahn, J. (2005) "Understanding bias in nonlinear panel models: Some recent developments." Invited lecture. Econometric Society World Congress, London.
- Aster, R., Borchers, B., and Thurber, C. (2004) *Parameter Estimation and Inverse Problems*. Burlington, MA: Academic Press.
- Athanasiadis, I. N., Karatzas, K. D. and Mitkas, P. A. (2007) "Classification techniques for air quality forecasting." Working paper. Accessed May 10, 2007:
[http://issel.ee.auth.gr/ktree/Documents/Root%20Folder/ISSEL/Publication s/AKM_besai_KARATZAS_revised.doc](http://issel.ee.auth.gr/ktree/Documents/Root%20Folder/ISSEL/Publication%20s/AKM_besai_KARATZAS_revised.doc).
- Atkinson, P. M., Clark, M. J., German, S. E. and Sear, D. A. (2003) "Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression." *Geographical Analysis* 35(1): 58-82.
- Barlevy, G. and Nagaraja, H. N. (2006) "Identification of search models with initial condition problems." Working paper. Accessed May 10, 2007:
<ftp://repec.iza.org/RePEc/Discussionpaper/dp2061.pdf>.
- Basu, S. and Thibodeau, T. G. (1998) "Analysis of spatial autocorrelation in housing prices." *Journal of Real Estate Finance and Economics* 17(1): 61-85.
- Bay Area Air Quality Management District (BAAQMD) (2005) "Bay Area 2005 ozone strategy and draft environmental impact report summary." Technical report. Accessed May 10, 2007:
http://www.baaqmd.gov/pln/plans/ozone/2003_workgroup/os_deirsummary.pdf.
- Beron, K. and Vijverberg, W. (2004) "Probit in a spatial context: a Monte Carlo analysis". In Anselin, L. Florax, R. and Rey, S. (Eds.), *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.
- Besner C. (2002) "A spatial autoregressive specification with a comparable sales weighting scheme". *Journal of Real Estate Research* 24: 193-211.
- Bester, C. A. and Hansen, C. (2007) "A penalty function approach to bias reduction in non-linear panel models with fixed effects." Accessed June 1, 2007:
http://faculty.chicagogsb.edu/christian.hansen/research/bh_penalizedfe_jul06.pdf

- Bhat, C. and Guo, J. (2004) "A mixed spatially correlated logit model: Formulation and application to residential choice modeling." *Transportation Research Part B* 38: 147-168.
- Boots, B.N. and Kanaroglou, P.S. (1988) "Incorporating the effects of spatial structure in discrete choice models of migration." *Journal of Regional Science* 28(4): 495-509.
- Briggs, R. (2006) "Spatial statistics." Lecture notes. Accessed May 10, 2007: <http://www.utdallas.edu/~briggs/poec6382/spatstat.ppt>.
- Caliper Corporation. (2004) *TransCAD GIS Software, version 4.7*. Caliper Corporation, Newton, Massachusetts.
- Campbell, J.B. (2002). *Introduction to Remote Sensing, 3rd Edition*. New York: The Guilford Press.
- Candau, J., Rasmussen, S. and Clarke, K. C. (2000) "A coupled cellular automaton model for land use/land cover dynamics". Accessed July 10, 2005: <http://www.geog.ucsb.edu/~kclarke/ucime/banff2000/533-jc-paper.htm>.
- CAPCO (Capital Area Planning Council), The University of Texas at Austin and ENVIRON International Corporation (2004) *Development of the September 13-20, 1999 Base Case Photochemical Model for Austin's Early Action Compact*. Technical report. Accessed May 10, 2007: [http://www.capco.state.tx.us/CAPCOairquality/NOV_30/eac_basecase_milestone_complete\(revised2\).pdf](http://www.capco.state.tx.us/CAPCOairquality/NOV_30/eac_basecase_milestone_complete(revised2).pdf).
- CAPCOG (Capital Area Council of Governments) (2006) *Information Clearinghouse-Geospatial Data*. Capital Area Council of Governments, Austin, Texas. Accessed May 10, 2007: http://www.capcog.org/Information_Clearinghouse/geospatial_main.asp.
- CARB (California Air Resource Board) (2007) *Latest Ozone Summary for Selected Regions (Preliminary Data)*. Accessed June 9, 2007: http://www.arb.ca.gov/aqmis2/ytd_ozone.php.
- Celeux, G., Forbes, F., Robert, C. P. and Titterton, D. M. (2006) "Deviance information criteria for missing data models." Technical report. Accessed May 10, 2007: <http://www.ceremade.dauphine.fr/~xian/cfirt05.pdf>.
- Chib, S. (1992) "Bayes regression for the tobit censored regression model." *Journal of Econometrics* 51(1-2): 79-99.
- City of Austin (2007). *City of Austin GIS Data Sets*. Accessed May 10, 2007: ftp://coageoid01.ci.austin.tx.us/GIS-Data/Regional/coa_gis.html.

- City of Austin (2006) *Employment Data*. Accessed May 10, 2007:
<http://www.ci.austin.tx.us/growth/employment.htm>.
- Clarke, K.C., Hoppen and S., Gaydos, L. (1996). "Methods and techniques for rigorous calibration of a cellular automaton model of urban growth". Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-25, 1996.
- Clarke, K. C. (1997). "Land transition modeling with deltatrons". Paper presented at the NCGIA Land Use Modeling Workshop. Accessed July 10, 2005:
<http://www.ncgia.ucsb.edu/conf/landuse97>.
- Cliff, A.D. and Ord., J. K. (1981) *Spatial Processes: Models and Applications*. London: Pion.
- Coughlin, C. C., Garrett, T. A. and Hernández-Murillo, R. (2003) "Spatial probit and the geographic patterns of state lotteries." Accessed March 1, 2006:
<http://research.stlouisfed.org/wp/2003/2003-042.pdf>.
- Cowles, M. K. (1996) "Accelerating Monte Carlo Markov Chain convergence for cumulative-link generalized linear models." *Statistics and Computing* 6(2): 101-110.
- Cowles, M. K. and Carlin, B. P. (1996) "Markov Chain Monte Carlo convergence diagnostics: A comparative review." *Journal of the American Statistical Association* 91: 883-896.
- Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.
- de la Barra, T. (1989) *Integrated Land Use and Transport Modeling*. Cambridge: Cambridge University Press.
- Dubin, R. A. (1988) "Estimation of regression coefficients in the presence of spatially autocorrelated error terms." *Review of Economics and Statistics* 70(3): 466-474.
- Dubin, R. A. (1992) "Spatial auto correlation and neighborhood quality." *Regional Science and Urban Economics* 22(3): 433-452.
- Dugundji, E. R. and Walker, J. L. (2005) "Discrete choice with social and spatial network interdependencies." *Transportation Research Record* 1921: 70-78.
- Environ (2007) *User's Guide for Comprehensive Air Quality Model with Extensions Version 4.40*. Accessed June 1, 2007:
http://www.camx.com/files/CAMxUsersGuide_v4.40.pdf
- EPA (Environmental Protection Agency) (2006) "8-hour ground-level ozone designations." Technical report. Accessed May 10, 2007:
<http://www.epa.gov/ozonedesignations/>.

- EPA (Environmental Protection Agency) (2003) "Deferral of effective date of non-attainment designations for 8-hour ozone national ambient air quality standards for early action compact areas." Technical Report. Accessed May 10, 2007:
http://www.epa.gov/ttnnaqs/ozone/eac/n_prm_eac_sign20031211.pdf.
- ESRI (2005) *ArcView GIS software, version 3.2*. Environmental Systems Research Institute Inc., Redlands California.
- Fiebig, D. G. (1999) "Seemingly unrelated regression." In Baltagi (Eds.), *Companion in Theoretical Econometrics*. Oxford: Basil Blackwell.
- Foody, G. M. (2002) "Status of land cover classification accuracy assessment." *Remote Sensing of Environment* 80(1): 185- 201.
- Fotheringham, A.S., Brunson, C. and Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Hoboken, NJ: John Wiley and Sons.
- Frazier, C. (2004) *Spatial Econometric Models for Land Use/Land Cover Data: Theory and Application using Satellite Images for the Austin, Texas Region*. Masters Thesis, The Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.
- Frazier, C. and Kockelman, K. (2005) "Spatial econometric models for panel data: Incorporating spatial and temporal data". *Transportation Research Record* 1902: 80-90.
- Friedman, M. S., Powell, K. E., Hutwagner, L., Graham, L. M. and Teague, W. G. (2001) "Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma." *Journal of American Medical Association* 285(7): 897-905.
- Fruhirth-Schnatter, S. and Wagner, H. (2005) "Gibbs sampling for discrete-valued data." Conference on Bayesian Applications and Methods in Marketing, Linz September 20-21, 2005
- Gao, O. H. and Niemeier, D. (2005) "Temporal/spatial statistical analyses of truck transportation and ozone dynamics." Group research at the University of California, Davis Accessed July 19, 2007:
<http://dn.engr.ucdavis.edu/GroupResearch/material/slides/oliver.pdf>.
- Gelfand, A. E. and Smith, A. F. M. (1990) "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association* 85: 398-409.

- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., and Raton, B. (2004) *Bayesian Data Analysis (2nd Edition.)*. Florida: Chapman and Hall/CRC Press.
- Geweke, J. (1993) "Bayesian treatment of the independent Student-t linear model." *Journal of Applied Econometrics* 8(S): 19-40.
- Girard P. and Parent E. (2001) "Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring." *Technometrics* 43(2): 180-191.
- Greene, W. (2000) *Econometric Analysis*. Upper Saddle River: Prentice-Hall.
- Guldmann, J.M. and Kim, H.Y. (2001) "Modeling air quality in urban areas: A cell-based statistical approach." *Journal of Planning Literature* 16(1): 80-163.
- Guttorp, P. (2006) "Setting environmental standards: A statistician's perspective." *Environmental Geosciences* 13(4): 261-266.
- Hamilton, G., Currat, M., Ray, N., Heckel, G., Beaumont, M. and Excoffier, L. (2005) "Bayesian estimation of recent migration rates after a spatial expansion." *Genetics* 170(1): 409-417.
- Hunt, J.D. and Abraham, J.E. (2003) "Design and application of the PECAS land use modeling system". *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management*, Sendai, Japan.
- Jensen, R.J. (1996). *Introductory Digital Image Processing*. New Jersey: Prentice Hall.
- Jensen, J.R. (2007). *Remote Sensing of The Environment: An Earth Resource Perspective, 2nd Edition*. Prentice Hall.
- Johnson, V. E. and Albert, J. H. (1999) *Ordinal Data Modeling*. New York: Springer.
- Johnston, D. M., and Timlin, D. (2000). "Spatial data accuracy and quality assessment for environmental management." In Heuvelink, G. B. M. Lemmens, M. J. P. M. (Eds.), *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Delft: Delft University Press.
- Kakamu, K. and Wago, H. (2007) "Bayesian spatial panel probit model with an application to business cycle in Japan." Working paper. Accessed May 10, 2007:
http://www.mssanz.org.au/modsim05/proceedings/papers/kakamu_2.pdf

- Klier, T. and McMillen, D. P. (2007) "Clustering of auto supplier plants in the U.S.: GMM spatial logit for large samples." *Journal of Business and Economic Statistics* (forthcoming). Accessed May 10, 2007: <http://tigger.uic.edu/~mcmillen/papers/Clustering%20of%20Auto%20Supplier%20Plants.pdf>.
- Kockelman, K. M. (1997) "Travel behavior as a function of accessibility, land use mixing, and land balance: Evidence from the San Francisco Bay Area." *Transportation Research Record* 1607: 117-125.
- Koop, G. M., Poirier, D. J. and Tobias, J. L. (2007) *Bayesian Econometric Methods*. New York: Cambridge University Press.
- Kotz, S. and Johnson, N. (1982) *Encyclopedia of Statistical Sciences*. New York: John Wiley & Sons
- LeSage, J. P. (1999) "The theory and practice of spatial econometrics." Manuscript. Accessed May 10, 2007: <http://www.spatial-econometrics.com>.
- LeSage, J. P. (2000) "Bayesian estimation of limited dependent variable spatial autoregressive models." *Geographical Analysis* 32(1): 19-35.
- Lillesand, T.M., Kiefer, R.W., and Chipman, J.W. (2003). *Remote Sensing and Image Interpretation, 5th Edition*. Hoboken, NJ: Wiley.
- Lin, C. Y. C. (2007) "A spatial econometric approach to measuring pollution externalities: An application to ozone smog." Working paper. Accessed May 10, 2007: http://www.des.ucdavis.edu/faculty/Lin/airqual_ext_paper.pdf.
- Loibl, W., Winiwarter, W., Kopsca, A., Zueger, J. and Baumann, R. (1994) "Estimating the spatial distribution of ozone concentrations in complex terrain." *Atmospheric Environment* 28(16): 2557-2566.
- Magnussen, S. (1993) "Bias in genetic variance estimates due to spatial autocorrelation." *Theoretical and Applied Genetics* 86(2-3): 349-355.
- Malczewski, J. and Poetz, A. (2005) "Residential burglaries and neighborhood socioeconomic context in London, Ontario: Global and local regression analysis." *Professional Geographer* 57(4): 516-529.
- Mathworks (2006), MatLab 2006a. The Mathworks, Inc. Natwick, Massachusetts.
- McDonald-Buller, E.C., Wiedinmyer, C., Kimura, Y. and Allen, D. T. (2001) "Effects of land use data on dry deposition in a regional photochemical model." *Journal of Air and Waste Management Association* 51(8): 1211-1218.

- McFadden, D. L. (1994) (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, MA: MIT Press.
- McKay, M. D. (1995) "Evaluating prediction uncertainty". Technical Report NUREG/CR-6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory.
- McMillen, D. P. (1995) "Spatial effects in probit models: A Monte Carlo investigation." In Anselin, L. and Florax, R. (Eds.), *New Directions in Spatial Econometrics*. Heidelberg: Springer-Verlag.
- McMillen, D. P., and McDonald, J. F. (1998) "Suburban subcenters and employment density in metropolitan Chicago." *Journal of Urban Economics* 43(2): 157-180.
- Mennis, J. (2006) "Mapping the results of Geographically Weighted Regression." *The Cartographic Journal* 43(2): 171-179.
- Miyamoto, K., Vichiensan, V., Shimomura, N. and Paez, A. (2004). "Discrete choice model with structuralized spatial effects for location analysis" *Transportation Research Record* 1898: 183-190.
- Moran, P.A.P. (1950) "Notes on continuous stochastic phenomena." *Biometrika* 37(1-2): pp.17-23.
- Munroe, D., Southworth, J. and Tucker, C. M. (2001) "The dynamics of land-cover change in western Honduras: Spatial autocorrelation and temporal variation". *Conference Proceedings. American Agricultural Economics Association. AAEA-CAES 2001 Annual Meeting*. Accessed July 10, 2004: http://agecon.lib.umn.edu/cgi-bin/pdf_view.pl?paperid=2611
- Nelson, G. C., and Geoghegan, J. (2002) "Deforestation and land use change: Sparse data environments." *Agricultural Economics* 27(3): 201-216.
- Nelson, G. C., and Hellerstein, D. (1997). "Do roads cause deforestation: Using satellite images in econometric analysis of land use". *American Journal of Agricultural Economics* 79: 80-88.
- Neyman, J. and Scott, E. L. (1948) "Consistent estimates based on partially consistent observations." *Econometrica* 16(1): 1-32.
- Niemeier, U., Granier, C., Kornblueh, L., Walters, S. and Brasseur, G. P. (2006) "Global impact of road traffic on atmospheric chemical composition and on ozone climate forcing." *Journal of Geophysical Research* 111(9): D09301.01- D09301.13.
- Olmo, J. C. (1995) "Spatial estimation of housing prices and locational rents." *Urban Studies* 32(8): 1331-1344.

- O'Sullivan, D. and Unwin, D. J. (2003). *Geographic Information Analysis*. Toronto: John Wiley and Sons.
- Paez, A.(2005) "Local analysis of spatial relationships: A comparison of GWR and the expansion method." *Lecture Notes in Computer Science* 3482: 162-172.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J. and Deadman, P. (2003) "Multi-agent systems for the simulation of land-use and land-cover change: A review". *Annals of the Association of American Geographers* 93(2): 314-340.
- Pinkse, J. and Slade, M.E. (1998) "Contracting in space: An application of spatial statistics to discrete-choice models." *Journal of Econometrics* 85(1): 125-54.
- Pinkse, J., Slade, M.E. and Shen, L. (2005) "Dynamic spatial discrete choice using one step GMM: An application to mine operating decisions." *Spatial Economic Analysis* 1(1): 53-99.
- Pont, V. and Fontan, J. (2001) "Comparison between weekend and weekday ozone concentration in large cities in France." *Atmospheric Environment* 35(8): 1527-1535.
- Richards, J.A., and Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction, 4th Edition*. New York: Springer.
- Schrank, D. and Lomax, T. (2005) *The 2005 Urban Mobility Report*. Accessed June 1, 2007: http://tti.tamu.edu/documents/mobility_report_2005.pdf.
- Scott, S.L. (2004) "Data augmentation, frequentistic estimation, and the Bayesian analysis of multinomial logit models." Working paper. Accessed May 10, 2007: <http://citeseer.ist.psu.edu/719437.html>.
- Shi, H., Laurent, E. J., LeBouton, J., Racevskis, L., Hall, K. R., Donovan, M., Doepker R. V., Walters, M. B., Lupi, F. and Liu, J. (2006) "Local spatial modeling of white-tailed deer distribution." *Ecological Modelling* 190(1-2): 171-189.
- Smith, T. E. and LeSage, J. P. (2004) "A Bayesian probit model with spatial dependencies." In Pace, R. K. and LeSage, J. P. (Eds.), *Advances in Econometrics Volume 18: Spatial and Spatiotemporal Econometric*. Oxford: Elsevier Ltd.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van der Linde, A. (2002) "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society B* 64(4): 583-616.

- Sun, D., Tsutakawa, R.K. and Speckman, P. L. (1999) "Posterior distribution of hierarchical models using CAR(1) distributions." *Biometrika* 86(2): 341-350.
- Tanner, M. and Wong, W. (1987) "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association* 82: 528-549.
- TCEQ (Texas Commission on Environmental Quality) (2007) *Daily Maximum Eight-Hour Ozone Averages*. Accessed June 9, 2007: http://www.tceq.state.tx.us/cgi-bin/compliance/monops/8hr_monthly.
- Tobler, W. R. (1970) "A computer model simulating urban growth in the Detroit region." *Economic Geography* 46(2): 234-240.
- Townshend, J. R. G. (1992) "Land cover." *International Journal of Remote Sensing* 13: 1319- 1328.
- Vanasse, A., Niyonsenga, T., Courteau, J., Gregoire, J. P., Hemiari, A., Loslier, J. and Benie, G. (2005) "Spatial variation in the management and outcomes of acute coronary syndrome." *BMC Cardiovascular Disorders* 5: 21.
- Waddell, Paul. (2002) "UrbanSim: Modeling urban development for land use, transportation, and environmental planning". *The Journal of the American Planning Association* 68(3): 297-314.
- Wallerman1, J., Vencatasawmy, C. P., and Bondesson, L. (2006) "Spatial simulation of forest using Bayesian state-space models and remotely sensed data." 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Lisbon, Portugal. July 5-7, 2006.
- Wang, X., Carmichael, G., Chen, D., Tang, Y. and Wang, T. (2005) "Impacts of different emission sources on air quality during March 2001 in the Pearl River Delta (PRD) region." *Atmospheric Environment* 39(29): 5227-5241.
- Wang, X. and Kockelman, K. (2005) "Occupant injury severity using a heteroscedastic ordered logit model: distinguishing the effects of vehicle weight and type." *Transportation Research Record* 1908: 195-204.
- Wang, X. and Kockelman, K. (2006a) "Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: application to crash rates in China." *Transportation* 34(3): 281-300.
- Wang, X. and Kockelman, K. (2006b) "Tracking land cover change in a mixed logit model: recognizing temporal and spatial effects." *Transportation Research Record* 1977: 112-120.

- Wang, X. and Kockelman, K. (2006c) "A comparison of simulation techniques: maximum simulated likelihood estimation with correlated observations." Presented at the RSAI's 53rd Annual Meeting, Toronto.
- Wear, D. N. and Bolstad, P. (1998) "Land-use changes in southern Appalachian landscapes: Spatial analysis and forecast evaluation." *Ecosystems* 1: 575-594.
- Wilkinson, G. G. (1996) "Classification algorithms —where next?" In Brivio E., Brivio, P. A. and Rampini, A. (Eds.), *Soft Computing in Remote Sensing Data Analysis*. Singapore: World Scientific.
- Wiedinmyer, C. (1999) *Biogenic Hydrocarbons in Texas: Source Characterization and Chemistry*. Ph.D. Dissertation, The University of Texas at Austin.
- Wooldridge, J. M. (2005) "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity." *Journal of Applied Econometrics* 20(1): 39-54.
- Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Yi, I. and Vishniac, E. T. (1993) "Stochastic analysis of the initial condition constraints on chaotic inflation." *Physical Review D* 47(12): 5280-5294.
- Zhang, L. J., Bi, H. Q., Cheng, P. F. and Davis, C. J. (2004) "Modeling spatial variation in tree diameter-height relationships." *Forest Ecology and Management*. 189(1-3): 317-329.
- Zhao, F. and Park, N. (2004) "Using geographically weighted regression models to estimate annual average daily traffic." *Transportation Research Record* 1879: 99-107.
- Zhou, B. and Kockelman, K. (2007) "Predicting the distribution of households and employment: a seemingly unrelated regression model with two spatial processes." Forthcoming, *Journal of Regional Science*.

APPENDIX A. DERIVATION OF HYPERPARAMETERS FOR β 'S CONDITIONAL POSTERIOR DISTRIBUTION

Section 3.3.4.1 outlines the derivation of β 's conditional posterior distribution, but leaves further details about calculating its hyperparameters for this appendix. Since the calculation of θ , λ and U 's hyperparameters uses a similar method, it is important to understand how this step is derived (i.e., how Equation (3.42) leads to (3.43)). As a supplement, this appendix provides the details for the hyperparameters' calculation.

According to Equation (3.42),

$$p(\beta|\Theta_\beta) \propto \exp\left[-\frac{1}{2}(\beta-c)'H^{-1}(\beta-c)-\frac{1}{2}(U^\lambda-\Delta\theta-X\beta)'\Omega^{-1}(U^\lambda-\Delta\theta-X\beta)\right]$$

The term inside the exponential function can be simplified as following:

$$\begin{aligned} & -\frac{1}{2}\left[(\beta-c)'H^{-1}(\beta-c)+(U^\lambda-\Delta\theta-X\beta)'\Omega^{-1}(U^\lambda-\Delta\theta-X\beta)\right] \\ & = -\frac{1}{2}\left[\begin{aligned} & (\beta'H^{-1}\beta-c'H^{-1}\beta-\beta'H^{-1}c+c'H^{-1}c) \\ & +\left(\begin{aligned} & (U^\lambda-\Delta\theta)'\Omega^{-1}(U^\lambda-\Delta\theta)-\beta'X'\Omega^{-1}(U^\lambda-\Delta\theta) \\ & -(U^\lambda-\Delta\theta)'\Omega^{-1}X\beta+\beta'X'\Omega^{-1}X\beta \end{aligned}\right) \end{aligned}\right] \end{aligned} \quad (A.1)$$

Using C to indicate all items that do not involve β (i.e., constant terms with respect to β), then Equation (A.1) can be further simplified to

$$\begin{aligned}
& -\frac{1}{2} \begin{bmatrix} \boldsymbol{\beta}' \mathbf{H}^{-1} \boldsymbol{\beta} - \mathbf{c}' \mathbf{H}^{-1} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{H}^{-1} \mathbf{c} \\ -\boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\Omega}^{-1} (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta}) - (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta})' \boldsymbol{\Omega}^{-1} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{C} \end{bmatrix} \\
& = -\frac{1}{2} \begin{bmatrix} \boldsymbol{\beta}' (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} + \mathbf{H}^{-1}) \boldsymbol{\beta} - \left((\mathbf{U}^\lambda - \Delta \boldsymbol{\theta})' \boldsymbol{\Omega}^{-1} \mathbf{X} + \mathbf{c}' \mathbf{H}^{-1} \right) \boldsymbol{\beta} \\ -\boldsymbol{\beta}' (\mathbf{X}' \boldsymbol{\Omega}^{-1} (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta}) + \mathbf{H}^{-1} \mathbf{c}) + \mathbf{C} \end{bmatrix} \tag{A.2}
\end{aligned}$$

Letting $\mathbf{A} = \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} + \mathbf{H}^{-1}$ and $\mathbf{b} = \mathbf{X}' \boldsymbol{\Omega}^{-1} (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta}) + \mathbf{H}^{-1} \mathbf{c}$, then Equation (A.2) can be expressed as follows:

$$\begin{aligned}
& -\frac{1}{2} (\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} - \mathbf{b}' \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{b} + \mathbf{C}) \\
& = -\frac{1}{2} (\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} - \mathbf{b}' \mathbf{A}^{-1} \mathbf{A} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{A} \mathbf{A}^{-1} \mathbf{b} + \mathbf{C}) \tag{A.3} \\
& = -\frac{1}{2} \left[(\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b})' \mathbf{A} (\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b}) \right] + \mathbf{C} + \frac{\mathbf{b}' \mathbf{A}^{-1} \mathbf{b}}{2}
\end{aligned}$$

From Equations (A.1), (A.2) and (A.3), it can be obtained that

$$\begin{aligned}
p(\boldsymbol{\beta} | \boldsymbol{\Theta}_\beta) & \propto \exp \left[-\frac{1}{2} (\boldsymbol{\beta} - \mathbf{c})' \mathbf{H}^{-1} (\boldsymbol{\beta} - \mathbf{c}) - \frac{1}{2} (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta} - \mathbf{X} \boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{U}^\lambda - \Delta \boldsymbol{\theta} - \mathbf{X} \boldsymbol{\beta}) \right] \\
& = \exp \left\{ -\frac{1}{2} \left[(\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b})' \mathbf{A} (\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b}) + \mathbf{C} - \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \right] \right\} \tag{A.4} \\
& \propto \exp \left\{ -\frac{1}{2} \left[(\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b})' \mathbf{A} (\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{b}) \right] \right\}
\end{aligned}$$

which is Equation (3.43).

APPENDIX B. DERIVATION OF HYPERPARAMETERS FOR U_{ikt} 'S CONDITIONAL POSTERIOR DISTRIBUTION

This appendix shows details of this calculation for U_{ikt} ($\forall i, k, 0 < t < T$) and explains how Equation (3.74) leads to (3.77).

The second item (un-truncated part) in Equation (3.74) is

$$\exp\left\{-\frac{1}{2\nu_i}\left[(U_{ikt} - \lambda U_{ikt-1} - \theta_i - \mathbf{X}_{ikt}\boldsymbol{\beta})^2 + (U_{ikt+1} - \lambda U_{ikt} - \theta_i - \mathbf{X}_{ikt+1}\boldsymbol{\beta})^2\right]\right\},$$

which can be rewritten as

$$\begin{aligned} & \exp\left\{-\frac{1}{2\nu_i}\left[U_{ikt}^2 - 2U_{ikt}(\lambda U_{ikt-1} + \theta_i + \mathbf{X}_{ikt}\boldsymbol{\beta}) + (\lambda U_{ikt-1} + \theta_i + \mathbf{X}_{ikt}\boldsymbol{\beta})^2\right.\right. \\ & \quad \left.\left.+ \lambda^2 U_{ikt}^2 - 2\lambda U_{ikt}(U_{ikt+1} - \theta_i - \mathbf{X}_{ikt+1}\boldsymbol{\beta}) + (U_{ikt+1} - \theta_i - \mathbf{X}_{ikt+1}\boldsymbol{\beta})^2\right]\right\} \\ & \propto \exp\left\{-\frac{1}{2\nu_i}\left[(1 + \lambda^2)U_{ikt}^2 - 2U_{ikt}(\lambda U_{ikt-1} + \theta_i + \mathbf{X}_{ikt}\boldsymbol{\beta} + \lambda(U_{ikt+1} - \theta_i - \mathbf{X}_{ikt+1}\boldsymbol{\beta}))\right]\right\} \\ & = \exp\left\{-\frac{1 + \lambda^2}{2\nu_i}\left[U_{ikt}^2 - 2U_{ikt}\frac{U_{ikt-1} + \lambda U_{ikt+1} + (1 - \lambda)\theta_i + (\mathbf{X}_{ikt} - \lambda\mathbf{X}_{ikt+1})\boldsymbol{\beta}}{1 + \lambda^2}\right]\right\} \quad (\text{B.1}) \\ & \propto \exp\left\{-\frac{1 + \lambda^2}{2\nu_i}\left[U_{ikt} - \frac{U_{ikt-1} + \lambda U_{ikt+1} + (1 - \lambda)\theta_i + (\mathbf{X}_{ikt} - \lambda\mathbf{X}_{ikt+1})\boldsymbol{\beta}}{1 + \lambda^2}\right]^2\right\} \end{aligned}$$

Just as Equation (3.77) shows, this is a normal distribution with mean

$$a_{ikt} = \left[\lambda U_{ikt+1} + \lambda U_{ikt-1} + (1 - \lambda)\theta_i + (\mathbf{X}_{ikt} - \lambda\mathbf{X}_{ikt+1})\boldsymbol{\beta}\right]/(1 + \lambda^2)$$

and variance $b_{ikt} = \nu_i/(1 + \lambda^2)$.

APPENDIX C. MATLAB CODE FOR THE CORE COMPONENT OF THE DSOP MODEL

This appendix provides Matlab code for the DSOP model's main function (i.e. the function that achieves iterative sampling). This part of the Matlab code represents the major contribution of this study. Of course, the code used to complete this dissertation is far more extensive than this main function, but other parts of the code are not presented here because they are already well understood and not very inspiring. The un-presented code includes sub-modules for reading data, printing results, generating simulated sample data, calculating deviance information criteria (DIC), calculating parameters' marginal effects, predicting future dependent variable values and generating graphs.

The function also uses various established routines. Most of these routines are built in Matlab. All others can be found from LeSage's (1999) spatial econometric toolbox (<http://www.spatial-econometrics.com>), including the following:

- `norm_rnd`: generating random vectors from multivariate normal distribution.
- `chis_rnd`: generating random numbers from chi-square distribution.
- `normt_rnd`: generating random numbers from truncated normal distribution.
- `draw_rho`: updating ρ via univariate numerical integration.
- `semip_eigs`: computing eigenvalues for the weight.
- `semip_parse`: parsing input arguments.

- semip_lndet: computing the log determinant of \mathbf{B}_ρ .

The following is the main function for the DSOP model estimation, showing how the parameters and variables are iteratively sampled:

```
function results=DSOP(y,x,W,T,M,mobs,nS,ndraw,nomit,prior)

% Prepare prior and matrix determinant values for the range of
rho values
[c,H,lamda0,D,sige,alpha,tao,rval,b,G,rho,rmin,rmax,a0,d0,ldetflag,metflag,eflag,order,iter,detval,inform_flag]=semip_parse(prior,n,nS,k);
[rmin,rmax,time1]=semip_eigs(eflag,W,rmin,rmax,M);
[detval,time2]=semip_lndet(ldetflag,W,rmin,rmax,0,order,iter);
rv=detval(:,1);
nr=length(rv);

% Storage for draws
bsave=zeros(ndraw-nomit,k);
thetasave=zeros(ndraw-nomit,M);
lamdasave=zeros(ndraw-nomit,1);
psave=zeros(ndraw-nomit,1);
ssave=zeros(ndraw-nomit,1);
vsave=zeros(ndraw-nomit,M);
gamasave=zeros(ndraw-nomit,nS+1);

% Give initial values and compute commonly used terms
L=zeros(N,M);
cmobs=0;
for m=1:M
    L(cmobs+1:cmobs+mobs(m),m)=1;
    cmobs=cmobs+mobs(m);
end
delta=kron(ones(T,1),L);
HI=inv(H);
HIc=HI*c;
DI=inv(D);
DIlamda0=DI*lamda0;
d0I=inv(d0);
d0Ia0=d0I*a0;
U=y;
U0=zeros(N,1);
U1=[U0;U(1:(T-1)*N)];
invomega=ones(N*T,1);
invV=ones(N,1);
Bp=eye(M);
```

```

theta=ones(M,1);
lamda=0;

% Start sampling
iter=1;
while (iter <= ndraw);
% Update beta
srtinvo=sqrt(invomega);
xs=kron(ones(1,k),srtinvo).*x;
Ainv=inv(xs'*xs + HI);
Ulamda=U-lamda*U1;
zmt=srtinvo.*(Ulamda-delta*theta);
bb=xs'*zmt + H1c;
Ainvb=Ainv*bb;
bhat=norm_rnd(Ainv) + Ainvb;

% Update theta
delome=kron(ones(1,M),srtinvo).*delta;
Athetainv=inv((1/sige)*Bp'*Bp + delome'*delome);
ztheta=srtinvo.*(Ulamda-x*bhat);
btheta=delome'*ztheta ;
Abtheta=Athetainv*btheta ;
theta=norm_rnd(Athetainv) + Abtheta ;

% Update lamda
Alamda=0;
blamda=0;
srtinvV=sqrt(invV);
Ltheta=L*theta;
for t=1:T
    index=N*(t-1)+1;
    indexN=N*t;
    uvsrinv=srtinvV.*U1(index:indexN,:);
    zvsr=srtinvV.*(U(index:indexN,:)-
    x(index:indexN,:)*bhat-Ltheta);
    Alamda=Alamda+uvsrinv'*uvsrinv ;
    blamda=blamda+uvsrinv'*zvsr;
end
Alamda=inv(Alamda+DI);
blamda=blamda+DIlamda0;
Alf=Alamda*blamda;
lamda=normt_rnd(Alf, Alamda, -1,1);

% Update rho (using univariate integration)
C0=theta'*theta;
Wtheta=W*theta;
C1=theta'*Wtheta;
C2=Wtheta'*Wtheta ;
rho=draw_rho(detval,C0,C1,C2,sige,rho);
Bp=eye(M) - rho*W;

```

```

% Update sige
term1=theta'*Bp'*Bp*theta + 2*tao;
chi=chis_rnd(1,M + 2*alpha);
sige=term1/chi;

% Update vi (and form invomega, invV)
cobs=0;
for i=1:M
    obs=mobs(i,1);
    Tee=0;
    for t=1:T
        index=N*(t-1)+cobs;
        e=Ulamda(index+1:index+obs)-
            x(index+1:index+obs,:)*bhat-ones(obs,1)*theta(i);
        ee=e'*e;
        Tee=Tee+ee;
    end;
    chi=chis_rnd(1,rval+obs*T);
    vi(i,1)=(Tee + rval)/chi;
    invV(cobs+1:cobs+obs,1)=ones(obs,1)/vi(i,1);
    cobs=cobs+obs;
end
vi=vi/vi(1,1);
invV=invV*vi(1,1);
invomega=kron(ones(T,1), invV);

%Update gama
for s=2:nS
    inds=find (y(1:N*T)==s-2);
    inds1=find (y(1:N*T)==s-1);
    inf=max([gamahat(s-1) max(U(inds))]);
    sup=min([gamahat(s+1) min(U(inds1))]);
    gamahat(s)=normt_rnd(b(s),G(s,s),inf,sup);
end;

%Update U0
cobs=0;
for i=1:M
    AU0=inv(lamda^2/vi(i,1)+d0I);
    parbu0=lamda/vi(i,1);
    obs=mobs(i);
    for j=1:obs
        index=cobs+j;
        bU0=parbu0*(U(index)-theta(i)-
            x(index,:)*bhat)+d0Ia0;
        AbU0=AU0*bU0;
        U0(index)=norm_rnd(AU0)+AbU0;
    end
    cobs=cobs+obs;
end

```

```

end

cobs=0;
%Update Ut, t is 1
for i=1:M
    obs=mobs(i);
    bUt=vi(i,1)/(1+lamda^2);
    for j=1:obs
        index=cobs+j;
        indexN=N+index;
        aUt=(lamda*U(indexN)+lamda*U0(index)+(1-
        lamda)*theta(i,1)+(x(index,:)-
        lamda*x(indexN,:))* bhat)/(1+lamda^2);
        for s=0:nS-1
            if (y(index)==s)
                U(index)=normt_rnd(aUt,bUt,gamahat(s+1),ga
                mahat(s+2));
            end
        end
    end
    cobs=cobs+obs;
end

%Update Ut, t is between (2, T-1)
for t=2:T-1
    cobs=0;
    for i=1:M
        bUt=vi(i,1)/(1+lamda^2);
        obs=mobs(i);
        for j=1:obs
            index=cobs+j;
            indexN=t*N+index;
            indexN1=indexN-N;
            indexN2=indexN1-N;
            aUt=(lamda*U(indexN)+lamda*U(indexN2)+(1-
            lamda)*theta(i,1)+(x(indexN1,:)-
            lamda*x(indexN,:))* bhat)/(1+lamda^2);
            for s=0:nS-1
                if (y(indexN1)==s)
                    U(indexN1)=normt_rnd(aUt,bUt,gamahat(s+1),
                    gamahat(s+2));
                end
            end
        end
        cobs=cobs+obs;
    end
end

%Update UT
t=T;

```

```

cobs=0;
for i=1:M
    obs=mobs(i);
    for j=1:obs
        index=cobs+j;
        indexN=t*N+index;
        indexN1=indexN-N;
        indexN2=indexN1-N;
        aUTf=(lamda*U(indexN2)+theta(i,1)+x(indexN1,:
        )* bhat);
        for s=0:nS-1
            if (y(indexN1)==s)
                U(indexN1)=normt_rnd(aUTf,vi(i),gamahat(s+1
                ),gamahat(s+2));
            end
        end
    end
    cobs=cobs+obs;
end

U1=[U0;U(1:(T-1)*N)];

if iter>nomit
    bsave(iter-nomit,1:k)=bhat';
    thetasave(iter-nomit,1:M)=theta';
    lamdasave(iter-nomit,1)=lamda;
    psave(iter-nomit,1)=rho;
    ssave(iter-nomit,1)=sige;
    vsave(iter-nomit,1:M)=vi';
    gamasave(iter-nomit,1:nS+1)=gamahat';
end;

iter=iter+1;
end;

results.bdraw=bsave;
results.thetadraw=thetasave;
results.lamdadraw=lamdasave;
results.pdraw=psave;
results.sdraw=ssave;
results.vsave=vsave;
results.gama=gamasave;

```

VITA

Xiaokun Wang was born in Dalian, China on July 17, 1979, the daughter of Shumei Xiao and Shuan Wang. She is married to Wenhan Hu and has one daughter, Isabella Hu. Xiaokun received her B.S. and M.S. degrees in civil engineering at Tsinghua University, China. During that period, she participated in several projects, which involved metropolitan transportation planning, passenger volume forecasting for light rail transit and traffic management system design.

She started her Ph.D. studies at the University of Texas at Austin in the Fall of 2003, as a graduate student research assistant. During her four years of doctoral study, she developed a major interest in integrated transport-land use planning for sustainable urban and regional systems, which involves spatial econometric models and rigorous use of satellite images. Xiaokun has been a contributor to research projects sponsored by the National Cooperative Highway Research Program (NCHRP), the Environmental Protection Agency (EPA), the National Science Foundation (NSF) and the Texas Department of Transportation (TxDOT). The topics of her research range from traffic safety analysis to spatial input-output modeling and statistical methods for transportation data. She is an author of several published papers on such topics.

Permanent Address: Xiehe Hutong No. 13, Beijing, China 100005

This dissertation was typed by the author.