9-5-2013

An Economic Analysis of Evolving Health Hazards

Justin Tevie

Justin Tevie
Candidate

Economics
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Alok Bohara, Chairperson

Robert Valdez (co-chair)

Robert Berrens

Guoyi Zhang

Bruce Hofkin

AN ECONOMIC ANALYSIS OF EVOLVING HEALTH HAZARDS

by

JUSTIN TEVIE

B.Sc., Business Administration, University of Ghana
MDE., Development Economics, Dalhousie University
MA., Economics, University of Alberta

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Economics

The University of New Mexico
Albuquerque, New Mexico

July 2013

DEDICATION

To my parents

ACKNOWLEDGEMENTS

First, I would like to thank my co-chairs, Dr. Robert Valdez and Dr. Alok Bohara, I don't think I could have done this without their support and weekly meetings. The two of you took over my dissertation when it was in complete disarray. I thank my committee members, Dr. Robert Berrens, Dr. Bruce Hofkin and Dr. Zhang for all their suggestions and advice. I would also like to thank all of the professors in the economics departments for their support over the years. I would like to thank Professor Blake Boursaw of the College of Nursing for his assistance in building the bio-economic model and the solution of the optimal control model. I also thank Dr. Gabe Sanchez, Dr. Charlene Porsild and Dr. Caroline Richardson all of the Robert Wood Johnson Foundation Center for Health Policy for their kind words and support. I also would like to thank Dr. Nina Wallerstein and Dr. Lorenda Belone for their support as well. Finally, I would also like to thank Dr. Ron Harstad, Dr. Isaac Miller and Dr. Oksana Loginova of the University of Missouri for motivating me.

AN ECONOMIC ANALYSIS OF EVOLVING HEALTH HAZARDS

By

Justin Tevie

B.Sc., Business Administration, University of Ghana
MDE., Development Economics, Dalhousie University
MA., Economics, University of Alberta
PhD., Economics, University of New Mexico

ABSTRACT

Health hazards (e.g., West Nile virus and antibiotic resistance) by their nature are detrimental to the health of mankind and are a vexing problem for society. Health authorities' awareness of the rising health care costs associated with these health hazards highlights the need to undertake research in these areas. This dissertation presents a series of papers on these health hazards.

Chapter 2 develops a spatial filtering panel data count model to examine the factors that contributed to the high prevalence of human West Nile virus (WNV) in California and Colorado using county-level data from 2003 to 2007. An econometric analysis was performed using a random effects negative binomial model to analyze the economic (income and home foreclosures) and biological (mosquitoes) factors associated with human WNV. Tests reveal the presence of spatial autocorrelation in the dependent variable (human WNV). The presence of this phenomenon implies that WNV in neighboring counties do impact the presence of WNV in adjacent counties. Consequently, the random effects negative binomial model is augmented with a spatially-lagged dependent variable and a spatial filtering term to correct for this problem and obtain unbiased estimates of the variance. Specification tests also show that income and home foreclosures are endogenous, i.e., home foreclosures, income and human WNV counts are determined jointly. Hence an instrumental variable (IV) technique is applied to the spatial filtering and spatial lag random effects negative binomial models to obtain consistent estimates. The former model is preferred because it is parsimonious in terms of a model selection criterion. Tests of over-identification (validity tests) reveal that the excluded instruments are indeed exogenous and for that matter valid. A number of hypotheses are tested regarding the economic and biological variables. The findings indicate that West Nile virus is higher in counties characterized by a low median income, high home foreclosures and high number of mosquito breeding sites. It is recommended that counties that exhibit these economic and biological characteristics should be allocated a higher percentage of resources for surveillance and monitoring of the disease.

Chapter 3 is devoted to disease mapping and presentation of the variography of the various human WNV risk measures. It employs Geographic Information Systems (GIS) mapping tools to create thematic risk or hazard maps that visually depict the predicted probabilities of human WNV and the standardized morbidity ratios. The predicted probabilities were generated from the IV spatial filtering random effects negative binomial model. The hazard maps may ultimately assist policy makers in identifying areas of high and low West Nile virus risk, allocating scarce resources, and disease etiology. Variograms are estimated using geo-statistical methods to examine the spatial structure of the various risk measures. In this regard, both isotropic and anisotropic (directional) variograms are generated using exponential and Gaussian methods. They show the presence of strong spatial patterns in observed West Nile counts and the standardized morbidity ratios, but no spatial patterns in the model residuals. This study demonstrates how econometric methods can be used concurrently with GIS tools to inform public policy on the transmission of human West Nile virus.

Chapter 4 builds a dynamic bio-economic model to study the impact of animal antibiotic use on the evolution of antibiotic resistance in humans. It reveals striking similarities between the theory of exhaustible resources in economics and antibiotic resistance. Antibiotic resistance is modeled as an exhaustible resource (common pool resource) extracted (used) over time. Each time an antibiotic is used it lowers the level of the resource (antibiotic effectiveness) by a small amount and thus raises the cost of using subsequent doses of an antibiotic. This process will continue and the next dose will lower the level of the resource even further making it more costly for future use of the drug. In other words, as more and more antibiotics are used the effectiveness of the drug dwindles

over time. The planner's problem is therefore to find the optimal use of antibiotics in animals and humans over time and this necessitates the use of capital-theoretic methods. Consequently, an optimal control model is developed to examine the trade-offs between current antibiotic use in humans and animals and future antibiotic effectiveness. The results reveal that antibiotics should be used in the animal industry to the point where the immediate net marginal benefit is just counterbalanced by the long-term cost in terms of dwindling drug effectiveness. The results of the simulation exercise show that antibiotic effectiveness decreases over time because of an accumulation of resistance to the drug by bacteria. Also the shadow value of antibiotic effectiveness decreases over time because of the decreasing levels of effectiveness. Sensitivity analyses show that increased use of antibiotics in the animal industry drastically reduces the level of antibiotic effectiveness and its shadow value in a given period.

The results of this dissertation could assist health policy makers in the allocation of scarce resources. The findings underscore the importance of factors such as income, home foreclosures and the number of mosquito pools in the transmission of human WNV. Thematic maps of the standardized morbidity ratios and predicted probabilities provide information on areas of high and low WNV risks. The optimal control model provides an insightful perspective on how to allocate antibiotic resources between animal use and human medicine.

Table of Contents

List of Figures xiii

List of Tables xiv

List of Acronyms xv

Chapter 1: Introduction 1

1.1 Health Hazards 1

1.2 Geographic Variation in Human West Nile Virus 4

1.3 Disease Mapping 4

1.4 Economics of Antibiotic Resistance 5

1.5 Contributions of this Dissertation 5

Chapter 2: Examination of the Geographic Variation in Human West Nile Virus 7

2.1. Introduction 7

2.2 Background Literature 12

2.2.1Economic Risk Factors 12

2.2.2 Environmental Risk Factors 15

2.2.3 Climatic Risk Factors 17

2.2.4 Geographic Risk Factors 19

2.2.5 Literature Synthesis 21

2.3 Transmission Cycle of West Nile Virus 21

2.4 Economics of West Nile Virus Control 22

2.4.1 Economics of Vector Control 23

2.4.2 Economic Losses Associated with WNV 24

2.5 Data Sources 25

2.6 Econometric Modeling 29

2.6.1 Random Effects Negative Binomial (RENB) Model 29

2.6.2 Empirical Model Specification 31

2.6.2.1 Role of Income 32

2.6.2.2. Role of Home Foreclosures 33

2.6.2.3 Role of Mosquitoes 34

2.6.2.4 Role of Population Density 35

2.6.2.5 Role of Climatic Factors 36
2.6.2.6 Role of Time Fixed Effects 37
2.6.3 Statement of Hypotheses 37
2.6.4 Endogeneity 38
2.6.5 Spatial Autocorrelation 40
2.6.5.1 Spatial Lag Method 42
2.6.5.2 Spatial Filtering Technique 43
2.7 Estimation and Results 46
2.7.1 Spatial Autocorrelation 48
2.7.2 California Regression Results 49
2.7.2.1 Non-Spatial Random Effect Negative Binomial Regression 49
2.7.2.2 Spatial Lag Random Effects Negative Binomial Regression 50
2.7.2.3 IV Spatial Lag Random Effects Negative Binomial Regression 50
2.7.2.4 Spatial Filtering Random Effects Negative Binomial Regression 51
2.7.2.5 IV Spatial Filtering Random Effects Negative Binomial Regression 51
2.7.3 Colorado Regression Results 52
2.7.3.1 Non-Spatial Random Effects Negative Binomial Regression 52
2.7.3.2 Spatial Lag Random Effects Negative Binomial Regression 52
2.7.3.3 IV Spatial Lag Random Effects Negative Binomial Regression 53
2.7.3.4 Spatial Filtering Random Effects Negative Binomial Regression 53
2.7.3.5 IV Spatial Filtering Random Effects Negative Binomial Regression 54
2.8 Discussion and Conclusions 79
Chapter 3: Disease Mapping and Variography 82
3.1 Introduction 82
3.2 Thematic Maps 84
3.2.1 Standardized Morbidity Ratio 84
3.2.2 Predicted Probabilities 92
3.3 Variogram Analysis 94
3.3.1 Observed Human West Nile Virus 97
3.3.2 Standardized Morbidity Ratio 100
3.3.3 Model Residuals 102
3.4 Discussion and Conclusions 105

Chapter 4: Economics of Antibiotic Resistance: Impact of Animal Antibiotic Use on the Evolution of Antibiotic Resistance in Humans 107
4.1 Introduction 107
4.2 Background Literature 110
4.2.1 Literature Synthesis 113
4.3 Animal Antibiotic Use 114
4.3.1 Benefits of Using Antibiotics in Animal Production 114
4.3.2 Human Health Risks 115
4.3.2.1 Contamination of Food Products 116
4.3.2.2 Transmission of Resistant Genes 116
4.3.2.3 Disease Transmission by Farm Workers 116
4.3.2.4 Contamination by Animal Waste 117
4.4 Antibiotic Resistance: A Primer 117
4.5 An Epidemiological Model of Infectious Diseases 118
4.6 Mathematical Model 121
4.7 Economics of Antibiotic Resistance 123
4.7.1 Antibiotic Resistance as an Exhaustible Resource 123
4.7.2 An Economic Approach to Antibiotic Resistance 124
4.8 Derivation of the Singular Controls 128
4.8.1 Derivation of the Double Singular Control Path 129
4.9 Simulation Methods and Results 131
4.9.1 Illustrative Example: Tetracycline Use in Humans and Animals 132
4.9.2. Sensitivity Analysis 133
4.10 Discussion and Conclusions 138
Chapter 5: Discussion and Conclusions 142
5.1 Summary of Dissertation 142
5.2 Geographic Variation in Human West Nile Virus 142
5.3 Disease Mapping and Variography 143
5.4 Economics of Antibiotic Resistance 143
5.5 Policy Implications 144
Appendices 147
Appendix A: General Solution for Effectiveness 148

Appendix B: Derivation of Threshold Value of Effectiveness 149

Appendix C: Explanation of Optimal Control Units 150

Appendix D: Stata, R and Excels Codes 152

Appendix E: Supplementary IV Spatial Filtering RENB Regressions……. 181

References 185

List of Figures

Figure 2-1: Comparison of Yearly Infections: The Two States and US 8
Figure 2-2: Comparison of Yearly Fatalities: The Two States and US 8
Figure 2-3: Comparison of Economic Risk Factors (2003-2007) 28
Figure 2-4: Comparison of Biological Risk Factors (2003-2007) 28
Figure 2-5: Spatial Distribution of Human WNV in California (2003-2007) 71
Figure 2-6: Spatial Distribution of Income in California (2003-2007) 72
Figure 2-7: Spatial Distribution of Home Foreclosures in California (2003-2007) 73
Figure 2-8: Spatial Distribution of Mosquito Pools in California (2003-2007) 74
Figure 2-9: Spatial Distribution of Human WNV in Colorado (2003-2007) 75
Figure 2-10: Spatial Distribution of Income in Colorado (2003-2007) 76
Figure 2-11: Spatial Distribution of Home Foreclosures in Colorado (2003-2007) 77
Figure 2-12: Spatial Distribution of Mosquito Pools in Colorado (2003-2007) 78
Figure 3-1: Raw SMR Estimates for California 88
Figure 3-2: Raw SMR Estimates for Colorado 88
Figure 3-3: Empircal Bayes Poisson-Gamma SMR Estimates for California 89
Figure 3-4: Empirical Bayes Poisson-Gamma Estimates for Colorado 89
Figure 3-5: Empirical Bayes Log-Normal SMR Estimates for California 90
Figure 3-6: Empirical Bayes Log-Normal SMR Estimates for Colorado 90
Figure 3-7: Empirical Bayes Global Marshall SMR Estimates for California 91
Figure 3-8: Empirical Bayes Global Marshall SMR Estimates for Colorado 91
Figure 3-9: Predicted Probabilities for California 93
Figure 3-10: Predicted Probabilities for Colorado 94
Figure 3-11: West Nile Virus Isotropic Variogram for California 97
Figure 3-12: West Nile Virus Isotropic Variogram for Colorado 98
Figure 3-13: West Nile Virus Anisotropic Variogram for California 98
Figure 3-14: West Nile Virus Anisotropic Variogram for Colorado 99
Figure 3-15: Standardized Morbidity Ratio Isotropic Variogram for California 100
Figure 3-16: Standardized Morbidity Ratio Isotropic Variogram for Colorado 101
Figure 3-17: Standardized Morbidity Ratio Anisotropic Variogram for California 101
Figure 3-18: Standardized Morbidity Ratio Anisotropic Variogram for Colorado 102
Figure 3-19: Model Residuals Isotropic Variogram for Colorado 103
Figure 3-20: Model Residuals Isotropic Variogram for Colorado 103
Figure 3-21: Model Residuals Anisotropic Variogram for California 104
Figure 3-22: Model Residuals Anisotropic Variogram for Colorado 104
Figure 4-1: Trajectory of Antibiotic Resistance 135
Figure 4-2: Optimal Time Path of the Shadow Value of Effectiveness 135
Figure 4-3: Time Path of Infections 136
Figure 4-4: Time Path of the Shadow Value of Infections 136
Figure 4-5: Time Paths of Antibiotic Effectiveness with Varying Animal Antibiotic Use 137
Figure 4-6: Optimal Time Paths of the Shadow Value of Effectiveness with Varying Animal Antibiotic Use 137

List of Tables

Table 2-1: Factors Used to Assess Risk of Human WNV in California 26
Table 2-2: Factors Used to Assess Risk of Human WNV in Colorado 27
Table 2-3: Global Spatial Autocorrelation Test on WNV counts 49
Table 2-4: Non-Spatial Random Effects Negative Binomial Regression for California 55
Table 2-5: Spatial Lag Random Effects Negative Binomial Regression for California 56
Table 2-6: IV Spatial Lag Random Effects Negative Binomial Regression for California 57
Table 2-7: Spatial Filtering Random Effects Negative Binomial Regression for California 58
Table 2-8: IV Spatial Filtering Random Effects Negative Binomial Regression for California ... 60
Table 2-9: Non-Spatial Random Effects Negative Binomial Regression for Colorado 62
Table 2-10: Spatial Lag Random Effects Negative Binomial Regression for Colorado 63
Table 2-11: IV Spatial Lag Random Effects Negative Binomial Regression for Colorado 64
Table 2-12: Spatial Filtering Random Effects Negative Binomial Regression for Colorado 65
Table 2-13: IV Spatial Filtering Random Effects Negative Binomial Regression for Colorado... 67
Table 2-14: Selected Eigenvectors for Spatial Filtering Random Effects Negative Binomial Regression 68
Table 2-15: Selected Eigenvectors for IV Spatial Filtering Random Effects Negative Binomial Regression 68
Table 2-16: Sign and statistical significance of the economic and biological covariates for California 69
Table 2-17: Sign and statistical significance of the economic and biological covariates for Colorado 70
Table 4-1: Parameter Values and Description Used in Simulations Exercises 141

List of Acronyms

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CDC	Centers for Disease Control
EM	Expectation Maximization
GIS	Geographic Information Systems
GLM	Generalized Linear Model
IV	Instrumental Variable
RENB	Random Effects Negative Binomial
SMR	Standardized Morbidity Ratio
WNV	West Nile Virus
USDOC	US Department of Commerce

Chapter 1: Introduction

1.1 Health Hazards

A health hazard is an activity, event or condition that has the potential to cause acute or chronic illness or even death if exposure to such event or condition is not mitigated. One difficulty for researchers and practitioners is developing an economic framework for studying these hazards. Often, these hazards are considered within an epidemiological framework. Health hazards are studied from a resource allocation and management framework in this dissertation. Economic theory is primarily concerned with the allocation of scare resources that have alternative uses. This dissertation focuses on the major determinants of these hazards. The determinants of health hazards are used as inputs in decisions relating to the allocation of resources. Devoting more resources to the mitigation of these hazards could have significant opportunity costs. This research is primarily motivated by this basic economic principle that resources are limited and this necessitates the need for an optimal allocation.

The West Nile virus (WNV) is a health hazard because it negatively impacts the health of the population leading to a loss of work days, a decline in economic output, and an increase healthcare cost. Past research on the study of WNV focused on the role of climatic factors (Gibbs et al. 2006; Winter et al. 2008; Ruiz et al. 2004). Others focused on the role of biological factors such as dead birds in WNV transmission (Patnaik 2007; Kwan et al. 2008; Nielson et al. 2008), while a few focused on economic factors such as per capita income and mortgage delinquencies (Harrigan et al. 2010; Reisen et al. 2007). These studies were limited to only one county and also were conducted in a non-economic framework. A purpose of this dissertation is to examine the role of economic

(income and home foreclosures) and biological (mosquito breeding sites) factors in WNV transmission. The approach adopted in this research is novel in the sense that it explores how these economic and biological factors can aid in the allocation of scarce resources to mitigate WNV.

The study of WNV is important for three reasons. First, WNV is the leading cause of arboviral diseases in the US and the average number of confirmed cases per year between 1999 and 2008 was 2,896 (Lindsey et al. 2010). It was estimated that about 80 percent of all WNV human infections are asymptomatic, i.e., symptoms do not manifest in the individual (Mostashari et al. 2001). Second, they impose some economic costs on individuals and governments. Individuals who are infected with WNV face medical and non-medical costs. These include pharmacy/medical supplies, diagnostics, productivity losses and premature deaths, cost of insecticide spraying and transportation to visit a healthcare provider. Government agencies face costs relating to mosquito vector surveillance and control programs. As select examples, Zohrabian (2004) indicated that West Nile cost the state of Louisiana about $20.1 million in 2002 and the Huffington Post (2012) reported that on average the areal spraying costs in a single county in California is about $1 million. Third, they impose an externality on society. Undertaking preventive actions or policies may affect the likelihood that other people become infected. For example, vector control strategies such as the spraying of mosquito breeding sites with insecticides will lessen the probability that others will be bitten by an infected mosquito. This is something that the sprayer of the sites ignores in his actions, which is referred to as "pure infection externality" by Gersovitz and Hammer (2004).

Antibiotic resistance is another example of a health hazard because it reduces the effectiveness of an antibiotic (increase in resistance) and could render it ineffective overtime leading to rising bacterial infections. Previous studies on the study of antibiotic use have been undertaken solely within an economic, biological or epidemiological framework. As noteworthy examples, Seechi and Babcock (2002) studied optimal animal antibiotic use within an economic framework and Massad et al. (2008) studied optimal antibiotic use within a biological framework. The approach pursued in this dissertation is innovative because it integrates biological and epidemiological information into a resource allocation framework in economics. Antibiotic effectiveness is treated as an exhaustible resource that dwindles overtime. Thus, the economic model presented provides a set of conditions that characterize the optimal use of antibiotic resources.

The study of antibiotic resistance is important from an economic standpoint for two reasons. First, economic cost to society resulting from antibiotic resistance includes higher mortality rates, longer hospital stays and the use of more expensive antibiotics. Coast and Smith (2003) estimated that the cost of antibiotic resistance in the United States amounts to approximately $7 billion annually. Second, the use of antibiotics imposes an external cost on society. For example, in using antibiotics, individuals often ignore the effect of their actions on the future effectiveness of the drug. This results in declining levels of antibiotic effectiveness, which is a valuable resource to society.

The analyses of these health hazards is organized into three chapters namely the geographic variation in West Nile virus, disease mapping (e.g., incidence of West Nile) and the economics of antibiotic resistance.

1.2 Geographic Variation in Human West Nile Virus

This chapter primarily examines the role of income and home foreclosures in the transmission of WNV. It is important for policy makers to understand how these economic factors and WNV are linked to ensure that scarce resources are allocated as efficiently as possible to mitigate the disease. The basic economic tenet of this chapter is that median household income and the number of home foreclosures can aid in resource allocation. Areas with a lower median household income have a higher prevalence of WNV because its residents suffer disproportionately from adverse neighborhood and environmental conditions. Counties with a lower number of home foreclosures have a lower incidence of WNV because the number of unmaintained properties and neglected swimming pools are lower. On the basis of this hypothesis, more resources should be allocated to areas with low median household income and a high number of home foreclosures.

1.3 Disease Mapping

The objective of this chapter is to provide a geographical distribution of the WNV health hazard. The approach is to map relative risk that reflects the number of people who are infected with the virus (morbidity) in a given period of time. A map of the relative risk provides a visual summary of high-risk areas which are the focus of monitoring and surveillance efforts. Thus, they will ultimately aid in the allocation of scarce health resources for monitoring and surveillance. These maps are also helpful in assisting policy makers, among other things, in identifying the risk factors associated with the spatial distribution of WNV and in disease etiology.

1.4 Economics of Antibiotic Resistance

This chapter develops a dynamic bio-economic model to study antibiotic resistance. This model is quite innovative because it integrates biological and epidemiological components into a resource management framework. The use of this model is appealing because it provides an insightful perspective on the inter-temporal trade-offs that exist between current antibiotic use in animals and humans and future drug effectiveness. Second, it allows us to study the dynamic interaction between animal antibiotic use and drug effectiveness. The results of this model provide guidelines on the optimal allocation of antibiotic resources between animal production and human medicine. Also, it provides an insight into the long-term balance between antibiotic use and antibiotic effectiveness.

1.5 Contributions of this Dissertation

This research undertakes three types of analyses on health hazards focusing on the geographic variation in WNV, disease mapping and economics of antibiotic resistance. The analyses are presented in the following three chapters: chapter 2, chapter 3, and chapter 4.

Chapter 2 is devoted to examining the factors associated with WNV transmission. The results indicate that WNV is negatively impacted by median household income. A novel finding of this research is that home foreclosures have a significant positive impact on the prevalence of WNV. Chapter 3 uses medical data to map the relative risk of WNV. The results indicate that areas close to each other seem to have similar relative risk values than those farther apart. This phenomenon is interesting and suggests that policy coordination is necessary in order to mitigate WNV. Chapter 4 develops an economic model to illustrate the trade-offs that result from using antibiotics in animal production

and in human medicine. Antibiotics effectiveness is treated as a non-renewable resource that is gradually extracted. An interesting finding is that increasing animal antibiotic use results in decreasing drug effectiveness or increasing antibiotic resistance.

This dissertation provides important results for advising health authorities on these health hazards. The results suggest that more resources should be allocated to areas that have low median income levels and high number of home foreclosures for WNV monitoring and surveillance. This research highlights the importance of neighborhood conditions in WNV prevalence. Poor environmental conditions, caused by economic hardships, could serve as breeding grounds for the mosquito vector. This suggests that maintenance of foreclosed homes should be encouraged because of the presence of standing water in swimming pools which serves as a breeding ground for mosquitoes. The findings of the antibiotic resistance research indicate that measures to reduce animal antibiotic use such as prescription requirements, would greatly assist in prolonging the potency of an antibiotic. This is because prescriptions can potentially decrease the volume of animal antibiotic use thereby increasing drug effectiveness.

Chapter 2: Examination of the Geographic Variation in Human West Nile Virus

2.1. Introduction

West Nile virus (WNV) is a mosquito-borne disease transmitted principally by mosquitoes in the *Culex* genus-*pipiens*, *quinquefasciatus* and *tarsalis*. Several bird species serve as reservoir hosts (Kramer et al. 2007). Since 1999, when the virus was first detected in New York, it has spread westward to all 48 contiguous states. Between 1999 and 2011, more than 30,000 humans have been infected with the virus, and 400 of these resulted in deaths (CDC 2011). The states of California and Colorado recorded some of the highest cases of WNV in the United States (US) between 2003 and 2007 (CDC 2011). During this time period, these two states accounted for approximately 28.2% of all WNV cases and 20% of all fatalities (CDC 2011). These states have consistently ranked either first or second in terms of WNV infections during the time period under study (CDC 2011). These facts are visually shown in Figures 2-1 and 2-2. This observation is very interesting because these states constitute approximately 7.34% of the total land area of US, but account for almost 13.6% of the US population. It is somehow intriguing why the incidence of WNV is disproportionately higher in these states.

These states are quite diverse in terms of geographical, climatic, demographic, socioeconomic and environmental factors. So they merit separate investigations as to the determinants of WNV incidence in these states. They also offer an important opportunity to evaluate the potential risk factors associated with the transmission of WNV to humans at the county level. Several studies on the determinants of WNV in these states have focused on geographic and climatic factors (Winters et al. 2008; Mongoh et al. 2007; Patnaik et al. 2007).

Figure 2-1: Comparison of Yearly Infections: The Two States and US

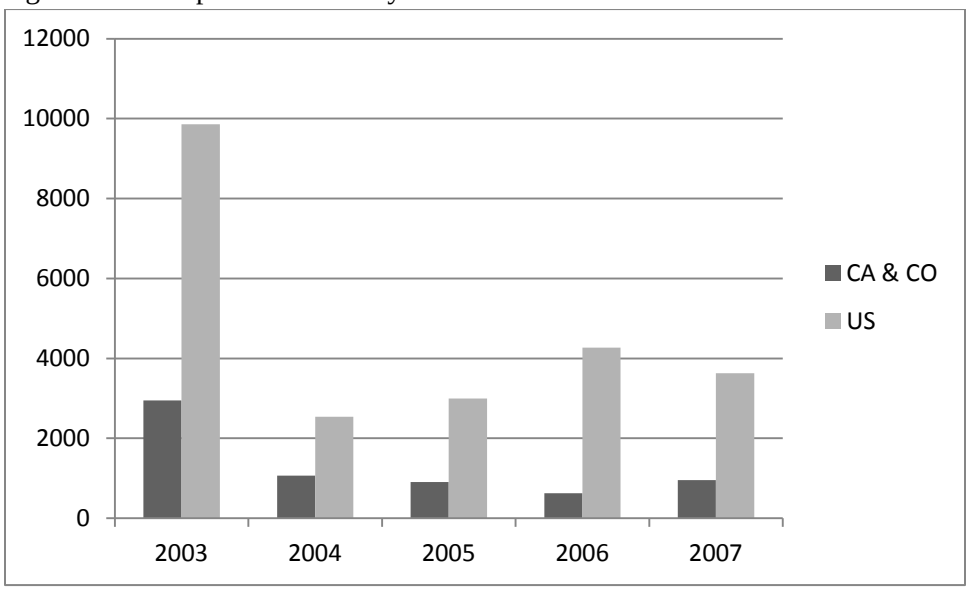
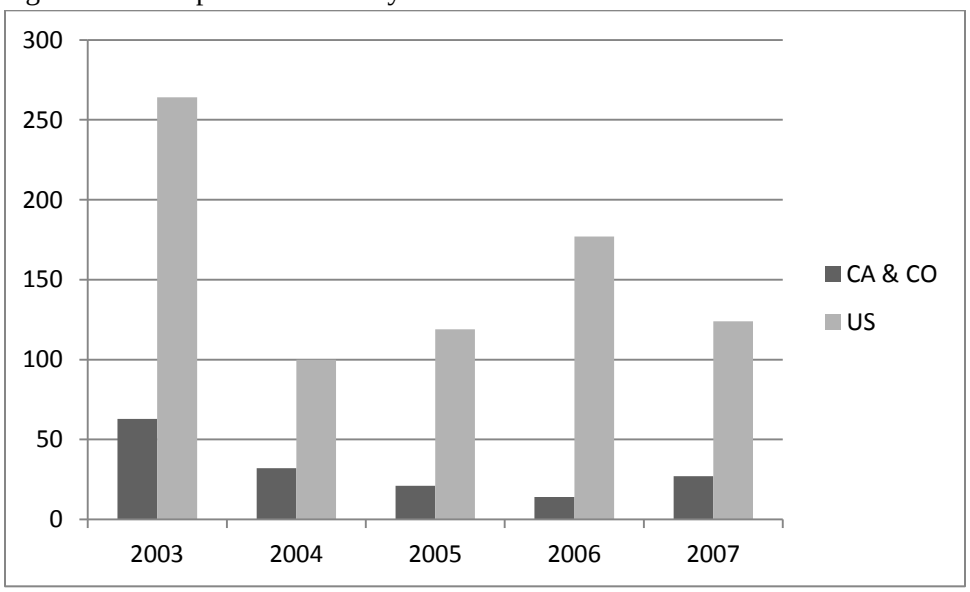


Figure 2-2: Comparison of Yearly Fatalities: The Two States and US



However, the role of economic factors in the transmission of the disease seems to be neglected in the literature. A number of WNV studies conducted in California (Reisen et al. 2008; Harrigan et al. 2010) focused on the role of economic conditions (per capita

income and mortgage delinquencies) in the transmission of WNV. These studies, however, were limited to only two counties in California; Orange and Kern. Similarly studies by Patnaik (2007), Kwan et al. (2008), Nielson et al. (2008) and Carney et al. (2011) which examined the role of dead birds and mosquitoes in WNV transmission in Colorado and California were restricted to a few counties. It is reasonable to argue that the presence of the *Culex* mosquito is necessary for WNV transmission, but not sufficient for an outbreak or spread of the disease. To improve our understanding of the potential risk factors that may amplify the incidence of WNV in these states, this study models WNV using data from all counties in both states. It is plausible to speculate that certain social, economic, demographic, environmental/biological, and ecological factors may be responsible for amplification/propagation of the disease.

A variety of approaches have been employed in the literature to examine the determinants of WNV such as logistic regression models (Gibbs et al. 2006; Winter et al. 2008; Ruiz et al. 2004) and principal component analysis (Mongoh et al. 2007). A critical review of the literature reveals some weaknesses of these studies. In fact, most did not adequately address some or all of estimation issues such as spatial autocorrelation, endogeneity, the count nature of human WNV, area-specific heterogeneity and the panel nature of the data. Albeit some studies (Messina et al. 2011; Linard et al. 2007) have used negative binomial techniques to address some of these issues, they fail to adequately address the panel nature of the data and potential endogeneity issues. The methodological approaches adopted in this dissertation overcome these weaknesses. This dissertation employs a panel data count model to study the key determinants of human WNV. The use of panel data has several advantages. First, they contain more variation and information

than cross-sectional data. Second, panel data leads to an increase in efficiency because of the availability of greater degrees of freedom. Third, we can control for the effects of missing or unobserved variables by the use of a random heterogeneity term.

This study contributes to the health economics literature in a number of ways. First, it employs a random effects negative binomial model to primarily study the importance of economic factors in the transmission of WNV in California and Colorado. Many studies have focused on the significance of climatic and geographic factors, however the role of factors such as income and home foreclosures have been under-studied in the literature. Second, it addresses the issue of spatial autocorrelation in the dependent variable within the context of a panel count data model. The presence of spatial autocorrelation in the dependent variable will cause the variance to be biased if not corrected. It uses a spatially- lagged dependent variable and employs a spatial filtering technique to remove spatial autocorrelation from the model. Third, it applies an instrumental variable technique to the spatial lag and spatial filtering random effects negative binomial models to correct for endogeneity in income and home foreclosures. The presence of endogeneity will cause the estimates to be biased and inconsistent if not rectified.

The overall objective of this chapter is to determine the important drivers of human WNV. The specific objectives are: (1) primarily examine the importance of economic (income and home foreclosures) factors in the transmission of human WNV in the states of California and Colorado using county-level panel data from 2003 to 2007. The role of biological factors in the propagation of the disease is also investigated. In other words, it tests several hypotheses regarding economic and biological factors in the

transmission of human WNV. The analysis in this chapter uses the number of mosquito pools (mosquito breeding sites) as the only biological factor because a supplementary analysis conducted using dead birds showed that they do not strongly impact the transmission of WNV. An econometric analysis is performed to analyze the determinants of human WNV within a random effects negative binomial model framework. Based on a mix of theory and empirical studies climatic and spatial factors are also controlled for; and (2) Determine the nature of spatial autocorrelation and how important spatial factors are in explaining human WNV patterns overtime and construct an econometric model that incorporates such spatial pattern in order to provide reliable estimates. A spatial filtering model is proposed as a non-parametric method to remove spatial autocorrelation. The econometric model proposed in this paper is one possible approach that can be used to study the determinants of human WNV. This dissertation proposes that the use of this model will not only emphasize the role of economic and biological factors in human WNV amplification, but will also provide policy makers with a more appropriate framework for policy making purposes. In particular, the results show that income, home foreclosures and the number of mosquitoes breeding sites contribute significantly to the risk of human WNV. An empirical methodology is employed based on data obtained through archival, library, internet searches, and official State and Federal policy documents.

The organization of this chapter is as follows. To underscore the dearth of studies on the role of particularly economic factors in the transmission of WNV in Colorado, California and the US in general, a detailed review of the relevant literature is provided in section 2. In section 3, the transmission cycle of WNV is reviewed. The economic costs

associated with WNV control and the economics of vector control are discussed in section 4. In section 5, a brief discussion of the data sources are presented. The econometric models employed in this paper are discussed in section 6, were the relative differences in models are described. Also the role of economic factors in and the determination of health outcomes are discussed. The results of this study are presented in detail in section 7. In section 8, a discussion of the empirical evidence is undertaken and the conclusions and policy relevance are summarized.

2.2 Background Literature

The purpose of this section is to review studies on WNV with a view to highlighting the lack of research on the role income and home foreclosures in the transmission of WNV.

2.2.1Economic Risk Factors

Socioeconomic factors are hypothesized to be associated with human diseases because they may explain health disparities that may exist among certain segments of society (Nazroo 2003). In fact, the significance of socioeconomic factors is still the subject of considerable debate. Some authors indicate that they make a minimal contribution to explaining health disparities (Wild and McKeigue 1997).

Messina et al. (2011) examined the spatial distribution of human WNV cases in the greater Chicago area from 2002 to 2006. The study area included Cook and Dupage counties because they had a relatively higher number of WNV cases than other counties. The authors identified certain environmental and socioeconomic factors that could serve as predictors of WNV. These include elevation, vegetation/land cover, housing age, race, age, income, the number of mosquito infection rate and April to August precipitation.

Generalized linear models (GLM) were used to model the determinants of WNV for each year and for all years. In particular, a Poisson or negative binomial model was used depending on whether the data displayed over-dispersion or not. Moran's *I* test revealed the presence of both positive and negative local spatial autocorrelation. The final models estimated were accordingly corrected for spatial autocorrelation. Their findings showed that elevation, vegetation and housing density had a statistically significant negative impact on the incidence of WNV, while race and income had a statistically significant positive effect on WNV.

The study by Harrigan et al. (2010) focused on the role of economic factors in predicting the high incidence of human WNV in Orange County, California. This county has a reputation for being a hot spot in terms of the disease from 2004-2008. The variables used in their study included per capita income, number of neglected swimming pools, population density, elevation, temperature and a vegetation cover index. They employed a spatial modeling approach in order to determine the factors that could predict the incidence of WNV in Orange County. Their findings showed that per capita income explained the largest variation in WNV in both mosquitoes and humans. Their results also suggest that poorer communities (low-income neighborhoods) created environmental conditions conducive to the survival of the mosquito vector. The argument was that neglected swimming pools on foreclosed homes collect pockets of water that could aid in vector amplification. Additionally, poorer communities may not have well-functioning drainage systems which could further aid in the multiplication of the mosquito vector.

Reisen et al. (2008) used surveillance and survey data to examine the factors responsible for the high prevalence of WNV in Kern County, California, using 2007 data.

They were interested in investigating whether the economic downturn and accompanying variable mortgage rates contributed to the outbreak of the disease. Their findings indicated that economic factors overwhelmingly accounted for the spike in WNV from 2006 to 2007. The aerial survey showed the existence of a large number of abandoned swimming pools on foreclosed residential properties that were infested with algae. They concluded that the housing crisis in 2007 that emanated from adjustable mortgage rates and sub-prime lending were responsible for this trend. During this period Kern County saw a 300% increase in mortgage delinquencies, and a 276% increase in human WNV cases. The astronomical increase in the disease was attributed to the rise in the number of neglected swimming pools on foreclosed homes that serve as a breeding site for the mosquito vector.

Ruiz et al. (2007) examined the determinants of the high incidence of human WNV cases in the urban areas of Chicago and Detroit using 2002 data. They identified 18 risk factors pertaining to socioeconomic, environmental and landscape factors. Socioeconomic factors included median family income, median age of the population, average age of housing, and the fraction of housing built in each decade from 1940 to 2000. Environmental factors included vegetation and land cover variables. In order to isolate which of these factors that could potentially explain most of the variation in human WNV cases, a principal component analysis was undertaken to select core variables that were unrelated. One important aspect of their study was the presence of spatial patterns in the data. In other words, tracts that displayed similar WNV cases were clustered together. They partitioned Chicago and Detroit into five urban classes: city, low income; city, high income; inner suburbs; outer suburbs; urban, no-man's land. Analysis

of variance (ANOVA) methods were then used to reveal that there existed significant differences in the means of WNV cases for all urban classes. Age of the housing unit and land cover variables were significant in explaining the prevalence of WNV in the two cities.

2.2.2 Environmental Risk Factors

The environment may assist in creating biological or etiological conditions for the amplification of both vector and host. Earlier studies on the potential environmental risk factors are comparable to the analysis in this essay because the most dominant mosquito species in those studies are *Culex pipiens* and *Culex tarsalis*. Kwan et al. (2010) undertook a study to determine the ecological and environmental factors responsible for the emergence and persistence of WNV in the Los Angeles area between 2003 and 2008. The study area focused on the Los Angeles Basin, San Fernando Valley, San Gabriel Valleys, and the Santa Clarita Valley. The researchers were interested in investigating whether significant differences existed in the means of mosquito abundance, dead birds, sentinel chickens, temperature and rainfall. The ANOVA method revealed that there were significant differences in WNV cases whether mean temperatures were classified by year or by region. Also the ANOVA method showed that mean mosquito abundance (trap counts) was not statistically different between the Los Angeles Basin and Santa Clarita Valley. The mean difference in dead birds was found to be significantly different for each of the years.

In another study conducted for the period 2002-2006, DeGroote et al. (2008) examined several factors that could account for the high incidence of WNV in Iowa using census block data. The authors used a variety of geographic, demographic, and climatic

variables as potential risk factors. The main technique of investigation was ANOVA. This method was used to identify factors that could explain differences in mean WNV cases between census blocks with WNV cases and those without WNV cases. The results revealed that population density, stream density, road density, land cover, irrigation, mean precipitation, and mean temperatures were statistically significant in explaining differences in mean WNV between census blocks. Tests revealed the presence of positive spatial autocorrelation. This means that adjoining census tracts have similar WNV counts.

In a similar study Patnaik et al. (2007) conducted a study in the Colorado counties of Adams, Arapahoe, and Douglas in 2002. They were interested in isolating those environmental factors that could reliably predict the incidence of WNV cases. The environmental factors considered were birds that tested positive for WNV, the number of mosquito pools, the equine population, and the number of sentinel chickens. Using a spatial model developed using geographic information systems (GIS) they found that dead birds were responsible for predicting a vast majority of human cases. Mosquito pools and the equine population also predicted a reasonable proportion of human cases, but not as successfully as dead birds.

Carney et al. (2011) developed a dynamic continuous area space-time (DYCAST) model to predict human cases of WNV based on dead birds in the California counties of Orange, Riverside and San Bernardino because these counties accounted for about 85% of all human cases in 2004. The results showed that the DYCAST model was 91% accurate in predicting human WNV cases using information on dead birds. In addition, the model identified high-risk populations about 5 weeks in advance of an outbreak.

In another study Nielson et al. (2008) used GIS tools as well as spatial statistics and surveillance systems to study the risk factors associated with human WNV infection in Davis (California). Their results showed that dead birds and equine cases were responsible for the high WNV counts in 2005. In 2006, on the other hand, mosquito pools were associated with the high counts of WNV.

2.2.3 Climatic Risk Factors

The impact of climate change and the environment on infectious diseases has been the subject of considerable debate, speculation and extensive study for centuries. Shope (1991) cites Jacob Henles 1840 study entitled "*On Miasmata and Contagia*" in which Henles asserts that

> Heat and moisture favor the production and propagation of the infusoria and the molds, as well as the miasmata and contagia, therefore miasmatic – contagious diseases are most often endemic in warm moist regions and epidemic in wet summer months

Climate change is one of several variables that affect the rate of infectious diseases. It may affect the geographical distribution of zoonotic diseases prevalent in the United States. McMichael et al. (1996) contend that many of the biological organisms and processes associated with the propagation of infectious diseases are specifically influenced by variations in climatic variables such as temperature, precipitation and drought. McMichael et al. (1996) suggest that the impact of climate change on zoonoses can be understood by examining the direct effects of climate on vector biology, indirect effects of climatic variables on vectors, and the effect of climate change on the distribution of vector-borne diseases. The climatic variables that directly affect the biology of the vector, pathogen and host include temperature, humidity and precipitation.

The studies on the relationship between climate change and WNV in the US have focused on three main climatic variables: temperature, precipitation and drought. Wimberly et al. (2008) examined the role of climatic and land use variables in the transmission of WNV in the Northern Great Plains of the US (North Dakota, South Dakota, Nebraska, and some areas of Wyoming and Montana) using county-level data for 2002 and 2003. They modeled WNV using a two-step process. In the first step of the modeling the determinants of WNV, five different models were run using a second-trend surface model of temperature and precipitation. In this stage, they accounted for spatial error effects generated by a conditional autoregressive process (CAR). This was to correct for any biases that might be present due to spatial autocorrelation. In the second step of their modeling exercise, the final model included land cover variables in addition to the climatic variables used in step one. Cluster analysis revealed that all three states in general formed one cluster. Their results indicated that temperature, irrigated lands, rural areas and wetlands had a substantial positive impact on the prevalence of WNV, while precipitation had a significant negative impact on WNV prevalence.

Mongoh et al. (2007) studied the environmental and ecological determinants of equine WNV cases in North Dakota in 2002. Potential risk factors which they posited could be responsible for the occurrence of WNV in horses include birds, human cases, mosquito pools, temperature, rainfall, distance from water bodies, and elevation. A principal component analysis was undertaken to create a set of uncorrelated variables and to determine which components accounted for most of the variation (or were significant). The results of principal components regression showed that human WNV cases, positive mosquito pools, mean temperature, mean elevation, mean precipitation, and distance to

water bodies were statistically significant in explaining WNV in horses. Specifically, positive mosquito pools, mean temperature, and distance to water bodies were negatively related to equine WNV cases, while human WNV cases, mean elevation and mean precipitation were positively related to the occurrence of WNV in horses.

2.2.4 Geographic Risk Factors

Gibbs et al. (2006) undertook a WNV study to examine the factors that determined the serostatus of avian/bird collection sites in Georgia between 2002 and 2004. The authors identified variables such as land cover, elevation, human population density, and climatic variables as potential determinants. They employed a logistic regression method in conjunction with GIS tools to determine which of these variables could plausibly explain the geographical distribution of WNV in Georgia. The results of the logistic regression showed that temperature, housing density, land use in urban/suburban areas, and elevation were statistically significant in predicting the geographical distribution of WNV in Georgia. In particular, the risk of WNV was found to be higher in urban/suburban areas and lower in mountainous areas.

Brown et al. (2008) examined the determinants of WNV in the eight Northeastern US states using county-level data from 1999-2006. They were particularly interested in the role of ecological factors in the transmission of the disease. The main variable of interest to the researchers was land use. To begin with, they classified land use as either urban or forest. Urban land included the following classes: low-intensity residential, high-intensity residential, commercial/industrial/transportation, and urban/recreational grasses. Land use devoted to deciduous, evergreen and mixed forests were classified as forests. The statistical methods they employed took into account the effect of spatial proximity. Their

findings suggest that urban counties where forest cover is less than 38%, were 4.4 times more likely to have above-median WNV incidence than rural counties where forest cover is greater than 70%.

Winters et al. (2008) developed a predictive spatial model to study the geographic factors associated with the prevalence of human WNV cases in western and eastern Colorado using data from 2002-2006. The environmental factors considered in this study include elevation, July precipitation, heating-degree days in August, a vegetation cover index, September snow amounts, and March temperatures. A multivariate logistic model was employed to assist in predicting the incidence of WNV cases in both western and eastern Colorado. The models accurately predicted the high incidence of WNV in 27% of areas in eastern Colorado, compared to only 12% in western Colorado.

Ruiz et al. (2004) analyzed factors associated with the incidence of human WNV in the Chicago area in 2002. This study was conducted in Cook and Dupage counties because they had the highest number of human WNV cases. The geographic and environmental risk factors considered included population density, income, race, age, elevation, distance to dead birds, and mosquito abatement efforts. Discriminant analysis was the main tool used in this study to select the set of explanatory variables to be included in the binary logit model. Tests of spatial autocorrelation were undertaken using the local version of Moran's *I* test. The results showed that in general positive spatial autocorrelation was present. So census tracts with similar WNV cases were adjacent to each other. Their findings show that each of the potential risk factors included in the logit model were statistically significant in predicting WNV.

2.2.5 Literature Synthesis

The literature review revealed several important facts. First, the role of climatic and geographic factors in amplification of the disease seems to be well-addressed. These are important determinants of the geographical distribution of WNV. However, their presence alone may not be sufficient to explain the amplification of the disease in California and Colorado. Second, the impact of factors of supreme interest in this study, income and home foreclosures have not been adequately addressed. The motivation for this study is to contribute to the literature by filling the void. Third, different methodological approaches have been proposed to examine the important risk factors that aid in WNV amplification. These approaches did not take into account the count and panel nature of the data and did not adequately address spatial autocorrelation in human WNV. The econometric model proposed to study the determinants of WNV amplification in California and Colorado corrects for these shortcomings. Clearly, there is no single econometric model which is comprehensive enough to incorporate all the potential risk factors associated with WNV amplification and transmission. However, it can be argued that the econometric model proposed in this essay offers a better explanation of the relationship between human WNV incidence and the economic factors identified in this study than the models used in previous studies.

2.3 Transmission Cycle of West Nile Virus

The transmission of West Nile from mosquitoes to humans involves three organisms namely a vector, host and infectious agent. In epidemiology, a vector is an insect or any living carrier that transmits an infectious agent or pathogen (Roberts et al. 2008). In the US, common vectors that transmit WNV to humans or other animals are mosquitoes in

the Genus *Culex*. A vector is not only a crucial aspect of some parasite's life cycles, but also transmits the parasite directly to subsequent hosts. A host is an organism upon which another organism lives as a parasite (McMichael et al. 1996). A reservoir host is a host in which viable infectious agents remain and from which infection of individuals may occur. Turell et al. (2002) explain that in the case of WNV the primary transmission cycle involves avian hosts. Humans and horses are considered to be accidental or dead-end hosts and develop very low levels of viremia which is not sufficient to infect mosquitoes (Hayes et al. 2005). An infectious agent or a pathogen or parasite is a biological organism that causes diseases. In the case of WNV the pathogen is a virus. The manner in which vector-borne diseases are transmitted to the human population depends on the characters and requirements of three living organisms – the pathologic agent, vector and human host and reservoir hosts. Typically, the vector injects the infectious agent into the blood stream of the host when it feeds on the blood of the host. For example, the female *Culex* mosquito (the vector for WNV), inserts its proboscis under the skin and feeds on its host blood. The parasites the mosquito carries, located in its salivary glands, are transmitted directed into the blood stream of the host.

2.4 Economics of West Nile Virus Control

The economic aspect of disease control is important in most epidemiological studies. In response to the rapid increase in WNV infections throughout the US, there has been a commensurate increase in the number of mosquito vector surveillance and control programs. These mosquito and surveillance programs are costly and time consuming. Economic information on costs and benefits assist policy makers in reinforcing and maintaining prevention and control measures, overhauling current surveillance systems,

and initiating other alternative policy interventions (Zohrabian et al. 2004). In the absence of a vaccine, surveillance of WNV is essential because it is the primary means of monitoring seasonal outbreaks of the disease so that prevention and control programs can be introduced. The California Department of Public Health (CDPH), the Mosquito and Vector Control Association of California (MVCAC), the University of California at Davis, and the Colorado Department of Public Health and Environment (CDPHE) undertake surveillance programs in California and Colorado. The cost to these agencies stem from mosquito abatement (including larval control), surveillance and education, communication services, and veterinary diagnostics and entomological services. According to the CDC (2011), the objectives of surveillance and monitoring for WNV include: (1) To evaluate the public health consequences of the disease; (2) Assess the need for public health intervention programs; (3) Identify potential risk factors responsible for transmission of the disease; (4) Determine which populations can be classified as high risk; and (5) identify which regions are in urgent need of interventions.

2.4.1 Economics of Vector Control

The study of dynamic economic control of vector-borne diseases has generated a lot of interest in recent years. The economic benefit of vector control derives from reducing vector longevity and thereby reducing disease transmission. The formal approach in economics is to examine the vector control problem within a capital-theoretic framework where the vector population is treated as a state variable. The objective of such studies is to find the optimal insecticide application which is then subsequently used to find the optimal path of the vector population.

Gersovitz and Hammer (2004) studied the control of vector-borne diseases using the susceptible-infections-susceptible (SIS) model. They specify equations explaining the evolution of the number of infected people, the number of infected mosquitoes, and the total population of mosquito vector. Their research examined the effects of various policy interventions (bed nets, provision of prophylactic drugs, insecticide spraying and provision of drugs that hasten recovery) on controlling the vector population. The objective of the decision maker is to choose policy intervention parameters that maximize a net benefit function over time. Policies for which the marginal benefits of reducing the vector population exceed the marginal costs may be deemed feasible. Brown et al. (2009) examined the optimal control of the mosquito vector with respect to malaria. To begin with, they specify equations that describe the way malaria prevalence and susceptible allele evolve over time. Next they specify a function that represents the total cost of infections and indoor residual spraying (IRS). The choice variable of the policymaker is the fraction of households covered by IRS. The policymaker's objective is to minimize the total present value of costs by choosing the optimal IRS policy.

2.4.2 Economic Losses Associated with WNV

The economic losses that accrue to individuals from a WNV infection include both medical and non-medical costs (Zohrabian et al. 2004). Medical costs include both inpatient and outpatient treatment costs—pharmacy/medical supplies, diagnostics, room and board, surgical services, intensive care and outpatient rehabilitation facilities and equipment. Non-medical costs include productivity losses, caused by illness and premature death, and transportation to visit a health care provider. Individuals or

households also incur significant cost on prevention measures such as insecticide spraying and using treated nets.

2.5 Data Sources

Data used in this study were collected from several sources. Data on temperature ($TEMP$), precipitation ($PRECIP$) and the drought index ($PDSI$) were collated from the National Climatic Data Center (NCDC 2011). County-level climatic data was not readily available so information from weather stations in each county was used to calculate annual climate data. An arithmetic average was used to calculate the climatic variables in counties with several weather stations. Information on human WNV (HV) and the number of positive mosquito pools ($MOSQUITO$) were acquired from the Centers for Disease Control (CDC 2011). Data on income ($INCOME$) and population density ($POPDENSE$) were taken from the US Department of Commerce/ Census Bureau (USDOC 2011). Data on home foreclosures ($FORCLOSE$) was acquired from Data Quick News and the Colorado Department of Local Affairs. Tables 2-1 and 2-2 contain the descriptive statistic of each of these variables for each state. Three trends are obvious from a detailed analysis of the data. First, they show that the data on WNV cases are highly asymmetrical in nature. Clearly, there is evidence of over-dispersion in the data because the variances are larger than the means. Second, the mean of WNV counts in Colorado is about 1.5 times that of California. Third, the values of the economic and biological variables for California are much larger than those for Colorado (Figures 2-3 and 2-4). The spatial distribution of human WNV cases and the key economic and biological variables for California and Colorado are shown in Figures 2-5 to 2-8 and Figures 2-9 to 2-12 respectively.

Table 2-1: Factors Used to Assess Risk of Human WNV in California

Variable	Description	Mean	S.D.	Min	Max
HV^{CDC}	human WNV case counts per county per year	8.355	29.530	0	331
$INCOME^{CB}$	Natural log of median household income ($)	10.750	.246	10.258	11.341
$FORCLOSE^{DQ}$	Natural log of home foreclosures plus 1(in thousands of houses)	4.004	3.668	0	10.603
$MOSQUITO^{CDC}$	Number of positive mosquito pools	14.679	43.901	0	408
$POPDENSE^{CB}$	Natural log of population density	4.508	1.993	.412	9.73
$PRECIP^{NCDC}$	Mean annual precipitation in inches	2.184	1.015	.270	4.283
$TEMP^{NCDC}$	Mean annual temperature in Fahrenheit	57.912	4.029	44.733	66.4
$PDSI^{NCDC}$	Palmer Drought Severity Index ranges from 0 to -5	-.241	2.053	-5.040	4.215
D2004	A dummy variable that equals one if year is 2004 and zero otherwise	.2	.400	0	1
D2005	A dummy variable that equals one if year is 2005 and zero otherwise	.2	.400	0	1
D2006	A dummy variable that equals one if year is 2006 and zero otherwise	.2	.400	0	1
D2007	A dummy variable that equals one if year is 2007 and zero otherwise	.2	.400	0	1

Data Sources: ^{CDC}-Centers for Disease Control (2011), ^{NCDC}-National Climatic Data Center (2011), ^{CB}-US Census Bureau, ^{DQ}-Data Quick News

Table 2-2: Factors Used to Assess Risk of Human WNV in Colorado

Variable	Description	Mean	S.D.	Min	Max
HV^{CDC}	Human WNV case counts per county per year	13.365	51.777	0	546
$INCOME^{CB}$	Natural log of median household income ($)	10.637	.288	9.986	11.458
$FORCLOSE^{COLA}$	Natural log of home foreclosures plus 1(thousands of houses)	2.903	2.669	0	8.975
$MOSQUITO^{CDC}$	Number of positive mosquito pools	5.593	21.051	0	247
$POPDENSE^{CB}$	Natural log of population density	2.509	1.797	-.368	8.227
$PRECIP^{NCDC}$	Mean annual precipitation in inches	1.337	.185	.803	1.62
$TEMP^{NCDC}$	Mean annual temperature in Fahrenheit	46.816	2.760	40.225	51.666
$PDSI^{NCDC}$	Palmer Drought Severity Index ranges from 0 to -5	-.772	2.204	-5.705	3.005
D2004	A dummy variable that equals one if year is 2004 and zero otherwise	.2	.400	0	1
D2005	A dummy variable that equals one if year is 2005 and zero otherwise	.2	.400	0	1
D2006	A dummy variable that equals one if year is 2006 and zero otherwise	.2	.400	0	1
D2007	A dummy variable that equals one if year is 2007 and zero otherwise	.2	.400	0	1

Data Sources: ^{CDC}-Centers for Disease Control (2011), ^{NCDC}-National Climatic Data Center (2011), ^{CB}-US Census Bureau, ^{COLA}-Colorado Department Of Local Affairs

Figure 2-3: Comparison of Economic Risk Factors (2003-2007)

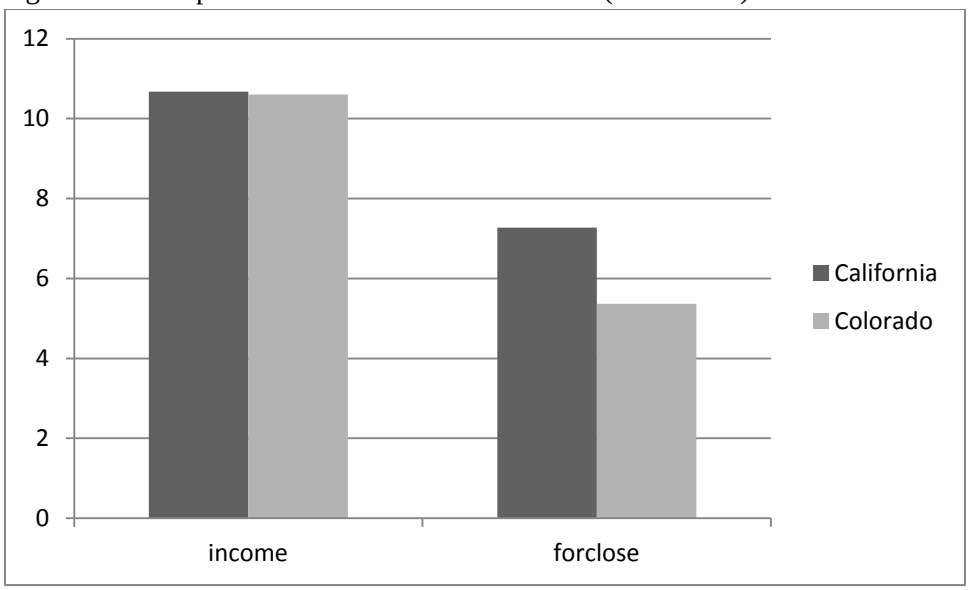
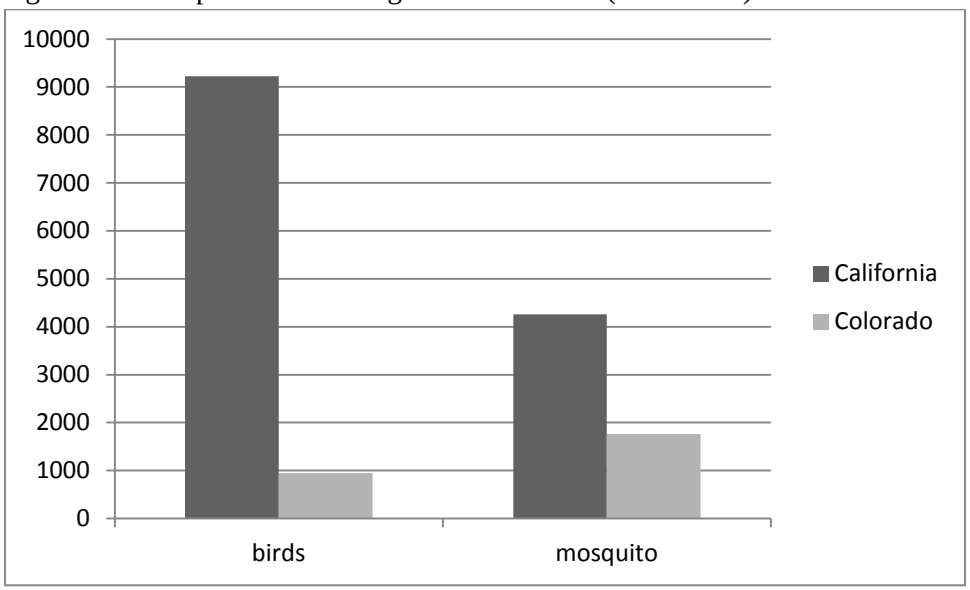


Figure 2-4: Comparison of Biological Risk Factors (2003-2007)



2.6 Econometric Modeling

The choice of the econometric model was motivated by the need to correct for the panel nature of the data. In dealing with the human WNV data, it is natural to model the human WNV case counts using a negative binomial model. However, because the data contains county-specific effects which may vary, it makes econometric sense to introduce random county specific effects into the relationship.

2.6.1 Random Effects Negative Binomial (RENB) Model

The specification of a RENB model is theoretically appealing because it can account for dependence that can exist between WNV in counties through the use of a county-specific random heterogeneity term. This random term could capture hidden or omitted variables such as quality of surveillance or early warning systems, response time in removing dead birds and spraying of mosquito pools and other monitoring technology available.

Hausman, Hall and Griliches (1984) extend the negative binomial model to accommodate random effects. The conditional negative binomial distribution is given as:

$$\Pr[Y_{it} = y_{it} \backslash \delta_i] = \frac{\Gamma(y_{it} + \lambda_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)} \left(\frac{\delta_i}{1 + \delta_i}\right)^{\lambda_{it}} \left(\frac{1}{1 + \delta_i}\right)^{y_{it}} \quad (2.1),$$

where $i(= 1, \ldots . N)$ indexes counties, $t(1, \ldots \ldots T)$ indexes a given time period, $\lambda_{it} = \exp(X_{it}\beta)$, $\delta_i$is the dispersion parameter and y_{it} is the number of WNV counts in any county in a given year. The moments (i.e. mean and variance) of this conditional distribution are:

$$E[Y_{it} \backslash \delta_i] = \lambda_{it}\delta_i \quad (2.2)$$

$$Var[Y_{it} \backslash \delta_i] = \frac{\lambda_{it}(1+\delta_i)}{{\delta_i}^2} \tag{2.3}$$

In the analysis of panel data pertaining to negative binomial models, the individual effects relate to the distribution of the dispersion parameter, not the mean rate. Following Hausman, Hall and Griliches (1984), the expected number of WNV counts can be written as:

$$\widetilde{\lambda_{\imath t}} = \lambda_{it}\delta_i = \exp(\boldsymbol{X}_{it}\beta + u_i) \tag{2.4}$$

where u_i is the random effects across counties and $\exp(u_i) \sim Gamma\left(\frac{1}{\delta}, \delta\right)$, $\boldsymbol{X}_{it}$ is a vector of covariates and β is the coefficient vector to be estimated. The RENB model accounts for variation over time by allowing the dispersion parameter to follow a beta distribution such that:

$$\frac{1}{1+\delta_i} \sim Beta(r, s) \tag{2.5}$$

with mean and variance respectively as:

$$\frac{r}{r+s}, \frac{rs}{(r+s)^2 + (r+s+1)}$$

r and s are parameters of the beta distribution. The joint probability density function can be expressed as:

$$\Pr(Y_{i1} = y_{i1}, \ldots\ldots, Y_{iT} = y_{iT}]$$

$$= \frac{\Gamma(r+s)\Gamma(r+\sum_t \lambda_{it})\Gamma(s+\sum_t y_{it})}{\Gamma(r)\Gamma(s)\Gamma(\mathrm{r+s}+\sum_{\mathrm{t}}\lambda_{\mathrm{it}}+\sum_{\mathrm{t}}\mathrm{y_{it}})} \times \prod_t^T \frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)} \tag{2.6}$$

2.6.2 Empirical Model Specification

The general econometric model for studying the determinants of WNV can be specified as:

$$HV_{it} = f(exp\left({X_{it}}^{E}, {X_{it}}^{BIO}, X^{POP}, {X_{it}}^{CL}, u_i \right)) \tag{2.7}$$

The economic variables (income and home foreclosures) are represented by the vector X^E, the biological variable (mosquito pools) is represented by the vector X^{BIO}, the climatic variables (precipitation, temperature and drought) are represented by the vector X^{CL} , population density is represented by the vector $X^{POP}$and $u_i$is the random effects term.

The econometric model specified uses human WNV (HV) as the dependent variable and home foreclosures ($FORCLOSE$), income ($INCOME$), mosquito pools ($MOSQUITO$), and population density ($POPDENSE$). Additional covariates,Z, are also controlled for. The following random effects negative binomial model is proposed to examine the determinants of human WNV transmission:

$$\begin{aligned} HV_{it} = \exp(\beta_0 + \beta_1 INCOME_{it} + \beta_2 FORCLOSE_{it} + \beta_3 MOSQUITO_{it} \\ + \beta_4 POPDENSE_{it} \quad + {Z_{it}}'\theta + u_i) \end{aligned} \tag{2.8}$$

$i(1, \ldots . . N)$ indexes counties, t $(1, \ldots \ldots T)$ indexes time, and $u_i$is the random effects term. The control vector Z' comprises climatic variables including annual precipitation ($PRECIP$), annual temperature ($TEMP$), annual drought ($PDSI$), and time fixed effects (dummies) for 2004 to 2007 ($D2004$, $D2005$, $D2006$ and $D2007$). The choice of controls is driven by a mix of theory and empirical findings of numerous WNV studies.

2.6.2.1 Role of Income

A direct link between income and human WNV is relatively easy to establish. Individuals earn income in the form of wages and salaries by supplying their services on the labor market. Several authors generally agree that income is a major determinant of health (Nazroo 2003 and Ettner 1996). However, it is very difficult to establish a direction of causality between income and health. Notwithstanding these issues, economists generally agree that income is positively related to health. It is therefore plausible to argue that the absence of health (disease/illness) is negatively related to income. According to the Center on Social Disparities in Health (CSDH 2010), income can directly affect an individual's medical care, housing options, nutrition and physical activity options, education, neighborhood conditions, and stress level. Residents of low-income communities suffer disproportionately from adverse environmental factors, lack of maintenance of property (houses), and poor drainage (Harrigan et al. 2010). Another reason to account for income is that it could serve as a proxy for surveillance efforts (Brown et al. 2008). Previous studies on animal rabies used county-level per capita income as a surrogate for investment in surveillance and laboratory testing (Childs et al. 2007). Poverty has been linked to income as well because malnutrition can lead to a compromised immune system (Farmer 1999). In most cases it is reasonable to hypothesize that income is negatively related to disease transmission (Messina et al. 2011; Linard et al. 2007). The variable $INCOME$ is included in the model to account for this effect.

2.6.2.2. Role of Home Foreclosures

The relationship between home foreclosures and WNV is not conclusive and is still the subject of considerable debate. In economics, homes are considered an investment, and individuals and families buy homes to get a favorable return on their investment in the future. Home foreclosures were considered a potential risk factor in the transmission of human WNV during the housing crisis that began in 2004 and culminated in the financial crisis of 2007 (Reisen et al. 2008; Harrigan et al. 2010). The Federal Reserve (2008) stated that the hardest hit states were California, Arizona, Nevada, Colorado, Florida and some New England states. The argument was that the economic downturn and accompanying housing market crisis which began in 2004 adversely affected the economy of the US in general, and particularly the states of California and Colorado. The combined effect was a growing number of neglected swimming pools on foreclosed properties in particular across Southern California. This was attributed to the high number of home foreclosures and delinquent mortgages because home owners could not afford their mortgages. Most of these neglected swimming pools collected small pockets of water and served as breeding grounds for mosquitoes (Reisen et al. 2008; Harrigan et al. 2010). According to Reisen et al. (2008), WNV cases escalated by 276% in Kern County in the Summer of 2007. This was blamed on the sluggish economy that obtained in 2007, and the accompanying slowdown in the housing market. As a result residents could not afford their mortgage payments leading to an increase in foreclosures. The result was an increase in WNV because abandoned swimming pools collected pockets of water that served as breeding grounds for the mosquito vector and other pests. Harrigan et al. (2010) found a similar association between WNV incidence in Orange County and

neglected swimming pools on foreclosed residential properties. Based on these studies by Harrigan et al. (2010) and Reisen at al. (2008), it is reasonable to hypothesize a positive relationship between home foreclosures and the transmission of WNV. In order to further verify and clarify the role of home foreclosures on WNV transmission, the variable $FORCLOSE$ is included in the model.

2.6.2.3 Role of Mosquitoes

The general consensus is that the presence of the Culex mosquito vector is necessary for the outbreak of WNV (Messina et al. 2011). Within the *Culex* genus, there are certain species local to specific regions of the US. In western US, the *tarsalis* and *quinquefasciatus* species are found in abundance (DeGroote et al. 2008). The *Culex tarsalis* is found in both Colorado and California, but the *Culex quinquefasciatus* is mostly restricted to California. The habitat location is different for each species of the *Culex* genus. The *Culex tarsalis* is the predominant vector in rural settings (Epstein et al. 2001), although Reisen et al. (2008) reported their presence in urban areas of California. Their survival depends on the presence of surface water created by rainfall and irrigation projects. *Culex tarsalis* is mainly responsible for transmitting the virus to humans in Colorado (CDPH 2002). They are known to feed at dawn and dusk. During the day, they usually rest in dark and secluded areas such as under the roof eaves, storm drains, porches and tall shrubs. Their habitat includes areas characterized by standing water such as irrigated fields, old tires, flower pots, hoof prints or small pockets of water on the ground. *Culex pipiens* is mostly found in urban areas and lays its eggs in stagnant water. This includes areas with poor drainage, sewage treatment lagoons/pools, urban catch basins, and containers found on the compounds of many low-cost houses (Savage and

Miller 1995; Huhn et al. 2005). Mosquitoes get infected by feeding on a bird with the virus in its blood stream. Mosquitoes then spread the virus to new hosts by biting another bird or mammal. The biological evidence from the foregoing discussion suggests that the number of mosquitoes breeding sites may be positively related to human WNV. The variable $MOSQUITO$ is included to control for vector abundance.

2.6.2.4 Role of Population Density

Population variables are included in this and many studies in epidemiology because they reflect the degree of exposure to a particular disease. In particular, the population can be used to calculate or measure the relative risk of a disease in each county. A limited number of studies have explicitly modeled the effect of population density on the transmission of WNV in California and Colorado (Harrigan et al. 2010). A high population density might increase the likelihood of vector abundance and the probability that an individual is bitten by an infected mosquito. In this regard, Trawinski and Mackay (2010) hypothesize that vector (mosquito) abundance is positively correlated with population density. Taylor et al. (1956) also found that areas with higher cases of WNV and mosquito abundance in Egypt also had higher population densities. From the foregoing discussion it is reasonable to hypothesize a positive relationship between human WNV transmission and population density. The variable $POPDENSE$ is included as a covariate to account for this effect. The $POPDENSE$ variable makes the use of an exposure/offset variable in the RENB regressions redundant. This is because the log of population, which is a component of $POPDENSE$, serves that purpose.

2.6.2.5 Role of Climatic Factors

Precipitation plays a significant role in the life cycle of insects such as mosquitoes. The presence or absence of breeding sites may depend on the amount of precipitation in an area. Guber et al. (2001) enumerates a range of possible channels through which rainfall can impact the diffusion of vector-borne diseases. These methods include: increased surface water can provide breeding sites for vectors; low rainfall can also increase breeding sites by slowing river flow; increased rainfall can increase vegetation and allow expansion in population of vertebrate host; flooding may eliminate habitat for both vectors and vertebrate and flooding may force vertebrate hosts into closer contact with humans. Some studies (Wimberly 2008; Winters et al. 2008) found evidence that precipitation had a negative impact on WNV, while Mongoh et al. (2006) and Messina et al. (2011) found evidence to the contrary. The variable $PRECIP$ is included to control for the effect of precipitation. Changes in temperature have been known to affect the distribution of many arthropod vectors because the minimum and maximum temperatures affect their geographical distribution. Hunter (2003) posits that temperatures can affect both the distribution of the vector and the effectiveness of pathogen transmission through the vector. Gubler et al. (2001) enumerates a range of possible mechanisms by which changes in temperature can greatly impact the risk of transmission of vector-borne diseases. These effects include increase or decrease in survival of vector, changes in rate of vector population growth, changes in feeding behavior, changes in susceptibility of vector to pathogens, changes in incubation period of pathogen, changes in seasonality of vector activity, and changes in seasonality of pathogen transmission. Thus, it could be hypothesized that the effect of temperature on WNV can be either negative or positive. A

number of empirical studies found a positive relationship between temperature and WNV (Winters et al. 2008; Mongoh et al. 2006; Gibbs et al. 2006). The variable $TEMP$ is included to control for this effect. Drought may also impact the incidence of human WNV. A number of authors suggest that drier conditions increase the contact between vectors and reservoir hosts thereby leading to increased WNV transmission during periods of drought (Gibbs et al. 2006; Brown et al. 2011; DeGroote 2008). The variable $PDSI$ is included to control for the effect of drought.

2.6.2.6 Role of Time Fixed Effects

It is sometimes assumed that regression coefficients are the same in subsets of the data. This assumption is occasionally violated because these subsets of the data are structurally different. Events that occur in each year may have unequal impact on counties. Because of the panel structure of the data, time effects are controlled for. The time effects are captured by the dummy variables$D2004$, $D2005$, $D2006$ and $D2007$. They could also represent technological changes or county policies that may affect WNV over time.

2.6.3 Statement of Hypotheses

The following hypotheses regarding the economic (income and home foreclosures) and biological (mosquito pools) variables are tested. The former is the primary hypothesis, while the latter is the secondary hypothesis.

Hypothesis 1: $\beta_{INCOME} < 0$

This hypothesis states that income will have a significant negative impact on the incidence of WNV. Under this hypothesis the regression coefficient on $INCOME$ is expected to be negative.

Hypothesis 2: $\beta_{FORCLOSE} > 0$

This hypothesis states that home foreclosures will have a positive effect on WNV transmission. Under this hypothesis the regression coefficient on $FORCLOSE$ is expected to be positive.

Hypothesis 3: $\beta_{MOSQUITO} > 0$

This hypothesis states that the number of mosquito breeding sites will have a significant positive impact on the incidence of WNV. Under this hypothesis the regression coefficient on $MOSQUITO$ is expected to be positive. Summaries of all three hypotheses for each of the estimated models are provided in Tables 2-16 and 2-17.

2.6.4 Endogeneity

The model specified in equation (2.8) is appropriate only if all the explanatory variables are exogenous. This is potentially a problem due to the possibility of simultaneity bias caused by endogeneity between WNV counts and income and home foreclosures. Put simply, income and home foreclosures could be determined jointly with human WNV. Ettner (1996) and Biddle et al. (2007) argue that potential endogeneity may exist between income and human WNV. They further explained that finding instruments for income to overcome the endogeneity problem could prove to be a challenge. In other words, it is generally difficult to find a set of variables that are highly correlated with income, but not directly associated with health. It can be argued straightforwardly that the benefits that residents of a county derive from the consumption of goods and services depend on their state of health. At the same time only healthier people can supply their services on the labor market and earn income. Foreclosures, on the other hand, depend on a variety of

socioeconomic variables that directly impacts the transmission of human WNV. The impacts of these variables on WNV are dependent on other socioeconomic variables as well as policy relating to health and welfare systems. To correct this problem both income and home foreclosures are instrumentalized. In other words, we search for instruments highly correlated with income and home foreclosures, but uncorrelated with human WNV and then implement a two-step instrumental variable procedure. The array of economic, biological, climatic and geographic variables used as instruments include: mosquito pools ($MOSQUITO$), birds ($BIRD$), equine case ($EQUINE$), population density ($POPDENSE$), precipitation ($PRECIP$), temperature ($TEMP$), drought ($PDSI$), education ($EDUCATE$), the unemployment rate ($UNRATE$), urbanization ($URBAN$), the number of airports ($AIRPORT$), elevation ($ELEVATE$), number of roads ($ROAD$), the log of population density ($POPDENSE$) and the time fixed effects ($D2004, D2005, D2006$ and $D2007$). In the first step, $INCOME$ and $FORCLOSE$ are regressed on the instruments described above to obtain the predicted values $INCHAT$ and $FORHAT$ respectively. In the second step, these predicted values are used to replace $INCOME$ and $FORCLOSE$ in equation (2.8) to solve the endogeneity problem.

2.6.4.1 Specification Tests for Endogeneity and Validity of the Instruments

To test for endogeneity of income and home foreclosures, a version of the specification test due to Hausman (1983) was carried out. This test is asymptotically equivalent to the original Hausman test and can be generalized to non-linear models such as generalized linear models (GLM). The first step of this test involves obtaining the residuals of a RENB regression of income and home foreclosures on all the exogenous variables. In the second step, these residuals are used in a RENB regression to test the null hypotheses that

the coefficients on the included residuals are jointly zero. The computed chi-square statistics for California are 14.57, 15.02, 15.05 and 15.06 for models 1 to 4 respectively. The computed chi-square statistics for Colorado are 100.95, 114.95, 115.02 and 116.06 for models 1 to 4 respectively. On the basis of these statistics, the hypotheses are rejected at the 10% level of significance, thus income and home foreclosures are endogenous.

Test for validity of the instruments (over-identification test) was conducted using Sargan's statistic. Identification will be achieved if the instruments are uncorrelated with the error term, but correlated with income and home foreclosures. To implement this test, the residuals from the instrumental variable RENB estimation were regressed on the instruments. The joint null hypothesis of Sargan's is that the excluded instruments are valid so they are rightly excluded (i.e. uncorrelated with the error terms) from the estimated equation. The test statistic calculated is nR^2 which follows a chi-square distribution with the degrees of freedom equal to the over-identifying restrictions. The calculated statistics for California are 0.90, 0.95, 0.96 and 0.98 for models 1 to 4 respectively. The calculated statistics for Colorado are 0.50, 0.90, 0.95 and 0.96 for models 1 to 4 respectively. On the basis of these statistics, the null hypotheses cannot be rejected at the 10% level of significance, so the excluded instruments are valid.

2.6.5 Spatial Autocorrelation

Shope (1991) argues that climate change impacts the spatial distribution of vector-borne diseases because they affect: the current geographic distribution of the disease; the scope of non-human hosts and reservoir hosts; temperature-related vector and parasite development, and adaptive processes pertaining to reservoir and parasite interactions;

capacity for migration of vectors and parasites; and the current seasonality of transmission. Winds contribute significantly to the passive dispersal of flying insects and mosquitoes, therefore wind speed and wind direction will affect the distribution of the vector. Hack (1955) states that insect vectors such as the *Culex* mosquitoes and black flies have been known to be dispersed hundreds of kilometers from their original location.

In this study, spatial autocorrelation may be caused by the fact that the geographical distribution of WNV is determined by factors that transcend county boundaries. Several authors have argued that these factors pertain to climate, environment, vegetation, hydrology and human activities (Kiszewski et al. 2004; Sattler et al. 2005). Spatial autocorrelation occurs when events in one location are dependent on events in another location (spatial externality/spill over). In particular, WNV infections in one county may spill over into an adjoining county. In other words, counties closer to each other may display similar infections rates than those farther apart. Anselin (2001) explains that spatial autocorrelation exists because residuals from different geographical units which are spatially correlated contain common omitted environmental, physical or economic factors. Anselin (2002) alludes to the fact that in the case of linear models, spatial lag and spatial error models can be used to correct for endogenous and exogenous spatial autocorrelation. However, these models cannot be readily applied to generalized linear models (GLM) such as RENB models that rely on maximum likelihood estimation. Equation (2.8) does not address potential spatial autocorrelation that could arise when geographic data is used. Spatial autocorrelation emerges because human WNV in one county could be determined not only by economic, environmental, demographic, climatic

and geographic variables, but also by human WNV in adjoining counties. This dissertation employs two methods for correcting for spatial autocorrelation—the spatial lag method and the spatial filtering technique. The former method is a parametric method because it incorporates the structure of spatial autocorrelation into the model. It is also the preferred method employed by economists to fix the spatial dependency problem. The latter method is more of a geo-statistical technique and is viewed as a non-parametric method for correcting spatial autocorrelation.

2.6.5.1 Spatial Lag Method

Tackling the problem of spatial autocorrelation in the dependent variable within a GLM framework could be quite challenging because this issue is apparently under-studied in the econometrics literature (Lambert et al. 2010). An innovative and instructive approach was adopted in this study to address this problem. The method employed in this paper is straight forward and is motivated by the spatial lag model (simultaneous autoregressive lagged response model) specification for correcting spatial autocorrelation in the dependent variable in linear models. Because of the presence of spatial autocorrelation in human WNV counts, equation (2.8) can be augmented with a spatially-lagged dependent variable on the right hand side. The spatial lag version of equation (2.8) is:

$$HV_{it} = exp(\beta_0 + \beta_1 INCOME_{it} + \beta_2 FORCLOSE_{it} + \beta_3 MOSQUITO_{it} + \beta_4 POPDENSE_{it} + \rho SLAG_{it} + Z'_{it}\,\theta + u_i) \quad (2.9)$$

where ρ is the autoregressive coefficient that corresponds to the spatially-lagged dependent variable ($SLAG$) derived as $SLAG_{it} = W \times HV_{it}$ and W is an $n \times n$ spatial weights matrix that represents the contiguity relationship among counties. It represents

the effect of WNV in neighboring counties on the average WNV counts in adjoining counties. Each entry in W contains the length of a given county's border that is shared by another county. The presence of a lagged dependent variable as a covariate means that estimated coefficients would be biased and inconsistent due to endogeneity. For a particular county, this model is represented by:

$$HV_1 = \exp(\rho[w_{11}HV_1 + w_{12}HV_2 + w_{13}HV_3 + \cdots + w_{1n}HV_n] + X_1\beta + u_1) \qquad (2.10)$$

The method employed in this paper to correct for spatial autocorrelation is a two-step instrumental variable approach which has been applied by Linard et al. (2007), Lambert et al. (2010) and Brown et al. (2011) in generalized linear models. In the first the step, the spatially-lagged dependent variable is regressed on a number of instruments using a RENB model to obtain the linear prediction (predicted values) of the dependent variable in adjoining spatial units. In the second stage, this linear prediction is included as an additional covariate in the RENB regressions. This linear prediction is used as a proxy for the spatial effect of one county on a neighboring county. This technique indirectly corrects for spatial autocorrelation in the dependent variable. To implement this method, $SLAG$ is regressed on the instruments described above to obtain the predicted values, $SLAGHAT$, in the first stage. In the second step, these predicted values are used to replace $SLAG$ in equation (2.9) to solve the endogeneity problem.

2.6.5.2 Spatial Filtering Technique

The second technique employed to correct for spatial autocorrelation is spatial filtering which is based on the spatial eigenvector mapping method. The rationale behind this method is that the configuration of spatial data points on a map, are reflected in covariates that capture spatial effects at different spatial scales. This is a non-parametric

method for correcting spatial autocorrelation (Patuelli et al. 2009). Spatial filtering techniques have been developed and implemented by several authors (Griffith 1981; Haining 1991; Getis and Griffith 2002; Tiefelsdorf and Griffith 2006). One merit of these techniques is that the variable of interest can be partitioned into spatial and non-spatial components which could then be used as covariates within a generalized linear model (GLM) framework. The spatial filtering technique employed in this paper is due to Griffith (2000). It is based on the formula for computing the Moran's *I* statistic and eigenvector decomposition of a modified or centered spatial weights matrix. This transformed spatial weights matrix reflects the latent spatial autocorrelation inherent in the dependent variable (Griffith 2005). The eigenvectors extracted are by definition orthogonal and uncorrelated to each other. The idea is to extract those eigenvectors such that consecutive values of Moran's *I* values are maximized so as to minimize spatial autocorrelation in the model residuals. In other words, eigenvectors are judiciously selected to minimize spatial autocorrelation in the model residuals. The first eigenvector has the greatest value of Moran's *I*, the second eigenvector has the second highest value of Moran's I and so on. The steps involved in deriving the spatial filter are as follows:

(1) It begins with constructing an $n \times n$ spatial weights matrix W that contain elements, w_{ij}. The spatial weights matrix selected for this dissertation, and the one commonly used in epidemiological studies, emphasizes geographical distances. It is based on the length of a given county's border shared by an adjoining county. The contiguity matrix is row-standardized so that the elements in each row sum to unity.

(2) Transform the spatial weights matrix based on Moran's *I* as follows:

$$\left(I - \frac{11'}{n}\right) W \left(I - \frac{11'}{n}\right) \tag{2.11}$$

where I is an $n \times n$ identity matrix, 1 is a column vector of ones and n is the number of contiguous spatial units. This transformation is mathematically equivalent to the numerator in the formula for Moran's I value:

$$I = \frac{n \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_i \sum_j w_{ij}\right) \sum_i (x_i - \bar{x})^2} \tag{2.12}$$

where x_i is the value of the dependent variable at spatial unit i, $\bar{x}$ is the average and w_{ij} is an element of the spatial weights matrix.

(3) Extract the eigenvectors of the transformed spatial weights matrix.

(4) Select k (less than n) candidate eigenvectors to be included in the spatial filter based on a criterion that uses Moran's *I*. Include eigenvectors in the filter as long as

$$\frac{I_i}{\max(I_i)} > 0.25 \tag{2.13}$$

A total of 19 candidate eigenvectors were extracted for California, while a total of 20 eigenvectors were extracted for Colorado.

(5) Select the subset of all candidate eigenvectors to be used as components of the spatial filter ($SFILTER$) or spatial covariates. The optimal number of eigenvectors was chosen based the stepwise backward elimination model selection procedure using a 5% level of significance. A full list of the selected eigenvectors is provided in Tables 2-14 and 2-15.

Griffith (2003) contends that the stepwise selection procedure is appropriate because it maximizes the percentage of variance in the geo-referenced dependent variable accounted

for by the selected eigenvectors. He further asserts that restricting attention to those eigenvectors with a Moran's ratios in excess of 0.25 ensures that each eigenvector accounts for at least some degree of spatial autocorrelation in the geo-referenced dependent variable. The combined use of these criteria produces regression residuals devoid of spatial autocorrelation.

The components of $SFILTER$ are then used as additional covariates in equation (2.8) to correct for spatial autocorrelation in the dependent variable. The spatial filtering RENB model is specified as follows:

$$HV_{it} = exp(\beta_0 + \beta_1 INCOME_{it} + \beta_2 FORCLOSE_{it} + \beta_3 MOSQUITO_{it} + \beta_4 POPDENSE_{it} + SFILTER_i\gamma + Z'_{it}\,\theta + u_i) \quad (2.16)$$

where $SFILTER_i$ is an array of selected eigenvectors (spatial filter components), denoted by VEC, of the transformed spatial weights matrix for county i. The spatial filter can be perceived as a proxy for omitted or missing variables from the regression (Patuelli et al. 2009). Getis and Griffith (2002) contend that all the n eigenvectors extracted represent all the possible orthogonal map patterns. In other words, they represent a kaleidoscope of all possible map patterns. Specifically the first two principal eigenvectors extracted are often associated with North-South and East-West patterns respectively. Eigenvectors with intermediate values of Moran's *I* typically exhibit regional patterns, while eigenvectors with extremely low values of Moran's *I* are associated with local map patterns.

2.7 Estimation and Results

Estimation of the parameters of the RENB model was undertaken by the technique of maximum likelihood. The maximum likelihood associated with the RENB is as follows:

$$lnL = \sum_{i=1}^{n} w_i \left\{ ln\Gamma(r+s) + ln\Gamma\left(r + \sum_{k=1}^{n_i} \lambda_{ik}\right) + ln\Gamma\left(s + \sum_{k=1}^{n_i} y_{ik}\right) - ln\Gamma(r) - ln\Gamma(s) \right.$$

$$- ln\Gamma\left(r + s + \sum_{k=1}^{n_i} \lambda_{ik} + \sum_{k=1}^{n_i} y_{ik}\right)$$

$$\left. + \sum_{t=1}^{n_i} [ln\Gamma(\lambda_{it} + y_{it}) - ln\Gamma(\lambda_{it}) - ln\Gamma(y_{it} + 1)] \right\}$$

where w_i is the weight for the ith group and all other variables are as previously defined. All estimations were carried out in STATA. Pearson's correlation and variance inflation factors (VIF) were employed to ascertain the degree of multi-collinearity among the potential risk factors identified in this study. In conformity with tradition, only those risk factors whose correlations were not in excess of 0.8 were retained in all the models (Messina et al. 2011).

The econometric analysis estimates five different regressions using HV as the dependent variable- non-spatial RENB regression, spatial lag RENB regression, IV spatial lag RENB regression, spatial filtering RENB regression and IV spatial filtering RENB regression. Four models were estimated under each regression. The baseline model (Model 1) excludes climatic and biological variables. In Model 2, $PRECIP$ and $MOSQUITO$ are included. In Model 3, only the climatic variable $TEMP$ is included. Finally in Model 4, only the $PDSI$ climatic variable is included. The AIC and BIC, which are model selection criteria, reveal that the IV spatial filtering RENB regressions, which correct for both endogeneity and spatial autocorrelation, are the most preferred. A conservative significance level of 10% was chosen to verify that residuals in these

models display spatial independence. On this basis, we cannot reject the null hypothesis that the residuals are spatially independent. Because the IV regressions include generated regressors ($FORHAT$ and $INCHAT$), the estimated standard errors are not reliable. In order to obtain reliable and robust estimates, a bootstrap procedure was implemented. The number of replications was initially set at 100 and gradually increased to 1000. In all cases, the results were consistent and robust. Perhaps, it is worthwhile pointing out that the results of Models 2 to 4 could be used as a robustness check of the stated hypotheses. The results show that the random effects term, u_i, follows a beta distribution with shape parameter values r and s provided in each table. In fact, in all Tables $s > r$ and this implies that the distribution is not symmetric. Wackerly et al. (2002, p. 183) posit that if $s = r = 3$, the beta distribution looks like a normal distribution. If $s > r$, as is the case here, the distribution is left skewed and not normal.

2.7.1 Spatial Autocorrelation

To test for spatial autocorrelation a spatial weights/ contiguity matrix, W, must be constructed. Separate matrices had to be constructed for both California and Colorado because the two states are not contiguous. With the number of contiguous counties in California equal to 58, the dimensions of W for California are 58×58 with each element of W denoted by w_{ij}. With the number of contiguous counties in Colorado equal to 63, the dimensions of W for Colorado are 63×63 with each element of W denoted by w_{ij}. While there are several options[1] in designing W this study uses a type of the spatial

[1] Bailey and Gatrell (1995) suggest the simplest form of the spatial weights matrix: $w_{ij} = 1$ if geographical units (here states) *i* and *j* share a common border and $w_{ij} = 0$ otherwise. Another possibility to let each element of W be a distance decay function where the spatial weights matrix is constructed using a simple exponential decay model where $w_{ij}$is defined as $w_{ij} = \exp(-\lambda d_{ij})$ where d_{ij} is the Euclidean distance between the centroids of areas *i* and *j*, and λ is a constant to be estimated from the data.

weights matrix in which each cell contains the length of a given county's border that is shared by another county[2]. Moran's *I* (Moran 1948) test was undertaken to verify the presence/absence of spatial autocorrelation in West Nile virus counts in each county. A positive and significant z-value (low p-value) for Moran's *I* is indicative of positive spatial autocorrelation (Anselin 1983). The results of this test are presented in Table 2-3. It shows that the z-values for Moran's *I* are positive and significant, indicating the presence of positive spatial autocorrelation in the WNV counts.

Table 2-3: Global Spatial Autocorrelation Test on WNV counts

Year	Moran's *I* Test for California			Moran's *I* Test for Colorado		
	Moran's *I* statistic	z-value	p-value	Moran's *I* statistic	z-value	p-value
2003	0.056	3.431	0.000	0.435	5.224	0.000
2004	0.439	4.875	0.000	0.432	5.406	0.000
2005	0.205	2.446	0.007	0.182	3.214	0.001
2006	0.318	3.774	0.000	0.284	3.992	0.000
2007	0.203	2.519	0.006	0.290	3.498	0.000
All years	0.350	8.552	0.000	0.537	14.848	0.000

2.7.2 California Regression Results

2.7.2.1 Non-Spatial Random Effect Negative Binomial Regression

The results of this regression are contained in Table 2-4. In each of the models the coefficients on $INCOME$ are statistically significant at the 1% level, with values ranging from -1.734 to -2.181. The negative relationship is expected, suggesting that WNV prevalence is higher in counties with a lower median income. The coefficient on $MOSQUITO$ is positive and significant at the 1% level and its value is 0.010. This result

[2]This can be considered the Rook criterion (locations sharing a boundary) or the queen criterion (locations sharing a vertex).

suggests that a high number of mosquito breeding sites increases the prevalence of human WNV. The $POPDENSE$ coefficients are significant at the 1% level in all models. The $PDSI$ coefficient is negative and significant at the 1% level

2.7.2.2 Spatial Lag Random Effects Negative Binomial Regression

Table 2-5 reports the results of this regression. The coefficient on $FORCLOSE$ is significant in only Model 1 at the 10% level with a value of 0.084. The sign of the coefficient suggest a positive relationship between human WNV and home foreclosures. The coefficients on $INCOME$ are negative and significant at the 1% level in all models, with values ranging from -2.160 to -2.571. The coefficients on $MOSQUITO$ are significant at the 1% level in all models and range from 0.004 to 0.005. Thus human WNV incidence is higher in counties with a higher number of mosquito breeding sites. The coefficients on $POPDENSE$ are positive and significant at either the 1% level in all models, with values ranging from 0.314 to 0.339. This suggests that WNV prevalence is higher in counties with higher population densities.

2.7.2.3 IV Spatial Lag Random Effects Negative Binomial Regression

The results of this regression are presented in Table 2-6. The estimated coefficients on $FORHAT$ are positive and significant at either the 5% or 10% level in all models, with values ranging from 0.180 to 0.301. This result suggests that home foreclosures were a significant contributory factor in the high incidence of human WNV in California. The coefficients on $INCHAT$ are negative and statistically significant at either the 5% or 10% level in Models 1-3, with values ranging from -1.444 to -2.385. This result implies that a high median county income reduces the prevalence of human WNV. The coefficient on

$POPDENSE$ is positive and significant at the 10% level in only Model 1 with a value of 0.300. The coefficients on $SLAGHAT$ are positive and significant at either the 1% level in all models, with values ranging from 0.498 to 0.597. Hence human WNV prevalence in one county is dependent on WNV incidence in neighboring counties.

2.7.2.4 Spatial Filtering Random Effects Negative Binomial Regression

Table 2-7 presents the results of this regression. The coefficient on $FORCLOSE$ is positive and statistically significant at the 10% level in only Model 1 with a value of 0.091. The coefficients on $INCOME$ are statistically significant at the 1% or 5% level in Models 1-4, with values ranging from -2.004 to -2.875. The coefficients on $MOSQUITO$ are significant at the 1% level in all models, with values ranging from 0.010 to 0.011. The coefficients on $POPDENSE$ are statistically significant at the 1% level in all models, with values ranging from 0.296 to 0.411. Of all the $SFILTER$ components included in the regression, only the eigenvectors $VEC3$, $VEC16$, $VEC29$, $VEC33$ and $VEC34$ are statistically significant at the 1% , 5% or 10% level.

2.7.2.5 IV Spatial Filtering Random Effects Negative Binomial Regression

The results of this regression are presented in Table 2-8. The results of this regression are the most preferred because not only do they correct for both spatial autocorrelation and endogeneity, but they also have the lowest AIC values. The estimated coefficients on $FORHAT$ are positive and statistically significant at the 5% level in all models, with values ranging from 0.382 to 0.413. The expected positive relationship suggests that WNV prevalence is higher in counties with a higher number of foreclosed homes. The coefficients on $INCHAT$ are negative and significant at either the 1% or 10% level in l

Models 1-3, with values ranging from -1.206 to -4.183. This result provides evidence that median county income is an important factor in WNV transmission in California. The coefficient on the biological factor ($MOSQUITO$) is highly significant at either the 5% or 10% level in all models with a value of 0.005. This result provides evidence that WNV prevalence is higher in counties with a higher number of mosquito pools. The coefficient on $POPDENSE$ is significant the 5% level in Model 1 only. Only the eigenvectors $VEC4$, $VEC15$, $VEC16$ and $VEC16$ are significant at either the 1%, 5% or 10% level. Because this was the most preferred regression, supplementary regressions were run to check the robustness of the primary hypotheses. These additional regressions included an interaction term between $MOSQUITO$ and $POPDENSE$ as an additional covariate and the significance of the economic variables did not change. Thus, the results were robust using all four model specifications. The results of these supplementary regressions are presented in Appendix E.

2.7.3 Colorado Regression Results

2.7.3.1 Non-Spatial Random Effects Negative Binomial Regression

The results of this regression are presented in Table 2-9. The coefficient on $MOSQUITO$ is positive and significant at the 1% level in all models with a value of either 0.004 or 0.005. The coefficients on $POPDENSE$ are highly significant at either the 1% or 5% level in all models, with values ranging from 0.265 to 0.639. Only the climatic variables, $TEMP$ and $PDSI$, are statistically significant at the 1% level.

2.7.3.2 Spatial Lag Random Effects Negative Binomial Regression

These results are presented in Table 2-10. They reveal that the coefficients on $POPDENSE$ are positive and significant at either the 1% or 5% level in all models, with

values ranging from 0.285 to 0.447. The spatial lag term, $SLAGHAT$, is statistically significant in Models 1-4 at the 1% level with values ranging from 0.763 to 1.112. The coefficient on $TEMP$ is significant at the 1% level and its value is 0.210.

2.7.3.3 IV Spatial Lag Random Effects Negative Binomial Regression

These results are shown in Table 2-11. They indicate that the economic variables, $FORHAT$ and $INCHAT$ are statistically significant at the 1% level. The coefficients on $FORHAT$ are positive, with values ranging from 2.009 to 2.709 in all models. This result provides evidence that counties with a higher human WNV prevalence also have a higher number of home foreclosures. The coefficients on $INCHAT$ are consistently negative in all models and range from -3.896 to -4.518. This result suggests that median county income is an important factor in WNV transmission in Colorado. The coefficient on $SLAGHAT$ is positive and significant at the 1% level in Model 1 indicating the presence of spatial spillover effects.

2.7.3.4 Spatial Filtering Random Effects Negative Binomial Regression

Table 2-12 presents the results of this regression. The estimated coefficient on $FORCLOSE$ is positive and significant at the 10% level in only Model 1 with a value of 0.066. The coefficients on $INCOME$ are positive and significant at the 5% or 10% level in Models 1 and 3 only. This is contrary to what is expected theoretically. The coefficient on $MOSQUITO$ is statistically significant at the 1% or 5% level, with values ranging from 0.002 to 0.004. The coefficients on $POPDENSE$ are consistently positive and statistically significant at the 1% level in all models, with values ranging 0.365 to 0.654. All the components of $SFILTER$ are statistically significant at either the 1%, 5% or 10%

level, except the eigenvector $VEC28$ and $VEC29$. The drought index, $PDSI$, is positive and significant at the 1% level with a value of 0.293.

2.7.3.5 IV Spatial Filtering Random Effects Negative Binomial Regression

The results of this regression are presented in Table 2-13. This regression is the most preferred and parsimonious because it has the lowest AIC value and also corrects for both spatial autocorrelation and endogeneity. The coefficients on $FORHAT$ are consistently positive and statistically significant at the 1% level, with values ranging from 2.631 to 2.926. This provides evidence that home foreclosures contributed significantly to the high incidence of human WNV in Colorado. The coefficients on $INCHAT$ are consistently negative and statistically significant at the 1% level in all models, with values ranging from -3.706 to -4.573. This result implies that human WNV is lower in counties with a higher median income. The coefficient on $MOSQUITO$ is positive and statistically significant at the 10% level with a value of 0.007 in Models 3 and 4. These results provide evidence that human WNV is higher in counties with a higher number of mosquito pools or breeding sites. All the components of $SFILTER$ are significant at either the 1% or 10% level. Because this regression was the most parsimonious and preferred, additional regression were run to verify that the economic variables were robust to different model specifications. An interaction term involving $MOSQUITO$ and $POPDENSE$ were included as extra regressors and the significance of the economic variables did not change. This suggests that the results were robust to various specifications. The results of these additional regressions are provided in Appendix E.

Table 2-4: Non-Spatial Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	0.073	0.064	0.048	0.025
	(1.27)	(1.03)	(0.76)	(0.41)
INCOME	-2.181	-1.734	-1.797	-2.079
	(2.84)***	(2.23)**	(2.33)**	(2.78)***
MOSQUITO		0.010	0.010	0.010
		(9.37)***	(9.59)***	(11.37)***
POPDENSE	0.274	0.327	0.326	0.451
	(2.43)**	(2.69)***	(2.80)***	(3.46)***
D2004	2.879	2.790	2.801	2.550
	(4.67)***	(4.59)***	(4.61)***	(4.16)***
D2005	4.155	3.451	3.589	4.457
	(6.87)***	(5.67)***	(5.96)***	(6.81)***
D2006	3.522	3.191	3.229	3.443
	(5.77)***	(5.23)***	(5.30)***	(5.63)***
D2007	3.501	3.336	3.198	2.486
	(5.59)***	(5.29)***	(5.13)***	(3.77)***
PRECIP		0.245		
		(1.54)		
TEMP			-0.003	
			(0.09)	
PDSI				-0.257
				(3.32)***
CONSTANT	17.916	12.582	13.970	16.239
	(2.28)**	(1.58)	(1.62)	(2.13)**
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1112.89	1083.319	1085.854	1075.842
BIC	1149.589	1127.358	1129.892	1119.88
r	0.794	1.023	1.090	1.145
	S.E. (0.169)	S.E. (0.253)	S.E. (0.285)	S.E. (0.293)
s	1.693	2.350	2.888	2.932
	S.E. (0.699)	S.E. (1.140)	S.E. (1.488)	S.E. (1.407)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

Absolute value of z statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
S.E. denotes the standard errors

Table 2-5: Spatial Lag Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	0.084	0.079	0.081	0.035
	(1.70)*	(1.36)	(1.42)	(0.63)
INCOME	-2.571	-2.160	-2.316	-2.246
	(4.40)***	(3.36)***	(3.65)***	(3.47)***
MOSQUITO		0.006	0.006	0.007
		(5.93)***	(5.76)***	(6.30)***
POPDENSE	0.314	0.331	0.321	0.399
	(3.83)***	(3.30)***	(3.32)***	(3.67)***
D2004	-0.943	-0.591	-0.571	-0.176
	(1.18)	(0.70)	(0.68)	(0.22)
D2005	-0.847	-0.388	-0.230	0.846
	(1.02)	(0.45)	(0.27)	(0.90)
D2006	-0.366	0.234	0.252	0.781
	(0.49)	(0.30)	(0.33)	(0.99)
D2007	-0.546	0.340	0.142	0.189
	(0.71)	(0.43)	(0.18)	(0.24)
SLAGHAT[c]	1.234	1.008	1.011	0.875
	(8.80)***	(6.38)***	(6.42)***	(5.45)***
PRECIP		0.277		
		(2.13)**		
TEMP			-0.048	
			(1.45)	
PDSI				-0.151
				(2.05)**
CONSTANT	25.010	19.699	24.746	20.610
	(4.18)***	(2.98)***	(3.45)***	(3.10)***
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1082.513	1053.28	1056.065	1053.996
BIC	1122.881	1100.989	1103.773	1101.704
r	3.162	2.050	2.240	1.957
	S.E. (2.354)	S.E. (0.900)	S.E. (0.994)	S.E. (0.758)
s	23.707	7.981	9.853	7.507
	S.E. (27.364)	S.E. (6.322)	S.E. (7.543)	S.E. (5.156)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
c-represents instrumentalized version of $SLAG$.

Instruments are bird, mosquito, net migration precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevate, urban, popdense, roads, log of area and D2004-D2007

S.E. denotes the standard errors

Table 2-6: IV Spatial Lag Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
FORHAT[a]	0.180	0.253	0.301	0.238
	(1.97)**	(1.67)*	(1.68)*	(1.77)*
INCHAT[b]	-2.385	-1.444	-1.650	-1.474
	(2.30)**	(1.71)*	(1.82)*	(1.40)
MOSQUITO		0.002	0.002	0.002
		(0.90)	(0.84)	(1.02)
POPDENSE	0.300	0.163	0.145	0.161
	(1.73)*	(0.77)	(0.69)	(0.93)
D2004	1.010	0.925	0.979	1.104
	(0.44)	(0.26)	(0.22)	(0.32)
D2005	1.546	1.306	1.484	2.021
	(0.67)	(0.38)	(0.34)	(0.59)
D2006	1.934	1.721	1.812	2.066
	(0.85)	(0.49)	(0.41)	(0.60)
D2007	1.811	1.690	1.641	1.526
	(0.79)	(0.48)	(0.37)	(0.44)
SLAGHAT[c]	0.597	0.578	0.561	0.498
	(3.88)***	(3.28)***	(2.86)***	(3.57)***
PRECIP		0.152		
		(1.46)		
TEMP			-0.047	
			(1.51)	
PDSI				-0.099
				(1.32)
CONSTANT	36.045	24.967	30.493	26.750
	(3.41)***	(1.87)*	(2.63)***	(2.19)**
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1047.9351	1048.576	1048.6282	1048.5851
BIC	1088.3037	1096.2845	1096.3366	1096.2936
r	1.783	1.800	1.755	1.839
	S.E.	S.E.	S.E.	S.E.
	(16.419)	(17.232)	(15.951)	(1.568)
s	4.907	4.951	4.717	5.312
	S.E.	S.E.	S.E.	S.E.
	(48.987)	(50.135)	(47.553)	(8.800)
Moran's I on Residuals (p value)	0.193	0.196	0.194	0.194

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

a, b, c-represent instrumentalized versions of $FORCLOSE$, $INCOME$ and $SLAG$respectively.

Instruments are bird, mosquito, net migration precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevate, urban, popdense, roads, log of area and D2004-D2007

S.E. denotes the standard errors

Table 2-7: Spatial Filtering Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	0.091	0.035	0.079	0.064
	(1.69)*	(0.57)	(1.36)	(0.97)
INCOME	-2.875	-2.174	-2.794	-2.004
	(3.74)***	(2.94)***	(3.67)***	(2.57)**
MOSQUITO		0.010	0.011	0.011
		(8.72)***	(9.90)***	(11.09)***
POPDENSE	0.308	0.336	0.296	0.411
	(2.77)***	(3.04)***	(2.65)***	(3.16)***
D2004	2.834	2.786	2.777	2.485
	(4.60)***	(4.58)***	(4.56)***	(4.04)***
D2005	4.304	3.651	3.796	4.535
	(7.10)***	(5.79)***	(6.30)***	(6.91)***
D2006	3.648	3.299	3.409	3.452
	(5.98)***	(5.39)***	(5.60)***	(5.64)***
D2007	3.617	3.375	3.299	2.346
	(5.79)***	(5.01)***	(5.30)***	(3.53)***
PRECIP		0.155		
		(0.58)		
TEMP			-0.009	
			(0.26)	
PDSI				-0.254
				(3.16)***
VEC3	2.299	2.615	2.279	
	(2.15)**	(2.66)***	(2.33)**	
VEC16	-1.781	-2.691	-1.113	-0.793
	(1.65)*	(2.70)***	(1.17)	(0.74)
VEC25	-1.047	-1.419	-1.287	
	(1.03)	(1.62)	(1.41)	
VEC29	-2.322	-2.756	-2.551	-2.941
	(2.32)**	(2.92)***	(2.65)***	(2.80)***
VEC46	2.745	2.307	3.043	
	(2.58)***	(2.33)**	(3.25)***	
VEC1		-1.481		1.381
		(0.65)		(1.14)
VEC11		0.400		
		(0.47)		
VEC13		-0.264		
		(0.28)		
VEC15		-1.419		
		(1.53)		
VEC21		0.605		
		(0.71)		
VEC26		1.647		
		(1.53)		

Table 2-7 (cont.): Spatial Filtering Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
VEC27		-0.802		
		(1.04)		
VEC33			-1.668	
			(1.79)*	
VEC34		2.901		
		(3.09)***		
CONSTANT	24.774	16.960	24.659	15.412
	(3.14)***	(2.22)**	(2.89)***	(1.93)*
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1104.085	1073.117	1070.276	1072.563
BIC	1159.133	1164.864	1136.334	1127.611
r	1.037	3.041	1.766	1.257
	S.E. (0.245)	S.E. (1.516)	S.E. (0.649)	S.E. (0.338)
s	3.539	19.349	7.857	3.397
	S.E. (1.599)	S.E. (14.781)	S.E. (5.159)	S.E. (1.709)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

z statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

S.E. denotes the standard errors

Table 2-8: IV Spatial Filtering Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
FORHAT[a]	0.406	0.382	0.413	0.387
	(2.16)**	(2.39)**	(2.33)**	(2.52)**
INCHAT[b]	-4.183	-1.206	-1.234	-1.012
	(3.72)***	(1.69)*	(1.70)*	(0.95)
MOSQUITO		0.005	0.005	0.005
		(2.28)**	(1.81)*	(2.36)**
POPDENSE	0.417	0.022	0.005	-0.001
	(2.25)**	(0.11)	(0.03)	(0.01)
D2004	3.288	2.870	2.861	2.743
	(1.09)	(0.75)	(0.63)	(0.84)
D2005	4.294	3.652	3.665	3.968
	(1.41)	(0.96)	(0.80)	(1.24)
D2006	4.114	3.413	3.405	3.454
	(1.33)	(0.89)	(0.74)	(1.06)
D2007	4.157	3.335	3.290	2.899
	(1.34)	(0.88)	(0.72)	(0.88)
PRECIP		0.037		
		(0.33)		
TEMP			-0.023	
			(0.63)	
PDSI				-0.111
				(1.39)
VEC4	-3.263			
	(4.08)***			
VEC6	1.638			
	(1.20)			
VEC15	-1.630			
	(1.76)*			
VEC16	-2.201	-1.728	-1.725	-1.696
	(1.74)*	(1.64)	(1.53)	(1.67)*
VEC46	2.402			
	(2.22)**			
VEC47	1.319			
	(1.21)			
VEC41				-0.013
				(0.01)
CONSTANT	48.989	20.222	21.861	17.605
	(4.12)***	(1.65)*	(1.79)*	(1.44)
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1044.021	1055.9445	1055.7214	1055.9667
BIC	1102.7391	1103.6529	1103.4298	1107.3451

Table 2-8 (cont.): IV Spatial Filtering Random Effects Negative Binomial Regression for California

	Model 1	Model 2	Model 3	Model 4
r	2.111	1.668	1.656	1.787
	S.E. (0.714)	S.E. (0.492)	S.E.(0.422)	S.E. (0.958)
s	7.474	4.689	4.598	5.414
	S.E. (4.172)	S.E. (2.413)	S.E. (4.910)	S.E. (5.846)
Moran's I on Residuals (p value)	0.196	0.200	0.200	0.195

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
a, b-represent instrumentalized versions of $FORCLOSE$ and $INCOME$ respectively

Instruments are bird, mosquito, net migration, precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevation, urban, popdense, road, log of area and D2004-D2007

S.E. denotes the standard errors

Table 2-9: Non-Spatial Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	0.041	0.050	0.010	0.032
	(0.74)	(0.89)	(0.24)	(0.59)
INCOME	-0.733	-0.990	-0.512	-0.727
	(1.13)	(1.52)	(0.77)	(1.15)
MOSQUITO		0.005	0.004	0.004
		(3.47)***	(2.70)***	(2.64)***
POPDENSE	0.308	0.342	0.639	0.265
	(2.49)**	(2.76)***	(5.53)***	(2.30)**
D2004	-2.193	-2.478	-2.209	-2.878
	(10.19)***	(8.26)***	(12.06)***	(12.89)***
D2005	-2.480	-2.631	-2.759	-3.631
	(10.67)***	(10.43)***	(13.29)***	(12.62)***
D2006	-2.183	-2.371	-2.401	-2.875
	(10.12)***	(10.13)***	(13.44)***	(13.23)***
D2007	-1.437	-1.722	-1.309	-2.666
	(7.44)***	(8.05)***	(8.12)***	(10.44)***
PRECIP		1.155		
		(1.57)		
TEMP			0.401	
			(6.35)***	
PDSI				0.275
				(5.12)***
CONSTANT	8.297	9.420	-13.727	9.420
	(1.25)	(1.44)	(1.68)*	(1.45)
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1341.671	1334.327	1294.175	1312.54
BIC	1379.196	1379.358	1339.206	1357.571
r	0.745	0.842	1.079	0.877
	S.E.	S.E.	S.E.	S.E.
	(0.150)	(0.184)	(0.225)	(0.181)
s	0.860	1.102	1.295	0.955
	S.E.	S.E.	S.E.	S.E.
	(0.268)	(0.419)	(0.389)	(0.304)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

Absolute value of z statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

S.E. denotes the standard errors

Table 2-10: Spatial Lag Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	-0.059	-0.058	-0.051	-0.067
	(1.31)	(1.25)	(1.25)	(1.54)
INCOME	-0.305	-0.478	-0.161	-0.149
	(0.47)	(0.72)	(0.24)	(0.22)
MOSQUITO		-0.000	-0.000	-0.001
		(0.26)	(0.03)	(0.63)
POPDENSE	0.285	0.296	0.447	0.293
	(2.33)**	(2.43)**	(3.51)***	(2.35)**
D2004	-0.240	-0.422	-0.619	0.230
	(0.77)	(1.08)	(1.77)*	(0.45)
D2005	-0.043	-0.107	-0.682	0.582
	(0.12)	(0.27)	(1.55)	(0.90)
D2006	-0.192	-0.284	-0.672	0.261
	(0.63)	(0.82)	(1.85)*	(0.54)
D2007	0.057	0.001	-0.124	0.592
	(0.25)	(0.00)	(0.47)	(1.17)
SLAGHAT[c]	0.964	0.954	0.763	1.112
	(9.04)***	(8.01)***	(5.49)***	(6.73)***
PRECIP		0.595		
		(0.90)		
TEMP			0.210	
			(2.87)***	
PDSI				-0.073
				(1.14)
CONSTANT	1.832	2.883	-9.511	-0.421
	(0.28)	(0.43)	(1.16)	(0.06)
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1272.577	1275.686	1268.383	1275.245
BIC	1313.855	1324.469	1317.167	1324.029
r	1.098	1.113	1.169	1.121
	S.E.	S.E.	S.E.	S.E.
	(0.233)	(0.238)	(0.247)	(0.238)
s	0.961	1.024	1.040	0.936
	S.E.	S.E.	S.E.	S.E.
	(0.282)	(0.316)	(0.294)	(0.269)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

c-represents instrumentalized version of $SLAG$.

Instruments are bird, mosquito, net migration, precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevation, urban, popdense, road, log of area and D2004-D2007

S.E. denotes the standard errors

Table 2-11: IV Spatial Lag Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
FORHAT[a]	2.009	2.709	2.557	2.637
	(4.50)***	(5.10)***	(6.11)***	(6.62)***
INCHAT[b]	-3.896	-4.504	-4.360	-4.518
	(2.89)***	(3.63)***	(3.28)***	(3.66)***
MOSQUITO		0.005	0.005	0.005
		(1.17)	(1.53)	(1.39)
POPDENSE	0.076	-0.058	0.000	-0.040
	(0.46)	(0.32)	(0.00)	(0.24)
D2004	-1.640	-2.079	-2.204	-2.135
	(3.43)***	(4.54)***	(4.77)***	(2.87)***
D2005	-1.683	-2.335	-2.433	-2.304
	(3.25)***	(4.71)***	(5.08)***	(2.18)**
D2006	-1.724	-2.307	-2.383	-2.296
	(3.17)***	(5.22)***	(5.78)***	(3.11)***
D2007	-1.562	-2.365	-2.348	-2.332
	(4.05)***	(5.47)***	(6.08)***	(2.71)***
SLAGHAT[c]	0.433	0.203	0.176	0.215
	(2.83)***	(1.28)	(1.26)	(0.84)
PRECIP		-0.374		
		(0.40)		
TEMP			0.047	
			(0.51)	
PDSI				-0.005
				(0.04)
CONSTANT	37.825	44.438	40.426	44.149
	(2.74)***	(3.46)***	(2.54)**	(3.39)***
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1237.283	1229.362	1229.213	1229.723
BIC	1278.562	1278.145	1277.996	1278.506
r	1.792	1.991	1.948	2.001
	S.E.	S.E.	S.E.	S.E.
	(0.739)	(0.590)	(0.565)	(0.623)
s	2.828	3.400	3.248	3.461
	S.E.	S.E.	S.E.	S.E.
	(1.247)	(1.266)	(1.294)	(1.457)
Moran's I on Residuals (p value)	0.127	0.128	0.128	0.128

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

a, b, c-represent instrumentalized versions of $FORCLOSE$, $INCOME$ and $SLAG$respectively.

Instruments are bird, mosquito, net migration, precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevation, urban, popdense, road, log of area and D2004-D2007
S.E. denotes the standard errors

Table 2-12: Spatial Filtering Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
FORCLOSE	0.066	0.058	0.060	0.069
	(1.68)*	(1.36)	(1.47)	(1.34)
INCOME	1.760	0.734	1.683	-1.073
	(2.05)**	(1.00)	(2.01)**	(2.23)**
MOSQUITO		0.002	0.003	0.004
		(2.12)**	(2.23)**	(3.11)***
POPDENSE	0.365	0.556	0.514	0.654
	(3.04)***	(4.22)***	(3.85)***	(6.76)***
D2004	-3.053	-3.020	-2.930	-2.748
	(18.65)***	(13.63)***	(14.64)***	(11.92)***
D2005	-3.179	-3.087	-3.091	-3.621
	(17.84)***	(16.79)***	(16.54)***	(12.31)***
D2006	-2.637	-2.604	-2.623	-2.756
	(17.32)***	(17.04)***	(16.73)***	(12.31)***
D2007	-1.831	-1.821	-1.880	-2.673
	(12.80)***	(12.38)***	(11.10)***	(10.76)***
PRECIP		0.331		
		(0.57)		
TEMP			0.014	
			(0.17)	
PDSI				0.293
				(5.89)***
VEC1	10.989	9.485	10.072	
	(8.55)***	(7.59)***	(5.79)***	
VEC2	2.579		2.703	
	(1.86)*		(1.94)*	
VEC11	3.985	2.489	3.348	0.877
	(3.34)***	(2.23)**	(2.82)***	(1.08)
VEC21	-3.970	-4.216	-4.087	-2.239
	(3.61)***	(3.91)***	(3.63)***	(2.84)***
VEC28	1.505		1.018	
	(1.23)		(0.86)	
VEC29	-1.596	-1.474	-1.584	-1.219
	(1.36)	(1.36)	(1.46)	(1.60)
VEC35	-3.508	-2.399	-2.750	-0.273
	(3.03)***	(2.09)**	(2.36)**	(0.33)
VEC58	-3.633	-2.984	-3.263	
	(3.27)***	(2.80)***	(2.99)***	
VEC4		3.246	3.122	6.037
		(2.35)**	(2.06)**	(5.78)***
CONSTANT	-17.744	-7.979	-18.211	10.944
	(2.01)**	(1.08)	(1.97)**	(2.17)**
Observations	315	315	315	315

Table 2-12 (cont.): Spatial Filtering Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
Number of locations	63	63	63	63
AIC	1264.235	1262.436	1261.042	1300.218
BIC	1331.781	1333.735	1339.846	1364.012
r	1.387	1.525	1.496	2.621
	S.E. (0.317)	S.E. (0.362)	S.E. (0.359)	S.E. (1.249)
s	1.570	2.077	2.009	12.731
	S.E. (0.540)	S.E. (0.798)	S.E. (0.785)	S.E. (10.347)
Moran's I on Residuals (p value)	0.000	0.000	0.000	0.000

z statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
S.E. denotes the standard errors

Table 2-13: IV Spatial Filtering Random Effects Negative Binomial Regression for Colorado

	Model 1	Model 2	Model 3	Model 4
FORHAT[a]	2.631	2.854	2.926	2.961
	(9.80)***	(6.36)***	(6.89)***	(7.95)***
INCHAT[b]	-3.737	-4.573	-3.706	-4.071
	(3.28)***	(3.83)***	(2.46)**	(3.91)***
MOSQUITO		0.006	0.007	0.007
		(1.07)	(1.85)*	(1.96)*
POPDENSE	0.326	0.135	0.205	0.055
	(1.63)	(0.60)	(1.15)	(0.36)
D2004	-2.692	-2.453	-2.643	-2.733
	(10.96)***	(7.01)***	(9.73)***	(13.88)***
D2005	-2.968	-2.854	-2.953	-3.066
	(14.49)***	(11.53)***	(13.43)***	(11.75)***
D2006	-2.923	-2.768	-2.934	-2.985
	(9.46)***	(10.30)***	(11.82)***	(11.80)***
D2007	-2.689	-2.736	-3.023	-3.066
	(12.21)***	(9.18)***	(9.89)***	(10.87)***
PRECIP		-0.637		
		(0.68)		
TEMP			-0.030	
			(0.31)	
PDSI				0.024
				(0.30)
VEC2	4.392		4.707	4.005
	(3.56)***		(3.83)***	(2.94)***
VEC4	3.269	2.554	3.320	
	(2.98)***	(1.68)*	(2.41)**	
VEC6	3.319	1.937	3.175	2.753
	(2.86)***	(2.02)**	(3.06)***	(2.48)**
VEC22	1.866	1.597	1.607	1.643
	(1.82)*	(1.76)*	(1.80)*	(1.68)*
CONSTANT	35.232	45.016	36.123	39.195
	(3.02)***	(3.74)***	(2.02)**	(3.65)***
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1221.647	1221.554	1205.827	1213.728
BIC	1274.183	1277.842	1265.869	1270.017
r	2.438	2.552	2.856	2.464
	S.E. (1.351)	S.E. (4.443)	S.E. (7.898)	S.E. (0.905)
s	6.376	6.075	7.300	5.042
	S.E. (3.206)	S.E. (10.964)	S.E. (21.131)	S.E. (1.873)
Moran's I on Residuals (p-value)	0.125	0.130	0.125	0.123

z statistics in parentheses, * significant at 10%, ** significant at 5% *** significant at 1%
a, b-represent instrumentalized versions of $FORCLOSE$ and $INCOME$ respectively

Instruments are bird, mosquito, net migration, precip, temp, pdsi poverty, unemployment rate, education, airport, equine, elevation, urban, popdense, road, log of area and D2004-D2007

S.E. denotes the standard errors

Table 2-14: Selected Eigenvectors for Spatial Filtering Random Effects Negative Binomial Regression

Model	Selected Eigenvectors	Level of Significance
	California	
1	VEC3, VEC1, VEC25, VEC29, VEC46	5%
2	VEC1, VEC3, VEC11, VEC13, VEC15, VEC16,VEC21, VEC25,	5%
3	VEC26, VEC27, VEC29, VEC34, VEC46	5%
4	VEC1, VEC16, VEC29	5%
	Colorado	
Model	Selected Eigenvector	
1	VEC1, VEC2, VEC11, VEC21, VEC28, VEC29, VEC35, VEC37, VEC58	5%
2	VEC1, VEC4, VEC11,VEC21, VEC29, VEC35, VEC58	5%
3	VEC1, VEC2, VEC4, VEC11, VEC21, VEC28, VEC29, VEC35,	5%
4	VEC58	5%
	VEC4, VEC11,VEC21,VEC29,VEC35	

Table 2-15: Selected Eigenvectors for IV Spatial Filtering Random Effects Negative Binomial Regression

Model	Selected Eigenvectors	Level of Significance
	California	
1	VEC4, VEC6, VEC15, VEC16, VEC46, VEC47	5%
2	VEC16	5%
3	VEC16	5%
4	VEC16, VEC41	5%
	Colorado	
Model	Selected Eigenvector	
1	VEC2, VEC4, VEC6, VEC22	5%
2	VEC4, VEC6,VEC22	5%
3	VEC2, VEC4, VEC6, VEC22	5%
4	VEC2,VEC6,VEC22	5%

Table 2-16: Sign and statistical significance of the economic and biological covariates for California

Non-Spatial Random Effects Negative Binomial Regression	Model 1	Model 2	Model 3	Model 4
Home Foreclosures	+	+	+	+
Income	-***	-**	-**	-***
Mosquito		+***	+***	+***
Spatial Lag Random Effects Negative Binomial Regression				
Home Foreclosures	+*	+	+	+
Income	-***	-***	-***	+***
Mosquito		+***	+***	+***
IV Spatial Lag Random Effects Negative Binomial Regression				
Home Foreclosures	+**	+*	+*	+*
Income	-**	-*	-*	-
Mosquito		+	+	+
Spatial Filtering Random Effects Negative Binomial Regression				
Home Foreclosures	+*	+	+	+
Income	-**	-***	-***	-**
Mosquito		+***	+***	+***
IV Spatial Filtering Random Effects Negative Binomial Regression (Preferred Model)				
Home Foreclosures	+**	+**	+**	+**
Income	-***	-*	-*	-
Mosquito		+**	+*	+**

*indicates significance at the 10% level, **indicates significance at the 5% level and *** indicates significance at the 1% level.

Table 2-17: Sign and statistical significance of the economic and biological covariates for Colorado

Non-spatial Lag Random Effects Negative Binomial Regression	Model 1	Model 2	Model 3	Model 4
Home Foreclosures	+	+	+	+
Income	-	-	-	-
Mosquito		+***	+***	+***
Spatial Lag Random Effects Negative Binomial Regression				
Home Foreclosures	-	-	-	-
Income	-	-	-	-
Mosquito		+	+	+
IV Spatial Lag Random Effects Negative Binomial Regression				
Home Foreclosures	+***	+***	+***	+***
Income	-***	-***	-***	-***
Mosquito		+	+	+
Spatial Filtering Random Effects Negative Binomial Regression				
Home Foreclosures	+*	+	+	+
Income	+**	+	+**	-**
Mosquito		+**	+**	+***
IV Spatial Filtering Random Effects Negative Binomial Regression (Preferred Model)				
Home Foreclosures	+***	+***	+***	+***
Income	-***	-***	-***	-***
Mosquito		+	+*	+*

*indicates significance at the 10% level, **indicates significance at the 5% level and *** indicates significance at the 1% level.

Figure 2-5: Spatial Distribution of Human WNV in California (2003-2007)

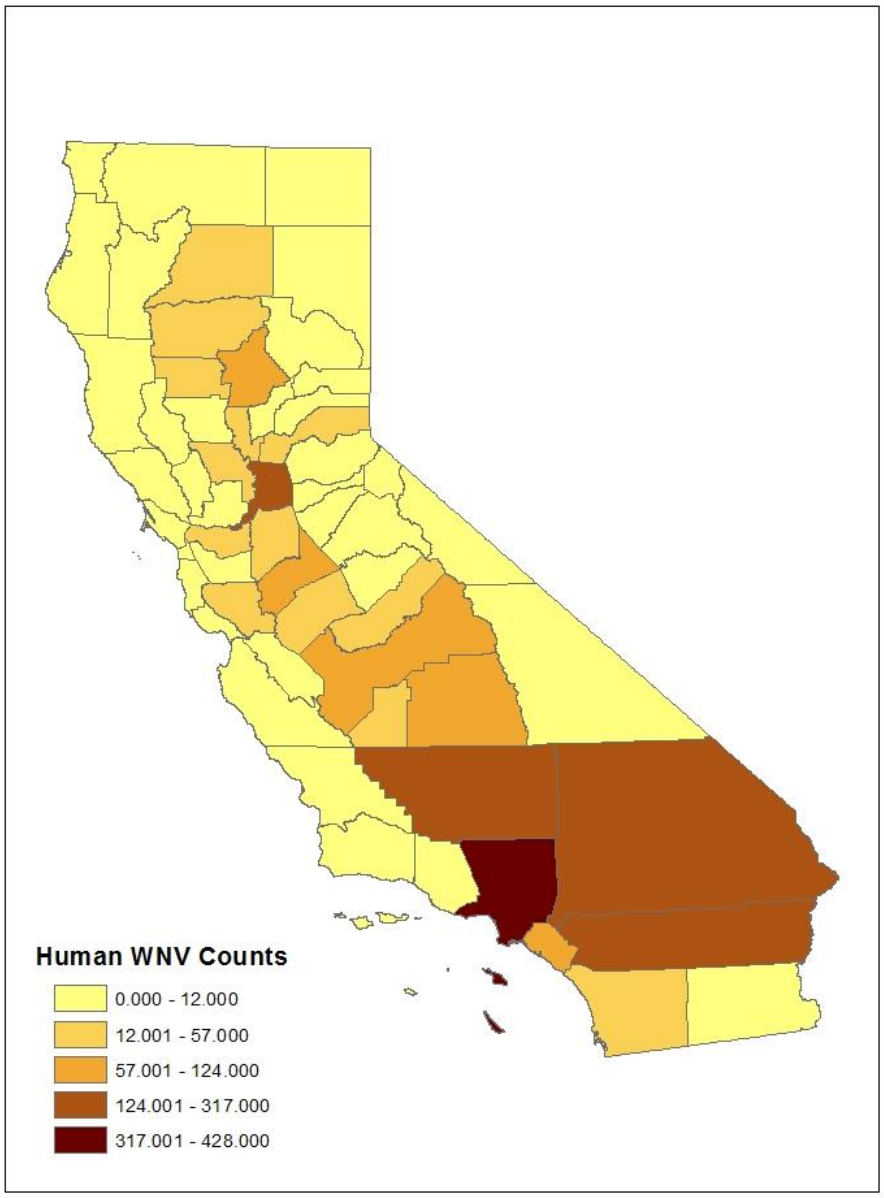


Figure 2-6: Spatial Distribution of Income in California (2003-2007)

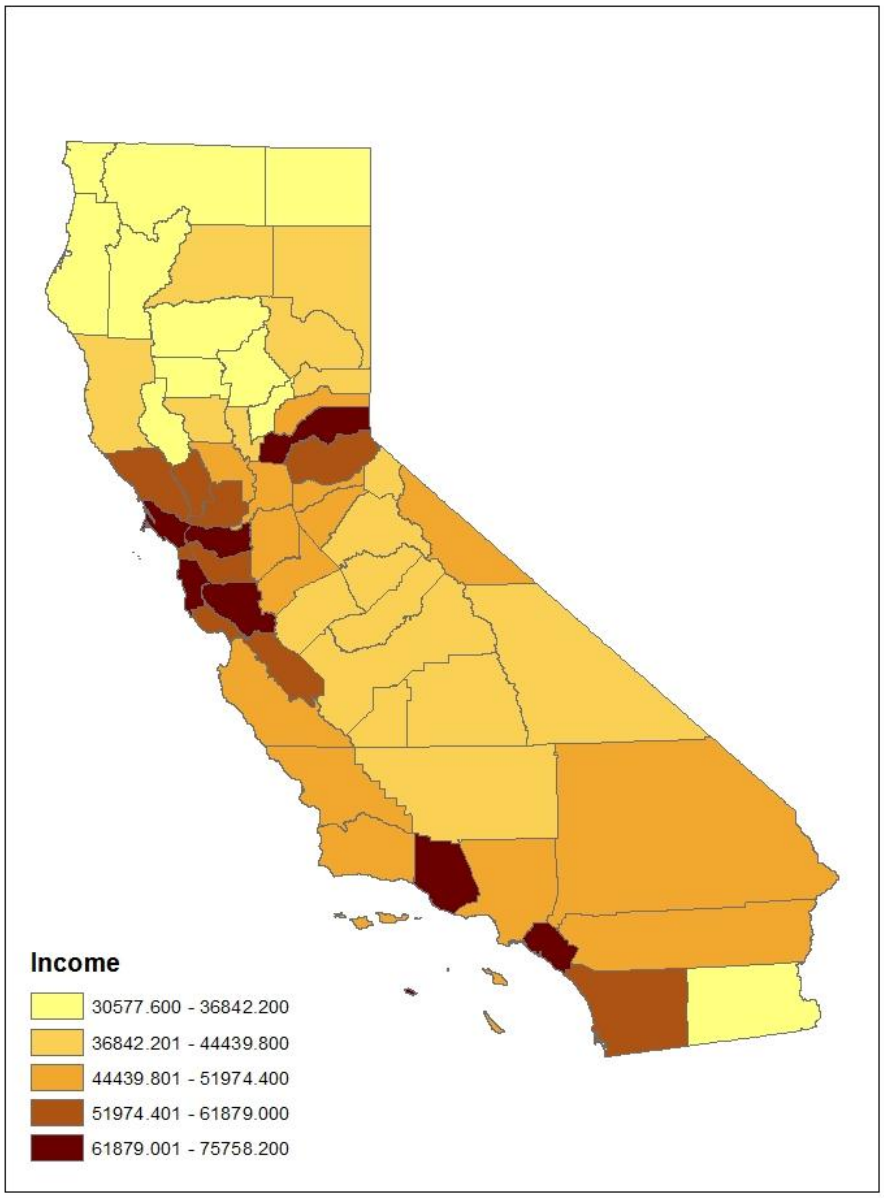


Figure 2-7: Spatial Distribution of Home Foreclosures in California (2003-2007)

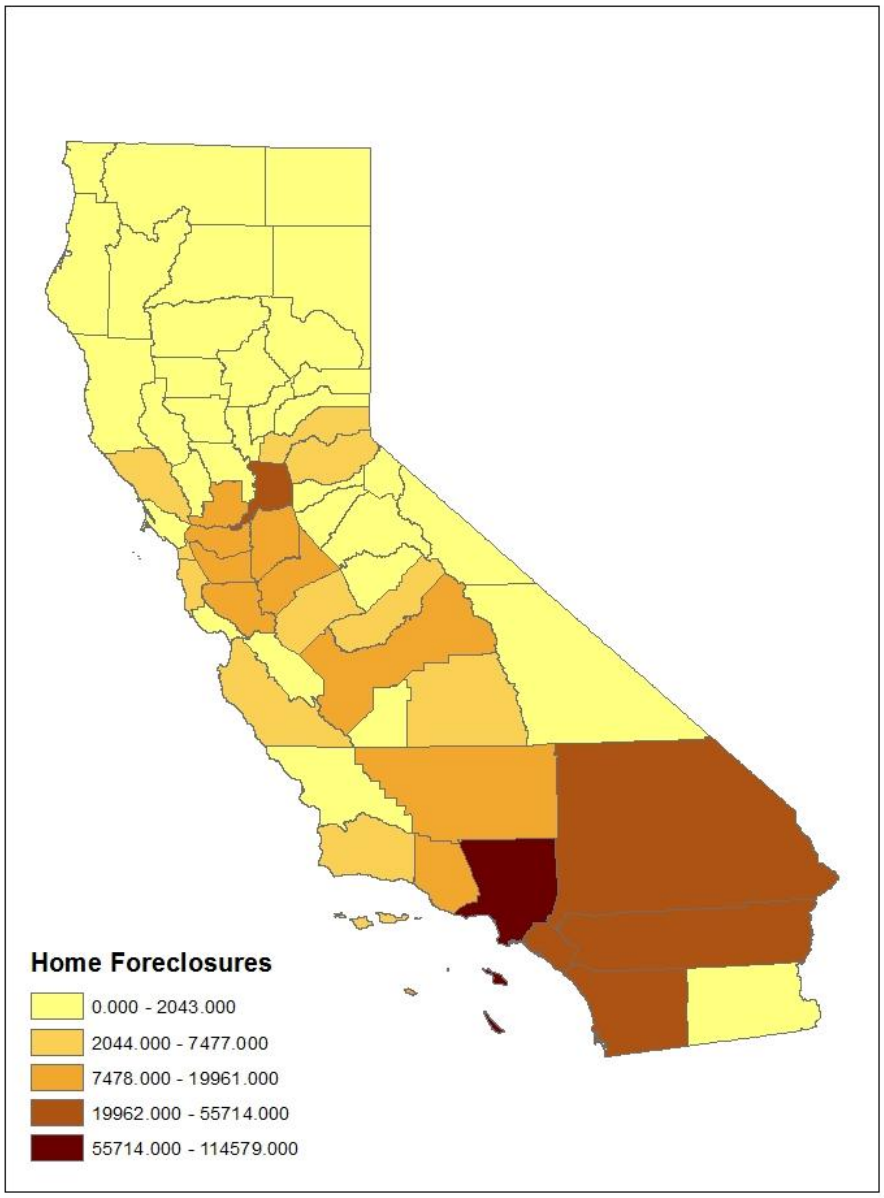


Figure 2-8: Spatial Distribution of Mosquito Pools in California (2003-2007)

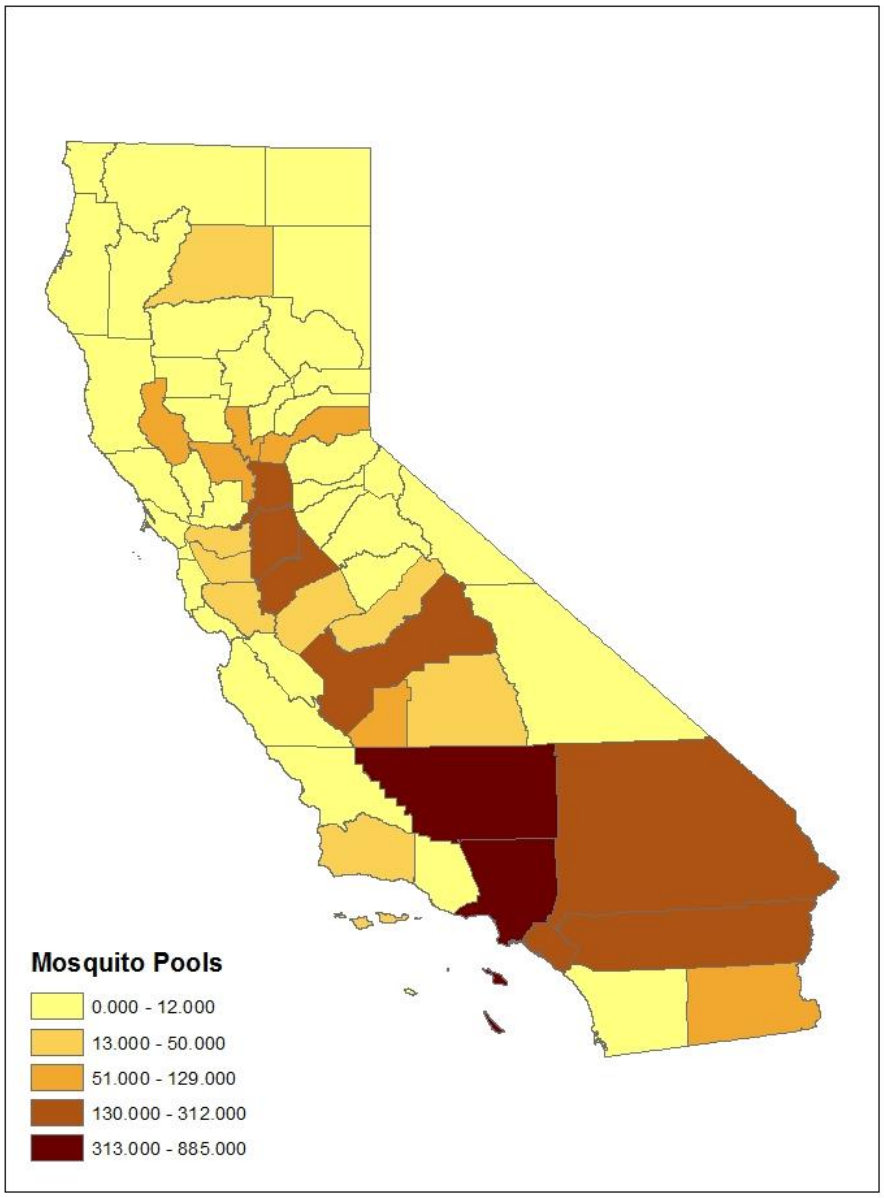


Figure 2-9: Spatial Distribution of Human WNV in Colorado (2003-2007)

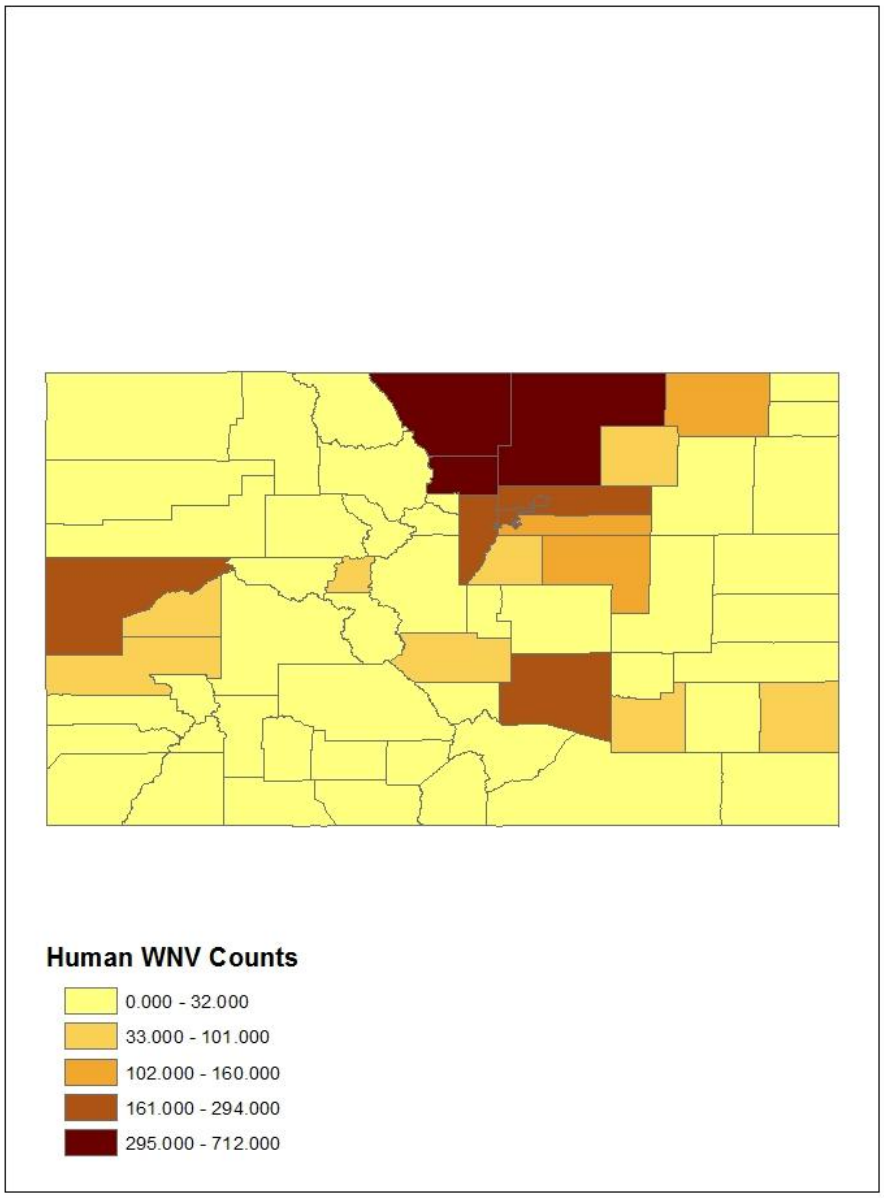


Figure 2-10: Spatial Distribution of Income in Colorado (2003-2007)

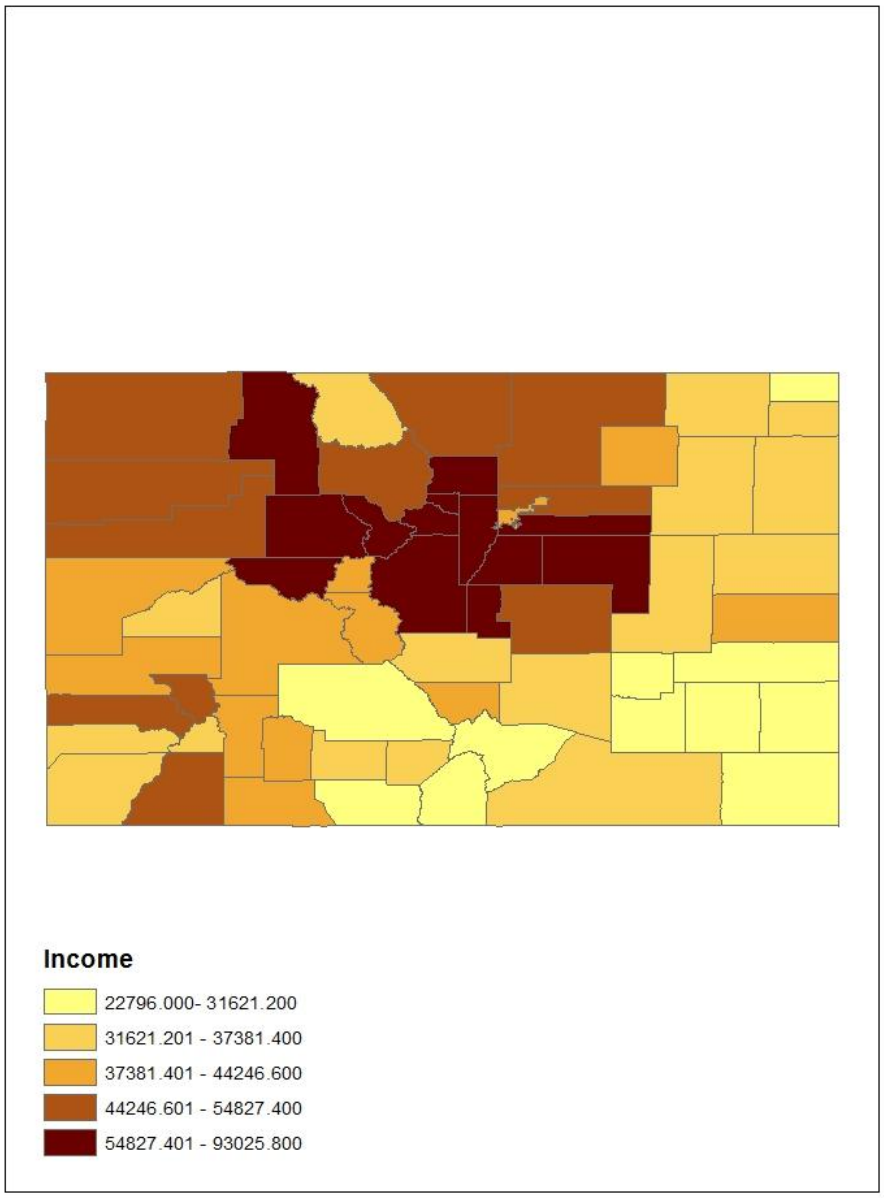


Figure 2-11: Spatial Distribution of Home Foreclosures in Colorado (2003-2007)

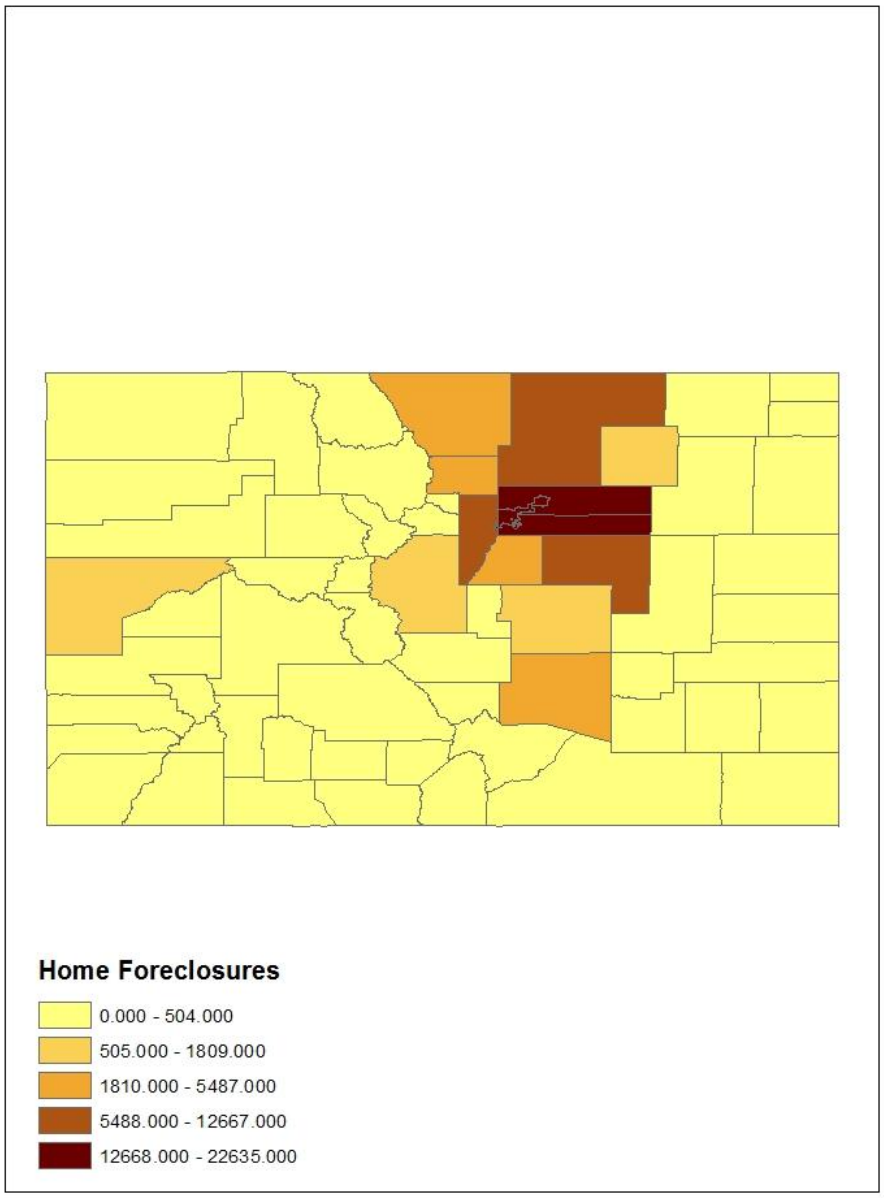
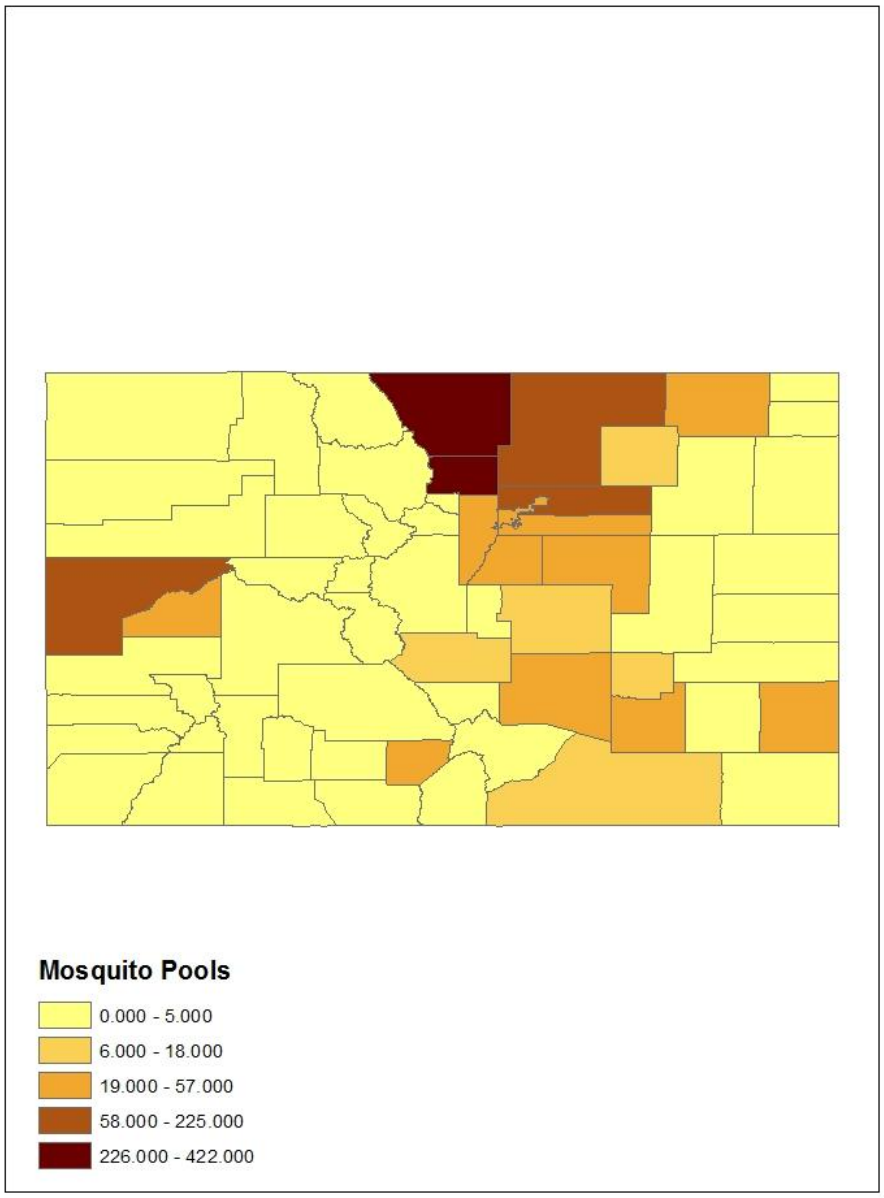


Figure 2-12: Spatial Distribution of Mosquito Pools in Colorado (2003-2007)



2.8 Discussion and Conclusions

The results of all the IV random effects negative binomial regressions for each state show some important findings. Specifically, the risk of human WNV is high in counties characterized by a high number of home foreclosures, low median income and a large number of mosquito pools. One interesting trend that emerged from the results is that the impact of the economic factors on the risk of WNV transmission is much stronger than those of the biological factor.

Counties with high median incomes have lower poverty levels all else equal. Residents of counties with a high median income level could afford to live in more decent neighborhoods with improved housing conditions. This fact is consistent with the finding of Harrigan et al. (2010) who found that per capita income was negatively associated the incidence of WNV in Orange County, California. He argued that poor neighborhoods provided an environment conducive to mosquito breeding. If income is used as a proxy for surveillance effort (Brown et al. 2008) and not as an index of poverty, the results still make epidemiological sense. Counties with higher surveillance efforts experienced a lower incidence of human WNV. The home foreclosure hypothesis is supported for both California and Colorado. These results are consistent with findings of Reisen et al. (2008) and Harrigan et al. (2010) who found that delinquent mortgages and neglected swimming pools were contributory risk factors in WNV transmission in Kern and Orange counties respectively. During the housing crises that began in 2004 there were significant increases in mortgage delinquencies and home foreclosures in several parts of California and Colorado. Mosquitoes are the primary vector responsible for the transmission of WNV and their populations are usually measured by the number of

mosquito pools or breeding sites. The presence of mosquito pools or WNV-infected mosquitoes are a necessary condition for human infection. Hence counties with a large number of mosquito pools are likely to have a higher risk of human WNV.

Moran's test revealed the presence of spatial autocorrelation, while the Hausman specification test indicates that home foreclosures and income are endogenous. To correct these problems, an instrumental variable technique is applied to the spatial lag and spatial filtering regressions. Diagnostic tests on the residuals obtained from these regressions show that they are spatially independent. The AIC and BIC model selection criteria indicate that the IV spatial filtering model is the most preferred. The time dummies show that shocks that occurred in each year affected each state differently. The evidence showed that WNV in Colorado was significantly lower in subsequent years after 2003. California on the other hand had significantly higher WNV cases between 2004 and 2007 relative to 2003. The period from 2004-2007 was characterized by the housing bubble that lead to high number of home foreclosures and delinquent mortgages.

The spatial dependency term showed that spatial spillovers or externalities were present in the data. Counties with similar WNV cases were adjacent to each other. From a policy making perspective, one political unit cannot act alone to address the problem of WNV. For example, suppose one county enacts policy interventions to reduce the risk of WNV but adjoining counties do not follow suit, policy interventions in the first county would be less effective. Thus the presence of spatial dependency signifies that a coordinated policy approach is recommended to address the problem of WNV, rather an individual effort by a single political unit. The results also suggest more resources should be allocated to counties that exhibit certain economic and biological characteristics such

as a high number of home foreclosures, a low median income and a high number of mosquito pools. This will enable them to undertake a comprehensive surveillance and monitoring program to mitigate the disease. Perhaps it is worth pointing out that although this study was conducted at the county level, the policy relevance may be more appropriate at the census track or block level.

Chapter 3: Disease Mapping and Variography

3.1 Introduction

The aim of disease mapping is to provide a visual representation of the geographical distribution of the risk or hazard of a particular disease in a given study area. The traditional approach is to map risk that reflect actual deaths resulting from a disease (mortality) or number of people who suffer from the disease (morbidity) in a given period of time for the population at risk. Disease mapping (production of disease atlases) has a long history in medical geography and epidemiology and can be traced back to the 1800s (Walter and Birnie 1991). These maps provide an insightful and visual summary of spatial or areal pattern of a particular disease or other measures of health outcomes. They are generally used for descriptive purposes, for monitoring and surveillance with a view to assessing which areas are high risk, identifying risk factors responsible for a particular disease, assisting in policy formulation and allocation of health care resources, and etiological disease hypotheses (causes of a disease). Knowledge of the etiology of a disease creates a useful platform for cost-effective preventive and public health services. It will also assist in the discovery of cost-effective strategies to treat the disease. From a historical perspective, early disease maps were often devoted to the depiction of the spatial pattern of infectious diseases such as cholera and yellow fever in the US and Europe. As a select example, Snow (1855) developed a spot map to demonstrate a cholera outbreak in London.

Variograms are geo-statistical techniques used to detect spatial autocorrelation in regional variables. They are quite useful in examining the spatial structure of a variable because they show how the spatial structure varies over a certain distance. Compared to

correlation or covariance functions that show the degree of similarity between two variables, variograms indicate the degree of dissimilarities between two variables as a function of lag distances. In some respect they can be used to depict the nature of local spatial autocorrelation. They are used here to examine the spatial structure of observed human WNV, the standardized morbidity ratio and the model residuals in both California and Colorado. The estimation and fitting methods used to generate the variograms are also described in this chapter.

The main objectives of this chapter are as follows: (1) Develop thematic maps of the standardized morbidity ratio and predicted probabilities from the spatial filtering random effects negative binomial model. This model was chosen because it had a relatively lower AIC value and was considered to be the most parsimonious. These maps may ultimately assist policy makers in the surveillance and monitoring of human WNV with a view to identifying high and low risk areas; (2) Develop spatial variograms to study the spatial structure of the standardized morbidity ratios, the observed human WNV counts and the model residuals. This chapter allows the integration of econometric methods and Geographic Information Systems (GIS) to produce human WNV risk maps for both states. The organization of this chapter is as follows. Section 2 presents thematic maps that show the spatial visualization of various risk measures. Section 3 describes the variogram analysis performed to uncover the spatial structure of the data.

3.2 Thematic Maps

Maps were created using GIS tools to visually display the risk of WNV in both states. These maps were restricted to the standardized morbidity ratio and the predicted probabilities

3.2.1 Standardized Morbidity Ratio

The relative risk of a disease is often measured by the standardized morbidity ratio (SMR). This is a summary measure used in most epidemiological studies and is defined as

$$\hat{\theta}_i = SMR_i = \frac{O_i}{E_i} \tag{3.1}$$

where O_i and E_i are the observed and expected number of WNV counts in area i respectively. A relative risk greater than 1 means that observed human WNV counts exceeds expected human WNV counts. Counties with SMR values in excess of 1 are of interest to decision makers because they can be classified as high risk areas.

The expected number of counts is defined by the formula

$$E_i = \frac{N_i \sum_{i=1}^{m} O_i}{\sum_{i=1}^{C} N_i} \tag{3.2}$$

where N_i is the population of area i ($i = 1, \ldots . m$). The raw estimates of the SMR do have a number of drawbacks. Specifically, they provide inaccurate estimates when the number of cases is small (Clayton and Kaldor 1987). The variances of the estimates are inversely related to E_i so there will be high sampling variability for regions with small or large populations. In other words the SMRs on a map tend to be dominated by areas with

small populations that have extreme values of the SMRs. To rectify this problem, it is not unusual to smooth the raw rates using a variety of methods. There are a variety of methods available in the literature to obtain smoothed estimates of the relative risk using the empirical Bayes method. These methods differ mainly by the distributional assumptions made about the relative risk and include the Poisson-Gamma method, the Log-normal method and Marshall's Global method.

In the Poisson Gamma distribution, estimates of the SMR are derived using a two-level distribution. The observed values are assumed to be independent Poisson random variables with means $\theta_i E_i$ conditional on the θ_i's . Second, the θ_i's are assumed to be independently distributed a priori as Gamma random variables with parameters α and β. Mathematically,

$$O_i \backslash \theta_i, E_i \sim Poisson(\theta_i, E_i)$$

$$\theta_i \sim Gamma(\alpha, \beta)$$

with mean, $E(\theta_i) = {}^{\alpha}/_{\beta} = \mu$ and variance, $Var(\theta_i) = {}^{\alpha}/_{\beta^2}$ where α and β are slope and scale parameters respectively. The Bayes estimator of θ_i is then given by:

$$\hat{\theta}^B = \frac{O_i + \alpha}{E_i + \beta}$$

$$= (1 - B_i) SMR_i + B_i \mu$$

where $B_i = \frac{\beta}{E_i + \beta}$. If α and β are replaced with their estimates $\hat{B}_i$, $\hat{\alpha}$ and respectively, the empirical Bayes estimate is obtained as

$$\hat{\theta}_i^{EB} = \left(1 - \hat{B}_i\right) SMR_i + \hat{B}_i \hat{\mu} \tag{3.3}$$

where $\hat{B}_i = \frac{\hat{\beta}_i}{E_i + \hat{\beta}}$ and $\hat{\mu}$ and $\hat{\sigma}^2$ are obtained using method of moments estimates as

$$\hat{\mu} = \frac{\sum_{i=1}^{m} \hat{\theta}_i E_i}{\sum_{i=1}^{m} E_i}$$

$$\hat{\sigma}^2 = s^2 - \frac{\hat{\mu}}{1/m \sum_i^m E_i}$$

where

$$s^2 = \frac{\sum_{i=1}^{m} E_i (\hat{\theta}_i - \widehat{\mu})}{\sum_{i=1}^{m} E_i}$$

An alternative estimator proposed by Clayton and Kaldor (1987) assumes that the logarithm of the relative risk (log (θ_i) follows a multivariate normal distribution with mean $\emptyset$ and variance σ^2, i.e.,

$$\beta_i = \log(\theta_i) \sim MN(\emptyset, \sigma^2)$$

Under this method, the log of the relative risk is $\log\left(\left(\theta_i + \frac{1}{2}\right)\right)/E_i$to ensure that the estimates are non-negative. The estimate of the relative risk is given as

$$\hat{\beta}_i = b_i = \frac{\hat{\theta} + \left(O_i + \frac{1}{2}\right)\hat{\sigma}^2 \log\left[\frac{(O_i + \frac{1}{2})}{E_i}\right] - \hat{\sigma}^2/2}{1 + (O_i + \frac{1}{2})\hat{\sigma}^2} \tag{3.4}$$

where $\widehat{\emptyset}$ and $\hat{\sigma}^2$ are estimates of the prior mean and variance estimated as

$$\widehat{\emptyset} = 1/n \sum_{i=1}^{n} b_i = \bar{b}$$

$$\hat{\sigma}^2 = 1/n \left\{ \hat{\sigma}^2 \sum_{i=1}^{n} \left[1 + \hat{\sigma}^2 (O_i + \frac{1}{2}) \right]^{-1} + \sum_{i=1}^{n} (b_i - \hat{\theta})^2 \right\}$$

The Expectation-Maximization (EM) algorithm is used to obtain estimates of the mean and variance. Estimates of b_i are successively iterated using previous formulae until it converges. In the limit, the estimate of θ_i is given as $\hat{\theta}_i = \exp(\hat{\beta}_i)$.

Marshall (1991) developed the Marshall's global empirical Bayes estimator of the relative risk. He assumes that the relative risks θ_i have a common mean μ and variance σ^2, but they do not follow any particular distribution. The method of moment estimator is used to derive an estimator of the relative risk as

$$\hat{\theta}_i = \hat{\mu} + C_i(SMR_i - \hat{\mu}) \tag{3.5}$$

where

$$\hat{\mu} = \frac{\sum_{i=1}^{n} O_i}{\sum_{i=1}^{n} E_i}$$

$$C_i = \frac{s^2 - \hat{\mu}/\bar{E}}{s^2 - \hat{\mu}/\overline{E + \hat{\mu}/E_i}}$$

and $\bar{E}$ is the mean of the E_i's and s^2 is the unbiased estimate of the variance of the SMR_i's. The raw SMR estimates are shown in Figures 3-1 and 3-2 for both states.

Figure 3-1: Raw SMR Estimates for California

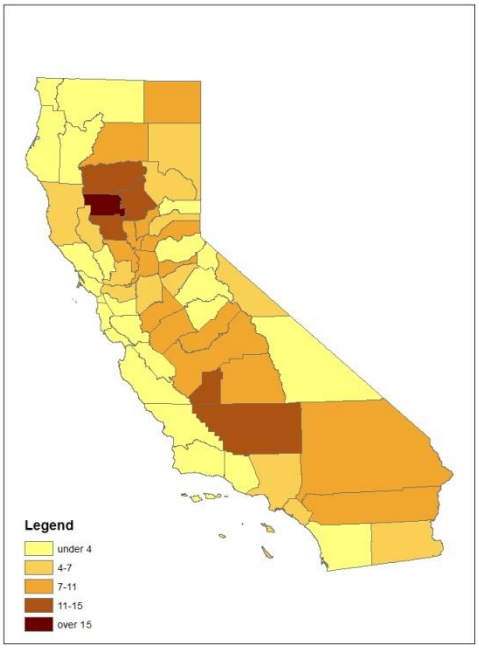


Figure 3-2: Raw SMR Estimates for Colorado

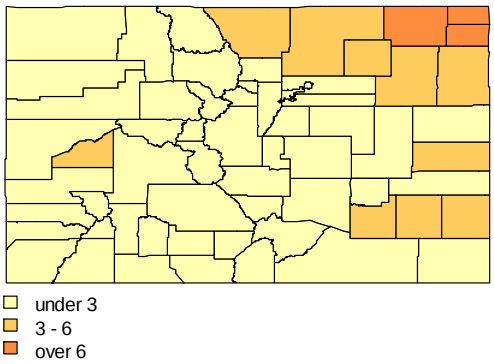


Figure 3-3: Empirical Bayes Poisson-Gamma SMR Estimates for California

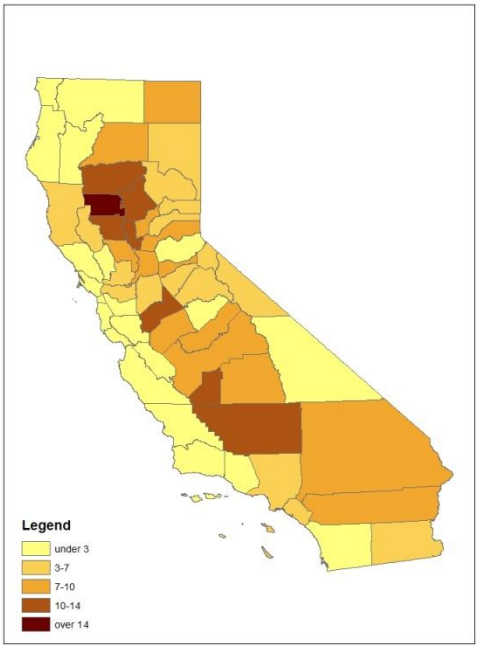


Figure 3-4: Empirical Bayes Poisson-Gamma Estimates for Colorado

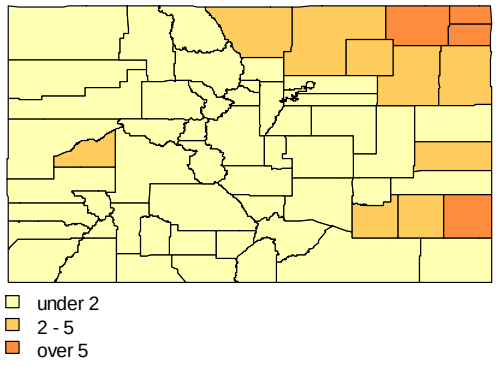


Figure 3-5: Empirical Bayes Log-Normal SMR Estimates for California

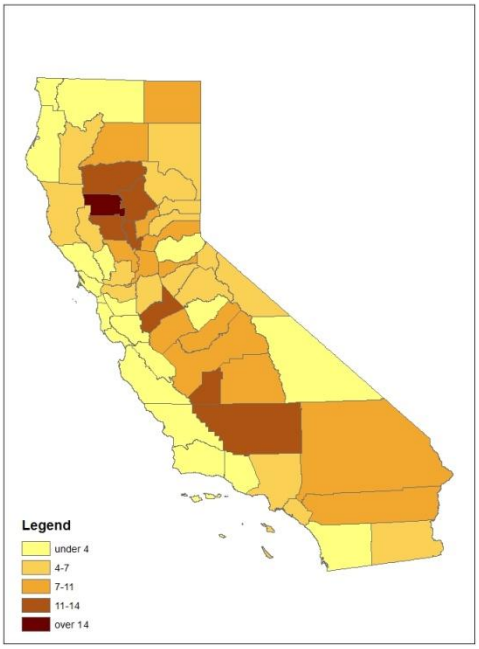


Figure 3-6: Empirical Bayes Log-Normal SMR Estimates for Colorado

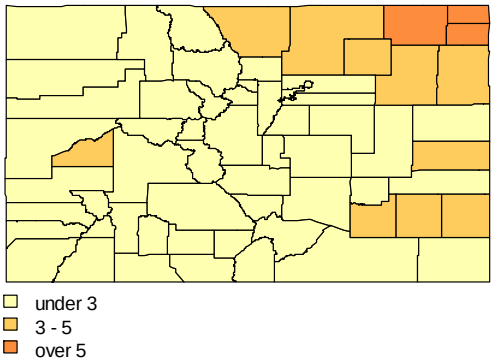


Figure 3-7: Empirical Bayes Global Marshall SMR Estimates for California

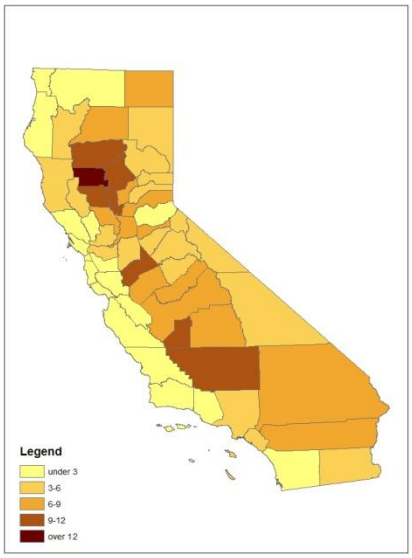


Figure 3-8: Empirical Bayes Global Marshall SMR Estimates for Colorado

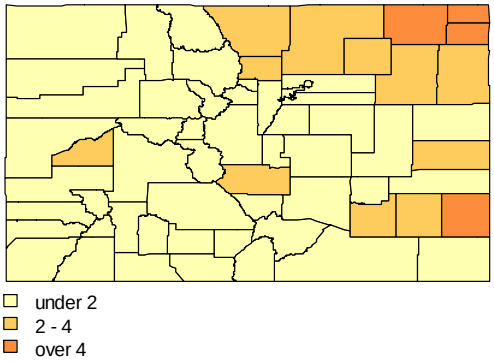


Figure 3-1 shows the raw SMRs each county in California. This map indicates that there is a tendency for the SMR to cluster. Noticeable grouping of counties with moderately high SMRs (between 1.292 and 3.265) occur in the Southern and middle regions of

California. In general most counties in the state have a relatively low SMR with the exception of Glenn county which has a SMR greater than 15. The raw SMR estimates obtained for each county in Colorado are mapped in Figure 3-2. The map shows that counties with similar SMR values tend to be clustered together especially in the Central and North-east portions of the state. Most counties have a SMR between 0.028 and 0.50, with the exception of Logan, Sedgwick, Phillip and counties that show relative risks over 6. The smoothed SMR estimates using the various empirical Bayes methods are shown in Figures 3-3 to 3-8. The smoothed estimates for the most part are consistent with the raw estimates. The spatial patterns shown for both states are similar for those obtained for the raw SMR values.

3.2.2 Predicted Probabilities

Probability maps are increasingly becoming an alternative way to visually depict the risk of a particular disease. Albeit less popular the SMRs, they can provide useful information about the spatial distribution of the probability of death (mortality) or infection (morbidity) from a disease. Probability maps developed by Choyonowski (1959) are a suitable method of visually representing the significance of observed values of disease counts. They show the probability of a count exceeding the observed value given the assumptions we have made about the model. Alternatively, they depict the p-value of the observed number of counts produced by the current model. In this study, the predicted probabilities are generated under the assumption that the observed number of cases follows a negative binomial distribution. In particular, the predicted probabilities generated using the spatial filtering negative binomial model derived in Chapter 2 as:

$$HV_{it} = exp(\beta_0 + \beta_1 INCOME_{it} + \beta_2 FORCLOSE_{it} + \beta_3 MOSQUITO_{it} + \beta_4 POPDENSE_{it} + SFILTER_i\gamma + Z'_{it}\,\theta + u_i) \quad (3.6)$$

The maps of predicted probabilities for each county in California and Colorado are mapped in Figures 3-9 and 3-10 respectively. They show that the values range from 0 to 0.2 for California and 0 to 0.12 for Colorado. Figure 3-9 reveals that California counties with similar predicted probabilities tend to be clustered together. This is especially true in the Northern and Southern regions of the state were counties with predicted probabilities between 0.016 and 0.05 are clustered together. Figure 3-10 on the other hand reveals less spatial clustering of the probabilities in Colorado counties. They are grouped into five classes to make the map more comprehensible. It is worth pointing out that Sacramento County in California and Adams, Routt, Kiowa, Crowley and Otero counties in Colorado had the highest predicted risk in their respective states.

Figure 3-9: Predicted Probabilities for California

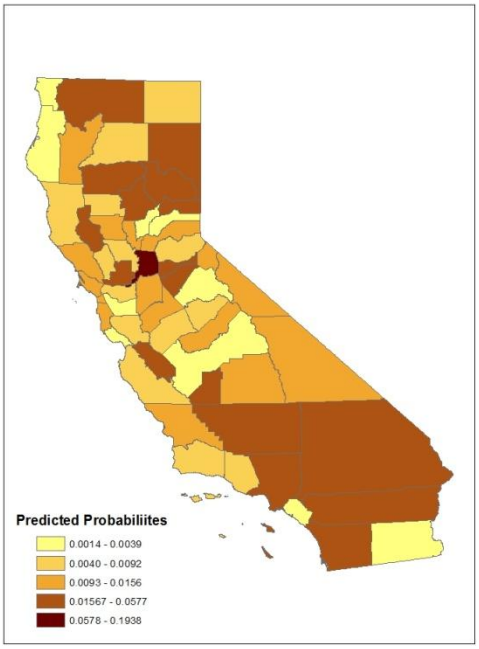
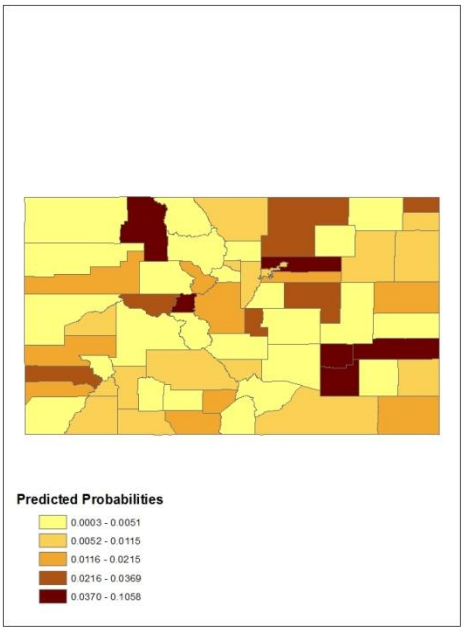


Figure 3-10: Predicted Probabilities for Colorado



3.3 Variogram Analysis

The variogram is a geostatistical tool used to investigate the spatial structure of a particular variable of interest. Mathematically a variogram is defined as:

$$E[Z(\boldsymbol{u}) - Z(\boldsymbol{u} + \boldsymbol{h})]^2 = 2\gamma(\boldsymbol{h}) \quad (3.7)$$

Following geostatistical convention $\boldsymbol{u}$ is a vector of spatial coordinates, $Z(\boldsymbol{u})$ is a variable that is dependent on spatial location, $\boldsymbol{h}$ is a vector of separation distances (lag vector) measured in kilometers and $Z(\boldsymbol{u} + \boldsymbol{h})$ refers to the lag of $Z(\boldsymbol{u})$. If h is replaced by $|h|$ then spatial autocorrelation is assumed to be isotropic so this phenomenon behaves the same way in all directions, otherwise spatial autocorrelation is anisotropic. It is further assumed that the mean of the process is stationary. This is mathematically represented as:

$$E[Z(\boldsymbol{u})] = \mu \tag{3.8}$$

If equations (3.7) and (3.8) are satisfied, the process is said to be intrinsically stationary. Clearly, the variogram depends on both the magnitude and direction of the separation distance. The variogram is estimated from the sample data using the formula:

$$\widehat{\gamma(\boldsymbol{h})} = \frac{1}{2N(\boldsymbol{h})} \sum_{i=1}^{N(\boldsymbol{h})} [Z(\boldsymbol{u}_i + \boldsymbol{h}) - Z(\boldsymbol{u}_i)]^2 \tag{3.9}$$

where $N(\boldsymbol{h})$ is the number of sample data pairs separated by lag h and all other variables are as previously defined. The coordinates of the spatial locations represent the centroids of each county and the lags are measured in kilometers.

The empirical variogram has three components namely the sill, range and nugget. The sill is the semivariance value at which the variogram tapers off. The range of a variogram is the distance or lag at which correlations are effectively zero. Alternatively, it is the lag distance at which the variogram reaches the sill value and beyond the range autocorrelations are effectively zero. Finally, the nugget refers to the value of the variogram for lag distances close to zero. It is essentially the intercept value of the variogram. If the variogram is a positive constant for all lag distances, then the variables separated by a lag are uncorrelated and there is no spatial structure or spatial autocorrelation. In other words, if the variogram is a horizontal line for all values of h, then spatial autocorrelation does not exist.

The empirical variogram formula presented so far only provides a useful conceptual or theoretical framework for studying variograms and is not adequate to statistically estimate it. In order to obtain estimates of the variogram, explicit parametric

models must be used. The most commonly used parametric models used in fitting variograms are the exponential, spherical and Gaussian models. The spherical model is the simplest and widely used model in modeling variograms. The spherical model assumes that correlations are zero at very large distances. This is an unrealistic assumption given that some degree of correlation will exist even at large distances. It is represented by the following equation:

$$\gamma(h;r,s,a)=\begin{cases}0, & h=0\\ a+(s-a)\left(\frac{3h}{2r}-\frac{h^3}{2r^3}\right), & 0<h\le r\\ \sigma^2, & h>r\end{cases} \tag{3.10}$$

where $r, s > 0$ and $a \geq 0$ are parameters and $s \geq a$.

The exponential model overcomes the weaknesses of the spherical model and is defined as

$$\gamma(h;r,s,a)=\begin{cases}0, & h=0\\ a+(s-a)(1-\exp(-3hr), & h>0\end{cases} \tag{3.11}$$

The Gaussian model is represented by the formula:

$$g(h)=c\left(1-\exp\left(\frac{3\mathrm{h}^2}{\mathrm{a}^2}\right)\right) \tag{3.12}$$

3.3.1 Observed Human West Nile Virus

To verify that observed human WNV counts have a tendency to cluster in both states, variograms are estimated and fitted to investigate this phenomenon. The variograms fitted for observed human WNV are shown in Figures 3-11 to 3-14 for California and Colorado.

Figure 3-11: West Nile Virus Isotropic Variogram for California

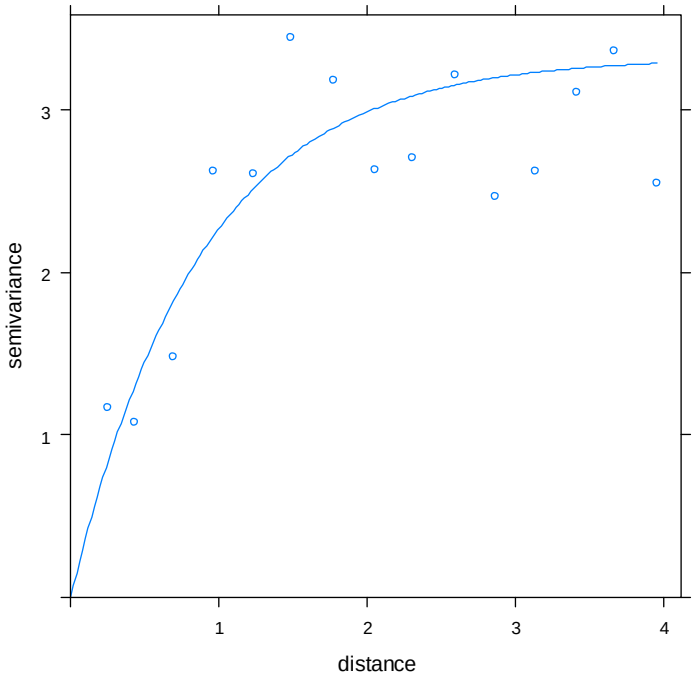


Figure 3-12: West Nile Virus Isotropic Variogram for Colorado

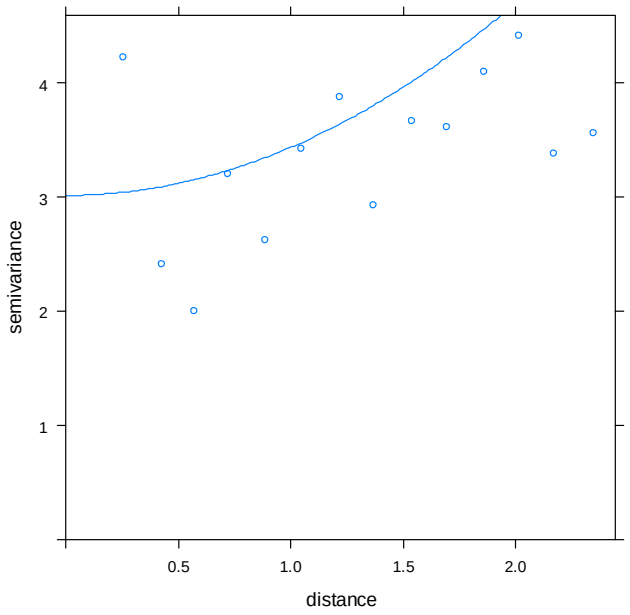


Figure 3-13: West Nile Virus Anisotropic Variogram for California

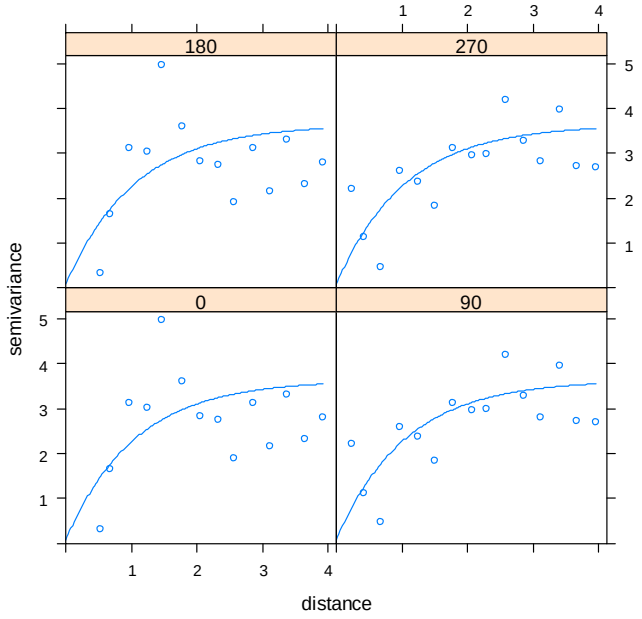
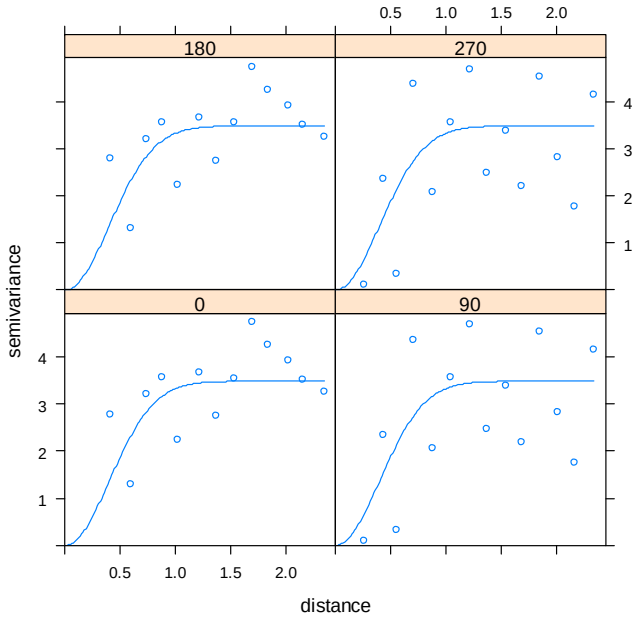


Figure 3-14: West Nile Virus Anisotropic Variogram for Colorado



They were fitted using exponential models to quantify the strength of spatial autocorrelation. Figures 3-11 and 3-12 depict the case where the variogram plot is omnidirectional or isotropic. They show that human WNV has a strong spatial structure in both states. The semivariance increases as the lag or separation distance increases. This is indicative of a strong spatial structure in WNV. To investigate whether different spatial autocorrelation structures exist in different directions, anisotropic variograms are developed. Spatial autocorrelation is examined along the principal cardinal points 0^o (North), 90^o (East), 180^o (South) and 270^o (West). These variograms are plotted in Figures 3-11 and3-12 and they do not show overwhelming evidence of anisotropy.

3.3.2 Standardized Morbidity Ratio

In Figures 3-15 and 3-16, isotropic variograms were fitted to investigate the spatial structure of the relative risk (SMR) in both states. They show that the semivariance rises steadily with the lag value. This suggests that counties with similar values of SMR tend to cluster or be close to each other, while counties with dissimilar SMR values are farther from each other. The presence of this phenomenon indicates a strong spatial trend in this variable. The directional variograms are plotted in Figures 3-17 and 3-18 do not show any evidence that the spatial structure of the SMRs is different along the four principal axis (North, South, East and West).

Figure 3-15: Standardized Morbidity Ratio Isotropic Variogram for California

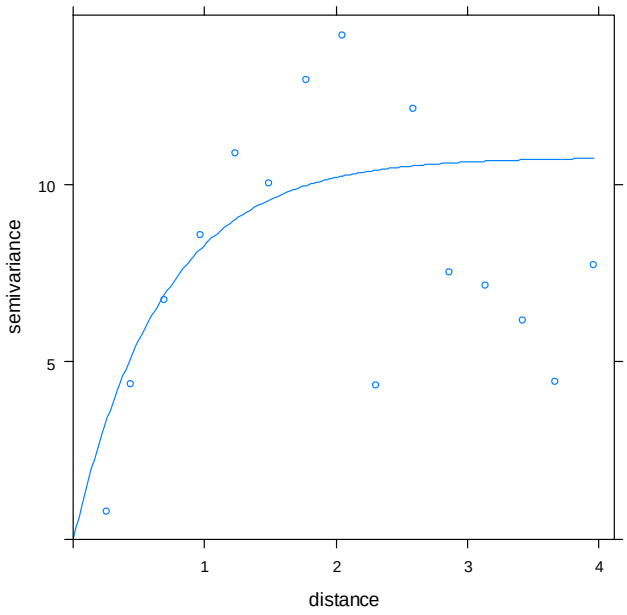


Figure 3-16: Standardized Morbidity Ratio Isotropic Variogram for Colorado

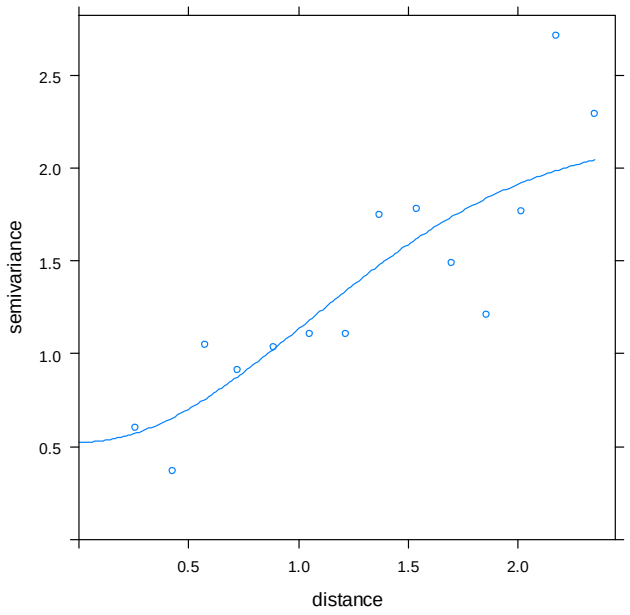


Figure 3-17: Standardized Morbidity Ratio Anisotropic Variogram for California

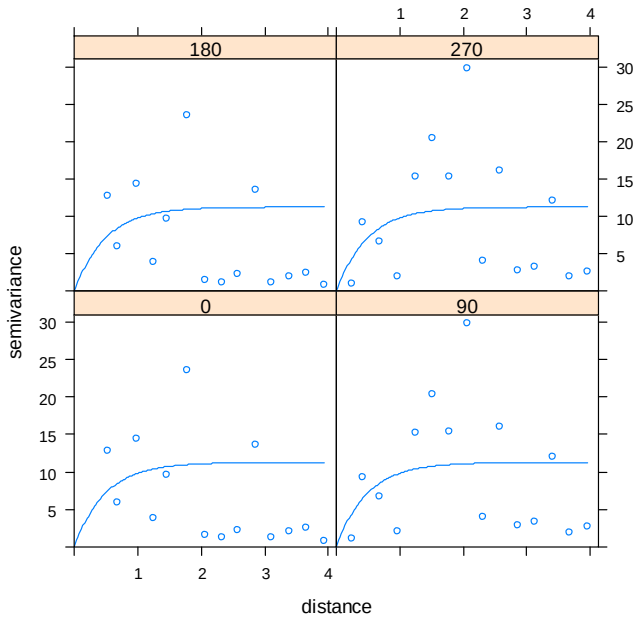
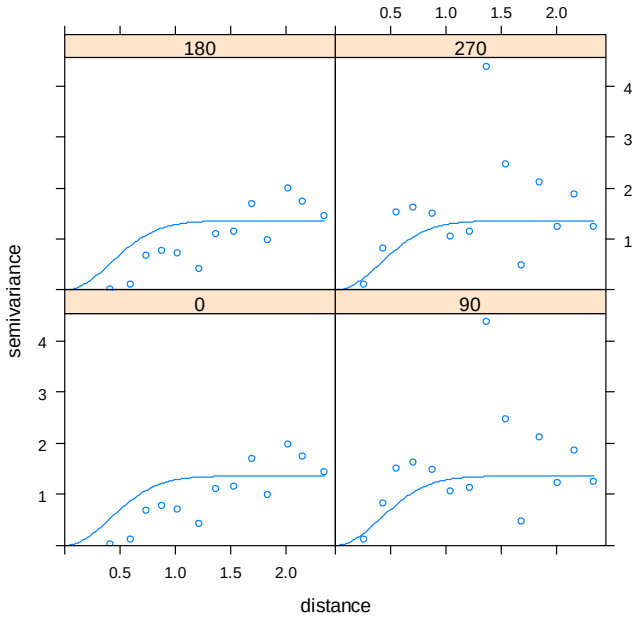


Figure 3-18: Standardized Morbidity Ratio Anisotropic Variogram for Colorado



3.3.3 Model Residuals

To verify that the residuals from the spatial filtering random effects negative binomial model are devoid of spatial autocorrelation, isotropic variograms are fitted for both states and plotted in Figures 3-19 and 3-20. They show that the semivariance exhibits pure spatial independence and is not a function of the lag distance. In other words, the empirical semivariance of the model residuals reveals that the spatial pattern present in the observed WNV counts was removed by the spatial filtering RENB model which was also the most parsimonious and preferred model. Thus the range of the empirical variogram is the same regardless of the lag distance. This result is not surprising at all because the spatial filtering model by construction removes the spatial component from the residuals. The directional (anisotropic) variograms shown in Figures 3-21 and 3-22 also reveal the absence of spatial autocorrelation and do not indicate that spatial autocorrelation is different along the four cardinal points.

Figure 3-19: Model Residuals Isotropic Variogram for Colorado

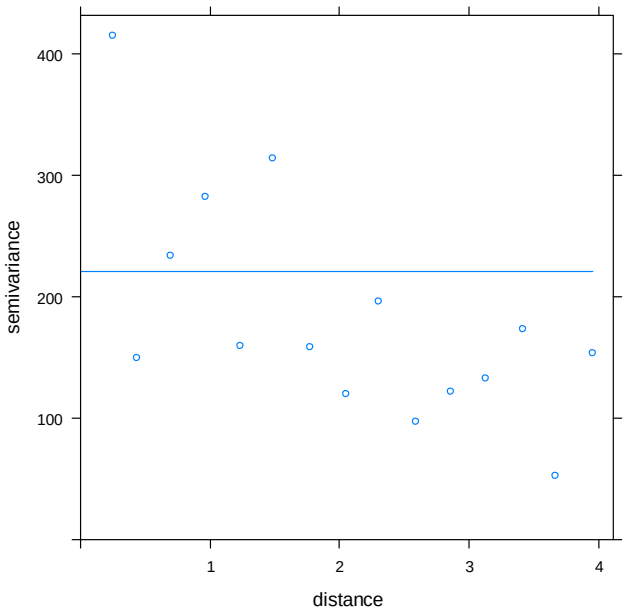


Figure 3-20: Model Residuals Isotropic Variogram for Colorado

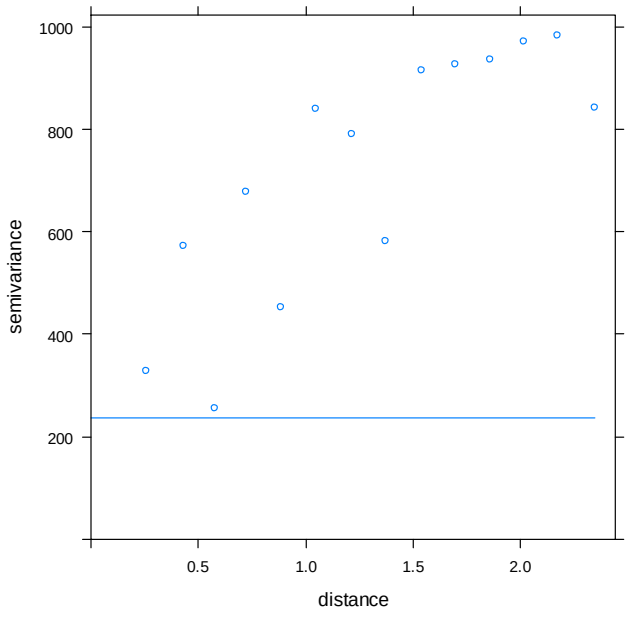


Figure 3-21: Model Residuals Anisotropic Variogram for California

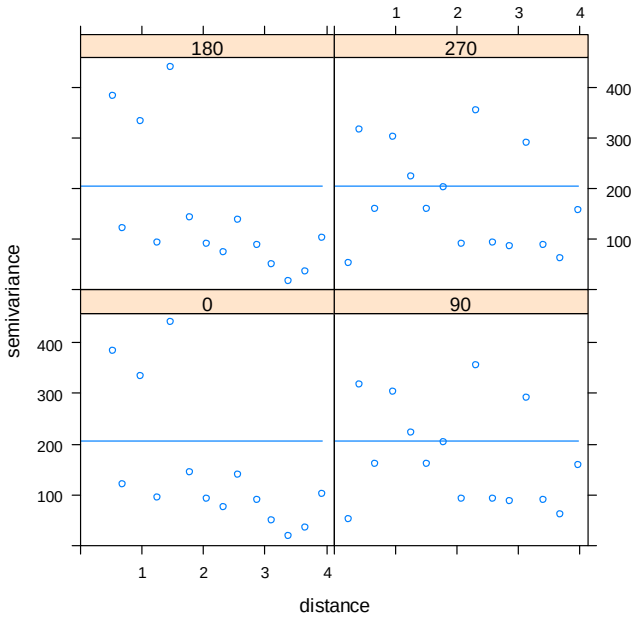
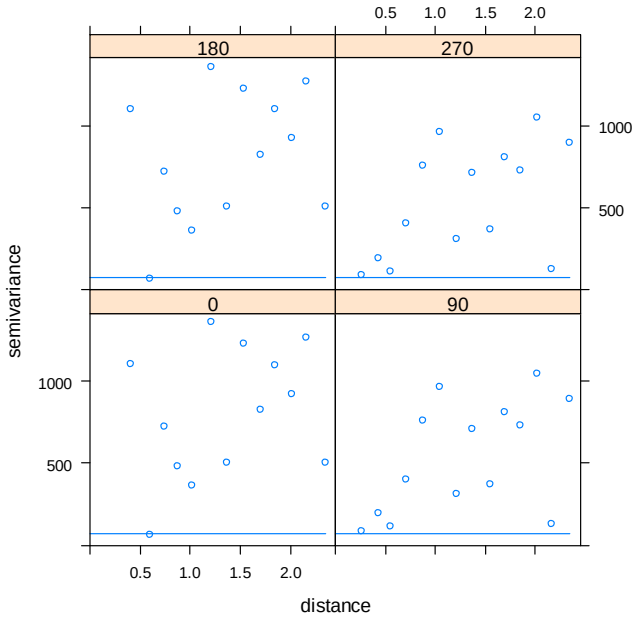


Figure 3-22: Model Residuals Anisotropic Variogram for Colorado



3.4 Discussion and Conclusions

The findings of this chapter reveal some important facts. The predicted probability maps show that Sacramento county in California and Crowley, Kiowa and Otero counties in Colorado have extremely high values. These can be classified as high-risk or high priority areas by health authorities. The standardized morbidity ratios in California were much higher than those of Colorado. The maps of the relative risks show that about 50% of counties in California do have a SMR greater than 1. This suggests that most counties face a greater relative risk of human WNV. Glenn county in California had the highest SMR (16.93), while Monterey county had the lowest SMR (0.012) in the state. In Colorado on the other hand, less than 50% of the counties had a relative risk in excess of 1. Phillips and Sedgwick counties had SMRs in excess of 7 which were the highest in the state. These maps can assist policy makers in the formulation of intervention strategies to control the disease. Adequate resources should be allocated to counties with high predicted probabilities and relative risk values for disease monitoring and surveillance because they represent high-risk areas. The variograms constructed indicate that there is a strong spatial presence in the distribution of the observed human WNV counts and SMRs. However, they do show that the model residuals are spatially independent. They also reveal that the spatial structure of human WNV counts and the SMRs are independent of the direction in which the variograms are constructed. This study provides an example of how GIS mapping tools can be combined with econometric methods to produce hazard maps of counties at risk of human WNV infections. These maps can assist health officials in the formulation of long-term mitigation plans and in the

development of least-cost routes of delivering medical supplies to high-risk areas in the event of an outbreak of the disease.

Chapter 4: Economics of Antibiotic Resistance: Impact of Animal Antibiotic Use on the Evolution of Antibiotic Resistance in Humans

4.1 Introduction

As with numerous other countries and regions, there is an increasing concern over the emergence of drug-resistant strains of bacteria in the United States (US) (Seechi and Babcock 2002). The importance of this "antibiotic resistance crisis" at both the national and international levels is seen through the resulting decline in antibiotic effectiveness (antithesis of antibiotic resistance). Some of the contributory factors hypothesized to be responsible for the crisis in the US emanated from the use of antibiotics in animals and over-consumption of antibiotics in humans (Levy 1992; Smith et al. 2002; Graham et al. 2007). A prominent feature of this crisis was the inappropriate use of antibiotics in animals for growth promotion and prophylaxis (APUA 2012; Smith et al. 2002). The reduction in the effectiveness of antibiotics is a growing concern and the public health consequence is that it raises the economic costs to society. Costs associated with antibiotic resistance include higher mortality rates, longer hospital stays and the necessity to use different or more expensive antibiotics (Sihapi 2008; Bishit et al. 2009). Antibiotics are a precious resource because they are used to treat infectious diseases in humans. It is therefore imperative that interventions be promulgated to prolong the effectiveness of these drugs. In this regard, a number of measures have been proposed and tried with varying success. They include the use of vaccines, controls or reductions of antibiotic use or recycling of different antibiotics (APUA 2012; Bonhoeffer et al. 1997).

The study of antibiotic use in the animal industry in the US is critical for several reasons. First, animal antibiotic use accounts for about 70% of all antibiotics used in the

US (Grace Communication Foundation 2011; Graham et al. 2007; Seechi and Babcock 2002; Smith et al. 2002; Mellon etal. 2001). Second, prescriptions are not required for the use of some antibiotics in the animal industry. A vital link between animal antibiotic use and the emergence of resistance in humans is often conjectured or hypothesized (Miller et al. 2002; Seechi and Babcock 2002) because of these issues. This hypothesis is reflected and reinforced by recent studies that show that there are many pathways through which resistant bacteria can be transferred to humans. These include farm workers (US GAO 2004; Smith et al. 2002; Simonsen et al. 1998), farm produce (US DHHS 2007; Chadwick et al. 1996;) and soil and water sources (US GAO 2004; Chaplin et al. 2004; Chee-Sanford et al. 2001). The foregoing discussion underscores the urgent need to address the problem of animal antibiotic use.

Several mathematical models have been developed by researchers to evaluate the extent to which bacteria (commensal and non-commensal) could be transferred to humans and the rate at which they could contribute to reduced antibiotic effectiveness. Select examples include the works of Miller et al. (2002) and Seechi and Babcock (2002). Notwithstanding these efforts, there is an apparent absence of models to study these issues within an epidemiological model of infectious diseases. It is imperative that a valid framework that can generate reliable predictions concerning antibiotic effectiveness be pursued.

This chapter develops a dynamic bio-economic model to study the vital link between animal antibiotic use and antibiotic resistance in humans. It contributes to the economics literature on antibiotic resistance in three ways. First, it investigates the issue of animal antibiotic use within an epidemiological model of infectious diseases known as

the susceptible-infected-susceptible (SIS) model. Second, it uses the results from the SIS model to generate predictions about the time path of antibiotic effectiveness. Third, it considers the dynamic interaction between animal antibiotic use and antibiotic effectiveness in humans. The specific objectives of this chapter are to use simulation methods to: (1) examine the impact of animal antibiotic use on the trajectory of antibiotic effectiveness; and (2) examine the impact of animal antibiotic use on the optimal time path of the shadow value of antibiotic effectiveness. This model is proposed as one possible framework that a decision maker can use to formulate long-run antibiotic use policies. Clearly, there is no singular model that will provide a panacea for all the problems of antibiotic resistance because of the complexity of the issue. It can however be argued that in the present context the proposed model is suitable in addressing the issue at stake. This conviction is supported by the results of this study which suggest that the predictions of the model are consistent with the stylized facts on antibiotic resistance. The main argument of this chapter is that the extensive use of antibiotics in animal production causes bacteria to evolve resistance which can then be transferred to humans through several channels or pathways.

The organization of the reminder of this chapter is as follows. The next section presents a survey of the relevant literature and the gaps/voids in the literature are also highlighted. In section 3, the use of antibiotics in the animal industry is discussed in detail where the relative benefits and costs (human health risks) of antibiotic use are also presented. The biological mechanisms responsible for the evolution of antibiotic resistance are discussed in section 4. The SIS model of infectious diseases, which forms the basis of the resistance model in this chapter, is presented in detail in section 5. Section

6 develops a dynamic mathematical model to study the impact of animal antibiotic use on the evolution of resistance in humans. The model developed analyzes the role of antibiotic use in humans and animals in the determination of antibiotic resistance. Section 7 presents an optimal control model to analyze the trade-offs between current antibiotic use in both human medicine and animal production and future antibiotic effectiveness. The discussion also determines the economic conditions that characterize the optimal use of antibiotics in human and animals. In section 8, the mechanics of deriving the singular optimal control for both human and animal uses are presented. The results of the simulation analyses are discussed in section 9. In section 10, a discussion of the results and the concluding remarks are presented.

4.2 Background Literature

Hueth and Regev (1974) were the first to model the resistance problem. They employed a pest management framework with a view to illustrating the similarities between the economics of pest resistance and the theory of exhaustible resources in economics. In their seminal paper, Hueth and Regev modeled the resistance problem using a discrete time capital theoretic approach. The decision maker (planner) is assumed to make choice decisions over chemical pest and non-pest inputs. The planner's problem is to maximize a net benefit function subject to the equations of motion of the system relating to potential plant product, pest population density and the stock of pest susceptibility. They concluded that pesticide use is optimal when the marginal profits are equated to the shadow value of pest susceptibility. Laxminarayan and Brown (2001) use an epidemiological model known as the susceptible-infected-susceptible (SIS) model due to Kermack and McKendrick (1927) to address the resistance problem. A continuous time

optimal control model is developed and nets benefits are maximized by choosing the fraction of infected people treated with two different antibiotics subject to equations of motion for antibiotic effectiveness and infections. The authors conclude that in a hospital setting where two antibiotics are used it makes more sense to use the least costly antibiotic first. Their simulation results revealed that antibiotic effectiveness decreases over time i.e. antibiotic resistance increases over time.

Turning to the issue of modeling the relationship between animal antibiotic use and the development of resistance in animals, there seem to be some research in this area albeit under-studied. A review of the literature indicates that Brown and Layton (1996) were the first to model this relationship. In their seminal paper "Resistance economics: social cost and the evolution of antibiotic resistance" they used general functions to illustrate the externality that arises when farmers and individuals make decisions regarding antibiotic use independently. Their idea was to model the effect of animal antibiotic use on drug effectiveness (bacteria susceptibility). In their model both individuals and farmers maximize private net benefits given a common pool of resistance. The conditions that characterize the solution to the individual's and farmer's problems indicate that too much antibiotics are consumed over time. The social planner's approach internalized the externality by maximizing society's net benefit function given a certain level of resistance. This approach yields the conventional result that antibiotic consumption is higher when individuals and farmers make private decisions than when a social planner makes decisions. In other words, bacteria susceptibility to antibiotics is higher when a social planner makes decisions than when decisions are made by private agents. McNamara and Miller (2002) also modeled the externality associated with

antibiotic use by both humans and animals. They modeled the externality problem within a static framework using utility maximization and profit maximization models. The consumer maximizes utility by choosing a combination of antibiotics and food subject to a budget constraint. The animal producer on the other hand chooses antibiotics and grains to maximize her profits subject to the neoclassical production function used to produce food for humans. The problem is that both consumers and producers do not take into account the social cost of their actions. The social planner's program takes into account the externality created by both individuals and producers making decisions independently based on private costs and benefits. The social planner will maximize net benefits to society as a whole by choosing plans for consumers and producers simultaneously. The conditions that characterize the optimal solution to the planner's problem imply that the use of antibiotics will be much lower than when consumers and producers make decisions independently. In their model on the evolution of human commensal bacteria, Smith et al. (2002) illustrated the medical consequences of antibiotic use in animals. To illustrate the relationship between animal antibiotic use and resistance in humans, their models are based on the use of Vancomycin in hospitals and the use of Avoparcin in the agricultural industry for growth promotion. They asserted that Vancomycin and Avoparcin are different names for the same antibiotic. Hence pathogens resistant to Vancomycin should also be resistant to Avoparcin. They use simulation exercises to demonstrate the potential health consequences of antibiotic use in agriculture. The results revealed that the rate of antibiotic resistance in humans to be much higher with antibiotic use in animals. In addition, the increased use of antibiotics in humans for medical purposes contributed to the increased rate of antibiotic resistance in humans. The authors

conclude that drug effectiveness in humans can be prolonged by delaying the use of antibiotics in animals. Secchi and Babcock (2002) developed an optimal control model to study the trade-offs between the current use of antibiotics in humans and animals and the future use of antibiotics in both humans and animals. In particular, they were interested in how antibiotics used as growth promoters affect drug efficacy (bacteria susceptibility). In their model, the utility that individuals derive from their health state depends on net income and the effectiveness of the antibiotic. The evolution of resistance is posited to depend on the proportion of animals given antibiotics, individual use of antibiotics, the proportion of humans treated, and the rate at which resistance is transmitted from animals to humans. The social planner maximizes the discounted utility of the treated and untreated overtime. Their findings suggest that animal antibiotic use contributed to increased resistance in humans. Their results are also consistent with standard economic principle that animal antibiotic use is warranted if the net marginal benefits exceed the marginal costs. In addition, the results support the stylized fact that antibiotic effectiveness decreases over time. Furthermore, the findings showed that human use of antibiotics partly contributed to increased resistance.

4.2.1 Literature Synthesis

The literature review revealed a couple of important facts and some gaps. The role of animal antibiotic use has been acknowledged as a significant contributory factor in the evolution of antibiotic resistance in humans. In addition, a number of studies (Brown and Layton 1996; Secchi and Bacock 2002; McNamara and Smith 2002) have modeled explicitly how animal antibiotic use impacts the development of antibiotic resistance in humans. However, some specific issues seem to be under-studied in the literature. First,

the relationship between animal antibiotic use and the evolution of antibiotic resistance in humans has seldom been studied within an epidemiological model of infectious diseases such as the SIS model. This model was developed by Kermack and McKendrick (1927) to study the evolution of antibiotic effectiveness and human infections and would be an ideal candidate for the current investigation. Second, the dynamic interaction between animal antibiotic use and antibiotic effectiveness seems to be under-investigated. The objective of this research is to fill in gaps in the present literature.

4.3 Animal Antibiotic Use

This purpose of this section is to put into perspective the benefits and human health hazards that result from the use of antibiotics in the animal industry.

4.3.1 Benefits of Using Antibiotics in Animal Production

Antibiotics are used in the animal industry for non-therapeutic/sub-therapeutic purposes (growth promotion and prophylaxis) and therapeutic purposes (treatment of diseases). A large proportion of antibiotics used in the animal industry are for growth promotion and prophylactic purposes. Antibiotics are used in animal production because they confer several benefits to producers and society as a whole. First, they are used to maintain the health and welfare of animals (Ziv 1986; Gustafson and Bowen 1997; Levitt 2011; Grace Communications Foundation 2011). In this regard, antibiotics are administered to improve the gastrointestinal tract and absorptive processes of farm animals, i.e. to improve the internal ecology of the animal (NRC 1999). The antibiotic is applied indiscriminately to both healthy and sick animals. It is usually placed in their feed or drinking water daily and this practice is repeated over a protracted period of time. Second, animal antibiotic use is encouraged because it promotes growth, weight gain and

feed efficiency. This ultimately improves the overall quality of the carcass on the market. Also the economics of agriculture makes it prudent to use antibiotics. From an economic perspective, the therapeutic use of antibiotics to treat a disease or infection is justified. In most cases the economic benefit of treating the animal with antibiotics exceeds the economic costs. The economic benefits are realized in terms of feed efficiency and performance, i.e., the growth rate of the animal. Research shows that the use of antibiotics in agriculture results into an increase in efficiency and performance by up to 15% (NRC 1999). The economic incentives for using antibiotics in agriculture are to maximize profits (minimize cost) so as to reduce the cost of meat to society (CBS 2010). Hence it is cheaper to apply antibiotics to the entire herd for disease prevention and growth promotion than to administer them individually. Lastly, animal antibiotic use is warranted because they fight infectious diseases that can be transmitted to humans. As select examples, the Council for Agricultural Science and Technology (CAST 1981) reported that the use of antibiotics such as chlortetracycline, oxytetracycline and tylosin in the animal industry led to a drastic decline in the incidence of liver abscesses.

4.3.2 Human Health Risks

The health risks associated with the use of antibiotics in animals are many and varied. There are several pathways through which resistant bacteria can be transferred to humans. These include residues of antibiotics on foods, transfer of resistant genes, and zoonotic organisms, transmission by farmer workers and contamination by animal waste. These arguments are seen in Gustafson and Bowen (1997) who argue that these health hazards could be on the rise.

4.3.2.1 Contamination of Food Products

Antibiotics are usually administered through drinking water in poultry and turkey, and administered through injections or feed in cattle. Occasionally, when dairy products such as eggs and milk are packed, they are tainted with small amounts of residues from farms. Resistant bacteria are sometimes present in these residues and humans acquire these resistant bacteria when they consume these products or do not cook them properly (US DHHS 2007). This poses a health hazard to humans and the animal industry could take steps to reduce the residue on food products by banning certain antibiotics.

4.3.2.2 Transmission of Resistant Genes

The main issue of concern is whether human health is compromised by the presence of resistant bacteria in animals. Some evidence has been found to suggest that antibiotic resistance in humans is linked to bacteria resistance in animals (CAST 1981; Smith et al. 2002; Office of Technology Assessment 1995). In an experimental study in Massachusetts, Levy (2002) concluded that a resistant gene found in farm animals was responsible for spreading resistant strains of *Escherichia coli* to farm family members. These *E. coli* persisted in the environment and spread to several animals including humans. A reduction in the volume of antibiotics used in the animal industry could greatly improve human health.

4.3.2.3 Disease Transmission by Farm Workers

Farm workers are directly exposed to a myriad of bacteria in their work environment so they can be an important channel through which resistant bacteria are transferred to humans. The United States General Accounting Office (US GAO 2004) argued that farm workers are sometimes exposed to resistant bacteria that cause them to contract food-

borne illnesses. They can spread these organisms to the general public when they get sick and go to the hospital.

4.3.2.4 Contamination by Animal Waste

A number of authors notably O'Brien (2002) and Summers (2002) have argued that release of farm waste that contaminates the water supply system and the soil also poses a serious environmental problem. These studies demonstrate that there is the potential for a greater degree of interaction between animal and human ecosystems. The use of organic manure also poses a serious health threat and could compound the problem because some bacteria in the feces of farm animals get transferred on to arable lands (Levitt 2011).

4.4 Antibiotic Resistance: A Primer

Antibiotics are a group of drugs that inhibit or decelerate the growth of bacteria (Levy 1998). There are two types of bacteria identified in the medical literature-susceptible (sensitive) bacteria and resistant (insensitive) bacteria. Resistant bacteria have evolved to the point where antibiotics are ineffective against them. Levy (1998) contends that resistant bacteria evolve in such a way that they develop proteins that shield them from attacks by antibiotics. Put simply, they release proteins that render antibiotics ineffective. In some cases resistant strains of bacteria develop enzymes that degrade antibiotics (Levy 1998).

The evolution of antibiotic resistance is caused by three main mechanisms: natural selection, genetic recombination mechanisms, and mutation (Massad et al. 2007; Sipahi 2008; Palumbi 2001; Levy 1998). Natural selection occurs because nature endows all bacteria with the ability to fight antibiotics. This is the classic example of "survival of

the fittest" whereby the least susceptible bacteria (most resistant bacteria) to a given antibiotic multiply rapidly giving rise to a vast number of other resistant bacteria. Resistant bacteria with the least fitness cost are those who survive bombardment with antibiotics. From an economic perspective, fitness cost can be viewed as a type of opportunity cost. This is so because the resistant strains thrive in the presence of antibiotics, but the cost of not using antibiotics is that they perish rapidly. Mutation is also another mechanism through which bacteria can evolve. Mutations can lead to the development of new traits or the reinforcement of ones already present (Levy 1998). Another way in which bacteria can acquire resistant genes is through the transfer of plasmids. Plasmids are small strands of DNA that assist bacteria in surviving a harsh or unfavorable environment. There are several examples of bacteria that have evolved to become resistant. *Staphylococcus aureus* (bacteria responsible for staph infections), which once responded well to Methicillin and Vancomycin, is now often resistant to both antibiotics. Strains of *Mycobacterium tuberculosis* (the most common cause of tuberculosis) have developed resistance to antibiotics that were previously successful against it. *Escherichia coli*, which was once treated successfully with Tetracycline is now often resistant to it.

4.5 An Epidemiological Model of Infectious Diseases

The fundamental model used in the analysis of infectious diseases is the Susceptible-Infected-Susceptible -SIS Model (Kermack and McKendrick 1927). It is also the key building block of the bio-economic model presented in this chapter. The model describes the process by which the population moves between the susceptible and infected stages via infection and treatment. The model assumes that the total population (N) can be

partitioned into two sub-populations namely susceptible (S) – the population that is in good health, but susceptible to infection- and the infected population (I). The model presented in this paper makes the following assumptions: (1) the force of infection or the rate of transmission (β) determines the rate at which the population moves from S to I; (2) the human population can be infected by two strains of bacteria – one that is sensitive to the antibiotic and one that is resistant; (3) the proportion of the population who are infected with the sensitive strain of bacteria are cured faster through antibiotic treatment, while those infected with a resistant strain however recover at a less rapid rate; (4) immunity of an individual is not allowed so an individual can become susceptible to an infection after treatment; (5) super-infections are not permitted so an individual is not susceptible to any secondary infection; (6) resistance has been introduced in advance so that a small number of individuals carry the resistant strain; (7) only one antibiotic is prescribed to treat infections. Therefore, at any time t, $\mathrm{N_t = S_t + I_t = S_t + I_w + I_r}$ where S_t represents the uninfected fraction of the population, I_t denotes the fraction of the population infected, I_w denotes the fraction of the population infected with the sensitive type, and I_r is the proportion of the population infected with the drug-resistant strain. Since some of the members of the uninfected population will become infected through contagion, the SIS model assumes that the term $\mathrm{\beta S_t I_t}$ incorporates this idea. Assuming a constant population, the susceptible sub-population dynamics is described by the differential equation

$$\frac{\mathrm{dS}}{\mathrm{dt}} = -\beta \mathrm{S}(\mathrm{I_w} + \mathrm{I_r}) + \mathrm{r_w I_w} + \mathrm{r_r I_r} + \mathrm{f I_w} = \dot{S} \qquad (4.1)$$

where, $\beta S(I_w + I_r)$ is the rate of addition to the infected population via both the resistant and sensitive strains, $r_w I_w$ is the total number of those infected with the sensitive strain that recover naturally, $r_r I_r$ is the total number of those infected with the resistant strain that recover naturally, fI_w is the total number of those infected with the sensitive strain that are treated with antibiotics, f is the fraction of the infected population treated with a single antibiotic, r_w is the rate of recovery of the sensitive strain in the absence of treatment, and r_r is the rate of recovery of the resistant strain in the absence of antibiotics. An inherent assumption is that the populations infected by both sensitive and resistant strains have equal access to treatment. It is worth pointing out that N_t can be normalized to 1 so that $I + S = 1$ and $dI/dt = -dS/dt$. It is appropriate that fitness cost be introduced into the analysis. Fitness cost measures the idea that the resistant strain survives in the presence of antibiotics so depriving them of antibiotics represents a biological cost and is measured as $\Delta r = r_r - r_w$. If the resistance bacteria perish at a much faster rate than the sensitive bacteria, the fitness cost is positive and effectiveness is considered as a renewable resource. On the other hand, if both the resistant strain and the sensitive strain clear at the same rate the fitness cost is zero. So effectiveness is considered as a non-renewable resource. The evolution of the sub-populations infected with the sensitive and resistant strains are described by the differential equations

$$\frac{dI_w}{dt} = \beta S I_w - r_w I_w - f I_w \tag{4.2}$$

$$\frac{dI_r}{dt} = \beta S I_r - r_r I_r \tag{4.3}$$

and antibiotic effectiveness is expressed by the fraction

$$w = \frac{I_w}{I} = \frac{\mathrm{I_w}}{\mathrm{I_r} + \mathrm{I_w}} \tag{4.4}$$

In other words, w is the ratio of those infected with the sensitive strain to the total number of infections. Intuitively, if the antibiotic is effective in curing the disease this fraction should be higher. Combining equations (4.1)-(4.4), the two fundamental equations of the SIS model can be written as

$$\frac{dI}{dt} = \frac{dI_w}{dt} + \frac{dI_r}{dt} = (\beta S + w\Delta r - r_r - wf)I = \dot{I} \tag{4.5}$$

$$\frac{dw}{dt} = (f - \Delta r)w(w - 1) = \dot{w} \tag{4.6}$$

These two equations describe the evolution of the infected sub-population and the effectiveness of the antibiotic over time. In this study, effectiveness is being considered as a non-renewable natural resource because $\Delta r = 0$. Consequently, the rate of recovery of an infected person from susceptible strain is identical to that of a resistant strain (i.e. $r_w = r_r = r$). It can be implied from equation (4.6) that the antibiotic effectiveness decreases over time which is consistent with the stylized facts on resistance.

4.6 Mathematical Model

The model presented in this section develops a framework to assess the impact of animal antibiotic use on the evolution of antibiotic resistance in humans. This is an adaptation of the SIS model developed to study the evolution of infectious diseases in humans. This model is modified to handle a potential link between animal antibiotic use and the emergence of antibiotic resistance in humans. Compared to the SIS model there are some

principal differences with this resistance model. First, the term $ku_a(t) + zu_h(t)$ (explained below) is included in the effectiveness equation to account for the potential transmission of resistance bacteria from animals to humans. This specification is consistent with those of Brown and Layton (1996) and McNamara and Miller (2002) in the sense that the rate at which antibiotic effectiveness evolves over time is linear in both human and animal antibiotic use. Second, the term u_h replaces f in the infections equation and these two terms are mathematically equivalent. This notation is used in order to be consistent with the notational scheme of section 4.7. The two fundamental equations of the model that describe the evolution of effectiveness and infections are modifications of (4.5) and (4.6) and presented as follows:

$$\frac{dw}{dt} = w(t)\big(ku_a(t) + zu_h(t)\big)(w(t) - 1) = \dot{w} \tag{4.7}$$

$$\frac{dI}{dt} = [\beta S(t) - r - w(t)u_h(t)]I(t) = \dot{I} \tag{4.8}$$

where k is the marginal contribution of animal antibiotic use to human resistance, z is the marginal contribution of human antibiotic use to human resistance, $u_h(t)$ is the fraction of the infected human population treated with antibiotics, $u_a(t)$ is the tonnage of antibiotics used in the animal industry, and all other variables are as previously defined. If $k = 0$, then animal antibiotic use has no impact on drug effectiveness and only human use impacts effectiveness. On the other hand if $k > 0$, then animal antibiotic use impacts antibiotic effectiveness and the magnitude of u_a now becomes an important policy variable. Because k is now an important determinant of policy, policy makers must determine the optimal u_a or the conditions that characterize the optimal use of u_a.

4.7 Economics of Antibiotic Resistance

The objective of this section is to model the antibiotic resistance problem as an exhaustible resource problem. An optimal control model is subsequently developed to determine the optimal use of antibiotics in both animals and humans.

4.7.1 Antibiotic Resistance as an Exhaustible Resource

Antibiotic resistance and exhaustible resources in economics share some salient features. First, it is common to conceptualize antibiotic effectiveness as a large pool (reserve) of a valuable resource such as oil (Brown and Layton 1996). Every time antibiotics are used (extracted), it lowers the level of the reserve by a small amount. This raises the cost of using antibiotics for subsequent doses prescribed. This process will continue for the next dose prescribed and subsequent doses will continue to lower the level of the reserve or pool of effectiveness. Another way of modeling this problem is to consider a common pool of bacteria that are capable of becoming resistant. As these bacteria populations are assaulted with antibiotics they evolve resistance. This means that the next dose prescribed will be more expensive. This process will continue making subsequent doses more and more expensive to prescribe. To put things in perspective, increasing use of antibiotics leads to increasing resistance which translates into rising treatment costs. It is also possible to conceive of a situation where the cost of treatment is so prohibitive that antibiotics are no longer useful. Second, antibiotics are quite unusual in the sense that their current use jeopardizes their future effectiveness because bacteria evolve resistance to them. In other words, there is a trade-off between current antibiotic use and future drug effectiveness. Third, the decision to use more of the resource today has two effects. The use of the resource influences the rate at which the stock is changing not only today, but

also its availability in subsequent periods of time and the flow of net benefits to society. Finally, the antibiotic resistance is a dynamic problem and the derivation of the conditions that characterize the optimal use of antibiotics necessitates the use of optimal control techniques.

4.7.2 An Economic Approach to Antibiotic Resistance

The ultimate goal of this economic analysis is to formulate a net benefit function ($NB(t)$) and derive conditions that characterize the optimal use of antibiotics. The benefit obtained by humans from the use of each antibiotic can be written as $b_h w(t) u_h(t) N(t) I(t)$, where b_h represents the dollar value per benefit associated with each successful treatment using the antibiotic and all other variables are as previously defined. The total cost incurred by humans from using antibiotics is $c_1 u_h(t) N(t) I(t) + c_2 N(t) I(t)$. The cost of a human treatment is$c_2 N(t) I(t) u_h(t)$, and the cost associated with a human infection is given as $c_1 N(t) I(t)$ and where c_1 is the unit cost of treatment using antibiotics and c_2 is the unit cost of an infection. The benefit obtained by animal producers from the use of each antibiotic is $b_a u_a(t)$, where b_a represents the dollar value per benefit associated with one ton of animal antibiotic use. The cost associated with animal antibiotic use is $c_a u_a(t)$, where c_a is the unit cost per ton of antibiotic use. The state variables associated with this problem are the level of effectiveness and the level of infection. The state equations (equations of motion) are given by equations (4.9) and (4.10).

$$\frac{dw}{dt} = w(t)\big(k u_a(t) + z u_h(t)\big)(w(t) - 1) = \dot{w} \tag{4.9}$$

$$\frac{dI}{dt} = [\beta(1 - I(t)) - r - w(t) u_h(t)] I(t) = \dot{I} \tag{4.10}$$

It is important to note that $S(t)$ has been replaced with $1 - I(t)$ in (4.10) since $N(t) = S(t) + I(t) = 1$. The control variables are the fraction of the human population treated with antibiotics u_h, and the tonnage of antibiotics used in the animal industry, u_a. Mathematically, the resistance problem is formulated as

$$\max_{u_a, u_h} \int_{t=0}^{t=\infty} NB(t) = e^{-\delta t}\{ [b_h w(t) u_h(t) I(t) - c_1 u_h I(t) + c_2 I(t)] + [\, b_a u_a(t) - c_a u_a]\} dt$$

s.t.

$$\frac{dw}{dt} = w(t)\big(k u_a(t) + z u_h(t)\big)(w(t) - 1) = \dot{w} \tag{4.11}$$

$$\frac{dI}{dt} = [\beta(1 - I(t)) - r - w(t) u_h(t)] I(t) = \dot{I} \tag{4.12}$$

Disregarding dependence on time, the current-value Hamiltonian is:

$$\begin{aligned} H &= [(b_h w I - c_1 I) u_h + (b_a - c_a) u_a - c_2 I + \mu_1 w (k u_a + z u_h)(w - 1) + \mu_2 (\beta(1 - I) \\ &\quad - r - w u_h) I] \qquad (4.13) \\ &= \sigma_h u_h + \sigma_a u_a + \mu_2 [\beta(1 - I) - r] - c_2 I \end{aligned}$$

where $\sigma_h = b_h w I - c_1 I + \mu_1 w z (w - 1) - \mu_2 w I$ and $\sigma_a = b_a - c_a + \mu_1 k w (w - 1)$ are called switching functions because they determine whether the controls switch from the upper values to the lower values. The Hamiltonian can be interpreted as follows: the total rate of increase in the value of antibiotic resources is the sum of net returns at time t plus the marginal value of drug effectiveness and the marginal value of infections.

Application of Pontryagin's Maximum Principle yields the following necessary conditions:

$$\frac{\partial H}{\partial u_h} = \sigma_h = b_h wI - c_1 I + \mu_1 zw(w-1) - \mu_2 wI \gtreqless 0 \quad (4.14)$$

$$\frac{\partial H}{\partial u_a} = \sigma_a = b_a - c_a + \mu_1 kw(w-1) \gtreqless 0 \quad (4.15)$$

$$\frac{\partial H}{\partial w} = b_h I u_h + \mu_1 (ku_a + zu_h)(2w-1) - \mu_2 u_h I = \delta\mu_1 - \dot{\mu}_1 \quad (4.16)$$

$$\frac{\partial H}{\partial I} = (b_h w - c_1)u_h - c_2 + \mu_2(\beta - 2\beta I - r - wu_h) = \delta\mu_2 - \dot{\mu}_2 \quad (4.17)$$

where μ_1 and μ_2 are the co-state variables for w and I. The transversality conditions are:

$$\lim_{t\to\infty} \mu_1 we^{-\delta t} = 0 \quad (4.18)$$

$$\lim_{t\to\infty} \mu_2 Ie^{-\delta t} = 0 \quad (4.19)$$

Equation (4.14) can be re-arranged as:

$$b_h wI - c_1 I = \mu_1 zw(1-w) + \mu_2 wI \quad (4.20)$$

The economic interpretation of (4.20) is: the immediate net marginal benefit of changing the fraction of the human population treated with antibiotics equals the long-term marginal cost of changing the fraction of human population treated with antibiotics. The immediate net marginal benefit has two components: the marginal benefit of curing infections ($b_h wI$) and the marginal cost of each infection ($c_1 I$). The marginal cost also has two components: marginal cost of reducing the value of effectiveness $\mu_1 zw(1-w)$ and the marginal cost of increasing infections ($\mu_2 wI$). Note that μ_1 is the shadow value of

a unit of effectiveness and μ_2 is the shadow value of a unit of infections. It is worth noting that $\mu_1 > 0$ since antibiotic effectiveness can be considered as good capital or beneficial, while $\mu_2 < 0$ since infections could be considered as bad capital or harmful. Equation (4.15) can also be re-arranged as:

$$b_a - c_a = \mu_1 wk(1-w) \tag{4.21}$$

The economic interpretation of (4.21) is: the immediate net marginal benefit of changing the tonnage of antibiotics used equals the long-term marginal cost of changing the tonnage of antibiotics used. Note that the marginal cost of decreased effectiveness is $(\mu_1 wk(1-w)$. In other words, the net benefits of animal antibiotic use would have to be counterbalanced by the cost to humans in terms of decreasing drug effectiveness. If $(b_a - c_a) < \mu_1 wk(1-w)$, then a ban on animal antibiotics use could be justified solely on economic grounds. On the other hand, if $b_a - c_a > \mu_1 wk(1-w)$ then animal antibiotic use is justified strictly on economic grounds. The threshold value of w beyond which a ban on animal antibiotic use is justified on economic grounds is derived in Appendix B.

The first-order conditions for maximization of the Hamiltonian provides information that will assist in finding the path of antibiotic effectiveness, infections and the co-state variables. Since the Hamiltonian is linear in the controls, u_a and u_h, the optimal control will be a combination of the Most Rapid Approach Path (bang-bang control) and a singular solution. This is illustrated as follows:

$$u_h = \begin{cases} {u_h}^{max} = 1 & if\ \sigma_h > 0 \\ {u_h}^* \in [0,1] & if \sigma_h = 0 \\ {u_h}^{min} = 0 & if\ \sigma_h < 0 \end{cases} \tag{4.22}$$

$$u_a = \begin{cases} {u_a}^{max} = A & if\ \sigma_a > 0 \\ {u_a}^{*} \in [0, A] & if \sigma_a = 0 \\ {u_a}^{min} = 0 & if\ \sigma_a < 0 \end{cases} \quad (4.23)$$

where A is the tonnage of antibiotics produced in the US in a single year. The optimal paths for the co-state variables from (4.16) and (4.17) are as follows:

$$\dot{\mu}_1 = \frac{\partial \mu_1}{\partial t} = \delta\mu_1 - b_h I u_h - \mu_1(ku_a + zu_h)(2w - 1) + \mu_2 u_h I \quad (4.24)$$

$$\dot{\mu}_2 = \frac{\partial \mu_2}{\partial t} = \delta\mu_2 - (b_h w - c_1)u_h + c_2 - \mu_2(\beta - 2\beta I - r - wu_h) \quad (4.25)$$

Equations (4.24) and (4.25) illustrate how the shadow prices of antibiotic effectiveness and infections evolve over time. These shadow prices could increase or decrease over time depending on whether the levels of antibiotic effectiveness and infections increase or decrease over time.

4.8 Derivation of the Singular Controls

From equation (4.22) ${u_h}^*$ is the optimal trajectory of the fraction of infected people treated with antibiotics that follows the singular path. Along the singular path, the immediate net marginal benefit of using antibiotics in human medicine is counterbalanced by the long-term costs in terms of decreasing drug effectiveness and rising human infections. Likewise from equation (4.23) ${u_a}^*$ is the optimal trajectory of animal antibiotic use and tracks the singular path. Along the singular path, the immediate net benefit of animal antibiotic use is just counterbalanced by the long-term costs in terms of reduced drug effectiveness. Singular solutions (sometimes referred to as singular arcs) arise in optimal control problems in which the Hamiltonian is linear in the control. In such cases, a direct application of Pontryagin's Maximum Principle breaks down and

cannot give a complete solution. Rather the singular solution is found by successively differentiating the switching functions with respect to time until the control explicitly appears in the equation. After this differentiation is performed $\dot{I}$, $\dot{w}$, $\dot{\mu}_1$, and $\dot{\mu}_2$ are substituted for using equations (4.11), (4.12), (4.24) and (4.25).This equation can then be set equal to zero and the control can be solved for. If this equation does not depend on the controls, u_a and u_h, then another differentiation will have to be performed. Because the problem at hand involves two controls (control vector), restrictions are placed on u_a in (4.26) and likewise restrictions are imposed on u_h in (4.28). These restrictions take the form of the assertion that the control cannot appear in an odd order derivative. In other words,

$$\frac{\partial}{\partial t}\frac{\partial^N}{\partial t^N}\left(\frac{\partial H}{\partial u_a}\right) = 0$$

in equation (4.28) and

$$\frac{\partial}{\partial t}\frac{\partial^N}{\partial t^N}\left(\frac{\partial H}{\partial u_h}\right) = 0$$

in equation (4.26) and $N = 1$ is the order of the singular arc. Substituting these restrictions into (4.26) and (4.28) yields a joint system of two equations which are both linear in u_a and u_h. The singular paths are then derived using standard linear algebra.

4.8.1 Derivation of the Double Singular Control Path

The double singular path is the solution path that contains the singular controls for both animals and humans. Application of the Maximum Principle yields equation (4.14). Differentiation of (4.14) with respect to time yields

$$\frac{\partial}{\partial t}\left(\frac{\partial H}{\partial u_a}\right) = kw^2\left(\frac{\partial \mu_1}{\partial t}\right) + k(2\mu_1 w)\left(\frac{\partial w}{\partial t}\right) - k\left(w\frac{\partial \mu_1}{\partial t} + \mu_1\frac{\partial w}{\partial t}\right) = 0 \quad (4.26)$$

Using equations (4.11), (4.12) and (4.24) in equation (4.26), the singular solution for u_h is derived as:

$$u_h{}^* = \frac{-\delta\mu_1}{I(\mu_2 - b_h)} \quad (4.27)$$

Application of the Maximum Principle yields equation (4.15). Differentiation of (4.15) with respect to time yields:

$$\frac{\partial}{\partial t}\left(\frac{\partial H}{\partial u_h}\right) = b_h\left[I\frac{\partial w}{\partial t} + w\frac{\partial I}{\partial t}\right] - c_1\frac{\partial I}{\partial t} + z\left[w^2\frac{\partial \mu_1}{\partial t} + 2w\mu_1\frac{\partial w}{\partial t}\right] - z\left[w\frac{\partial \mu_1}{\partial t} + \mu_1\frac{\partial w}{\partial t}\right]$$
$$- \left[I\left(w\frac{\partial \mu_2}{\partial t} + \mu_2\frac{\partial w}{\partial t}\right) + \mu_2 w\frac{\partial I}{\partial t}\right] = 0 \quad (4.28)$$

Using equations (4.11), (4.12), (4.24) and (4.25) in (4.28), the singular control for u_a is derived as follows:

$$u_a{}^* = -\frac{\Gamma_1}{\Gamma_3} \quad (4.29)$$

where Γ_1 and Γ_3 are defined as follows:

$$\Gamma_1 = -c_1(\beta - 2\beta I - r)I + w\Gamma_2 + \delta\mu_1 w^2 \quad (4.30)$$

$$\Gamma_3 = k(b_h - \mu_2)I(w^2 - w) \quad (4.31)$$

and

$$\Gamma_2 = -z\delta\mu_1 + (\beta b_h - r b_h + \delta\mu_2 + c_2)I - (\beta b_h + \beta\mu_2)I^2 \quad (4.32)$$

Cliff (1999) asserts that if the first-order conditions are successively differentiated with respect to time, then the controls can only appear in an even-order derivative. By the implicit function theorem, this assertion is represented by the expression:

$$\frac{\partial}{\partial \boldsymbol{u}}\left[\frac{\partial^{2N}}{\partial t^{2N}}\left(\frac{\partial H}{\partial \boldsymbol{u}}\right)\right] \neq 0 \tag{4.33}$$

where $\boldsymbol{u}$ is a vector of controls and N is the order of the singular arc. From the foregoing discussion, a necessary condition for the singular solution to exist is known as the Generalized Legendre-Chebsch condition which states that

$$(-1)^{N}\frac{\partial}{\partial \boldsymbol{u}}\left[\frac{\partial^{2N}}{\partial t^{2N}}\left(\frac{\partial H}{\partial \boldsymbol{u}}\right)\right] \geq 0 \tag{4.34}$$

4.9 Simulation Methods and Results

In order to numerically analyze the solutions to the optimal control problem, the model needs to be parameterized or calibrated. Parameter values for parameters defined in section 4.6 are available in the literature. For the fraction of humans treated with antibiotics, u_h is derived as 0.003 using reported cases of *E. c*oli infection, number of physician visits and number hospitalized. The tonnage of antibiotics used in agriculture ranged from 1.4 million pounds to 10.45 million pounds. For the benefit per ton of using antibiotics in the animal industry, this is imputed to be between $30.1 to $40.6 using information on the benefits and net return for pigs. The marginal contribution of animal antibiotic use to human resistance was estimated to be 0.7 using information provided on the fraction of antibiotics used in the animal industry. Likewise, the marginal contribution of antibiotic use in human medicine to human resistance was imputed to be 0.3 from data on the fraction of antibiotics used in treating humans. The fraction of humans infected

was derived from reported cases of *E. c*oli. The cost of a human infection was calculated as the opportunity cost of a day's work and was between $68 and $70. With an annual discount rate of 0.10, the monthly rate was calculated to be 0.008. The parameter values, their description and sources are provided in Table 4-1. The general solution to antibiotic effectiveness is provided in Appendix A. The simulations of the time paths of antibiotic effectiveness and its shadow value (equations 4.11 and 4.24) were performed in Excel. The Excel code is provided in Appendix D. To evaluate the full impact of animal antibiotic use on drug effectiveness and its shadow value, the composite variable ($g = ku_a$), is used in the analyses that follows to represent the total volume of animal antibiotic use.

4.9.1 Illustrative Example: Tetracycline Use in Humans and Animals

To illustrate the key aspects of this model, the parameter values used or estimated are consistent with current knowledge of the use of tetracycline in human medicine and animal production and the spread of *Escherichia coli* in human society. There has been widespread use of tetracycline for prophylaxis and growth promotion in the animal industry since it was approved in 1948 (Tadesse et al. 2012).

The time paths of effectiveness and the shadow value of antibiotic effectiveness are shown in Figures 4-1 and 4-2. Figure 4-1 indicates that initially when the antibiotic is introduced or given to humans, the level of effectiveness (low resistance) is high. Over time bacteria begin to develop resistance to the new drug and it begins to lose its potency thus effectiveness declines over time. Figure 4-2 shows the time path of the shadow value of effectiveness. Initially, the shadow value of effectiveness is high because of the high level of antibiotic effectiveness. Over time as antibiotic effectiveness declines, its shadow

value decreases accordingly. The trajectory of infections is depicted in Figure 4-3. The Figure indicates that the introduction of antibiotics or a new drug to human society lowers the level of infection initially. Over time however, the level of infection starts to increase but tapers off as antibiotics become less effective. The time path of the shadow value of infection is displayed in Figure 4-4. The shadow value increases in absolute value as the level of infections goes down and tapers off as infections level off.

4.9.2. Sensitivity Analysis

To investigate the sensitivity of antibiotic effectiveness and its shadow value to changes in the total volume of animal antibiotic use, the baseline case will be the reference model. The baseline case corresponds to the scenario where g is set equal to its initial value.

In the first experiment, the sensitivity of the optimal paths of antibiotic effectiveness to changes in the total volume of animal antibiotic use is investigated. Figure 4-5 plots the time paths of antibiotic effectiveness under the baseline scenario and the scenarios where animal antibiotic use is rising. The graph reveals that antibiotic effectiveness is much lower at every point in time under the new scenarios than in the baseline case all else equal. In other words, increasing use of antibiotics in animal production, ceteris paribus, ultimately leads to increasing antibiotic resistance in humans. This finding supports the main thesis of this paper and is consistent with the views shared by Levy (2002) and Secchi and Babcock (2002) that animal antibiotic use contributes to increased resistance in humans.

In the second experiment, the sensitivity of the optimal path of the shadow value of drug effectiveness to changes in the total volume of animal antibiotic use is

investigated. Figure 4-6 reveals that the shadow value of antibiotic effectiveness declines when higher values of animal antibiotic use are introduced. This is because the new drug loses its effectiveness over time so its shadow value has to decrease.

Figure 4-1: Trajectory of Antibiotic Resistance

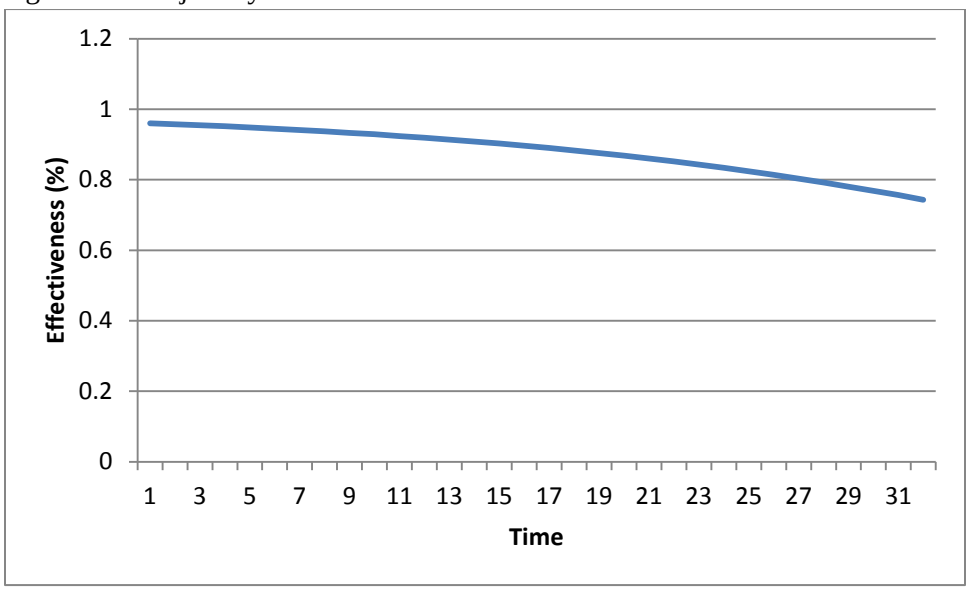


Figure 4-2: Optimal Time Path of the Shadow Value of Effectiveness

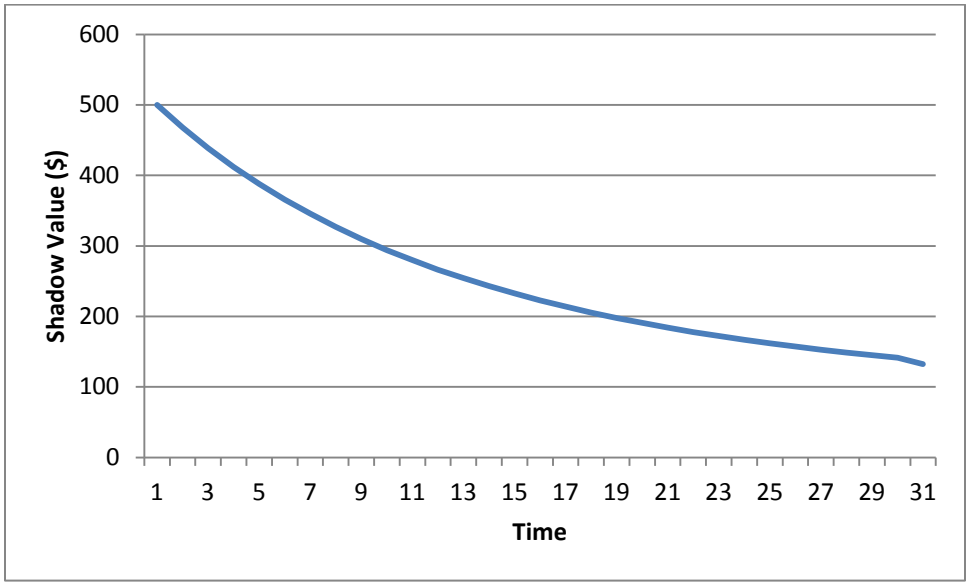


Figure 4-3: Time Path of Infections

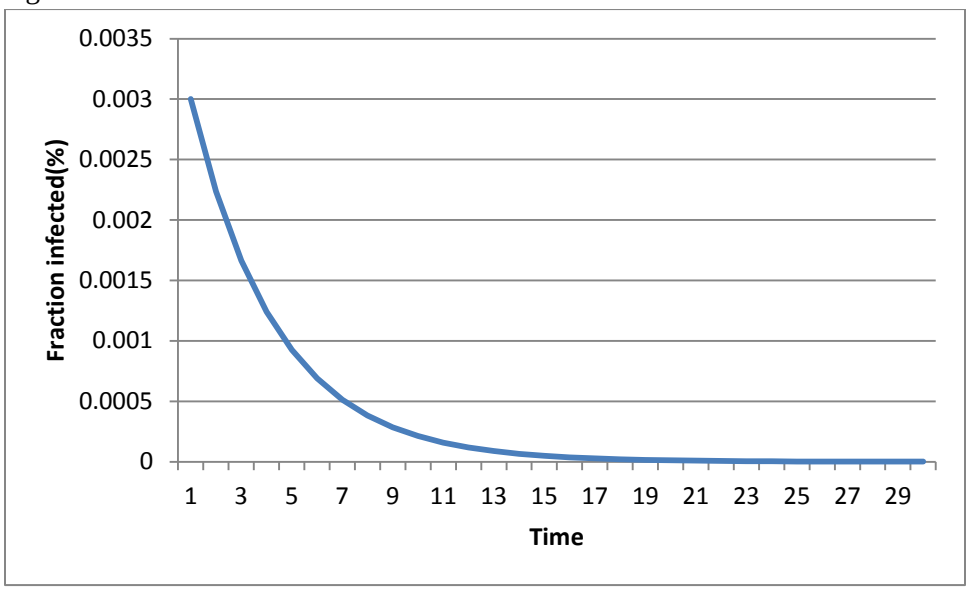


Figure 4-4: Time Path of the Shadow Value of Infections

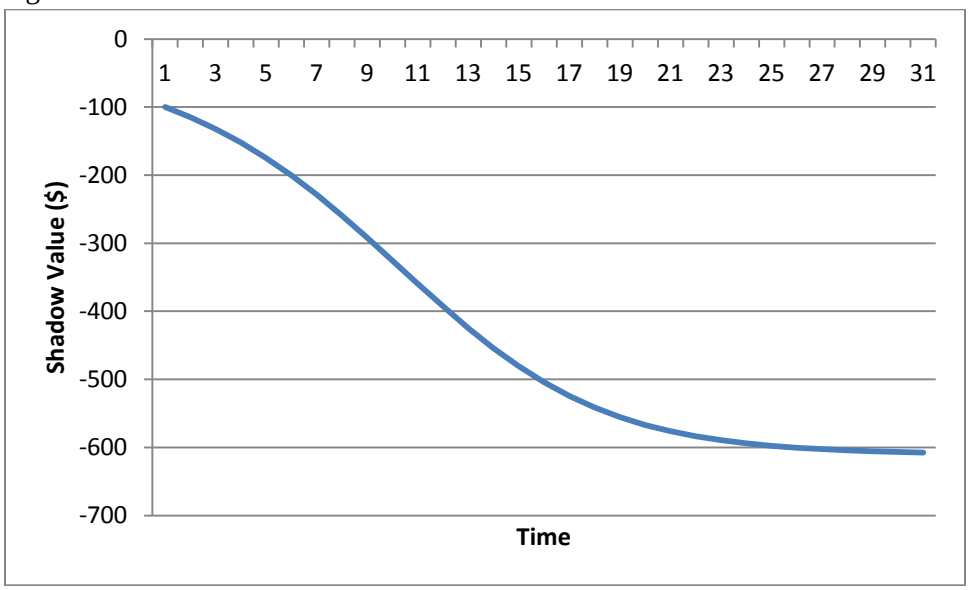
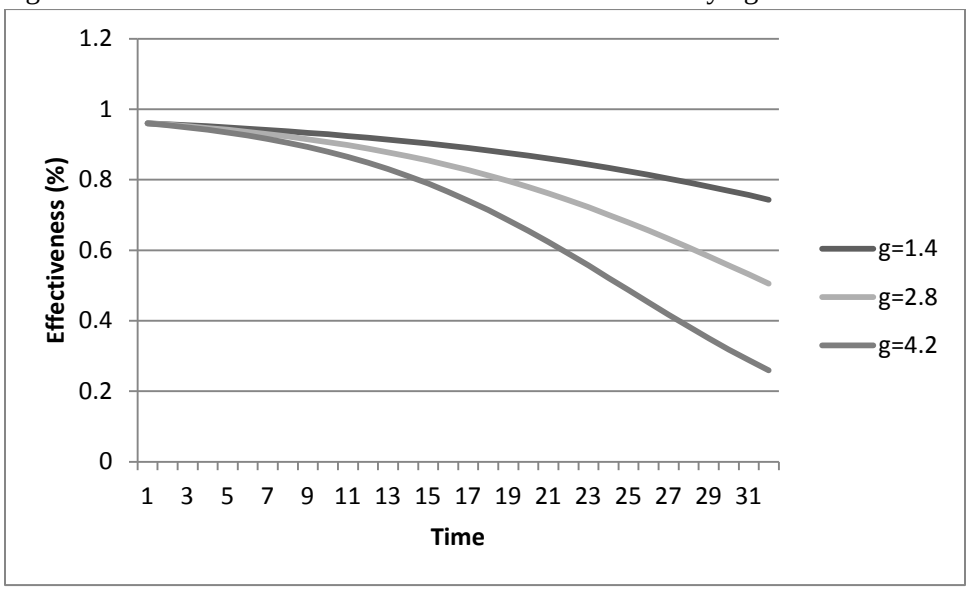
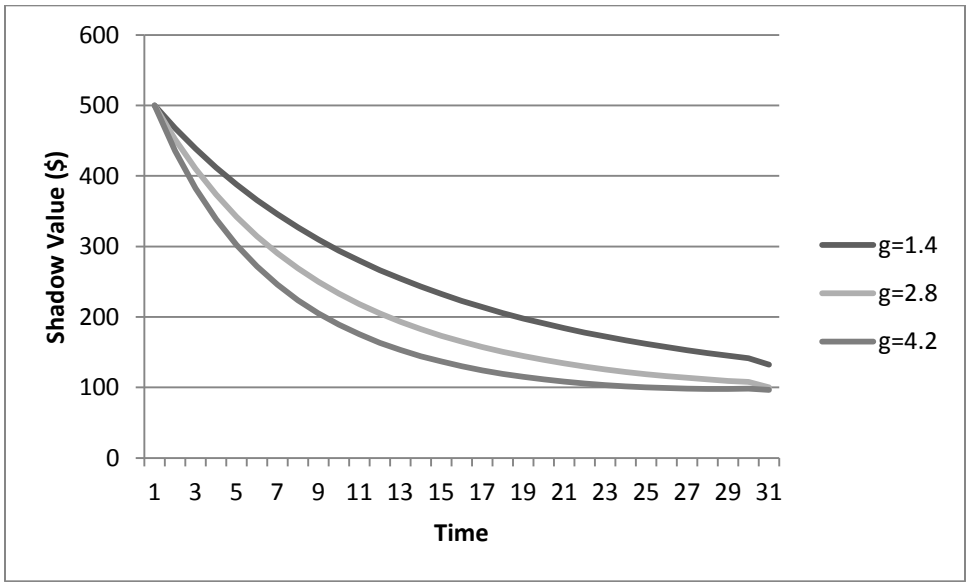


Figure 4-5: Time Paths of Antibiotic Effectiveness with Varying Animal Antibiotic Use



g represents the volume of animal antibiotic use.

Figure 4-6: Optimal Time Paths of the Shadow Value of Effectiveness with Varying Animal Antibiotic Use



g represents the volume of animal antibiotic use.

4.10 Discussion and Conclusions

This research provides an example of a study that combines biological theory, epidemiological knowledge and economic analysis. The purpose of this study was to examine the impact of animal antibiotic use on the evolution of antibiotic resistance in humans. The findings indicate that antibiotic resistance is a classic example of extraction of a non-renewable natural resource. As more and more antibiotics are used in animal production, the effectiveness of the antibiotic (natural resource) dwindles over time. This study developed an optimal control model to examine the trade-offs between current antibiotic use and future antibiotic effectiveness. A number of instructive findings are revealed when the economic objectives of optimal control are applied to the model and simulation exercises are performed. The main finding of this study is that antibiotic effectiveness declines with increased use of antibiotics in animal production. So to ameliorate the excessive depletion of this resource or to arrest the increase in resistance animal antibiotic use should be reduced.

One finding of this research consistent with the arguments seen in Secchi and Babcock (2002) and Laxminarayan and Brown (2001) is that it is beneficial to use an antibiotic so long as the net marginal benefits exceed the long-term marginal costs. However, they differ significantly from the conclusion reached by Bonhoeffer et al. (1997) and Massad et al. (2008). The former argued that a single antibiotic should be used: (1) until the resistant bacteria constitutes a specific percentage of the total population; and (2) if it minimizes the population of infected hosts (maximizes the population of uninfected hosts). The latter contends that an antibiotic should be administered only if it minimizes the mortality rate and hospitalization time. Clearly,

these criteria completely disregard economic considerations. Ignoring the economic aspects could change the decision-making outcome. In using antibiotics both economic and biological considerations must be taken into account. Economics would dictate that a particular antibiotic be used so long as the marginal benefits exceed the marginal costs.

A corollary of this result is that antibiotics should be used in the production of animals so long as the immediate net marginal benefit more than counterbalances the long-term costs in terms of increased antibiotic resistance in humans. Otherwise a ban on animal antibiotic use could be justified on economic grounds. This will make much more sense especially if the antibiotic is of high value in human medicine. One might argue that if the marginal contribution of animal antibiotic use to antibiotic resistance is greater than zero (i.e. $k > 0$), then a ban or partial ban on animal antibiotic use is justified. The results of the model suggests that the determinants of a ban on animal antibiotic use include the magnitude of k, the net marginal benefit to animal producers, and the cost in terms of decreasing effectiveness of drugs. On the other hand, if this marginal contribution to antibiotic resistance is negligible (i.e.,$k \approx 0$), then animal antibiotic use should be encouraged.

A planner can use information obtained from the model to determine the critical level of effectiveness of a particular antibiotic beyond which an intervention is necessary. Palumbi (2001) contends that antibiotic resistance is usually observed anywhere from a few months to two decades after a new antibiotic is introduced. The decision maker can determine the volume of antibiotics to be used in the animal industry based on the rate of decrease in antibiotic effectiveness (increase in antibiotic resistance). Since antibiotic use imposes a negative externality on society, a Pigouvian tax could be imposed by the

planner. This tax could be based on the marginal external cost imposed by animal producers or on the marginal contribution of animal antibiotic use to resistance in humans. This would encourage the planner to discourage the current low-value use (use of antibiotics to treat viral infections in humans and use for growth promotion in animals) of antibiotics and preserve the effectiveness of antibiotics for high-value use (use of antibiotics to treat infectious diseases in humans) in the future. The results could also be of assistance to a central planner at a pharmaceutical firm. The planner can use information on drug effectiveness and the patent to decide whether it is worthwhile to develop a new drug. For example suppose an antibiotic has a patent of 20 years but after 10 years of use the drug effectiveness is 15%, the planner has to decide whether to develop a new drug or allow the drug to remain on the market until its patent expires.

Table 4-1: Parameter Values and Description Used in Simulations Exercises

Parameter	Description	Source	Value	Range
u_h	Fraction of the human population treated with antibiotics	Calibrated from Mead et al. (1999)	0.003	0.003-0.009
u_a	Tonnage on antibiotics used in the animal industry (in thousands of tons)	Secchi and Babcock (2002) and Graham et al.(2007)	1.4	1.4-10.45
b_h	Benefit for each successful treatment of infected humans ($)	Laxaminarayan and Brown (2001)	500	200-2000
b_a	Benefit per ton of using antibiotics in animal industry (in thousands of $)	Calibrated from Cromwell (2002)	30.1	30.1-40.6
c_1	Cost per treatment of a human infections ($)	UNM Pharmacy (2012)	4.24	4.24-5.50
c_2	Cost of a human infected ($)	Frenzen et al. (2005)	68	68-70
c_a	Cost per ton antibiotics (in thousands of $)	Cromwell (2002)	30	30-40
w	Antibiotic Effectiveness (fraction)	Laxaminarayan and Brown (2001)	0.96	0.81-0.96
k	Marginal contribution of animal antibiotic use to human resistance (fraction)	Calibrated from Graham et al. (2007)	0.70	0.13-0.80
z	Marginal contribution of human antibiotic use to human resistance (fraction)	Calibrated from Graham et al. (20070	0.30	0.2-0.87
$I(t)$	Fraction of human population infected with the disease	Calibrated from Mead et al. (19990	0.0002	0.0002-0.00028
$S(t)$	Fraction of humans susceptible, but not infected	Calibrated from Mead et al. (1999)	0.9900	0.9972-0.9998
δ	Social discount	Laxaminarayan and Brown (2001)	0.008	0.004-0.10

Chapter 5: Discussion and Conclusions

5.1 Summary of Dissertation

This dissertation analyzed two health hazards namely West Nile virus and antibiotic resistance within an economic framework. This dissertation begins by examining the economic and biological factors associated with the geographic variation in human WNV in chapter 2. Disease mapping is undertaken using GIS tools to visually identify areas of high and low risk of WNV in chapter 3. Variogram analyses are also performed to examine the spatial structure of the standardized morbidity ratio, observed counts of WNV and the model residuals. The final chapter develops an optimal control model to study the inter-temporal trade-offs between current antibiotic use and future antibiotic effectiveness.

5.2 Geographic Variation in Human West Nile Virus

The primary goal of this analysis is to test a series of hypotheses relating to economic (home foreclosures and income) factors using a spatial filtering random effects negative binomial model. The hypotheses that income has a negative impact on the incidence of West Nile virus, home foreclosures are positively related to WNV and that mosquito pools have a positive effect on the prevalence of human WNV are supported in both states. The two main estimation problems encountered were endogeneity and spatial autocorrelation. Hence, an instrumental variable version of spatial filtering and spatial lag models are estimated to correct for these problems. The spatial filtering model is the preferred model based on the AIC and BIC model selection criteria.

5.3 Disease Mapping and Variography

Disease mapping and variogram analysis were undertaken in chapter 3 using GIS and geo-statistical tools. Thematic maps of the standardized morbidity ratios that the relative risks are generally higher in California counties than in Colorado counties. The probability maps, which show the probability of WNV being higher than the observed value, is higher in California than in Colorado. The variogram analysis reveals the presence of a strong spatial structure in the standardized morbidity ratio and observed counts of WNV in both states. However, they do reveal that the model residuals are devoid of spatial autocorrelation.

5.4 Economics of Antibiotic Resistance

The optimal control model illustrates that a trade-off exists between current antibiotic use in humans and animals and the future effectiveness of antibiotics. In particular, this chapter emphasizes the contribution of animal antibiotic use to the evolution of drug resistance in humans. The parameter values used to calibrate the model are derived from a variety of sources or imputed from existing data. A sensitivity and simulation analyses are conducted to investigate the impact of animal antibiotic use on the trajectory of antibiotic effectiveness and its shadow value. The simulation results show that increasing animal antibiotic use decreases the level of antibiotic effectiveness and its concomitant shadow value. The economic analysis reveals that there are certain conditions that characterize the optimal use of antibiotics resources in animals and humans. Antibiotics should be used in animal production to the point where the immediate net marginal benefits are exactly offset by the long-term marginal cost in terms of decreasing drug effectiveness. On the other hand, antibiotics should be used in human medicine to the

point where the immediate net marginal benefits are counterbalanced by the long-term costs in terms of decreasing drug effectiveness and rising infections.

5.5 Policy Implications

The issues examined in each chapter of this study have significant policy implications. The IV spatial filtering regression results do indicate that WNV prevalence is higher in counties that display certain economic and biological characteristics such as a low median income, a high number of home foreclosures and a high number of mosquito pools. Vector control strategies such as insecticide spraying and the use of mosquito nets that could reduce WNV infections are suggested by this research. This research highlights the importance of neighborhood conditions in WNV prevalence. Poor environmental conditions, caused by economic hardships, could serve as breeding grounds for the mosquito vector. It also suggests that maintenance of foreclosed homes should be encouraged. This is because it will prevent the existence of standing water in swimming pools which could serve as a breeding ground for mosquitoes. Thus, the results of this dissertation suggest that more resources should be allocated to counties that have a low median income, a high number of home foreclosures and a high number of mosquito pools for disease surveillance and monitoring in order to mitigate the disease. This will ultimately assist in reducing the incidence of West Nile virus.

Disease maps are useful to policy makers because they summarize the spatial variation in human WNV risk. The information provided by these maps can be used to formulate hypotheses about disease etiology, for monitoring and surveillance of areas classified as high risk, and in the allocation of scarce health resources. They can also be used to determine the minimum-cost path of delivering medical supplies to high-risk

areas in the event of a disease outbreak. Disease etiology could also assist policy makers in designing cost-effective preventive measures.

There is the need for policy makers to adopt a multidisciplinary approach-economics and epidemiology-in addressing the resistance problem. The results reveal how insights from antibiotic resistance fit well within the context of optimal control and policy making. Estimating the volume of animal antibiotic use and the initial level of effectiveness is crucial in formulating policies on antibiotic use. High volumes of animal antibiotic use and low magnitudes of antibiotic effectiveness translate into rising infections in the long-run. The results of the optimal control model can be useful to a decision maker concerned about the declining rate of antibiotic effectiveness. The decision maker can use the parameter values of the model to determine a threshold value of antibiotic effectiveness beyond which it may be necessary to switch to a new drug or enact policies to intervene in the decline in the effectiveness of a particular drug. Since antibiotics differ from each other with respect to the marginal benefit they bestow on the user, the marginal cost of use, and the rate of decline in effectiveness there may be the urgent need to intervene in the use of certain antibiotics quicker than others to prolong their effectiveness. The planner can therefore select the level of animal antibiotic use to derive a time path of antibiotic effectiveness consistent with its goal of conserving a particular antibiotic. Policies to regulate animal antibiotic use such as the prescription requirements are suggested by this research. Also since antibiotic use creates a negative externality on society, the planner can impose a tax equivalent to the marginal cost of the externality to discourage the overuse of antibiotics. The results could also be of assistance to a central planner at a pharmaceutical firm. The planner can use information on drug

effectiveness and the patent expiration date to decide whether it is worthwhile to develop a new drug to replace the old drug before the patent on the existing drug expires.

5.6 Directions of Future Research

The results of this study rested heavily on the spatial scales used in the analysis of WNV. This study employed county-level data to study the determinants of the high incidence of WNV in California and Colorado. It may be more meaningful studying the determinants of WNV using census tract or block level data. In terms of future research, it may be worthwhile further investigating the role of dead birds in WNV transmission. Also there are a couple of extensions that could be made to antibiotic resistance model. First, an equation could be incorporated to describe the evolution of resistance in animals. Second, the model can be modified to handle the use of two antibiotics. Simulations can then be performed to determine which antibiotic deteriorates faster in effectiveness so that its use can be discontinued. Third, the assumption of uniform antibiotic resistance is not quite realistic. Heterogeneity in resistance and infections could be introduced by specifying evolution equations for infections and effectiveness that vary by geographic region, distance to industrial farms and age.

Appendices

Appendix A: General solution for Effectiveness…….148

Appendix B: Threshold Value of Antibiotic Effectiveness…….149

Appendix C: Units and Measurements……....150

Appendix D: Stata, R and Excel Codes……....152

Appendix E: Supplementary IV Spatial Filtering RENB Regressions…….181

Appendix A: General Solution for Effectiveness

$$\frac{dw}{dt} = w(t)\big(ku_a(t) + zu_h(t)\big)(w(t) - 1) = \dot{w}$$

Let p= $ku^a(t) + zu^h(t)$ then ignoring time subscripts

$$wp(w-1) = \dot{w}$$

So $$pw^2 - pw = \dot{w}$$

Is an example of a Bernoulli equation so dividing by w^2 and re-arranging yields

$p = w^{-1}p + w^{-2}\dot{w}$

Let m=w^{-1} be a transformation. Then $\dot{m} = -w^{-2}\dot{w}$ and substituting into ** yields

$$\dot{m} - mp(t) = -p(t) \qquad (A1)$$

Multiply (A1) by an integrating factor $e^{A(t)}$. This yields the following equation:

$$\dot{m}e^{A(t)} - mp(t)e^{A(t)} = -p(t)e^{A(t)} \qquad (A2)$$

Next we need to find a $A(t)$ such that the left hand side of equation A2 equals the derivative of $me^{A(t)}$. The derivative of $me^{A(t)}$ is equal to $\dot{m}e^{A(t)} + m\dot{A}e^{A(t)}$. By making $A(t)$ satisfy $A(t) = a(t)$ and choosing $A(t) = \int a\,(t)dt$ will make A2 equivalent to the following equation:

$$\frac{d}{dt}(me^{A(t)} = p(t)e^{A(t)}$$

Hence $me^{A(t)}$ is an indefinite integral of $p(t)e^{A(t)}$, so there has to exist a constant C such that

$$me^{A(t)} = \int p(t)\, e^{A(t)}dt + C \qquad (A3)$$

Multiplying (A3) by $e^{-A(t)}$ we get

$$m = Ce^{-A(t)} + e^{-A(t)} \int p(t)\, e^{A(t)}dt$$

Hence the general solution for m is:

$$m = e^{-\int a(t)dt}\left(C + \int e^{\int a(t)dt}\, p(t)dt\right) \qquad (A4)$$

and $w = m^{-1}$.

Appendix B: Derivation of Threshold Value of Effectiveness

Equation (7.13) can be re-written in quadratic form as:

$\mu_1 k w^2 - \mu_1 k w + b_a - c_a = 0$. So the threshold hold is calculated as follows:

$$w_a = \frac{\mu_1 k \pm \sqrt{(\mu_1 k)^2 - 4(\mu_1 k)(b_a - c_a)}}{2\mu_1 k} \qquad (B1)$$

Appendix C: Explanation of Optimal Control Units

Parameters:

b_h=benefit associated with each treatment (dose) of the antibiotic measured in [$/ person]

b_a= benefit associated with each use of the antibiotic measured in [$/ton]

c_1= unit cost of human treatment (dose) [$/person]

c_2=unit cost of human infection [$/person]

c_a=unit cost per ton of administering antibiotics to animals measured [$/ton].

r=recovery rate/clearance rate [%]

β=transmission rate [%]

k=marginal contribution of animal use of antibiotics to resistance due to one more use of the drug [%]
z=marginal contribution of human use of antibiotics to resistance due to one more use of the drug [%]

Variables

$$N(t) = S(t) + I(t)$$

$S(t)$ =fraction of humans that are susceptible to infections [unitless]= $\left(\frac{susceptible\ population}{total\ human\ population}\right)$
Susceptible population is measured in [person/t] and total human population is in [person/time]
$I(t)$ =fraction of humans that are infected [unitless]= $\left(\frac{infected\ population}{total\ human\ population}\right)$
Infected population is measured in [person/t]

$I_r(t)$ = fraction of humans infected with the resistant strain
[unitless]= $\left(\frac{fraction\ of\ infected\ \ with\ resistant\ strain}{total\ infected}\right)$
Fraction infected with resistant strain is measured in [person/t] and total infected is measured in [person/t]

$I_w(t)$=fraction of humans infected with the susceptible strain
[unitless]= $\left(\frac{fraction\ infected\ with\ susceptible\ strain}{total\ infected}\right)$
Fraction infected with susceptible strain is measured in [person/t] and $I(t) = I_r(t) + I_w(t)$.

$w(t)$=effectiveness of antibiotic and is equal to the ratio of humans infected with the susceptible strain to the total infected population [unitless]= $\left(\frac{humans\ infected\ with\ susceptible\ strain}{total\ infected}\right)$

u_h=fraction of the infected population treated with antibiotics
[unitless]= $\left(\frac{treated}{total\ infected\ population}\right)$
Treated are measured in [person/t]
u_a=tonnage of antibiotics used in animal industry [tons/t]
Benefit to Humans

$$B_H = b_h w(t) u_h(t) N(t) I(t) \qquad (C1)$$

Benefit to Humans:

$$\frac{\$}{t} = \frac{\$}{treated} * [effectiveness\ ] * \left(\frac{treated}{total\ infected\ population}\right) * \left(\frac{total\ infected\ population}{total\ human\ population}\right)$$

$$\frac{s}{t} = \frac{\$}{treated}\left(\frac{treated}{t}\right) * effectiveness * u_h * I(t) = \frac{\$}{t}$$

Cost to Humans:

$$C_H = c_1 u_h(t) N(t) I(t) + c_2 N(t) I(t) \qquad (C2)$$

$$\frac{\$}{t} = \frac{\$}{treated} * \left(\frac{treated}{total\ infected\ population}\right) * \left(\frac{total\ infected\ population}{total\ human\ population}\right) + \frac{\$}{infected} * \left(\frac{infected}{total\ human\ population}\right)$$

$$\frac{\$}{t} = \frac{\$}{treated}\left(\frac{treated}{t}\right) * u_h * I(t) + \frac{\$}{infected}\left(\frac{infected}{t}\right) * I(t) = \frac{\$}{t}$$

Benefit to Animal Producers:

$$B_A = b_a u_a(t) \qquad (C3)$$

[$/time] = [$/ton]*[number of tons]

$$\frac{\$}{t} = \frac{\$}{ton} * \frac{ton}{t} = \frac{\$}{t}$$

Cost to Animal Producers: $C_A = c_a u_a(t)$ $\qquad (C4)$

[$/t] = [$/ton]*[number of tons]

$$\frac{\$}{t} = \frac{\$}{ton} * \frac{ton}{t} = \frac{\$}{t}$$

Appendix D: Stata, R and Excels Codes Stata Codes

```
*******************************************************************************
* Stata 9.2
*
* Dissertation Chapter 2: Examination of Geographic Variation in Human West Nile Virus
*
* WNVCA0307.do. This file estimates the Random Effects Negative Binomial Regressions for California
*
* It uses data on West Nile Virus and other climatic, economic and environmental variables from 2003-
*2007
*It reproduces Tables 2-4 to 2-8
*
*******************************************************************************
```

California Codes

```
clear
#delimit;
cap log close;
set mem 80m;
set more off;
set scrollbufsize 300000;
 log using E:\research\zoonotic, replace;
insheet using E:\research\WNVCA0307.csv;
gen migrate=netmigr/population;   // generates a migration rate
gen lnincome=ln(income);
gen elevate=elevation/1000;   // scales down elevation
gen lnpop=ln(population);   // generates the exposure variable
gen lnpop2=ln(lnpop);   // generates the offset variable
gen fcloseca=ln(fclose+1);   // transforms home foreclosures
gen newhuman=ln(human+1);
gen lnarea=ln(area);
tabulate year, gen(time);   // generates the time fixed effects
encode county, generate(location);   // generates the random effects variable
gen popdense=population/area;   // generates the population density
gen lnpopdense=ln(popdense);   // generates the log of population density
summarize human lnincome fcloseca bird mosquito migrate pwhite lnpop lnpopdense precip temp pdsi
urban elevate time2-time5;
summarize bird mosquito equine migrate precip temp pdsi educate poverty unrate airport urban elevate
lnpopdense lnarea road time2-time5;
corr bird mosquito equine fcloseca income urban educate poverty elevate migrate lnpop lnpopdense temp
precip pwhite pdsi;

**********Spatial Autocorrelation Test TABLE 2-3**********************************
spatwmat using E:\research\swm_ca.dta, name(W) eigenval(E) standardize; // reads W a 58X58 spatial
weights matrix
 set matsize 400;
 matrix I = I(5); // creates a 5X5 identity matrix
 matrix W1=I#W; // creates a new 290 X 290  matrix
 spatgsa human, weights(W1) moran geary; // Moran's I test on WNV cases

*******************************************************************************
*****
matrix I58=I(58);   // creates a 58 X 58 identity matrix
```

```
mkmat O;   // reads a 58 X  1 column of ones
matrix C1=I58-OO'/58;   // transforms the identity matrix
matrix C=C1*W*C1;   // creates a centered spatial weights matrix
matrix symeigen X v = C;  // extracts the eigenvectors from the centered spatial weights matrix C
```

**********TABLE 2-4 MODELS 1-4 (4 COLUMNS)**********

```
*Model 1 Non-spatial Random Effects Negative Binomial Regression (Equation 2.8)
xtnbreg human fcloseca lnincome lnpopdense time2-time5,i(location) iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(NB-IV) ctitle (Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat1, xb;
gen resh1=human-hhat1;   // generates model residuals
spatgsa resh1, weights(W1) moran geary; // spatial autocorrelation test on model residuals
estat ic;

*Model 2 Non-spatial Random Effects Negative Binomial Regression (Equation 2.8)
xtnbreg human  fcloseca lnincome mosquito lnpopdense precip time2-time5, i(location) iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(NB-IV) ctitle (Model 2) 3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat2, xb;   // generates the predicted values
gen resh2=human-hhat2;   // generates model residuals
spatgsa resh2, weights(W1) moran geary;// spatial autocorrelation test on model residuals
estat ic;

*Model 3  Non-spatial Random Effects Negative Binomial Regression (Equation 2.8)
xtnbreg human  fcloseca lnincome mosquito lnpopdense temp time2-time5, i(location) iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(NB-IV) ctitle (Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat3, xb;   // generates the predicted values
gen resh3=human-hhat3;   // generates model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
estat ic;

Model 4 Non-spatial Random Effects Negative Binomial Regression (Equation 2.8)
xtnbreg human  fcloseca lnincome mosquito lnpopdense pdsi time2-time5, i(location) iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(NB-IV) ctitle (Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;   // generates the predicted values
gen resh=human-hhat4;   // generates model residuals
spatgsa resh4, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
```

****

**********TABLE 2-5 MODELS 1-4 (4 COLUMNS)**********

```
*Model 1 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // instrumentalizes the spatial lag
predict slaghat1, xb;   // generates predicted values of spatial lag
adjust,exp;
xtnbreg human fcloseca lnincome  lnpopdense time2-time5 slaghat1, i(location) iterate(100) nolog;
outreg using G:\research\WNV2.doc, title(RE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat1, xb;  // generates model predicted values
gen resh1=human-hhat1;   // generates model residuals
spatgsa resh1, weights(W1) moran geary;  //spatial autocorrelation test on model residuals
```

```
estat ic;

*Model 2 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog; // instrumetnalize the  spatial lag
predict slaghat2, xb;
adjust,exp;
xtnbreg human fcloseca lnincome mosquito lnpopdense precip time2-time5 slaghat2, i(location)
iterate(100) nolog;
outreg using G:\research\WNV2.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat2, xb;   // generates model predicted values
gen resh2=human-hhat2;   // generates model residuals
spatgsa resh2, weights(W1) moran geary; //spatial autocorrelation test on the model residuals
estat ic;

*Model 3 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog;    //instrumentalizes spatial lag
predict slaghat3, xb;   / generates predicted spatial lag values
adjust,exp;
xtnbreg human fcloseca lnincome mosquito lnpopdense temp time2-time5 slaghat3, i(location) iterate(100)
nolog;
outreg using G:\research\WNV2.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat3, xb;   // generates model predicted values
gen resh3=human-hhat3;   // generates model residuals
spatgsa resh3, weights(W1) moran geary;  //spatial autocorrelation test on the model residuals
estat ic;

**Model 4
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog;    //instrumentalizes spatial lag
predict slaghat4, xb;   / generates predicted spatial lag values
adjust,exp;
xtnbreg human fcloseca lnincome  lnpopdense  mosquito pdsi time2-time5 slaghat4, i(location)iterate(100)
nolog;
outreg using E:\research\WNV2.doc, title(RE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;


**********************************************************************************************


**********TABLE 2-6 MODELS 1-4 (4 COLUMNS)**********
Model 1 IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseca  educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;  //  instrumentalizes home
foreclosures
predict fhat1, xb;   // generates the predicted values of home foreclosures
adjust,exp;
gen resf=fcloseca-fhat1;  // generates home foreclosure residuals
```

```
egen zresf=std(resf);   // generates the standardized home foreclosure residuals
regress lnincome  educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5;   //  instrumentalizes income
predict yhat1,xb;   // generates the predicted values of log of income
gen resinc=lnincome-yhat1;   // generates income residuals
egen zresinc=std(resinc);   //  generates the standardized income residuals
xtnbreg lag1 fcloseca  educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog;   // instrumentalizes the
spatial lag
predict slaghat1, xb;   // generates the predicted values of the spatial lag
adjust,exp;
gen respat=lag1-slaghat1;   // generates the spatial lag residuals
egen zrespat=std(respat);   // generates the standardized spatial lag residual
xtnbreg human fhat1 yhat1 lnpopdense time2-time5 slaghat1, i(location)
iterate(100)vce(bootstrap,rep(100))nolog;
outreg using G:\research\WNV3.doc, title(RE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on the model residuals
estat ic;

**Endogeneity Test**
xtnbreg human fcloseca lnincome bird mosquito migrate zresf zresinc zrespat time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf=zresinc=0;   // tests equality of standardized residuals

**Validity/Over-Identification Test**
xtnbreg human fhat1 yhat1 bird mosquito migrate time2-time5 slaghat1,
i(location)iterate(100)exposure(lnpop)nolog;
predict wnvhat,xb;   // generates the IV predicted values
adjust, exp;
gen reswnv=human-wnvhat;   // generates the IV residuals
egen zreswnv=std(reswnv);   // generates the standardized IV residuals
gen zreswnv1=zreswnv+1;   // transforms the standardized residuals
sum(zreswnv1);
xtnbreg zreswnv1 precip temp pdsi educate equine unrate poverty bird mosquito migrate airport pwhite
elevate urban time2-time5,i(location)iterate(100)vce(bootstrap,rep(5))nolog;   // regresses standardized
residuals on instruments
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // get R^2 from regression to calculate test statistic

*Model 2 IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   //instrumentalizes home
foreclosures
predict fhat2, xb;   // generates predicted values of home foreclosures
adjust, exp;
gen resf2=fcloseca-fhat2;   // generates the home foreclosure residuals
egen zresf2=std(resf2);   // generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of income
predict yhat2, xb;   // generates the predicted values of income
gen resinc2=lnincome-yhat2;   // generates the income residuals
egen zresinc2=std(resinc2);   // generates standardized income residuals
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog;   // instrumentalizes the spatial lag
```

```
predict slaghat2, xb;   // generates the spatial lag predicted values
adjust, exp;
gen respat2=lag1-slaghat2;   // generates the spatial lag residuals
egen zrespat2=std(respat2);   // generates standardized spatial lag residuals
xtnbreg human fhat2 yhat2 mosquito lnpopdense precip time2-time5 slaghat2, i(location) iterate(100)
vce(bootstrap,rep(100)) nolog;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat2, xb;   // generates the model predicted values
gen resh2=human-hhat2;   // generates the model residuals
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on the residuals
estat ic;

***Endogeneity Test
xtnbreg human fcloseca lnincome bird mosquito migrate zresf2 zresinc2 zrespat2 time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf2=zresinc2=0;  // tests for equality of standardized residuals

**Model 3 IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   //instrumentalizes home
foreclosures
predict fhat3, xb;   // generates the predicted values for home foreclosures
adjust, exp;
gen resf3=fcloseca-fhat3;   // generates home foreclosure residuals
egen zresf3=std(resf3);   // generates the standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of income
predict yhat3, xb;   // generates the predicted values of log of income
gen resinc3=lnincome-yhat3;    // generates the income residuals
egen zresinc3=std(resinc3);   // generates the standardized income residuals
xtnbreg lag1  educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5,i(location)nolog;  // instrumentalizes the spatial lag
predict slaghat3, xb;   // generates the predicted values of the spatial lag
adjust, exp;
gen respat3=lag1-slaghat3;   // generates the spatial lag residuals
egen zrespat3=std(respat3);  // generates the standardized spatial lag residuals
xtnbreg human fhat3 yhat3 mosquito lnpopdense temp time2-time5 slaghat3, i(location)  iterate(100)
vce(bootstrap,rep(100)) nolog;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat3, xb;   / generates the model predicted values
gen resh3=human-hhat3;   // generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on the model residuals
estat ic;

***Endogeneity Test
xtnbreg human fcloseca lnincome bird mosquito migrate zresf3 zresinc3 zrespat3 time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf3=zresinc3=0;   // equality test of standardized residuals

Model 4
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;
predict fhat4, xb;
adjust, exp;
reg lnincome fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5;
```

```
predict yhat4, xb;
xtnbreg lag1 fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5 ,i(location)iterate(100)nolog;
predict slaghat4, xb;
adjust, exp;
xtnbreg human fhat4 yhat4 lnpopdense mosquito pdsi time2-time5 slaghat4, i(location)
iterate(100)vce(bootstrap,rep(100)) nolog;
outreg using E:\research\WNV3.doc, title(RE-IV) ctitle(Model 4)3aster append;
*display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

**********TABLE 2-7 MODELS 1-4 (4 COLUMNS)**********

```
**Model 1 Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseca lnincome  lnpopdense time2-time5 e3 e16 e25 e29 e46, i(location) iterate(100)
vce(bootstrap, rep(100)) nolog;
outreg using G:\research\WNV4.doc, title(RE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
estat vce;
estat ic;

**Model 2 Spatial Filter Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseca lnincome mosquito lnpopdense precip time2-time5 e1 e3 e11 e13 e15 e16 e21 e25
e26 e27 e29 e34 e46, i(location) iterate(100) vce(bootstrap, rep(100)) nolog;
outreg using G:\research\WNV4.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat2, xb;   // generates the model predicted values
gen resh2=human-hhat2;   // generates the model residuals
spatgsa resh2, weights(W1) moran geary;   //spatial autocorrelation test on model residuals
estat ic;

**Model 3  Spatial Filter Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseca lnincome mosquito lnpopdense temp time2-time5 e3 e16 e25 e29 e33 e46,
i(location) iterate(100) vce(bootstrap, rep(100)) nolog;
outreg using G:\research\WNV4.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
estat ic;

***Model 4
xtnbreg human fcloseca lnincome  lnpopdense mosquito pdsi time2-time5 e1 e16 e29, i(location)
iterate(100) vce(bootstrap, rep(100)) nolog;
outreg using E:\research\WNV4.doc, title(RE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

**********TALE 2-8 MODELS 1-4 (4 COLUMNS)**********
**Model 1 IV Spatial Filter Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // instrumentalizes home foreclosures
predict fhat1, xb; // generates the home foreclosures predicted values
adjust,exp;
gen resf=fcloseca-fhat1; // generates the home foreclosure residuals
egen zresf=std(resf); // generates the standardized home foreclosure residuals
regress lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5; //instrumentalizes the log of income
predict yhat1,xb; // generates the predicted values of log of income
gen resinc=lnincome-yhat1; // generates income residuals
egen zresinc=std(resinc); // generates the standardized income residuals
xtnbreg human fhat1 yhat1 lnpopdense time2-time5 e4 e6 e15 e16 e46 e47, i(location) iterate(100) vce(bootstrap,rep(100)) nolog;
outreg using E:\research\WNV5.doc, title(RE-IV) ctitle(Model 1)3aster replace;
predict hhat1, xb; // generates the model predicted values
gen resh1=human-hhat1; // generates the model residuals
spatgsa resh1, weights(W1) moran geary; // spatial autocorrelation test on model residuals
estat ic;

***Endogeneity Test
xtnbreg human fcloseca lnincome bird mosquito migrate zresf zresinc zrespat time2-time5 lag1, i(location)exposure(lnpop)iterate(100)nolog;
test zresf=zresinc=0; // tests the equality of the standardized residuals

**Validity/Over-Identification Test
xtnbreg human fhat1 yhat1 bird mosquito migrate time2-time5 slaghat1, i(location)iterate(100)exposure(lnpop)nolog;
predict wnvhat,xb; // generates model predicted values
gen reswnv=human-wnvhat; //generates model residuals
egen zreswnv=std(reswnv); // standardized model residuals
gen zreswnv1=zreswnv+1; // transformed model residuals
sum(zreswnv1);
xtnbreg zreswnv1 precip temp pdsi educate equine unrate poverty bird mosquito migrate airport pwhite elevate urban time2-time5,i(location)iterate(100)vce(bootstrap,rep(5))nolog; // regress standardized residuals on instruments
display "Pseudo-R^2"=1-e(ll)/e(ll_0); // get R^2 to calculate test statistic

/*predict reshat, xb;
gen essres=reshat-1;
gen essres2=(essres)^2;
sum(essres2);
gen tssres=reswnv-1;
gen tssres2=(tssres)^2;
sum(tssres2);
predict reswnvhat, xb;
adjust, exp;
list reswnv;*/

**Model 2 IV Spatial Filter Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // instrumentalizes home foreclosures

```
predict fhat2, xb;   // generate predicted values of home foreclosures
adjust, exp;
gen resf2=fcloseca-fhat2;   // generates home foreclosures residuals
egen zresf2=std(resf2);   // generates the standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of income
predict yhat2, xb;   // generates the predicted values of log of income
gen resinc2=lnincome-yhat2;  // generates the log of income residuals
egen zresinc2=std(resinc2);   // generates the standardized income residuals
xtnbreg human fhat2 yhat2 mosquito lnpopdense precip time2-time5 e16, i(location) iterate(100)
vce(bootstrap,rep(100))  nolog;
predict hhat2, xb;   // generates the model predicted values
gen resh2=human-hhat2;   // generates the model residuals
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
estat ic;

***Endogeneity Test
xtnbreg human fcloseca lnincome bird mosquito migrate zresf2 zresinc2 zrespat2 time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf2=zresinc2=0;   // tests for equality of standardized residuals

**Model 3 IV Spatial Filter Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes home
foreclosures
predict fhat3, xb;   // generates the predicted values of home foreclosures
adjust, exp;
gen resf3=fcloseca-fhat3;   // generates the home foreclosures residuals
egen zresf3=std(resf3);   // generates the standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes the log of income
predict yhat3, xb;   // generates the predicted values of income
gen resinc3=lnincome-yhat3;   // generates the income residuals
egen zresinc3=std(resinc3);   // generates the standardized income residuals
xtnbreg human fhat3 yhat3 mosquito lnpopdense temp time2-time5 e16, i(location) iterate(100)
vce(bootstrap,rep(100)) nolog;
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   / generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on the residuals
estat ic;

**Endogeneity Test
xtnbreg human fcloseca lnincome bird mosquito migrate zresf3 zresinc3 zrespat3 time2-time5 lag1,
i(location)exposure(lnpop)iterate(100) nolog;
test zresf3=zresinc3=0;  // tests the equality of the standardized residuals
```

Model 4

```
xtnbreg fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;
predict fhat4, xb;
adjust, exp;
reg lnincome fcloseca educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5;
predict yhat4, xb;
xtnbreg human fhat4 yhat4  lnpopdense mosquito pdsi time2-time5 e16 e41, i(location) iterate(100)
vce(bootstrap,rep(100)) nolog;
outreg using E:\research\WNV5.doc, title(RE-IV) ctitle(Model 4)3aster append;
*display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

Colorado Codes

```
**********************************************************************************************
  Stata 9.2
  Dissertation Chapter 2: Examination of Geographic Variation in Human West Nile Virus
  WNVCO0307.do. This file estimates the Random Effects Negative Binomial Regressions for Colorado
  It uses data on West Nile Virus and other climatic, economic and environmental variables from 2003-
2007
  It reproduces Tables 2-9 to 2-13
**********************************************************************************************
clear
#delimit;
cap log close;
set mem 80m;
set more off; log using E:\research\zoonotic, replace;
insheet using E:\research\WNVCO0307.csv;
gen migrate=netmigr/population;
gen lnpop=ln(population);   // creates the exposure variable
gen lnpop2=ln(lnpop);   // generates the offset variable
gen lnincome=ln(income);   // generates the log of income
gen elevate=elevation/1000;   // scales down elevation
gen fcloseco=ln(fclose+1);   // transforms home foreclosures
gen newhuman=ln(human+1);
gen lnarea=ln(area);
tabulate year, gen(time);   // generates the time fixed effects
encode county, generate(location);   // generates the random effects
gen popdense=population/area;   // generates the population density
gen lnpopdense=ln(popdense);   // generates the log of population density
summarize human lnincome fcloseco bird mosquito migrate pwhite lnpop lnpopdense precip temp pdsi
urban elevate time2-time5;
summarize bird mosquito equine migrate precip temp pdsi educate poverty unrate airport urban elevate
lnpopdense lnarea road time2-time5;
corr bird mosquito fcloseco lnincome educate urban elevate migrate temp precip pdsi pwhite equine lnpop
lnpopdense;
```

********** TABLE 2-3 SPATIAL AUTOCORRELATION TEST ON WNV COUNTS*******

```
spatwmat using G:\research\swm_co.dta, name(W) eigenval(E) standardize; // reads W a 63 X 63 spatial
weights matrix
set matsize 400;
matrix K=W;
matrix I = I(5); // creates a 5 X 5 identity matrix
matrix list I;
matrix W1=I#W; // creates a 315 X 315 weights matrix using a Kroenecker product
spatgsa human, weights(W1) moran geary;   // spatial autocorrelation test on WNV counts
```

****

```
matrix I63=I(63);   // creates a 63 X 63 identity matrix
mkmat O;   // reads a 63 X 1 column of ones
matrix C1=I63-OO'/63;   // transforms the identity matrix
matrix C=C1*W*C1;   // creates a centered spatial weights matrix
matrix symeigen X v = C;  // extracts the eigenvectors from a centered  spatial weights matrix C
```

**********TABLE 2-9 MODELS 1-4 (4 COLUMNS)**********

```
**Model 1 Non-spatial Random Effects Negative Binomial Regressions (Equation 2.8).
xtnbreg human fcloseco lnincome lnpopdense time2-time5, i(location) iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(FE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the R2
estat ic;
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 2 Non-spatial Random Effects Negative Binomial Regressions (Equation 2.8).
xtnbreg human fcloseco lnincome mosquito lnpopdense precip time2-time5,
i(location)exposure(lnpop)iterate(100)nolog;
outreg using G:\research\WNV1.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the R2
estat ic;
predict hhat2, xb;   // generates model predicted values
gen resh2=human-hhat2;   // generates the model residuals
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 3  Non-spatial Random Effects Negative Binomial Regressions (Equation 2.8).
xtnbreg human fcloseco lnincome mosquito  lnpopdense temp time2-time5, i(location) iterate(100) nolog;
outreg using G:\research\WNV1.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the R2
estat ic;
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

Model 4
xtnbreg human fcloseco lnincome  lnpopdense mosquito pdsi time2-time5, i(location) iterate(100) nolog;
outreg using E:\research\WNV1.doc, title(FE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

****

**********TABLE 2-10 MODELS 1-4 (4 COLUMNS)**********

```
**Model 1 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9).
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   //  instrumentalizes the spatial lag
predict slaghat1, xb;   // generates the spatial lag predicted values
adjust, exp;
xtnbreg human fcloseco lnincome lnpopdense time2-time5 slaghat1, i(location) iterate(100) nolog;
outreg using G:\research\WNV2.doc, title(FE-IV) ctitle(Model 1) 3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the model R2
estat ic;
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 2 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9).
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes the spatial lag
predict slaghat2, xb;   // generates the spatial lag predicted values
adjust, exp;
xtnbreg human fcloseco lnincome mosquito precip time2-time5 slaghat2, i(location) iterate(100) nolog;
outreg using G:\research\WNV2.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the model R2
estat ic;
predict hhat2, xb;
gen resh2=human-hhat2;
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 3 Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9).
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes the spatial lag
predict slaghat3, xb;   // generates the spatial lag predicted values
adjust, exp;
xtnbreg human fcloseco lnincome mosquito lnpopdense temp time2-time5 slaghat3, i(location) iterate(100)
nolog;
outreg using G:\research\WNV2.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the model R2
estat ic;
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 4
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate urban
lnpopdense lnarea road time2-time5, i(location)nolog;
predict slaghat4, xb;
adjust, exp;
xtnbreg human fcloseco lnincome  lnpopdense mosquito pdsi time2-time5 slaghat4, i(location) iterate(100)
nolog;
outreg using E:\research\WNV2.doc, title(FE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
```

```
estat ic;

**********************************************************************************************

**********TABLE 2-11 MODELS 1-4 (4 COLUMNS)**********
**Model 1_IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes home
foreclosures
predict fhat1, xb;   // generates home foreclosures predicted values
adjust, exp;
gen resf=fcloseco-fhat1;   // generates home foreclosure residuals
egen zresf=std(resf);   // generates the standardized home foreclosures residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of  income
predict yhat1, xb;   // generates predicted values of income
gen resinc=lnincome-yhat1;   // generates the income residuals
egen zresinc=std(resinc);   // generates the standardized income residuals
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes the spatial lag
predict slaghat1, xb;   // generates the predicted values of the spatial lag
adjust, exp;
gen respat=lag1-slaghat1;   // generates the spatial lag residuals
egen zrespat=std(respat);  // generates the standaridzed spatial lag residuals
xtnbreg human fhat1 yhat1 lnpopdense time2-time5 slaghat1, i(location) iterate(100) vce(bootstrap,
rep(100)) nolog;
outreg using G:\research\WNV3.doc, title(FE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // calculates the model R2
estat ic;
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

***Validity/Over-Identifaction Tests
xtnbreg human fhat1 yhat1 bird mosquito migrate time2-time5 slaghat1,
i(location)iterate(100)exposure(lnpop)nolog;
predict wnvhat,xb;   // generates model predicted values
gen reswnv=human-wnvhat;   // generates model residuals
egen zreswnv=std(reswnv); // generates standardized IV regression residuals
gen zreswnv1=zreswnv+1; // transforms the generated IV residuals
xtnbreg zreswnv1 precip temp pdsi educate equine unrate poverty bird mosquito migrate airport pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // regresses residuals on
instruments
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // gives the model R2 to calculates test statistics
/*predict reshat, xb;
gen essres=reshat-1;
gen essres2=(essres)^2;
sum(essres2);
gen tssres=reswnv-1;
gen tssres2=(tssres)^2;
sum(tssres2);
predict reswnvhat, xb;
adjust, exp;
list reswnv;
```

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf zresinc zrespat time2-time5 lag1, i(location)exposure(lnpop)iterate(100)nolog;
test zresf=zresinc=0; // for equality of standardized residuals

**Model 2 IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog; // instrumentalizes home foreclosures
predict fhat2, xb; // generates the predicted values of home foreclosures
adjust, exp;
gen resf2=fcloseco-fhat2; // generates home foreclosures residuals
egen zresf2=std(resf2); // generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5; // instrumentalizes the log of income
predict yhat2, xb; // generates the predicted values of income
gen resinc2=lnincome-yhat2; // generates the residuals of income
egen zresinc2=std(resinc2); // generates standardized income residuals
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog; // instrumentalizes the spatial lag
predict slaghat2, xb; // generates the spatial lag predicted values
adjust, exp;
gen respat2=lag1-slaghat2; // generates the spatial lag residuals
egen zrespat2=std(respat2); // generates standardized spatial lag residuals
xtnbreg human fhat2 yhat2 lnpopdense mosquito precip time2-time5 slaghat2,i(location) iterate(100) vce(bootstrap, rep(100)) nolog;
display "Pseudo-R^2"=1-e(ll)/e(ll_0); // gives model R2
estat ic; // produces model diagnostics
predict hhat2, xb; // generates model predicted values
gen resh2=human-hhat2; // generates model residuals
spatgsa resh2, weights(W1) moran geary; // spatial autocorrelation test on model residuals

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf2 zresinc2 zrespat2 time2-time5 lag1, i(location)exposure(lnpop)iterate(100)nolog;
test zresf2=zresinc2=zrespat2=0; // equality test of standardized residuals

**Model 3 IV Spatial Lag Random Effects Negative Binomial Regressions (Equation 2.9)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // instrumentalizes the spatial lag
predict fhat3, xb; // generates the predicted values of the spatial lag
adjust, exp;
gen resf3=fcloseco-fhat3; // generates the spatial lag residuals
egen zresf3=std(resf3); // generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5; // instrumentalizes the log of income
predict yhat3, xb; // generates the predicted values of log of income
gen resinc3=lnincome-yhat3; / generates the income residuals
egen zresinc3=std(resinc3); // generates standardized income residuals
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog; // instrumentalizes the spatial lag
predict slaghat3, xb; // generates the predicted values of the spatial lag
adjust, exp;
gen respat3=lag1-slaghat3; // generates the spatial lag residuals
egen zrespat3=std(respat3); // generates standardized spatial lag residuals

```
xtnbreg human fhat3 yhat3 mosquito lnpopdense temp time2-time5 slaghat3, i(location) iterate(100)
vce(bootstrap, rep(100) )nolog;
outreg using G:\research\WNV3.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // gives the model R2
estat ic;   // gives the model diagnostics
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates the model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on the model residuals

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf3 zresinc3 zrespat3 time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf3=zresinc3=zrespat3=0;   // equality test of standardized residuals

**Model 4
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;
predict fhat4, xb;
adjust, exp;
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate urban
lnpopdense lnarea road time2-time5;
predict yhat4, xb;
xtnbreg lag1 educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate urban
lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;
predict slaghat4, xb;
adjust, exp;
xtnbreg human fhat4 yhat4 mosquito lnpopdense pdsi time2-time5 slaghat4, i(location) iterate(100)
vce(bootstrap, rep(100) )nolog;
outreg using E:\research\WNV3.doc, title(FE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

**********TABLE 2-12 MODELS 1-4 (4 COLUMNS)**********

```
**Model 1 Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseco lnincome  lnpopdense time2-time5 e1 e2 e11 e21 e28 e29 e35 e58, i(location)
iterate(100) nolog;
outreg using G:\research\WNV4.doc, title(FE-IV) ctitle(Model 1)3aster replace;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays the R2
estat ic;   // gives the model diagnostics
predict hhat1, xb;   // generates the predicted values of the model
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

**Model 2 Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseco lnincome mosquito lnpopdense precip time2-time5 e1 e4 e11 e21 e29 e35 e58,
i(location) iterate(100) nolog;
outreg using G:\research\WNV4.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays the R2
estat ic;   // gives model diagnostics
predict hhat2, xb;   // generates the model predicted values
gen resh2=human-hhat2;   // generates the model residuals
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
```

```
**Model 3 Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16).
xtnbreg human fcloseco lnincome mosquito lnpopdense temp time2-time5 e1 e2 e4 e11 e21 e28 e29 e35
e58, i(location) iterate(100) nolog;
outreg using G:\research\WNV4.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays model R2
estat ic;   // gives model diagnostics
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test

**Model 4
xtnbreg human fcloseco lnincome  lnpopdense mosquito pdsi time2-time5 e4 e11 e21 e29 e35, i(location)
iterate(100)nolog;
outreg using E:\research\WNV4.doc, title(FE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

**************************TABLE 2-13 MODELS 1-4 (4 COLUMNS)****************************

```
**Model 1 IV Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes home
foreclosures
predict fhat1, xb;   // generates home foreclosures predicted values
adjust, exp;
gen resf=fcloseco-fhat1;   // generates residuals from home foreclosures regression
egen zresf=std(resf);   //  generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of income
predict yhat1, xb;   // generates the predicted values of log of income
gen resinc=lnincome-yhat1;   // generates residuals from income regression
egen zresinc=std(resinc);   // generates standardized income residuals
xtnbreg human fhat1 yhat1 lnpopdense time2-time5 e2 e4 e6 e22, i(location) iterate(100) vce(bootstrap,
rep(100)) nolog;
outreg using G:\research\WNV5.doc, title(RE-IV) ctitle(Model 1)3aster replace;
predict hhat1, xb;   // generates the model predicted values
gen resh1=human-hhat1;   // generates the model residuals
spatgsa resh1, weights(W1) moran geary;   // spatial autocorrelation test on model residuals
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays the model R2
estat ic;   //gives the model diagnostics

***Validity/Over-Identifaction Tests
xtnbreg human fhat1 yhat1 bird mosquito migrate time2-time5 slaghat1,
i(location)iterate(100)exposure(lnpop)nolog;
predict wnvhat,xb;
gen reswnv=human-wnvhat;   // generates residuals from IV regression
egen zreswnv=std(reswnv);   // generates standardized IV residuals
gen zreswnv1=zreswnv+1;   // transform standardized residuals
sum(zreswnv1);
xtnbreg zreswnv1 precip temp pdsi educate equine unrate poverty bird mosquito migrate airport pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   //regresses standardized
residuals on instruments
```

```
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // gives the model R2 to calculates test statistics
/*predict reshat, xb;
gen essres=reshat-1;
gen essres2=(essres)^2;
sum(essres2);
gen tssres=reswnv-1;
gen tssres2=(tssres)^2;
sum(tssres2);
predict reswnvhat, xb;
adjust, exp;
list reswnv;*/

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf zresinc  time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf=zresinc=0;   // equality test of standardized residuals

**Model 2 IV Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5,i(location)iterate(100)nolog;   //instrumentalizes home
foreclosures
predict fhat2, xb;   // generates the predicted values of home foreclosures
adjust, exp;
gen resf2=fcloseco-fhat2;   // generates home foreclosures residuals
egen zresf2=std(resf2);   // generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;
predict yhat2, xb;
gen resinc2=lnincome-yhat2;   // generates income residuals
egen zresinc2=std(resinc2);   // generates standardized income residuals
xtnbreg human fhat2 yhat2 mosquito lnpopdense precip time2-time5 e4 e6 e22,i(location) iterate(100)
vce(bootstrap, rep(100)) nolog;
outreg using G:\research\WNV5.doc, title(FE-IV) ctitle(Model 2)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays model R2
estat ic;   // gives model diagnostics
predict hhat2, xb;   // generates model predicted values
gen resh2=human-hhat2;   // generates model residuals
spatgsa resh2, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf zresinc  time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf2=zresinc2=0;   // equality test of standardized residuals

**Model 3 IV Spatial Filtering Random Effects Negative Binomial Regressions (Equation 2.16)
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite
elevate urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;   // instrumentalizes home
foreclosures
predict fhat3, xb;   // generates the predicted values of home foreclosures
adjust, exp;
gen resf3=fcloseco-fhat3;   // generates home foreclosure residuals
egen zresf3=std(resf3);   // generates standardized home foreclosure residuals
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi pwhite elevate
urban lnpopdense lnarea road time2-time5;   // instrumentalizes log of income
predict yhat3, xb;   // generates the predicted values of log of income
gen resinc3=lnincome-yhat3;   // generates income residuals
```

```
egen zresinc3=std(resinc3);   // generates the standardized income residuals
xtnbreg human fhat3 yhat3 mosquito lnpopdense temp time2-time5 e2 e4 e6 e22, i(location) iterate(100)
vce(bootstrap, rep(100)) nolog;
outreg using G:\research\WNV5.doc, title(FE-IV) ctitle(Model 3)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);   // displays the model R2
estat ic;   // gives model diagnostics
predict hhat3, xb;   // generates the model predicted values
gen resh3=human-hhat3;   // generates model residuals
spatgsa resh3, weights(W1) moran geary;   // spatial autocorrelation test on model residuals

***Endogeneity Test
xtnbreg human fcloseco lnincome bird mosquito migrate zresf zresinc  time2-time5 lag1,
i(location)exposure(lnpop)iterate(100)nolog;
test zresf3=zresinc3=0;   // equality test of standardized residuals

**Model 4
xtnbreg fcloseco educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate
urban lnpopdense lnarea road time2-time5, i(location)iterate(100)nolog;
predict fhat4, xb;
adjust, exp;
reg lnincome educate equine unrate poverty bird mosquito migrate airport precip temp pdsi elevate urban
lnpopdense lnarea road time2-time5;
predict yhat4, xb;
xtnbreg human fhat4 yhat4  lnpopdense mosquito pdsi time2-time5 e2 e6 e22, i(location) iterate(100)
vce(bootstrap, rep(100)) nolog;
outreg using E:\research\WNV5.doc, title(FE-IV) ctitle(Model 4)3aster append;
display "Pseudo-R^2"=1-e(ll)/e(ll_0);
predict hhat4, xb;
gen resh4=human-hhat4;
spatgsa resh4, weights(W1) moran geary;
estat ic;
```

R Codes

```
##############################################################################################
#R 2.15.2
# Dissertation Chapter 3: Disease Mapping and Presentation of the Variogram
# CA_SMR.R, CO_SMR.R, CA_variogram.R. CO_variogram.R, CA_predict_prob.R, CO_predict_prob.R.
These #files estimates the Standardized Morbidity Ratios , the variograms and predicted probabilities for
California and #Colorado
# It uses data on West Nile Virus and population variables from 2003-2007
#It reproduces Figures 3-1 to 3-22
##############################################################################################
```

 # Calculating the Standardized Morbidity Ratios (SMR) for California

```
############################ Raw SMRs (Equation 4.1) ###################################
#read in data from the file cross_ca0307
cawnv<-read.table("E:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
```

```
#calculate the raw SMR
raw<-human/E
#list the SMR values
raw

################### Poisson-Gamma Empirical Bayes SMR (Equation 4.3) ######################
# read in data from the file cross_ca0307
cawnv<-read.table("E:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Poisson-Gamma SMR
smth<-empbaysmooth(human, E)
smth

## ####################Log-Normal Empircal Bayes SMR (Equation 4.4) ######################
# read in data from the file cross_ca0307
cawnv<-read.table("E:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Log-Normal SMR
smthln<-lognormalEB(human, E)
#exponentiate to get values
ebln<-exp(smthln$smthrr)
ebln

## #################Global Marshall Empircal Bayes SMR (Equation 4.5) ######################
# read in data from the file cross_ca0307
cawnv<-read.table("E:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Global SMR
smtheb<-EBest(human, E)
smtheb
```

#####Calculating the Standardized Morbidity Ratios (SMR) for Colorado##################

```
############### Raw SMR (Equation 4.1) ################################################
# read in data from the file cross_co0307
cownv<-read.table("E:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the Dcluster library
```

```
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the raw SMR
raw<-human/E
raw

## ################Poisson-Gamma Empirical Bayes SMR(Equation 4.3) ##############
# read in data from the file cross_co0307
cownv<-read.table("E:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Poisson-Gamma SMR
smth<-empbaysmooth(human, E)
smth

## ############Log-Normal Empirical Bayes SMR (Equation 4.4) ####################
# read in data from the file cross_co0307
cownv<-read.table("E:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the Dcluster library
library(DCluster)
#define the WNV data

wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Log-Normal SMR
smthln<-lognormalEB(human, E)
#Exponentiate to get value
ebln<-exp(smthln$smthrr)
ebln

## #############Marshall Global Empircal Bayes SMR (Equation 4.5)##########################
# read in data from the file cross_co0307
cownv<-read.table("E:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the Dcluster library
library(DCluster)
#define the WNV data
wnv<-data.frame(Observed=human)
#calculate the expected number of cases
wnv<-cbind(wnv, Expected=population*sum(human)/sum(population))
#calculate the Global SMR
smtheb<-EBest(human, E)
smtheb
```

################## Maps of the SMRs and Predicted Probabilities for California##############

```
## ####################Raw SMR (Equation 3.1) Figure 3-1#################################
#load the maptools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the intervals tools
library(classInt)
#read California shape file containing data
ca_map<- readShapePoly("E:/research/california_wnv_maps.shp")
#Assigns different shades of the color gray
grays=gray.colors(5,start = 1.00, end = 0.3)
#define breaks
brks<-classIntervals(ca_map$smr, n=5, style="equal")
brks<- brks$brks
#plot the map
plot(ca_map,col=grays[findInterval(ca_map$smr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft", legend=leglabs(round(brks)), fill=grays, bty="n")

## Poisson-Gamma Bayes Smoothed (Equation 3.3). This code produces Figure 3-3
#load the maptools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the interval tools
library(classInt)
# read the California shape file containing data
ca_map<- readShapePoly("E:/research/california_wnv_maps.shp")
#define different shades of the color gray
grays=gray.colors(5,start = 1.00, end = 0.3)
#define a class intervals and breaks points
brks<-classIntervals(ca_map$bayesmr, n=5, style="equal")
brks<-brks$brks
#plot the map
plot(ca_map,col=grays[findInterval(ca_map$bayesmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft", legend=leglabs(round(brks)), fill=grays, bty="n")

######## Log-Normal Empirical Bayes (Equation 3.4). This code produces Figure 3-5###############
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
load the interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/california_wnv_maps.shp")
#define different shades of the color gray
grays=gray.colors(5,start = 1.00, end = 0.3)
#define class intervals and break points
brks<-classIntervals(ca_map$lnsmr, n=5, style="equal")
brks<- brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$lnsmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)) , fill=grays, bty="n")
```

```
######## Marshall Global Empirical Bayes (Equation 3. 5) Figure 3-7############################
#load the maptools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/california_wnv_maps.shp")
#define different shades of the color gray
grays=gray.colors(5,start = 1.00, end = 0.3)
#define class intervals and break points
brks<-classIntervals(ca_map$ebsmr, n=5, style="equal")
brks<- brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$ebsmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)), fill=grays, bty="n")


###################### Predicted Probabilities from (Equation 3.6)
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/california_wnv_maps.shp")
#define different shades of gray and break points
grays=gray.colors(5,start = 1.00, end = 0.3)
#define intervals and break points
brks<-classIntervals(ca_map$prob, n=5, style="equal")
brks<- brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$prob,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)), fill=grays, bty="n")
```

##############Maps of the SMRs and Predicted Probabilities for Colorado#################

```
## Raw SMR (Equation 3.1). This produces Figure 3-1#######################################
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/colorado_wnv_maps.shp")
#define different shades of gray
grays=gray.colors(3,start = 1.00, end = 0.3)
#define intervals and break points
brks<-classIntervals(ca_map$smr, n=3, style="equal")
brks<- brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$smr,brks,all.inside=TRUE)],axes=F)
```

```
legend("bottomleft", legend=leglabs(round(brks)), fill=grays, bty="n")

## ###########Poisson-Gamma Bayes Smoothed (Equation 3.3) Figure 3-4#######################
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/colorado_wnv_maps.shp")
#define different shades of gray
grays=gray.colors(3,start = 1.00, end = 0.3)
#define intervals and break points
brks<-classIntervals(ca_map$bayesmr, n=3, style="equal")
brks<-brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$bayesmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft", legend=leglabs(round(brks)), fill=grays, bty="n")

## ########Log-Normal Empirical Bayes (Equation 3.4) Figure 3-6#############################
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/coorado_wnv_maps.shp")
#define different shades of the color gray
grays=gray.colors(3,start = 1.00, end = 0.3)
#define class intervals and break points
brks<-classIntervals(ca_map$lnsmr, n=3, style="equal")
brks<- brks$brks
#plot the SMR
plot(ca_map,col=grays[findInterval(ca_map$lnsmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)) , fill=grays, bty="n")

################ Marshall Global Empirical Bayes (Equation 3.5) Figure 3-8####################
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/colorado_wnv_maps.shp")
#define different shades of the color gray
grays=gray.colors(3,start = 1.00, end = 0.3)
#define class intervals and break points
brks<-classIntervals(ca_map$ebsmr, n=3, style="equal")
brks<- brks$brks
#plot the SMRs
plot(ca_map,col=grays[findInterval(ca_map$ebsmr,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)) , fill=grays, bty="n")
```

```
## #############Predicted Probabilities from Equation 3.6 Figure 3-10##########################
#load the map tools
library(maptools)
#load the color tools
library(RColorBrewer)
#load the class interval tools
library(classInt)
#read the California shape file containing data
ca_map<- readShapePoly("E:/research/colorado_wnv_maps.shp")
#define different shades
grays=gray.colors(3,start = 1.00, end = 0.3)
#define a class interval and break points
brks<-classIntervals(ca_map$prob, n=3, style="equal")
brks<- brks$brks
#plot the SMRs
plot(ca_map,col=grays[findInterval(ca_map$prob,brks,all.inside=TRUE)],axes=F)
legend("bottomleft",legend=leglabs(round(brks)), fill=grays, bty="n")
```

################################## Variogram Plots for California#####################

```
## #########Isotropic Variogram for Observed WNV using equation 3.11 Figure 3-11##############
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#loads the gstat tools
library(gstat)
#assigns coordinates to the data
coordinates(cawnv) <- c("x", "y")
#transforms WNV counts
lnhuman<-log(human+1)
#specify a model to fit
wnv.var2 <- variogram(lnhuman~1,Cressie=TRUE,data=cawnv)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)


## ####Anisotropic Variogram for Observed WNV using equation 3.11 Figure 3-13#################
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#loads the gstat tools
library(gstat)
#assigns coordinates the data
coordinates(cawnv) <- c("x", "y")
#transforms WNV counts
lnhuman<-log(human+1)
#specify a model to fit
wnv.var2 <- variogram(lnhuman~1, locations=~x+y,Cressie=TRUE,data=cawnv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)
```

```
## ######## Isostropic Variogram for the SMR using equation 3.11 Figures 3-11###################
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the gstat tools
library(gstat)
#assigns coordinates to the data
coordinates(cawnv) <- c("x", "y")
#transform SMR
lnsmr<-log(smr)
#specify a model to fit
wnv.var2 <- variogram(lnsmr~1, locations=~x+y,Cressie=TRUE,data=cawnv)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit a variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)

## #########Anisotropic Variogram for the SMR using equation 3.11 Figure 3-13###################
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the gstat tools
library(gstat)
#assign coordinates to the data
coordinates(cawnv) <- c("x", "y")
#transform the SMR
lnsmr<-log(smr)
#generate a scatterplot
variog<-variogram(lnsmr~1,data=cawnv)
hscat(lnsmr ~ 1, data=cawnv,breaks=c(0,5,10))
#specify a model to fit
wnv.var2 <- variogram(lnsmr~1, locations=~x+y,Cressie=TRUE,data=cawnv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)

## #########Isotropic Variogram for the Model Residuals from equation 3.6 Figure 3-19############
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the gstat tools
library(gstat)
#assign coordinates to the data
coordinates(cawnv) <- c("x", "y")
#generate  scatterplot
hscat(residuals ~ 1, data=residuals,breaks=c(0,5,10))
#specify a model to fit
wnv.var2 <- variogram(residuals~1, locations=~x+y,Cressie=TRUE,data=cawnv)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)
```

```
############ Anisotropic Variogram for the Model Residuals from equation 3.6 Figure 3-21########
#reads a data file cross_ca0307.csv
cawnv<-read.table("G:/research/cross_ca0307.csv",header=T,sep=",")
attach(cawnv)
#load the gstat tools
library(gstat)
#assign coordinates to the data
coordinates(cawnv) <- c("x", "y")
#generate a scatterplot
hscat(residuals ~ 1, data=cawnv,breaks=c(0,5,10))
#specify a model to fit
wnv.var2 <- variogram(residuals~1, locations=~x+y,Cressie=TRUE,data=cawnv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Exp", nugget=1, range=60)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)
```

######################Variograms for Colorado###################################

```
## #####Isotropic Variograms for Observed WNV Counts using equation 3.12 Figure 3-12###########
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#loads the gstat tools
library(gstat)
#assigns coordinates to the data
coordinates(cownv) <- c("x", "y")
#transforms WNV counts
lnhuman<-log(human+1)
#generate a scatterplot
hscat(human ~ 1, data=cownv, breaks=c(0,5,10,15))
#specify a model to fit
wnv.var2 <- variogram(lnhuman~1, locations=~x+y, Cressie=TRUE,data=cownv)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the model
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)
######## Anisotropic Variograms for Observed WNV Counts using equation 3.12 Figure 3-14########
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#loads the gstat tools
library(gstat)
#assigns coordinates to the data
coordinates(cownv) <- c("x", "y")
#transforms the WNV counts
lnhuman<-log(human+1)
#generates a scatterplot
hscat(human ~ 1, data=cownv, breaks=c(0,5,10,15))
#specify a model to fit
wnv.var2 <- variogram(lnhuman~1, locations=~x+y, Cressie=TRUE,data=cownv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
```

```
#plot the fitted variogram
plot(wnv.var2,fit.variog)

##########Isotropic Variograms for the SMR using equation 3.12 Figure 3-15###################
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
library(gstat)
#assigns coordinates to the data
coordinates(cownv) <- c("x", "y")
#specify a model to fit
wnv.var2 <- variogram(smr~1, locations=~x+y, Cressie=TRUE,data=cownv)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the variogram
plot(wnv.var2,fit.variog)

## Anisotropic Variograms for the SMR using equation 3.12. This produces Figure 3-17.
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#loads the gstat toos
library(gstat)
#assigns the coordinates to the data
coordinates(cownv) <- c("x", "y")
#specify a model to fit
wnv.var2 <- variogram(smr~1, locations=~x+y, Cressie=TRUE,data=cownv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)


#########Isotropic Variograms for Model Residuals derived from equation 3.6 Figure 3-20###########
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the gstat tools
library(gstat)
#assign coordinates to the data
coordinates(cownv) <- c("x", "y")
#specify a model to fit
wnv.var2 <- variogram(residuals~1, locations=~x+y, Cressie=TRUE,data=cownv)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)

##########Anisotropic Variograms for Model Residuals derived from equation 3.6 Figure 3-21#######
#reads a data file cross_co0307.csv
cownv<-read.table("G:/research/cross_co0307.csv",header=T,sep=",")
attach(cownv)
#load the gstat tools
```

```
library(gstat)
#assign coordinates to the data
coordinates(cownv) <- c("x", "y")
#specifies a model to fit
wnv.var2 <- variogram(residuals~1, locations=~x+y, Cressie=TRUE,data=cownv,alpha=c(0:3)*90)
model.variog <-vgm(psill=1, model="Gau",nugget=800,range=1)
#fit the variogram
fit.variog<-fit.variogram(wnv.var2, model.variog)
#plot the fitted variogram
plot(wnv.var2,fit.variog)
```

Calculating Predicted Probabilities for California

```
#reads a data file wnvco0307.csv
cawnv<-read.table("G:/research/wnvca0307.csv",header=T,sep=",")
attach(cawnv)
#generate the exposure and offset variables
lnpop<-log(population)
lnpop2=log(lnpop)
#transform home foreclosures
fcloseca<-log(fclose+1)
#transform income
lnincome<-log(income)
#generate a migration rate
migrate=netmigr/1000
#generate the year/time fixed effects dummies
dum<-factor(year)
dummies<-model.matrix(~dum)
dummies
#instrumentalize income
reg1<-
lm(lnincome~migrate+bird+mosquito+precip+temp+pdsi+poverty+unrate+education+airport+equine+elev
ation+area+urban+factor(year),data=cawnv)
lp<-predict(reg1)
#instrumentalize home foreclosures
zinbf<-
lm(fcloseca~migrate+bird+mosquito+precip+temp+pdsi+poverty+unrate+education+airport+equine+elevat
ion+area+urban+factor(year),data=cawnv,dist="negbin")
pf<-predict(zinbf,type="count")
exppf<-exp(pf)
zinb<-zeroinfl(human~pf+lp+lnpopdense+factor(year)+e4+e6+e15+e16+e46+e47,data=cawnv,
dist="negbin")
summary(zinb)
#generate predicted probabilities
Prediction<-predict(zinb,type="prob")
Prediction
sum_pred<-rowSums(Prediction)
sum_pred
prhat<-predprob(zinb, newdata=pr)
prhat
```

Calculating Predicted Probabilites for Colorado

```
#reads a data file wnvco0307.csv
cawnv<-read.table("G:/research/wnvco0307.csv",header=T,sep=",")
attach(cawnv)
#generate the exposure and offset variables
lnpop<-log(population)
```

```
#transform home foreclosures
fcloseca<-log(fclose+1)
#transform income
lnincome<-log(income)
#generate a migration rate
migrate=netmigr/1000
#generate time fixed effects dummies
dum<-factor(year)
dummies<-model.matrix(~dum)
dummies
#instrumentalize income
reg1<-
lm(lnincome~migrate+bird+mosquito+precip+temp+pdsi+poverty+unrate+education+airport+equine+elev
ation+area+urban+factor(year),data=cawnv)
lp<-predict(reg1)
instrumentalize home foreclosures
zinbf<-
lm(fcloseca~migrate+bird+mosquito+precip+temp+pdsi+poverty+unrate+education+airport+equine+elevat
ion+area+urban+factor(year),data=cawnv,dist="negbin")
pf<-predict(zinbf,type="count")
exppf<-exp(pf)
#run negative binomial regression
zinb<-zeroinfl(human~pf+lp+lnpopdense+factor(year)+e4+e6+e15+e16+e46+e47,data=cawnv,
dist="negbin")
summary(zinb)
#generate predicted probabilites
Prediction<-predict(zinb,type="prob")
Prediction
sum_pred<-rowSums(Prediction)
sum_pred
prhat<-predprob(zinb, newdata=pr)
prhat
```

Excel Code

```
##############################################################################
```

#Chapter 4: Economics of Antibiotic Resistance: Impact of Animal Antibiotic Use on the #Evolution of Resistance in Humans.

The file resistance_simulation_1 contains data obtained from several sources and the results of #the simulations. It also contains the code used to simulate the time paths of antibiotic #effectiveness using Euler's method.

```
##############################################################################
```

The time path of antibiotic effectiveness is represented by the differential equation:

$$\frac{dw}{dt} = w(t)\big(ku_a(t) + zu_h(t)\big)(w(t) - 1) = \dot{w} \qquad (1)$$

Initial values of the parameters are given as follows:

$u_h = 0.001$, $g = ku_a = 1.4$, $\beta = 0.001$, $I_t = 0.00029$, $S_t = 0.99971$, $b_h = 2000$, $b_a = 30.1$, $c_1 = 4.24$, $c_2 = 68$, $c_a = 30$, $u_a = 2$, $k = 0.70$, $z = 0.3$, $\delta = 0.004$, $h = 0.10$, $w_t = 0.96$

Euler's method in approximating the solution to (1) is $w + hw'$, where w' represents the differential equation in (1). The Excel code is obtained by typing

=$B$16*(G43)^2-G43*$B$16

into the appropriate cell. Repeating this for $g = ku_a = 2.8$ and $g = ku_a = 4.2$ produces Figure 4.5 when the fill down command is used forward the system

This code is used to simulate the time paths of the shadow value of antibiotic effectiveness using Euler's method. The time path of the shadow value of antibiotic effectiveness is represented by the differential equation:

$$\dot{\mu}_1 = \frac{\partial \mu_1}{\partial t} = \delta\mu_1 - b_h I u_h - \mu_1(ku_a + zu_h)(2w - 1) + \mu_2 u_h I \quad (2)$$

Euler's method in approximating the solution to (2) is $\mu_1 + h\mu_1'$, where μ' represents the differential equation in (2). The Excel code is obtained by typing

=($B$15-($B$13*$B$2+$B$14*$B$1)*(2*E43-1))*P43-
$B$8*G83*$B$1+P6*$B$1*G83

into the appropriate cell. Repeating this for $g = ku_a = 2.8$ and $g = ku_a = 4.2$ produces Figure 4.6 when the fill down command is used to forward the system

Appendix E: Supplementary IV Spatial Filtering RENB Regressions

IV Spatial Filtering RENB for California

	Model 1	Model 2	Model 3	Model 4
FORHAT	0.448	0.302	0.324	0.315
	(2.66)***	(1.82)*	(1.67)*	(1.91)*
INCHAT	-2.680	-1.515	-1.571	-1.232
	(2.15)**	(1.68)*	(1.66)*	(1.15)
LNPOPDENSE	0.182	0.139	0.127	0.094
	(0.89)	(0.67)	(0.56)	(0.48)
MOSPOP	0.000	0.002	0.001	0.001
	(1.23)	(0.88)	(0.66)	(0.74)
D2004	2.994	2.881	2.876	2.743
	(0.89)	(0.87)	(0.66)	(0.89)
D2005	3.936	3.635	3.667	3.950
	(1.17)	(1.11)	(0.84)	(1.29)
D2006	3.727	3.429	3.437	3.456
	(1.09)	(1.04)	(0.78)	(1.12)
D2007	3.668	3.392	3.353	2.917
	(1.07)	(1.04)	(0.77)	(0.94)
VEC4	-2.949			
	(3.88)***			
VEC6	1.404			
	(1.11)			
VEC15	-1.473			
	(1.66)*			
VEC16	-2.118	-1.642	-1.646	-1.620
	(1.99)**	(1.58)	(1.44)	(1.63)
VEC46	2.103			
	(2.13)**			
VEC47	1.331			
	(1.15)			
MOSQUITO		0.015	0.015	0.014
		(1.30)	(1.01)	(1.16)
PRECIP		0.049		
		(0.43)		
TEMP			-0.018	
			(0.49)	
PDSI				-0.113
				(1.43)
VEC41				0.053
				(0.05)
CONSTANT	33.363	23.765	24.863	20.639
	(2.52)**	(1.86)*	(1.89)*	(1.59)
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1063.487	1082.009	1081.533	1080.793
BIC	1125.875	1133.388	1132.911	1135.841
r	1.506	1.069	1.085	1.196
	S.E. (4.871)	S.E. (2.510)	S.E. (2.810)	S.E. (3.118)
s	5.756	2.619	2.756	3.346
	S.E. (19.961)	S.E. (7.108)	S.E. (8.268)	S.E. (10.008)
Moran's I on Residuals (p value)	0.193	0.208	0.205	0.197

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
MOSPOP is an interaction term between MOSQUITO and POPDENSE

IV Spatial Filtering RENB for California

	Model 1	Model 2	Model 3	Model 4
FORHAT	0.406	0.404	0.440	0.418
	(2.16)**	(2.43)**	(2.41)**	(2.51)**
INCHAT	-4.183	-1.428	-1.415	-1.196
	(3.72)***	(1.68)*	(1.67)*	(1.11)
LNPOPDENSE	0.417	0.025	0.002	-0.009
	(2.25)**	(0.12)	(0.01)	(0.05)
D2004	3.288	2.925	2.912	2.803
	(1.09)	(0.89)	(0.68)	(0.91)
D2005	4.294	3.742	3.738	4.014
	(1.41)	(1.14)	(0.87)	(1.33)
D2006	4.114	3.494	3.473	3.516
	(1.33)	(1.06)	(0.81)	(1.15)
D2007	4.157	3.413	3.367	3.007
	(1.34)	(1.04)	(0.79)	(0.97)
VEC4	-3.263			
	(4.08)***			
VEC6	1.638			
	(1.20)			
VEC15	-1.630			
	(1.76)*			
VEC16	-2.201	-1.799	-1.794	-1.774
	(1.74)*	(1.67)*	(1.59)	(1.69)*
VEC46	2.402			
	(2.22)**			
VEC47	1.319			
	(1.21)			
MOSPOP		0.001	0.001	0.001
		(1.59)	(1.21)	(1.65)*
PRECIP		0.023		
		(0.21)		
TEMP			-0.023	
			(0.63)	
PDSI				-0.102
				(1.29)
VEC41				0.019
				(0.02)
CONSTANT	48.989	22.315	23.830	18.846
	(4.12)***	(1.78)*	(1.90)*	(1.48)
Observations	290	290	290	290
Number of location	58	58	58	58
AIC	1044.021	1058.4467	1056.1735	1058.7184
BIC	1102.7391	1106.1551	1100.2121	1110.0968
r	1.292	1.277	1.270	1.320
	S.E. (4.054)	S.E. (5.429)	S.E. (4.521)	S.E. (4.117)
s	3.042	2.833	2.795	3.178
	S.E. (10.048)	S.E. (13.073)	S.E. (10.879)	S.E. (11.032)
Moran's I	0.196	0.200	0.200	0.196

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
MOSPOP is an interaction term between MOSQUITO and POPDENSE

IV Spatial Filtering RENB for Colorado

	Model 1	Model 2	Model 3	Model 4
FORHAT	2.789	2.860	2.937	2.975
	(9.39)***	(7.52)***	(6.44)***	(8.67)***
INCHAT	-3.555	-4.582	-3.718	-4.084
	(2.83)***	(3.82)***	(2.58)***	(3.54)***
LNPOPDENSE	0.243	0.135	0.203	0.056
	(1.31)	(0.52)	(1.22)	(0.34)
MOSPOP	0.001	0.000	0.001	0.001
	(1.67)*	(0.07)	(0.11)	(0.30)
D2004	-2.589	-2.464	-2.656	-2.773
	(10.11)***	(7.43)***	(11.01)***	(11.99)***
D2005	-2.901	-2.865	-2.971	-3.124
	(16.39)***	(13.10)***	(14.29)***	(12.13)***
D2006	-2.866	-2.779	-2.953	-3.036
	(12.26)***	(10.61)***	(12.57)***	(13.41)***
D2007	-2.874	-2.751	-3.049	-3.142
	(11.20)***	(9.84)***	(9.07)***	(9.65)***
VEC2	4.559		4.704	4.000
	(3.59)***		(3.73)***	(2.81)***
VEC4	3.312	2.535	3.298	
	(2.81)***	(1.72)*	(2.72)***	
VEC6	3.284	1.935	3.167	2.743
	(3.47)***	(1.91)*	(2.78)***	(2.62)***
VEC22	1.502	1.603	1.623	1.666
	(1.86)*	(1.87)*	(1.64)	(2.06)**
MOSQUITO		0.007	0.010	0.013
		(0.36)	(0.41)	(0.63)
PRECIP		-0.627		
		(0.69)		
TEMP			-0.030	
			(0.31)	
PDSI				0.028
				(0.37)
CONSTANT	33.217	45.097	36.250	39.348
	(2.54)**	(3.78)***	(2.08)**	(3.28)***
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1204.822	1223.534	1207.769	1215.312
BIC	1261.11	1283.575	1271.562	1275.353
r	2.739	2.559	2.871	2.487
	S.E. (1.017)	S.E. (4.122)	S.E. (10.145)	S.E. (0.821)
s	6.883	6.073	7.308	5.041
	S.E. (3.134)	S.E. (9.924)	S.E. (27.732)	S.E. (1.723)
Moran's I on Residuals (p value)	0.125	0.129	0.125	0.122

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
MOSPOP is an interaction term between MOSQUITO and POPDENSE

IV Spatial Filtering RENB for Colorado

	Model 1	Model 2	Model 3	Model 4
FORHAT	2.631	2.821	2.877	2.928
	(9.80)***	(6.22)***	(7.21)***	(7.80)***
INCHAT	-3.737	-4.536	-3.666	-4.054
	(3.28)***	(3.83)***	(2.48)**	(3.92)***
LNPOPDENSE	0.326	0.140	0.213	0.059
	(1.63)	(0.62)	(1.21)	(0.38)
D2004	-2.692	-2.412	-2.605	-2.695
	(10.96)***	(6.92)***	(9.55)***	(13.78)***
D2005	-2.968	-2.808	-2.903	-3.010
	(14.49)***	(10.82)***	(13.25)***	(11.16)***
D2006	-2.923	-2.719	-2.881	-2.934
	(9.46)***	(10.04)***	(11.85)***	(11.03)***
D2007	-2.689	-2.665	-2.935	-2.977
	(12.21)***	(9.04)***	(10.35)***	(10.46)***
VEC2	4.392		4.685	3.981
	(3.56)***		(3.75)***	(2.91)***
VEC4	3.269	2.637	3.381	
	(2.98)***	(1.72)*	(2.52)**	
VEC6	3.319	1.950	3.206	2.761
	(2.86)***	(2.03)**	(3.03)***	(2.47)**
VEC22	1.866	1.579	1.564	1.634
	(1.82)*	(1.73)*	(1.82)*	(1.67)*
MOSPOP		0.001	0.001	0.001
		(1.06)	(1.77)*	(1.84)*
PRECIP		-0.670		
		(0.71)		
TEMP			-0.027	
			(0.28)	
PDSI				0.020
				(0.25)
CONSTANT	35.232	44.684	35.582	39.025
	(3.02)***	(3.75)***	(2.03)**	(3.66)***
Observations	315	315	315	315
Number of location	63	63	63	63
AIC	1221.647	1222.055	1206.626	1215.023
BIC	1274.183	1278.344	1266.667	1271.311
r	2.438	2.521	2.798	2.430
	S.E. (1.351)	S.E. (4.145)	S.E. (6.430)	S.E. (0.903)
s	6.376	6.083	7.228	5.035
	S.E. (3.206)	S.E. (10.350)	S.E. (17.563)	S.E. (1.936)
Moran's I	0.125	0.130	0.125	0.123

z statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
MOSPOP is an interaction term between MOSQUITO and POPDENSE

REFERENCES

Alliance for the Prudent Use of Antibiotics. 2012. Informing Public Policy on Agricultural Use of Antimicrobials in the United States: Strategies Developed by an NGO. http://www.tufts.edu/med/apua/research/completed_projects_11_2340293929.pdf (accessed on January 20 2012).

Anselin, L. 1983. *Spatial Econometrics: Methods and models*. Dordrecht: Kluver Academic Publishers

Anselin, L. 2001. "Spatial Effects in Econometric practice in environmental and resource economics" *American Journal of Agricultural Economics* 83: 705-710.

Anselin, L. 2002. "Under the hood: Issues in the specification and interpretation of spatial regression models." *Agricultural Economics* 27: 247-267.

Bailey, T.C., and A.C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Warwickshire: Pearson

Bates, J., Jordens, J.Z., and D.T. Griffiths. 1994. Farm Animals as a Putative Reservoir for Vancomycin-Resistant Enterococcal Infection in Man. *Journal of Antimicrobial Chemotherapy* 34: 507-516.

Bisht, R., Katiyar, A., Singh, R., and Mittal, P. 2009. Antibiotic Resistance-A Global Issue of Concern. *Asian Journal of Pharmaceutical and Clinical Research* 2: 34-39.

Bibble, N., Kennedy, S., and McDonald, J.T. 2007. Health Assimilation Patterns Amongst Australian Immigrants. *The Economic Record* 83: 16-30.

Bonhoeffer, S.M., and Lipsitch, M. 1997. Evaluating treatment protocols to prevent antibiotic resistance. *Proceedings in the National Academy of Science* 94:12106-11.

Brown, G., and Layton, D.F. 1996. Resistance economics: Social Cost and the Evolution of Antibiotic Resistance. *Environmental and Development Economics* 1:349-355.

Brown, H.E., Childs, J.E., Duik-Wasser, M.A., and Durland, F. 2008. Ecological factors associated with West Nile virus transmission in Northeastern United States. *Emerging Infectious Diseases* 14: 1539-1545.

Brown, Z., Dickson, K., and Kramer, R. 2009. The bioeconomics of insecticide resistance in the optimal control of vector-borne diseases.

Brownstein, J.S., Rosen, H., Purdy, D., Miller, J.R., Merlino, M., Mostashari, F., and Fish, D. 2002. Spatial Analysis of West Nile Virus: Rapid Risk Assessment of an Introduced Vector-Borne Zoonosis. *Vector Borne and Zoonotic Diseases* 2:157-163.

California Department of Public Health, Mosquito and Vector Control Association of California, and University of California, Davis. 2011. California Mosquito-Borne Virus

Surveillance and Response Plan. http://www.westnile.ca.gov/resources.php (accessed on December 30 2011).

Campbell, G.L., Marfin, A.A., Lanciotti, R.S., and Gubler, D.J. 2002. West Nile virus. *Lancet Infectious Diseases* 2: 519-29.

Carney, R.M., Ahearn, S.C., McConchie, A., Glaser, C., Jean, C., Barker, C., Park, B., Padgett, K., Parker, E., Aquino, E., and Kramer, V. 2011. Early Warning System for West Nile Virus Risk Areas, California, USA. *Emerging Infectious Diseases* 17: 1445-1454.

CDC. 2011.West Nile Virus-Statistics, Surveillance and Control Archive http://www.cdc.gov/ncidod/dvbid/westnile/city_states.htm (accessed on May 2011).

Center on Social Disparities in Health. 2011. Why should physicians be concerned with the social determinant of Health. University of California, San Francisco.

Chadwick, P.R., Woodford, N., Kaczmarski, E.B., Gray, S., Barreli, R.A., and Oppenheim, B.A. 1996. Glycopeptide-resistant enterococci isolated from uncooked meat. *Journal of Microbial Chemotherapy* 38: 908-909.

Chapin et al. 2004. Airborne Multi-drug Resistant Bacteria Isolated From a Concentrated Swine Feeding Operation." *Environmental Health Perspectives* 113: 137-142.

Choynowski, M. 1959. Map based on probabilities. *Journal of the American Statistical Society* 54: 385-388.

Animal Antibiotic Overuse Hurting Humans?. http://www.cbsnews.com/stories/2010/02/09/eveningnews/main619153. (accessed on September 25 2011).

Charron, D. 2005, *Zoonoses: How climate change affects the modes by which diseases are passed from animals to humans*, Division of Infectious Food borne, Waterborne and Zoonotic Diseases at the Public Health Agency of Canada.

Chase, J.M., and Knight, T.M. 2003. Drought-induced mosquito outbreaks in wetlands. *Ecology Letters* 6: 1017-1024.

Chee-Sanford, J.C., Aminov, R.I., Krapac, I.J., Garrigues-Jeanjean, N., and Mackie, R.I. 2001. Occurrence and Diversity of Tetracycline Resistant Genes in Lagoons and Groundwater Underlying Two Swine Production Facilities. Applied Environmental Microbiology 201: 1494-1502.

Childs, J.E., Krebs, J.W., Real, L.A., and Gordon, E.R. 2007. Animal-based National Surveillance for Zoonotic Disease: Quality, Limitations, and Implications of a Model System for Monitoring Rabies. Prev. Vet Med 78: 246-261.

Choynowski, M. 1959. Map based on probabilities. *Journal of American Statistical Society* 54: 385-388.

Coast J., and Smith. R.D. 2003. Antimicrobial resistance, cost and containment. *Expert Rev. Anti Infect. Ther* 1:241-251.

Cohen, F.L., and Tartasky, D. 1997. Microbial Resistance to Drug Therapy: A Review. *American Journal of Infection Control* 25: 51-64.

Cohen, M.L., and Tauxe, R.V. 1986. Drug-Resistant Salmonella in the United States: An Epidemiologic Perspective. *Science* 234: 964-969.

Colorado Department of Local Affairs. 2011. Foreclosure Reports and Statistics http://www.colorado.gov/cs/Satellite?c=Page&childpagename=DOLA-Main%2FCBONLayout&cid=1251596868532&pagename=CBONWrapper (accessed on December 10 2011).

Colorado Department of Public Health and Environment. 2002. West Nile Virus and Mosquito-Borne Viruses in Colorado. http://www.cdphe.state.co.us/dc/Zoonosis/zoonosis.asp (accessed on December 18 2011).

Clayton, D., and Kaldor, J. 1987. Empirical Bayes estimates of the age-standardized relative risks for use in disease mapping. *Biometrics* 43: 671-681.

Coast, J., and Smith R.D. 2003. Antimicrobial resistance, cost and containment. *Expert Rev. Antibiotic Infection Ther* 1: 241-251.

Council for Agricultural Science and Technology (CAST). 1981. Antibiotics in Animal Feeds. Report No.88, Ames, Iowa.

Cromwell, G.L. 2002. Why and How Antibiotics are Used in Swine Production. *Animal Biotechnology* 13: 7-27.

Culyer, A. 1971. The Nature of the Commodity 'Health Care' and its Efficient Allocation. *Oxford Economic Papers* 23: 189-211.

D'Agata, E.M.C., Magal, P., Oliver, D., Ruan, S., and Webb, G.F. 2007. Modeling antibiotic resistance in hospitals: The impact of minimizing treatment duration. *Journal of Theoretical Biology* 249: 487-499.

Data Quick News. 2011. Real Estate News and Custom Data. http://dqnews.com/Articles/2011/News/California/CA-Foreclosures/RRFor111018.aspx (accessed on December 11 2011).

DeGroote, J.P., Sugumaran, R.S., Brend, S.M., Tucker, B.J., and Bartholomay, L.C. 2008. Landscape, demographic, entomological, and climatic associations with human

disease incidence of West Nile virus in the state of Iowa, USA. *International Journal of Health Geographics* 7: 1-16.

Epstein, P.R., and Delfilippo, C. 2001. West Nile Virus and Drought. *Global Change and Human Health* 2: 2-4

Erens, B., Primatesta, P., and Prior, G. 2001. Health Survey for England 1999: the health of minority ethnic groups. London: The Stationery Office.

Ettner, S. 1996. New Evidence on the Relationship between Income and Health. *Journal of Health Economics* 15: 6-85.

Fagerberg, D.J., and Quarles, C.L 1979. "Antibiotic Feeding, Antibiotic Resistance and Alternatives." American Hoechst Corporation, Animal Health Division. Somerville, New Jersey.

Federal Reserve. Mortgage Delinquencies and Foreclosures. http://ww.federalreserve.gov/newsevents/speech/bernanke2008050a.htm (accessed on September 26 2011).

Farmer, P. 1999. *Infections and Inequality: The Modern Plagues*. Berkeley: University of California Press.

Feinman, S.E. 1979. The Transmission of Antibiotic Resistant Bacteria to People and Animals. In CRC Handbook Series in Zoonoses: Antibiotics, Sulfonamides, and Public Health. Boca Raton: CRC Press.

Fey et al. 2000. Ceftriaxone-Resistant Salmonella Infection Acquired by a Child from Cattle. *New England Journal of Medicine* 324: 1242-49.

Fortin, M.J., and Dale, M. 2005. Spatial analysis: A guide for ecologists. Cambridge: Cambridge University Press.

Frenzen, P.D., Drake, A., and Agulo, F.J. 2005. Economic cost of illness due to Escherichia coli O157 infections in the United States. Economic Research Service, U.S. Department of Agriculture, Washington, DC

Gersovitz, M., and Hammer, J.S. 2004. The economical control of infectious diseases. *The Economic Journal* 492: 1-27.

Getis, A., and Griffith, D.A. 2002. Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis* 36: 90-114.

Gibbs, S.E.J., Wimberly, M.C., Madden, M., Masour, J., Yabsley, M.J., and Stallknecht, D.E. 2006. Factors affecting the geographic distribution of West Nile virus in Georgia, USA. *Vector-Borne and Zoonotic Diseases* 6: 73-82.

Goddard, L.B., Roth, A.E., Reisen, W.K., and Scott, T.W. 2002. Vector Competence of California Mosquitoes for West Nile Virus. *Emerging Infectious Diseases* 8: 1385-1391.

Gomez, A., Kilpatrick, A.M., Kramer, L.D., Dupuis, A.P.I., J.G. Maffei et al. 2008. Land use and West Nile virus seroprevalence in wild animals. *Emerging Infectious Diseases* 14: 962-965.

Grace Communication Foundation (2011), 'Antibiotics', [Available at] http://www.sustainabletable.org/issues/antibiotics/ (accessed on September 25 2011).

Graham, J.P., Boland, J.J., and Silbergeld, E. 2007. Growth Promoting Antibiotics in Food Animal Production: An Economic Analysis. *Public Health Reports* 122: 79-87.

Griffith, D. A. 1981. Towards a Theory of Spatial Statistics: A Rejoinder. *Geographical Analysis* 13: 91-93.

Griffith, D.A. 1996. Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data. *The Canadian Geographer* 40: 351-367.

Griffith, D.A. 2000. A Linear Regression Solution to the Spatial Autocorrelation Problem. *Journal of Geographical Systems* 2: 141-156.

Griffith, D.A. 2003. Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization. Springer, New York.

Grundmann, H. M., De-Souza, A., Boyce, M., and Tiemersma, E. 2006. Emergence and resurgence of Methicillin-Resistant Staphylococcus Aureus as a public health threat. *Lancet* 368: 874-885.

Gubler, D.J., Reiter, P., and Ebi, K.L. 2001. Climate variability and change in the United States: potential impacts on vector-and rodent-borne diseases, *Environmental Health Perspective* 109, 223-233.

Gustafson, R.H., and Bowen, R.E. 1997. Antibiotic Use in Animal Agriculture. *Journal of Applied Microbiology* 83: 531-41.

Hack, W.H. 1955. Study of the biology of Triatomainfestans. *Annals of Regional Medicine* 4, 125-147.

Haining, R. 1991. Bivariate Correlation and Spatial Data. *Geographical Analysis* 23: 210-227.

Hanberger et al. 1997. High Incidence of Antibiotic Resistance Among Bacteria in four Intensive Care Units at a University Hospital in Sweden. *Scandinavian Journal of Infectious Diseases* 29: 607-614.

Harrigan, R.J., Thomassen, H.A., Cummings, R.F., Buermann, W., Kahn, M.E., and Smith, T.B. 2010. Economic Conditions Predict the Prevalence of West Nile Virus. *Plosone* 5:1-8.

Hausman, J., Hall, B., and Griliches, Z. 1984. Econometric Models for Count Data with Applications to the Patents-R&D Relationship. *Econometrica* 52: 909-938.

Hayes, E.B., Sejvar, J.J., Zaki, S.R., Lanciotti, R.S., Bode, A.V., and Campbell, G.L. 2005. Virology, pathology, and clinical manifestations of West Nile virus disease. *Emerging and Infectious Diseases* 11: 1174-9.

Hayes, R.D., Lamotte, L.C., and Hess, A.D.1960, Enzootic eastern encephalitis activity, *Massachusetts Mosquito News* 20, 85-87.

Hess, A.D., Cherubin, C.E., and Lamotte, L.C. 1963. Relation of temperature to activity of western and St. Louis encephalitis viruses, *American Journal of Tropical Medicine and Hygiene* 12, 657-667.

Hiramatsu, K. 2001. Vancomycin-resistant Staphylococcus aureus: a new model of antibiotic resistance. *Lancet Infect. Dis*. 1:147-155.

Holmegerg, S.D., Solomon, S.L., and Blake, P.A. 1987. Health and economic impacts of antimicrobial resistance. *Review of Infectious Diseases* 9: 1065-1078.

Hueth, D., and U. Regev. 1974. Optimal Agricultural Pest Management with Increasing Pest Resistance. *American Journal of Agricultural Economics* 56:543-552.

Huffington Post. 2012. West Nile Virus: Dallas , Texas Attack Plan Includes Aerial Assault Spray. http://huffingtonpost.com/2012/08/17/dallas-west-nile-virus_n_1791551.html (accessed on September 6 2012).

Huhn, G.D., Austin, C., Langkop, C., Kelly, K., Lucht, R. Lampman, R. Novak, L. Haramis, R. Boker, S. Smith et al. 2005. The emergence of West Nile virus during a large outbreak in Illinois in 2002.*American Journal of Tropical Medicine and Hygiene* 72:768-776.

Hunter, P.R. 2003. "Climate change and water-borne and vector-borne disease."*Journal of Applied Microbiology* 94, 37S-46S.

Institute of Medicine of the National Academies. 2003. The Resistance Phenomenon in Microbes and Infectious Disease Vectors: Implications for Human Health and Strategies for Containment—Workshop Summary. Washington D.C.: The National Academies Press

Kermack, W.O., and McKendrick, A.G. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society* A115: 700-21.

Kitron, U. 1998. Landscape ecology and epidemiology of vector-borne diseases: tool for spatial analysis. *Journal Medical Entomology* 35:435-445.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E., and Sachs, J. 2004. A global index representing the stability of malaria transmission. *American Journal of Tropical Medicine and Hygiene.* 70: 486-498.

Kramer, LD., Li, J., and Shi, P. 2007. West Nile Virus. *Lancet Neural* 6:171-178.

Kumar, D., Prasad, G.V., Zaltzman, J., Levy, G.A., and Humar, A. 2004. Community-acquired West Nile virus infection in solid-organ transplant recipients. *Transplantation* 77: 399-402.

Kwan, J.L., Kluh, S., Madon, M., and Reisen, W. 2010. West Nile Virus Emergence and Persistence in Los Angeles, California, 2003-2008. *American Journal of Tropical Medicine and Hygiene* 83: 400-412.

La Deau, S.L., Marra, P.P., Kilpatrick, A.M., and Calder, C.A. 2008. West Nile Virus Revisited: Consequences for North American Ecology. *BioScience* 58: 937-46.

Lambert, D.M., Brown, J.P., and Florax, R.J.G.M. A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics* 40: 241-252.

Laxminarayan, R., and Brown, G.M. 2001. Economics of Antibiotic Resistance: A Theory of Optimal Use. *Journal of Environmental Economics and Management* 42:183-206.

Levitt, Tom. 2011. Overuse of Drugs in Animal Farming Linked to Growing Antibiotic Resistance in Humans. *The Ecologist* 1-3.

Levy, S.B. 1992. Antibiotics, Animals and the Resistant Gene Pool. In The Antibiotic Paradox. New York: Plenum Press.

Levy, S.B. 1998. Multidrug Resistance- A Sign of the Times. *New England Journal of Medicine* 338: 1376-1378.

Levy, S.B. 2001. Antibiotic resistance: Consequences of inaction. *Clin Infect Dis* 33: 124-129.

Linard, C., Lamarque, P., Heyman, P., Ducoffre, G., Luyasu, V., Tersago, K., Vanwambeke, S.O., and Lambin, E.F. 2007. Determinants of the geographic distribution of Puumala virus and Lyme borreliosis infections in Belgium. *International Journal of Health Geographics* 6: 1-14.

Lindsey, N.P., Staples, J.E., Lehman, J.A., and Fischer, M. 2010. Surveillance for Human West Nile Virus Disease – United States, 1999-2008. *Morbidity and Mortality Weekly Report* 59:1-17.

Lothrop B., Meyer, R.P., Reisen, W.K., and Lothrop, H.D. 1995. Occurrence of *Culex (Melanoconion) erraticus* (Diptera: Culicidae) in California. *Journal of American Mosquito Control Association* 11: 367-8.

Lyons, R.W., Samples, C.L., DeSilva, H.W., Ross, K.A., Julian, E.M., and Checko, P.J. 1980. An Epidemic of Resistant Salmonella in a Nursery: Animal-to-Human Spread. *Journal of America Medical Association* 243: 546-547.

Marshall, R.J. 1991. Mapping disease and mortality rates using Empirical Bayes estimators. *Applied Statistics* 40: 283-294.

Massad, E., Burattini, M.N., and Coutinho, F.A.B. 2008. An optimization model for antibiotic use. *Applied Mathematics and Computation* 201: 161-167.

McCabe, G.J., and J.E. Bunnell. 2004. Precipitation and the occurrence of Lyme disease in the Northeastern United States, *Vector-Borne and Zoonotic Diseases* 4, 143-148.

McGowan, J.E. 1983. Antimicrobial resistance in hospital organisms and its relation to antibiotic use. *Journal of Infectious Diseases* 5: 1033-1048.

McLennan, W., and Madden, R. 1999. The health and welfare of Australia's aboriginal and Torres Strait Islander peoples. Commonwealth of Australia: Australian Bureau of Statistics.

McMichael, A.J., Haines, A., Slooff, R., and Kovats, S. 1996. Eds. Climate Change and Human Health, WHO, Geneva.

McNamara, P.E., and G.Y. Miller. 2002. Pigs, People, and Pathogens: A Social Welfare Framework for the Analysis of Animal Antibiotic Use Policy. *American Journal of Agricultural Economics* 84: 1293-1300.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., and Tauxe, R.V. 1999. Food-Related Illnesses and Death in the United States. *Emerging Infectious Diseases* 5: 607-625.

Mellon, M., Benbrook, C., and Benbrook, K.L. 2001. Hogging It: Estimates of Antimicrobial Abuse in Livestock. (Union of Concerned Scientist), Cambridge, Massachusetts.

Messina, J.P., Brown, W., Amore, G., Kitron, U.D., and Ruiz, M.O. 2011. West Nile Virus in the Greater Chicago Area: A Geographic Examination of Human Illness and Risk from 2002 to 2006. *Urisa Journal*: 5-18.

Meyer, R.P., Martinez, V.M., Hill, B.R., and Reisen, W.K. 1988. *Aedesthelcter* from the lower Colorado River in California. *Journal of American Mosquito Control Association* 4: 366-7.

Mongoh, M.N., Khaitsa, M.L., and Dyer, N.W. 2007. Environmental and ecological determinants of West Nile virus occurrence in horses in North Dakota, 2002. *Epidemiological Infections* 135:57-66.

Moran, P.1948. The interpretation of statistical maps. *Journal of Royal Statistical Society* 10: 243-251.

Mostashari, F., Bunning, M.L., Kisutani, P.T., et al. 2001. Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. *Lancet* 358: 261-4.

National Oceanic and Atmospheric Administration. 2011. Index of Public Data. http://www.ncdc.noaa.gov/pub/data/cirs (accessed on October 10 2011).

National Research Council. 1999. The Use of Drugs in Food Animals. Washington, D.C.: National Academy Press.

Nash, D., Mostashari, F., Fine, A., et al. 2001. The outbreak of West Nile virus infection in the New York City area in 1999. *New England Journal of Medicine* 344: 807-14.

Nazroo, J. 2001. Ethnicity, class and health. London: Policy Studies Institute.

Nazroo, J. 2003. The structuring of ethnic inequalities in health: economic position, racial discrimination and racism. *American Journal of Public Health* 93: 277-84.

Nielson, C.F., Armijos, M.V., Wheeler, S., Carpenter, T.E. Boyce, W.M., Kelley, K., Brown, D., Scott, T.W., and Reisen, W.K. 2008. Risk Factors Associated with Human Infection during the 2006 West Nile Virus Outbreak in Davis, a Residential Community in Northern California. *American Journal of Tropical Medicine and Hygiene* 78: 53-62.

O'Brien, T.F. 2002. Emergence, Spread, and Environment Effect of Antimicrobial Resistance: How Use of an Antimicrobial Anywhere Can Increase Resistance to Any Antimicrobial Anywhere Else. *Clinical and Infectious Diseases* 34: S78-S84.

Office of Technology Assessment. 1995. Impact of Antibiotic Resistance Bacteria. OTA-H-629. Washington D.C.: US Government Printing Office.

Palumbi, S.R. 2001. Humans as the World's Greatest Evolutionary Force. *Science* 293: 1786-1790.

Pamuk, E., Makuc, D., Heck, K., Reuben, C., and Lochner, K. 1998. Socioeconomic status and health chartbook. Hyattsvile, Maryland: National Center for Health Statistics.

Pan American Health Organization. 2001. Equity in Health: from an ethnic perspective. Washington, D.C: Pan American Health Organization.

Patnaik, J.L., Juliusson, L., and Vogt, R.L. 2007. Environmental Predictors of Human West Nile Virus Infections, Colorado. *Emerging Infectious Diseases* 13:1788-1790.

Patuelli, R., Griffith, D.A., Tiefelsdorf, M., and Nijkamp, P. 2009. The Use of Spatial Filtering Techniques: The Spatial and Space-Time Structure of German Unemployment Data.Tinbergen Institute Discussion Paper: TI 2006-049/3.

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., and Mish, E.F. The Mathematical Theory of Optimal Processes. 1962. New York

Reisen, W.K., Takahasi, R.M., Carroll, B.D., and Quiring, R. 2008. Delinquent Mortgages, Neglected Swimming Pools, and West Nile Virus, California. *Emerging Infectious Diseases* 14: 1747-1749.

Roberts, L.S., Janovy, J., and Schmidt, G.D. 2008. Foundations of Parasitology. New York: McGraw Hill.

Ruiz, M. O., Tedesco, C., McTighe, T.J., and Austin, C. 2004. Environmental and Social Determinants of Human Risk during a West Nile Virus Outbreak in the Greater Chicago area, 2002. *International Journal of Health Geographics* 3: 1-11.

Ruiz, M.O., Walker, E.D., Foster, E.S., Haramis, L.D., and Kitron, U.D. 2007. Association of West Nile virus illness and urban landscapes in Chicago and Detroit. *International Journal of Health Geographics*6:1-11.

Sattler, M.A., Mtasiwa, D., Kiama, M., Premji Z., Tanner, M., Killeen, G.F., and Lengeler, C. 2005. Habitat characterization and spatial distribution of Anopheles sp. Mosquito larvae in Dar es Salaam (Tanzania) during an extended period. *Malar* 4: 4.

Savage, H., and B. Miller. 1995. House Mosquitoes of the USA., Culex pipiens complex. *Wing Beats . Florida Mosquito Control Associations* 6:8-9.

Schulze, A.L., and R.A. Jordan. 1996. Seasonal and long-term variations in abundance of adult Ixodesscapularis in different coastal plain habitats in New Jersey. *Journal of Medical Entomology* 33, 963-970.

Secchi, S., and B.A. Babcock. 2002. Pearls Before Swine? Potential Trade-offs Between the Human and Animal Use of Antibiotics. *American Journal of Agricultural Economics* 84: 1279-1286.

Sejvar, J.J., and A.A. Marfin. 2006. Manifestation of West Nile neuroinvasive disease. *Review of Medical Virology* 16:209-24.

Shaman, J., Day, J.F., and Stieglitz, M. 2005. Drought-induced Amplification and Epidemic transmission of West Nile Virus in Southern Florida. *Journal of Medical Entomology* 42: 134-41.

Shope, R.E. 1980. Arbovirus-related encephalitis. *Yale Journal of Biology and Medicine* 53: 93-99.

Shope, R.E. 1991. Global climate change and infectious diseases. *Environmental Health Perspectives* 96: 171-174.

Sihapi, R.S. 2008. Economics of Antibiotic Resistance. *Expert Review of Anti-Infective Therapy* 6: 523-539.

Simonsen G.S., Haaheim, H., Dahl, K.H., Cruse, H., Loveseth, A., Olsvik, O., and Sundsfjord, A. 1998. Transmission of VanA-type vancomycin-resistant enterococci and vanA resistance elements between chicken and humans at avoparcin-exposed farms. *Microbial Drug Resistance* 4: 313-318.

Smith, D.L., Harris, A.D., Johnson, J.A., Silbergeld, E.K., and Morris, J.G. 2002. Animal Antibiotic Use has An Early but Important Impact on the Emergence of Antibiotic Resistance in Human Commensal Bacteria. 2002. *Proceedings in the National Academy of Sciences* 95: 6434-6439.

Snow, J. 1855. *On the mode of communication of cholera*, 2nd edition. The Commonwealth Fund, New York.

Summers, A.O. 2002. Generally Overlooked Fundamentals of Bacterial Genetics and Ecology. *Clinical and Infectious Diseases* 34: S85-S92.

Tadesse, D.A., Zhao, S., Tong, E., Ayers, S., Singh, A., Bartholomew, M.J., and McDermott, P.F. 2012. Antimicrobial Drug Resistance in *Escherichia coli* from Humans and Food Animals, United States, 1950-2002. *Emerging and Infectious Diseases* 18:741-749.

Takeda, T., Whitehouse, C.A., Brewer, M., Gettman, A.D., and Mather, T.N. 2003. Arbovirus surveillance in Rhode Island: Assessing potential ecologic and climatic correlates. *Journal of the American Mosquito Control Association* 19:179-189.

Taylor, R.M., Work, T.H., Hurlbut, H.S., and Rizk F. 1956. A Study of ecology of West Nile virus in Egypt. *American Journal of Tropical Medicine and Hygiene* 5: 579-620.

Tiefelsdorf, M., and Griffith, D.A. 2007. Semi-parametric Filtering of Spatial Autocorrelation: The Eigenvector Aproach. *Environment and Planning A* 39: 1193-1221.

Tisdell, C. 1982. Exploitation of Techniques That Decline in Effectiveness with Use. *Public Finance* 37: 428-437.

Trawinski, P.R., and Mackay, D.S. 2010. Identification of environmental covariates of West Nile virus vector mosquito population abundance. *Vector Borne Zoonotic and Diseases* 10: 515-26.

Turell, M.J., Dohm, D.J., Sardelis, M.R., Oguinn, M.L., Andreadis, T.G., and Blow, J.A. 2005. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit West Nile virus. *Journal of Medical Entomology* 42: 57-62.

United States Census Bureau. 2011. People and households. http://census.gov/popest/data/historical/index.html (accessed on October 2 2011).

United States Department of Health and Human Services (USDHHS). 1985. Report of the Secretary on Black and Minority Health. Washington, D.C.: US Department of Health.

United States Department of Health and Human Services (US DHHS) and Center for Disease Control and Prevention (CDC). 2007. National Antimicrobial Resistance Monitoring System (NARMS): Frequently Asked Question about Antibiotic Resistance. http://www.cdc.gov/narms/faq_antiresis.htm (accessed on September 26 2010).

United States General Accounting Office (US GAO). 2004. Report to Congressional Requesters No.04-490 Antibiotic Resistance: Federal Agencies Need to Better Focus Efforts to Address Risk to Humans from Antibiotic Use in Animals.

University of New Mexico Pharmacy. 2012. University of New Mexico, Albuquerque, New Mexico, USA.

Wackerly, D.D., Mendenhall, W., and Scheaffer, R.L. 2002. Mathematical Statistics with Applications, 6th Edition. Duxbury, Pacifica Grove California.

Walter, S.D., and Birnie, S.E. 1991. Mapping mortality and morbidity patterns: An international comparison. *International Journal of Epidemiology* 20: 678-689.

Walton, J.R. 1986. Impact of Antibiotic Resistance in Animal Production on Public Health. *Journal of Animal Science* 62: 74-85.

Watson, J.T., Pertel, P.E., Jones, R.C., et al. 2004. Clinical characteristics and functional outcomes of West Nile Fever. *Annals of Internal Medicine* 141: 360-5.

Webb, G.F., D'Agata, E.M.C., Magal, P., and Ruan, S. 2005. A model of antibiotic-bacterial epidemics in hospitals. *Proceedings in the National Academy of Sciences* 102:13343-13348.

Wild, S., and Mckeigue, P. 1997. Cross sectional analysis of mortality by country of birth in England and Wales. *British Medical Journal* 314: 705-710.

Williams, D.R. 2001. Racial variations in adult health status: patterns, paradoxes and prospects. In: Smelser, N.J., Wilson, W.J., and Mitchel, F. (Eds.) America becoming: racial trends and their consequences, Washington D.C., National Academy Press.

Witte, W., and I. Klare. 1995. Glycopeptide-Resistant Enterococcus Faecium Outside Hospitals: A commentary. *Microbial Drug Resistant* 1: 259-263.

Witte. 1998. Medical Consequences of Antibiotic Use in Agriculture. *Science* 279: 949-1096.

Wimberly, M.C., Hildreth, M.B., Boyte, S.P., Lindquist, E., and Kightlinger, L. 2008. Ecological Niche of the 2003 West Nile Virus Epidemic in the Northern Great Plains of the United States. *Plosone* 3:1-7.

Winters, A.M., Eisen, R.J., Delorey, M.J., Fischer, M., Nasci, R.S., Zielinski-Gutierrez, E., Moore, C.G., Pape, W.J., and Eisen Lars. 2010. Spatial Risk Assessments Based on Vector-Borne Disease Epidemiologic Data: Importance of Scale for West Nile Virus Disease in Colorado. *American Journal of Tropical Hygiene and Medicine* 82: 945-953.

Winters, A.M., Eisen, R.J., Lozano-Fuentes, S., Moore, C.G., Pape, W.J., and Eisen, L. 2008. Predictive Spatial Models for Risk of West Nile Virus Exposure in Eastern and Western Colorado. *American Journal of Tropical Medicine and Hygiene* 79:581-590.

Work, T.H., Hurlbut, H.S., and Taylor, R.M. 1955. Indigenous wild birds of the Nile delta as potential West Nile virus circulating reservoirs. *American Journal of Tropical Medicine and Hygiene* 4: 872-88.

Zohrabian, A., Meltzer, M.I., Ratard, R., Billah, K., Molinari, N.A., Roy, K., Douglas Scott II, R., and Peterson, L.R. 2004. West Nile virus economic impact, Louisiana , 2002. *Emerging Infectious Diseases* 10: 1736-1744.

Ziv, G. 1986. Therapeutic Use of Antibiotics in Farm Animals. In: W.A. Moats (ed.) Agricultural Uses of Antibiotics, Washington D.C.: American Chemical Society.