A REGIONAL POLICY SIMULATION AND FORECAST MODEL FOR THE STATE OF OKLAHOMA: A MAXIMUM ENTROPY APPROACH

By

STEVEN R. MILLER

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1995

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2005

A REGIONAL POLICY SIMULATION AND

FORECAST MODEL FOR THE STATE

OF OKLAHOMA: A MAXIMUM

ENTROPY APPROACH

Dissertation Approved:

Dan S. Rickman

Dissertation Adviser

Lee C. Adkins

Ron L. Moomaw

Francis Epplin

A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGEMENTS

Foremost, I would like to recognize my parents, Gerald and Linda Miller, for their support and guidance and for instilling in me the desire and the discipline to pursue this goal. Who would've thought their old adage, suffering builds character is true? Further mention goes to my siblings Randall Miller and Christine Joly, whose support and praise means so much to me.

A heartfelt thank you goes out to Dan S. Rickman, who saw a potential that I did not see and set me out to reach beyond my imagination. Dan, you have an incredible way of finding the gift in knowledge just by asking the right question.

Further recognition goes out to my dissertation committee, Dan S. Rickman, Lee C. Adkins, Ron L. Moomaw, and Francis Epplin, for their time effort, and guidance in this study. Their insight, instruction, and inspiration will continue to be reflected in all my future endeavors.

A special bow of thanks goes to best friend Susanne Rassouli-Currier who kept me on track. Further gratitude is owed to Kevin Currier. The Currier's continual support and hospitality proved invaluable in my endeavor. Both are to be admired as great teachers.

For Linda and Stephen Carter; let it be known that you two touched my very soul and gave me sunshine. Your gifts I'll always treasure.

Group recognition goes out to the Accounting faculty of OSU, who kept a poor starving grad student full of beer: T. Sterling Wetzel, Gary K. Meek, Janet Kimbrell, and Kevin Murphey. I hope you weren't expecting me to pay that five-year running bar tab. I'm broke! I studied Economics for crying out loud!

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

II. SURVEY OF BAYESIAN METHODS IN FORECASTING REGIONAL ECONOMIES 4

INTRODUCTION 4
THE BAYESIAN PARADIGM 8
REGIONAL BAYESIAN VECTOR AUTOREGRESSIVE MODELS AND THE MINNESOTA PRIOR 15
VAR INTEGRATION OF INPUT-OUTPUT AND ECONOMETRIC MODELS 25
BAYESIAN SPATIAL MODELS 36
OTHER APPLICATIONS..47
SUMMARY, TRENDS, AND CONCLUSION..50

III. ENTROPY ESTIMATION OF THE OKLAHOMA POLICY SIMULATION AND FORECASTING MODEL..57

INTRODUCTION 57
REGIONAL CGE POLICY MODELING 58
PARSIMONIOUS REGIONAL FORECASTING MODELS 60
REGIONAL STRUCTURAL MODELS FOR POLICY ANALYSIS AND FORECASTING..61
RECENT SYSTEM ESTIMATION APPROACHES..63
RECENT APPLICATIONS OF ENTROPY METHODS IN ECONOMICS..........65
MAXIMUM ENTROPY AND CROSS ENTROPY ESTIMATION 68
CONCLUSION..77

IV. SPECIFYING AND ESTIMATING A REGIONAL POLICY SIMULATION AND FORECAST MODEL FOR THE STATE OF OKLAHOMA 79

INTRODUCTION 79
OUTPUT BLOCK 82
PRODUCTION BLOCK 98
INCOME BLOCK 104

POPULATION BLOCK 109
HISTORIC FIT 111
DATA ISSUES 112
ESTIMATION 117
PARAMETER COEFFICIENT ESTIMATES 124
CONCLUSION 126

V. POLICY SIMULATIONS AND FORECASTS OF THE OKLAHOMA POLICY SIMULATION AND FORECAST MODEL 129

INTRODUCTION 129
EXISTENCE OF FIXED-POINT SOLUTIONS TO PROJECTIONS 129
PROJECTION EQUATIONS AND SYSTEMS FORECASTS 132
FORECAST MODEL SPECIFICATIONS AND FORECASTS 134
IN-SAMPLE PREDICTIONS 135
OUT-OF-SAMPLE FORECASTS 137
PRODUCTION BLOCK 144
INCOME BLOCK 145
POPULATION BLOCK 147
OUT-OF-SAMPLE COMPARISONS 147
IMPACT MULTIPLIERS AND POLICY SIMULATIONS 151
IMPACT MULTIPLIERS 152
ELASTICITY OF RESPONSE 160
CONCLUSION 166

VI. CONCLUSION 168

REFERENCES 173

APPENDICES 185

APPENDIX I 185
APPENDIX II 188
APPENDIX III 202
APPENDIX IV 205
APPENDIX V 210
APPENDIX VI 212
APPENDIX VII 213

LIST OF TABLES

Table Page

4.1 Prior Parameter Support Sets, Probabilities, and Estimates 128

5.1 Oklahoma Policy Simulation and forecasting Model: Percent Errors and Mean Absolute Percent Error..136

5.2 Oklahoma Policy VAR Model: Percent Errors and Mean Absolute Percent Errors 137

5.3 Projections of Key Indicators..139

5.4 Regionalized Components of Final Demands and Output..................................140

5.5 Correlations of State to Nation 143

5.6 Production Block 146

5.7 Income Components 146

5.8 Out-of-Sample MAPES to Actuals...149

5.9 Relative Student's *t*-Statistics of Equality of Relative Growth Rates 151

5.10 Calculated Output and Employment Multipliers 159

5.11 Key Simulation Elasticities...162

LIST OF FIGURES

Figure Page

3.1 Partial Transactions Table....83

5.1 Growth Rates of Employment 141

5.2 Growth Rates of Wages 142

5.3 Growth Rates of Income Components....145

CHAPTER I

INTRODUCTION

A common problem for regional economists is that empirical models designed for policy analysis tend to perform poorly in forecasting applications in comparison to parsimonious models, while empirical forecasting models tend to be inadequate policy analysis tools. Parsimonious models such as vector autoregression (VAR) models, or simple reduced-form econometric models, tend to forecast well, but they are unable to offer insights into the impacts of policy decisions. Structurally elaborate models for policy analysis, such as regional computable general equilibrium (CGE) models, necessitate extensive parameterization that requires data beyond what is routinely available in time series form. These models therefore are almost exclusively formulated as static models that are calibrated to a benchmark-year data set with no ability to track or forecast time series. An ideal model of a regional economy would marry the policy analysis strengths of regional CGE models with the forecasting capabilities of parsimonious models.

Increasingly, regional analysts are turning to Bayesian methods of integrating economic structure to otherwise atheoretical forecasting models. Though econemetric models continue to be implemented and improved for regional forecasting and policy analysis, the paucity of regional economic data renders econometric methods of estimating regional structure arduous. Paring down the structure of the model gives rise to ad-hoc restrictions that may lead to model bias from misspecification and lead to poor model forecasts.

Since data alone are not sufficient in determining the structure of regional economies, methods of combining sample and non-sample information into model specification is increasingly sought. One ideal method is through Bayesian estimation of otherwise atheoretical model structures. The atheoretical structure emphasizes the role of past observations on current observations, while Bayesian priors impose economic structure to these atheoretical structures. Indeed the resulting specifications advantage from accurate forecasts while allowing for full structural responses to changes.

The current study proceeds by surveying this integrating process to reveal the progress and opportunities for integrating atheoretical structures and economic theory with Bayesian methods. Therefore a survey of Bayesian estimation in regional forecasting models is introduced.

Though Bayesian estimation is integral in combining sample and non-sample information in estimating economic structure, Bayesian methods are limited in the difficulty of specifying Bayesian estimators for large-scale forecasting and theoretical models. An alternative Bayesian methodology to estimation, called entropy estimation, is introduced that eases the complexity constraint of traditional Bayesian methods allowing the estimation of large-scale, complex economic structures.

A structural policy simulation and forecasting model is constructed and estimated employing a variant of the entropy methodology that allows the simultaneous fit of the model structure to economic time-series. Within this structure is the essence of the economic linkages of a static general equilibrium model that facilitates the fit of data over time. This historical fit facilitates forecasting where the full time-path responses of structural change can be captured for policy and development analysis.

In summary, the proposed model integrates the forecasting accuracy of a regional VAR/econometric model with a policy-relevant structure that is representative of that associated with a regional CGE model. It extends current regional modeling by estimating the model employing a maximum entropy (ME) approach. The ME approach can be used to estimate models that contain numerous parameters in cases where data are limited. The ME approach also allows for calibrating the model to the time-series movement of key variables in true dynamic fashion, and for imposing Bayesian-type prior information. It facilitates this by specifying the estimation problem as non-linear programming problem.

CHAPTER II

SURVEY OF BAYESIAN METHODS IN FORECASTING REGIONAL ECONOMIES

INTRODUCTION

The goal of any regional forecasting system is to systematically make the best possible judgment about future events. Good forecasts are vital to good decision-making, and the better the forecast, the better-informed decision makers are. Furthermore, the higher are the stakes, the more vital good decision-making is. Given this, it is no wonder that so much effort has been applied to the development of accurate forecasting systems.

Two formidable constraints exist in creating viable regional forecasts. The first and most severe constraint is the paucity of usable regional economic data. The second constraint is the properties of econometric estimators. The second constraint is arguably less binding if sufficient regional data are available, but the absence of such data puts limitations on the procedures used to model regional economic behavior.

The paucity of good economic data has been a persistent problem for economists and is an especially acute problem for regional economists. Many variables available at the national level are not available at the regional level, and the more disaggregate the region of study the more limited regional economic data and industry detail becomes. Therefore regional econometric models are restricted by availability of histories as well as detail of the regional data. This makes regional models less functional than their national counterparts and restricts their ability to accurately represent economic relationships dictated by economic theory.

Limitations to effective procedures of estimating econometric relationships are especially problematic for situations where regional data is scarce. Traditional ordinary least squares relies on asymptotic properties for inferences and small sample estimators espouse a high degree of variation. In small samples, ordinary least squares is subject to over-fitting where too many parameters are estimated relative to the observations available to estimate those parameters. In such cases, the degrees of freedom are eroded leading to estimators with poor sampling performance. Furthermore, since regional economic variables tend to share co-movements, traditional least squares procedures suffer from multicollinear regressors. This is a common problem associated with estimating relationships from explanatory variables that share co-movements across time.

This chapter discusses Bayesian methods of correcting for such data deficiencies in estimating economic models. Primarily, this section presents applications of Bayesian methods of estimation used in regional models where regional models comprise multiple equations of multiple endogenous variables with or without direct feedback relationships.

Regional econometric models consist of a system of stochastic equations and identities. Individual industries comprising regional econometric models can be modeled with atheoretical relationships, entailing no linkages predicted by theory, or with structural relationships, where linkages from economic theory are incorporated into equation estimation. Missing the important linkages that capture economic relationships, atheoretical models tend to perform well in short-term forecasts but poorly relative to structural models in long-term forecasts. The relative long-term forecast success of structural models over atheoretical models can be attributed to additional use of information contained in structural linkages. Valid structural linkages assure that the related variables do not meander in some random-walk manner, but rather share co-movements with other economically related variables.

Sims (1980) cautions against imposing undue structure in econometric models when noting that incorrectly imposed structure causes model misspecification. Particular misspecification errors occur when the extent of endogeneity of included variables is questionable. Rather than imposing structure on econometric models, Sims recommends treating each variable symmetrically and allowing the estimation procedure to determine the extent of endogeneity of the model. He advocates vector autoregressive models (VAR) as alternatives to structural econometric models in estimating econometric systems. Stock and Watson (2001) review the VAR methodology and investigate their use in forming dynamic impact analysis, while Cromwell and Hannan (1993) do likewise for regional applications.

VARs are very general representations of data generating processes that facilitate ease in specifying, estimating, and forecasting (Zellner 1979). They allow a full range of structural relationships limited only by the inclusion of endogenous variables. Furthermore, VAR models allow the data to empirically dictate the economic structure. Therefore VAR models are relatively easy to implement having no structural relationships to set a priori. While forecasts from VARs are inexpensive, there are several inherent problems with VAR models. Economists rarely have sufficient time-series data to construct regional VAR models. Even when sufficiently long histories do exist, structural breaks render older observations irrelevant, limiting economists to smaller historical series. Because constructing VARs, with even a small number of variables, requires a considerable number of observations, VARs tend to be overparameterized, having too many parameters to estimate from too few data points. Such overparameterized systems lead to good in-sample fit but poor out-of-sample forecast performance (Litterman 1986c). Overparameterized models cannot distinguish the systematic relationships (signal) comprising the data generating process from the random variation (noise) when fitting model parameters. Therefore VARs tend to be smaller than structural models implying that they use less information (Fair and Shiller 1990).

VAR models are characterized as a system of equations of endogenous variables in lags of all system variables. For example, an n variable, unrestricted VAR in reduced form follows,

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{nt} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) & \cdots & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \cdots & A_{2n}(L) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}(L) & A_{n2}(L) & \cdots & A_{nn}(L) \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{nt} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ \vdots \\ U_{nt} \end{bmatrix}, \tag{2.1}$$

where Y_{it} is the i^{th} endogenous variable at time t, C_i is the constant term in equation i, U_{it} is a stochastic term assumed to be characterized as white noise, and the matrix of parameter terms A_{ij} are estimated with least squares. The vector lag operator L is defined as the vector lag operators $[L^1\ L^2\ \cdots\ L^s]'$, where $L^k Y_{it} = Y_{i(t-k)}$ such that for equation i,

$$Y_{it} = C_i + \sum_{j=1}^{n} \sum_{k=1}^{s} a_{ijk} L^k Y_{i(t-k)} + U_{it}\,. \tag{2.2}$$

Equation (2.2) specifies Y_{it} is as a function of its own s lags and the s lags of the other n endogenous variables in the system. Because each equation in the traditional reduced form VAR model has the same regressors, each equation can be estimated separately using lease squares with no loss in efficiency (Judge *et al.* 1988, pp. 450). While the strength of the VAR specification is the ease of formulating the model, two empirical weaknesses are noted. Since each equation in the VAR system requires $n \times s+1$ parameter estimates, there is a rapid decline of degrees of freedom for each endogenous variable added to the system. Furthermore, since many economic variables share co-movements over time, VAR models generally suffer from a high degree of multicollinearity. The Bayesian framework has been employed to mitigate these weaknesses in VAR estimation.

This chapter reviews the current state of Bayesian methods in creating regional forecasting systems. Bayesian methods have been applied to a host of estimation problems in regional analysis. The third section surveys Regional Bayesian vector autoregressive (BVAR) models, which draw primarily on the use of the Minnesota prior specification (Doan *et al.* 1984) and forms the keystone of Bayesian forecasting. These models are mostly atheoretical in that prior distributions generally do not reflect economic theory. The forth section surveys Bayesian applications that impose economic theory on parameter estimates through informative prior distributions. The sixth section surveys recent applications of accounting for spatial location of regions.

THE BAYESIAN PARADIGM

All forecasting models combine sample and non-sample information to derive forecasting equations. The process of combining these sources of information is one of the most controversial topics in applied econometrics and econometric forecasting. The controversy surrounds two competing statistical paradigms. The frequentists or traditionalists treat equation parameters as unknown constants and rely on repeated sampling for estimation, while the subjectivists or Bayesians treat equation parameters as unknown variables and rely on the combination of prior information and the data for estimation. Succinctly, the frequentists focuse on the probabilities of various possible sample outcomes resulting from a given population while the Bayesians view an observed sample as given and consider the probabilities of various populations from which the sample might have come (Kmenta 1986, pp. 192).

In formulating model equations, the frequentists rely on classical estimation methods that entails including only a few explanatory variables suggested by theory when formulating the forecast equations. This is equivalent to claiming no prior information concerning the relationship of included variables and absolute knowledge of no relationship for excluded variables (Litterman 1986b). Rather than employing these extreme options, Bayesians introduce prior information that accounts for the expected value of the estimate and the degree of confidence in their expectation of that estimate. The subjectivists postulate that imposing informal exact prior restrictions creates more formidable bias than the formal prior restrictions imposed by Bayesian methods (Poirier 1995, pp. 482). The frequentists contend that prior information may not be conveniently expressed in a formal prior, in which case it is better to incorporate such prior information in a thoughtful, ad-hoc way (Kennedy 1998, pp. 215).

The paucity of regional times-series data makes regional forecasting ideally suited for Bayesian methods since Bayesian estimation offers an objective means of correcting for insufficient quality and quantity of data (West and Theil 1991). The stage for Bayesian forecasting is set by Friedman (1953, pp. 8-9) and supported in Zellner (1985) in stating that, "[t]he only relevant test of the validity of a hypothesis is comparison of its predictions with experience." Hence, regardless of the methods employed in forecasting, the acid test of the efficacy of the model is its ability to forecast well.

Bayesian methods have been applied extensively over the past 40 years for national economic forecasting (Zellner 1985) and, to a lesser and more recent extent, to regional economic forecasting models. Evidence of the usefulness of Bayesian methods in regional forecasting models can be found in their application to business cycle forecasting (DeJong *et al.* 2000; LeSage 1991; Otrok and Whiteman 1998), to vector autoregression and error correction forecasting models (LeSage 1990; Litterman 1980, 1986b; Liu 2002; McNees 1986), and to forecast model selection (Geweke 2001; LeSage and Rey 2002; Rickman and Miller 2002). In fact, Bayesian methods of estimation and forecasting have a seemingly unlimited set of applications.

The first application of Bayesian methods to forecasting regional data is found in West and Theil (1991), who employed a Stein-like shrinkage estimator to forecast industry employment for 20 Florida MSAs. Their purpose for using Stein effects is to help mitigate deficiencies in the quality and quantity of sub-national data. Since West and Theil's seminal paper on regional forecasting, a host of other Bayesian applications in regional forecasting models has emerged. The general rationale for these efforts centers on the need to augment deficient regional data for forecasting purposes.

The building block of the Bayesian paradigm is Bayes' theorem. Using Poirier's (1995) notation, set y equal to some vector of observations, X equal to some matrix of predetermined explanatory variables of y, and let θ be some vector of parameters that together with X describes the observation vector y. Bayes' theorem can be written as,

$$g(\theta \mid y, X) = \frac{f(y \mid \theta, X) g(\theta \mid X)}{f(y \mid X)} \propto f(y \mid \theta, X) g(\theta \mid X), \tag{2.3}$$

where $g(\theta \mid y, X)$ is known as the posterior probability distribution, $g(\theta|X)$ is the prior distribution depicting subjective beliefs of the values the parameters can take and is independent of the sample, $f(y\,|\theta\,, X)$ is the conditional probability of the observed y's given the values of θ and X generally known as the likelihood function, and $f(y|X)$ is the marginal likelihood of y and does not depend on θ. The denominator, $f(y|X)$, is known and invariant to θ and ensures that $g(\theta \mid y, X)$ integrates to unity.[1]

The posterior distribution is the product of the data and subjective prior beliefs. To see this, note that an equivalent statement of Bayes' theorem can be stated as,

$$\textit{Posterior Distribution } \boldsymbol{\mu} \textit{ Likelihood Function} \times \textit{Prior Distribution}, \tag{2.4}$$

which is an equivalent statement of the relationship in Equation (2.3). Bayes' theorem combines information from two sources to derive the posterior distribution. The relative influence of these sources of information on determining the posterior distribution depends on the relative precision sources take. The stronger the researcher's belief in their prior knowledge, the more precise the prior distribution relative to the likelihood and the more influence the prior has on the posterior estimates. Furthermore the larger the sample size, the less weight is placed on the prior. By the theory of large numbers, as the sample size increases the variance of the likelihood estimator decreases placing more weight on the likelihood function relative to the prior distribution. For large samples, the Bayesian and the classical approaches tend to converge as all precision is established on the likelihood function determined by the data (Dorfman 1997).

Prior distributions are generally classified as being either non-informative, reflecting no prior expectations, or informative, where some outside prior expectations are specified. Non-informative Bayesian priors rely on Jeffreys' (1967) vague or

[1] The presence of stochastic regressors, X, as found in forecasting applications is generally assumed away in application. For detailed descriptions of applying Bayesian methods refer to any combination of (1971), Box and Tao (1973), and Poirier (1995).

indifference priors which distributes equal likelihood to all values of the parameter across the parameter support. For Equation (2.1), this is specifying the prior distributions as,

$$g(\theta) = \begin{cases} \dfrac{1}{a}, & 0 < \theta < a, \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Equation (2.5) states that equal prior probability is given for all possible values of the parameters θ, but it does not constitute a proper probability distribution function since the integral over all possible values of θ does not equal unity. Such continuous distribution in which the integral over the support of the distribution does not sum to unity is known as an improper distribution (Birkes and Dodge 1993, pp. 146).

Though, a proper prior distribution is sufficient for a unique solution to the posterior (Press 1989, pp. 29), the posterior distribution is invariant up to a multiplicative constant, and henceforth an improper non-informative prior is sufficient for a unique solution to the posterior distribution. By noting that the maximum likelihood estimator of the parameter θ of a linear function is equivalent to the least squares estimate (Judge *et al.* 1988, pp. 223-224), and recalling that $f(y|\theta,X)$ is the likelihood function for the linear relationship in parameters θ, it can easily be seen that imposing non-informative priors into the Bayesian Equation (2.3) gives the least squares estimates of θ. To see this, note that $f(y)$ in Equation (2.3) is the marginal likelihood of y and is some function of the known observations only and therefore constant. Restating Equation (2.3) as,

$$\begin{aligned}
&g(\theta \mid y, X) = f(y)^{-1} f(y \mid \theta, X) g(\theta) \\
&g(\theta|y,X) = cf(y \mid \theta, X) g(\theta), \text{ where } \mathrm{c} = f(y)^{-1} \text{ and is a constant,} \\
&g(\theta|y,X) = c' f(y \mid \theta, X), \text{ where } c' = cg(\theta) \text{ and constant, therefore,} \\
&g(\theta|y,X) \propto f(y \mid \theta, X).
\end{aligned}$$

By specifying $g(\theta)$ as a constant non-informative prior, it is absorbed into the constant term of the posterior distribution. Since the posterior distribution is invariant to a multiplicative constant the non-informative prior does not influence the optimal values of θ, but only the value of the posterior distribution for all $\theta \in \mathbb{R}^n$. For a non-informative prior, the resulting posterior distribution is the likelihood function of the data and completely dominated by the sample information. For a more complete proof and discussion refer to Judge et al. (1988, pp. 281-288).

Informative priors use subjective information to determine the distribution the researcher expects the model parameters to take prior to estimation. This information can come from a variety of sources such as prior research findings, or expert opinion. In most cases this prior information is with some uncertainty, which is reflected in the specification of the prior variance. A more common measure of uncertainty is the prior precision, which is simply the inverse of the variance. Practitioners that have a high degree of certainty *a priori* of the values the parameters should take will place a low degree of variance or a high degree of precision on the prior distribution. As described below, this causes the estimator to place greater weight on the prior distribution relative to the data.

For convenience, informative priors are usually specified as conjugate prior distributions, which simplify the math required in calculating the posterior distribution. A conjugate prior distribution is a distribution that when multiplied by the specified likelihood function, creates a distribution of the form of the prior distribution, while a natural conjugate prior distribution produces a distribution in the form of the likelihood function (Poirier 1995, pp. 291). In either case, the choice of prior distribution is often made on the basis of computational convenience rather than on the belief of the representation of the prior on the actual distribution. Convenience priors have the benefit of producing analytically tractable solutions to posterior distributions that take a recognizable form of a known probability distribution, thereby saving the analyst the need to analytically integrate over these complex functions.

Increases in computational power of modern computers have decreased the practitioner's reliance on conjugate priors. Recent, computationally intensive methods of integration rely on computational power and limit the need to restrict prior specification to those of convenience priors. Rather than deriving statistics from the posterior distribution using calculus, Monte Carlo methods of numerical integration employ pseudo-random number generation to investigate the empirical properties of posterior probability distributions. Though numerical methods of deriving results from Bayesian inference leads to approximation error, the precision of the estimates is determined by the number of samples drawn from the posterior distribution, and so is within the analyst's control. The advent of numerical methods of estimating posterior distributions gives researchers greater flexibility in specifying prior distributions avoiding complexities of intractable posterior distributions.

Because Bayesian applications of regional forecasting are well established and because new applications continue to be developed, the time is ripe to review these applications and assess their merits over traditional methods. Furthermore, the current body of literature is broad enough and mature enough to reflect trends in applications of Bayesian forecasting techniques at the regional level. Therefore the current study surveys the current literature on Bayesian applications of regional forecasting by describing issues relating to Bayesian applications, describing the methods of and the rationale for their applications, and reporting the results of those applications. This study further reflects on the trend of these applications as sequenced over time and assesses the direction of growth as well as gaps in the current literature.

The following section presents the regional Bayesian vector autoregressive model employing an atheoretical prior distribution devised by Litterman that creates a marriage of univariate time-series forecasting with structural forecasting methods. The resulting estimator is the keystone to many applications of regional models to be presented in subsequent sections.

REGIONAL BAYESIAN VECTOR AUTOREGRESSIVE MODELS AND THE MINNESOTA PRIOR

Noting weaknesses in the vector autoregressive (VAR) models of Sims, Litterman (1986a) presents the Bayesian variant, BVAR models, employing what is now known as the Minnesota prior. Litterman's BVAR methodology incorporates the forecaster's prior beliefs of the values the coefficients a_{ijk} in Equation (2.2) should take. Litterman's priors are atheoretical in that the priors do not specify any theoretical underpinnings of the data, but merely act to control for weaknesses in the VAR methodology. His prior specification takes the form of normally distributed prior densities, which can be completely defined by means and variances.

Understand the Minnesota prior first requires introducing the method of imposing this prior on estimation. Litterman's BVAR system is estimated by an application of Theil's mixed estimation. In the traditional reduced form VAR, each equation regresses a dependent variable on lags of itself and other variables in the system. Therefore, each equation has identical regressors, and because there is no loss in efficiency in estimating each equation individually using lease squares , each equation is estimated independently and takes the form,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\text{, where } \mathbf{u} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right), \tag{2.7}$$

where **y** is a $T\times 1$ vector of observations on the dependent variable, **X** is a $T\times m$ matrix of predetermined explanatory variables with rank m, and **u** is a $T\times 1$ vector of disturbances, where, $E[\mathbf{u}] = 0$, $E[\mathbf{u}\,\mathbf{u}'] = \boldsymbol{\sigma}^2\mathbf{I}$. The matrix of estimated coefficients, **β,** is an $m\times 1$ vector and assumed to be normally distributed, where m ($= n\cdot k + 1$) is the product of the number of variables in the VAR and the number of lags plus the intercept. Extraneous information not derived from the data is represented as linear stochastic restrictions of the form,

$$\mathbf{r} = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\upsilon}\text{, where } \boldsymbol{\upsilon} \sim N\left(\mathbf{0}, \sigma^2\boldsymbol{\psi}\right) \tag{2.8}$$

where **r** is an $m\times 1$ vector of prior means, **R** is an m dimensional identity matrix, $\boldsymbol{\upsilon}$ is an m dimensional vector of stochastic restriction error terms, where $E[\boldsymbol{\upsilon}] = 0$, $E[\boldsymbol{\upsilon}\,\boldsymbol{\upsilon}'] = \boldsymbol{\psi}$, and $\boldsymbol{\psi}$ is a positive-definite, non-singular, symmetric matrix of the prior expected variance-covariance matrix of the prior means.

Theil's mixed estimation is to apply Aitken's (1935) generalized least squares to the system of stacked Equations (2.7) and (2.8) forming,

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{r}\end{bmatrix} = \begin{bmatrix}\mathbf{X}\\ \mathbf{R}\end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix}\text{, where } \begin{bmatrix}\mathbf{e}\\ \mathbf{v}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{r}-\mathbf{R}\boldsymbol{\beta}\end{bmatrix}, \sigma^2\begin{bmatrix}\mathbf{I_T} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\psi}\end{bmatrix}\right). \tag{2.9}$$

Theil's generalized least squares estimator of the parameter vector **β** in Equation (2.9) is,

$$\boldsymbol{\beta}_{\mathbf{Theil}} = \left(\mathbf{X'X} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{R}\right)^{-1}\left(\mathbf{X'y} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{r}\right), \tag{2.10}$$

and can be viewed as a weighted average of the means and prior distribution for the vector of coefficients **β** (Birkes and Dodge 1993, pp. 167). Expanding Equation (2.10) and multiplying the first term by $(\mathbf{X'X})^{-1}(\mathbf{X'X})$ gives,

$$\boldsymbol{\beta}_{\mathbf{Theil}} = \left(\mathbf{X'X} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{R}\right)^{-1}\left(\mathbf{X'X}\right)\boldsymbol{\beta}_{\mathbf{ols}} + \left(\mathbf{X'X} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{R}\right)^{-1}\mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{r}\,. \tag{2.11}$$

This arrangement of Equation (2.10) shows that $\boldsymbol{\beta}_{\mathbf{Theil}}$ is a weighted average of the least squares estimator $\boldsymbol{\beta}_{\mathbf{ols}}$ and the prior predicted means $\mathbf{r}$ with weighting matrices, $\left(\mathbf{X'X} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{R}\right)^{-1}\mathbf{X'X}$ and $\left(\mathbf{X'X} + \mathbf{R'}\boldsymbol{\psi}^{-1}\mathbf{R}\right)^{-1}\mathbf{R'}\boldsymbol{\psi}^{-1}$. The weighting matrices are both positive definite and sum to the identity matrix. The prior covariance matrix $\boldsymbol{\psi}$ is specified by prior beliefs and drives the weights. In the extreme event of no confidence in the prior restriction, the prior covariance matrix approaches infinity such that Equation (2.10) verges on to $\boldsymbol{\beta}_{\mathbf{Theil}} = \mathbf{I}\cdot\boldsymbol{\beta}_{\mathbf{ols}} + \mathbf{0} = \boldsymbol{\beta}_{\mathbf{ols}}$, resulting in OLS estimates. In the opposite extreme of perfect confidence in the prior means, the prior covariance matrix approaches the zero matrix such that Equation (2.10) converges to $\boldsymbol{\beta}_{\mathbf{Theil}} = \mathbf{0} + \mathbf{I}\cdot\mathbf{r} = \mathbf{r}$. That is Theil's parameter estimates revert back to the imposed restriction with no variance. In practice prior precision is set somewhere in between the two extremes offering a flexible channel in which to impress non-sample information on parameter estimates.

Litterman integrates a random-walk Bayesian prior to the self specifying structure found in VAR estimates through Theil's GLS. Litterman notes that many economic time-series follow a random-walk implied by the efficient markets hypothesis,

$$Y_{it} = Y_{i(t\text{-}1)} + U_{it}. \tag{2.12}$$

The random-walk model implies that aside from past observations, there is no method of determining what future values an economic time-series will take. For each equation in Equation (2.1), the Minnesota prior specifies a prior distribution based on the belief that the series espouses a random-walk, where the specified prior mean for the single lag coefficient for equation i, $a_{ij,1}$ is set equal to unity, and all remaining prior means on own-lags and cross-lags are set equal to zero. Therefore the Minnesota prior specifies a one-to-one relationship between future values of a series with it's immediate past value and assigns a prior value of zero for all subsequent lags and cross-equation relationships. Drift can be included in the Minnesota prior by defining a diffuse prior distribution for the constant a constant term C_i,

$$Y_{it} = C_i + Y_{i(t-1)} + U_{it} \tag{2.13}$$

This specification says that the best guess of a variable's current value is the value the variable took last period. Forecasts from this overly simple specification often fit as well as complex structural models (Crone and McLaughlin 1999).

In Litterman's specification, the elements of the prior variance-covariance matrix $\boldsymbol{\psi}$ are specified as,

$$\lambda_{ijl}^2 = \left(\theta_i f_i(j,l) g_i(l) \left(S_i / S_j\right)\right)^2. \tag{2.14}$$

The hyperparameter θ determines the overall tightness of the prior variances and reflects confidence in the prior means. For cases with certainty about the set of prior means, this parameter should be near zero. The set of parameters $f(i,j)$ specifies the relative tightness on variable i to variable j. The decay rate parameter $g(s)$ specifies the rate of decay of uncertainty over lags. The decay rate incorporates greater confidence that the prior means equal the specified prior mean of zero as the lag length increases.

The parameter values in Equation (2.13) reflect the prior weight placed on the prior random-walk means in estimation. Large parameter values imply greater precision and impose the random-walk priors with more certainty. The coefficients relating Y_{it} to lags of itself, a_{iik}, are scale invariant such that multiplying both the left- and the right-hand side of equation i by some constant, leaves the own-lag coefficients unchanged.

Therefore the scaling term $S_i/S_j=1$ for $i=j$. For cross-variable specification, the ratio S_i/S_j scales the variances to correct for different units of measures. The Minnesota prior systematically places more confidence that the actual means of the own- and cross-lags are close to zero the further the lag through the decay rate parameter $g(s)$. The decay parameter is usually specified as a harmonic decay as, $g(s) = s^{-1}$, where s is the lag length. Doing so asserts that the less important a lag is believed to be, the more confident the forecaster should be that the lag's true coefficient value is zero and the tighter the precision values should be. Finally the relative tightness parameter, $f(i,j)$ is generally specified as unity for own lags where $i=j$ and some value less than unity otherwise (Litterman 1986b). This specification augments the VAR model by emphasizing the autoregressive properties of the prior means

The usual procedure in specifying the hyperparameter is to begin with a small value that imposes the random-walk on estimation firmly. Then sequentially increase it such that the weight is diverted to the least squares estimator. The performance of the new specification is compared to the prior by out-of-sample model performance. The process is continued until the forecaster is satisfied that the specification produces the model with the best out-of-sample fit (Crone and McLaughlin 1999).

In regional BVAR models where national variables are introduced to the VAR, it is common to specify one-way relationships from national variables to state variables. This specification is achieved through the $f(i,j)$ matrix of Equation (2.14), where a large value, specifying a loose prior of zero, is chosen for national variables in regional equations, and a small value, specifying a tight prior of zero, is chosen for regional variables in national equations. This is the Litterman "circle-star" structure where star (national) variables are specified to influence both star and circle (regional) variables, and circle variables influence only circle variables (Doan *et al.* 1984).

Several benefits are derived from the Bayesian estimation of VAR forecasting models. Namely, incorporating stochastic restrictions in estimation recovers a degree of freedom for every restriction, partially mitigating the sample size problems in regional models. This property of BVARs allows the estimation of systems that, because of insufficient histories, cannot be estimated in unrestricted VARs. Less apparent is the corrective properties through shrinkage-like estimation; particularly of the Minnesota priors. The benefits of shrinkage and ridge regressions when forecasting models with multicollinear regressors is well documented (Birkes and Dodge 1993, pp. 173-187; Vinod 1978). The shrinkage-like properties of the Minnesota priors come about from specifying all same- and all cross-lag prior means as zero, (except the same variable, one lag coefficient) thereby shrinking the estimated coefficients toward zero in a similar manner of ridge estimators.

Furthermore, the Minnesota prior specified BVAR encompasses the univariate AR model and the structural UVAR model as special cases, adjusting on the precision of the priors. This specification enables the desirable properties of both specifications while controlling for multicollinear relationships among explanatory variables. The strengths of these benefits are evident in increased forecast accuracy found in the following studies.

Several studies do not benchmark their results with alternative model specifications. Amirizadeh and Todd (1984) and Todd (1984) forecasts non-farm employment, earned income, and retail sales applying the atheoretical Minnesota prior in BVAR models of the five states comprising the Ninth Federal Reserve District. National drivers drive their models through the circle-star specified priors. Crone and McLaughlin (1999) compare BVAR forecasts of a system composed of the City of Philadelphia, the Philadelphia MSA, and the nation. The Minnesota prior with "circle-star" structure is used to guide unidirectional causality from the nation to the MSA to the city.

Other studies compare BVAR model performance against alternative specifications. Puri and Soydemir (2000) employ the Minnesota BVAR model to forecast employment for three key industries and aggregate employment for five Southern California counties. Rather than assigning prior variances informally, the author's apply Theil's *U* statistic (1966) in systematically selecting the optimal in-sample model fit with national drivers. They further find that loose priors provide estimates that outperform more restrictive priors suggesting that the strength of the Minnesota prior is its shrinkage like properties. Two papers make comparisons with exogenous drivers in the circle-star structure. Kinal and Ratner (1986) compare forecast accuracy from a Minnesota prior specified BVAR, its unrestricted VAR model, univariate ARIMA, and multivariate ARIMA models employing a transfer function . Fullerton (2001) uses the BVAR specification as a benchmark comparison for forecasts of a borderplex econometric model. Doing so raises the bar over other atheoretical benchmark models that his model must surpass.

Shoesmith (1990) compares the out-of-sample forecast performance of quarterly BVAR models and VAR models for North Caroline, New York, and Texas when the models are misspecified. Drawing from prior research that shows that forecasts from unrestricted VAR models are sensitive to model specification while BVAR models are much less sensitive (Shoesmith 1988), Shoesmith sets out to show that BVAR models mitigate the effect of model misspecifications on forecast performance. The test consists of drawing a random series from a standard normal distribution and either substituting it or adding it to the national variables in two models thereby creating three misspecified models. National variables are chosen to represent a wide range of national economic activity. The author then compares both short- and long-run forecasts and finds that the Minnesota prior with Litterman's "circle-star" structure outperforms VAR models in ninety percent of the trials. Because the BVAR model mitigates the effects of multicollinearity, the BVAR models outperform the VAR models for virtually every long-run forecast and for the vast majority of short-run forecasts. The results also show that forecast performance varies significantly across model specification for VAR models while BVAR models do not appear sensitive to the inclusion or exclusion of irrelevant national variables. This is because the Minnesota specified BVAR model shrinks estimates toward a univariate AR model thereby de-emphasizing the role of these irrelevant variables in forecasts.

LeSage (1990) and Shoesmith (1992) expand on the VAR methodology by incorporating an error correction term to the estimation equations. The vector error components model (VEC) is a straightforward generalization of the VAR framework that takes advantage of long-run relationships across variables characterized as linear combinations with reduced orders of integration. VEC models are formed from the same VAR system specification in Equation (2.1) in differences and augmented by an error correction term. The error correction term measures the distance between variables from their equilibrium states. Engle and Granger (1987) show that omitting the error-correction term from the estimation of the VAR in differences leads to model misspecification if a cointegrating relationship across any subsystem of variables exists. Since a VEC model with no cointegrating relationship reduces to a VAR model in differences the VAR model is a special case of the more general VEC model.

Because VAR models are nested in VEC models, comparing forecast performance entails testing the contribution of specifying cointegrating relationships in forecasting models. On theoretical grounds, VEC models are expected to outperform VAR models in long-run forecast horizons since VEC models incorporate long-run relationships in model estimation. If short-run dynamics lead to long-run relationships, VEC models are also expected to outperform unrestricted VAR models in short-run forecasts (Engle and Granger 1987; Engle and Yoo 1987; Granger 1986).

Following the neoclassical labor demand model, LeSage (1990) constructs a VAR system in man-hours of employment, nominal wages and prices for 50 industries in Ohio. VAR estimates of industries that espouse a cointegrating relationship across variables will be biased without augmenting the models with an error-correction term, (Engle and Granger 1987) so all VARs are replicated in the VEC format by the additional error-correction term. Forecasts from unrestricted estimates and Bayesian restricted estimates, with the Minnesota prior, of the 50 VAR and similarly specified VEC models are compared. While Shoesmith (1992) compares unrestricted VAR and VEC models in personal income, retail sales, and a host of national variables as drivers, designed to capture a broad range of national economic activity. Forecasts from these unrestricted models are compared to forecasts made by similarly specified Bayesian variants with the Minnesota prior and circle-star specification.

Both studies find that forecasts from the Bayesian specification of the VEC model (BVEC) outperform those by the BVAR models for long-run forecasts while the Minnesota prior specified BVAR generally produces better short-term forecasts. This is expected since cointegrated series exhibit a long-run tendency to revert to some equilibrium relationship. The empirics show that imposing this long-run relationship on short-term dynamics restricts the short-term forecasts even when a long-run cointegrating relationship exists. This finding counters Engle and Yoo (1987), and Granger's (1986) expectation of improved short-run prediction with cointegration. Relative long-run forecast performance for variables where cointegrating relationships exists, show an advantage to the BVEC formulation. They further find that the BVEC specification often produces superior long-run forecasts over the Minnesota prior VAR specification even for those industries where a cointegrating vector does not exist. This result is not especially troubling since an over-specified model does not induce forecast bias but the added multicollinear relations can cause forecast deterioration only if the collinear relationship does not hold over time.

The current state of research indicates that the application of Bayesian methods to regional VAR modeling contributes to out-of-sample forecast performance. Greater forecast performance over traditional VAR models can be attributed to the shrinkage-like property of the Bayes estimator in reducing the effects of multicollinearity in estimation, and the stochastic imposition of the AR specification, which has been shown to outperform more complex structural forecasting models (Fair and Shiller 1990). Furthermore, the Minnesota prior BVAR embodies, as a special case, the unrestricted VAR and the univariate AR models, depending on the tightness of the priors. The

Minnesota BVAR methodology offers a range of flexibility in specifying the degree of cross-variable dependence. An alternative to applying the atheoretical Minnesota prior is to structure prior distributions based on economic theory. The structure of input-output models is one such source of theoretical structure that has been applied to regional econometric models.

VAR INTEGRATION OF INPUT-OUTPUT AND ECONOMETRIC MODELS

Early applications of Bayesian methods in regional modeling focused on the need to assist statistical estimators in capturing patterns across time as exemplified by the atheoretical Minnesota prior. The attributes of the simple random-walk prior specification of the Minnesota prior BVAR prompted researchers to seek more theoretical priors to impart interindustry linkages to the otherwise atheoretical VAR framework. This step is quite logical since the unrestricted VAR is instrumental in objectively identifying linkages, and the Minnesota prior BVAR mitigates the effects of multicollinear relationships across explanatory variables. If it is true that the strength of the Minnesota BVAR prior is the shrinkage-like property of the estimator, then a systematic way of introducing industry linkages back into the estimation will improve forecasts from VARs while retaining the interindustry interaction.

While parsimonious atheoretical models may be affective in short-run forecasting models, long-run forecasting and policy analysis models require greater detail to capture the economic structure underlying the data. A way of employing Bayesian methods to otherwise atheoretical models was desired for imposing economic structure and therefore economic theory. Input-output models, with their linear general equilibrium relationships, present an opportune way of combining theory in an objective way.

Input-output (IO) models have long held a presence in regional economic analysis. Several reasons contribute to the continued application of IO models in regional analysis. First is the appeal of a completely interlinking set of relationships across industries in a region that forms the basis of IO models. Furthermore, regional IO models present intuitive measures that are easy to comprehend and put to use. More recently, IO models have won favor by many regional analysts because of the low costs of implementing regional IO models produced from the competitive industry of ready made IO software (Brucker *et al.* 1990; Hastings and Brucker 1993).

Nevertheless, IO models alone are imperiled with several weaknesses stemming from their inert assumptions and structure. For instance, IO models are developed around the assumptions of linear production technologies, constant returns to scale, homogeneous consumption functions, and price inflexibility (Rey 2000). Furthermore, IO models are strictly static and offer no time-path responses to changes in final demands. Regional econometric models are not restricted to such assumptions, thereby offering greater flexibility in specifying the underlying theory of the model and modeling dynamic responses to changes. Though regional econometric models offer greater opportunities for policy analysis, their use as a regional modeling tool remains somewhat limited (Rey 1997).

While the similarities between IO models and large-scale macroeconometric models have long been recognized by modelers, empirical work on integrating the two has not. Gerking (1976) explored estimating IO linkages econometrically, while Klein (1989) discussed incorporating IO models into large-scale macroeconometric models and LeSage and Magura (1991) apply information contained in the national IO table in specifying a national VAR employment system for forecasting. More recently, IO models have facilitated the specification and estimation of regional econometric models in Glennon and Lane (1990), Treyz, Rickman, and Shao (1992), Fawson and Criddle (1994), and Magura (1987).

The intuition for impressing the IO table into econometrically estimated regional models becomes clear when considering the structure of closed IO models. A closed IO model is a linear representation of interindustry transactions and industry final demands. A representation equation may take the form as,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} + \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} \tag{2.15}$$

or

$$\mathbf{Y} = \mathbf{AY} + \mathbf{F}.$$

Aside from the exogenous final demand driver **F**, Equation (2.15) captures the essence of the VAR model of Equation (2.1), differing only in contemporaneous rather than lag relationships and in the stochastic error term. The IO transactions table does not have a time subscript because the IO transactions table is specified with no time element. The direct input coefficient a_{ij} represents the dependence of output in industry j on the inputs from industry i similar to the dynamic $A_{ij}(L)$ coefficient in the VAR equations.

Placing exclusion restrictions on the appropriate cross industry relationships coefficients of Equation (2.1) found in the representative coefficients a_{ij} of Equation (2.15) creates a structural set of equations that can be modeled with econometric models, assuming the data for such a model exists. Where data is lacking, mixed estimation recovers the ability to econometrically model the set of relationships found in Equation (2.15) through imposing IO relationships in the form of prior expectations.

A striking benefit of specifying prior distributions from the direct requirements matrix is the ability to impose economic structrure on the VAR framework. IO models offer point estimate impacts with no time-path response, while econometric models offer testable time-path responses to changes in the system but generally lacks the ability to model complex structure at the regional level. Integrating interindustry relationships into the atheoretical regional VAR model provides structural relationships to the forecasting model. The resulting models produce impact multipliers that are both testable and have a time-path response (Rey 2000).

Though several non-Bayesian applications of integrating regional IO and econometric models have materialized[2], two non-Bayesian approaches have been adopted by regional analysts within the Bayesian VAR specification. The first follows Magura (1987) in utilizing a representative IO table to guide in anticipated $A_{ij}(L)$ coefficient values. Since the Bayesian methodology allows full enclosure of the IO transactions table, it is an intuitive way of incorporating the direct requirements matrix into the VAR forecasting model. In Magura's strategy, industries that are primary users of the output of the dependent variable industry are included as explanatory variables to capture relevant intermediate demand linkages from the IO transactions table. The resultant VAR system of industry forecast equations are then structurally determined by the IO table. Critical of this method, is that this ad-hoc method of exclusion restrictions may lead to misspecification errors by excluding otherwise key explanatory variables from

[2] See West and Jackson (1998), West (1995) for surveys of integrating IO models with econometric models for policy simulation and Rey (1997) and Rey (1998) for surveys for forecasting models.

estimation.

The second approach follows Moghadam and Ballard (1987) in collapsing the IO table into a single variable that captures the interindustry relationships. This approached, known as the I-SAMIS method, integrates IO models with econometric models by summarizing interindustry linkages by a set of interindustry demand variables (*IDV*) that estimate the demands for output of one industry from other industries within the region. The *IDV* variable for industry *j* in time *t* is computed as,

$$IDV_{it} = \sum_{j=1}^{n} r_{ij} X_{it} \text{ , for } i \neq j, \tag{2.16}$$

where the vector X_{it} is the total regional output for industry *i* in time *t*, and r_{it} is the direct requirements matrix from the national or regional IO transactions table. This variable of aggregate interindustry relationships is then included as an explanatory variable in the time series equations of industry *i* to capture interindustry linkages[3]. This approach has the benefit of capturing interindustry linkages without imposing unattainable data requirements and introducing multicollinearity, but suffers in that cross-industry parameter restrictions are implicitly assumed in the aggregation (LeSage and Rey 2002; Rickman and Miller 2002) and explicitly fixes intermediate demand to industry output (West 1995).

The first integration strategy, within the BVAR context, is found in Magura (1990). Magura starts with the Minnesota prior specification of zero means on all but the own-first lag coefficients with advancing precision on lag length. The prior precision parameter *f*(*i*,*j*) of Equation (2.14) is altered based on the national IO table; decreasing

[3] For a comprehensive study of non-Bayesian methods of integrating regional IO and econometric models, see Rey (2000). For a generalization of the *I-SAMIS* methodology, see Glennon and Lane (1990).

precision where the IO table reflects interindustry linkages and increasing precision where no interindustry linkage appears. By decreasing the precision more weight is applied to the data in estimating the slope coefficients while increasing precision places more weight on the zero slope prior of the random-walk prior.

Comparing out-of-sample forecast performance from this prior against various atheoretical univariate and multivariate AR, or VAR models reveals that regional employment in basic industries are best modeled with an atheoretical univariate AR specifications, while forecasts of non-basic industries favors the BVAR model with IO priors. This finding concludes that relaxed prior precision on the Minnesota prior random-walk restriction facilitates employment forecasts of non-basic industries while stringent prior precisions on the random-walk restriction facilitates basic industry employment forecasts. This is not surprising given that basic industries are subject to changes outside of the region while the health of non-basic industries depends on local economic conditions.

While Bayesian estimates from this specification transcends the ad-hoc misspecification issue associated with exclusion restrictions based on the transactions table of Magura (1987), Magura's BVAR specification suffers some short-commings. Magura's reliance on intermediate demand linkages omits the induced effects of final demand linkages found in national IO transactions tables. In addition, employing the national IO table, though suitable as a first approximation, omits regional specific information that can otherwise be imparted by a regionalized IO table. Furthermore, Magura's specification does not link his model to the national and world economy typically found in regional econometric models.

Partridge and Rickman (1998) employ ready made regionalized IO tables reported by IMPLAN rather than relying on the national IO table as a first best approximation to the interindustry linkages. They further capture the full set of endogeneities by solving the IO relationships for reduced-form employment relationships that entails both intermediate and regional final demand components. Long-run responses relating changes in one industry to changes in another are calculated as elasticities with simple log-derivatives of the reduced form equations. Similar elasticity responses are calculated for export final demand components. These elasticity responses are then incorporated into the random-walk means specification as the precision variable, $f(i,j)$ in Equation (2.14). They further use IO information in forming prior means in a separate analysis for comparisons. These prior means are set with diffuse precisions to account for uncertainty of timing and lack of prior studies to appropriate values. Finally, they link their regional models to the nation through the exogenous domestic and world export final demand components for forecasting purposes.

The authors note that forecast performance depended on three factors, the length of the forecast, whether the industry was primary or tertiary, and if there existed an economic turning point in the forecast horizon. Supporting the findings of Magura (1990) and LeSage and Magura (1991), forecast accuracy of models restricting interindustry linkages such as AR and the Minnesota prior BVAR models deteriorated with the forecast horizon showing that interindustry linkages are important in long-run forecasting. Similarly, tertiary employment forecasts benefit from interindustry linkages relative to basic industries. Further analysis shows that interindustry linkages facilitate forecasting turning points. Where AR specifications rely solely on past trends, capturing correlations across industries facilitates cross industry relationships essential in capturing downturns across industries.

In two articles Rickman (2001) and Rickman (2002) compare forecast performance from this specification over differing closure assumptions. Depending on the extent of final demand inclusion in calculating prior means restrictions, Rickman creates prior means entailing Type I, Type II, and the extended multipliers described in Batey and Rose (1990). The first article focuses on out of sample predictive power, while the second discusses the estimated multiplier responses by imposing prior means in two polar extremes. Imposing tight priors from the IO relationship imposes near exacting restrictions to the IO table multipliers provided by IMPLAN, while the opposite of diffuse prior precision gives ordinary least squares estimates. Distributing the weights evenly between the prior distribution and the likelihood function gives mixed estimation results allowing a combination of prior means and the data to determine interindustry relationships. Forecasts based on these three estimators are carried out to assess which will produce the best impact multipliers. Results of these exercises are diverted until later.

Using similar analysis, LeSage and Rey (2002) and Rickman and Miller (2002) examine the relative performance of Bayesian and non-Bayesian approaches to integrating IO and econometric models in forecasting. LeSage and Rey (2002) forecast employment for 20 industries in 88 counties in Ohio and compares forecast performance of four groups of models. The first group collapses intermediate demand linkages into a single *IDV* variable (Moghadam and Ballard 1987), the second econometrically identifies industries to include in the calculation of the *IDV* variable (Glennon and Lane 1990;

Glennon *et al.* 1987; Glennon *et al.* 1986), the third specifies a BVAR with IO linkages as priors, and the forth econometrically specifies interindustry linkages for inclusion in a VAR.

Likewise, Rickman and Miller (2002) present three sets of overlapping strategies in forecasting employment in 30 state level industries. The first strategy entails VAR models from the I-SAMIS model of collapsing demand components into an *IDV* variable. The second set compares forecast accuracy over different degrees of endogeneity of the model specification and selection criteria. The final set compares forecasts from econometrically specified models to those in which variable inclusion is determined by the IO transactions table.

Both papers employ the Bayesian model selection in econometrically selecting variables for inclusion. The selection criteria follows from the Bayesian model averaging algorithm of Raftery*, et al.* (1997). The BMA procedure follows a Markov chain Monte Carlo methodology that generates a process that moves through model space in search of an optimal combination of explanatory variables based on the fit of the model. Conditional probabilities, conditioned on calculated posterior distributions, are assigned to model specifications by comparing model posterior distributions to the subset-best distribution, thereby placing more chance of drawing the most probable model specification. The procedure is continued, where draws of a variable into the model are based on how often that variable shows up in past model draws. The procedure is similar to the stepwise regression procedure in that it accounts for model specification uncertainty, but it benefits in applying posterior probability densities rather than the flawed coefficient of determination measure in comparing model specifications.

The cumulative results of these four articles show that IO linkages increase relative forecast performance for long-term forecasting and for forecasts of tertiary employment. Partridge and Rickman (1998) and Rickman (2001) further show that the univariate autoregressive model produces relatively more accurate forecasts for employment in goods-producing sectors. Furthermore, Partridge and Rickman and Rickman and Miller (2002) show that inclusion of the interindustry linkages facilitates forecasts of turning points, and that imposing the technical coefficients on the prior means rather than the prior variances does not generally improve forecast performance, but it does improve turning point forecasts.

Comparing results across model closure assumptions shows that the more parsimonious Type I multiplier closure assumption tends to produce the best out-of-sample forecasts. The more endogeneity imposed on the estimated equation, the greater the forecast errors tended to be, even for long-term forecasts. This result is robust for both the I-SAMIS specification and the individual industry VAR specifications. Counter intuitive, parsimonious models that entail some interindustry linkages produce better long- and short-term forecast relative to their more detailed counterparts.

Since the results indicate that the Type I multiplier concept should be applied to regional employment forecasts, and that limiting the set of interindustry intermediate demands increases forecast performance, it appears instructive to further test the extent to which interindustry variables should enter the equations. LeSage and Rey (2002) find that the Bayesian model selection procedure to econometrically screen cross-industry inclusion of explanatory variables outperforms using the IO transactions tables as prior weights in the Minnesota prior framework and all methods of collapsing interindustry linkages into a single *IDV* variable. Rickman and Miller show that using intermediate demand linkages alone to select five explanatory industries in each equation outperforms forecasts methods employing the *IDV* variable. Conversely to LeSage and Rey, they further show that selecting the interindustry linkages from the IO transactions table outperforms econometrically screening cross-industry relationships for inclusion. Since the regions of study in both papers do not overlap, differences in result may be attributed to regional specific characteristics.

This series of articles shows that imposing restriction through the Bayesian framework consistently improves out of sample forecast performance. Furthermore improvements in forecast accuracy are found in incorporating the industry-by-industry transactions table into the random-walk prior specification. Partridge and Rickman (1998) and Rickman (2001) find that increases in forecast accuracy from these IO specified priors are the result of both the shrinkage-like properties of the random-walk assumption and the use of IO information by comparing weighted and un-weighted priors.

Though this series of articles shows that Bayesian methods of estimation produce forecasts with relatively smaller out-of-sample errors, Rickman (2001) warns that the Minnesota-type prior can lead to biased aggregate forecasts. Forecast errors of unbiased industry forecasts will tend to cancel out over aggregation. That is positive errors are offset by negative errors in aggregation. Biased industry forecasts may not have this attribute if industry forecasts are biased negatively or positively over the complete set of disaggregate industries. Therefore biased forecasters that produce more precise industry forecasts, may lead to worse aggregate forecast than their less precise but unbiased counterparts.

Rickman and Miller further cautions on the use of intermediate demand linkages for policy analysis. Though the authors show that restricting explanatory variables to five intermediate demand relationships produce the best out-of-sample forecasts, the multipliers implied by such parsimonious representations can give inadmissible multipliers. The multipliers implied by the *IDV* framework are much more consistent with expectations.

BAYESIAN SPATIAL MODELS

Because regional economies are interdependent, spatial relationships across contiguous regions are instrumental in modeling any one region. The econometric rationale for this is that spatially related industries exhibit co-movement over time due to common influences. It seems intuitive that capturing this co-movement across regions can facilitate more accurate regional forecasts. Recognition of the interdependence of regions sprung from Isard's *Channels of Synthesis* (1960), while capturing the interregional relationship with spatial econometric methods has developed largely around the work of Anselin (1980; 1988b).

Central to spatial econometrics is the construction of the contiguity matrix. The contiguity matrix for first order spatial autocorrelation[4] is constructed on the basis of binary indicators between spatial units such that the structure of neighbors is expressed by 0-1 values (Anselin 1988a, Pg. 17). The dimension of the square contiguity matrix is the number of possible interrelated regions, where each region is represented by

[4] The term first order is in reference to the time-dependent first order autoregression, where in the spatial autocorrelated model, the first order case is to include the dependence on those regions that are immediately contingent. Second order autocorrelation would entail those regions contingent and those regions contingent with those immediately contingent.

corresponding row and column. The matrix element comprising the row-column intersection of contiguous regions is set to one. A region is not contiguous with itself, so the diagonal is set equal to zero. To form contiguous weights requires standardizing the contiguity matrix by dividing the row elements by the sum of the respective row such that each row of the contiguous weighting matrix sums to one.

What is considered a contiguous relationship becomes somewhat murky. Consider the contiguous relationships across counties. Generally contiguous neighbors are neighboring counties that share borders and or corners. Potter County in Texas is a good example of a square county with four counties sharing its border and four counties sharing corners. Inclusion of the corner counties as contiguous is subjective. Tulsa County in Oklahoma, at the other extreme, is oddly shaped and shares borders with seven counties. No county intersects Tulsa County at the corner without also sharing a border, though Muskogee County, which does not border or share a corner with Tulsa County, very nearly does and would have if it were not for strange politicking. Muskogee County is 3.5 miles short of intersecting Tulsa County while Pawnee county borders Tulsa County by less than 2 miles. In the case of Tulsa County, constructing the contiguity matrix requires a good deal of discretion. Regardless of the shape and proximity of spatial systems, care must be given to account for relevant regions that share some causal relationship across boarders.

Failure to account for contiguous relationships by applying ordinary least squares to estimates produces prediction errors that vary systematically over space. Since the errors espouse some form of systematic variation across space, the error is said to be spatially autocorrelated. In time-series models, autocorrelation causes consistent but inefficient estimates henceforth, correcting for autocorrelation entails correcting for inefficiency in parameter estimates. Spatial autocorrelation can be analogously modeled. To exemplify, consider the spatial error model (SEM),

$$\begin{aligned} &\mathbf{y} = \mathbf{X\beta} + \mathbf{u} \\ &\mathbf{u} = \boldsymbol{\rho}\mathbf{W} + \boldsymbol{\varepsilon} \\ &\boldsymbol{\varepsilon} \sim N(\mathbf{0},\sigma^2\boldsymbol{I}_{\mathrm{n}}), \end{aligned} \tag{2.17}$$

where $\mathbf{y}$ is the n vector of cross-sectional observations for a region, $\mathbf{X}$ is an $n\times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, $\boldsymbol{\rho}$ is a vector of correlation terms to be estimated and describes the relationships of errors across contiguous regions, $\mathbf{W}$ is the $n\times n$ contiguity matrix described above, and $\boldsymbol{\varepsilon}$ is a vector of Gaussian disturbances (LeSage 1997). This specification is analogous to a moving average error process over space rather than over time, and results in a conditional covariance matrix of $\mathbf{y}$ given $\mathbf{X}$ that is non-scalar. The OLS estimator for $\boldsymbol{\beta}$, though unbiased, is inefficient. Efficient estimates of the vector of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ can be derived with the Cochrane-Orcutt procedure of iterative solves (1949).

Anselin (1980) shows that this approach is appropriate only in a limited class of spatial processes. Another class of spatially dependent models that results in biased parameter estimates is the spatial autoregressive models (SAR) represented as,

$$\begin{aligned} &\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X\beta} + \boldsymbol{\varepsilon} \\ &\boldsymbol{\varepsilon} \sim N(\mathbf{0},\sigma^2\boldsymbol{I}_{\mathrm{n}}), \end{aligned} \tag{2.18}$$

where $\mathbf{y}$ is the n element vector of dependent variables, $\mathbf{X}$ $n\times k$ matrix of explanatory variables, $\mathbf{W}$ is a known $n\times n$ matrix of contiguous weights, and ρ is a correlation coefficient on the spatially lagged dependent variable to be estimated. Equation (2.18) is analogous to the lagged dependant variable regression model. Anselin (1980, pp. 58) recommends standardizing the variables for estimating Equation (2.18) where the estimated parameter ρ measures the variation in the vector of observations in $\mathbf{y}$ that are explained by the average neighboring observations (LeSage 1997) and the error term $\boldsymbol{\varepsilon}$ has the usual Gaussian distribution. Anselin shows that the parameter estimates of Equation (2.18) using ordinary least squares produces biased and inconsistent estimates (1980, pp. 58) resulting from simultaneity across explanatory variables.

Anselin (1980, Pg. 9) identifies two broad motivations for the special treatment of space in econometric models leading to the application of different methods for inclusion of spatial relationships. The first is accounting for spatial dependence, which is the functional relationship between events occurring in two points in space. Omitting these interregional relationships across space, results in the omission of systematic information useful in producing accurate forecasts. The second motivation is accounting for heterogeneous relationships across regions. Including interregional relationships in estimation equations is not sufficient to account for contiguous relationships. Lack of uniformity across space requires that particular features of each interrelated region be accounted for in model specifications. Heterogeneous regions are expected to have abruptly or slowly evolving changes in the spatial relationships over time. The traditional estimation assumption of constant relationship over time is inappropriate in this setting. The first motive is ideally suited for VAR estimation where parameter estimates are free to denote spatial interdependence. The second motivation is best suited for time-varying parameter estimates. The Kalmon filter is one approach to this heterogeneous relationship problem and is beyond the scope of this inquiry.

A single variable VAR representation of a multi-regional system is characterized as an n variable VAR in n regions. A multiple variable VAR representation of a multiregional system models k regionally specific variables in n regions resulting in an $n \times k$ variable VAR. Adding s lags will require estimating ($n \times k \times s$+1) coefficients per equation. Such a large VAR system will generally suffer from the over-fitting and sever multicollinearity for which the Bayesian VAR framework was designed to correct. Therefore several authors have pursued Bayesian methods in fitting multiregional VARs to data.

Bayesian solutions to Anselin's spatial framework have been addressed by augmenting multiregional BVAR forecasting systems with contiguous regional variables. In such a multiregional setting, regional economic variables such as employment for the n included regions are included as single equations in the n variable VAR system. Spatial relationships are integrated into the BVAR with the Litterman's random-walk prior means specification replacing the prior variance-covariance matrix with a modified contiguity matrix that accounts for contiguous relationships. Specifically, this replaces the relative weight matrix f with the contiguity matrix w, producing a prior variance-covariance matrix of Equation (2.14) as,

$$\lambda_{ijk}^{2} = \left(\theta\, w(i,j)\, g(s)\left(S_i / S_j\right)\right)^{2}. \qquad (2.19)$$

The contiguity matrix, is characterized has having ones down the main diagonal and in positions associated with contiguous entities. Values less than one are placed on all off-diagonal positions associated with non-contiguous entities. This spatial allocation of weights relaxes zero mean restrictions of the random-walk specified Minnesota prior for those variables expected to influence the current observation, and tightens estimation to the prior means of zero for those postulated to be without relationship.

In this, LeSage (LeSage 1989) models monthly average hourly wages by industry for Ohio's four major MSAs in a VAR framework for forecasting. The strategy follows LeSage and Magura (1986) in that empirically linking regional industry employment to neighboring industry employment provides useful information unique to other regional variables and national variables. Twenty sets of VARs are constructed for 20 industry sectors. Similarly a BVAR variant is devised based on the Minnesota prior random-walk means but incorporates the contiguity matrix in assigning prior precision. Contiguous relationships are not determined by spatial proximity, but rather by general outside knowledge of interspatial linkages across the 20 industries. Tightness and decay parameters are assigned across industry groups to reflect the type of industry and historic knowledge of their respective AR structure. Out-of-sample predictions shows that this BVAR model with contiguous weights gives superior forecasts over univariate AR and UVAR models, giving strong evidence that UVAR models are not viable forecasting models for spatially related models.

Though LeSage finds that his spatial BVAR model outperforms unrestricted VARs, his study fails to discern the source of forecast improvement. The relative performance of the contiguous priors found by LeSage may be the result of the shrinkage properties of the random-walk prior and not a function of the contiguous information. To discern the two requires comparing symmetrically specified Minnesota priors against similarly specified priors that also account for contiguous relationships. If the two priors produce equally compelling forecasts, then the shrinkage property of the random-walk prior is accountable for the forecast performance. Otherwise, an improvement in forecasts from contiguous priors over symmetric priors indicates that the induction of spatial information aids in forecast precision.

Three studies pursue this line of inquiry. Pan and LeSage (1995) employ contiguous priors to BVAR equations for forecasting corn production of 15 major corn producing states along two strategies. The first strategy is simply to form a 15 state VAR system of corn production where each state is regressed across the remaining 14 states. The second strategy develops 15, two-variable VARs in corn production and price with deterministic lagged acreage and production of the 14 remaining states as exogenous to the system. Contiguous priors are used to introduce co-movements across neighboring regions by relaxing weights to the set of random-walk prior means. LeSage and Pan (1995) expand the prior study of corn production by adding a third model to the second forecasting procedure that imposes a non-contiguous weight matrix of the Minnesota prior of contiguous state output. Further evidence is shed in Dowd and LeSage (1997) in modeling contiguous price-level relationships across states. The authors test contiguous relationships and estimate state BVAR models along contiguous states using the Minnesota and contiguous random-walk priors. The authors further compare results to unrestricted VAR models to assess the benefits of the random-walk prior.

In all three studies, the authors find that relaxing the Minnesota prior means restriction of zero on contiguous state variables increases forecast performance and supports the importance of contiguous regions' information in forecasting regional economies. Dowd and LeSage also reinforce the benefits to shrinkage estimation by comparing results to unrestricted VAR forecasts. Hence BVAR estimators with contiguous priors benefit from both the shrinkage property of the random-walk prior in correcting for multicollinearity and the systematic imposition of contiguous information in estimation that aids in capturing co-movements across neighboring regions.

LeSage and Pan further generalize the contiguous BVAR by including an error correction term in the Bayesian Error Correction (BECM) framework. A cointegrating relationship across contiguous regions will cause miss-specification bias if not accounted for. Generally it seems plausible that regional variables should some equilibrium growth relationship, so LeSage and Pan further generalize the contiguous BVAR of Pan and LeSage by including an error correction term in the Bayesian Error Correction (BECM) framework. Contrasting BECM and BVAR forecasts, testing the inclusion of contiguity relationships, and testing the incorporation of contiguity weight matrix over the Minnesota prior specification shows that including the contiguous relationships contribute to forecast accuracy and that relaxing of the AR1 restrictions of the Minnesota prior specification for contiguous regions extends those accuracy gains.

Though not directly reported in LeSage and Pan's two papers, the Bayesian gains, from the contiguous priors, to forecast accuracy tends to erode for long forecast horizons. Rickman and Miller (2002) find that inclusion of inter-industry structural linkages aids in producing accurate long-run forecast. Though Rickman and Miller specified inter-industry linkages as structure, it is conceivable that modeling inter-regional structure should produce similar results. The random-walk structure of the priors in LeSage and Pan's papers restricts the inclusion of the spatial structure while the unrestricted VAR places greater emphasis on modeling the full intra-regional structure indicating that industry structure facilitates long-run forecast accuracy while inter-regional structure does not. The reasoning for the prior can be found in Anselin's second motivation for accounting for inter-regional relationships; heterogeneous relationships.

Further evidence as to why forecast accuracy systematically erodes over forecasts horizons relative to unrestricted VARs are found in a subsequent paper by LeSage and Krivelyova (1999) discussed below. Prior VAR applications that integrate Bayesian priors to impose spatial structure on the model have relied on the random-walk prior means specification and a prior variance structure that allows nearby entities to exert more influence than more spatial entities. LeSage and Krivelyova (1999) develop a set of prior means for use in forecasting spatial models that place greater influence on one-lag contiguous observations and less emphasis on the own one lag observations. The authors note that their spatial prior is more appropriate for estimating regional models since the priors generalize the priors observed to this point by entail both spatial means and weights. They apply their prior specification to BVAR monthly forecasts employment for eight states

Noting that the first order spatial autoregressive model for cross-sectional data specified in Anselin (1988a) as,

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n), \end{aligned} \tag{2.20}$$

suggesting a prior distribution different from the spatial random-walk process discussed to this point. In Equation (2.20), $\mathbf{y}$ is an $n\times 1$ vector of dependent variables of regions, ρ is a scalar coefficient to be estimated, and $\mathbf{W}$ is an $n \times n$, first-order weight matrix with zeros along the main diagonal values representing shared influences of contiguous regions. The weight matrix is row-normalized such that the rows sum to unity producing a spatial lagged variable $\mathbf{Wy}$ that gives the average values of neighboring regions. Because this prior specification creates a series as an average value of neighboring regions, it is necessary to scale or transform the series such that they have similar magnitudes.

Equation (2.20) suggests a prior means specification different from the Minnesota prior, which emphasizes the random-walk. A reduced form VAR framework, which adds a time element to Equation (2.20), suggests placing emphasis on first lags of observations from contiguous regions and less influence on lags greater than one. Rather than specifying a random-walk prior, the spatial autocorrelation model suggests a random-walk prior of the average of the first-order autoregressive influences from contiguous regions. The prior variance structure is based on the belief that non-contiguous influences should have no direct influence on a region, and there is a decline in influence over time such that distant lags should have no influence on prediction, and that variation in lags of contiguous regions captures the full-extent of lag influences. The prior variance structure is designed such that own lags, geographically distant lags, and temporally distant lags have tight prior variances imposing prior means of zero tightly, while the first lags of contiguous regions have loose prior variances imposing loose prior means that distributes influence equally across contiguous regions.

The authors compare out-of-sample forecasts for all industry-state employment combinations and find that overwhelmingly, this Bayesian spatial prior outperforms other spatial priors employing AR1 means. They also find that this specification outperforms error correction and Bayesian error correction models based on the random-walk means. They further find that adding a cointegrating term to this spatial structure does not increase model performance, suggesting that the emphasis on contiguous regions captures cointegrating relationships. Even relaxing the zero restriction on the own-lag coefficient does not increase forecast performance, highlighting the importance of contiguous regions in forecasting regional economies.

In this regard, the BVAR method of accounting for spatial correlation has left a point unaddressed. The SEM model of spatial autocorrelation, gives unbiased yet inefficient parameters, while the SAR model of spatial correlation, results in biased coefficient estimates if not corrected for. Anselin devises efficient and unbiased estimators for both. Since Bayesian estimators are necessarily biased to the data, does Bayesian integration improve upon the maximum likelihood estimators of Anselin, or do they offer convenience at the expense of biasedness? A hint toward the answer appears throughout the literature. That is the random-walk prior means mentioned throughout reveals that the shrinkage property of the BVAR lead to superior out-of-sample forecasts. The attributes of shrinkage estimators in macroeconomic forecasting is well known (Diebold 1998). The questions as to whether the forecasts BVAR models introduced here benefit from the inclusion of relevant outside information, or if forecast predictions are improved through mitigating multicollinear effects needs more attention.

If employing both the IO table and the contiguity matrix in formulating the Bayesian priors for BVAR models improve forecast accuracy in isolation, it seems intuitive that entailing both simultaneously may contribute further gains in forecast accuracy. Along this line of inquiry, Magura (1998) integrates both national inter-industry structure and inter-regional relationships in specifying prior weights within the random-walk prior means specification for forecasting Ohio and the surrounding states' employment. The exercise limits industries to four tradable industries excluding tertiary industries that tend to not trade over regions. This results in a 20 variable VAR system that is estimated in its unrestricted form, BVAR with Minnesota priors, and three random-walk priors models: one with interindustry, another with spatial, and the third with a combined industry and special linkages as prior weights. Industry forecasts are compared across all models to discern the gains to the Minnesota prior, industry linkages, spatial linkages, and combined industry-spatial linkages. The findings favor combining industry and spatial information in model estimation.

Magura's limitation to tradable industries leaves open the possibilities that his findings are not general. Restricting the analysis to tradable goods producing sectors, that have deep inter-industry linkages and deep spatial linkages, fails to capture the general benefit of combining IO and spatial information in estimation. A more general set of industries has yet to be analyzed.

OTHER APPLICATIONS

As previously noted, there are endless ways to apply Bayesian methods to model specification. So far this review of applications have found Bayesian priors used to impose diss-information in the form of the random-walk prior of the Minnesota prior, to integrate empirical structure to VAR models through the IO table, and finally to impose spatial structure to model estimation with the contiguous priors. This section reviews other, less common Bayesian applications in modeling regional economies.

One such approach has been the integration of national and regional information in estimation. LeSage and Magura (1988) employ Bayesian mixed estimation in combining pooled time series estimates to a regional forecasting model of SMSA employment in levels. Their regional forecasting model entails several regional and national variables that capture leading and contemporaneous relationships thought to describe the eight Ohio SMSAs. The authors estimate the pooled coefficients for the eight SMSAs using ordinary least squares and apply those estimated coefficients to individual SMSA OLS estimates using the Bayesian mixed estimation,

$$\overline{\boldsymbol{\beta}}_{\mathbf{i}} = \left(\mathbf{X}_i{}'\mathbf{X}_i + \lambda_i \mathbf{I}\right)^{-1}\left(\mathbf{X}_i{}'\mathbf{X}_i\hat{\boldsymbol{\beta}}_i + \lambda_i\tilde{\boldsymbol{\beta}}\right). \tag{2.22}$$

The mixed estimator $\overline{\boldsymbol{\beta}}_i$ is the weighted average of the OLS estimator for industry i $\hat{\boldsymbol{\beta}}_i$ and the pooled estimator $\tilde{\boldsymbol{\beta}}$. The value of the weighting parameter λ_i is determined by the value of the parameter that gives the best out of sample forecast.

The authors compare the forecast performance of the pooled Bayesian estimates against a random-walk specification, and identically specified OLS and Ridge regression models. They find that the Ridge and Bayesian mixed estimation models outperform the more traditional AR and OLS models. They further find evidence that Ridge regression models tended to outperform the Bayesian mixed estimation models, suggesting that the shrinkage like properties of the Bayesian mixed estimation is the source of forecast efficiency in Bayesian applications. But strength of this application is its ability to mitigate deficient regional data through Bayesian means. Bayesian estimators recover a degree of freedom for every estimated coefficient.

The paucity of regional data has led many regional modelers to turn to national or pooled national estimates in specifying regional equations (Jones and Whalley 1989; Kraybill *et al.* 1992; Treyz *et al.* 1992). Doing so leaves out regional specific relationships for which a regional model is designed to replicate. Rickman (1995) shows that multipliers derived by REMI, a widely used regional modeling system differ from multipliers that are derived from local data. But the paucity of this local data hinders analysts' ability to derive them in the first place. As an alternative, Rickman advocates use of national or pooled relationships in Bayesian estimates of regional equations. The Bayesian estimator corrects for insufficient degrees of freedom while national priors offer a first best guess value for parameters to take.

In this Rickman re-estimates the pooled nationally estimated equations in the REMI (Treyz *et al.* 1992) system with identically specified regionally specific equations for the Las Vegas economy. Bayesian priors are specified from the native pooled estimates of REMI to varying degrees of precision. Diffuse priors return OLS estimates while tight priors replicate the pooled estimates of REMI. He then calculates impact multipliers and finds substantial variations across those produced by REMI and those from local data.

Otrok and Whiteman (Otrok and Whiteman 1998) seek to identify an unobservable regional coincident and leading indicator for forecasting that summarily present the projected state of the economy for the state of Iowa. Motivating the study is the difficulty in expressing economic projections in a usable form for casual users. Their method follows that of Stock and Watson (1989; 1992) but differs in the method of identifying the latent dynamic factor. Rather than employing the Kalmon filter to estimate the unobservable indicator, the authors employ the Chib-Greenberg Markov-chain procedure (Chib and Greenberg 1994) to estimate the unobserved factor. Through the Markov-chain, a posterior distribution of the unobserved variable is derived from the conditional posterior distributions. The parameters are estimated using the means of the posterior distribution.

Otrok and Whiteman (1997) employ a noninformative Bayesian prior in an ongoing multivariable VAR of real and nominal income, employment and population. In forecasting tax revenues conditioned on the state VAR, the authors account for structural changes in tax code, and produce forecast probability distributions by specifying asymmetric loss functions and sampling from the posterior distribution. These draws numerically calculate the expected loss from the posterior distributions. The asymmetry of the loss function accounts for increased costs of over predicting tax revenues relative to under predicting them. The result is prediction distributions that accurately reflect asymmetric costs of failure and that corrects for data and reporting deficiencies in tax revenue data.

SUMMARY, TRENDS, AND CONCLUSION

This chapter has reviewed several applications of Bayesian methods of integrating non-data information into data estimation. Within the VAR methodology, Bayesian methods of integrating the random-walk prior to model estimates reveals unambiguous forecast improvement. Experimentation shows that much of that improvement rests on the shrinkage-like property of the Minnesota prior, where Bayesian estimates shrink coefficient estimates toward zero for own- and cross-variable lags.

Bayesian methods offer solutions to a host of empirical problems for regional model builders. Bayesian restrictions recover degrees of freedom lost from parameter estimation, therefore allows the estimation of models that otherwise are rendered impossible by data constraints. Furthermore, given that unrelated economic variables tend to share co-movements, the Bayesian methodology, along with the random-walk prior, produces shrinkage-like estimation that has been shown to mitigate the ill affects of multicollinearity. Within the Bayesian paradigm, the extent of outside information influence is under the control of the practitioner. The Bayesian VAR methodology grants flexible means of combining prior and observational information to form estimates that entail more than mechanistically determined values.

While frequentists are apprehensive about this joint determination of estimates that give rise to biased estimates, the true test of methodological merit lies in out-of-sample performance. Unambiguous out-of-sample forecast enhancements are found in all cases of imposing the Minnesota prior to VAR models. Further objections on objectivity grounds are generally unfounded in the application of Litterman's prior that treats all parameter coefficients symmetrically, imposing only an atheoretical random-walk process as prior information in estimation.

The desire to incorporate economic structure in forecasting models has led to several adaptations of the random-walk prior. It is rather difficult to maintain objections of the random-walk prior in light of its success in enhancing forecast accuracy. Nonetheless, the random walk prior does little to facilitate structure in forecasting models. Within the shadow of parsimonious forecasting models is the desire to capture not only trends in data but also correlation and causal relationships that will aid in understanding and capturing dynamic linkages within the economy. Forming objective, asymmetric priors that facilitate some outside source of information about co-movements in data is desirable for capturing these dynamic linkages. Capturing inter-industry wage, employment, and output relationships within VAR models has relied on national or regional input-output tables. Initial applications relied on the random-walk prior means, and prior precisions on those means representing input-output relationships. Exclusionary restrictions based on the input-output table entail tight prior precisions on the prior means of zero, and loose precision on relevant linkages.

Partridge and Rickman (1998) generalize this approach by employing the input-output model to specify prior means rather than precision. Rather than strict adherence to the random-walk prior, the authors calculate cross-industry responses from the input-output table and impose these responses as prior means in BVAR estimation. This facilitates differing closure assumptions by incorporating different degrees of endogeneity in setting prior means. Aside from increasing forecast accuracy, this approach presents the opportunity to test the soundness of the input output matrix. By testing the input-output table against historical data, the validity of the input-output linkages is verifiable. This appears to be a promising application to testing the structure of other general equilibrium models such as computable general equilibrium models.

Avoiding the criticism fronting Bayesian estimation, Bayesian applications have been employed in selecting appropriate linkages in regional forecasting models. Markov-chain, Monte-Carlo models test posterior distributions across different model specifications to locate the combination of right hand side variables that best replicate the dependent variable. This procedure is analogous to step-wise regression but avoids the deficiencies associated with compare measures of the determination of variation. The final selected model specification can be estimated with least-squares methods or can be estimated with mixed estimation using priors derived from the posterior distributions of tested models. The posterior distribution can be used as weights in mixed estimation to assign prior probabilities to inclusive model formulations through Bayesian model averaging. Non-relevant variables are weighted heavily toward exclusion based on the while others are weighted toward their weighted probable representation based on the Markov-chain posteriors.

Similarly to incorporating the input-output table, spatial proximity has been applied in specifying prior precisions on the random-walk prior means. On the assumption of co-movements across economic variables across regions, relaxing the atheoretical random-walk structure for contiguous regions is perceived to improve forecast precision. The intuition arises in that the interdependence of open economies are translated into co-movement across economic variables. Forecasting applications that utilize these co-movements will benefit in the inclusion of additional relevant information. In the Bayesian VAR application, relaxing the random-walk prior for spatial relationships offers both the shrinkage-like qualities of the Minnesota priors while advantaging from relevant spatial interdependence.

Bayesian vector error-correction models have been specified for the random-walk prior means with symmetric precisions based on the Minnesota prior and based on informative precision priors of the input-output table as well as the spatial contiguity matrix. Cointegrating relationships are expected to be found across non-basic industries to both basic and non-basic industries but not generally so across basic to basic industries. Likewise manufacturing industries that produce tradable goods are likely to see cointegrated relationships across regions, while non-basic industries are likely to be insulated from inter-regional relationships. Generally Bayesian applications are not suitable methods of estimating cointegrating relationships since estimates will be biased. But if capturing cointegrating relationships across industries and regions benefit forecast performance, it is practical to employ it.

Bayesian application offers a host of opportunities for imposing economic theory on otherwise atheoretical estimates. Limiting the support space of coefficients with prior expectations aids proper estimation of those coefficients with no prior expectations. To exemplify, consider a single linear equation to be estimated with two co-linear explanatory variables. If prior knowledge of possible values of one of the coefficients can be derived from prior studies or theory, then a restriction on the parameter space can be incorporated on this coefficient based on theory. The second coefficient is specified with a non-informative prior. Coefficient estimates of such a specification reduce the problem of multicollinear relationships by limiting the variability of the dependant variable and leaving the second coefficient to be determined by the remaining variability. Smaller support space and greater prior precision on the coefficient with expected values leaves more variation associated with the second at the possible expense of biased estimation if the prior is indeed false.

In the VAR approach where all prior coefficient values except the own first lag, are place at zero corrects for multicollinearity by imposing zero restrictions with greater confidence on those coefficients expected to take on zero values. The remaining variability is left for those coefficients on recent lags. The same concept applies to imposing economic theory. Coefficients assumed to be exclusionary are tightly bound toward zero allowing those coefficients assumed to be influential to capture co-variation in estimation.

Bayesian applications also mitigate degrees of freedom problems which are essentially extreme cases of multicollinearity. Restrictions placed on parameter estimates returns a degree of freedom lost to estimation. Alternative methods of mitigating degree of freedom shortages at the regional level has been the application of pooled time series methods where no systematic way exists to emphasize the importance of the local region to the over-all pooled coefficients. LeSage and Magura (1988) base their Bayesian pooled estimator on this bases. Though the authors find that forecast improvement is accounted for only in the shrinkage properties of the estimator, the finding is limited in that explanatory variables are limited to leading indicators. No attempt as been employed to impose national economic linkages to local Bayesian pooled estimation. For example, if increased auto manufacturing is found to cause an increase in steel production nationally, an associated relationship should be expected locally and the use of this estimated national coefficient can help guide local model specification.

No study to date tests the implied structural relationships resulting from BVAR models with informative priors. This is not surprising given that the applications presented above focus on forecast accuracy of alternative informative priors. Informative priors that affectively capture the structure of the underlying economy should aid in policy simulations as well as long-run forecasting. But the question of policy analysis is hardly covered in the current regional BVAR literature as it is in popular macroeconomics literature (Sims 1982; Sims 1986; Walsh 2003). An affirmative finding of accurate policy analysis with the successful forecasting strengths of informative BVAR models would clearly indicate a superior modeling paradigm sought by regional economists (Treyz *et al.* 1992).

Partridge and Rickman (1998) and Rickman (2002) hint at such a test by comparing estimated policy shocks to those derived from the IO table. An alternative methodology is the application of the Blanchard-Qua decomposition (Blanchard and Quah 1989) which separates out short- and long-term components of change. Intuitively, the policy impacts, or multipliers, found in computable general equilibrium and input-output models, reflect equilibrium-to-equilibrium responses to change. The implication is that these impact measures only reflect long-run adjustments and that disequilibrium transition period entails the lengthy adjustment process. If this is the case, time-series policy impact models based on such multipliers should return the same multipliers once the disequilibrium adjustment process is exhausted.

There clearly exists other structural information that can be useful in imposing structure on the otherwise atheoretical BVAR model with Minnesota priors. As computing capacity increases, regional analyst are increasingly turning to computable general equilibrium models to capture regional economic structure. These models are generalizations of the fixed-price input-output models, and represent a more generalized set of priors in estimation.

CHAPTER III

ENTROPY ESTIMATION OF THE OKLAHOMA POLICY SIMULATION AND FORECASTING MODEL

INTRODUCTION

Regional models for policy analysis and forecasting differ in approaches to model calibration. Policy analysis models are generally calibrated from outside sources based on optimizing behavior. First order conditions for optimizing behavior are derived and elasticity responses to changes in prices are specified from a host of sources that may or may not accurately reflect those responses of the region under study. Forecasting models tend to abstract from economic theory and optimizing behavior in favor or capturing historical relations over time. These models may or may not entail economic theory as the primary goal is to be able to replicate and interpolate historical series. While theoretical forecasting models capture economic relationships through correlations across theoretically related variables, forecasts from such models tend to be inferior to atheoretical forecasting models that rely on correlations with past observations in interpolating future expectations. This chapter explores the differences between policy and forecasting models and postulates a merger in the two that offers superior application of both policy analysis and forecasting. It concludes by presenting an estimation procedure known as entropy in estimating such models.

A common problem for regional economists is that empirical models designed for policy analysis tend to perform poorly in forecasting applications in comparison to parsimonious models, while empirical forecasting models tend to be inadequate policy analysis tools. Parsimonious models such as vector autoregression (VAR) models, or simple reduced-form econometric models, tend to forecast well, but they are unable to offer insights into the impacts of policy decisions. Structurally elaborate models for policy analysis, such as regional computable general equilibrium (CGE) models, necessitate extensive parameterization that requires data beyond what is routinely available in time-series form. These models therefore are almost exclusively formulated as static models that are calibrated to a benchmark-year data set with no ability to track or forecast time-series. The 'ideal' model of a regional economy would marry the policy analysis strengths of regional CGE models with the forecasting capabilities of parsimonious models.

REGIONAL CGE POLICY MODELING

Because of its detailed theoretical structure derived from neoclassical economic theory, regional CGE models allow for the examination of a plethora of regional policy issues (Partridge and Rickman 1988). The high degree of endogeneity that results from the general equilibrium structure of a CGE model provides insights into policy impacts that simpler partial equilibrium models fail to capture (Pereira and Shoven 1988). Unfortunately, implementation of a CGE models require the specification of a large number of parameters, which are often unavailable; this requires the use of 'best guess' values at the expense of model precision (West 1995), while requiring the routine use of comparative static CGE models. Specifically, insufficient time-series data requires calibration to a benchmark-year data set (Mansur and Whalley 1984), and ignoring any available time-series information imparts bias into the model that is difficult to detect (Jorgenson 1984; McKitrick 1995, 1998). The absence of a mapping between the CGE model and time-series data for a region begs the question of whether the model accurately describes the region under study. In addition, static CGE models are inadequate for applications where the time-path of returns and costs are crucial to policy maker decisions.

Dynamically-sequenced CGE models have been proposed as alternatives to comparative static CGE models, however, with few exceptions, *regional* CGE models have been static (Partridge and Rickman, 1998).[5] Nevertheless, most dynamic CGE models rely on the consumption component to create the time-path responses of policy shocks (Pereira and Shoven 1988), in which there is a trade-off between current consumption and savings for future consumption. The time-path responses are obtained by either simulating a balanced growth path through sequencing equilibria over time, by augmenting existing supplies of labor and capital, or by choosing parameters of the dynamic CGE model that produce a reasonable time paths based on known co-movements between variables (Partridge and Rickman 1988).

Several weaknesses exist in the dynamically-sequenced approach. With regionally mobile capital, the link between regional savings and long-run growth becomes tenuous, creating doubt about the wisdom of making the inter-temporal consumption choice the dominant source of a *regional* model's dynamic properties (Partridge and Rickman 1988). In addition, dynamic CGE models impose inter-temporal market clearing with perfect foresight, and require calibration over a time-path that is subjective depending upon the arbitrarily imposed timing of factor expansion (Partridge

[5] Exceptions include McGregor, Swales and Ping (1996) and West and Deepak (2001).

and Rickman 1988; Pereira and Shoven 1988). Finally, the sequencing of static short-run equilibria is typically made on the assumption that each period reaches equilibrium. In short, dynamically-sequenced CGE models are not ‘truly’ dynamic in the sense that they do not allow for disequilibrium for any discrete time periods, and the time dimension is arbitrarily imposed. This makes them ill-suited for forecasting, and less able to demonstrate that they reflect known co-movements in the economy for historical time periods.

PARSIMONIOUS REGIONAL FORECASTING MODELS

Regional econometric models consist of systems of either recursive or simultaneous stochastic equations and identities (Bolton 1985). Unlike regional CGE models that are typically calibrated to a benchmark-year data-set, regional econometric models estimate parameters with time-series data. Econometric estimation contributes to model fit to the time-series data for the region, allowing them to be used for forecasting. Nevertheless, econometric model structure is limited by the availability of time-series data, which often leads to the omission of key general equilibrium features of regional economies. For example, regional econometric models often follow Keynesian theory by implicitly assuming perfectly elastic supply and fixed prices, resulting in endogenous impacts proportionate to exogenous change in demand (Partridge and Rickman 1988). This parsimony limits the range of policy uses regional econometric models can address.

Vector autoregressive models are generally employed in modeling economic systems and are generally considered generalizations of structural econometric models (Zellner, 1979). VAR models are generalizations of their structural counterparts because imposing exclusion restrictions on the VAR system to capture theoretical economic structure reverts the VAR to a structural econometric system. Sims (1980) cautions against imposing undue structure in econometric models noting that imposition of incorrect structure causes model misspecification, particularly where the extent of endogeneity of variables is questionable. Rather than imposing structure on econometric models, Sims recommends treating each variable symmetrically and allowing the estimation procedure to determine the extent of endogeneity of the model. He advocates the use of vector autoregressive models (VAR) as alternatives to structural econometric models, and argues that VARs are innate foundations for modeling dynamic relationships.

All variables entering a regional VAR model are considered endogenous, which greatly reduces the time costs of developing the model. Although they have enjoyed some success in forecasting (Litterman 1986a; Sims 1986), reduced-form VARs are not well adept to policy analysis. The structural vector autoregression (SVAR) model has been proposed as a solution to the limitations of the VAR for use in policy analysis (Sims 1986). But that brings back, full-circle, to Sim's claim of incredible restrictions in which the VAR methodology was designed to correct.

REGIONAL STRUCTURAL MODELS FOR POLICY ANALYSIS AND FORECASTING

An ‘ideal’ regional model would encompass structural attributes found in regional CGE models, while maintaining the dynamic fit of a regional econometric/VAR model. To be sure, the marriage is under way, as attempts to incorporate more structure into econometric models for policy analysis has ‘blurred’ the distinction between regional econometric models and regional CGE models (West, 1995). For example, some econometric models include neoclassical production functions and price responsive product demands. Yet, data considerations limit the amount of structure that can be incorporated, and the additional structure often comes at the expense of forecast accuracy.

A noteworthy attempt at such a marriage is the widely used model by Regional Economic Models, Inc. (REMI) (Treyz *et al.* 1992). The REMI model has its origins in the Massachusetts Economic Policy Analysis Model (MEPA) for the state of Massachusetts (Treyz *et al.* 1980). The structure of the MEPA model was subsequently used to construct similar models for other states, becoming publicly available in 1980 under the name REMI (Treyz *et al.* 1981). The first generation of commercially available REMI models, often referred to as the TFS modeling approach, is described in Treyz and Stevens (1985). In his review, Bolton (1985) observes that relative to other econometric models, the MEPA/TFS modeling system "…is a world apart in complexity, reliance on interindustry linkages, and modeling philosophy."

The TFS/REMI model bears some resemblance to CGE models in that it includes price responsive demands and supplies in the product and factor markets, interindustry transactions, and endogenous final demands. Unlike static regional CGE models, the REMI model integrates econometrically estimated parameters, and does not require all markets to clear continuously (Treyz *et al.* 1992). The econometric parameters, along with those exogenously specified, determine the time paths of economic responses to policy shocks.

However, the REMI model still falls short of the 'ideal' regional model. For one, many parameters are estimated outside the full set of general equilibrium constraints placed upon them as a system. Although these statistically estimated parameters may improve the fit of each equation, the model taken together may not be consistent with the economy it is intended to represent (Arndt *et al.* 2002), and may limit forecast accuracy.[6] Moreover, due to limited time-series data for a region, econometric estimates obtained from pooling time-series data for cross-sections are routinely used. Pooled estimates may be biased for a particular region, reducing the forecast accuracy of the model for that region (Rickman 1995). Many other parameters used from benchmark input-output tables/data are imposed as exact restrictions. The benchmark year may not be representative of other years in the region, and employing input-output information in the form of stochastic restrictions (Bayesian priors), rather than exact restrictions, improves forecast performance (Rickman 2001, 2002; Rickman and Miller 2002).

In short, because of the absence of system-wide calibration, the REMI model cannot be demonstrated to be representative of the economy under study, which is a legitimate concern of policy makers. In addition, the methods used to parameterize the REMI model have been shown to be inferior to other approaches in terms of forecast accuracy. To both improve model forecast accuracy and demonstrate its policy analysis capabilities, an approach is required that calibrates the entire model to movements of key variables in the local economy. Fortunately, recent advances in computer intensive computation have made system estimation of complex dynamic systems more feasible.

RECENT SYSTEM ESTIMATION APPROACHES

System approaches have become routinely used to calibrate dynamic stochastic general equilibrium (DSGE) macroeconomic models (Kim and Pagan 1995). Calibration

[6] Rickman and Treyz (1993) examined the out-of-sample forecast performance of the REMI model and several versions with alternative labor market closures. However, the parameterization of the versions was fixed, not being adjusted to improve forecast accuracy.

of DSGE models proceeds by specifying parameters of the model and simulating data from the model, and comparing properties of the simulated data with actual data. If the simulated data correspond well to the actual data, the DSGE model is considered to be representative of the underlying economy. Parameters can be adjusted to improve the 'fit' between the simulated and actual data, thereby, dynamically calibrating the model. Advances in dynamic calibration of DSGE models have come in the form of Bayesian approaches (Dejong *et al.* 1994, 1996), which purport to provide sounder statistical frameworks to evaluate model performance and provide a formalized means for incorporating available prior information into calibration of the model. Nevertheless, the DSGE models implemented are usually small-scale, and typically are linearized for estimation purposes. Implementing a regional model for both forecasting and policy analysis necessitates an estimation procedure that can yield parameter estimates for a nonlinear large-scale model, which incorporates general equilibrium restrictions, based on data that is limited or missing for many structural variables.

An approach advocated for use where data is limited and the number of parameters to be estimated is large, is the maximum entropy (ME) approach (Golan *et al.* 1996). Arndt *et al.* (2002) applied the generalized maximum entropy (GME) approach to calibrating a large-scale static CGE model for Mozambique. In this formalism, Arndt *et al.* calibrate the static model to a base year and then adjust elasticities of substitution so that the model better tracks the historical data for key target variables.[7] Their approach is similar to that of Jorgenson (1984) and McKitrick (1995) in that parameters are simultaneously estimated using time-series data, which produces parameter estimates that

[7] Robertson, Tallman and Whiteman (Robertson *et al.* 2002) also use an entropy approach to impose moment restrictions derived from theoretical models on forecasts produced by atheoretical macroeconomic VAR models.

are consistent with the overall system. Nevertheless, their method of simulating the time-path of the economy involves a series of equilibrium solves of a static model. While their entropy approach results in improved model performance for comparative static policy analysis, the model is not truly dynamic, and therefore precludes accurate quantitative assessments of regional economic dynamics for policy simulation or forecasting.

RECENT APPLICATIONS OF ENTROPY METHODS IN ECONOMICS

The method of maximum entropy and its cousin the general maximum entropy (generalizes the ME approach by taking into account noisy data) are versatile tools for economists. These robust approaches to estimation require minimal distributional assumptions, yet can solve ill-posed problems that are not possible with traditional methods. The ME formalisms are versatile and employ all available data, all relevant constraints, and prior information in the estimation process without requiring the input information to be complete (Arndt *et al.* 2002). Because the ME principle is so versatile, it has been applied to many problems where information must be extracted from data. Golan *et al.* (1996) have made important contributions to ME applications in econometric models that has initiated a variety of literature applying the ME principle to econometric problems.

One such application is the recovery of information from incomplete economic data. Golan *et al.* (1994) apply entropy in recovering flows from a multisectoral SAM framework. The method applied is similar to the RAS method of bi-proportional adjustment. The problem is to employ a complete data matrix for a particular year and estimate a matrix of flows based only on the row and column sums of the next year. Because this problem requires the use of prior expectations of the distributions of observations, the cross-entropy formalism is applied. The authors specified the priors, **q** from the completed table of the prior year and estimated the succeeding year's completed table as those values that minimized the Kullback-Leibler Information Criterion objective function subject to the constraints of the row and column totals.

Paris and Howitt (1998) propose an application of ME to the specification of flexible functional form (FFF) production functions to farm level data from European Union member countries. Empirically estimated FFF production functions often require a great number of restrictions for estimation since unrestricted estimates often give inappropriate curvature conditions (Diewert and Wales 1987). Since imposing exact restrictions to assure proper second-order conditions may give biased results, application of stochastic restrictions, may produce useful results with proper second-order conditions. The use of Bayesian stochastic restrictions to FFF production functions are uncommon, most probably because establishing Bayesian prior restrictions of such productions functions is difficult. The CE framework is a practicable method of estimating FFF production functions with proper curvature conditions.

Paris and Howitt (1998), apply a two-stage process to recover FFF cost functions from very little data with the ME formalism. The three-stage process determines the functional form and estimates the form's parameters even though the data is insufficient and results in an ill-posed problem. This process exactly reproduces the base period so no outside estimates are required to calibrate the model. The authors further show the estimated model is robust to policy simulations.

Rather than estimating a production system, Golan *et al.* (2001) apply the ME formalism to estimating an almost ideal demand system (AIDS) of the Mexican market for meat products. The ME formalism allows the AIDS to be estimated with binding nonnegativity constraints. Because the AIDS framework requires estimating budget shares, the ME formalism is applied to recover the unknown parameters of a nonlinear, censored demand system. The authors note three benefits derived from applying the ME formalism to AIDSs over previous methods. The first is that the system can be estimated with binding nonnegativity constraints without requiring the usual two-stage process of estimation. The second advantage is that the estimates are robust to assumptions of the distribution of the error terms. Especially relevant to regional modeling, the third benefit is that the ME method is applicable when data availability impedes other methods.

Robertson *et al.* (2002) apply the CE formalism to impose moment restrictions on simulated forecast distributions. The process is unique in that the authors apply importance sampling from the synthetic probability measure described in Csisrár (1975). The result is an expected forecast sampling distribution that satisfies the set of moment conditions that are imposed by economic theory or prior beliefs. Applications of this process include forecasting and policy analysis, where the extent of the structure imposed is up to the modeler.

Arndt *et al.* (2002) attempt to transcend the calibration and estimation problem in the absence of sufficient economic data by combining calibration with entropic estimation to simultaneously capture economic structure and the historical record in model estimation. Because their model is calibrated to a base year, the model still suffers the bias imposed with exact calibration to a base year (McKitrick 1998; Roberts 1994). Furthermore, their dynamic specification relies on a series of equilibrium solutions of the model rather than allowing the data to fully dictate the time-path responses. A complete method of estimating CGE parameters without calibration is in order to fully track history.

The current research proposes a model that integrates the forecasting accuracy of a regional econometric model with a policy-relevant structure that is representative of that associated with a regional CGE model. It extends current regional modeling by estimating the modeled relationships employing a maximum entropy (ME) approach that produces parameter estimates that satisfy the full set of general equilibrium constraints. The ME approach can be used to estimate models that contain numerous parameters in cases where data are limited. This approach allows for calibrating the model to the time-series movements of key variables in true dynamic fashion, while imposing Bayesian-type prior information; in short, the approach provides a sound empirical foundation for the model's quantitative predictions for both policy analysis and forecasting.

MAXIMUM ENTROPY AND CROSS ENTROPY ESTIMATION

This section introduces the theory behind entropic estimation and applies this to systems estimation. As described below, maximum entropy, ME and cross-entropy, CE are themselves merely ways in which probabilities of discrete events can be assigned. In its most basic form, the ME approach seeks to distribute equal probabilities to all prespecified discrete events, unless evidence supports otherwise. The CE approach seeks to distribute probabilities to all prespecified discrete events to the exact prior distribution imposed, unless evidence supports otherwise. In generalizing the ME and CE approach, the discrete set of events are mapped into parameter space identifying the structure of the data generating process.

The ME method follows Shannon (1948) and Jaynes (1957a; 1957b), who built on the work of Boltzman on the second law of physics. The ME formalism is founded on information theory and seeks to transform the empirical moments of the observed data into a distribution of probabilities. This section describes the foundation of the ME procedure and follows closely to Golan *et al.* (1996).

To formulate the development of ME, suppose an experiment of N trials with M discrete possible outcomes. Denote the number of times that an outcome m is observed as N_m, such that,

$$\sum_m N_m = N, \quad N_m > 0 \text{ and } m = 1, 2, \ldots, M. \tag{3.1}$$

There are M^N possible outcomes since there are N trials with M possible outcomes. The probability of outcome m converges in probability as,

$$\frac{N_m}{N} \xrightarrow{p} p_m \text{, or } N_m \xrightarrow{p} p_m N \text{, for } m = 1, 2, \ldots, M. \tag{3.2}$$

Counting principles show that the number of ways a set of N_m can be realized is characterized as the multinomial coefficient,

$$W = \frac{N!}{\prod_m N_m!} = \frac{N!}{Np_1!Np_2!\ldots Np_M!}. \tag{3.3}$$

Forming the monotonic, log transformation of W produces,

$$\ln W = \ln N! - \sum_{m=1}^{M} \ln N_m!. \tag{3.4}$$

Applying Stirling's approximation gives,

$$\ln W \approx N \ln N - N - \sum_{m=1}^{M} N_m \ln N_m + \sum_{m=1}^{M} N_m, \tag{3.5}$$

and further substituting Equation (3.1) into Equation (3.5) gives,

$$\ln W \approx N \ln N - \sum_{m=1}^{M} N_m \ln N_m. \tag{3.6}$$

Applying the second relationship in Equations (3.1) to Equation (3.6) and rearranging gives,

$$N^{-1} \ln W \approx -\sum_{m=1}^{M} p_m \ln p_m = H(\mathbf{p}) \text{, where } -p_m \ln p_m = \{0 : p_m = 0\}, \tag{3.7}$$

which is Shannon's measure. The probability measure for event *m*, p_m is restricted to be between zero and unity, and the sum of the *m* probabilities must sum to one. This is a measure of uncertainty of the combination of events and is maximized when,

$$p_1 = p_2 =,\ldots,= p_M = 1/M. \tag{3.8}$$

That is, Shannon's measure is maximized when the probabilities of the *M* discrete events are equally distributed.

By maximizing Shannon's measure subject to the restriction imposed by the data, we maximize the number of ways in which we can observe the probabilities that is consistent with the data (Jaynes 1957a, 1957b). The maximum of Shannon's measure is consistent with Laplace's notion of the *principle of insufficient reason*, which states that mutually exclusive events should be assigned equal probabilities in the absence of evidence to the contrary (Jaynes 1957a; Press 1989, pp. 47; Sinn 1980). Because the ME formalism assigns equal probabilities in the absence of contrary restrictions, the solution is the most conservative estimates in terms of Laplace's principle of insufficient reason. The means of imparting evidence derived from the data is presented below.

A related measure to Shannon's measure is the Kullback-Leibler Information Criterion (KLIC), of relative entropy. The KLIC is applied to a variant of the ME formalism known as the cross-entropy (CE) formalism. The CE formalism deviates from the ME formalism in how prior knowledge is presented in the estimation process. Rather than maximizing Shannon's measure of uncertainty, the CE method proposes minimizing the Kullback-Leibler distance (Good 1963; Kullback 1959) from the specified priors. The Kullback-Leibler Information Criterion (KLIC) is specified as,

$$I(\mathbf{p},\mathbf{q}) = \sum_{m=1}^{M} p_m \ln\left({p_m}/{q_m} \right), \text{ where } p_m \ln\left(p_m / q_m\right) = \{0 : p_m, q_m = 0\}, \qquad (3.9)$$

where the M-dimensional vector **p** is as defined above and the M-dimensional vector of **q** is the prior expected value of **p**. The elements of the probability **p** and prior **q** vectors are restricted to be between zero and one and to sum to unity. The extent of shrinkage on the estimates of **p** toward the prior is determined by the distribution of the prior vector **q**. If the M values of **q** are equal, then the *KLIC* reduces to Shannon's measure. If the bulk of the prior probability mass is centered on a single q_m, then estimate of p_m will be shrunk toward q_m. Through this, the CE formalism is an efficient method of imposing stochastic restrictions without specifying prior and posterior distributions since CE makes no distributional assumptions.

Mapping the probabilities into parameter support space generalizes the ME and CE formalisms. This is done by assuming the parameter to be estimated, β_k, as a discrete random variable with a compact support of M>2 discrete monotone increasing sequence of states, z_m. Because the pre-specified support set defines the upper and lower bounds, $[z_1, z_M]$, that β_k can take, care must be exercised to include the true parameter values. The less theory tells us about the true parameter value the wider the support bounds should be. Mapping the discrete probabilities into β_k recovers parameter estimates by expressing β_k as a convex combination of the z_m discrete states,

$$\beta_k = \sum_{m=1}^{M} z_{k,m} p_{k,m}. \tag{3.11}$$

Parameter estimates are specified as weighted averages of the support space $\mathbf{z}_k$ weighted by the probabilities $\mathbf{p}_k$. The optimum values of p_{km} from optimizing Shannon's measure or the KLIC subject to actual observations produces optimal values of β_k that fully utilizes the combination of data and prior information in estimation.

Since the CE formalism entails the combination of prior information with parameter estimates consistent with the data, the CE formalism runs parallel with Bayesian estimation methods that shrink parameter estimates toward prespecified expected values. Though shrinkage techniques coax parameter estimates away from the maximum likelihood values, biased parameter estimates may improve forecast performance by reducing the forecast error variance (Birkes and Dodge 1993). Such shrinkage estimation has had a long history in nonstructural model forecasting. A case in point is the Minnesota prior specification of Bayesian vector autoregression models (Doan et al. 1984). In an excellent introduction to the new challenges of macroeconomic forecasting, Francis Diebold (1998) speculates that shrinkage estimation will play a key role in estimating DSGE models.

Following the notation of Arndt, Robinson and Tarp (Arndt *et al.* 2002), consider the following representation of a structural economic model to be estimated,

$$F(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{B}) - \mathbf{e}_t = \mathbf{0}, \text{ for } \forall\ t \in T, \tag{3.12}$$

where F is a vector valued function that includes the model structure and dynamic linkages, $\mathbf{Y}$ is a vector of endogenous variables, $\mathbf{X}$ is a vector of exogenous variables, $\mathbf{B}$ is a K dimensional vector of behavioral parameters to be estimated, and $\mathbf{e}$ is a vector of stochastic error terms. Lagged values of both endogenous and exogenous variables can enter any given time t as contemporaneously exogenous variables such that $\mathbf{X}_t$ can be separated into exogenous variables that enter the model exogenously, $\mathbf{X}_t^{\mathbf{X}_t}$ lagged endogenous variables, $\mathbf{X}_t^{\mathbf{Y}_{t-1}}$. Therefore, substituting $\mathbf{X}_t^{\mathbf{X}_t}$ and $\mathbf{X}_t^{\mathbf{Y}_{t-1}}$ into Equation (3.12) and subscripting time t gives the theoretical model to be estimated as,

$$F\left(\mathbf{Y}_t, \mathbf{X}_t^{\mathbf{X}_t}, \mathbf{X}_t^{\mathbf{Y}_{t-1}}, \mathbf{B}\right) - \mathbf{e}_t = \mathbf{0}, \text{ for } \forall\ t \in T \tag{3.13}$$

Traditional econometric methods of estimating such a system require a two-stage process. Entropy estimation allows the simultaneous estimation of all parameters with simultaneous restrictions to be carried out in a single step.

The method of estimating the K behavioral parameters, $\mathbf{B}$, follows Golan *et al.* (1996). The k^{th} behavioral parameter B_k is treated as a discrete random variable with compact support and $2 \leq \text{M} \leq \infty$ possible outcomes. B_k is stated as a convex set of the upper and lower bounds of its support, $z_{k,1}$ and $z_{k,M}$, and all the support points in between as,

$$\beta_k = \sum_{m=1}^{M} p_{k,m} z_{k,m}, \text{ for } 0 \leq p_{k,m} \leq 1,\ M \geq 2, \text{ and } \sum_{m=1}^{M} p_{k,m} = 1, \tag{3.14}$$

In compact matrix notation, Equation (3.14) is,

$$\boldsymbol{\beta} = \mathbf{Zp} = \begin{bmatrix} \mathbf{z'}_1 & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{z'}_2 & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{z'}_K \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ . \\ \mathbf{p}_K \end{bmatrix}, \text{ for } 0 \leq p_{k,m} \leq 1 \text{ and } \sum_{m=1}^{M} p_{k,m} = 1, \tag{3.15}$$

where $\mathbf{Z}$ is ($K \times KM$), $\mathbf{p}$ is ($KM \times 1$), and $\mathbf{z}_k$ is ($M \times 1$) in dimension support set for the associated probability $p_{k,m}$. In Equation (3.15) the support set $\mathbf{Z}$ is determined in advanced such that the behavioral parameters are recovered by finding appropriate values of the probability-set $\mathbf{p}$.

The vector of error terms, **e**, is treated in a similar manner when solving for the optimal behavioral parameter values. The error terms are assumed to be discrete random variables with compact support and $2 \leq J \leq \infty$ possible outcomes,

$$e_t = \sum_{j=1}^{J} v_{t,j} w_{t,j}, \text{ for } 0 \leq w_{t,j} \leq 1, J \geq 2, \text{ and } \sum_{j=1}^{J} w_{t,j} = 1, \tag{3.16}$$

In compact matrix notation Equation (3.16) is:

$$\mathbf{e} = \mathbf{V}\mathbf{w} = \begin{bmatrix} \mathbf{v'}_1 & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{v'}_2 & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{v'}_T \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ . \\ \mathbf{w}_T \end{bmatrix}, \text{ for } 0 \leq w_{t,j} \leq 1 \text{ and } \sum_{j=1}^{J} w_{t,j} = 1, \tag{3.17}$$

where **V** is $(T \times TJ)$, **w** is $(TJ \times 1)$, and $\mathbf{v}_i$ is $(J \times 1)$ in dimension. The support set in Equation (3.17), **V** is determined a priori such that the errors of the system can be recovered using the estimated probability of the convex set.

Substituting Equations (3.15) and (3.17) into Equation (3.12) gives,

$$F(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{Zp}) - \mathbf{V}_t \mathbf{w}_t = \mathbf{0}, \text{ for } \forall\ t \in T. \tag{3.18}$$

There exist an infinite number of combinations of **p** and **w** that satisfy Equation (3.18) and yield valid estimates of the probability weights. Shannon's (Jaynes 1957a, 1957b; Shannon 1948) entropy measure, $H(\mathbf{p},\mathbf{w})$, assures a unique solution given the data,

$$H(\mathbf{p}, \mathbf{w}) \equiv -\sum_k p_k \ln p_k - \sum_t w_t \ln w_t. \tag{3.19}$$

Shannon's measure, $H(\mathbf{p},\mathbf{w})$ quantifies the degree of uncertainty in the distribution of probabilities (Golan *et al.* 1996, pp. 10). Systems estimation is made through Shannon's measure subject to the fit of the model and the consistency constraints specified as,

$$
\begin{aligned}
&\max_{\mathbf{p},\mathbf{w}} H(\mathbf{p},\mathbf{w}) \equiv -\mathbf{p}'\ln\mathbf{p} - \mathbf{w}'\ln\mathbf{w} \\
&\qquad s.t. \\
&F(\mathbf{Y},\mathbf{XZp}) - \mathbf{Vw} = \mathbf{0} \\
&\mathbf{1}_K = (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} \\
&\mathbf{1}_T = (\mathbf{I}_T \otimes \mathbf{1}'_J)\mathbf{w}.
\end{aligned}
\tag{3.20}
$$

Equation (3.9) seeks to maximize the claim of absolute disinformation of the structure of the underlying parameters, while Equation (3.20) augments this objective function with sample information in the form of optimization constraints. The first constraint imposes information in the form of the systems model fit against Shannon's measure, which is maximized with perfect disinformation. The second and third restrictions are the summation restrictions that assure the probabilities sum to one. The vector $\mathbf{1}_K$ is a K dimensional unit vector, and $\mathbf{I}_K$ is an identity matrix of size K. A similar vector and identity matrix of T dimension are created for the error terms. The symbol, $\otimes$, is the Kroneker product of matrices. The solution to the constrained optimization of Equation (3.20) offers the best-fit relationship between $\mathbf{y}_t$ and $\mathbf{x}_t$. The parameter coefficients and the error terms are recovered as $\mathbf{B} = \mathbf{Zp}$ and $\mathbf{e} = \mathbf{Vw}$, respectively.

The KLIC measure imposes shrinkage like properties to the distribution estimates. The estimated coefficients are shrunk toward the values that economic theory imposes through properly setting prior probabilities. The KLIC objective measure for the generalized cross entropy problem is specified as,

$$
l(\mathbf{p},\mathbf{w}) \equiv +\sum_k p_k \ln(p_k / q_k) + \sum_t w_t \ln(w_t / u_t), \tag{3.21}
$$

where $l(\mathbf{p},\mathbf{w})$ is minimized when $p_k = q_k$, $\beta k \in K$, and $w_t = u_t$, $\forall t \in T$. The GCE formalism therefore seeks to minimize the KLIC subject to the fit of the model and the summing up constraints of the probabilities,

$$
\begin{aligned}
&\min_{\mathbf{p},\mathbf{w}} l(\mathbf{p},\mathbf{w}) \equiv \mathbf{p}'\ln(\mathbf{p}/\mathbf{q}) + \mathbf{w}'\ln(\mathbf{w}/\mathbf{u}) \\
&\qquad\qquad s.t. \\
&F(\mathbf{Y},\mathbf{XZp}) - \mathbf{Vw} = \mathbf{0} \\
&\mathbf{1}_K = (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} \\
&\mathbf{1}_T = (\mathbf{I}_T \otimes \mathbf{1}'_J)\mathbf{w}.
\end{aligned}
\tag{3.22}
$$

Equation (3.22) seeks to minimize the log difference of q_k from p_k for all k explanatory variables and the log difference of u_t from w_t, for all t subject to the data and summing up constraints. The first restriction in Equation (3.22) is merely the statement of relationship between $\mathbf{y}_t$ and $\mathbf{x}_t$. The second and third restrictions are the summation restrictions that assure the probabilities sum to one. The solution to the constrained optimization problem of Equation (3.22) gives values of the behavioral parameters and error terms that optimizes the KLIC measure, $l(\mathbf{p},\mathbf{w})$, where the parameter coefficients and the error terms are recovered as $\mathbf{B} = \mathbf{Zp}$ and $\mathbf{e} = \mathbf{Vw}$, respectively.

The model structure to be estimated is imposed on the objective function through the moment constraints of Equation (3.18). Without such constraint Shannon's measure would be optimized where all states are equally likely and the KLIC would be optimized at the prior means specified by the prior probabilities. The data moments impose information to the contrary of absolute disinformation, in the case of Shannon's measure, and absolute prior knowledge in the case of the KLIC. The systems fit restriction is comparable to the exact identifying restrictions on single-year, static CGE models proposed by Shoven and Whalley (1972; 1984). Since the proposed model is to fit multiple years, the system of identifying relations must account for non-exact fits for particular years. The calibration process of dynamic general equilibrium models of Kydland and Prescott (1982) entails calibrating successive years, adjusting the calibrated parameters for the combined fit over all years (Hansen and Heckman 1996). This procedure is analogous to econometric estimation where the loss function is subjectively chosen by the modeler in deciding what the best overall fit of the data is.

CONCLUSION

The present study incorporates entropy estimation of the model described in Chapter IV over time. Relative to least-squares methods, entropy estimation has the advantage of utilizing all relevant outside knowledge about the parameter coefficient values. Loss functions are created for both parameter coefficients and prediction errors, where the full set of identifying restrictions can be posed with error. Formulating a loss function about the prior expected parameter values to be estimated corresponds to specifying a prior distribution for parameter estimates in Bayesian methods of estimation. For example, in specifying prior expected values of linear slope parameters, Bayesian analysts often assume lower and upper probable bounds postulated to represents one's own beliefs of the true, or population values, of the slope coefficients. Reinterpreting the upper and lower bounds as confidence intervals with confidence level alpha, a complete prior expected distribution can be created representing the expectations of the population data generating function.

Analogously, Cross-Entropy specification of prior beliefs requires at least a compact closed set representing an upper and lower absolute bound. A two-point support set can only represent the first and second moment about the mean prior and estimate of the parameter. Higher moments for prior specification and for estimation may be added by the addition of more support points offering a greater breadth of distribution assumptions (Golan *et al.* 1996, pp. 87).

In Chapter IV a three-point support set is built for both parameter coefficients and error terms allowing the structure of symmetric loss distributions corresponding to approximations of the normal distribution. Prior weights are initially assigned symmetrically, where subsequent trial and error warranted some alteration. The structure of the model is then assigned as the data moment restrictions to the CE formalism where optimizing the KLIC returns parameter estimates used to calibrate the model for forecasting and policy simulation.

CHAPTER IV

SPECIFYING AND ESTIMATING A REGIONAL POLICY SIMULATION AND FORECAST MODEL FOR THE STATE OF OKLAHOMA

INTRODUCTION

A common problem for regional economists is that empirical models designed for policy analysis tend to perform poorly in forecasting applications in comparison to parsimonious models, while empirical forecasting models tend to be inadequate policy analysis tools. Parsimonious models such as vector autoregression (VAR) models, or simple reduced-form econometric models, tend to forecast well, but they are unable to offer insights into the impacts of policy decisions. Structurally elaborate models for policy analysis, such as regional computable general equilibrium (CGE) models, necessitate extensive parameterization that requires data beyond what is routinely available in time series form. These models therefore are almost exclusively formulated as static models that are calibrated to a benchmark-year data set with no ability to track or forecast time series. The 'ideal' model of a regional economy would marry the policy analysis strengths of regional CGE models with the forecasting capabilities of parsimonious models.

The proposed model integrates the forecasting accuracy of a regional VAR/econometric model with a policy-relevant structure that is representative of that associated with a regional CGE model. It extends current regional modeling by estimating the model employing a maximum entropy (ME) approach. The ME approach can be used to estimate models that contain numerous parameters in cases where data are limited. The ME approach also allows for calibrating the model to the time-series movement of key variables in true dynamic fashion, through imposing Bayesian-type prior information; in short, the approach provides a sound empirical foundation for the model's quantitative predictions for both policy analysis and forecasting.

An 'ideal' regional model would encompass structural attributes found in regional CGE models, while maintaining the dynamic fit of a regional econometric/VAR model. As West (1995) has observed, attempts to incorporate more structure into econometric models for policy analysis has 'blurred' the distinction between regional econometric models and regional CGE models; for example, some econometric models include neoclassical production functions and price responsive product demands. Yet, data considerations limit the amount of structure that can be incorporated, and the additional structure often comes at the expense of forecast accuracy.

A noteworthy attempt at such a model is the widely used model by Regional Economic Models, Inc. (REMI) (Treyz, et al. 1992). The REMI model has its origins in the Massachusetts Economic Policy Analysis (MEPA) Model (Treyz, et al. 1980); the structure of the MEPA model was subsequently used to construct similar models for other states, becoming publicly available in 1980 (Treyz, et al. 1981). The first generation of commercially available REMI models, often referred to as the TFS modeling approach, is described in Treyz and Stevens (1985). In his review, Bolton (1985) observes that relative to other econometric models, the MEPA/TFS modeling system "…is a world apart in complexity, reliance on interindustry linkages, and modeling philosophy."

The TFS/REMI model bears some resemblance to CGE models in that it includes price responsive demands and supplies in the product and factor markets, interindustry transactions, and endogenous final demands. Unlike static regional CGE models, the REMI model integrates econometrically estimated parameters, and does not require all markets to clear continuously (Treyz, et al. 1992). The econometric parameters, along with those exogenously specified, determine the time-path of economic responses to policy shocks.

However, the REMI model still falls short of the 'ideal' regional model. For one, many parameters are estimated outside the full set of general equilibrium constraints placed upon them as a system. Although statistically estimated parameters may improve the fit of each equation, the model taken together may not be consistent with the economy it is intended to represent and may limit forecast accuracy (Arndt, et al. 2002). Moreover, due to limited time-series data for a region, econometric estimates obtained from pooling time-series data for cross sections are routinely used. Pooled estimates may be biased for a particular region and are found to reduce the forecast accuracy of the model for that region over regional estimates (Rickman 1995). Many other parameters used from benchmark input-output tables/data are imposed as exact restrictions. The benchmark year may not be representative of other years in the region, and employing input-output information in the form of stochastic restrictions (Bayesian priors), rather than exact restrictions, improves forecast performance (LeSage and Magura 1991; Magura 1987, 1990; Rickman 2001, 2002; Rickman and Miller 2002).

In short, because of the absence of system-wide calibration, the REMI model cannot be demonstrated to be representative of the economy under study, which is a legitimate concern of policy makers. In addition, the methods used to parameterize the REMI model have been shown to be inferior to other approaches in terms of forecast accuracy. To both improve model forecast accuracy and demonstrate its policy analysis capabilities, an approach is required that calibrates the entire model to movements of key variables in the economy. Fortunately, recent advances in computation have made system estimation of complex dynamic systems more feasible.

OUTPUT BLOCK

Building the regional simulation model requires the combined use of a benchmark set of input-output (IO) relationships and time-series observations on those series that are readily available with continuous histories. In the following section the IO table is specified that defines the structure of the regional economic system. To place this structure in the form of time-series observations requires placing observable and non-observable historical time-series into the structure of the IO table. Ideally a continuous history of IO tables can be stacked to track historical changes alleviating the need for fitting outside estimates to the structure. Because of frequent data revisions, the expense of complete IO tables, and the limiting restrictions placed on the assumptions of the Leontief system, a full set of IO tables that tracks history is not feasible. Therefore, following the presentation of the IO structure, focus is turned to the fitting of time-series to the IO structure.

The IO table is both an analytical framework for analyzing impacts of policy, and a descriptive tool for identifying key inter-sectorial linkages (Richardson 1972). In this latter context, use of the descriptive relationships is employed in developing the structure of final demands and production linkages based on a benchmark year. The production

Figure 3.1: Partial Transactions Table.

$$
\begin{array}{ccccccccccccccccccc}
 & Q_{1,1} & + & Q_{1,2} & + & C_1 & + & I_1 & + & SLG_1 & + & FED_1 & + & X_1 & - & M_1 & = & Q_1 \\
+ & Q_{2,1} & + & Q_{2,2} & + & C_2 & + & I_2 & + & SLG_2 & + & FED_2 & + & X_2 & - & M_2 & = & Q_2 \\
 & & & & & & & & & & & & & & & & & \\
+ & EC_1 & & EC_2 \\
+ & YPROP_1 & & YPROP_2 \\
+ & \underline{IBT_1} & & \underline{IBT_2} \\
= & Q_1 & & Q_2
\end{array}
$$

$$
\begin{array}{ccccccccccccc}
VA_i & = & EC_i & + & YPROP_i & + & IBT_i \\
FD_i & = & C_i & + & I_i & + & SLG_i & + & FED_i & + & X_i & - & M_i
\end{array}
$$

$Q_{i,j}$	Intermediary inputs from industry i to industry j.
Q_i	Output in industry i.
C_i	Consumption component of final demand of industry i.
I_i	Investment component of final demand of industry i.
SLG_i	State & local final demand for output of industry i.
FED_i	Federal government final demand for output of industry i.
M_i	Imports of goods for final demand for industry i.
X_i	Export component of final demand for goods of industry i.
EC_i	Employee compensation in industry i.
$YPROP_i$	Property type income earned in industry i.
IBT_i	Indirect business tax paid from producers in industry i.
VA_i	Total value-added production of industry i.
FD_i	Final demand of industry i.

function interpretation of the benchmark IO table is utilized to specify the complex set of linkages across industries. Figure 1 presents a “dog-leg” two-sector IO table that omits the institution-by-institution linkages found in the more complete social accounting matrices (SAM) representation.

The IO table is useful in describing the linkages comprising economic transactions in terms of value. The rows in the table specify the sales of industry *i* to all other industries, for intermediate demands, and sales for final consumption, for investment, to state & local and federal governments, and for exports less imports. The columns show the purchases of industry *j* from industry *i*, for intermediate inputs, and from primary factor inputs such as capital labor and land. The primary factor inputs comprise payments to factor inputs, or total value-added output, while sales to consumption investment, government and net exports reflect expenditures of income, or GSP. In equilibrium, total industry output (column sums) must equal total industry demand (row sums). Furthermore, since industry intermediate inputs must equal intermediate supply, total value-added output must equal total GSP. This completes the equilibrium requirement that output equates with demand.

Total regional output is defined as the sum of regional intermediate and final demands as defined by the regional counterpart to the national accounts identity. The regional accounts identity equation over time, *t*, is defined as,

$$Q_{it} = INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it} + X_{it} - M_{it},$$

where *Q* is total output and is defined as the sum of intermediate goods produced in the region and of the final demand components, *C*, *I*, *SLG*, *FED*, and *X-M*. These final demand components are defined as *C*, consumption expenditures, *I*, investment, *SLG* and *FED*, state & local government and federal government expenditures respectively, and *X-M* as net exports.

Regional imports, *M*, are defined as the intermediate and final demands imported from outside the region. Defining imports as a proportional value, *m*, of total local demand gives,

$$M_{it} = m_{it} INT_{it} + m_{it} C_{it} + m_{it} I_{it} + m_{it} SLG_{it} + m_{it} FED_{it}.$$

Substituting for *M* in the regional accounts identity gives,

$$Q_{it} = (1 - m_{it}) \sum_{j=1}^{N} Q_{ijt} + (1 - m_{it}) C_{it} + (1 - m_{it}) I_{it} + (1 - m_{it}) SLG_{it} + (1 - m_{it}) FED_{it} + X_{it},$$

further defining $(1\text{-}m_{it}) = rpc_{it}$ and simplifying gives,

$$Q_{it} = (INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}) rpc_{it} + X_{it}.$$

This is the regional account identity that is common in the regional IO literature (Richardson 1972).

To get historical series structured on the IO table requires fitting observable and estimates of non-observable time-series values to the output equation defined above. Since regional total output data is not collected and distributed on a regular basis, developing a time-sequenced model requires the regional accounts identity to be stated in value-added output or employment terms that are readily available. Industry value-added output, VA_i, is derived by assuming some non-constant proportional relationship between total output and value-added output, $\lambda_{VA,it} = VA_{it}/Q_{it}$, such that,

$$VA_{it} = Q_{it}\lambda_{VA,it} = \left(\left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}\right)rpc_{it} + X_{it}\right)\lambda_{VA,it}.$$

Furthermore, value-added measures are transformed into employment terms by some proportional relationship of employment to value-added output, $epv_{it} = L_{it} / VA_{it}$, which is easily calculated given available data from the Bureau of Economic Analysis. Substituting this relationship for industry value-added output above gives,

$$L_{it} = VA_{it} \cdot epv_{it} = \left(\left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}\right)rpc_{it} + X_{it}\right)\lambda_{VA,it} \cdot epv_{it}. \tag{4.1}$$

Equation (1) simply depicts employment derived from total output where the Bureau of Economic Analysis (BEA) reports industry employment L_{it} and value-added output annually. Employment reports are generally considered more accurate measures of economic activity than value-added output given the long duration of value-added output data revisions.

The BEA does not readily report historical estimates of intermediate inputs. Therefore estimates of historical intermediate inputs must be made. Intermediate inputs, Q_{ij}, are assumed proportionately related to total industry output by defining the technical requirements coefficient, as $a_{ij} = Q_{ij} / Q_i$. The $n{\times}n$ matrix of technical requirements coefficients make up what is known as the technical coefficients matrix in the IO literature. Defining the variable INT_i as the row sum of intermediate inputs gives,

$$INT_i = \sum_{j=1}^{n} Q_{ij}.$$

Further, substituting the technical coefficient relationship for Q_{ij} and adding the time subscript gives,

$$INT_{it} = \sum_{j=1}^{n} a_{ij} Q_{jt}. \tag{4.2}$$

Equation (2) is used to estimate intermediate demands over history. The technical coefficient, a_{ij}, is assumed fixed at the benchmark-year such that changes in intermediate demands are driven by a constant proportional relationship with total regional output.

Similar to intermediate demands, annual values of industry final demand components must be estimated, as complete histories are not readily available. Historical estimates of such are derived from fixed proportional relationships of regional endogenous drivers and national exogenous drivers. Define Y_i as the region's final demand component for industry i and Y, without the industry subscript, as the aggregate final demand component over all industries. Further define the variable Z as the regional driver of the local final demand component, then,

$$\hat{Y}_{it} = \rho_i \cdot \gamma_Y \frac{\hat{Z}_t}{Z_t^U} Y_t^U, \text{ where } \gamma_Y = \frac{Y/Z}{Y^U/Z^U} \text{ and } \rho_i = Y_i/Y,$$

where variables with no time subscript, t, are benchmark-year observations derived from the IO relationship, the character hats ($\hat{X}$) denote system estimated variables, and the U superscript denotes national exogenous drivers. The benchmark ratio ρ_i denotes the proportion of the industry final demand to the total final demand components, and the ratio γ_Y scales the estimated final demand component to that of the benchmark year. For completeness, each final demand component equation is described, starting with the private consumption component.

The private consumption component of final demand is characterized as being some constant proportion to real disposable income. Let b represent the average regional marginal propensity to consume such that $b = C / RYD$, and $0 < b < 1$ where C is the total consumption component of final demand in the region. Further let consumption demand for industry i be some fixed proportion of total consumption, then,

$$C = b(RYD), \text{ and}$$
$$C_i = pce_i \cdot C, \text{ where } \sum_{i=1}^{n} pce_i = 1.$$

Substituting the first into the second gives,

$$C_i = pce_i \cdot b(RYD)$$

Dividing this by an equally specified national consumption equation, rearranging, and adding time subscripts give,

$$C_{it} = pce_i \cdot \gamma_C \frac{RYD_t}{RYD_t^U} C_t^U = pce_i \cdot \gamma_C \cdot RYD_t \cdot b_t^U, \tag{4.3}$$

where $\gamma_C = b/b^U$. The first relationship implies that relative regional real disposable income to the nation, RYD_t / RYD_t^U, and total national consumption C_t^U drives regional consumption of industry i. A more intuitive way of viewing this is found in the second relation that implies the regional propensity to consume follows the nation's propensity to consume, b_t^U, such that an increase in the national marginal propensity to consume or an increase in the region's real disposable income results in an increase in regional consumption of good *i*.

Regional investment demand of industry *i* is expanded into fixed residential investment and capital investment components since different economic drivers are assumed to drive the two components. First, total expenditures on fixed residential investment, *IR*, is assumed to be some proportion of total real disposable income, *RYD*,

$$IR = k_R \cdot RYD,$$

and fixed residential investment of industry *i* is assumed to account for a fixed proportion of total fixed residential investment,

$$IR_i = inv_{R_i} IR, \text{ where } \sum_{i=1}^{n} inv_{R_i} = 1.$$

Substituting the prior into the latter gives industry expenditures on fixed residential investment as a function of real disposable income,

$$IR_i = inv_{R_i} \cdot k_R \cdot RYD.$$

Defining the same for the nation, taking the ratio of the region to the nation and rearranging gives,

$$IR_i = inv_{IR_i} \cdot \gamma_{IR} \frac{RYD}{RYD^U} IR^U,$$

where $\gamma_{IR} = k_R / k_R^U$ and $inv_{R_i}^U IR^U$ is substituted for IR_i^U. Assuming that γ_{IR} and inv_{R_i} are invariant over time, and adding a time subscript gives,

$$IR_{it} = inv_{IR_i} \cdot \gamma_{IR} \frac{RYD_t}{RYD_t^U} IR_t^U, \tag{4.4a}$$

which implies that regional fixed residential investment expenditures on output from industry *i* are related to the relative real disposable income to the nation and the level of national fixed residential investment.

Regional expenditure on non-residential investment is assumed to be proportionately related to regional output, and industry expenditures on non-residential investment is assumed to be proportionately related to total regional non-residential investment such that,

$INR = k_{NR} \cdot GSP$, and

$INR_i = inv_{NR_i} INR$, where $\sum_{i=1}^{n} inv_{NR_i} = 1.$

Combining the two, defining the same for the nation, taking the ratio of regional to national, rearranging and adding a time subscript gives,

$$INR_{it} = inv_{NR_i} \cdot \gamma_{NR} \frac{GSP_t}{GDP_t} INR_t^U. \tag{4.4b}$$

The time invariant parameter γ_{NR} defines the benchmark observations of the ratio of regional to national proportion of output invested in productive capital and inventories, $\gamma_{NR} = k_{NR} / k_{NR}^U$. Equation (4.4b) implies that regional expenditures on industry i change as non-residential investment demand responds to changes in the relative level of output to the nation and the nation's level of total non-residential investment. If state output grows at the same rate as the nation, then the growth in regional non-residential investment grows at the rate of the nation.

Total regional investment is the sum of residential fixed investment and capital investment,

$$I_{it} = inv_{IR_i} \cdot \gamma_{IR} \frac{\hat{RYD}_t}{RYD_t^U} IR_t^U + inv_{NR_i} \cdot \gamma_{NR} \frac{\hat{GSP}_t}{GDP_t} INR_t^U. \tag{4.4}$$

An increase in relative personal income, value-added output, national fixed residential investment, or national expenditures on non-residential investment will reflect in increases regional investment.

The government component of final demand is expanded into state & local (*SLG*) and federal (*FED*) government components based on different drivers. Federal government final demands are not contingent on local economic activity, so are therefore driven solely by the nation. This regional final demand component is the sum of regional federal government industry demands.

Assuming some constant proportional relationship of industry demand to total regional federal government final demand, the federal government component of regional industry final demand is specified as,

$FED_i = fgov_i \cdot FED$, where $\sum_{i=1}^{n} fgov_i = 1$,

and *FED* is total federal government expenditures in the region in the benchmark year. Similarly defining national federal government industry expenditures, taking the ratio of the two and solving for the regional federal government component of industry final demand gives,

$$FED_i = \frac{fgov_i}{fgov_i^U} \frac{FED}{FED^U} FED_i^U, \text{ or}$$

$$FED_i = fgov_i \cdot \gamma_{FED} \cdot FED^U,$$

where γ_{FED} is the ratio of total regional to total national federal government final demands, $\gamma_{FED}=FED/FED^U$. The parameters $fgov_i$ and γ_{FED} are held constant at the benchmark-year observation. Adding time subscripts, the federal government component of industry *i* final demand gives,

$$FED_{it} = fgov_i \cdot \gamma_{FED} \cdot FED_t^U, \tag{4.5}$$

and the total regional federal government final demand is $FED_t = \sum_{t=1}^{n} FED_{it}$.

The total state & local government component of final demand is assumed driven by and proportional to the region's population. For the benchmark year, this gives,

$$SLG = k_{SLG} \cdot N,$$

where k_{SLG} relates total regional state & local expenditures to population, N. Further defining industry regional state & local government expenditures as some proportion of total regional state & local government expenditures gives,

$$SLG_i = gov_i \cdot SLG, \text{ where } \sum_{i=1}^{n} gov_i = 1.$$

Substituting the first into the latter and dividing by the same for the nation gives,

$$\frac{SLG_i}{SLG_i^U} = \frac{gov_i}{gov_i^U} \frac{k_{SLG}}{k_{SLG}^U} \frac{N}{N^U},$$

defining $\gamma_{SLG} = k_{SLG} / k_{SLG}^U$ and $SLG_i^U = gov_i^U \cdot SLG^U$, adding a time subscript, and rearranging gives,

$$SLG_{it} = gov_i \cdot \gamma_{SLG} \frac{N_t}{N_t^U} SLG_t^U, \tag{4.6}$$

where parameters gov_i and γ_{SLG} are held constant at the benchmark-year observation. Equation (4.6) implies that regional state & local government expenditures change with the region's relative population to the nation, and the nation's level of state & local government expenditures. Total regional state & local government expenditure is the sum of the industry expenditures, $SLG_t = \sum_{t=1}^{n} SLG_{it}$.

The export component of regional final demand is the sum of domestic and international exports. The superscripts N and W denote the region's exports to the nation and the world respectively. The region's industry i exports to the nation is specified as some function of the share of the region's contribution to total industry output for the nation and the relative costs to the nation as,

$$X_{it}^{N} = S_i^N \cdot RAC_i \left(\frac{1}{RAC_{it}} \right) \cdot VA_{it}^{U}, \text{ where } S_i^N = \left(\frac{X_i^N}{VA_i^U} \right), \tag{4.7a}$$

where the share parameter S_i^N is calculated from the benchmark-year regional and national IO tables and assumed invariant over time, and the ratio of industry comparative costs of production, RAC_{it}, relates comparative disadvantage to decreasing export demand. The benchmark-year RAC_i scales the equation to match the benchmark-year observation. Equation (4.7a) shows that if the selling price of domestically produced goods increases relative to the nation export demand from the region will decline. The cost functions are derived in the production block and described below.

International exports of industry i for the region are modeled as a constant fixed share of total national exports as follows,

$$X_{it}^{W} = S_i^W \cdot X_t^U, \text{ where } S_i^W = \left(\frac{X_i^W}{X^U} \right), \tag{4.7b}$$

where X^W denotes the region's exports to areas outside of the nation and X^U is the nation's exports. The share parameter S_i^W is calculated from the benchmark-year regional and national IO tables as the share of region's international exports to national exports and is invariant over time. Total industry exports from the region is expressed as the sum of the nation and world export final demands,

$$X_{it} = S_i^N \cdot RAC_i \left(\frac{1}{RAC_{it}} \right) \cdot VA_{it}^{U} + S_i^W \cdot X_t^U, \tag{4.7}$$

where the region's exports respond to changes in national economic activity and national exports and changes in the region's comparative production advantage.

All final demand components are designed to reproduce the benchmark-year observations for the benchmark year if the drivers themselves equal the benchmark-year values. During estimation the drivers may take on values different from the benchmark-year observations since all relationships must simultaneously hold. Therefore care is exercised in estimation to coax the benchmark-year solution toward the benchmark-year observations. That is the associative loss function, defined below, becomes more sensitive to missing the benchmark-year observations than it is on other years.

Returning to the regional accounts identity in employment terms, Equation (4.1), for any given year in which the complete IO accounts are available, this can be stated as an identity by definition. Estimating a complete history, based on a benchmark-year set of observations, will induce error over actual employment observations and requires estimating a historical relationship. Restating Equation (1) in a statistical form to be estimated as,

$$\hat{L}_{it} = \left(\phi_{FDi,1} \left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it} \right) rpc_{it} + \phi_{FDi,2} \cdot X_{it} \right) \lambda_{VA,it} \cdot epv_{it}, \qquad (4.8)$$

where the linear coefficients $\Phi_{FDi,k}$ map employment for local and export demands into total industry employment and character hats (^) denote systems estimated variables. Estimates of the final demand components (*INT*, *C*, *I SLG*, *FED*, and *X*) are described above. Aggregate employment is simply the sum of predicted industry employment or,

$$\hat{L}_t = \sum_{i \in N} \hat{L}_{it} . \qquad (4.9)$$

The historical values of *rpc*, λ_{VA}, and *epv* are systems estimated for optimal model fit as described below, while the linear coefficients, $\Phi_{FD,ik}$ are estimated based on the best fit of the equation to the historical series given *rpc*, λ_{VA}, and *epv*. Since the parameter coefficients are simultaneously estimated with the variable fit of the system, the parameter coefficients are also estimated for the optimal model fit.

For Equation (4.8) to track the historical series that it is designed to mimic requires some method of referencing the actual historic series. To impose this mimicking behavior a shrinkage-like relationship is imposed that shrinks the prediction values toward the actual observed values through the entropy function from defining the prediction error as,

$$L_{it} - \hat{L}_{it} = le_{it}, \tag{4.10}$$

where the character hat, $\hat{L}_{it}$ denotes the model predicted industry employment, and L_{it} denotes actual observations. The difference between actual and predicted industry employment is le_{it}. Estimates of le_{it} are provided in the entropy function where optimizing the entropy will result in shrinking le_{it} toward zero for all observations.

The multiplicands of Equation (4.8), *rpc*, *epv*, and λ_{VA} are estimated over time along with the parameter estimates of $\Phi_{FDi,1}$ and $\Phi_{FDi,2}$ maximizing the systems fit. In fact the complete systems fit is required to estimate Equation (4.8) since *rpc*, *epv*, and λ_{VA} are identified by the system itself for any given year. Similar to the static CGE model of Shoven and Whalley (1972; 1984), the structure of the model is instrumental in identifying the values within the model. Without the full structure of the model, Equation (4.8) can not be estimated.

Prior to model estimation, bounds and expected values must be specified for the coefficient support space. Expected values of the parameters $\Phi_{FDi,1}$ and $\Phi_{FDi,2}$ can be deduced from the structure of this Equation (4.8). All variables in Equation (4.8) can be identified for the benchmark year in which the identity-form of Equation (4.8), Equation (4.1) holds. For the benchmark year, the slope coefficients $\Phi_{FDi,1}$ and $\Phi_{FDi,2}$ equal unity. If the benchmark-year IO table is an accurate representation of the regional economy, the estimates of the regional final demand components are reasonable, and the relationships are stable over time, then coefficient estimates will be near unity. Otherwise the coefficient estimates will differ from unity. Assuming that Equation (4.8) is a sufficient representation of the regional economy, the best prior estimate of the slope coefficients should be centered at unity with tight bounds as shown in Table 3.1

The proportion of demand filled by local production responds to changes in relative selling prices of goods and services. If locally produced goods become more expensive relative to nationally produced goods then regional customers are likely to switch to the relatively less expensive nationally produced goods. Therefore the proportion of final and intermediate demands filled by regional output, or the industry regional purchase coefficient (rpc_i), responds to changes in relative selling prices of locally produced goods to those of the nation. Furthermore, since regional preferences for locally produced goods and services, and industry production change over time, a trend component is added to capture systematic changes in industry regional purchase coefficient over time. The regional purchase coefficient is modeled as,

$$rpc_{it} = rpc_i + \phi_{rpc,i,1} \cdot time_t + \phi_{rpc,i,2} \cdot AC_{it}, \tag{4.11}$$

where rpc_i is the benchmark-year observation, AC_{it} is the average industry cost of production relative nation described below, time is a unit counter by year, and the parameters $\Phi_{rpc,ii}$ are coefficients to be estimated.

The benchmark-year industry rpc_i is reported from the construction of the benchmark-year IO table and historical values of the *rpc* is systems estimated with Equation (4.8) to optimize the total fit of the model. Since industry regional purchase coefficients are not observable overtime, the linear coefficients $\boldsymbol{f}_{rpc,i,k}$ are simultaneously estimated along with the left-hand-side values of the *rpc* to maximize the overall model fit.

Economic theory is silent on the expected relationship of industry regional purchase coefficients to time, therefore a diffuse prior centered at zero is in order for $\Phi_{rpc,i,2}$. The law of demand guides the coefficient values for the relative average cost relationships. As local production cost increases relative to the nation, locally produced goods become less competitive and the demand for locally produced goods and services will tend to decline. Therefore, an inverse relationship is expected between the industry regional purchase coefficients and industry relative average costs. That is, *a priori* $\Phi_{rpc,i,1}$ is expected to be negative, and the prior probabilities and support set is shifted toward negative values as shown in Table 3.1.

The ratio of value-added output to total output by industry, $\lambda_{VA} = VA_{it} / Q_{it}$, is assumed to adjust gradually over time. This ratio relates the total proportion of total output attributable to value-added output excluding intermediate inputs. For modeling purposes, λ_{VA} is stated as a function of time as,

$$\lambda_{VA,it} = \lambda_{VA,i} + \phi_{VA,i,1} \cdot time_t + \phi_{VA,i,2} \cdot time_t^2, \tag{4.12}$$

where $time_t$ takes the value of zero on the benchmark-year restricting the value of λ_{VA} to the known benchmark value for the benchmark year.

The ratio of value-added output to total output, λ_{VA}, can only be identified within the employment Equation (4.8) that can only be estimated with a complete system fit as described below. Therefore prior expected values for the slope coefficients can not be ascertained by the data. Since neither data nor theory give insight to the ratio of value-added output to total output over time, diffuse priors should be used for the slope coefficients centered on zero with wide bounds. Similar to estimating industry regional purchase coefficients, the parameters $\Phi_{VA,i,k}$ are systems estimated such that values of λ_{VA} depend on the optimal fit of the model.

Industry labor productivity is captured in the variable *epv*, which measures the employment per dollar of value-added output, $epv_{it} = L_{it} / VA_{it}$. Productivity enhancements are captured by a decrease in this variable. While epv_{it} is well defined from this relationship and solving Equation (4.8) for epv_{it}, returns the same, a similar equation for *epv* comes from the production block and links the production side of the economy with the demand side discussed here.

Predictions of industry employment and employment per value-added output imply predictions for industry value-added output *VA*. Industry value-added output is the industry contribution to gross state product, *GSP* and is calculated as,

$$\hat{VA}_{it} = \frac{1}{epv_{it}} \cdot L_{it}. \tag{4.13}$$

Decreases in labor productivity; or rather increases in *epv,* while holding employment constant, results in decreases in *VA*. Total value-added output, or gross state product, *GSP*, is simply the sum of industry value-added output or,

$$VA_t = \sum_{i \in N} VA_{it}. \tag{4.14}$$

Increases in labor productivity or increases in labor, as defined by increases in final demand components, will lead to increases in industry value-added output and increases in total *GSP*.

BEA reports industry value-added output estimates and those reported values are used in guiding the model estimation. The difference between predictions of industry value-added output and those reported from the BEA are minimized subject to the complete model fit. Define the difference between predicted and actual values as,

$$VA_{it} - \hat{V}A_{it} = ve_{it}. \tag{4.15}$$

Similar to Equation (4.10), entropy estimation minimizes, or shrinks the values of ve_{it} toward zero.

PRODUCTION BLOCK

The production block assumes that factor demands are derived from cost minimization. The model assumes three factor inputs to production; labor (L_{it}), capital (K_{it}) and fuel (F_{it}) with prices W_{it}, r_{it}, and e_{it} respectively. Factor input prices are annualized rates per unit input. A Cobb-Douglas production function with constant returns to scale transforms factor inputs into value-added output. Furthermore, perfect capital markets are assumed such that factor costs are strictly linear in factor inputs. This gives the cost minimization problem as,

$$\min_{L_i, K_i, F_i} Cost_{it} = W_{it} L_{it} + r_{it} K_{it} + e_{it} F_{it}$$
$$s.t. \qquad VA_{it} = A_{it} L_{it}^{\alpha_i} K_{it}^{\beta_i} F_{it}^{\gamma_i},$$

where the first equation is the primitive objective function and the second is called the optimization constraint. For constant returns, the sum of the parameter exponents in the production function must sum to one, or $\alpha_i + \beta_i + \gamma_i = 1$. The variable A_{it} is a measure of total factor productivity. A higher value of total factor productivity implies greater output for any given level of input and is termed Hick's neutral because an increase in this term does not lead to a change in factor intensities.

Industry conditional factor demand functions for time t are derived from cost minimization as,

$$L_{it} = \frac{\alpha_i}{W_{it}} VA_{it},$$
$$K_{it} = \frac{\beta_i}{c_{it}} VA_{it}, \text{ and}$$
$$F_{it} = \frac{\gamma_i}{e_{it}} VA_{it}.$$

Substituting these values into the optimization constraint, and noting that the sum of the exponents equal unity gives,

$$VA_{it} = A_{it}\left(\frac{\alpha_i}{W_{it}}\right)^{\alpha_i}\left(\frac{\beta_i}{r_{it}}\right)^{\beta_i}\left(\frac{\gamma_i}{e_{it}}\right)^{\gamma_i} VA_{it}.$$

Further, solving the industry conditional labor demand function for value-added output, $VA_{it} = (W_{it}/\alpha_{it})\, L_{it}$ and substituting this for the right hand side VA_{it} gives,

$$VA_{it} = A_{it}\left(\frac{\alpha_i}{W_{it}}\right)^{\alpha_i}\left(\frac{\beta_i}{r_{it}}\right)^{\beta_i}\left(\frac{\gamma_i}{e_{it}}\right)^{\gamma_i} \frac{W_{it}}{\alpha_i} L_{it}, \text{ or by rearranging,}$$

$$\frac{L_{it}}{VA_{it}} = \frac{1}{A_{it}}\left(\frac{W_{it}}{\alpha_i}\right)^{\alpha_i - 1}\left(\frac{r_{it}}{\beta_i}\right)^{\beta_i}\left(\frac{e_{it}}{\gamma_i}\right)^{\gamma_i}.$$

Doing the same for the nation and dividing by the nation gives,

$$\frac{\dfrac{L_{it}}{VA_{it}}}{\dfrac{L_{it}^U}{VA_{it}^U}} = \frac{\dfrac{1}{A_{it}}\left(\dfrac{W_{it}}{\alpha_i}\right)^{\alpha_i-1}\left(\dfrac{r_{it}}{\beta_i}\right)^{\beta_i}\left(\dfrac{e_{it}}{\gamma_i}\right)^{\gamma_i}}{\dfrac{1}{A_{it}^U}\left(\dfrac{W_{it}^U}{\alpha_i^U}\right)^{\alpha_i^U-1}\left(\dfrac{r_{it}^U}{\beta_i^U}\right)^{\beta_i^U}\left(\dfrac{e_{it}^U}{\gamma_i^U}\right)^{\gamma_i^U}}.$$

By assuming equal factor cost shares for the nation and region ($\alpha_i = \alpha_i^U$, $\beta_i = \beta_i^U$, and $\gamma_i = \gamma_i^U$) the prior reduces to,

$$\frac{\dfrac{L_{it}}{VA_{it}}}{\dfrac{L_{it}^U}{VA_{it}^U}} = \frac{epv_{it}}{epv_{it}^U} = \frac{A_{it}^U}{A_{it}}\left(\frac{W_{it}}{W_{it}^U}\right)^{\alpha_i-1}\left(\frac{r_{it}}{r_{it}^U}\right)^{\beta_i}\left(\frac{e_{it}}{e_{it}^U}\right)^{\gamma_i}, \text{ or}$$

$$epv_{it} = RF_{it}^{-1} \cdot RLC_{it}^{\alpha_i-1} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i} \cdot epv_{it}^U, \tag{4.16}$$

where $RF_{it} = A_{it} / A_{it}^U$ is the industry relative total factor productivity at time *t*, and RLC_{it}, RCC_{it}, and RFC_{it} are relative labor, capital and fuel costs respectively. If total factor productivity increases locally relative to the nation (RF_{it} increases), then local production will require fewer workers and epv_{it} will decrease given the national productivity measure, epv_{it}^U.

Cost minimization implies that as the cost of one input increases compared to the costs of others, the firm will substitute other inputs in place of the now comparatively more expensive input. Likewise, if the relative cost of one input increases compared to the nation, the region will substitute other inputs for the now comparatively more expensive input. Therefore cost minimization implies that an increase in labor cost compared to the nation will induce firms to substitute capital and energy for labor, thereby decreasing epv_{it}. Increases in the relative costs of other factor inputs, *RCC*, and *RFC*, will induce firms to shift to the relatively less expensive labor, increasing epv_{it}.

Solving the cost minimization derivation of industry *epv*, Equation (4.16), for *RF* gives,

$$RF_{it} = \frac{epv_{it}^{U}}{epv_{it}} RLC_{it}^{\alpha-1} \cdot RCC_{it}^{\beta} \cdot RFC_{it}^{\gamma}. \tag{4.17}$$

Predictions of RF_{it} maintain all the properties of cost minimization discussed above. An increase in employment cost, RLC_{it}, or required employment per value-added output, VA_{it}, will reduce relative total factor productivity, RF_{it}.

The Relative cost functions, RLC, RCC, and RFC represent the regional cost of the respective input compared to the nation. For example, relative labor cost is modeled simply as the ratio of regional wage rates to the nation,

$$RLC_{it} = {W_{it}}/{W_{it}^{U}}, \tag{4.18}$$

where W_{it} and W_{it}^{U} are average industry annual wage rates for the region and nation respectively. In estimation, it is assumed that both *RCC* and *RFC* are constant and equal to unity. The latter two are free to vary with policy analysis.

Industry average cost, AC_i, similarly relies on the assumption of regional cost minimization. The industry cost function, which assumes cost minimizing combination of inputs, is used to measure the regional cost of production. Substituting the conditional factor demands into the primal cost function, noting that constant returns to scale imply that the sum of the share parameters equal unity, and simplifying gives,

$$Cost_{it} = \frac{1}{A_{it}^{-1}} \left(\frac{W_{it}}{\alpha_i} \right)^{\alpha_i} \left(\frac{r_{it}}{\beta_i} \right)^{\beta_i} \left(\frac{e_{it}}{\gamma_i} \right)^{\gamma_i} VA_{it}.$$

Dividing both sides of the cost function by *VA* to get cost per dollar of value-added output, or the average cost of production in industry *i*, gives,

$$AC_{it} = \frac{Cost_{it}}{VA_{it}} = \frac{1}{A_{it}}\left(\frac{W_{it}}{\alpha_i}\right)^{\alpha_i}\left(\frac{r_{it}}{\beta_i}\right)^{\beta_i}\left(\frac{e_{it}}{\gamma_i}\right)^{\gamma_i}.$$

AC is non-decreasing in factor prices *W*, *r*, and *e*, and decreasing in total factor productivity.

Relative industry average cost to the nation, *RAC*, is specified by calculating the industry AC function for the nation and taking the ratio of the region to the nation to get,

$$\frac{AC_{it}}{AC_{it}^U} = \frac{\frac{Cost_{it}}{VA_{it}}}{\frac{Cost_{it}^U}{VA_{it}^U}} = \frac{\frac{1}{A_{it}}\left(\frac{W_{it}}{\alpha_i}\right)^{\alpha_i}\left(\frac{r_{it}}{\beta_i}\right)^{\beta_i}\left(\frac{e_{it}}{\gamma_i}\right)^{\gamma_i}}{\frac{1}{A_{it}^U}\left(\frac{W_{it}^U}{\alpha_i^U}\right)^{\alpha_i^U}\left(\frac{r_{it}^U}{\beta_i^U}\right)^{\beta_i^U}\left(\frac{e_{it}^U}{\gamma_i^U}\right)^{\gamma_i^U}}.$$

Continuing the assumption that national and regional shares are equal and simplifying gives,

$$RAC_{it} = \frac{A_{it}^U}{A_{it}}\left(\frac{W_{it}}{W_{it}^U}\right)^{\alpha_i}\left(\frac{r_{it}}{r_{it}^U}\right)^{\beta_i}\left(\frac{e_{it}}{e_{it}^U}\right)^{\gamma_i},$$

or by substitution,

$$RAC_{it} = \left(1/RF_{it}\right)\cdot RLC_{it}^{\alpha_i}\cdot RCC_{it}^{\beta_i}\cdot RFC_{it}^{\gamma_i}. \tag{4.19}$$

RAC is non-decreasing in relative factor prices *RLC*, *RCC*, and *RFC*, and increases with increases in relative total factor productivity *RF*.

Calculations of relative industry labor costs rely on estimates of industry wage rates. Estimates of the wage rate rely on the local labor market conditions, the local price index, national wage rate and local amenities. A lagged endogenous relationship is captured with a lag term such that the industry wage rate is described as,

$$W_{it} = f\left(REO_t, CP_t, W_{it}^U, Amenities_t, W_{it-1}\right)$$

Since measures of local amenities are not directly observable, regional amenity effects are captured by the intercept term, which is allowed to change with time. The estimation equation for industry wage rate is specified as,

$$W_{it} = \phi_{Wi} + \phi_{Wi,REO} REO_t + \phi_{Wi,CP} CP_t + \phi_{Wi,W^U} W_{it}^U + \phi_{Wi,time} time_t + \phi_{Wi,lag} W_{it-1}. \tag{4.20}$$

Relative employment opportunity, *REO,* measures the nature of employment opportunity in the region relative to that of the nation as described below. An increasing *REO* indicates that employment opportunities in the region are growing faster than that of the nation. The source of increased employment opportunities can come from two offsetting changes; an increase in demand for workers and a decrease in the supply of workers. If an industry labor market is historically characterized by changes in demand for workers, *REO* will enter the equation positively. Otherwise, supply changes will enter the wage equation negatively. Therefore it is not possible to sign the *REO* coefficient $\Phi_{Wi,REO}$ a priori, and a diffuse prior about zero is used.

Local industry wage rates should respond positively to changes in wage rates at the nation, therefore Φ_{Wi,W^U} should be in the neighborhood of unity with wide positive bounds. Furthermore, regional wage rates should poses a positive autocorrelation with past values of itself but not be explosive such that $\Phi_{Wi,lag}$ should be in the range of $0 \leq \Phi_{Wi,lag} < 1$.

Variations of wage rates across regions should reflect, among other things, differences in local amenities (Beeson and Eberts 1989; Hoehn *et al.* 1987; Roback 1982). Residents in amenity rich regions are willing to forego higher wages offered in other regions to retain access to the benefits of living in an amenity rich region. Regional amenity and productivity effects on regional industry wage rates are modeled as,

$\Phi_{Wi}+\Phi_{Wi,time}\cdot time_t$, where the inclusion of the time variable allows for systematic changes over time.

Since actual values of W_{it} are known over histories, comparisons of the model predictions to actuals are instrumental in judging the fit of the model. To assure that the wage relationships actually track history, a shrinkage estimator is added that minimizes the historical difference between the model prediction and the actual observations as,

$$W_{it} - \hat{W}_{it} = we_{it}. \tag{4.21}$$

Similar to Equations (4.10) and (4.13) the difference between the observed and the predicted values, we_{it}, are minimized through the entropy function such that the predicted values of wage rates track historical values known to exist.

Wages are further subject to the consistency constraint that the regional aggregate wage rate, or overall wage rate, is a weighted average of industry wage rates weighted by the proportion of employment by industry. The overall wage rate is calculated as,

$$W_t = \sum_{i=1}^{N} W_{it} L_{it} \Big/ L_t \, . \tag{4.22}$$

Wages make up a sizeable proportion of total regional income. Equation (4.22) ensures that aggregate wage & salary disbursement (WSD_t) is simply the sum of industry disbursements (WSD_{it}). Both are described below in the income block.

INCOME BLOCK

Personal income is calculated by BEA component. The fundamental component of personal income is wage & salary disbursements. Total wage & salary disbursement (WSD) is obtained as the sum of the product of industry wage rates and employment. The remaining personal income components are calculated based on the corresponding national component, economic activity, scaled to the benchmark-year component value of personal income. These personal income components for estimation include property income; transfer payments; social insurance payments; other labor income; and taxes.

A region-specific deflator transforms nominal total personal income into real regional income. This index is calculated as a function of the national consumer price index and the relative regional to national cost of production.

To begin, personal income is the sum of wage & salary disbursements, proprietor's income, other labor income, and property type income less contributions to social insurance,

$$YP_t = WSD_t + YOL_t + YPROP_t - TWPER_t^{8}. \tag{4.23}$$

The variable YP_t is personal income, WSD_t is non-farm wage & salary disbursements, YOL_t is other labor income and farm and non-farm proprietary income, $YPROP_t$ is dividends, interest, and income from real-estate, and $TWPER_t$, is contributions to social insurance. All components of personal income are measured in thousands of current dollars.

Wage & salary disbursement, *WSD*, is measured in thousands and is the sum of industry wage & salary disbursements,

$$WSD_t = {W_t \cdot L_t}/{1{,}000}. \tag{4.24}$$

Industry wage & salary disbursement is computed as the product of the industry wage rate and employment as,

$$WSD_{it} = {W_{it} \cdot L_{it}}/{1{,}000}, \tag{4.25}$$

[8] Residential adjustment for place of work is abstracted away from the model formulation as its proportion to total state employment is miniscule. Smaller regional specifications will warrant full attention to residential adjustments.

where W_{it} is the average wage rate and L_{it} is labor in industry i at time t.

Projections of the remaining components of personal income are modeled with regional and national drivers as,

$$Y_t = \left(\frac{Y/X}{Y^U/X^U} \right) \left(X_t / X_t^U \right) \cdot Y_t^U,$$

where Y_t is the personal income component to be estimated and X_t is the appropriate datum for that component at time t. The absence of a t subscript implies benchmark-year observations, and the U superscript denotes national values. The first multiplicative relationship in parenthesis is a scaling function that scales the national and regional driver to the benchmark-year value of the personal income component Y. This formulation assures that benchmark-year values are returned for benchmark years for drivers that similarly replicate benchmark-year observations. The relationship in the second parenthesis forms the response to regional specific economic activity relative to the nation. If the regional economic activity grows faster than the nation, the region's income component, Y_t will grow faster than the nation's Y_t^U. The final multiplicand is simply the national driver. Holding all else constant, the regional income component should grow at the pace of the nation. For completeness, the remaining personal income components, *YOL*, *YPROP*, and *TWPER*, along with *TAX* and *TRAN* are presented.

Other labor income, *YOL*, is driven by national other labor income and relative employment as,

$$YOL_t = \frac{YOL/L}{YOL^U/L^U} \cdot \left(L_t / L_t^U \right) \cdot YOL_t^U \,. \tag{4.26}$$

Property-type income *YPROP*, is driven its national counterpart and by population as,

$$YPROP_t = \frac{YPROP / N}{YPROP^U / N^U} \cdot \left(N_t / N_t^U \right) \cdot YPROP_t^U . \tag{4.27}$$

Finally, contributions to social insurance *TWPER*, is driven by the national *TWPER* and *WSD*, as,

$$TWPER_t = \frac{TWPER / WSD}{TWPER^U / WSD^U} \cdot \left(WSD_t / WSD_t^U \right) \cdot TWPER_t^U . \tag{4.28}$$

The proceeding projection and estimation equations give all the components necessary to compute regional personal income through Equation (4.23).

To model actual spending income requires netting out taxes from personal income to get disposable personal income, or net personal income. Treating income transfers as a negative tax, disposable income is defined as,

$$YD_t = YP_t - NTAX_t, \tag{4.29}$$

where NTAX is net taxes and is defined as tax payments and non-tax payments to state, local, and federal government less government transfers to businesses and households,

$$NTAX_t = TAX_t - TRAN_t. \tag{4.30}$$

The BEA measure of tax payments is the personal tax and non-tax payments of the PSI series 50 and includes taxes on income, transfers, and personal property and non-tax payments of donations, fees, fines and forfeitures. Projections of local area tax payments are structured like projections on personal income components with wage & salary disbursement as the regional driver as,

$$TAX_t = \frac{TAX/WSD}{TAX^U/WSD^U} \cdot \left(WSD_t / WSD_t^U \right) \cdot TAX_t^U . \tag{4.31}$$

Transfer payments are payments made to persons for which no services are performed. They include payments by governments and businesses to individual and nonprofit institutions, and are reported in the PSI series 50. State population drives projections of transfer payments,

$$TRAN_t = \frac{TRAN/N}{TRAN^U/N^U} \cdot \left(N_t / N_t^U \right) \cdot TRAN_t^U . \tag{4.32}$$

Holding all else constant, both tax and transfer payments move in step with the national tax and transfer payments respectively. If regional conditions change relative to the nation, tax and transfer payments will respond accordingly.

Real disposable income is disposable income adjusted by the regional specific cost of living index as,

$$RYD_t = YD_t / cpi_t , \tag{4.33}$$

where cpi_t is a regional specific price index that differs from the nation to the extent that regional production cost differ.

The local cost of living index, cpi_{it} is calculated as the weighted average of relative local selling price of production to the US, weighted by the distribution of local final demands,

$$cpi_t = \sum_{i=1}^{N} \left(FD_i / FD \cdot Sp_{it} \right) \cdot cpi_t^U . \tag{4.34}$$

The weights, FD_i / FD are the proportion of total final demand component of total output for industry i calculated in the benchmark year. The selling price of regionally produced goods and services, Sp_{it} reflects the costs of regional production relative to the nation.

The industry relative selling price Sp_i is assumed to directly reflect costs of production such that an increase in relative average costs, RAC_{it}, will result in an increase in the relative selling price. Since the relative selling price is a ratio, it is centered on unity. The relative selling price of import competing goods and services are assumed to equal the national price since locally produced goods must compete at a national level. The relative selling price of non-goods producing industries are assumed to reflect the comparative cost advantage or disadvantage the region has over the nation, such that,

$$Sp_{it} = \begin{cases} 1 & \text{if } i = \text{goods producing} \\ RAC_{it} & \text{if } i = \text{nongoods producing} \end{cases}. \tag{4.35}$$

This structure reflects that local producers producing transportable goods are strictly price takers in the national economy, and that local producers of non-transportable goods and services do have some degree of market power. Finally, the ratio of domestic price index to the national consumer price index is calculated as,

$$CP_t = {cpi_t}/{cpi_t^U}, \tag{4.36}$$

which is centered on one. It measures the extent to which the regional price index changes relative to that of the nation.

POPULATION BLOCK

The population block follows closely to the migration assumption found in Harris and Todaro (1970; Todaro 1969) generalized with (Plaut 1981). Harris and Todaro limit migration motives to economic gains. Plaut expands upon the Harris-Todaro model by recognizing that relative amenities also influence migration decisions. Accepting Plaut's modification, local migration responds to changes in the relative labor markets and changes in relative amenities. If local market conditions favor employees through increasing employment opportunities or through increasing wages, then economic migration into the region should increase. Furthermore, if local amenities are viewed favorably relative to amenities offered in other locations, then migrants seeking better living environments will migrate to this area.

Purely economic migrants respond to changes in the labor market conditions. The Harris-Todaro interpretation of economic migration predicts that economic migrants will relocate to regions that offer greater expected income, where expected income is the expected value of the aggregate wage rate times the proportion of the population actually employed. Relative wage rate is defined as the relative aggregate wage rate,

$$RWR_t = {W_t}/{W_t^U}, \tag{4.37}$$

and relative employment opportunity is substituted by the relative employment rate to the nation, as,

$$REO_t = \frac{L_t / N_t}{L_t^U / N_t^U}. \tag{4.38}$$

The latter reflects the relative probability of finding employment in the region. Expected relative earning is defined as *RWR·REO*. The assumption is that migration continues as long as expected relative earnings is greater than one (Harris and Todaro 1970; Todaro 1969).

Rather than modeling each specific amenity variable's effect on migration, amenity effects are captured as residual determinants of migration. That is migration is driven by purely economic factors such as the relative employment rates and wage rates, and driven by non-economic factors,

$$N_t = \phi_{N0} + \phi_{N,REO} \cdot REO_t + \phi_{N,RWR} \cdot RWR_t + \phi_{N,time} \cdot time_t + \phi_{N,Nlag} \cdot N_{t-1}. \tag{4.39}$$

The effect of expected relative earnings on regional migration is estimated as $\Phi_{N,REO} \cdot REO + \Phi_{N,RWR} \cdot RWR$ where the coefficients reflect the relative contribution to the migration decision of changes in the respective employment opportunities. The remaining terms capture hysteric and amenity affects. The combined intercept and time response, ($\Phi_{N0} + \Phi_{N,time} \cdot time$), allows amenity effects to change over time while the lag response captures inertia in migration.

Model predictions of population should mimic historical observations of population. The difference between the actual observation and model's predicted value is,

$$N_t - \hat{N}_t = ne_t, \tag{4.40}$$

where the absolute value of net is minimized in estimation. Optimizing model fit partially entails minimizing the values of ne_t.

HISTORIC FIT

Four equations form the shrinkage relationships that reduce the difference between observed values and actual values observed in the estimation period. Aside from the forty equations that describe the model's relationships and prediction errors, four additional equations are added that define the error terms in Equations (4.10), (4.15), (4.21), and (4.39) as entropy error terms,

$$le_{it} = \sum_{j=1}^{J} v_{Lit,j} w_{Lit,j}, \tag{4.41}$$

$$ve_{it} = \sum_{j=1}^{J} v_{VAit,j} w_{VAit,j}, \tag{4.42}$$

$$we_{it} = \sum_{j=1}^{J} v_{Wit,j} w_{Wit,j}, \text{ and} \tag{4.43}$$

$$ne_{t} = \sum_{j=1}^{J} v_{Nt,j} w_{Nt,j}. \tag{4.44}$$

As described in Chapter III the vector of support sets **w** in the error equations are centered about zero. With the prior support probabilities also centered about zero, the entropy function is optimized when the error is as close to zero as possible for all observations while allowing the remaining forty sets of equations to hold. Stacking the vectors le_{it}, ve_{it}, we_{it}, and ne_t over all i and t forms the error vector **e** in Equation (4.16) of Chapter III.

DATA ISSUES

Data for estimation and forecasting comes from a wide range of sources. Industrial sector data, including value-added output, and employment, along with income components, and tax & transfer payments, originates with the Bureau of Economic Analysis (BEA). Regional income components are derived by the CA05 series of the Regional Economic Information System (REIS) CD-ROM in SIC industries. SIC industry employment is obtained by the CA25 series of the REIS CD-ROM and is aggregated to the model's aggregation scheme. And state population is derived from series CA1-3. Wage rates are calculated from series CA06 as total industry employee compensation normalized to wage & salary disbursement and divided by industry employment giving an annualized rate.

Estimates of interindustry linkages are derived from the 1997 Census of Business. The Census of Business is conducted every five years and produces a detailed survey concerning the purchases of intermediate goods, payments to factors (labor, capital, land and entrepreneurship), and taxes of companies in the United States. The Census of Business does not report regional relationships but rather national statistics. Adjustments to the national survey data are required to apply the Census of Business to any particular region. The Minnesota IMPLAN Group Inc. Impact Analysis for Planning (IMPLAN) regionalizes data for the state of Oklahoma. The IMPLAN program uses recent state-level employment data to scale national-level industrial data down to the size of a state for policy analysis. Though the IMPLAN program is a policy analysis tool in its own, it is merely the starting point for the model described above.

The IMPLAN data set comprises the full-regionalized IO and social accounting matrices used in static IO models of policy analysis. Production relationships are all derived from IMPLAN's SAM matrix for Oklahoma. Factor shares, intermediate demand linkages, total output, total value-added output, and intermediate inputs are revealed in the IMPLAN data set. Demand components for the region are also represented including all intermediate and final demand components. Housed within the IMPLAN databank are labor values, factor incomes, household income distribution, and price indices in high detail. Furthermore, a comparable data set accompanies the Oklahoma data for the nation, in which national relationships are derived.

Since IMPLAN data is single year and relies on the five-year interval of the Census of Business survey, inter-census years are extrapolated from employment figures produced from the BEA. The last such survey was 1997. The 1999 IMPLAN data set is used in the accompanying model. All benchmark-year observations are made from this benchmark year. Time-series histories of relevant state employment, value-added output, income, and taxes are reported in REIS CD-ROM of the BEA. Since these sources are subject to revisions, the actual BEA reported values differ from those reported from the IMPLAN program. All relevant series are normalized to those values found in IMPLAN for consistency.

Employment histories are aggregated into their respective industries. These values are checked against IMPLAN's reported aggregated industry employment for the benchmark year, then normalized such that the benchmark-year equates with the IMPLAN value. Total employment is then calculated as the aggregate of industry employment.

The model uses three factor inputs, capital/land, labor, and energy. IMPLAN does not treat energy as a factor input. To adopt energy as a factor input, a separate IMPLAN data set is created isolating energy as a single industry. Taking energy out of the industry-by-industry IO table, it is then added as a component of factor payments. Since the row entries denote expenditures per dollar of sales, the proportional relationship is retained and the measure of energy input is comparable to that of capital and labor. To allocate the energy sector demands for other industry output, industry input into energy, is factored into the new energy factor row proportional to the industry purchase of energy inputs. This results in total energy expenditures by industry that also includes the energy sector's expenses on other sector inputs.

Constant returns to scale in the production process are assumed and a Cobb-Douglas production function is used. Therefore factor share coefficients, α_i, β_i, and γ_i are simply calculated as the proportion of total value-added paid to the respective factor in the benchmark year. Factor shares are assumed to remain constant over time. A similar set of national factor shares are derived. The accompanying forecasting model assumes identical production processes for Oklahoma and the U.S. so the average of the two are used with the restriction that they sum to unity for constant returns to scale.

All reported historical industry values are aggregated into their respective aggregate industry class. Values of the benchmark-year observations are compared to IMPLAN's, then normalized to equate with that value. From these histories, initial-valued series of industry total output can be derived from the benchmark ratio of value-added output to total output. From which, the difference between industry output and value-added output gives initial values of intermediate inputs over time. Since the ratio of value-added output to total output is subject to change over time, these initial values are not the final estimates, but they do serve as valid starting points when estimating the full system. A similar process is applied to national data.

The regional purchase coefficient is derived from the IMPLAN data set for the benchmark year. Initial-valued observations of the regional purchase coefficients over time are calculated by solving Equation (11) for rpc_{it} assuming a constant benchmark-year value of λ_{VA} and calculating epv_{it} as $epv_{it} = L_{it}/VA_{it}$. These initial values of rpc_{it} are the starting values used in optimizing the model.

Starting values of relative labor costs, *RLC*, wage rates, *RWR*, and employment opportunity, *REO*, are calculated from known observations reported by the BEA. From which, starting values of relative total factor productivity, *RF*, relative average cost of production, *epv*, and the full set of price and relative price index, *CP*, can be derived.

For the production block equations, initial values of RF_{it} from observable time-series are used to calculate initial values of RAC_{it}, which are then used to calculate initial values of Sp_{it}, CP_t, and cp_{it}. These initial historical values are then used as starting points to the optimization problem over all time periods *t*.

Though not conforming to the full system fit, the starting values are all derived from the structure of the model, and therefore should be in the neighborhood of the full model solution. Therefore, the starting values are instrumental in placing upper and lower bounds on the feasible solution sets by year. Adding and subtracting some multiple of the starting value gives the upper and lower bounds respectively for the solution space. In estimation, care must be given that the bounds are not part of the solution. That is, model solutions on the bounds indicate that either the system is non-stable or that the bounds are too tight for a proper solve. Efforts to solve the system show that some bounds required widening while others required tightening to facilitate a solution.

Along with model endogenous variables, parameter estimates, and error predictions require upper and lower bounds and initial starting values. These are defined by the three-point support sets and prior probabilities. Table 3.1 presents these three-point support sets and the prior expected probabilities that define the prior values of the systems estimated parameters. The prior expected value of the parameter is found by the dot product of the 3×1 support set and the 3×1 probability set for a weighted average of the support, weighted by the mutually exclusive probabilities of the support point being the true value. For symmetric support points and weights, this value is the second or middle support point.

Error terms are also estimated to record and minimize the difference between model predictions and known values of those variables. Unlike static general equilibrium models that require all equations fit with certainty, parameter values are identified by the single year fit of the equations (Shoven and Whalley 1972, 1984), the accompanying model requires single parameter estimates for all years in the sample period to hold. Such dynamic systems, at best, overidentify the parameters. Parameters that fit the model exactly for one year's data may not necessarily fit a second year's data leading to the necessity of adding an error term to one or more of the equations. In the current application, more than one error term is added to the model to facilitate the overall fit of the model.

ESTIMATION

The 28 system coefficients described above and the corresponding error terms are systems estimated with cross-entropy described in Chapter III. The modeling environment used is the GAMS IDE software version 2.0.26.8, Build VIS 21.3 138 (Brooke *et al.* 1998) , while optimization is solved numerically through the MINOS 5.51 system (Murtagh and Saunders 1987; Murtagh *et al.* 2002).

The MINOS package is a highly useful optimization package based on the MINOS algorithm for medium to large-scale optimization problems with or without equality or inequality constraints. It is also a very general and flexible environment for solving linear or nonlinear optimization problems subject to linear or nonlinear constraints. The problem presented in Equation (3.11) of Chapter III is to maximize a nonlinear objective function subject to the nonlinear constraints that make up the structure of the model. For solving such nonlinear problems in the objective function and constraints, the MINOS package implements a *Projected Augmented Lagrangian Algorithm*. Augmented Lagrangian algorithms are based on successive optimizations of the augmented Lagrangian function subject to the linearly approximated constraints. MINOS augments the Lagrangian function with the addition of a quadratic penalty function that measures the square of the differences between the linearly approximated values of the constraints and the primitive values of the constraints (Murtagh *et al.* 2002; Nocedal and Wright 1999, pp. 524). In its simplest description, this algorithm linearly approximates the nonlinear constraints at some initial value and optimizes the augmented objective function subject to those linearly approximated constraints as a subproblem. Once the subproblem is sufficiently solved with a standard reduced-gradient algorithm, a new subproblem is optimized by linearly approximating the constraints at the previous solved values. The procedure is iterated until the difference in successive subproblem solves is sufficiently small.

Because the problem presented here constitutes a nonlinear objective function with a set of nonlinear equality constraints, assessing whether a global or a local optimum has been reached is problematic. The general procedure for assessing whether a local optimum has been reached is to alter the starting values and resolving such that the iterative procedure moves the solution over a different region of the solution space. Though no systematic method is employed to evaluate the sensitivity of the solution to various starting values, several alternative bounds and starting values are attempted throughout the estimation process and results are compared. Generally, parameter estimates are robust over several attempts. Though parameter estimates are robust over several starting values, the ability of the MINOS program to find a feasible solution is not.

There is no guarantee that MINOS will be able to find a global optimal that satisfies the full set of constraints. Experiments with the current model found that most starting values and systems of bounds and constraints will either reach a consistent solution or fail to converge. It is further found that changing the solver parameters such as step-size and penalty parameters, alters the speed of convergence but not the optimal solution.

As with any complex optimization packages, there are several parameters that alter the way the solver progresses. At the core of the *Projected Augmented Lagrangian Algorithm* for nonlinear constraints, is the penalty parameter that adjusts the importance of the difference of the linearized and primal constraint values on the augmented Lagrangian. Though parameter estimates are robust to the specified penalty parameter, the ability of the MINOS algorithm to attain a feasible solution and the speed of convergence is not. Locating a feasible solution is contingent upon properly specifying a penalty parameter. In application, it is found that only through trial and error can an appropriate penalty parameter be found that will not cause MINOS to fail to reach a feasible optimum.

Further experimentation revealed that the penalty parameter and other solver parameters interact such that finding a successful combination of solver parameters for the MINOS program to reach a feasible optimal solution required a great deal of trial and error. Altering prior parameter values, starting values, and/or bounds often required altering the solver parameters for a feasible solution without significantly changing the final solution.

Aside from experimenting with solver parameters, the process of formulating a solvable model required experimentation with the inclusion and exclusion of slack terms within the model structure. The general attempt is to minimize the inclusion of slack terms in the model while allowing the solver to attain a feasible solution. Inclusion of too many slack terms left variables indeterminate, while inclusion of too few left the model equations too rigid to solve. Experiments favored adding a slack variable to the equations for industry rpc_{it}, Equation (4.11), allowing the constraints to relax rigidities in the model. Without these slack variables in place, a solution is not feasible. Adding slack variables to the income block did not have an effect on the ability of the model to solve.

Further rigidities across few equations posed an unexpected problem to the ability of the MINOS solver to reach an optimal solution. The ability of the MINOS solver to find an optimal solution given the four equations for industry and aggregate wage rates and wage & salary disbursements Equations (4.20, 4.21, 4.24, and 4.25) was somewhat nonsensical and related to the rigidity of these four equations at initial iterations of the solver. That is if the solver is able to get through the initial hurdle of finding an initial interim solution to these six equations the solver is able to complete the solution given that no other infeasible or nonoptimal constraints in the solution exists.

More interesting is that the form of Equation (4.24) makes a difference in the ability of the solver to find a solution but has no bearing on the value of the solution itself. Equation (4.24) is equivalently stated as,

$$WSD_t = {W_t \cdot L_t}\big/{1000} = \sum_{i=1}^{N} WSD_{it}.$$

Proof: since $W_t = \Sigma\, W_{it} \cdot L_{it} / L_t$ and $WSD_{it} = W_{it} \cdot L_{it} /\ 1000$, substituting W_t into the second term and substituting WSD_{it} into the last gives,

$$WSD_t = \frac{\sum_{i=1}^{N} W_{it} \cdot L_{it}}{1000} = \frac{\sum_{i=1}^{N} W_{it} \cdot L_{it}}{1000},$$

which verifies that the two are equivalent statements. Stating Equation (4.24) as $WSD_t = \Sigma WSD_{it}$ creates a more formidable hurdle for MINOS to surmount than stating it as $WSD_t = W_t \cdot L_t / 1000$. Nonetheless, those solutions that were attainable with this second formulation were not different from those from Equation (4.24). In some instances, a feasible solution required that Equation (4.24) be specified as the sum of the industry wage & salary disbursements. Other instances required a slack variable be added to aggregate *WSD* while others did not. The final solution required both specifications to exist coincidently for a feasible solution. My thoughts are that I might as well err on the side of over-specifying than under-specifying the relationships.

The four equations specifying *WSD* implies the solution to industry and aggregate employment. This was found to create a source of rigidity in the system that could not be alleviated without the specification of a slack variable, defined over the range of ± 800 and taking a value of zero at the optimal solution, onto Equation (4.9) for aggregate employment. The difficulty arises because aggregate employment is defined within Equations (4.20, 4.21, 4.24, and 4.25). Proof: From before, define,

$$WSD_t = {W_t \cdot L_t}\big/{1000} = \sum_{i=1}^{N} WSD_{it}.$$

Defining $L_t = \Sigma\, L_{it}$ and dropping the first equality gives,

$$W_t \cdot \frac{\left(\sum_{i=1}^{N} L_{it}\right)}{1000} = \sum_{i=1}^{N} WSD_{it}\,.$$

Since $W_t = \Sigma\, W_{it}{\cdot}L_{it}/L_t$, and $WSD_{it} = W_{it}{\cdot}L_{it}/1000$, substituting these for W_t and WSD_{it}, gives,

$$\sum_{i=1}^{N} \frac{W_{it} \cdot L_{it}}{L_t} \cdot \frac{\left(\sum_{i=1}^{N} L_{it}\right)}{1000} = \sum_{i=1}^{N} \frac{W_{it} \cdot L_{it}}{1000},$$

divide by $(\Sigma\, W_{it}{\cdot}L_{it} / L_t) / 1000$ to get

$$\sum_{i=1}^{N} L_{it} = L_t,$$

completing the proof. Technically, the aggregate employment Equation (4.9) is redundant in that it is fully defined elsewhere. But a feasible solution could never be attained from eliminating Equation (4.9).

Further sensitivity issues arise with equal specifications of single equations. Equations with endogenous denominator terms may hinder MINOS from finding a feasible solution. Shifting the endogenous denominators to the left hand side of the equation by multiplication can often result in a feasible solution. For example, industry relative average cost is defined as,

$$RAC_{it} = \left(1\big/_{RF_{it}} \right) \cdot RLC_{it}^{\alpha_i} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i}.$$

Restating this as,

$$RAC_{it} \cdot RF_{it} = RLC_{it}^{\alpha_i} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i},$$

will often allow the MINOS program to find an optimal, feasible solution when it could not with the prior specification. The problem is not related to a division-by-zero problem since RF_{it} is bounded sufficiently far from zero. Several such relationships exist and much experimentation was done to locate the optimum combinations of specifications, slack variables, boundaries, starting values, support sets, and solver parameters that gave robust and consistent feasible solutions. As unsettling as this is, some experiments showed that the latter specification above hampers a feasible solution for some equations when the prior does not. There just does not seem to be any rhyme or reason to what hinders and what facilitates a solution. It often changes for the same equation under different bounds, other equation specifications, slack variable locations, and solver parameters.

Table 4.1 displays the support sets, prior probabilities, implicit prior means and variances, and the system estimates of the 29-parameter coefficients. The support sets are initiated from theory or least-squares estimation and adjusted through experimentation. From the combination of solution attempts it was found that widening some support sets while contracting others were instrumental in attaining a solution. Multiplying the support set with the corresponding prior probabilities (**q** in Chapter III) gives the implicit prior values of the estimates. The cross-entropy objective function will shrink estimates of the parameter coefficients to these values to the extent that the data allows. Implicit prior standard deviations (Std. Dev.) of the parameter coefficients are derived from the 99% confidence bounds of the normal distribution, and solving for the implicit standard deviation.

Finally, the parameter coefficient estimates are presented in the last column of Table 4.1. Estimation is made with the complete set of equations shown in Appendix I as constraints to the optimization problem,

$$
\min_{\mathbf{p},\mathbf{w}} l(\mathbf{p},\mathbf{w}) \equiv \mathbf{p}'\ln(\mathbf{p}/\mathbf{q}) + \mathbf{w}'\ln(\mathbf{w}/\mathbf{u})
$$

$$
\begin{aligned}
& s.t. \\
& F(\mathbf{Y},\mathbf{XZp}) - \mathbf{Vw} = \mathbf{0} \\
& \mathbf{1}_K = (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} \\
& \mathbf{1}_T = (\mathbf{I}_T \otimes \mathbf{1}'_J)\mathbf{w}.
\end{aligned}
$$

The 43 equations over $i \in G,N$ represent the system of restrictions $F(\mathbf{Y},\mathbf{XZp})\text{-}\mathbf{Vw}=\mathbf{0}$. Minimizing the objective function returns parameter coefficient point estimates and error estimates from the compared observed series.

The error support sets and prior probabilities are set by first specifying large bounds around zero, then by adjusting through symmetric increases or decreases around zero. The optimal error bounds were found to be plus or minus fifteen percent of the observed variable bounds. Since the error terms shrink model predicted values to actuals the bounds for them can rely on historical observations.

PARAMETER ESTIMATES

Estimated coefficients for Equation (4.8), *a priori*, are expected to take values near unity. The surprisingly high coefficients for goods producing sectors indicate that the benchmark-year estimates may not accurately represent the structure of the Oklahoma economy. Care should be applied when interpreting these coefficients since there is no intercept coefficient to act as a scale adjustment. The theoretical relationship does not call for one, so any scale adjustment must necessarily be made with slope adjustments. Several reasons may exist to cause this discrepancy. Firstly, the relationship across final demands and employment may be understated such that the *epv* for the benchmark year for goods manufacturing may be artificially low. It could also be an indication that the benchmark-year value of intermediate goods production is overstated, or that the regional purchase coefficient is suppressed for the benchmark year. Nonetheless, the purpose for stating Equation (4.8) as a statistical relationship is to account for such measurement inaccuracies, not necessarily identifying them.

Coefficient estimates for the wage rate equations reported in Table 4.1 are particularly interesting. Since the estimates are systems made, coefficient estimates depend not only on the right-hand-side variables, but also all the variables that affect those variables, including the left-hand-side variable being estimated. This is the result of internalizing the full set of feedback effects in estimation. Evident from the parameter estimates is that wages in goods producing sectors respond to variables quite differently than the non-goods producing sectors. Noteworthy is that the coefficients mapping wage rates to relative employment opportunity enters the two wage equations with opposite signs. Rationalizing the signs requires inquiring as to the source of the employment opportunities. Relative employment opportunity is calculated as ratios of employment to population relative to the nation. Both a decrease in supply and increase in demand of workers relative to the nation cause relative employment opportunity to increase while having opposite affects on wages. If industry wages are driven by local supply shocks, *REO* will enter the wage equation inversely, if driven by demand shocks, it will enter positively. The market conditions in goods producing sectors can be quite different from that in non-goods producing sectors leading to opposing signs as seen here. For instance, the negative sign associated with *REO* in non-goods producing sectors implies supply side changes in the labor market, while the positive relation for goods producing sectors signifies a labor demand-pull market. Furthermore, the time coefficients reflect the general decline in goods manufacturing employment in the U.S. Though both goods and non-goods producing sector reflect a negative relationship between wages and time, this negative relationship is more acute for goods producing sectors. Furthermore, coefficients to the relative cost of living index, *CP*, reflect that nominal wages respond positively to increases in the standard of living.

The estimated population equation reflects the Harris-Todaro (1970; Todaro 1970) theory of economic migration and augments it with Plaut's (1981) expression of non-economic migration. Coefficient estimates reflect the expected positive response to the general economic welfare of the region. If either wages or employment opportunities arise locally relative to the nation, net migration is expected to be positive. Furthermore the magnitudes of the coefficients tend to support the notion of risk aversion in economic migrants (Greenwood 1975). Since the coefficient for *REO* is greater than that of *RWR*, migrants are more sensitive to job availability than to wage differentials.

CONCLUSION

This chapter presents the model equations defining the structure of the Oklahoma Policy and Forecasting Model, describes data sources and cleaning, and explains model assumptions. This structure is then substituted as the moment restrictions in the cross-entropy problem that allows the model parameter coefficients to be estimated such that the model maintains optimal fit over history through minimizing the errors of key observable variables. Once estimates of these parameter coefficients are defined, the regional model is completely defined which has more structure and economic content than the traditional econometrically estimated model. The contention is that this process of completely structuring the relationships and estimating key parameter coefficients based on historic fit of the model will offer greater flexibility for analyzing the time-path response to policy changes that are not possible with static policy analysis models.

In application, it was found that optimizing such a complex non-linear objective with complex, non-linear restrictions is extremely taxing on the optimization algorithm. Nonetheless the ability of the estimated system to replicate history was found to be robust over many attempted model specifications, and the estimated parameter coefficients were nearly as robust. Furthermore, the fully estimated system fully replicates the base year observations with minimal coaxing; assuring the full set of equations and data properly conforms to the model specification.

What remains is to forecast the system and to perform policy simulations assessing the ability of the model to produce viable policy responses. This and in-sample diagnostics checks are presented in the next Chapter V.

Table 4.1: Prior Parameter Support Sets, Probabilities, and Estimates

	Prior Support Sets			Prior Probabilities			Implicit Priors		Estimates
	1	2	3	1	2	3	Mean	Std. Dev.	
$f_{FD,G,1}$	0.75	1	1.25	0.05	0.9	0.05	1	3.50E-01	1.17115
$f_{FD,G,2}$	0.75	1	1.25	0.05	0.9	0.05	1	3.50E-01	1.18511
$f_{FD,N,1}$	0.75	1	1.25	0.05	0.9	0.05	1	3.50E-01	1.0954
$f_{FD,N,2}$	0.75	1	1.25	0.05	0.9	0.05	1	3.50E-01	0.98952
$f_{W,G}$	-2000	0	2000	0.1	0.8	0.1	0	2.80E+03	173.89238
$f_{W,G,REO}$	-150	0	150	0.1	0.8	0.1	0	2.10E+02	0.49445
$f_{W,G,CP}$	-150	0	150	0.1	0.8	0.1	0	2.10E+02	5.16958
f_{W,G,W^U}	0	1	2	0.1	0.8	0.1	1	1.40E+00	0.12785
$f_{W,G,time}$	-2500	0	2500	0.1	0.8	0.1	0	3.50E+03	-86.44293
$f_{W,G,lag}$	0	0.5	1	0.1	0.8	0.1	0.5	7.00E-01	0.91616
$f_{W,N}$	-2000	0	2000	0.1	0.8	0.1	0	2.80E+03	-831.93193
$f_{W,N,REO}$	-150	0	150	0.1	0.8	0.1	0	2.10E+02	-8.10858
$f_{W,N,CP}$	-150	0	150	0.1	0.8	0.1	0	2.10E+02	-2.73789
f_{W,N,W^U}	0	1	2	0.1	0.8	0.1	1	1.40E+00	0.82397
$f_{W,N,time}$	-2500	0	2500	0.1	0.8	0.1	0	3.50E+03	-243.3425
$f_{W,N,lag}$	0	0.5	1	0.1	0.8	0.1	0.5	7.00E-01	0.09659
f_{N}	-8E+06	0	8E+06	0.1	0.8	0.1	0	1.12E+07	825553.16
$f_{N,REO}$	-4000	0	10000	0.1	0.8	0.1	600	1.40E+04	589.42946
$f_{N,RWR}$	-4000	0	10000	0.1	0.8	0.1	600	1.40E+04	370.38196
$f_{N,time}$	-8000	0	8000	0.1	0.8	0.1	0	1.12E+04	7953.40068
$f_{N,lag}$	-1	0	1	0.1	0.8	0.1	0	1.40E+00	0.73517
$f_{VA1,G}$	-0.1	0	0.1	0.15	0.7	0.15	0	1.40E-01	-0.006472
$f_{VA2,G}$	-0.1	0	0.1	0.15	0.7	0.15	0	1.40E-01	0.000074
$f_{VA1,N}$	-0.1	0	0.1	0.15	0.7	0.15	0	1.40E-01	0.004925
$f_{VA2,N}$	-0.1	0	0.1	0.15	0.7	0.15	0	1.40E-01	0.00027
$f_{rpc1,G}$	-0.025	-0.015	0.02	0.1	0.8	0.1	-0.01	4.90E-02	0.002571
$f_{rpc2,G}$	-0.1	-0.05	0	0.1	0.8	0.1	-0.05	7.00E-02	-0.022445
$f_{rpc1,N}$	-0.025	-0.015	0.02	0.1	0.8	0.1	-0.01	4.90E-02	0.000818
$f_{rpc2,N}$	-0.1	-0.05	0	0.1	0.8	0.1	-0.05	7.00E-02	-0.000055

CHAPTER V

POLICY SIMULATIONS AND FORECASTS OF THE OKLAHOMA POLICY SIMULATION AND FORECAST MODEL

INTRODUCTION

Chapter IV described the process of estimating parameter coefficients for the Oklahoma Policy Simulation and Forecast Model. Within this system, several equations were defined by their parameter estimates while others were held as identities. This chapter applies the set of equations, the parameter coefficients, and the predicted values of unobservable variables defined in Chapter IV to a new system of equations for forecasting. National drivers remain exogenous to the system and the equations in the forecasting system take the form of that found in Chapter IV.

The first section of this chapter explores the numerical solution to a system of non-linear equations, and defines the sufficient condition for this solution to exist. The second section presents the system of forecasting equations, tests the condition for a solution, and presents the forecasting system. The third section reviews the forecasting experiment and is followed by policy simulations that can be applied over the forecast horizon. The final section concludes.

EXISTENCE OF FIXED-POINT SOLUTIONS TO PROJECTIONS

Chapter IV presented a square system of nonlinear equations to be estimated. Unlike estimation, projecting the model does not entail simultaneously estimating parameter coefficients and fitting the set of nonlinear equations in the form of optimization constraints. Parameter coefficients are fixed throughout the forecasting horizon such that forecasting requires the simultaneous fit of the *N* endogenous variables over *N* equations defined in the estimation stage.

Projecting the model forecast equations beyond the sample years creates forecasts. The mere existence of a solution to the estimation process implies a solution exists for the out-of-sample period. As described in Chapter IV, efforts to estimate model parameters while the full set of equation constraints hold failed, and industry slack variables had to be added to the regional purchase coefficient equations, the existence of a solution to the full set of model equations has not been established. So this section presents the sufficient conditions necessary for a solution of single-year solves of the system of non-linear equations to exist. The existence of an out-of-sample solution requires a fixed-point solution to the system of equations.

The existence of this solution is verified by a generalized contraction-mapping theorem. Defining the system of linear or nonlinear equations as,

$$
\begin{aligned}
f_1(z_1, z_2, \ldots, z_n) &= 0 \\
f_2(z_1, z_2, \ldots, z_n) &= 0 \\
&\vdots \\
f_n(z_1, z_2, \ldots, z_n) &= 0,
\end{aligned}
$$

defines a set of n equations, f_i, in n unknown endogenous variables z_i. Those equations estimated in Chapter IV retain the estimated parameter coefficients derived from the system fit throughout the forecasting and policy simulation stage. Similarly, since the error terms denoting deviations from observed values are assumed a random process with expected value of zero, their out-of-sample values are set to their expected values and therefore dropped from the forecasting equations. Therefore all functional relationships are treated as identities in the forecasting system. A transformation G: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ defines the vector system of relationships as a mapping of n endogenous variables, into itself as,

$$
\begin{aligned}
z_1 &= g_1(z_1, z_2, \ldots, z_n) \\
z_2 &= g_2(z_1, z_2, \ldots, z_n) \\
&\vdots \\
z_n &= g_n(z_1, z_2, \ldots, z_n).
\end{aligned}
$$

An equivalent vector relation is stated as,

$G(\mathbf{z})$: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ where $\mathbf{z} \in \mathbb{R}^n$,

where $G(\mathbf{z})$ represents the vector functional relationships mapping the endogenous variables $\mathbf{z}$ into themselves. A solution $\mathbf{z}^*$ is said to exist if there exists a vector $\mathbf{z}^*$ such that $\mathbf{z}^* = G(\mathbf{z}^*)$. The existence of such a solution is assured through the well known *Banach's Fixed Point*, or *Contraction Mapping Theorem* which states that a contraction mapping on a complete metric space has a unique fixed point (Bartle and Sherbert 1992, pp. 368; Sydsaeter *et al.* 1999, pp. 36, 119). Therefore determining the existence of a fixed-point solution for the system at hand requires defining and testing the system for existence of a contraction mapping.

An n dimensional mapping into itself, $G(\mathbf{z})$: $\mathbb{R}^n \rightarrow \mathbb{R}^n$, is a contraction mapping if there exists some constant k in [0,1) such that

$\| G(x) - G(y) \| \leq k \bullet \| x - y \|$,

where $\|\cdot\|$ denotes the Euclidean distance between x and y, for all $x, y = \mathbf{z}: \in \mathbb{R}^n$ and is sufficient to define a complete metric space (Sundaram 1996, pp. 366). If such k exists, then the sequence $\{\mathbf{z}^{(l)}\}_{l=0}^{\infty}$ converges to a unique fixed point by the fixed-point algorithm,

$\mathbf{z}^{(l)} = G(\mathbf{z}^{(l\text{-}1)})$,

$\mathbf{z}^*$ in $\mathbb{R}^n$ such that $G(\mathbf{z}^*) = \mathbf{z}^*$. Moreover, for any starting point, $\mathbf{z}^{(0)} \in \mathbb{R}^n$, $\mathbf{z}^{(l)} \rightarrow z^*$ as l

$\rightarrow \infty$, where $\mathbf{z}^{(l)} = G(\mathbf{z}^{(l-1)})$ for $l \in \mathbb{N}$ (Burden and Faires 1989, pp. 531). Ortega (1972, pp. 153) provides proof of the theorem, while Moore (1977) proposes that convergence of the fixed point algorithm is proof of the existence of a local solution.

PROJECTION EQUATIONS AND SYSTEMS FORECASTS

The forecasting equations defining the projection relationships retain the exact structure of the estimating equations. Several supporting equations in the estimation stage are not present in the projection stage. For instance, the projection stage assumes that all relationships hold with certainty. In a different form, the expected value of the error terms described in Equations (4.41-4.42) is zero for all $t \in T_f$ where T_f $\mathbb{N}$ is the forecast horizon.

The complete set of projection equations are presented in Appendix II. For the current two-sector model, there exist 54 equations and 54 unknowns for each forecast period. The forecast horizon comprises 11 annual forecast periods from the year 2000 to 2010. Together, there are 54x11 = 594 variables to solve for in batches of 54, where each set of 54 solved variables are solved as a complete system fit of the closed 54 set equations. That is, they comprise the unique fixed point solution to the system for any given forecast year.

The system of equations has several lagged relation terms that link prior-year predictions to current-year predictions, but no lead relation terms. Therefore, for each $t \in T_f$, a solution to t-1 must be complete for a solution of t to be complete. Since there is no lead, or forward-looking relationships, the converse is not necessary. Therefore, the solution to the full set of T_f projection years can be performed sequentially from the lowest to the highest forecast year.

Since the solution proceeds sequentially over time with a lagged relationship, a set of starting values are required for $t_1 = \min\{T_f\}$. As described in chapter IV, the benchmark year is the last year of estimation and the solution is devised such that the benchmark year is fully replicated. For t_1 the lagged reference year, t-1, is the benchmark-year 1999, and the lagged relationships for the first year projection is simply the benchmark year values that are known to replicate the benchmark year observations.

Because there are equal number equations and endogenous variables, the usual counting rule holds for a unique solution to a system of equations but does not guarantee a solution (Sydsaeter *et al.* 1999, see pp. 38 for details). Proof of the existence of a unique solution rests on the contraction mapping theorem described above and requires a solution of the contractivity constant, k. Verification that $k \in [0,1)$ is sufficient for there to be a unique fixed-point solution to the set of projection equations for a single forecast period.

The calculation of the contractivity constant k is derived from the contraction mapping theorem itself. Take the relationship, $\| G(x) - G(y) \| \leq k \| x - y \|$. Replacing the inequality with equality and dividing by $\| x - y \|$ gives,

$$k = \frac{\|G(x) - G(y)\|}{\|x - y\|}.$$

Analytically, it is difficult to actually calculate a value of k from $G(x)$. Numerically the solution is inexpensive. Select any two vectors (x,y) within the domain of G and solve $G(x)$ and $G(y)$, then solve for k by substituting these vectors into the equation above. The solution of k is invariant to different vectors (x,y) as long as x and y are contained in the domain of G.

For the current context, y is defined as the benchmark values of the 54 endogenous equations and the vector of functional relationships, G, is defined as the 54 equations in Appendix II. Values of x are derived as $x = G(y)$ and $G(x)$ is defined from this solution. The constant k is then simply calculated as the ratio of the two Euclidean distances and found to be .97, which is within the necessary bounds required to define the system as a contraction mapping. Since the function G is a contraction mapping, by the Contraction Mapping Theorem, there exists a unique solution $\mathbf{z}^*$ such that $\mathbf{z}^* = G(\mathbf{z}^*)$ (Sundaram 1996, pp. 293).

The current application requires eleven years of consecutive fixed-point solutions to create a time-series projection over the forecast horizon. Convergence of the fixed-point algorithm confirms the solution of the n equations in n unknowns for a single year vector of unknowns, but not for the complete forecast horizon. Since the model is not forward looking and the equation parameters are fixed over time, the vector solution $\mathbf{z}^*$ of any projected year is independent of other years and the constant k holds for all years. Therefore for $k \in [1,0)$ for one observation is sufficient for all projected years to have a fixed point solution[9]. Appendix IV presents the MATLAB 6.1 program to create the Gauss-Seidel forecast projections.

FORECAST MODEL SPECIFICATIONS AND FORECASTS

The equations defined in Chapter IV define the estimation structure of the model. This section describes the similar structure of the forecast equations. The difference comes about because in forecasting with the model, structures placed to coax the estimation toward observable known values are not necessary. Furthermore, much of the

[9] Sharp changes in national drivers from one year to the next may cause a single year to not solve, but such a sharp change over time is not considered likely and following Moore (1977) proof of the contraction is found in the solution of all forecast periods.

haphazard complexities associated with estimation do not exist for prediction. For example, there is no need to be concerned with multiplying out a divisor to facilitate a solution in projections. In fact, it is essential that all equations have a single left-hand side variable, as the solution method employing a fixed-point algorithm requires mapping the set of variables onto itself. Therefore a square system of N unknown variables in N equations is imperative for a solution to the projection.

There are $N = 52$ variables and equations to estimate within the system. That is, there are 52 variables that feed back into the system and several that do not. For a system solution to exist, there must be a fixed-point value of the N-vector $\mathbf{z}^*$ such that $\mathbf{z}^* = G(\mathbf{z}^*)$. Projections are derived by iterating the fixed-point algorithm,

$$\left\{\mathbf{z}^{(l)} = G\left(\mathbf{z}^{(l-1)}\right)\right\}_{l=0}^{\infty},$$

to convergence.

Two sets of model predictions are made. One set replicates historical values to assess the dynamic fit of the in-sample predictions to the in-sample observations. This set of model predictions tests to ability of the model to replicate history. The second set of model predictions is the projection of the sample into 11 future, or out-of-sample predictions. Both sets are dynamic projections such that lagged relationships enter as past predicted values not actual observations.

IN-SAMPLE PREDICTIONS

To test the model's ability to replicate history, comparisons are made over known observations. Since known data is exhausted up to the benchmark year, this requires a comparison of in-sample observations. This gives the model a bit of unfair advantage since the data points the model is charged with replicating are the same points used in

Table 5.1: Oklahoma Policy Simulation and Forecasting Model:
In-Sample Percent Errors and Mean Absolute Percent Error*

	1995	1996	1997	1998	1999	MAPE
L	1.38	0.00	0.00	-0.64	2.10	0.83
L_G	1.46	1.45	0.26	1.19	1.81	1.23
L_N	1.36	-0.32	-0.05	-1.05	2.17	0.99
VA	1.01	-0.99	-0.48	-0.11	2.90	1.10
VA_G	4.78	-2.36	-3.32	0.07	2.69	2.64
VA_N	-0.25	-0.50	0.53	-0.17	2.96	0.88
W	-0.26	-0.13	-0.05	-0.08	-0.03	0.11
W_G	1.54	3.29	2.38	-2.20	0.00	1.88
W_N	-0.96	-1.58	-0.98	0.44	0.00	0.79
N	-0.48	-0.42	-0.35	-0.28	0.00	0.31

*Percent errors are computed for in-sample prediction against actaual reporte values. MAPE is calculated as the average of the absolute percent errors over the reported range

estimating the parameter coefficients. The exercise is nonetheless instructional in assessing whether the model can replicate history without including the error terms and therefore a useful tool for forecasting.

Table 5.1 shows the model percent prediction errors for the in-sample period of 1995 to 1999 and the mean percent squared errors (MAPE) of this range. The in-sample results show that there is reason to believe that the fixed point algorithm on the complete set of equations in Appendix III does produce results that mimic actual observations.

Care should be noted that the in-sample performance may be biased toward replicating the historical series, since the historical series the predictions are compared with are used in estimating the predictor. Notably, the equations for wage and population were particularly effective at replicating the benchmark-year observations. This is a reflection of the structure of the estimating relationship that seeks to replicate the benchmark-year observations.

To facilitate comparison of in-sample tracking of the simulation model to more generalized econometric methods, Table 5.2 presents the in-sample prediction MAPEs for a seven variable VAR estimated over the same period. The VAR system is made up of industry employment, wage rates, and value-added output, and population with one-year lags and national drivers. Estimation and prediction is facilitated in the EViews 4.1 programming environment and shown in Appendix V. Facilitating comparison, the in-sample MAPEs represent the same prediction advantage in predicting a subset of the estimation set. Comparing Table 5.1 and 5.2 shows that both compare equitably with neither showing any tendency for improved historical replication over the other.

OUT-OF-SAMPLE FORECASTS

As previously noted, the system of equations is solved for eleven out-of-sample years to form long-term projections of the model. These forecasts are conditional in that the forecasts are based on the condition that the national drivers take the value of their projections. National projections are the January 2000 *National Forecast Model* projections from Global Insights (2000) and are shown in Appendix V.

Modifying the fixed-point algorithm above to the Gauss-Seidel algorithm speeds convergence. The fixed-point algorithm described above updates variables between iterations. The Gauss-Seidel algorithm continuously updates the predicted values by immediately placing the new calculated value into the pool of **z** variables. For instance,

Table 5.2: Oklahoma Policy VAR Model:
In-Sample Percent Errors and Mean Absolute Percent Errors

	1995	1996	1997	1998	1999	MAPE
L	-0.68	0.03	0.91	1.21	-2.14	0.99
L_G	-0.13	-0.15	1.20	2.84	-5.86	2.04
L_N	-0.80	0.06	0.85	0.87	-1.44	0.80
VA	-0.40	-0.57	-0.13	3.95	-2.18	1.45
VA_G	-0.55	-1.04	0.84	6.97	-4.43	2.77
VA_N	-0.33	-0.37	-0.53	2.82	-1.38	1.09
W	0.10	-0.47	0.19	-0.15	1.65	0.51
W_G	-0.91	-0.16	1.59	-1.91	5.26	1.97
W_N	0.36	-0.54	-0.28	0.21	0.92	0.46
N	0.05	-0.02	0.08	-0.08	-0.59	0.16

rather than storing the new calculated values between iterations, the Gauss-Seidel algorithm employs new values in the calculation of the remaining variables within the iteration (Burden and Faires 1989, pp. 534). Depending on the ordering of the equations, this gives closer inter-iteration values to the fixed-point solution.

To properly order equations to facilitate convergence, variables that are both exogenous to the system and less dependent on the solutions of other endogenous variables should be solved first within the iterations. If not, highly endogenous variables will not benefit from the increased proximity to the fixed-point value of those variables. The ordering of equations is represented in Appendix IV. The equation numbers are retained from Chapter IV.

Forecasts are derived from the full set of equations by iterating on the Gauss-Seidel fixed-point algorithm. The system stopping rule is defined as the Euclidean distance between iterative solutions, or,

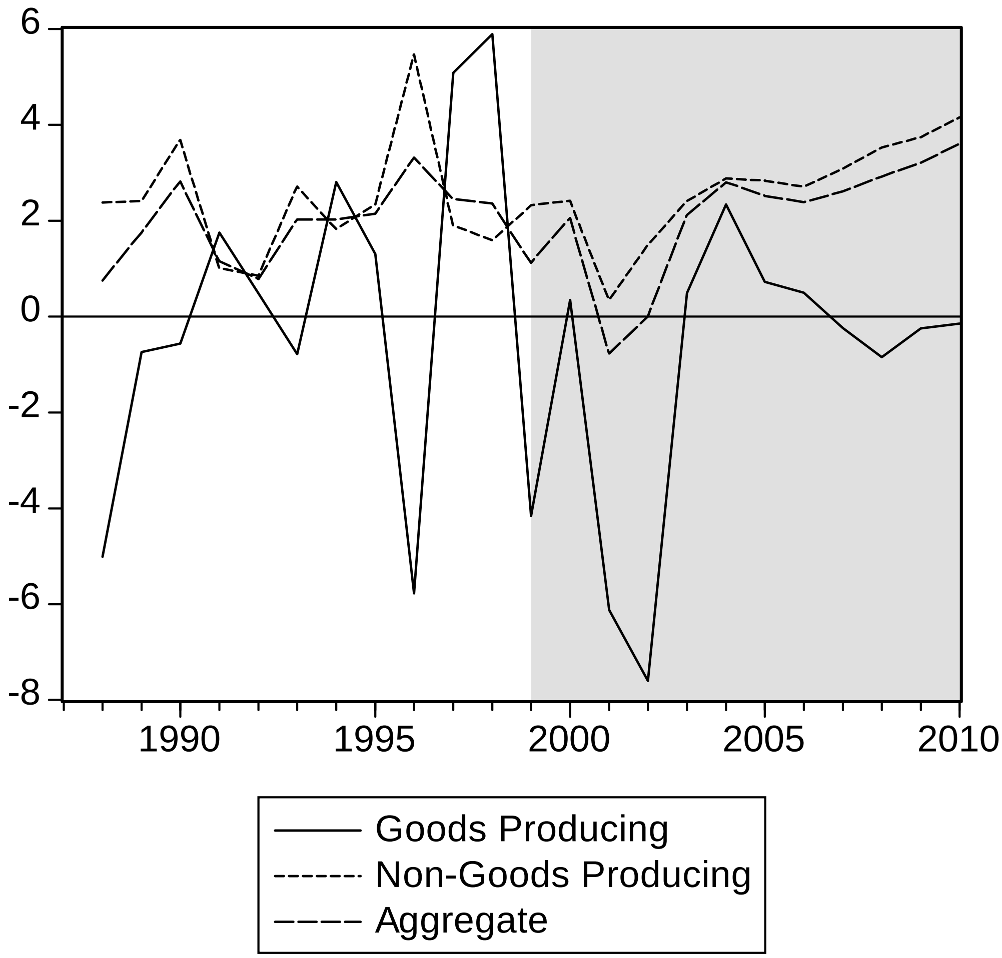
$\| x - y \| <$ 1.0E-9,

where y is the vector of new values and x is the vector of the previous solution. The complete conditional forecasts of the regional model for the forecast horizon 2001 to 2010 are reported in Appendix VI.

Table 5.3 shows the 11-year conditional forecasts from 2000 to 2010 for key economic variables. Of primary interest is how the model captures the recession of 2001. Figure 5.1 shows the percent change in industry and aggregate employment from 1987 through the projection period to 2010. As evident in the graph, goods producing employment declined in 1999 dragging down aggregate employment. The 2001

Tables 5.3 and 5.5 here

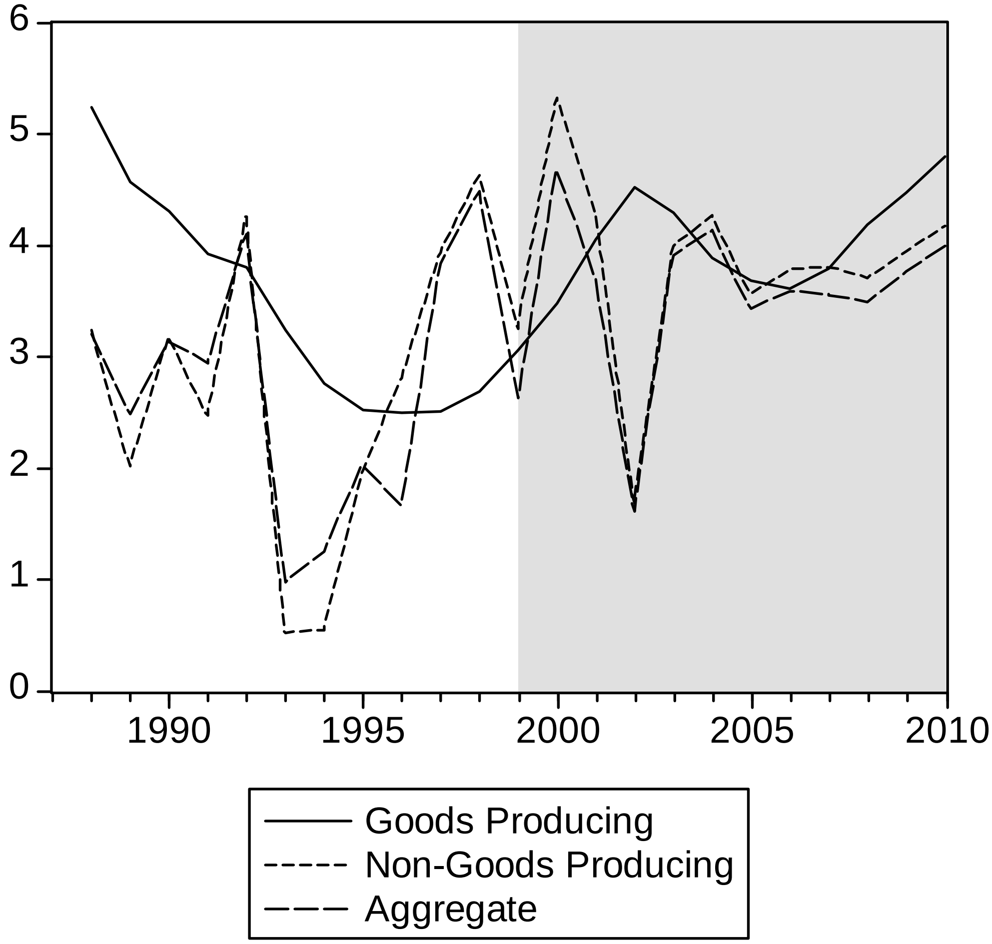
Figure 4.1: Growth Rates of Employment



recession is accurately captured by the projections of the national economy. That is, Global Insight's (2000) national projection for a 2001 decline drives the state model to similar declines.

In summary, the model forecast is predicting a growth rate in total employment that exceeds the sample period. Across the sample period of 1987 to 1999, the average growth rate in aggregate employment is 1.8 percent while the projection years of 2000 to 2010 project a mean growth rate of 2.1 percent. Goods producing industries are expected

Figure 4.2: Growth Rates of Wages



to shed more employment at an increasing rate over the projection period, but non-goods producing sectors will take up the slack for a net gain in aggregate employment growth.

Though employment is not driven directly by national employment, it is indirectly influenced through the ratio of national and regional employment per unit of value-added. An increase in national productivity is captured with a similar increase in regional productivity. Final demand components determine regional value-added output, which are driven by national final demand components. Together, productivity and final demand components determine regional employment.

Table 5.5: Correlations of State to Nation

	Sample	Projection
L	0.449	0.685
L_G	0.363	0.969
L_N	0.285	0.326
VA	0.422	0.941
VA_G	0.500	0.991
VA_N	-0.189	0.905
W	0.851	0.822
W_G	0.284	0.349
W_N	0.789	0.982
N	0.253	-0.284

Industry wage rates, on the other had, depend directly on a combination of national and regional drivers. Primarily the national/regional linkage is made with relative employment opportunities in the region relative to the nation. The estimated parameter coefficient relating relative employment opportunity to industry wage rates reflects the national/regional linkage to industry wage rates. Furthermore, national industry wages directly influence regional wages through the estimated parameter coefficient relating national wages to those of the region. Figure 5.2 shows the growth rates of industry and aggregate nominal wage rates. The model shows that the anticipated recession of 2001 will adversely affect non-goods production wages more than goods production. The model further projects nominal wage growth to exceed historical series, with growth in non-goods producing sector wages more than offsetting the losses in goods producing wages.

It is little surprise that the projected values of employment, wages, value-added output, and population are generally more correlated with national values than are the histories since the national values directly or indirectly drive the regional projections. Table 5.5 shows the correlation across these variables for the sample period and across the projections. With the exception of population, all projection correlations are consistent or stronger than actuals. Population is the exception because national population is not a driver of the regional population in the model, but it does indirectly influence it through relative employment opportunity, *REO*.

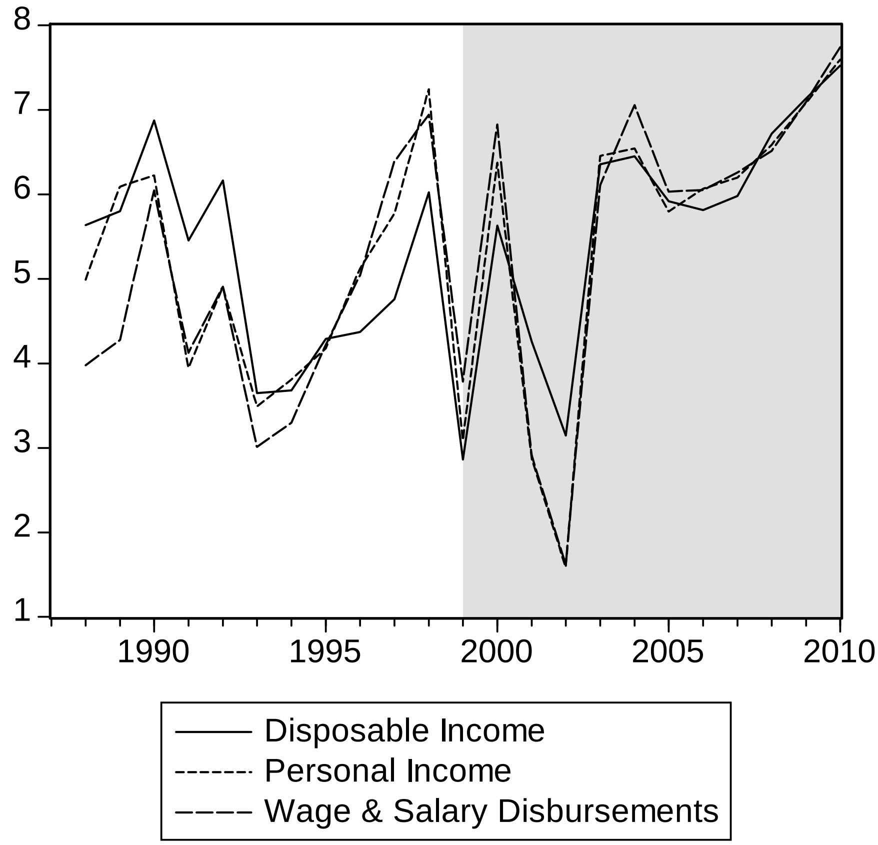
Table 5.4 shows regionalized aggregate and industry final demand and output projections. The sum of regionalized final demands defines value-added output, representing local and export demand for goods and services produced locally. The model is projecting a slowdown in the growth of all final demand components except for investment, which sees a short reversal. This exposes the highly sensitive nature of investment to economic shocks.

PRODUCTION BLOCK

Key components of the production block include labor productivity and the relative average cost of production. Table 5.6 presents the model projections of key production variables. The variables employment per value-added output, *epv*, and relative total factor productivity, *RF*, are two sides of the same coin. If less labor per dollar value of output is required, then labor productivity has increased. Since relative labor input is the only endogenous variable input in the production process, the same productivity increases measured in *epv* reflect in increases in *RF*. Relative capital and energy costs are allowed to deviate in policy simulations where both can alter profit-maximizing behavior and therefore total factor productivity.

Similarly, because the cost function of profit maximization is a function of variable factor payment, relative average cost, *RAC* is inversely related to total factor productivity *RF*. Furthermore, *RAC* is an inverse function of total factor productivity as shown in Equation (4.19) of Chapter IV. Therefore all the production relationships are

Figure 4.3: Growth Rates of Income Components



closely linked with *RLC* as the outside driver. This variable is determined by the wage rate equations and the national wage rates, but directly drives the remaining production equations.

INCOME BLOCK

Projections of the income block equations are shown in Table 5.7. The national slowdown in personal income growth is reflected in the region with a concurrent drop in Oklahoma income projections. Figure 5.3 shows the annual growth rates for wage and

INSERT TABLE 5.6 AND 5.7

salary disbursements, personal income, disposable personal income, and real disposable personal income over both the sample and the projection periods. Clearly evident is that the model is projecting a substantial increase in personal income over the forecast horizon, aside from the 2001 recessionary period, while growth in real personal income lags from the general increase in the region's price index.

POPULATION BLOCK

Population growth in the region is expected to make hefty gains. Comparing projection growth rates for the region to the nation, the average annual growth rate for the nation is expected to drop to 0.7% in the projection period from 1.0% in the sample, while the region is expected to maintain increase rates of population growth from 0.4% to 0.8%.

OUT-OF-SAMPLE COMPARISONS

This section compares forecast predictions of growth rates to out-of-sample actual observations. To facilitate comparisons, out-of-sample forecast accuracy of the current application is compared to forecast from the structurally specified Oklahoma State Econometric Model and a set of simplified vector autoregressive models. The three models represent three approaches to regional forecasting systems. The out-of-sample forecast horizon for comparison is limited to the years 2000 to 2003 by data release dates[10].

The Oklahoma State Econometric model (CAER, 2001) is structural in design, integrating economic theory and statistical relationships in specifying forecasting

[10] To facilitate out-of-sample comparisons, forecasts are compared to the BEA sources from which the estimation sample was derived. At current, the last BEA observation for wage rates, wage & salary disbursements, and employment are for the calendar year ended 2003, and 2002 for value-added estimates. Because of the short out-of-sample range for value-added, no comparisons are made for this measure.

relationships. These structural models tend to be expensive to design but effective as forecasting tools. The Mid-Year Update of the Oklahoma State Econometric model for 2001 produced a ten-year forecast that is the baseline for comparison. This forecasting model is a structural forecasting model designed to capture co-movements of key variables to the nation. Though it is structural it is not designed as a simulation tool for lacking the key linkages necessary to affectively capture policy responses.

A set of bivariate VAR models, with national drivers, are specified as atheoretical econometric forecasting models. These models are extremely general in structure and lack policy response linkages necessary for policy simulations. Large VAR models are argued to be affective policy analysis tools (Sims, 1982) but require substantial data histories for estimation. Bivariate VARs capture the essence of the methodology, without eroding the degrees of freedom necessary for estimation. Finally, the current application presents a completely structural approach to estimating structural relationships in forecasting models. All three approaches attempt to capture correlations across both time and inter-related variables to project the likely growth path of the economy into the foreseeable future.

Three bivariate VAR models are constructed through goods and non-goods manufacturing sectors for employment, wage rates, and value-added output. These variables represent the observable histories modeled with the current simulation model. Aggregates are further compared for total employment, wage rates, and value-added output.

The bivariate VAR specifications are as,

$$X_{it} = \sum_{i \in G,N} \sum_{s \in 1,2} a_{is} \cdot X_{it-s} + \sum_{i \in G,N} X_i^U,$$

where i is the industry sector, s is the lag length from time t, X_{it} is the endogenous VAR variable and X_{it}^U is the exogenous national driver. The parameter coefficient, a_{is}, is to be estimated with least squares and captures the co-movements across time and industries of the endogenous variables. Industry aggregates are estimated as the sum of industry estimates.

Out-of-sample mean absolute percent errors (MAPES) across three key economic measures and the three forecasting methodologies are compared in Table 5.8. The first column shows the MAPEs of the policy simulation model forecasts (PSM: CLOSE) discussed above, the second shows the MAPEs for the bivariate vector autoregressive models (VAR), the third column shows the MAPEs for the Oklahoma State Econometric Model (CAER), and discussion of the fourth column will be taken up later. The last row gives average MAPEs across all nine forecast variables and indicates favoritism for the CAER model while revealing that the bivariate VAR models as least accurate. Bold numbers indicate lowest MAPEs.

Table 5.8: Out-of-Sample MAPE to Actuals*

	PSM: CLOSE	VAR	CAER	PSM: OPEN
WSD_G	3.14	6.03	**2.59**	2.90
WSD_N	1.59	1.36	**1.09**	1.64
WSD	1.45	2.23	1.22	**1.20**
L_G	3.18	4.06	**2.39**	3.00
L_N	1.02	1.63	**0.98**	1.21
L	1.05	1.73	1.13	**0.88**
W_G	**3.33**	4.72	3.76	4.48
W_N	1.37	**0.51**	1.03	1.14
W	0.70	0.62	0.71	**0.56**
Mean	1.87	2.54	**1.66**	1.89

*Forecast horizon: 2000-2003; Bold indicates lowest observation.

Table 5.8 shows that the two structural models produce comparable out-of-sample MAPEs with the average MAPE for the CAER model showing an advantageous low overall average MAPE. The current model for Oklahoma generally offers lower out-of-sample precision than the structural state econometric model, while the bivariate VAR provided wildly varying MAPEs across sectors.

To be sure, a second comparison is made to statistically infer the comparative performances of the models. Since data revisions render the relative forecast values incomparable, the relative growth rates should remain comparable. That is, given identical national drivers, the projected growth rates of all three models should be comparable. To facilitate inference, the relative average growth rates are tested for equality with a two-sample t-test for equal means (Wackerly *et al.* 2002, pp. 492). With no assumption of equal variances, the test is specified as,

$$t = \frac{\overline{X}_a - \overline{X}_b}{\sqrt{s_a^2 + s_b^2 \big/ N}},$$

where a and b are the comparative models and $a \neq b$, $\overline{X}$ is the average projected growth rate from the year 2000 to 2003, s^2 is the sample variance, and N is the common sample size. The calculated t is distributed as a student's t-distribution with $v=2 \cdot N-2$ degrees of freedom under the null hypothesis of equal average growth rates. With $N=4$, $v=6$ and the two-tailed critical values are ±1.94 at confidence α = .10. The policy simulation, the VAR, and the CAER model projects equal growth rates to actuals under the null hypothesis.

Table 5.9 reports the calculated t-stats and compares mean growth projections against the CAER model. Table 5.9 shows that the test indicates scant evidence that the models are producing average growth rates different from the out-of-sample observations. More so, the signs of the t-statistics reveal a general tendency for over prediction. This is generally the product of estimating the sample over the economic growth spurt of the 1990's with the inclusion of time as right-hand-side variables.

IMPACT MULTIPLIERS AND POLICY SIMULATIONS

Up to this point, projections have been made of the baseline forecast. That is, the projections are conditional on the projections of the national economy assuming all endogenous components maintain their historical relationships. Forecasts from simulated changes in the regional relationships are compared to the baseline forecasts in what is generally termed impact analysis. These calculated impact responses measure the response of the regional economy to some exogenous change. To this effect, two types of comparisons are discussed below. The first is a unit specific response calculated as impact multipliers. Two sets of model closures are assumed to isolate the contributions of the current model to traditional IO model responses. The second type of comparison is

Table 5.9: Relative Student's t-Statistics of Equality of Relative Growth Rates*

	PSM: CLOSE	VAR	CAER	PSM: OPEN
WSD_G	-0.03	1.35	0.74	0.04
WSD_N	0.94	0.88	0.72	1.25
WSD	0.53	1.16	0.78	0.75
L_G	-0.18	1.12	1.23	-0.08
L_N	0.95	**1.84**	0.84	**1.46**
L	0.69	**1.87**	1.15	1.12
W_G	0.11	0.63	0.10	0.25
W_N	0.89	-0.22	0.41	0.63
W	0.75	0.42	0.44	0.43

*Forecast horizon: 2000-2003; Bold indicates significance at 10 percent confidence.

a unit-free measure of response to exogenous stimuli calculated as elasticity responses. Elasticity responses are useful in that elasticity responses can be compared over a wide range of studies given the unit-free measures.

IMPACT MULTIPLIERS

Impact multipliers can be readily attained from the model through counter-factual forecasts. The structure of the model facilitates analysis of total output and employment multipliers. Impact responses come from three sources; direct, indirect, and induced effects. The direct effect is the actual exogenous change in final demand that may arise from increased export demand or government projects. The indirect effect results from interindustry demand for inputs from other industries that produces a chain of responses across all industries. The combined direct and indirect effects are associated with the Type I multiplier responses through interindustry transactions found in traditional input-output models. Adding the induced effect expands the endogeneity in recognizing that greater regional output leads to more jobs and more regional income, some of which is spent locally. This leads to greater output and the continuation of the chain of responses found in closed input-output models.

In closed models as this, multiplier effects arise through two avenues. One is through the open-model linkages and the other is through model closure. In the input-output literature, open models only account for direct interindustry linkages through interindustry sales and generally associated with Type I impact multipliers. In these open models, regional households are exogenous to the system and merely contribute to total final demands. Since changes in household incomes resulting from increases in final demand do not contribute additional economic activity, changes in household incomes are considered leakages to open models and do not interact with the production sector.

Closed models impart the effect on households in changes to output and are associated with Type II impact responses. Regional households earn their incomes through labor services and spend their income as consumers. Increasing output requires more employment and household income, which is partially spent locally across all industries inducing further increases in the total regional output. Therefore, closed models treat households as endogenous to the system allowing households to interact with the production sector. The increased closure of endogenizing household income increases the multiplier size.

The current application extends the Type II responses to the extended multiplier responses of Batey and Rose (1990) that include induced investment and government expenditure responses. The additional closure offers greater feedback effects and lead to larger impact multipliers. To compare the current application to the traditional Type II multipliers reported by IMPLAN (Rickman and Schwer 1993, 1995) requires opening the model by the exclusion of these extended multiplier responses. Therefore, in contrast to the traditional terminology, openness refers to the restrictive Type II model closure and closeness refers to the extended model closure that up to this point has described the current model.

Restricting the model to Type II responses requires restricting investment and state & local government responses to exogenous behavior. To do so the investment and state & local final demand linkages, Equations (4.4 and 4.6 respectively) are restated as proportionally related to their respective national drivers as,

$$Z_{it} = \gamma_i Z_t^U,$$

where Z_{it} is the final demand component re-specified as exogenous for industry i, Z_t^U is the national aggregate driver, and γ_i is the calculated proportional relationship to the national aggregate driver calculated at the benchmark year as the ratio of Z_t and Z^U.

Specifying the current model to this relatively open Type II response specification does not certify equal comparisons of Type II multiplier responses to input-output models. If this re-specification erodes the model's ability to track historical values and forecast well, then the re-specified model will not produce valid multiplier responses. That is the accuracy of the multipliers are tested against the data by assessing the model's fit over time (Rickman 2002). In doing so, a second set of forecasts are derived by setting investment and state & local government final demand components exogenous, and these forecasts are compared to actual observations.

The forth columns of Tables 5.8 and 5.9 (PSM: OPEN) compare these projections to actual observations. Evident from comparing the PSM: CLOSE columns to PSM: OPEN, is that the open version of the model that replicates Type II model closure compares admirably to the closed version. The closed model does relatively poorly for goods producing industries tending to over predict where multiplier linkages are more pronounced. In fact there's a tendency for the open model to further over-predict all variables than the closed model. Nonetheless, the open formed model returns a comparable average MAPE over all industries, produces growth rates not significantly different from actuals, and produces MAPEs sufficiently lower than the bivariate VAR models. Accepting that both the open and the closed form of the model sufficiently tracks historical values and forecasts well, impact responses can by analyzed with the assurance that the model relationships reflect the historical pattern.

Impact studies generally report two types of impact responses. The first calculates the change in the value of total output given an exogenous increase in industry *i* output demand,

$$M_{Q'} = \frac{\Delta Q}{\Delta Q_i},$$

while the second measures the change in total employment given a change in industry *i* employment valued exogenous output,

$$M_{L'} = \frac{\Delta L}{\Delta L_i}.$$

Both measures reflect on the expected outcome of a change to some exogenous demand that causes a resulting chain of demand linkage responses. The greater the closure of the model the greater will be the multiplier.

Impact multipliers based on the current applications are analogous to those derived in the input-output literature in that a cumulative measure of total response is derived from an exogenous change in a final demand segment. Whether measured in output or in employment terms, impact multipliers are calculated as the ratio of the total effect of a change to the direct effect, or the change itself. Numerically, they are calculated by imposing a single-year shock to exports in the goods producing industry, projecting the model over the forecast horizon, summing the difference between the base projection and the prior over the forecast horizon and dividing by the value of the initial shock, or,

$$M_Q = \frac{\text{Total Effect}}{\text{Direct Effect}} = \frac{\text{Direct} + \text{Indirect} + \text{Induced}}{\text{Direct}} = \frac{\sum_{s=0}^{H} \Delta Q_{t+s}}{\Delta X_{it}},$$

where H is the forecast horizon, s is the forecast observation, ΔQ_t is the difference between the baseline projection and that of the shocked series, and ΔX_t is the shock at year t. The total effect is calculated as the sum difference in the counterfactual projection to that of the baseline projections.

Deriving employment multipliers requires transforming the change in final demand into employment terms. To do so, output is simply restated in employment terms similar to that shown in Equation (4.1) of Chapter IV. Converting output, Q_{it}, into employment terms requires first converting industry output into value-added output, VA_{it}, with the multiple λ_{VAit}. Next, value-added output is transformed into employment terms through the multiple epv_{it}, or,

$$L_{it} = Q_{it} \cdot \lambda_{VAit} \cdot epv_{it}.$$

Having the direct effect in labor terms the next step is to calculate the sum-difference of the counterfactual projections to the baseline projections and divide it by the initial change measured in employment terms to derive the employment multiplier,

$$M_L = \frac{\sum_{s=0}^{H} \Delta L_{t+s}}{\Delta(X_{it} \cdot \lambda_{VA,i} \cdot epv_i)} = \frac{\sum_{s=0}^{H} \Delta L_{t+s}}{\Delta L_{it}}.$$

This is identical to the output multiplier except that the numerator and denominator are stated in employment terms.

Impact multipliers derived from the present application do not suffer the criticisms of static input-output and CGE model multipliers, in that the derived multipliers are based on dynamic relationships and follow a time-path response. Where static models are limited to long-run multipliers, they cannot assess the timing of the impacts. Several attempts have been made to garner the full time-path responses with estimates of dynamic multipliers over well-defined time periods (see for example: Kraybill and Dorfman 1992; Krikelas 1992; LeSage and Reed 1989); these studies are limited by the econometric techniques that estimate the dynamic relationships. A short-coming of these attempts is the use of single equation estimation that leaves the absence of full-model closure in measuring multipliers. That is, the multiplier responses are partial, not general equilibrium estimates. General equilibrium estimates recognize the inter-dependence of industry sectors. The current application effectively tracks the time-path responses as well as the general equilibrium structure of a closed-formed CGE models relaxing the fixed-price restrictions of IO models.

Furthermore, the traditional Type II IO multipliers implicitly assume perfectly elastic supply responses implicit in long-run constant returns to scale (Isard *et al.* 1998, pp. 306) abstracting from short run adjustment. This is a primary reason for their denotation with long-run multipliers. In the short-run, capacity may not be sufficient for all industries to respond with constant costs to all changes. The current application allows for partial fixed responses in the labor market leaving capital and energy markets unconstrained. An increase in employment demand will result in wage pressures that will cause the strained wage market to increase wage rates. Since a proportion of regional income is spent on locally produced goods and services, this increase in wage rates reflects in greater demand for locally output sparking further increases in output.

This increase in wages will also erode the relative competitiveness of regional producers to national producers. Following the REMI model, increases in wage-costs to regional producers will impel local producers to increase the selling prices (Rickman and Schwer 1993). Cost minimizing consumers will be encouraged to switch to competing regional imports as reflected in a decline in the regional purchase coefficient, Equation (4.11).

Multipliers derived from the traditional IO table implicitly assume no price responses, but the current application allows the comparison of the two effects if compared with those multipliers based on the IO table. Impact responses that are greater than those of IMPLAN give evidence in support of the dominance of the wage impact on increased demand, while those smaller than IMPLAN's gives evidence in support of the dominance of resource constraints (Rickman and Schwer 1993).

Calculated open and closed model output and employment multipliers from the current policy simulation model and Type II impact multipliers of a similarly aggregated IMPLAN IO model are calculated and compared. The closed policy simulation model entails more closure than IMPLAN by endogenizing investment and state & local government final demand components. Therefore multipliers are calculated based on relatively more open specification that restrict feedback effects to the consumption component of final demand only. Limiting the feedback of investment and state & local government components to exogenous requires that impact responses be excluded from these components in counterfactual projections. This is facilitated by restricting these values to there baseline forecasts for all $s \in H$ and all $i \in N$.

Table 5.10 shows that the unrestricted open model projects greater impact responses than IMPLAN, indicating that the feedback responses, through induced consumption, investment, and state & local government final demands, outweigh the offsetting resource constraints. Relatively large impact multipliers for the open model to those of IMPLAN show that the income effect dominates the offsetting effect of the loss of relative competitiveness of regional producers from labor supply adjustments. The results counter those of Rickman and Schwer (1993, 1995) who find that the supply response dominates the REMI extended multipliers resulting in lower multiplier responses against a comparably specified IMPLAN model.

The more comparable model multipliers, that restricts closure to that of IMPLAN is not as revealing. By restricting feedback responses to the consumption component of final demand, the upward bias to IMPLAN's Type II multiplier is partially mitigated by restricting the feedback through investment and local government channels. This results in smaller multipliers than those for the closed model. Furthermore, the relative impacts to those of IMPLAN are not congruent in that basic production has a relatively larger impact than IMPLAN and non-basic has a relatively smaller impact. Succinctly, the offsetting resource constraint is shown through estimation to not be as great for manufacturing as it is for service industries. This result is consistent with Rickman and Schwer's (1993, 1995) findings that the REMI and IMPLAN models give consistent impact multipliers adjusting for industry classification, and closure assumptions.

Given that the present model gives greater multiplier responses is evidence to the feedback of prices that are absent from IO models. Increases in export demand cause an increase in derived labor demands for both the export and tertiary industry output. IO

Table 5.10: Calculated Output and Employment Multipliers

	PSM: OPEN		PSM: CLOSE		IMPLAN Type II	
	G	N	G	N	G	N
Output	2.664	2.412	2.318	2.001	2.248	2.153
Employ	4.470	2.177	3.710	1.867	3.540	1.997

models fail to capture the price effects on wages while the present model fully endogenizes price movements. The increased wages lead to greater regional income, some of which is spent locally but the offsetting decrease in the region's national competitiveness also reduces this demand. This wage response dynamic is not captured in IO models.

ELASTICITY OF RESPONSE

Several policy simulations can be performed to study the impact transmission mechanism within the model. These simulations rely on some policy change or exogenous change that alters variables within the model exogenously. For example, the Quality Jobs Program of Oklahoma results in lower employment costs by subsidizing Oklahoma firms with cash subsidies for new job creation. Furthermore investment subsidies are reflected in regional capital cost reductions. These simulation responses are measured in elasticity responses because elasticities are unit free measures of responses allowing comparisons across many different responses.

Elasticity responses are defined as the percent change in a variable given a percent change in another. In the current context of dynamic responses, and given that the full extent of a response is not fully accounted for until the full adjustment period is complete, the elasticity responses are calculated from a permanent, or persistent, exogenous change as,

$$\varepsilon_{Y,X} = \frac{\%\Delta Y_H}{\%\Delta X_s}, \text{ for all } s = 1, 2, \ldots, H,$$

where s is the forecast observation, H is the forecast horizon, Y is the endogenous variable of interest and X is the exogenous shock.

Bartik (1991) summarizes the elasticity responses across various exogenous shocks to state and MSA economies. Few studies are directly comparable given the breadth of model assumptions, geography considered, model closure, and measures that are used to calculate elasticity responses. Generally, most studies attempt to explain policy shifts on regional unemployment not employment itself. Comparisons of these studies to studies on employment responses are blurred by the implied migration responses found in unemployment responses. Nonetheless, they are useful as first approximations.

To facilitate comparison to the current model, Bartik concludes that the elasticity response of wages on changes in demand driven employment change is around .15 to .6 (Treyz and Stephens 1985, Topel 1986). Bartik further finds a wide band of elasticity responses of business activity to tax changes from past studies. The difficulty associated with testing such responses is that the mode of distributing proceeds of the tax, the reallocation of the tax, and the timing of the response empirically changes the response, such that a generalized finding is not generally achievable. For example, if it is true that firms and owners of firms seek avoid paying higher taxes, a relative increase in business or personal taxes will induce business activities to relocate to low tax regions, but if the government revenues are reallocated in such a way to create business amenity effects that fosters productivity gains, economic activity will be drawn to the region. The two mitigating effects may or may not be fully offsetting. Therefore the literature tends to indicate a range of elasticity responses of -.90 to .04, with most evidence indicating that the negative effect dominates (Testa 1989, Wasylenko 1988, Romans and Subrahmanyam 1979). Canto and Webb (1987) find that the elasticity response on personal income from changes in state incidence of tax is -.35.

Table 5.11 reports key elasticity responses to four exogenous changes derived from counterfactual projections from the model. The elasticity experiments consists of the elasticity of responses changes in capital and energy costs, changes in tax burden, and to changes in exogenous employment. The first three constitutes changes in factor costs of production while the forth column gives responses to government imposed net taxes and the fifth to exogenous growth.

Starting with factor price responses, Factor input relationships such as labor inputs, are determined in the indirect cost function, or simply the cost function, such that profit maximization implicitly states factor demands as conditional factor demands, conditioned on the vector of factor rental rates. As the cost of one factor input increases the profit maximizing response is to shift away from that factor and into the now relatively less expensive remaining factors. Berndt and Wood (1975) and Griffin and Gregory (1976) extend factor inputs to include energy. In the current context the factor elasticity of substitution is limited to unity by the application of the Cobb-Douglas production function, which imposes strict substitutability across labor, capital, and energy factors (Chung 1994, pp. 97).

Berndt and Wood (1975), and Griffin and Gregory (1976) estimate cross-price elasticities across OECD countries and U.S. manufacturing industries respectively. Where these studies attempt to estimate factor elasticities from the translog production function, they are limited in that estimates are produced within a partial-equilibrium framework and are limited to static relationships. The current application is limited in that the general-equilibrium framework relies on the restrictive Cobb-Douglas production function, but the inclusion of full-equilibrium constraints at estimation offers a formidable improvement over partial-equilibrium estimations.

Table 5.11 shows that the own-price labor elasticity is -.78, and the cross-price elasticities of labor to capital and energy costs are .02 and .01 respectively. The calculated own price is higher than those of the previous two studies that find -.47 and -.27 for Berndt and Wood and Griffin and Gregory respectively. The cross-price elasticities are more consistent with Berndt and Wood who finds .05 for capital costs and .03 for energy. Griffin and Gregory find .12 and .15 for capital and energy respectively.

The RLC shows that changes in relative labor costs lead to atomistic changes in employment per unit of value-added (*epv*) output since value-added output changes in near lockstep with labor inputs. In general, *epv* is expected to be inversely related to labor costs as labor productivity gains would be expected as capital and energy deepening takes place. Though the model gives slight evidence of this productive gain the adjustment is rigid because of the offsetting changes in both relative total factor productivity (RF_{it}) and relative labor costs (RLC_{it}). From Equation (4.16), *epv* is defined as,

$$epv_{it} = \frac{epv_{it}^{U}}{RF_{it}} RLC_{it}^{\alpha_i - 1} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i}, \tag{4.16}$$

where *RCC* and *RFC* is assumed one throughout the forecast period. The percent change from the baseline solution in epv_{it} is found by taking logs and the total differential,

Table 5.11: Key Simulation Elasticities

	RLC	RCC	RFC	TAX	Growth
L	-0.775	0.020	0.007	-0.087	1.646
VA	-0.773	0.009	-0.004	-0.084	1.738
W	-0.020	-0.029	-0.036	-0.012	0.153
WSD	-0.773	0.635	-0.016	-0.100	1.802
YP	-0.629	0.530	-0.005	-0.078	1.432
YD	-0.496	0.420	-0.003	-0.148	1.122

holding the exogenous variables constant, to get,

$$\frac{\Delta epv_{it}}{epv_{it}} = -\frac{\Delta RF_{it}}{RF_{it}} + (\alpha_i - 1)\frac{\Delta RLC_{it}}{RLC_{it}}$$, or in percent deviation,

$$\%\Delta epv_{it} \approx (\alpha_i - 1)\%\Delta RLC_{it} - \%\Delta RF_{it},$$

where the positive share parameter $\alpha_i < 1$. For $\%\Delta epv_{it}$ to be zero requires that

$\%\Delta RLC_{it} \approx (\alpha_i - 1)^{-1} \%\Delta RF_{it}$ which is found to be approximately true for all $t \in H$.[11]

Ultimately, changes in the relative labor costs lead to slightly less than one-to-one proportional changes in labor, value added output, and wage & salary disbursements. Despite this overly restrictive assumption, Griffin and Gregory (1976) show the Cobb-Douglas representation provides a good representation of substitutability across factor inputs.

Changes in capital costs reflect government policy toward subsidizing local production development as characterized by regional business recruitment packages of low business taxes, subsidized utility expenses, supplying developed land for factories, and even sharing construction expenses. Low variable capital costs translate into capital deepening and lower per-unit employment. Returns to capital are not limited to the region as benefits to high capital returns in the region are allowed to flow across the nation without cost.

Table 5.11 reports key industry and aggregate elasticity responses to increases in relative capital costs to the nation, *RCC*. The direct result of an increase in *RCC* is an increase in employment per unit of value-added output, and is a reflection of the

[11] Different orderings of the system of equations was attempted to coax the solution away from this apparent anomaly, but it was found that any ordering of epv_{it} after RLC_{it} caused a razor's edge solution that was unstable to the policy simulation but not to the baseline solution. For lack of a more appropriate alternative, the current solution was deemed appropriate.

substitution of labor for capital when capital costs increase. The elasticity response of employment to capital costs of .02 is lower that those found by Crihfield and Panggabean (1996) who report cross-price elasticities of about .45 for MSA production. Furthermore, value-added output is negatively related, as the rate of change in *epv* is greater than the rate of change in employment arising from increases in *RCC*;

$$\%\Delta VA_{it} \approx \%\Delta L_{it} - \%\Delta epv_{it}.$$

Since the conversion of increased capital costs into property-type income is disseminated over the nation and not limited to the state, there is no negative impact on local incomes, such that the positive output impact is derived through greater reliance on employment.

Similar to changes in relative costs of capital, changes in relative energy costs (*RFC*) cause a direct response to employment per value-added output but secondary affects to value-added output. Table 5.11 shows that the region's responses to changes in relative energy costs are similar to those for relative capital costs (*RCC*) but less pronounced since energy makes up a smaller share of total factor expenditures.

Given that personal income and total regional value-added output responds with opposite signs, the increase in income must represent increased demand for goods made outside of the region. Recall from Chapter IV that an increase in imported goods relative to regionally produced goods is the same as a decrease in the regional purchase coefficient. The results here indicate that the regional purchase coefficient moves in the same direction as *RCC* and *RFC*.

Wage responses to changes in factor costs are consistent across all factor costs. Since the wage linkage does not represent supply and demand linkages, wage responses are limited to projections of past correlations and can not discern supply and demand responses. As described in Chapter IV, the estimated response parameter of wages to relative labor opportunity reflects whether the segment's labor market is historically supply driven or demand driven. In comparison Bartik (1991, pp. 148-149) finds no long run relationship between employment growth and wage rates, but reports several surveys that show a positive relationship.

Increases in the personal tax burden decrease disposable income, consumption expenditures, and output, leading to further decreases in income through a reduction in relative employment opportunities and wages. Since an increase in personal tax burden decreases personal income, ultimately, the increase in tax revenue is partially offset by a reduction in tax receipts through reductions in wage and salary disbursements. This implies that the elasticity response of personal income to changes in taxes will be less than unity in absolute terms as shown in Table 5.11. The elasticity response of employment to taxes of .087 is consistent with those of local taxes found in Crihfield (1989) who finds -.07 for increases in the incident of local tax.

CONCLUSION

In this chapter, the system for projecting the forecast and simulation model was presented. The existence of the model solution required for projection is found through the contraction mapping theorem and verified with actual model solutions for all out-of-sample forecast periods. The time-series projections are driven by national projections and internal model linkages allowing the model to affectively capture the national economic correction of 2001. Furthermore, the anticipated continuation of national productivity gains of the 1990's is reflected in the regional projections of higher than average growth rates through the forecast period.

These projections from the model have the advantage in capturing not only trends in the data, but also the structural co-movements of economically related variables. The added attribute grants the ability to conduct policy simulations with the projects through counter-factual forecasts. comparing simulations to analogous IMPLAN simulations shows that impact multipliers exceeds those given by the more restrictive Type II multipliers of IMPLAN attributed to price responses. The wage price response dominates the regional selling price responses reflecting growth in real disposable personal income. Though policy impact multipliers are not tested against observations, generalizations of fixed price policy responses are generally considered to be preferred to fixed price responses (Isard *et al.* 1998, Chp. 7).

In general, sub-national elasticity responses of employment growth to changes in exogenous demand follow the findings in economic growth theory that postulates growth can only be altered in the short-run. In the long-run the growth rate of employment and economic activity reverts back to its long-run trend. Two offsetting effects can dominate and alter the long-run activity trend. First agglomeration effects imply that acceleration of regional economic growth causes productivity spillovers that will continue building causing a long-run increase in economic growth rate. Second, congestion effects act to reduce the rate of economic growth as increases in economic activity taxes the regions resources and capital structure. If the two are offsetting, then changes in the short-run growth of employment and economic activity will eventually revert to a long-run natural rate of growth. Bartik's survey supports this long-run natural rate of growth in that exogenous increases in economic activity lead to only short-term gains in economic growth (Bartik 1991, pp. 64 and 95).

CHAPTER VI

CONCLUSION

Bayesian methods offer opportunities to surmount the common problem for regional economists of the trade-off between policy analysis and forecast accuracy. It transcends this trade-off by accommodating sampling and non-sampling information in model estimation. Rather than exacting restrictions imposed by variable exclusion, this methodology imposes prior information only to the degree that the practitioner imposes allowing the two extreme cases of no prior influence and exacting restrictions be special cases of the general estimator.

As the survey of Bayesian applications shows, there exist virtually no limit to the extent in which Bayesian priors can represent economic theory. The full account of past application can only be a partial representation of the applicability of Bayesian methods of integrating non-sample information in estimation. Many opportunities exist to implement stochastic Bayesian restrictions to estimating economic relationships at the regional level.

One such Bayesian application, presented here, utilizes the entropy variant of Bayesian methods that eases computational difficulties of specifying Bayesian estimators and results in a simple non-linear math programming problem. In application, the analytical complexities of the Bayesian formalism is supplanted with numerical complexities and the anticipation of a fluid method of estimating the complete structure of a theoretical regional model over time was soon squelched[12]. Though not living up to its anticipation, the complete systems estimation through cross-entropy offers a viable alternative to traditional econometric approaches that do not fully account for the complete set of general equilibrium constraints. The Bayesian-like restrictions in estimation impose a varying degree of non-sample information on the otherwise sample estimation. It is also viable for estimating relationships that are ill-conditioned and otherwise not estimable with traditional methods. Though protracted to implement, the system estimates of the structural set of equations representing the Oklahoma economy produce viable forecasts and policy simulations.

The underlying structure of the model follows that of the commercially successful Regional Econometric Modeling Incorporated REMI model (Treyz *et al*. 1992). This structure is similar to static general equilibrium models in that it integrates both supply and demand linkages, but adds a time-adjustment mechanism that is estimated over historical data. Where the REMI model relies on national pooled estimates for parameterization, entropic estimation allows the structure to be fully estimated within the region of study. By transcending the need for national estimates, the full set of regional specific relationships can be represented. The benefit of doing so is that peculiar relationships that are specific to the region are incorporated in estimation increasing accuracy of both regional forecasts and simulations.

Contrasting the proposed model forecast accuracy to out-of-sample observations shows that this systems estimation methodology produces forecasts in line with a traditional econometrically estimated forecasting model and forecasts that are superior to

[12] On a theoretical account, the same could be said of other numerical Bayesian methods. See Dorfman (1997)

bivariate vector autoregression models. Though forecasts projections are not universally better than existing methods, the strength of the entropic estimation method, and the existing model, is the ability to estimate the inclusive structure of the model. This inclusive structure allows policy and growth analysis not available to traditionally estimated regional forecasting models.

Policy simulations show that derived multiplier responses are characteristic of those found in the IMPLAN model that is used to set benchmark values. Differences from the IMPLAN multipliers arise from increased market structure and time-path responses. The time-path responses are induced through estimating the economic structure over time, while the market structure generalizes the assumptions of the IMPLAN structure.

Similar to IMPLAN's calculated multipliers, the current model accounts for direct, indirect, and induced effects in calculating multiplier responses. But unlike IMPLAN's multipliers induced effects are extended to include induced investment and government expenditure responses and the elastic supply response assumption is relaxed to account for resource supply constraints.

Resource supply constraints are assumed to exist through labor market linkages. Supply pressures from rapid economic growth will likely lead to wage rate increases. Increasing wage rates create two offsetting responses. First, wage rate increases inject more income into the local economy that is then partially re-spent locally causing an increase in the total economic impact for a given exogenous change. Second, wage rate increases force businesses to cut back on employment dampening the economic impact for a given exogenous change. IMPLAN's Type II multiplier assumes that the two offsetting effects are netted out. The current model offers the alternative of letting the structure and data determine the offsetting effects through the complete set of model linkages.

Empirically it is found that the prior positive effect dominates exogenous changes in the goods producing sectors, while the latter effect dominates exogenous changes in the non-goods producing sectors. By specifying the model to comparative closure assumptions, the cumulative impact on exogenous changes in goods producing sectors give multiplier responses greater than, and exogenous changes in non-goods producing sectors give multiplier responses less than those of a comparably closed IMPLAN model.

Empirically the full-structural model is computationally expensive. A natural extension of the current model is to increase the level of desegregation including greater industry detail. Computationally, this is seen as impractical. The current experience shows that increasing structural complexity necessitates decreasing estimation parameters. Several issues contribute to this. First, moment restrictions of the current application that induce the structure to the model are non-linear and lead to more localized instability in estimation. Though locally optimum solutions were not generally encountered in this application, locally infeasible solutions where common. Second, unobserved or latent variables proved to be a particular source of contention in estimation. Increasing industry detail necessarily requires a one-for-one increase in the number of regional purchase coefficients and output per value added. This also increases the number of estimated final demand components five-to-one and two-to-one increase in production relationships. Small increases in industry detail lead to large increases in model structure.

Structurally, the model deviates from the commercially successful REMI model little. Complexities are added in estimating the relationships of final demand components to total final demand rather than restricting this relationship to unity. Estimation complexities rest on the complete closed-form estimation. An extension to the current application is to test if full model inclusion in estimation contributes to accurate parameter estimates over single-equation methods. If not, single-equation estimates, hard-coded into the model structure in estimation, will benefit model estimation by reducing the total number of systems-fit estimation coefficients. In the current context, these equations include Equations (4.16, 4.20, and 4.39) with appropriate proxies for relative prices, and relative total factor productivities. Imposing REMI-like restrictions of unity to final-demand component responses to aggregate final demand can further limit the number of parameters to estimate drawing less information contained in the data to estimates of the regional purchase coefficients and value-added to output parameters. This comes with the cost of less generalized results that depend absolutely on the validity of the benchmark-year input-output table.

A final extension is to relax the complexities of non-linear restrictions of the relationships on estimation. The generally econometric approach to relaxing non-linearities is a first-order Taylor-series expansion that linearizes non-linear relationships contingent on a specified point. Assuming an interior-point solution exists, the mathematical optimization will be assured a global solution with linear constraints by the nature of the entropy objective function (Golan, et al. 1996, pp. 101). By specifying such linear approximations, the computational complexities of systems estimation with entropy will be greatly reduced.

REFERENCES

Aitken, A. C. (1935). "On Least Squares and Linear Combinations of Observations." Proceedings of the Royal Statistical Society 55: 42-48.

Amirizadeh, H. and R. M. Todd (1984). "More Growth Ahead for Ninth District States." Federal Reserve Bank of Minneapolis Quarterly Review(8): 8-17.

Anselin, L. (1980). Estimation Methods for Spatial Autoregressive Structures. Regional Science Dissertation and Monograph Series #8. Ithaca, NY.

Anselin, L. (1988a). "Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches." International Regional Science Review 11(3): 279-316.

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

Arndt, C., S. Robinson and F. Tarp (2002). "Parameter Estimation for a Computable General Equilibrium Model: A Maximum Entropy Approach." Economic Modeling 19(3): 375-398.

Bartik, T. J. (1991). Who Benefits from State and Local Economic Development Policies? Kalamazoo, W.E. Uphohn Institute.

Bartle, R. G. and D. R. Sherbert (1992). Introduction to Real Analysis: Second Edition. New York, Jon Wiley & Sons Inc.

Batey, P. W. J. and A. Z. Rose (1990). "Extended Input-Output Models: Progress and Potential." International Regional Science Review 13(1-2): 24-49.

Beeson, P. and R. Eberts (1989). "Identifying Productivity and Amenity Effects in Interurban Wage Differentials." Review of Economics and Statistics 71(443-452): 443-452.

Berndt, E. R. and D. O. Wood (1975). "Technology, Prices, and the Derived Demand for Energy." The Review of Economics and Statistics 57(3): 259-268.

Birkes, D. and Y. Dodge (1993). Alternative Methods of Regression. New York, John Wiley & Sons.

Blanchard, O. and D. Quah (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances." American Economic Review 79: 655-673.

Bolton, R. (1985). "Regional Econometric Models." Journal of Regional Science 25(4): 495-518.

Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis. Reading, MA, Addison-Wesley Pub Co.

Brooke, A., D. Kendrick, A. Meeraus, R. Raman and R. E. Rosenthal (1998). GAMS: A User's Guide. Cologne.

Brucker, S. M., S. E. Hastings and W. R. Latham, III (1990). "The Variation of Estimated Impacts from Five Regional Input-Output Models." International Regional Science Review 13(1-2): 119-139.

Burden, R. L. and J. D. Faires (1989). Numerical Analysis: Fourth Edition. Boston, PWS-KENT Publishing Company.

Canto, V. A. and R. I. Webb (1987). "The Effect of State Fiscal Policy on State Relative Economic Performance." Southern Economic Journal **54**(1): 186-202.

Center for Applied Economic Research (2001). Oklahoma Economic Outlook, CEAR:Oklahoma State University. 2001.

Chib, S. and E. Greenberg (1994). "Bayes Inference in Regression Models with ARMA (p,q) Errors." Journal of Econometrics 64: 183-206.

Crihfield, J. B. (1989). "A Structural Empirical Analysis of Metropolitan Labor Demand." Journal of Regional Science **29**(3): 347-371.

Crihfield, J. B. and M. P. H. Panggabean (1996). "The Structure of Metropolitan Factor and Product Markets." Journal of Regional Science **36**(1): 17-41.

Chung, J. W. (1994). Utility and Production Functions: Theory and Applications. Oxford, Cambridge, Blackwell.

Cochrane, D. and G. Orcutt (1949). "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms." Journal of the American Statistical Association 44: 32-61.

Cromwell, J. B. and M. J. Hannan (1993). "The Utility of Impulse Response Functions in Regional Analysis: Some Critical Issues." International Regional Science Review 15(2): 199-122.

Crone, T. M. and M. P. McLaughlin (1999). A Bayesian VAR Forecasting Model for the Philadelphia Metropolitan Area. Federal Reserve Bank of Philadelphia Research Working Paper: 99/07. Philadelphia.

Csisrár, I. (1975). "I-divergence Geometry of Probability Distributions and Minimization Problems." Annals of Probability 3(1): 146-158.

DeJong, D. N., B. F. Ingram and C. H. Whiteman (2000). "A Bayesian Approach to Dynamic Macroeconomics." Journal of Econometrics 98(2): 203-223.

Dejong, D., B. Ingram and C. H. Whiteman (1994). Beyond Calibration. Working Papers from University of Iowa - Department of Economics. Iowa City, Iowa.

Dejong, D., B. Ingram and C. H. Whiteman (1996). "A Bayesian Approach to Calibration." Journal of Business and Economic Statistics 14(1): 1-9.

Diebold, F. X. (1998). "The Past, Present, and Future of Macroeconomic Forecasting." Journal of Economic Perspectives 12(2): 175-192.

Diewert, W. E. and T. J. Wales (1987). "Flexible Functional Forms and Global Curvature Conditions." Econometrica 55(1): 43-68.

Doan, T., R. B. Litterman and C. Sims (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions." Econometric Reviews 3(1): 1-100.

Dorfman, J. H. (1997). Bayesian Economics through Numerical Methods: A Guide to Econometrics and Decision-Making with Prior Information. New York, Berlin, Heidelberg, Springer-Verlag.

Dowd, M. R. and J. P. LeSage (1997). "Analysis of Spatial Contiguity Influences on State Price Level Formation." International Journal of Forecasting 13(2): 245-53.

Engle, R. F. and B. S. Yoo (1987). "Forecasting and testing in co-integrated systems." Journal of Econometrics 35(1): 143-159.

Engle, R. F. and C. W. J. Granger (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing." Econometrica 55(2): 251-276.

Fair, R. C. and R. J. Shiller (1990). "Comparing Information in Forecasts from Econometric Models." The American Economic Review 80(3): 375-389.

Fawson, C. and K. R. Criddle (1994). "A Comparative Analysis of Time Series Approaches to Modeling Intersectoral and Intercounty Employment Linkages in Rural Regional Labor Markets." Journal of Regional Science 34(1): 57-74.

Friedman, M. (1953). The Methodology of Positive Economics. Essays in Positive Economics. M. Friedman. Chicago, University of Chicago Press: 18-35.

Fullerton, T. M., Jr. (2001). "Specification of a Borderplex Econometric Forecasting Model." International Regional Science Review 24(2): 245-60.

Gerking, S. D. (1976). "Input-Output as a Simple Econometric Model." The Review of Economics and Statistics 58(3): 274-282.

Geweke, J. (2001). "Bayesian Econometrics and Forecasting." Journal of Econometrics 100(1): 11-15.

Glennon, D. and J. Lane (1990). Input-Output Restriction, Regional Structural Models and Econometric Forecasts. New Directions In Regional Analysis: Integrated And Multi-Regional Approaches. L. Anselin and M. Madden. London, Bellhave Press: 133-145.

Glennon, D., J. Lane and S. Johnson (1987). "Regional Econometric Models That Reflect Labor Market Relations." International Journal of Forecasting 3(2): 299-312.

Glennon, D., J. Lane, S. Johnson and E. Robb (1986). "Incorporating Labour Market Structure in Regional Econometric Models." Applied Economics 18(5): 545-555.

Global Insights (2000). National Forecast Model: Jan. 2000, Global Insights. 2000.

Golan, A., G. G. Judge and D. J. Miller (1996). Maximum Entropy Econometrics: Robust Estimation with Limited Data. New York, John Wiley & Sons.

Golan, A., G. G. Judge and S. Robinson (1994). "Recovering Information from Incomplete or Partial Multisectoral Economic Data." The Review of Economics and Statistics 76(3): 541-549.

Golan, A., J. M. Perloff and E. Z. Shen (2001). "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand." The Review of Economics and Statistics 83(3): 541-550.

Good, I. J. (1963). "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables." Annals of Mathematical Statistics 34(3): 911-934.

Granger, C. W. J. (1986). "Developments in the Study of Cointegrated Economic Variables." Oxford Bulletin of Economics and Statistics 48: 213-227.

Greenwood, M. J. (1975). "Research on Internal Migration in the United States: A Survey." Journal of Economic Literature 13(2): 397-433.

Hansen, L. P. and J. J. Heckman (1996). "The Empirical Foundations of Calibration." The Journal of Economic Perspectives 10(1): 87-104.

Harris, J. R. and M. P. Todaro (1970). "Migration, Unemployment , and Development: A Two Sector Approach." American Economic Review 60: 126-42.

Hastings, S. E. and S. M. Brucker (1993). An Introduction to Regional Input-Output Analysis. Microcomputer-based input-output modeling: Applications to economic development. D. M. Otto and T. G. Johnson. Boulder and Oxford, Westview Press.

Hoehn, J. P., M. C. Berger and G. C. Bolmquist (1987). "A Hedonic Model of Interregional Wages, Rents, and Amenity Values." Journal of Regional Science 4(27): 605-620.

Isard, W. (1960). Methods of Regional Analysis; An Introduction to Regional Science. New York, Technology Press of the Massachusetts Institute of Technology and Wiley & Sons.

Isard, W., I. J. Azis, M. P. Drennan, R. E. Miller, S. Saltzman and E. Thorbecke (1998). Method of Interregional and Regional Analysis. Brookfield, Ashgate.

Jaynes, E. T. (1957a). "Information Theory and Statistical Mechanics." Physics Review(106): 620-630.

Jaynes, E. T. (1957b). "Information Theory and Statistical Mechanics II." Physics Review(108): 171-190.

Jeffreys, H. (1967). Theory of Probability. London, Oxford University Press.

Jones, R. and J. Whalley (1989). "A Canadian Regional General Equilibrium Model and Some Applications." Journal of Urban Economics 25: 368-404.

Jorgenson, D. W. (1984). Econometric Methods for Applied General Equilibrium Analysis. Applied General Equilibrium Analysis. J. B. Shoven. New York, Cambridge University Press.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl and T. Lee (1988). Introduction to the Theory and Practice of Econometrics: Second Edition. New York, John Wiley & Sons.

Kennedy, P. (1998). A Guide to Econometrics: Fourth Edition. Cambridge, MA, MIT Press.

Kim, K. and A. R. Pagan (1995). The Econometric Analysis of Calibrated Macroeconomic Models. Handbook Of Applied Econometrics. Volume 1. Macroeconomics. M. R. Wickens. Oxford and Malden, Blackwell.

Kinal, T. and J. Ratner (1986). "A VAR Forecasting Model of a Regional Economy: Its Construction and Comparative Accuracy." International Regional Science Review 10(2): 113-1126.

Klein, L. R. (1989). Econometric Aspects of Input-Output Analysis. Frontiers of Input-Output Analysis. R. E. Miller, K. R. Polenske and A. Z. Rose. New York, Oxford University Press. 3-11.

Kmenta, J. (1986). Elements of Econometrics. New York, London, Macmillan.

Kraybill, D. S. and J. H. Dorfman (1992). "A Dynamic Intersectoral Model of Regional Economic Growth." Journal of Regional Science 32(1): 1-16.

Kraybill, D. S., T. G. Johnson and D. Orden (1992). "Macro-Economic Imbalances: A Multiregional General Equilibrium Analysis." American Journal of Agricultural Economics 74: 726-736.

Krikelas, A. C. (1992). "Why Regions Grow: A Review of Research on the Economic Base Model." Federal Reserve Bank of Atlanta: Economic Review(July/August): 16-29.

Kullback, S. (1959). Information Theory and Statistics. New York, John Wiley.

Kydland, F. E. and E. C. Prescott (1982). "Time to Build and Aggregate Fluctuations." Econometrica 50(6): 1345-1370.

LeSage, J. P. (1989). "Incorporating Regional Wage Relations in Local Forecasting Models with a Bayesian Prior." International Journal of Forecasting 5(1): 37-47.

LeSage, J. P. (1990). "A Comparison of the Forecasting Ability of ECM and VAR Models." The Review of Economics and Statistics 72(4): 661-671.

LeSage, J. P. (1991). "Analysis and Development of Leading Indicators Using a Bayesian Turning-Points Approach." Journal of Business and Economic Statistics 9(3): 305-316.

LeSage, J. P. (1997). "Regression Analysis of Spatial Data." Journal of Regional Analysis and Policy 27(2): 83-94.

LeSage, J. P. and A. Krivelyova (1999). "A Spatial Prior for Bayesian Vector Autoregressive Models." Journal of Regional Science 39(2): 297-317.

LeSage, J. P. and J. D. Reed (1989). "The Dynamic Relationship Between Export, Local, and Total Area Employment." Regional Science and Urban Economics 19(4): 615-636.

LeSage, J. P. and M. Magura (1986). "Econometric Modeling Of Interregional Labor Market Linkages." Journal of Regional Science 26(3): 567-577.

LeSage, J. P. and M. Magura (1988). "A Regional Payroll Forecasting Model That Uses Bayesian Shrinkage Techniques for Data Pooling." Regional Science Perspectives 18(1): 100-116.

LeSage, J. P. and M. Magura (1991). "Using Interindustry Input-Output Relations as a Bayesian Prior in Employment Forecasting Models." International Journal of Forecasting 7(2): 231-238.

LeSage, J. P. and S. J. Rey (2002). Restrictions in Integrated Econometric-Input Output Models. Trade, Networks and Hierarchies. G. J. D. Hewings, M. Sonis and D. E. Boyce, Springer-Verlag: 251-256.

LeSage, J. P. and Z. Pan (1995). "Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models." International Regional Science Review 18(1): 33-53.

Litterman, R. B. (1980). A Bayesian Procedure for Forecasting with Vector Autoregression. Department of Economics, Massachusetts Institute of Technology. Cambridge.

Litterman, R. B. (1986). "Forecasting With Bayesian Vector Autoregressions-Five Years of Experience." Journal of Business and Economic Statistics 4(1): 25-38.

Litterman, R. B. (1986a). "Forecasting Accuracy of Alternative Techniques: A comparison of U.S. Macroeconomic forecasts: Comment." Journal of Business and Economic Statistics 4(1): 17-19.

Litterman, R. B. (1986b). "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience." Journal of Business and Economic Statistics 4(1): 25-38.

Litterman, R. B. (1986c). "A Statistical Approach to Economic Forecasting." Journal of Business and Economic Statistics 4(1): 1-4.

Liu, S.-I. (2002). "Bayesian Forecasts for Cointegrated Models." Journal of Forecasting 21(3): 167-180.

Magura, M. (1987). "The Use of Input-Output Tables in Specifying Interindustry and Interregional Labor Market Linkages." Papers of the Regional Science Association 62: 117-123.

Magura, M. (1990). Using Input-Output Data in a Bayesian Vector Autoregressive Forecasting Model. New Directions In Regional Analysis: Integrated And Multi-Regional Approaches. M. Madden and L. Anselin. London, Bellhaven Press: 133-145.

Magura, M. (1998). "IO and Spatial Information as Bayesian Priors in an Employment Forecasting Model." Annals of Regional Science 32(4): 495-503.

Mansur, A. and J. Whalley (1984). Numerical Specification of Applied General Equilibrium Models: Estimation, Calibration, and Data. Applied General Equilibrium Analysis. J. B. Shoven. New York, Cambridge University Press.

McGregor, P. G., J. K. Swales and Y. Ping (1996). "A Long-Run Interpretation of Regional Input-Output Analysis." Journal of Regional Science 36(3): 479-501.

McKitrick, R. R. (1995). The Econometric Critique of Computable General Equilibrium Modeling: The Role of Parameter Estimation. University of British Columbia, Department of Economics Discussion Paper.

McKitrick, R. R. (1998). "The Econometric Critique of Computable General Equilibrium Modeling: The Role of Functional Forms." Economic Modeling 15(4): 543-73.

McNees, S. K. (1986). "Forecasting Accuracy of Alternative Techniques: A comparison of U.S. Macroeconomic forecasts." Journal of Business and Economic Statistics 4(1): 5-15.

Moghadam, K. and K. P. Ballard (1987). "Small Area Modeling of the Industrial Sector (SAMIS): An Integrated Econometric-Interindustry Approach." Environment and Planning A 20(5): 655-668.

Moore, R. E. (1977). "A Test for Existence of Solutions to Nonlinear Systems." SIAM Journal on Numerical Analysis 14(4): 611-615.

Murtagh, B. A. and M. A. Saunders (1987). MINOS 5.1 User's Guide, Report SOL 83-20. Department of Operations Research, Stanford University.

Murtagh, B. A., M. A. Saunders, P. E. Gill, R. Raman and E. Kalvelagen (2002). MINOS: A Solver for Large-Scale Nonlinear Optimization Problems. Cologne.

Nocedal, J. and S. J. Wright (1999). Numerical Optimization. New York, Springer-Verlag New York Inc.

Ortega, J. M. (1972). Numerical Analysis-A Second Course. New York, Academic Press.

Otrok, C. and C. H. Whiteman (1997). What to Do When the Crystal Ball is Cloudy: Conditional and Unconditional Forecasting in Iowa. Institute for Economic Research: The University of Iowa. Iowa City: 14.

Otrok, C. and C. H. Whiteman (1998). "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa." International Economic Review 39(4): 997-1014.

Pan, Z. and J. P. LeSage (1995). "Using Spatial Contiguity as Prior Information in Vector Autoregressive Models." Economic Letters 47(2): 137-142.

Paris, Q. and R. E. Howitt (1998). "An Analysis of Ill-Posed Production Problems Using Maximum Entropy." American Journal of Agricultural Economics 80(1): 124-138.

Partridge, M. D. and D. S. Rickman (1988). "Regional Computable General Equilibrium Modeling: A Survey and Critical Appraisal." International Regional Science Review 21(3): 205-48.

Partridge, M. D. and D. S. Rickman (1998). "Generalizing the Bayesian Vector Autoregression Approach for Regional Interindustry Employment Forecasting." Journal of Business and Economic Statistics 16(1): 62-72.

Pereira, A. M. and J. B. Shoven (1988). "Survey of Dynamic Computational General Equilibrium Models for Tax Policy Evaluation." Journal of Policy Modeling 10: 401-36.

Plaut, T. R. (1981). "An Econometric Model for Forecasting Regional Population Growth." International Regional Science Review 6(1): 53-70.

Poirier, D. J. (1995). Intermediate Statistics and Econometrics: A Comparative Approach. Cambridge, London, MIT Press.

Press, S. J. (1989). Bayesian Statistics: Principles, Models, and Applications. New York, John Wiley & Sons.

Puri, A. and G. Soydemir (2000). "Forecasting Industrial Employment Figures in Southern California: A Bayesian Vector Autoregressive Model." Annals of Regional Science 34(4): 503-514.

Raftery, A., D. Madigan and J. Hoeting (1997). "Bayesian Model Averaging for Linear Regression Models." Journal of the American Statistical Association 92(1997): 179-191.

Raynauld, J. (1988). "Canadian Regional Cycles: The Quebec-Ontario Case Revisited." The Canadian Journal of Economics 20(1): 115-128.

Rey, S. J. (1997). "Integrating Regional Econometric and Input-Output Models: An Evaluation of Embedding Strategies." Environment and Planning A 29: 1057-1072.

Rey, S. J. (1998). "The Performance of Alternative Integration Strategies for Combining Regional Econometric and Input-Output Models." International Regional Science Review 21(1): 1-36.

Rey, S. J. (2000). "Integrated Regional Econometric+Input-Output Modeling: Issues and Opportunities." Papers in Regional Science 79(3): 271-292.

Richardson, H. W. (1972). Input-Output and Regional Analysis. New York, John Wiley & Sons.

Rickman, D. S. (1995). "A Bayesian Analysis of the Use of Pooled Coefficients in a Structural Regional Economic Model." International Journal of Forecasting 11(3): 447-90.

Rickman, D. S. (2001). "Using Input-Output Information for Bayesian Forecasting of Industry Employment in a Regional Econometric Model." International Regional Science Review 24(2): 226-44.

Rickman, D. S. (2002). "A Bayesian Forecasting Approach to Constructing Regional Input-Output Based Employment Multipliers." Papers in Regional Science 81(4): 483-498.

Rickman, D. S. and S. R. Miller (2002). "An Evaluation of Alternative Strategies for Incorporating Inter-Industry Relationships into a Regional Employment Forecasting Model." The Review of Regional Studies 32(1): 133-147.

Rickman, D. S. and R. K. Schwer (1993). "A Systematic Comparison of the REMI and IMPLAN models: The Case of Southern Nevada." Review of Regional Studies **23**: 143-161.

Rickman, D. S. and R. K. Schwer (1995). "A Comparison of the Multipliers of IMPLAN, REMI, and RIMS II: Benchmarking ready-Made Models for Comparison." Annals of Regional Science **29**: 363-374.

Roback, J. (1982). "Wages, Rents, and the Quality of Life." Journal of Political Economy 90: 1257-1278.

Roberts, B. M. (1994). "Calibration Procedure and the Robustness of CGE Models: Simulations with a Model for Poland." Economics and Planning 27(3): 189-210.

Robertson, J., E. W. Tallman and C. H. Whiteman (2002). Forecasting Using Relative Entropy. FRB of Atlanta Working Paper. Atlanta.

Romans, T. and G. Subrahmanyam (1979). "State and Local Taxes, Transfers and Regional Economic Growth." Southern Economic Journal **46**(2): 435-444.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." Bell System Technical Journal 27: 379-423.

Shoesmith, G. L. (1988). Prior Specification, Variable Choice and Accuracy in VAR Regional Forecasting: A Case Study. Wake Forest University, Babcock Graduate School of Management, Discussion Paper. Winston-Salem, NC.

Shoesmith, G. L. (1990). "The Forecasting Accuracy of Regional Bayesian VAR Models with Alternative National Variable Choices." International Regional Science Review 13(3): 257-69.

Shoesmith, G. L. (1992). "Cointegration, Error Correction and Medium-Term Regional VAR Forecasting." Journal of Forecasting 11: 91-109.

Shoven, J. B. and J. Whalley (1972). "A General Equilibrium Calculation of the Effects of Differential Taxation of Income from Capital in the U.S." Journal of Public Economics 1: 281-321.

Shoven, J. B. and J. Whalley (1984). "Applied General-Equilibrium Models of Taxation and International Trade: An Introduction and Survey." Journal of Economic Literature 22(3): 1007-1051.

Sims, C. A. (1980). "Macroeconomics and Reality." Econometrica 48(1): 1-49.

Sims, C. A. (1982). "Policy Analysis with Econometric Models." Brookings Papers on Economic Activity 1982(1): 107-152.

Sims, C. A. (1986). "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis Quarterly Review 10(1): 2-16.

Sinn, H.-W. (1980). "A Rehabilitation of the Principle of Insufficient Reason." The Quarterly Journal of Economics 94(3): 493-506.

Stock, J. H. and M. W. Watson (1989). New Indexes of Coincident and Leading Economic Indicators. NBER Macroeconomics Annual 1989. Cambridge, The MIT Press: 351-394.

Stock, J. H. and M. W. Watson (1992). A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Performance. Federal Reserve Bank of Chicago Working Paper. Chicago.

Stock, J. H. and M. W. Watson (2001). "Vector Autoregression." Journal of Economic Perspectives 15(101-116).

Sundaram, R. K. (1996). A First Course in Optimization Theory. Cambridge, Cambridge.

Sydsaeter, K., A. Strom and P. Berck (1999). Economists' Mathematical Manual. Berlin, Heidelberg, New York, Springer-Verlag.

Testa, W. A. (1989). Metro Area Growth from 1976 to 1985: Theory and Evidence. Working Paper. Federal Reserve Bank of Chicago, January.

Theil, H. (1966). Applied Econometric Forecasting. Amsterdam, North Holland.

Todaro, M. P. (1969). "A Model of Labor Migration and Urban Unemployment in Less Developed Countries." The American Economic Review 59(1): 138-148.

Todaro, M. P. (1970). "Labor Migration and Urban Unemployment: Reply." American Economic Review 60(1): 187-88.

Todd, R. M. (1984). "Improving Economic Forecasting with Bayesian Vector Autoregression." Federal Reserve Bank of Minneapolis Quarterly Review 8(4): 18-29.

Topel, R. (1986). "Local Labor Markets." Journal of Political Economy **94**: S111-S143.

Treyz, G. I. and B. H. Stevens (1985). "The TFS Regional Modeling Methodology." Regional Studies 19(6): 547-62.

Treyz, G. I., A. F. Friedlaender and B. H. Stevens (1980). "The Employment Sector of a Regional Economic Policy Simulation Model." Review of Economics and Statistics 62(1): 63-73.

Treyz, G. I., B. H. Stevens and D. J. Ehrlich (1981). A State Core Forecasting and Policy Simulation Model. National Academy of Science: National Cooperative Highway Research Program. Washington D.C.

Treyz, G. I., D. S. Rickman and G. Shao (1992). "The REMI Economic-Demographic Forecasting and Simulation Model." International Regional Science Review 14(3): 221-53.

U.S. Department of Commerce, Bureau of the Census. (1977) Census of Business. Washington, Government Printing Office. 1997.

Vinod, H. D. (1978). "A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares." The Review of Economics and Statistics 60(1): 121-131.

Wackerly, D. D., William Mendenhall III and R. L. Scheaffer (2002). Mathematical Statistics with Applications: Sixth Ed. Pacific Grove, CA, Duxbury.

Walsh, C. E. (2003). Monetary Theory and Policy: Second Edition. Cambridge, MIT Press.

Wasylenko, M. J. (1988). Economic Development in Nebraska. Nebraska Comprehensive Tax Study Staff Paper No. 1, Metropolitan Studies Program, The Maxwell School, Syracuse University, revised.

West, C. T. and H. Theil (1991). "An Empirical Analysis of Stein Effects in Regional Forecasts." Empirical Economics 16(3): 363-73.

West, G. R. (1995). "Comparison of Input-Output, Input-Output +Econometric and Computable General Equilibrium Impact Models at the Regional Level." Economic Systems Research 7(2): 209-27.

West, G. R. and R. W. Jackson (1998). "Input-Output + Econometric and Econometric + Input-Output: Model Differences or Different Models?" Journal of Regional Analysis and Policy 28(1): 33-48.

Zellner, A. (1971). An Introduction to Bayesian Inference In Econometrics. New York, John Wiley & Sons.

Zellner, A. (1979). "Statistical Analysis of Econometric Models." Journal of the American Statistical Association 74(367): 628-643.

Zellner, A. (1985). "Bayesian Econometrics." Econometrica 53(2): 253-270.

APPENDIX I

MODEL ESTIMATION EQUATIONS

$$L_{it} = VA_{it} \cdot epv_{it} = \left(\left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}\right)rpc_{it} + X_{it}\right)\lambda_{VA,it} \cdot epv_{it} \quad (4.1)$$

$$INT_{it} = \sum_{i \in N} a_{ij} Q_{jt} \quad (4.2)$$

$$C_{it} = pce_i \cdot \gamma_C \frac{RYD_t}{RYD_t^U} C_t^U \quad (4.3)$$

$$I_{it} = inv_{IR_i} \cdot \gamma_{IR} \frac{RYD_t}{RYD_t^U} IR_t^U + inv_{NR_i} \cdot \gamma_{NR} \frac{GSP_t}{GDP_t} INR_t^U \quad (4.4)$$

$$FED_{it} = fgov_i \cdot \gamma_{FED} \cdot FED_t^U \quad (4.5)$$

$$SLG_{it} = gov_i \cdot \gamma_{SLG} \frac{N_t}{N_t^U} SLG_t^U \quad (4.6)$$

$$X_{it} = S_i^N \cdot RAC\left(\frac{1}{RAC_t}\right) \cdot VA_{it}^U + S_i^W \cdot X_t^U \quad (4.7)$$

$$L_{it} = \left(\phi_{FDi,1}\left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}\right)rpc_{it} + \phi_{FDi,2} \cdot X_{it}\right)\lambda_{VA,it} \cdot epv_{it} \quad (4.8)$$

$$L_t = \sum_{i \in N} L_{it} \quad (4.9)$$

$$L_{it} - \hat{L}_{it} = le_{it} \quad (4.10)$$

$$rpc_{it} = rpc_i + \phi_{rpc,1} \cdot time_t + \phi_{rpc,2} \cdot AC_t \quad (4.11)$$

$$\lambda_{VA,it} = \lambda_{VA,i} + \phi_{VA1i} \cdot time_t + \phi_{VA2i} \cdot time_t^2 \quad (4.12)$$

$$VA_{it} = \frac{1}{epv_{it}} \cdot L_{it} \quad (4.13)$$

$$VA_t = \sum_{i \in N} VA_{it} \quad (4.14)$$

$$VA_{it} - \hat{VA}_{it} = ve_{it} \quad (4.15)$$

$$epv_{it} = RF_{it}^{-1} \cdot RLC_{it}^{\alpha_i - 1} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i} \cdot epv_{it}^{U} + rf_{it} \quad \text{....................} (4.16)$$

$$RF_{it} = \frac{epv_{it}^{U}}{epv_{it}} RLC_{it}^{\alpha-1} \cdot RCC_{it}^{\beta} \cdot RFC_{it}^{\gamma} \quad \text{....................} (4.17)$$

$$RLC_{it} = W_{it} / W_{it}^{U} \quad \text{....................} (4.18)$$

$$RAC_{it} = \left(1 / RF_{it} \right) \cdot RLC_{it}^{\alpha_i} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i} \quad \text{....................} (4.19)$$

$$W_{it} = \phi_{Wi} + \phi_{Wi,REO} REO_t + \phi_{Wi,CP} CP_t + \phi_{Wi,W^U} W_{it}^{U} + \phi_{Wi,time} time_t + \phi_{Wi,lag} W_{it-1} \quad \text{....................} (4.20)$$

$$W_{it} - \hat{W}_{it} = we_{it} \quad \text{....................} (4.21)$$

$$W_t = \sum_{i \in N} W_{it} L_{it} / L_t \quad \text{....................} (4.22)$$

$$YP_t = WSD_t + YOL_t + YPROP_t - TWPER_t \quad \text{....................} (4.23)$$

$$WSD_t = W_t \cdot L_t / 1{,}000 \quad \text{....................} (4.24)$$

$$WSD_{it} = W_{it} \cdot L_{it} / 1{,}000 \quad \text{....................} (4.25)$$

$$YOL_t = \frac{YOL / L}{YOL^U / L^U} \cdot \left(L_t / L_t^U \right) \cdot YOL_t^U \quad \text{....................} (4.26)$$

$$YPROP_t = \frac{YPROP / N}{YPROP^U / N^U} \cdot \left(N_t / N_t^U \right) \cdot YPROP_t^U \quad \text{....................} (4.27)$$

$$TWPER_t = \frac{TWPER / WSD}{TWPER^U / WSD^U} \cdot \left(WSD_t / WSD_t^U \right) \cdot TWPER_t^U \quad \text{....................} (4.28)$$

$$YD_t = YP_t - NTAX_t \quad \text{....................} (4.29)$$

$$NTAX_t = TAX_t - TRAN_t \quad \text{....................} (4.30)$$

$$TAX_t = \frac{TAX / WSD}{TAX^U / WSD^U} \cdot \left(WSD_t / WSD_t^U \right) \cdot TAX_t^U \quad \text{....................} (4.31)$$

$$TRAN_t = \frac{TRAN / N}{TRAN^U / N^U} \cdot \left(N_t / N_t^U \right) \cdot TRAN_t^U \quad (4.32)$$

$$RYD_t = YD_t / cpi_t \quad (4.33)$$

$$cpi_t = \sum_{i \in N} \left(FD_i / FD \cdot Sp_{it} \right) \cdot cpi_t^U \quad (4.34)$$

$$Sp_{it} = \begin{cases} 1 & \text{if } i = \text{goods producing} \\ RAC_{it} & \text{if } i = \text{nongoods producing} \end{cases} \quad (4.35)$$

$$CP_t = cpi_t / cpi_t^U \quad (4.36)$$

$$RWR_t = W_t / W_t^U \quad (4.37)$$

$$REO_t = \frac{L_t / N_t}{L_t^U / N_t^U} \quad (4.38)$$

$$N_t = \phi_{N0} + \phi_{N,REO} \cdot REO_t + \phi_{N,RWR} \cdot RWR_t + \phi_{N,time} \cdot time_t + \phi_{N,Nlag} \cdot N_{t-1} \quad (4.39)$$

$$N_t - \hat{N}_t = ne_t \quad (4.40)$$

$$le_{it} = \sum_{j \in J} v_{Lit,j} w_{Lit,j} \quad (4.41)$$

$$ve_{it} = \sum_{j \in J} v_{VAit,j} w_{VAit,j} \quad (4.42)$$

$$we_{it} = \sum_{j \in J} v_{Wit,j} w_{Wit,j} \quad (4.43)$$

$$ne_t = \sum_{j \in J} v_{Nt,j} w_{Nt,j} \quad (4.44)$$

APPENDIX II

GAMS ESTIMATION PROGRAM

```
$eolcom #
set ful             "time"              / 1980 * 2010 /;
set smpl(ful)       "sample period"     / 1987 * 1999 /;
set hist(ful)       "history"           / 1980 * 1999 /;
set fcst(ful)       "forecast period" / 2000 * 2010 /;
set m               "parameter support set" / 1*3 /;
set j               "error support set"     / 1*3 /;
set ind             "industry ident"        / G N /;
set k               "parameters to be estimated"
/
         LG_g1*LG_g2
         LN_g1*LN_g2
         WG_g0*WG_g2
         WN_g0*WN_g2
         WG_g3*WG_g5
         WN_g3*WN_g5
         N_g0*N_g4
         lvaG_g1*lvaG_g2
         lvaN_g1*lvaN_g2
         rpcg_g1*rpcg_g2
         rpcn_g1*rpcn_g2
 /;

table z(k,m) "parameter support"
                         1          2           3
(LG_g1*LG_g2)          0.75       1.00        1.25
(LN_g1*LN_g2)          0.75       1.00        1.25
(WG_g0)               -2000       0.00        2000
(WN_g0)               -2000       0.00        2000
(WG_g1)               -0150       0.00        0150
(WN_g1)               -0150       0.00        0150
(WG_g2)               -0150       0.00        0150
(WN_g2)               -0150       0.00        0150
(WG_g3)                 0         1.00          2
(WN_g3)                 0         1.00          2
(WG_g4)               -2500       0.00        2500
(WN_g4)               -2500       0.00        2500
(WG_g5)                 0.0       0.50         1.0
(WN_g5)                 0.0       0.50         1.0
N_g0               -8000000       0.00      8000000
N_g1                  -4000       0.00        8000
N_g2                  -4000       0.00        8000
N_g3                  -8000       0000        8000
N_g4                  -2.00       0.00        2.00
(lvaG_g1*lvaG_g2)     -0.10       0.00        0.10
(lvaN_g1*lvaN_g2)     -0.10       0.00        0.10
rpcg_g1               -0.025     -0.015       0.02
rpcg_g2               -0.10      -0.05        0.00
rpcn_g1               -0.025     -0.015       0.02
rpcn_g2               -0.10      -0.05        0.00
;
```

```
table qq(k,m) "parameter probability priors"
                           1           2           3
(LG_g1*LG_g2)            0.05        0.90        0.05
(LN_g1*LN_g2)            0.05        0.90        0.05
(WG_g0*WG_g2)            0.15        0.70        0.15
(WN_g0*WN_g2)            0.15        0.70        0.15
(WG_g3*WG_g5)            0.10        0.80        0.10
(WN_g3*WN_g5)            0.10        0.80        0.10
(N_g0)                   0.10        0.80        0.10
(N_g1)                   0.10        0.80        0.10
(N_g2)                   0.10        0.80        0.10
(N_g3)                   0.10        0.80        0.10
(N_g4)                   0.10        0.80        0.10
(lvaG_g1*lvaG_g2)        0.15        0.70        0.15
(lvaN_g1*lvaN_g2)        0.15        0.70        0.15
(rpcg_g1*rpcg_g2)        0.10        0.80        0.10
(rpcn_g1*rpcn_g2)        0.10        0.80        0.10
;

set eqn "equations to estimate with entropy"
/
         L_G
         L_N
         W_G
         W_N
         ne
         vaeg
         vaen
         rpcg
         rpcn
/
;
table v(eqn,smpl,j)  "error support set"
                                1        2        3
(L_G).(1987*1999)           -2800        0       2800
(L_N).(1987*1999)           -2800        0       2800
(W_G).(1987*1999)           -13000       0       13000
(W_N).(1987*1999)           -10000       0       10000
(ne).(1987*1999)            -65000       0       65000
(vaeg).(1987*1999)         -900000      0.00     900000
(vaen).(1987*1999)         -900000      0.00     900000
(rpcg).(1987*1999)         -.01200      0.00     .01200
(rpcn).(1987*1999)         -.01200      0.00     .01200
;

table u(eqn,smpl,j) "error pobability priors"
                              1          2          3
(L_G).(1987*1999)            .10        .80        .10
(L_N).(1987*1999)            .10        .80        .10
(W_G).(1987*1999)            .10        .80        .10
(W_N).(1987*1999)            .10        .80        .10
(ne).(1987*1999)             .10        .80        .10
(vaeg).(1987*1999)           .10        .80        .10
(vaen).(1987*1999)           .10        .80        .10
(rpcg).(1987*1999)           .020       .96        .020
(rpcn).(1987*1999)           .020       .96        .020
;

*Calculated Sets
$include "c:\stevestuff\dis\gams\start.prn";
*National set and table of parameters
$include "c:\stevestuff\dis\gams\national.prn"  ;
*Regional set and table of parameters
```

```
$include "c:\stevestuff\dis\gams\regional.PRN"  ;
scalar bob1;
scalar bob2;
bob1=(p_L_G/regional('L_G','1999'));
bob2=(p_L_N/regional('L_N','1999'));
display bob1, bob2;

regional('N',smpl)=regional('N',smpl)*(p_pop/regional('N','1999'));
national('N',smpl)=national('N',smpl)*(p_pop_u/national('N','1999'));
regional('L_G',smpl)=regional('L_G',smpl)*(p_L_G/regional('L_G','1999'))
;
regional('L_N',smpl)=regional('L_N',smpl)*(p_L_N/regional('L_N','1999'))
;
regional('L',smpl)=regional('L_G',smpl)+regional('L_N',smpl);
national('L_G',smpl)=national('L_G',smpl)*(p_L_G_u/national('L_G','1999'
));
national('L_N',smpl)=national('L_N',smpl)*(p_L_N_u/national('L_N','1999'
));
national('L',smpl)=national('L_G',smpl)+national('L_N',smpl);
regional('VA_G',smpl)=regional('VA_G',smpl)*(p_VA_G/regional('VA_G','199
9'));
regional('VA_N',smpl)=regional('VA_N',smpl)*(p_VA_N/regional('VA_N','199
9'));
regional('VA',smpl)=regional('VA_G',smpl)+regional('VA_N',smpl);
national('VA_G',smpl)=national('VA_G',smpl)*(p_VA_G_u/national('VA_G','1
999'));
national('VA_N',smpl)=national('VA_N',smpl)*(p_VA_N_u/national('VA_N','1
999'));
national('VA',smpl)=national('VA_G',smpl)+national('VA_N',smpl);
regional('epv_G',smpl)=regional('L_G',smpl)/regional('VA_G',smpl);
regional('epv_N',smpl)=regional('L_N',smpl)/regional('VA_N',smpl);
national('epv_G',smpl)=national('L_G',smpl)/national('VA_G',smpl);
national('epv_N',smpl)=national('L_N',smpl)/national('VA_N',smpl);
regional('W',smpl)=(regional('L_G',smpl)*regional('W_G',smpl)+

regional('L_N',smpl)*regional('W_N',smpl))/regional('L',smpl);
regional('WSD_G',smpl)=regional('W_G',smpl)*regional('L_G',smpl)/1000;
regional('WSD_N',smpl)=regional('W_N',smpl)*regional('L_N',smpl)/1000;
regional('WSD',smpl)=regional('WSD_G',smpl)+regional('WSD_N',smpl);
regional('PY',smpl)=regional('WSD',smpl)+regional('YOL',smpl)+regional('
YDIR',smpl)-regional('TWPER',smpl);
regional('YD',smpl)=regional('PY',smpl)-
regional('TAX',smpl)+regional('TRAN',smpl);
regional('cpi',smpl)=national('cpi',smpl)*(1.65090214/national('cpi','19
99'));
regional('RYD',smpl)=(regional('YD',smpl)/regional('cpi',smpl));
display p_cpi;
display regional;

$include "c:\stevestuff\dis\gams\nat_pp.prn"
$include "c:\stevestuff\dis\gams\reg_pp.prn"
$include "c:\stevestuff\dis\gams\inverstime.prn"
display national,regional,invtime;
$include "c:\stevestuff\dis\gams\vatest1.txt";
*########################################################################
*Run fixed exogenous variables
$include "c:\stevestuff\dis\gams\finaldemand.ins";
$include "c:\stevestuff\dis\gams\nfinaldemand1.ins";
p_LamVA_G=regional('VA_G','1999')/NFD('Q_G','1999');
p_LamVA_N=regional('VA_N','1999')/NFD('Q_N','1999');
display p_LamVA_G,p_LamVA_N;
*########################################################################
vatest('LamVA_G',smpl)=regional('VA_G',smpl)/NFD('Q_G',smpl);
vatest('LamVA_N',smpl)=regional('VA_N',smpl)/NFD('Q_N',smpl);
```

```
DISPLAY VATEST;
vatest('LamVA_G',smpl)=(regional('VA_G',smpl)/NFD('Q_G',smpl))*(p_LamVA_
G/vatest('LamVA_G','1999'));
vatest('LamVA_N',smpl)=(regional('VA_N',smpl)/NFD('Q_N',smpl))*(p_LamVA_
N/vatest('LamVA_N','1999'));
display vatest;
*#######################################################################
$include "c:\stevestuff\dis\gams\production.ins";
$include "c:\stevestuff\dis\gams\py.ins";

scalar bnd;
bnd=.12;
v('L_G',smpl,'1')=-bnd*regional('L_G',smpl);
v('L_N',smpl,'1')=-bnd*regional('L_N',smpl);
v('L_G',smpl,'3')=bnd*regional('L_G',smpl);
v('L_N',smpl,'3')=bnd*regional('L_N',smpl);
v('L_G','1999','3')=0;
v('L_N','1999','3')=0;
v('L_G','1999','1')=0;
v('L_N','1999','1')=0;
bnd=.12;
v('W_G',smpl,'1')=-bnd*regional('W_G',smpl);
v('W_N',smpl,'1')=-bnd*regional('W_N',smpl);
v('W_G',smpl,'3')=bnd*regional('W_G',smpl);
v('W_N',smpl,'3')=bnd*regional('W_N',smpl);
v('W_G','1999','3')=0;
v('W_N','1999','3')=0;
v('W_G','1999','1')=0;
v('W_N','1999','1')=0;
v('ne','1999','1')=0;
v('ne','1999','3')=0;
*v('L_G','1987','1')=-.10*regional('L_G','1987');
*v('L_G','1987','1')= .10*regional('L_G','1987');
DISPLAY V;

########################################################################

variables
$ontext
         REOL(smpl)
         REOLL(smpl)
         RWRL(smpl)
         RWRLL(smpl)
         NL(smpl)
         NLL(smpl)
$offtext
         INT_G(smpl)
         INT_N(smpl)
         GSL_G(smpl)
         GSL_N(smpl)
         GF_G(smpl)
         GF_N(smpl)
         LD_G(smpl)
         LD_N(smpl)
         X_G(smpl)
         X_N(smpl)
         C_G(smpl)
         C_N(smpl)
         I_G(smpl)
         I_N(smpl)

         YPROP(ful)
         TWPER(ful)
         TRAN(ful)
```

```
        TAX(ful)
        NTAX(ful)
        YOL(ful)
        YP(ful)
        YD(ful)
        RYD(ful)
        WSD(ful)
        WSD_G(ful)
        WSD_N(ful)


        VA_G(smpl)
        VA_N(smpl)
        VA(smpl)
        L_G(smpl)
        L_N(smpl)
        L(smpl)
        LamVA_G(smpl)
        LamVA_N(smpl)
        epv_G(smpl)
        epv_N(smpl)
        RF_G(smpl)
        RF_N(smpl)
        AC_G(smpl)
        AC_N(smpl)
        sP_G(smpl)
        sP_N(smpl)
        cpi(smpl)
        CP(smpl)
        rpc_G(smpl)
        rpc_N(smpl)
        RLC_G(smpl)
        RLC_N(smpl)
        REO(smpl)
        RWR(smpl)

        W_G(smpl)
        W_N(smpl)
        W(smpl)
        WLag_G(smpl)
        WLag_N(smpl)
        N(smpl)
        NL(smpl)

        obj
        p(k,m)             parameter support space probabilities
        ww(eqn,smpl,j)      error support space probabilities
;
bnd=.20;

RLC_G.l(smpl)=pro('RLC_G',smpl);
RLC_G.lo(smpl)=.75*pro('RLC_G',smpl);
RLC_G.up(smpl)=1.25*pro('RLC_G',smpl);
RLC_N.l(smpl)=pro('RLC_N',smpl);
RLC_N.lo(smpl)=.75*pro('RLC_N',smpl);
RLC_N.up(smpl)=1.25*pro('RLC_N',smpl);
REO.l(smpl)=pro('REO',smpl);
REO.lo(smpl)=0.75*pro('REO',smpl);
REO.up(smpl)=1.25*pro('REO',smpl);

RWR.l(smpl)=pro('RWR',smpl);
RWR.lo(smpl)=.5;
```

```
CP.l(smpl)=pro('CP',smpl);
CP.lo(smpl)=.01;
rpc_G.l(smpl)=pro('rpc_G',smpl);
rpc_N.l(smpl)=pro('rpc_N',smpl);

rpc_G.lo(smpl)=.8*pro('rpc_G',smpl)$(.8*pro('rpc_G',smpl) gt
0.00)+0.1$(.8*pro('rpc_G',smpl) le 0.00);
rpc_N.lo(smpl)=.8*pro('rpc_N',smpl)$(.8*pro('rpc_N',smpl) gt
0.00)+0.1$(.8*pro('rpc_N',smpl) le 0.00);
rpc_G.up(smpl)=1.2*pro('rpc_G',smpl)$(1.20*pro('rpc_G',smpl) lt
1.00)+1.00$(1.20*pro('rpc_G',smpl) gt 1.00);
rpc_N.up(smpl)=1.2*pro('rpc_G',smpl)$(1.20*pro('rpc_N',smpl) lt
1.00)+1.00$(1.20*pro('rpc_N',smpl) gt 1.00);
rpc_G.lo(smpl)=.25;
rpc_N.lo(smpl)=.25;
rpc_G.up(smpl)=.99;
rpc_N.up(smpl)=.99;
rpc_G.fx('1999')=p_rpc_g;
rpc_N.fx('1999')=p_rpc_n;
LamVA_G.l(smpl)=pro('LamVA_G',smpl);
LamVA_N.l(smpl)=pro('LamVA_N',smpl);
LamVA_G.lo(smpl)=.75*pro('LamVA_G',smpl);
LamVA_N.lo(smpl)=.75*pro('LamVA_N',smpl);
LamVA_G.up(smpl)=1.25*pro('LamVA_G',smpl);
LamVA_N.up(smpl)=1.25*pro('LamVA_N',smpl);
*LamVA_G.fx('1999')=p_lamVA_G;
*LamVA_N.fx('1999')=p_lamVA_N;
epv_G.l(smpl)=regional('epv_G',smpl);
epv_N.l(smpl)=regional('epv_N',smpl);
epv_G.lo(smpl)=.75*regional('epv_G',smpl);
epv_N.lo(smpl)=.75*regional('epv_N',smpl);
epv_G.up(smpl)=1.25*regional('epv_G',smpl);
epv_N.up(smpl)=1.25*regional('epv_N',smpl);
VA_G.l(smpl)=regional('VA_G',smpl);
VA_N.l(smpl)=regional('VA_N',smpl);
VA.l(smpl)=regional('VA',smpl);
VA_G.lo(smpl)=.80*regional('VA_G',smpl);
VA_N.lo(smpl)=.80*regional('VA_N',smpl);
VA.lo(smpl)=.80*regional('VA',smpl);
VA_G.up(smpl)=1.20*regional('VA_G',smpl);
VA_N.up(smpl)=1.20*regional('VA_N',smpl);
VA.up(smpl)=1.20*regional('VA',smpl);
L_G.l(smpl)=regional('L_G',smpl);
L_N.l(smpl)=regional('L_N',smpl);


L.l(smpl)=regional('L',smpl);
L_G.lo(smpl)=(1-bnd)*regional('L_G',smpl);
L_N.lo(smpl)=(1-bnd)*regional('L_N',smpl);
L.lo(smpl)=(1-bnd)*regional('L',smpl);
L_G.up(smpl)=(1+bnd)*regional('L_G',smpl);
L_N.up(smpl)=(1+bnd)*regional('L_N',smpl);
L.up(smpl)=(1+bnd)*regional('L',smpl);
bnd=.25;
W_G.l(smpl)=regional('W_G',smpl);
W_N.l(smpl)=regional('W_N',smpl);
WLag_G.l(smpl)=regional('W_G',smpl);
WLag_N.l(smpl)=regional('W_N',smpl);
W.l(smpl)=regional('W',smpl);
N.l(smpl)=regional('N',smpl);
NL.l(smpl)=regional('N',smpl);
NL.lo(smpl)=0.75*regional('N',smpl);
NL.up(smpl)=1.25*regional('N',smpl);
```

```
cpi.l(smpl)=pro('cpi',smpl);
cpi.lo(smpl)=.1;

AC_G.lo(smpl)=.75*pro('AC_G',smpl);
AC_N.lo(smpl)=.75*pro('AC_N',smpl);
AC_G.up(smpl)=1.25*pro('AC_G',smpl);
AC_N.up(smpl)=1.25*pro('AC_N',smpl);
AC_G.l(smpl)=pro('AC_G',smpl);
AC_N.l(smpl)=pro('AC_N',smpl);
RF_G.l(smpl)=pro('RF_G',smpl);
RF_N.l(smpl)=pro('RF_N',smpl);
RF_G.lo(smpl)=.55;
RF_N.lo(smpl)=.55;
RF_G.up(smpl)=1.45;
RF_N.up(smpl)=1.45;

INT_G.l(smpl)=GFD('INT_G',smpl);
INT_N.l(smpl)=GFD('INT_N',smpl);
C_G.l(smpl)=GFD('C_G',smpl);
C_N.l(smpl)=GFD('C_N',smpl);
I_G.l(smpl)=GFD('I_G',smpl);
I_N.l(smpl)=GFD('I_N',smpl);
GSL_G.l(smpl)=GFD('GSL_G',smpl);
GSL_N.l(smpl)=GFD('GSL_N',smpl);
GF_G.l(smpl)=GFD('GF_G',smpl);
GF_N.l(smpl)=GFD('GF_N',smpl);
X_G.l(smpl)=GFD('X_G',smpl);
X_N.l(smpl)=GFD('X_N',smpl);

WSD.l(smpl)=regional('WSD',smpl);
WSD_G.l(smpl)=regional('WSD_G',smpl);
WSD_N.l(smpl)=regional('WSD_N',smpl);
WSD.lo(smpl)=(1-bnd)*regional('WSD',smpl);
WSD_G.lo(smpl)=(1-bnd)*regional('WSD_G',smpl);
WSD_N.lo(smpl)=(1-bnd)*regional('WSD_N',smpl);
WSD.up(smpl)=(1+bnd)*regional('WSD',smpl);
WSD_G.up(smpl)=(1+bnd)*regional('WSD_G',smpl);
WSD_N.up(smpl)=(1+bnd)*regional('WSD_N',smpl);
YPROP.l(smpl)=pinc('YPROP',smpl);
TWPER.l(smpl)=pinc('TWPER',smpl);
TRAN.l(smpl)=pinc('TRAN',smpl);
TAX.l(smpl)=pinc('TAX',smpl);
YOL.l(smpl)=pinc('YOL',smpl);
YP.l(smpl)=pinc('YP',smpl);
NTAX.l(smpl)=pinc('NTAX',smpl);
YD.l(smpl)=pinc('YD',smpl);
RYD.l(smpl)=pinc('RYD',smpl);
N.lo(smpl)=.95*(regional('N',smpl));
N.up(smpl)=1.05*(regional('N',smpl));

L_G.fx('1999')=regional('L_G','1999');
L_N.fx('1999')=regional('L_N','1999');
*epv_G.fx('1999')=regional('epv_G','1999');
*epv_N.fx('1999')=regional('epv_N','1999');
VA_G.fx('1999')=regional('VA_G','1999');
VA_N.fx('1999')=regional('VA_N','1999');
RF_G.fx('1999')=pro('RF_G','1999');
RF_N.fx('1999')=pro('RF_N','1999');

p.lo(k,m)=.0001;          p.up(k,m)=.999;            p.l(k,m)=1/3;
ww.lo(eqn,smpl,j)=.0001;  ww.up(eqn,smpl,j)=.999;
ww.l(eqn,smpl,j)=1/3;

variables
```

```
          ne(smpl)
*          le(smpl)
          eewsd(smpl)
          vaeg(smpl)
          vaen(smpl)
*not entropied
          rfg(smpl)
          rfn(smpl)
          err(smpl)
          rpcg(smpl)
          rpcn(smpl)
          fudgeg(smpl)
          fudgen(smpl)
          ex1
          ex2
;
rfg.lo(smpl)=-.025;
rfn.lo(smpl)=-.025;
rfg.up(smpl)= .025;
rfn.up(smpl)= .025;
err.lo(smpl)=-800;
err.up(smpl)= 800;
*le.up(smpl)= 260;
*le.lo(smpl)=-260;
fudgeg.lo(smpl)=-.1;
fudgen.lo(smpl)=-.1;
fudgeg.up(smpl)= .1;
fudgen.up(smpl)= .1;
rpcg.lo(smpl)=-.020;
rpcn.lo(smpl)=-.020;
rpcg.up(smpl)= .020;
rpcn.up(smpl)= .020;

ex1.lo=-1.5;
ex2.lo=-1.5;
ex1.up=0;
ex2.up=0;
equations
          tNL(smpl)
          tWLag_G(smpl)
          tWLag_N(smpl)

          eINT_G(smpl)
          eINT_N(smpl)
          eGSL_G(smpl)
          eGSL_N(smpl)
          eGF_G(smpl)
          eGF_N(smpl)
*          eLD_G(smpl)
*          eLD_N(smpl)
          eX_G(smpl)
          eX_N(smpl)
          eC_G(smpl)
          eC_N(smpl)
          eI_G(smpl)
          eI_N(smpl)

          eYPROP(ful)
          eTWPER(ful)
          eTRAN(ful)
          eTAX(ful)
          eYOL(ful)
          eYP(ful)
          eNTAX(ful)
```

```
          eYD(ful)
          eRYD(ful)
          eWSD(ful)
          eWSD_G(ful)
          eWSD_N(ful)
          pWSD(smpl)


          eepv_G(smpl)
          eepv_N(smpl)
          eVA_G(smpl)
          eVA_N(smpl)
          eVA(smpl)
*          wL_G(smpl)
*          wL_N(smpl)
          pL_G(smpl)
          pL_N(smpl)
          pL(smpl)
          pRF_G(smpl)
          pRF_N(smpl)
          eAC_G(smpl)
          eAC_N(smpl)
          esP_G(smpl)
          esP_N(smpl)
          ecpi(smpl)
          eCP(smpl)
          erpc_G(smpl)
          erpc_N(smpl)
          eRLC_G(smpl)
          eRLC_N(smpl)
          eREO(smpl)
          eRWR(smpl)

          pW_G(smpl)
          pW_N(smpl)
          pW(smpl)
          pN(smpl)
          pLamVA_G(smpl)
          pLamVA_N(smpl)



          fne(smpl)
          wne(smpl)
*          fewsd(smpl)
*          fle(smpl)
          fl_g(smpl)
          fl_n(smpl)
          fva_g(smpl)
          fva_n(smpl)
          fW_G(smpl)
          fW_N(smpl)

          objective        objective function to max
          add1(k)          parameter additivity constraint
          add2(eqn,smpl)   error additivitiy constraint
;

*####################################################################
tNL(smpl)..     NL(smpl)=e=regional('N','1986')$(ord(smpl) eq 1)+N(smpl-
1)$(not ord(smpl) eq 1);
tWLag_G(smpl)..   WLag_G(smpl)=e=regional('W_G','1986')$(ord(smpl) eq
1)+W_G(smpl-1)$(not ord(smpl) eq 1);
```

```
tWLag_N(smpl)..   WLag_N(smpl)=e=regional('W_N','1986')$(ord(smpl) eq
1)+W_N(smpl-1)$(not ord(smpl) eq 1);

eINT_G(smpl)..   INT_G(smpl)=e=
p_agg*(VA_G(smpl)/LamVA_G(smpl))+p_agn*(VA_N(smpl)/LamVA_N(smpl));
eINT_N(smpl)..   INT_N(smpl)=e=
p_ang*(VA_G(smpl)/LamVA_G(smpl))+p_ann*(VA_N(smpl)/LamVA_N(smpl));

eC_G(smpl)..     C_G(smpl)  =e=
p_pce_G*p_gamma_c*((RYD(smpl)/national('RYD',smpl))*national('CONS',smpl
));
eC_N(smpl)..     C_N(smpl)  =e=
p_pce_N*p_gamma_c*((RYD(smpl)/national('RYD',smpl))*national('CONS',smpl
));
eI_G(smpl)..     I_G(smpl)  =e=
p_invr_G*p_gamma_IR*(RYD(smpl)/national('RYD',smpl))*national('IR',smpl)

+p_invnr_G*p_gamma_INR*(VA(smpl)/national('VA',smpl))*national('INR',smp
l);
eI_N(smpl)..     I_N(smpl)  =e=
p_invr_N*p_gamma_IR*(RYD(smpl)/national('RYD',smpl))*national('IR',smpl)

+p_invnr_N*p_gamma_INR*(VA(smpl)/national('VA',smpl))*national('INR',smp
l);
eGSL_G(smpl)..   GSL_G(smpl)=e=
p_gov_G*p_gamma_gsl*(N(smpl)/national('N',smpl))*national('GSL',smpl);
eGSL_N(smpl)..   GSL_N(smpl)=e=
p_gov_N*p_gamma_gsl*(N(smpl)/national('N',smpl))*national('GSL',smpl);
eGF_G(smpl)..    GF_G(smpl) =e=
p_fgov_G*p_gamma_fed*national('GF',smpl);
eGF_N(smpl)..    GF_N(smpl) =e=
p_fgov_N*p_gamma_fed*national('GF',smpl);

eX_G(smpl)..     X_G(smpl)  =e=
p_S_G_N*((pro('AC_G','1999')/AC_G(smpl)))*national('VA_G',smpl)
                                  +p_S_G_W*national('EX',smpl);
eX_N(smpl)..     X_N(smpl)  =e=
p_S_N_N*((pro('AC_N','1999')/AC_N(smpl)))*national('VA_N',smpl)
                                  +p_S_N_W*national('EX',smpl);
*########################################################################
eYPROP(smpl)..   YPROP(smpl)=e=
((regional('YDIR','1999')/regional('N','1999'))/

(national('YPROP','1999')/national('N','1999')))

*(national('YPROP',smpl)/national('N',smpl))*N(smpl);
eTWPER(smpl)..   TWPER(smpl)=e=
((regional('TWPER','1999')/regional('WSD','1999'))

/(national('TWPER','1999')/national('WSD','1999')))

*(national('TWPER',smpl)/national('WSD',smpl))*WSD(smpl);
eTRAN(smpl)..    TRAN(smpl) =e=
((regional('TRAN','1999')/regional('N','1999'))
                         /(national('VP','1999')/national('N','1999')))

*(national('VP',smpl)/national('N',smpl))*N(smpl);
eTAX(smpl)..     TAX(smpl)  =e=
((regional('TAX','1999')/regional('WSD','1999'))

/(national('TAX','1999')*1000000/national('WSD','1999')))

*(national('TAX',smpl)*1000000/national('WSD',smpl))*WSD(smpl);
```

```
eYOL(smpl)..      YOL(smpl)  =e=
((regional('YOL','1999')/regional('L','1999'))

/(national('YOL','1999')/national('L','1999')))

*(national('YOL',smpl)/national('L',smpl))*L(smpl);
eYP(smpl)..       YP(smpl)   =e= WSD(smpl)+YOL(smpl)+YPROP(smpl)-
TWPER(smpl);
eNTAX(smpl)..     NTAX(smpl) =e= TAX(smpl)-TRAN(smpl);
eYD(smpl)..       YD(smpl)   =e= YP(smpl)-NTAX(smpl);
eRYD(smpl)..      RYD(smpl)  =e= YD(smpl)/cpi(smpl);
eWSD_G(smpl)..    WSD_G(smpl)=e= ((W_G(smpl)*L_G(smpl))/1000);
eWSD_N(smpl)..    WSD_N(smpl)=e= ((W_N(smpl)*L_N(smpl))/1000);
eWSD(smpl)..      WSD(smpl)  =e= W(smpl)*L(smpl)/1000;
pWSD(smpl)..      WSD(smpl)  =e= WSD_G(smpl)+WSD_N(smpl);
*######################################################################
eRLC_G(smpl)..    RLC_G(smpl)=e= (W_G(smpl)/national('W_G',smpl));
eRLC_N(smpl)..    RLC_N(smpl)=e= (W_N(smpl)/national('W_N',smpl));
eREO(smpl)..      REO(smpl)  =e=
(L(smpl)/N(smpl))/(national('L',smpl)/national('N',smpl));
eRWR(smpl)..      RWR(smpl)  =e= W(smpl)/national('W',smpl);
$ontext
eAC_G(smpl)..     AC_G(smpl) =e=
(1/RF_G(smpl))*(RLC_G(smpl)**(p_alpha_G));                #rlc into ac
causes problem
eAC_N(smpl)..     AC_N(smpl) =e=
(1/RF_N(smpl))*(RLC_N(smpl)**(p_alpha_N));
$offtext
eAC_G(smpl)..     AC_G(smpl)*RF_G(smpl) =e= (RLC_G(smpl)**(p_alpha_G));#
+rfg(smpl);
eAC_N(smpl)..     AC_N(smpl)*RF_N(smpl) =e= (RLC_N(smpl)**(p_alpha_N));#
+rfn(smpl);
*######################################################################
esP_G(smpl)..     sP_G(smpl) =e= 1;
esP_N(smpl)..     sp_N(smpl) =e= AC_N(smpl);
ecpi(smpl)..      cpi(smpl)  =e=
((p_FD_G/p_FD)*sP_G(smpl)+(p_FD_N/p_FD)*sP_N(smpl))*national('cpi',smpl)
;
eCP(smpl)..       CP(smpl)   =e= cpi(smpl)/national('cpi',smpl);
*######################################################################
pRF_G(smpl)..
RF_G(smpl)*epv_G(smpl)=e=(national('epv_G',smpl))*RLC_G(smpl)**(p_alpha_
G-1)        +rfg(smpl);
pRF_N(smpl)..
RF_N(smpl)*epv_N(smpl)=e=(national('epv_N',smpl))*RLC_N(smpl)**(p_alpha_
N-1)        +rfn(smpl);
$offtext
pRF_G(smpl)..
RF_G(smpl)=e=((national('epv_G',smpl))*RLC_G(smpl)**(p_alpha_G-
1))/epv_G(smpl)      +rfg(smpl);
pRF_N(smpl)..
RF_N(smpl)=e=((national('epv_N',smpl))*RLC_N(smpl)**(p_alpha_N-
1))/epv_N(smpl)      +rfn(smpl);
*######################################################################
pW_G(smpl)..      (W_G(smpl))=e=(sum(m,z('WG_g0',m)*p('WG_g0',m)))

+((REO(smpl))*(sum(m,z('WG_g1',m)*p('WG_g1',m))))

+((CP(smpl))*(sum(m,z('WG_g2',m)*p('WG_g2',m))))

+((national('W_G',smpl))*(sum(m,z('WG_g3',m)*p('WG_g3',m))))
                         +((national('time',smpl)-
1986)*(sum(m,z('WG_g4',m)*p('WG_g4',m))))
```

```
+((WLag_G(smpl))*(sum(m,z('WG_g5',m)*p('WG_g5',m))))
;
pW_N(smpl)..      (W_N(smpl))=e=(sum(m,z('WN_g0',m)*p('WN_g0',m)))

+((REO(smpl))*(sum(m,z('WN_g1',m)*p('WN_g1',m))))

+((CP(smpl))*(sum(m,z('WN_g2',m)*p('WN_g2',m))))

+((national('W_N',smpl))*(sum(m,z('WN_g3',m)*p('WN_g3',m))))
                          +((national('time',smpl)-
1986)*(sum(m,z('WN_g4',m)*p('WN_g4',m))))

+((WLag_N(smpl))*(sum(m,z('WN_g5',m)*p('WN_g5',m))))
;
pW(smpl)..        W(smpl)
=e=(W_G(smpl)*L_G(smpl)+W_N(smpl)*L_N(smpl))/L(smpl)
+err(smpl);

*##########################################################################
pN(smpl)..        N(smpl)    =e=  sum(m,z('N_g0',m)*p('N_g0',m))

+((L(smpl)/regional('N',smpl))/(national('L',smpl)/national('N',smpl)))*
sum(m,z('N_g1',m)*p('N_g1',m))
                                 +
(RWR(smpl))*sum(m,z('N_g2',m)*p('N_g2',m))
                                 + (national('time',smpl)-
1986)*sum(m,z('N_g3',m)*p('N_g3',m))
                                 +
NL(smpl)*sum(m,z('N_g4',m)*p('N_g4',m));

wne(smpl)..       ne(smpl)    =e= regional('N',smpl)-N(smpl);
*##########################################################################
pLamVA_G(smpl).. LamVA_G(smpl)=e=p_LamVA_G

+sum(m,z('lvaG_g1',m)*p('lvaG_g1',m))*(national('time',smpl)-1999)

+sum(m,z('lvaG_g2',m)*p('lvaG_g2',m))*sqr(national('time',smpl)-1999)
;
pLamVA_N(smpl).. LamVA_N(smpl)=e=p_LamVA_N

+sum(m,z('lvaN_g1',m)*p('lvaN_g1',m))*(national('time',smpl)-1999)

+sum(m,z('lvaN_g2',m)*p('lvaN_g2',m))*sqr(national('time',smpl)-1999)
;

*##########################################################################

erpc_G(smpl).. rpc_G(smpl)=e=p_rpc_G

+sum(m,z('rpcg_g1',m)*p('rpcg_g1',m))*(national('time',smpl)-1986)
                   +sum(m,z('rpcg_g2',m)*p('rpcg_g2',m))*(AC_G(smpl))
                   +sum(j,v('rpcg',smpl,j)*ww('rpcg',smpl,j));
;
erpc_N(smpl).. rpc_N(smpl)=e=p_rpc_N

+sum(m,z('rpcn_g1',m)*p('rpcn_g1',m))*(national('time',smpl)-1986)
                   +sum(m,z('rpcn_g2',m)*p('rpcn_g2',m))*(AC_N(smpl))
                   +sum(j,v('rpcn',smpl,j)*ww('rpcn',smpl,j));
;
*##########################################################################
fne(smpl)..      ne(smpl)     =e= sum(j,v('ne',smpl,j)*ww('ne',smpl,j));
fVA_G(smpl)..    regional('VA_G',smpl)-
VA_G(smpl)=e=sum(j,v('vaeg',smpl,j)*ww('vaeg',smpl,j));
```

```
fVA_N(smpl)..   regional('VA_N',smpl)-
VA_N(smpl)=e=sum(j,v('vaen',smpl,j)*ww('vaen',smpl,j));
fL_G(smpl)..   regional('L_G',smpl)-
L_G(smpl)=e=sum(j,v('L_G',smpl,j)*ww('L_G',smpl,j));
fL_N(smpl)..   regional('L_N',smpl)-
L_N(smpl)=e=sum(j,v('L_N',smpl,j)*ww('L_N',smpl,j));
fW_G(smpl)..    regional('W_G',smpl)-
W_G(smpl)=e=sum(j,v('W_G',smpl,j)*ww('W_G',smpl,j));
fW_N(smpl)..    regional('W_N',smpl)-
W_N(smpl)=e=sum(j,v('W_N',smpl,j)*ww('W_N',smpl,j));

*######################################################################
pL_G(smpl)..    L_G(smpl)=e=(

+sum(m,z('LG_g1',m)*p('LG_g1',m))*(INT_G(smpl)+C_G(smpl)+I_G(smpl)+GSL_G
(smpl)+GF_G(smpl))
                   *rpc_G(smpl)

+sum(m,z('LG_g2',m)*p('LG_g2',m))*(X_G(smpl)))*LamVA_G(smpl)
                   *epv_G(smpl)
;
pL_N(smpl)..    L_N(smpl)=e=(

+sum(m,z('LN_g1',m)*p('LN_g1',m))*(INT_N(smpl)+C_N(smpl)+I_N(smpl)+GSL_N
(smpl)+GF_N(smpl))
                   *rpc_N(smpl)

+sum(m,z('LN_g2',m)*p('LN_g2',m))*(X_N(smpl)))*LamVA_N(smpl)
                   *epv_N(smpl)
;
pL(smpl)..      L(smpl)=e=L_G(smpl)+L_N(smpl);
*######################################################################
$ontext
eepv_G(smpl)..   epv_G(smpl)=e=(1/RF_G(smpl))*(RLC_G(smpl)**(p_alpha_G-
1))*national('epv_G',smpl)  +rfg(smpL);
eepv_N(smpl)..   epv_N(smpl)=e=(1/RF_N(smpl))*(RLC_N(smpl)**(p_alpha_N-
1))*national('epv_N',smpl)  +rfn(smpl);
$offtext
eepv_G(smpl)..   epv_G(smpl)=e=L_G(smpl)/VA_G(smpl)  +rfg(smpL);
eepv_N(smpl)..   epv_N(smpl)=e=L_N(smpl)/VA_N(smpl)  +rfn(smpl);
*######################################################################
eVA_G(smpl)..    VA_G(smpl)=e=L_G(smpl)/epv_G(smpl);
eVA_N(smpl)..    VA_N(smpl)=e=L_N(smpl)/epv_N(smpl);
eVA(smpl)..      VA(smpl)=e=VA_G(smpl)+VA_N(smpl);

*######################################################################
objective..      obj        =e= 200*
(1*(sum(k,sum(m,p(k,m)*log(p(k,m)/qq(k,m))))

+1*sum(eqn,sum(smpl,sum(j,ww(eqn,smpl,j)*log(ww(eqn,smpl,j)/u(eqn,smpl,j
))))) ) );
add1(k)..           sum(m,p(k,m))  =e=  1;
add2(eqn,smpl)..    sum(j,ww(eqn,smpl,j))  =e=  1;

*######################################################################

model ent / all /;
option nlp=minos;
option domlim=1000000;
option iterlim=900000000;
*option sysout=on;
ent.workspace=150;
ent.reslim=60000;
*ent.optfile=1;
```

```
ent.scaleopt=1;
solve ent minimizing obj using nlp;
*#########################################################################
options decimals=8;
parameter betap(k)
          errorsp(eqn,smpl)
          betam(k)
          errorsm(eqn,smpl)
          e_hat_G(smpl)
          e_hat_N(smpl);
betap(k)=sum(m,p.l(k,m)*z(k,m));
errorsp(eqn,smpl)=sum(j,ww.l(eqn,smpl,j)*v(eqn,smpl,j));
betam(k)=sum(m,p.m(k,m)*z(k,m));
errorsm(eqn,smpl)=sum(j,ww.m(eqn,smpl,j)*v(eqn,smpl,j));
display V,Z;
display p.l,p.m,ww.l,ww.m;
display betap, betam;
display errorsp, errorsm;
```

APPENDIX III

ORDER OF MODEL EQUATIONS FOR GUASS-SEIDEL ALGORYTHM

$$\lambda_{VA,it} = \lambda_{VA,i} + \phi_{VAi,time} \cdot time_t + \phi_{VAi,time2} \cdot time_t^2 \quad (4.12)$$

$$X_{it} = S_i^N \cdot RAC\left(\frac{1}{RAC_t}\right) \cdot VA_{it}^U + S_i^W \cdot X_t^U \quad (4.7)$$

$$FED_{it} = fgov_i \cdot \gamma_{FED} \cdot FED_t^U \quad (4.5)$$

$$SLG_{it} = gov_i \cdot \gamma_{SLG} \frac{N_t}{N_t^U} SLG_t^U \quad (4.6)$$

$$I_{it} = inv_{IR_i} \cdot \gamma_{IR} \frac{RYD_t}{RYD_t^U} IR_t^U + inv_{NR_i} \cdot \gamma_{NR} \frac{GSP_t}{GDP_t} INR_t^U \quad (4.4)$$

$$INT_{it} = \sum_{j \in N} a_{ij} \frac{VA_{jt}}{\lambda_{VAj,t}}, \quad (4.2)$$

$$C_{it} = pce_i \cdot \gamma_C \frac{RYD_t}{RYD_t^U} C_t^U \quad (4.3)$$

$$L_{it} = \left(\phi_{FDi,1}\left(INT_{it} + C_{it} + I_{it} + SLG_{it} + FED_{it}\right) rpc_{it} + \phi_{FDi,2} \cdot X_{it}\right) \lambda_{VA,it} \cdot epv_{it} \quad (4.8)$$

$$L_t = \sum_{i \in N} L_{it} \quad (4.9)$$

$$epv_{it} = \frac{epv_{it}^U}{RF_{it}} RLC_{it}^{\alpha_i - 1} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i}, \quad (4.16)$$

$$VA_{it} = \frac{1}{epv_{it}} \cdot L_{it} \quad (4.13)$$

$$VA_t = \sum_{i=1}^{N} VA_{it} \quad (4.14)$$

$$W_{it} = \phi_{Wi} + \phi_{Wi,REO} REO_t + \phi_{Wi,CP} CP_t + \phi_{Wi,W^U} W_{it}^U + \phi_{Wi,time} time_t + \phi_{Wi,time} W_{it-1} \quad (4.20)$$

$$W_t = \sum_{i \in N} {W_{it} L_{it}} \Big/ {L_t} \text{ ......} (4.22)$$

$$WSD_{it} = {W_{it} \cdot L_{it}} \Big/ {1{,}000} \text{ ......} (4.25)$$

$$WSD_t = \sum_{i \in N} WSD_{it} \text{ ......} (4.24)$$

$$RLC_{it} = {W_{it}} \Big/ {W_{it}^U} \text{ ......} (4.18)$$

$$RWR_t = {W_t} \Big/ {W_t^U} \text{ ......} (4.37)$$

$$RF_{it} = \frac{epv_{it}^U}{epv_{it}} RLC_{it}^{\alpha_i - 1} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i} \text{ ......} (4.17)$$

$$RAC_{it} = \left(1 \Big/ RF_{it} \right) \cdot RLC_{it}^{\alpha_i} \cdot RCC_{it}^{\beta_i} \cdot RFC_{it}^{\gamma_i} \text{ ......} (4.19)$$

$$Sp_{it} = \begin{cases} 1 & \text{if } i = \text{goods producing} \\ RAC_{it} & \text{if } i = \text{nongoods producing} \end{cases} \text{ ......} (4.35)$$

$$rpc_{it} = rpc_i + \phi_{rpc,1} \cdot time_t + \phi_{rpc,2} \cdot AC_t \text{ ......} (4.11)$$

$$cpi_t = \sum_{i \in N} \left({FD_i} \Big/ {FD} \cdot Sp_{it} \right) \cdot cpi_t^U \text{ ......} (4.34)$$

$$CP_t = {cpi_t} \Big/ {cpi_t^U} \text{ ......} (4.36)$$

$$W_{it} = \phi_{Wi} + \phi_{Wi,REO} REO_t + \phi_{Wi,CP} CP_t + \phi_{Wi,W^U} W_{it}^U + \phi_{Wi,time} time_t + \phi_{Wi,lag} W_{it-1} \text{ ......} (4.20)$$

$$W_t = \sum_{i \in N} {W_{it} L_{it}} \Big/ {L_t} \text{ ......} (4.22)$$

$$REO_t = \frac{L_t / N_t}{L_t^U / N_t^U} \text{ ......} (4.38)$$

$$YPROP_t = \frac{YPROP \Big/ N}{YPROP^U \Big/ N^U} \cdot \left({YPROP_t^U} \Big/ {N_t^U} \right) \cdot N_t \text{ ......} (4.27)$$

$$TWPER_t = \frac{TWPER \Big/ WSD}{TWPER^U \Big/ WSD^U} \cdot \left({TWPER_t^U} \Big/ {WSD_t^U} \right) \cdot WSD_t \text{ ......} (4.28)$$

$$TRAN_t = \frac{TRAN/N}{TRAN^U/N^U} \cdot \left(TRAN_t^U / N_t^U \right) \cdot N_t \quad \text{(4.30)}$$

$$TAX_t = \frac{TAX/WSD}{TAX^U/WSD^U} \cdot \left(TAX_t^U / WSD_t^U \right) \cdot WSD_t \quad \text{(4.31)}$$

$$YOL_t = \frac{YOL/L}{YOL^U/L^U} \cdot \left(YOL_t^U / L_t^U \right) \cdot L_t \quad \text{(4.26)}$$

$$YP_t = WSD_t + YOL_t + YPROP_t - TWPER_t \quad \text{(4.23)}$$

$$NTAX_t = TAX_t - TRAN_t \quad \text{(4.30)}$$

$$YD_t = YP_t - NTAX_t \quad \text{(4.29)}$$

$$RYD_t = YD_t / cpi_t \quad \text{(4.33)}$$

APPENDIX IV

MATLAB PROJECTION PROGRAM

```
clear all


load blow32_11_d_todaro.mat
%blow32e_test11 todaro d
LG_g1  =      1.17115392
LG_g2  =      1.18510548
LN_g1  =      1.055396962
LN_g2  =      0.989519074
WG_g0  =      173.8923752
WG_g1  =      0.494450216
WG_g2  =      5.169576145
WN_g0  =      -831.9319289
WN_g1  =      -8.108580925
WN_g2  =      -2.737893979
WG_g3  =      0.127846229
WG_g4  =      -86.44292651
WG_g5  =      0.916158812
WN_g3  =      0.823968122
WN_g4  =      -243.3424988
WN_g5  =      0.096589167
N_g0   =      825553.1598
N_g1   =      589.4294556
N_g2   =      370.3819609
N_g3   =      7953.400684
N_g4   =      0.73516576
lvaG_g1       =      -0.006471975
lvaG_g2       =      7.36292E-05
lvaN_g1       =      0.0049251
lvaN_g2       =      0.000270035
rpcg_g1       =      0.002570914
rpcg_g2       =      -0.022445243
rpcn_g1       =      0.000817961
rpcn_g2       =      -0.000055


flag_pol=1;
flag_rates=1;
policy=ones(6,24);
policy(4:5,:)=0;

%output multilpiers
%policy(4,14:14)=1000;   %Goods
%policy(5,14:14)=1000;   %Goods

% TAX MULTIPLIERS in percents
%  policy(6,14:24)=1.05;

% COST RATIOS in percents
% %RLC
%  policy(1,14:24)=1.10;
% %RCC
%  policy(2,14:24)=1.05;
% %RFC
%  policy(3,14:24)=1.05;
nation=national;
regional=initial;
```

```
for yr=14:24

sse=100000;
count=1;
while sse>.000000010;                  %inner loop
if count==1
    if yr~=1
    regional(:,yr)=regional(:,yr-1);
    progress(:,count)=regional(:,yr);
    end
end
z0=regional(:,yr);          %Begining values

%################################################################################
regional(22,yr) = rparam(42) + lvaG_g1*(nation(41,yr)-1999) +
lvaG_g2*((nation(41,yr)-1999)^2);           %LamVA_G
    regional(23,yr) = rparam(43) + lvaN_g1*(nation(41,yr)-1999) +
lvaN_g2*((nation(41,yr)-1999)^2);           %LamVA_N
%################################################################################
regional(51,yr) = rparam(64)*(rparam(80)/regional(1,yr))*nation(44,yr)  +
rparam(65)*nation(13,yr)...
        + policy(4,yr);
%X_G
    regional(52,yr) = rparam(66)*(rparam(81)/regional(2,yr))*nation(45,yr)  +
rparam(67)*nation(13,yr)...
        + policy(5,yr);
%X_N
%################################################################################
regional(9,yr)  = rparam(20)*rparam(23)*nation(15,yr);
%GF_G
    regional(10,yr) = rparam(21)*rparam(23)*nation(15,yr);
%GF_N
%################################################################################
regional(11,yr) =
rparam(29)*rparam(25)*(regional(26,yr)/nation(28,yr))*nation(16,yr);
%GSL_G
    regional(12,yr) =
rparam(30)*rparam(25)*(regional(26,yr)/nation(28,yr))*nation(16,yr);
%GSL_N
%################################################################################
regional(13,yr) =
rparam(33)*rparam(27)*(regional(36,yr)/nation(36,yr))*nation(19,yr)...
        +  rparam(31)*rparam(26)*(regional(42,yr)/nation(43,yr))*nation(18,yr);
%I_G
    regional(14,yr) =
rparam(34)*rparam(27)*(regional(36,yr)/nation(36,yr))*nation(19,yr)...
        +  rparam(32)*rparam(26)*(regional(42,yr)/nation(43,yr))*nation(18,yr);
%I_N
%################################################################################
regional(15,yr) =
rparam(1)*(regional(43,yr)/regional(22,yr))+rparam(2)*(regional(44,yr)/regional(
23,yr));%INT_G
    regional(16,yr) =
rparam(5)*(regional(43,yr)/regional(22,yr))+rparam(6)*(regional(44,yr)/regional(
23,yr));%INT_N
%################################################################################
regional(3,yr)  =
(rparam(48))*(rparam(22))*((regional(36,yr)/nation(36,yr))*nation(1,yr));
%C_G
    regional(4,yr)  =
(rparam(49))*(rparam(22))*((regional(36,yr)/nation(36,yr))*nation(1,yr));
%C_N
%################################################################################
    regional(20,yr) =
(LG_g1*(regional(15,yr)+regional(3,yr)+regional(13,yr)+regional(11,yr)+regional(
9,yr))...

*regional(33,yr)+LG_g2*(regional(51,yr)))*regional(22,yr)*regional(7,yr);
%L_G
    regional(21,yr) =
(LN_g1*(regional(16,yr)+regional(4,yr)+regional(14,yr)+regional(12,yr)+regional(
10,yr))...
```

```
*regional(34,yr)+LN g2*(regional(52,yr)))*regional(23,yr)*regional(8,yr);
%L_N
    regional(19,yr) = regional(20,yr)+regional(21,yr);
%L
%################################################################################
    regional(7,yr)  =
nation(11,yr)*(1/regional(29,yr))*(regional(31,yr)^(rparam(3)-
1))*(policy(2,yr)^rparam(7))...
        *(policy(3,yr)^rparam(24));
%epv_G
    regional(8,yr)  =
nation(12,yr)*(1/regional(30,yr))*(regional(32,yr)^(rparam(4)-
1))*(policy(2,yr)^rparam(8))...
        *(policy(3,yr)^rparam(28));
%epv_N
%################################################################################
    regional(43,yr) = regional(20,yr)/regional(7,yr);
%VA_G
    regional(44,yr) = regional(21,yr)/regional(8,yr);
%VA_N
    regional(42,yr) = regional(43,yr)+regional(44,yr);
%VA
%################################################################################
    if yr==1
            regional(46,yr) = WG_g0 + WG_g1*regional(28,yr) +
WG_g2*regional(5,yr) + WG_g3*nation(48,yr)...
                + WG_g4*(nation(41,yr)-1986) + WG g5*regional(46,yr);
%W_G  first year
            regional(47,yr) = WN_g0 + WN_g1*regional(28,yr) +
WN_g2*regional(5,yr) + WN_g3*nation(49,yr)...
                + WN_g4*(nation(41,yr)-1986) + WN_g5*regional(47,yr);
%W_N  first year
    else
            regional(46,yr) = WG_g0 + WG_g1*regional(28,yr) +
WG_g2*regional(5,yr) + WG_g3*nation(48,yr)...
                + WG_g4*(nation(41,yr)-1986) + WG_g5*regional(46,yr-1);
%W_G  other years
            regional(47,yr) = WN_g0 + WN_g1*regional(28,yr) +
WN_g2*regional(5,yr) + WN_g3*nation(49,yr)...
                + WN_g4*(nation(41,yr)-1986) + WN_g5*regional(47,yr-1);
%W_N  other years
    end
    regional(45,yr) = (regional(46,yr)*regional(20,yr) +
regional(47,yr)*regional(21,yr))/regional(19,yr);
%W
%################################################################################
    regional(49,yr) = ((regional(46,yr)*regional(20,yr))/1000);
%WSD_G
    regional(50,yr) = ((regional(47,yr)*regional(21,yr))/1000);
%WSD_N
    regional(48,yr) =   regional(49,yr)+regional(50,yr);
%WSD
%################################################################################
    regional(31,yr) = (regional(46,yr)/nation(48,yr))*policy(1,yr);
%RLC_G
    regional(32,yr) = (regional(47,yr)/nation(49,yr))*policy(1,yr);
%RLC_N
%################################################################################
    regional(35,yr) = regional(45,yr)/nation(47,yr);
%RWR
%################################################################################
regional(29,yr) = (nation(11,yr)/regional(7,yr))*(regional(31,yr)^(rparam(3)-
1))*(policy(2,yr)^rparam(7))...
        *(policy(3,yr)^rparam(24));
%RF_G
    regional(30,yr) =
(nation(12,yr)/regional(8,yr))*(regional(32,yr)^(rparam(4)-
1))*(policy(2,yr)^rparam(8))...
        *(policy(3,yr)^rparam(28));
%RF_N
%################################################################################
```

```
    regional(1,yr)  =
(1/regional(29,yr))*(regional(31,yr)^rparam(3))*(policy(2,yr)^rparam(7))...
        *(policy(3,yr)^rparam(24));
%AC_G
    regional(2,yr)  =
(1/regional(30,yr))*(regional(32,yr)^rparam(4))*(policy(2,yr)^rparam(8))...
        *(policy(3,yr)^rparam(28));
%AC_N
%################################################################################
    regional(37,yr) = 1;
%SP_G
    regional(38,yr) = regional(2,yr);
%SP_N
%################################################################################
regional(33,yr) = rparam(60) + rpcg_g1*(nation(41,yr)-1986) +
rpcg_g2*regional(1,yr);                     %rpc_G
    regional(34,yr) = rparam(61) + rpcn_g1*(nation(41,yr)-1986) +
rpcn_g2*regional(2,yr);                     %rpc_N
%################################################################################
regional(6,yr)  =
((rparam(17)/rparam(19))*regional(37,yr)+(rparam(18)/rparam(19))*regional(38,yr)
)*...
        nation(2,yr);
%CPI
    regional(5,yr)  = regional(6,yr)/nation(2,yr);
%CP
%################################################################################
    if yr==1
        regional(26,yr) = N_g0 + N_g1*regional(28,yr) + N_g2*regional(35,yr) +
N_g3*(nation(41,yr)-1986)...
            + N_g4*regional(26,yr);
%N first year
    else
        regional(26,yr) = N_g0 + N_g1*regional(28,yr) + N_g2*regional(35,yr) +
N_g3*(nation(41,yr)-1986)...
            + N_g4*regional(26,yr-1);
%N other years
    end
%################################################################################
regional(28,yr) =
(regional(19,yr)/regional(26,yr))/(nation(20,yr)/nation(28,yr));
%REO
%################################################################################
regional(58,yr) =
((initial(58,13)/initial(26,13))/(nation(62,13)/nation(28,13)))*...
        (nation(62,yr)/nation(28,yr))*regional(26,yr);
%YPROP
    regional(41,yr) =
((initial(41,13)/initial(48,13))/(nation(42,13)/nation(50,13)))*...
        (nation(42,yr)/nation(50,yr))*regional(48,yr);
%TWPER
    regional(40,yr) =
((initial(40,13)/initial(26,13))/(nation(46,13)/nation(28,13)))*...
        (nation(46,yr)/nation(28,yr))*regional(26,yr);
%TRAN
    regional(39,yr) =
((initial(39,13)/initial(48,13))/(nation(37,13)/nation(50,13)))*...
        (nation(37,yr)/nation(50,yr))*regional(48,yr).*policy(6,yr);
%TAX
    regional(56,yr) =
((initial(56,13)/initial(19,13))/(nation(58,13)/nation(20,13)))*...
        (nation(58,yr)/nation(20,yr))*regional(19,yr);
%YOL
    regional(57,yr) = regional(48,yr)+regional(56,yr)+regional(58,yr)-
regional(41,yr);                        %YP
    regional(27,yr) = (regional(39,yr)-regional(40,yr));
%NTAX
    regional(55,yr) = regional(57,yr)-regional(27,yr);
%YD
    regional(36,yr) = regional(55,yr)/regional(6,yr);
%RYD
%################################################################################
```

```
z1=regional(:,yr);            %ending values
ze=z0-z1;
sse=ze'*ze;
count=count+1;
progress(:,count)=regional(:,yr);
if count>1000;
    break
end
end
[yr count-1]
end

if flag_pol==1;
    load baseline.mat;
       regional(59:61,:)=0;
       regional(60,:)=regional(43,:)./regional(22,:);
       regional(61,:)=regional(44,:)./regional(23,:);
       regional(59,:)=regional(60,:)+regional(61,:);
    chg1=regional-base;
    if flag_rates==1;
        chg1=chg1./base;
    end
    chg2=cumsum(chg1,2);
    chg=cat(1,chg1,zeros(1,24),chg2);
    wk1write('C:\stevestuff\dis\change.wk1',chg,1,1);
end

out=regional;
out2=((out(:,2:yr)-out(:,1:yr-1))./out(:,1:yr-1))*100;
[rows cols]=size(out2);
out2=cat(2,zeros(rows,1),out2);
[rows cols]=size(out2);
out2=cat(1,zeros(1,cols),out2);
out=cat(1,out,out2);
err=((regional(1:58,1:13)-initial)./initial)*100;
wk1write('C:\stevestuff\dis\out.wk1',out,1,1);
wk1write('C:\stevestuff\dis\regionalout.wk1',regional,1,1);
wk1write('C:\stevestuff\dis\error.wk1',err,1,1);
'done'
```

APPENDIX V: EVIEWS PROGRAMS FOR VAR MODELS

```
subroutine in_sample

smpl 1987 1999
 equation eW_G.ls @pch(W_G) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))  @pch(W_G_U) '@pch(N(-1))
 equation eW_N.ls @pch(W_N) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))  @pch(W_N_U)  '@pch(N(-1))

 equation eL_G.ls @pch(L_G) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))   @pch(L_G_U) '@pch(N(-1))
 equation eL_N.ls @pch(L_N) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))  @pch(L_N_U) ' @pch(N(-1))

 equation eVA_G.ls @pch(VA_G) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))  @pch(VA_G_U) ar(1)  '@pch(N(-1))
 equation eVA_N.ls @pch(VA_N) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1))
@pch(L_N(-1)) @pch(VA_G(-1)) @pch(VA_N(-1))  @pch(VA_N_U) ar(1)  '@pch(N(-1))

 equation eN.ls @pch(N) c @pch(W_G(-1)) @pch(W_N(-1)) @pch(L_G(-1)) @pch(L_N(-
1)) @pch(VA_G(-1)) @pch(VA_N(-1)) @pch(N(-1)) @pch(N_U)

!exists=@isobject("var1")
if !exists=1 then
    delete var1
endif

model var1
 var1.merge eW_G
 var1.merge eW_N
 var1.merge eL_G
 var1.merge eL_N
 var1.merge eVA_G
 var1.merge eVA_N
 var1.merge eN
smpl 2000 2010
 var1.solve


 genr w_0 =(l_g_0*w_g_0+l_n_0*w_n_0)/(l_g_0+l_n_0)
 genr l_0 =l_g_0+l_n_0
 genr va_0=va_g_0+va_n_0

 genr dw_g=100*abs(w_g-w_g_0)/w_g
 genr dw_n=100*abs(w_n-w_n_0)/w_n
 genr dw  =100*abs(w-w_0)/w

 genr dl_g=100*abs(l_g-l_g_0)/l_g
 genr dl_n=100*abs(l_n-l_n_0)/l_n
 genr dl  =100*abs(l-l_0)/l

 genr dva_g=100*abs(va_g-va_g_0)/va_g
 genr dva_n=100*abs(va_n-va_n_0)/va_n
 genr dva  =100*abs(va-va_0)/va

 genr dn   = 100*abs(n_0-n)/n

!exists=@isobject("t")
if !exists=1 then
    delete t
endif
```

```
 group t dw_g dw_n dw dl_g dl_n dl dva_g dva_n dva dn
' show t
 t.stats

 genr aw_g=(w_g-w_g_0)/w_g
 genr aw_n=(w_n-w_n_0)/w_n
 genr aw  =(w-w_0)/w

 genr al_g=(l_g-l_g_0)/l_g
 genr al_n=(l_n-l_n_0)/l_n
 genr al  =(l-l_0)/l

 genr ava_g=(va_g-va_g_0)/va_g
 genr ava_n=(va_n-va_n_0)/va_n
 genr ava  =(va-va_0)/va

 genr an   = (n_0-n)/n

!exists=@isobject("s")
if !exists=1 then
    delete s
endif

 group s aw_g aw_n aw al_g al_n al ava_g ava_n ava an
' show s

 group r w_g_0 w_n_0 l_g_0 l_n_0 va_g_0 va_n_0 'l_0 w_0  va_0 n_0
 show r

end sub

'################################################

subroutine out_sample

call clearv("emp")
smpl 1987 1999
 var emp.ls 1 2 @pch(L_G) @pch(L_N) @ @pch(L_G_U) @pch(L_N_U)
call clearv("empmod")
 emp.makemodel(empmod)
smpl 2000 2010
 empmod.solve
genr L_0=L_G_0+L_N_0

call clearv("va")
smpl 1987 1999
 var va.ls 1 2 @pch(VA_G) @pch(VA_N) @ @pch(VA_G_U) @pch(VA_N_U)
call clearv("vamod")
 va.makemodel(vamod)
smpl 2000 2010
 vamod.solve
genr VA_0=VA_G_0+VA_N_0

call clearv("wr")
smpl 1987 1999
 var wr.ls 1 2 @pch(W_G) @pch(W_N) @ @pch(W_G_U) @pch(W_N_U)
call clearv("wrmod")
 wr.makemodel(wrmod)
smpl 2000 2010
 wrmod.solve
genr W_0=(W_G_0*L_G_0+W_N_0*L_N_0)/L_0

smpl 1987 2010
 group r L_G_0 L_N_0 L_0 VA_G_0 VA_N_0 VA_0 W_G_0 W_N_0 W_0
 show r

subroutine clearv(string %name)
!exists=@isobject(%name)
if !exists=1 then
    delete %name
endif
endsub
```

APPENDIX VI: NATIONAL MODEL ASSUMPTIONS

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
CONS	6,649,634,579.00	6,965,228,138.00	7,158,515,696.00	7,594,609,153.00	8,027,910,686.00	8,451,302,127.00	8,893,579,817.00	9,363,210,942.00	9,886,140,621.00	10,461,570,809.00	11,101,180,640.00
cpi	1.7228	1.7739	1.8055	1.8480	1.8972	1.9483	2.0019	2.0586	2.1176	2.1775	2.2403
dcpi	0.0562	0.0512	0.0315	0.0426	0.0491	0.0511	0.0536	0.0567	0.0591	0.0599	0.0628
epv_G	0.0133	0.0125	0.0117	0.0112	0.0109	0.0105	0.0101	0.0096	0.0089	0.0084	0.0079
epv_N	0.0179	0.0175	0.0172	0.0163	0.0156	0.0149	0.0143	0.0137	0.0131	0.0125	0.0120
EX	1,102,875,000.00	1,047,052,500.00	940,232,500.00	1,026,020,000.00	1,145,235,000.00	1,248,670,000.00	1,353,732,500.00	1,459,340,000.00	1,576,652,500.00	1,707,425,000.00	1,838,220,000.00
G	1,740,975,000.00	1,836,165,000.00	1,931,502,500.00	2,025,002,500.00	2,128,357,500.00	2,230,105,000.00	2,329,085,000.00	2,428,237,500.00	2,531,017,500.00	2,641,030,000.00	2,761,520,000.00
GF	589,478,401.60	612,291,443.00	655,616,245.20	693,454,281.30	730,179,132.80	763,554,192.80	795,380,582.90	824,332,869.50	852,657,647.50	884,176,650.90	921,219,376.70
GSL	1,343,869,873.00	1,428,330,975.00	1,489,068,895.00	1,554,019,054.00	1,631,777,593.00	1,711,580,359.00	1,789,954,327.00	1,871,897,953.00	1,958,803,543.00	2,050,426,735.00	2,147,820,588.00
I	1,818,765,550.00	1,698,021,291.00	1,679,275,665.00	1,818,914,753.00	1,967,408,300.00	2,078,292,263.00	2,199,273,189.00	2,345,935,134.00	2,521,774,090.00	2,710,086,978.00	2,927,978,183.00
INR	1,381,393,350.00	1,239,367,571.00	1,218,017,006.00	1,344,582,416.00	1,477,008,542.00	1,567,289,938.00	1,669,889,296.00	1,796,165,276.00	1,950,161,013.00	2,117,793,825.00	2,303,395,298.00
IR	437,372,199.50	458,652,690.20	461,256,600.90	474,329,507.20	490,400,787.60	511,002,839.50	529,382,864.60	549,769,343.10	571,612,048.30	592,290,579.90	624,583,141.50
L	168,480,276.50	169,068,018.30	168,523,171.40	170,392,420.10	173,085,762.40	175,252,524.30	177,028,264.10	178,672,987.50	180,515,162.70	182,695,416.80	185,310,180.90
L_G	24,587,800.54	23,627,244.28	22,246,494.84	22,448,384.25	23,074,890.67	23,348,349.90	23,542,896.69	23,446,149.09	23,119,874.76	23,008,938.83	22,942,207.82
L_N	143,892,476.00	145,440,774.00	146,276,676.50	147,944,035.80	150,010,871.80	151,904,174.40	153,485,367.40	155,226,838.40	157,395,287.90	159,686,478.00	162,367,973.10
N	275,690,000.00	278,180,000.00	280,685,000.00	283,172,500.00	285,640,000.00	288,085,000.00	290,525,000.00	292,955,000.00	295,382,500.00	297,810,000.00	300,237,500.00
RYD	6,539,167,250.00	6,772,925,000.00	6,908,118,750.00	7,157,635,000.00	7,382,778,750.00	7,576,281,250.00	7,752,564,500.00	7,925,713,500.00	8,156,069,000.00	8,416,709,000.00	8,701,585,250.00
TAX	1,288.17	1,303.99	1,285.15	1,341.45	1,400.53	1,459.13	1,564.38	1,678.36	1,764.12	1,854.47	1,956.94
TAXGF	1,009.50	1,009.12	989.13	1,025.61	1,065.28	1,103.82	1,188.03	1,281.69	1,347.51	1,416.76	1,490.98
TAXGSL	278.65	294.84	296.02	315.85	335.25	355.31	376.34	396.66	416.60	437.71	465.96
time	2,000.00	2,001.00	2,002.00	2,003.00	2,004.00	2,005.00	2,006.00	2,007.00	2,008.00	2,009.00	2,010.00
TWPER	701.53	731.33	745.30	783.29	825.60	863.67	904.54	948.67	997.25	1,051.43	1,113.00
VA	9,872,925,000.00	10,194,060,000.00	10,411,600,000.00	11,069,172,500.00	11,748,392,500.00	12,392,952,500.00	13,077,590,000.00	13,807,770,000.00	14,619,087,500.00	15,505,985,000.00	16,483,322,500.00
VA_G	1,845,808,333.00	1,884,792,857.00	1,905,952,768.00	2,008,659,049.00	2,115,914,937.00	2,218,032,217.00	2,328,891,029.00	2,449,835,351.00	2,587,589,320.00	2,741,637,992.00	2,915,190,559.00
VA_N	8,027,116,667.00	8,309,267,143.00	8,505,647,232.00	9,060,513,451.00	9,632,477,563.00	10,174,920,283.00	10,748,698,971.00	11,357,934,649.00	12,031,498,180.00	12,764,347,008.00	13,568,131,941.00
VP	1,069.05	1,150.06	1,235.21	1,296.74	1,367.88	1,451.79	1,543.21	1,646.63	1,765.11	1,890.80	2,023.57
W	28,710.73	30,144.26	31,012.06	32,333.91	33,652.21	34,870.41	36,227.33	37,700.91	39,265.10	40,931.52	42,747.45
W_G	36,974.54	40,111.67	43,285.71	44,870.83	45,860.91	47,325.31	49,094.41	51,690.05	55,148.18	58,544.71	62,380.19
W_N	27,298.64	28,525.02	29,145.43	30,431.61	31,774.25	32,956.04	34,253.67	35,587.93	36,932.03	38,393.67	39,973.39
wsd	4,837,191,000.00	5,096,430,000.00	5,226,251,250.00	5,509,453,500.00	5,824,718,500.00	6,111,127,500.00	6,413,262,000.00	6,736,134,250.00	7,087,946,500.00	7,478,001,250.00	7,921,537,750.00
wsd_g	909,122,646.20	947,728,336.70	962,955,238.20	1,007,277,692.00	1,058,235,583.00	1,104,967,921.00	1,155,824,505.00	1,211,932,565.00	1,275,018,927.00	1,347,051,563.00	1,431,139,198.00
wsd_n	3,928,068,354.00	4,148,701,663.00	4,263,296,012.00	4,502,175,808.00	4,766,482,917.00	5,006,159,579.00	5,257,437,495.00	5,524,201,685.00	5,812,927,573.00	6,130,949,687.00	6,490,398,552.00
YD	7,030,990,750.00	7,420,843,250.00	7,679,399,000.00	8,134,935,000.00	8,598,735,500.00	9,041,592,000.00	9,487,603,250.00	9,958,771,000.00	10,530,450,000.00	11,166,935,000.00	11,871,420,000.00
YENTAFADJ	30,583,250.00	28,464,880.00	30,209,662.50	32,293,140.00	32,562,855.00	32,242,230.00	32,386,030.00	31,939,955.00	31,995,862.50	31,801,307.50	32,215,265.00
YENTNFADJ	684,389,250.00	714,989,600.00	756,969,250.00	813,523,550.00	854,184,000.00	900,359,250.00	954,368,350.00	1,009,401,200.00	1,065,876,500.00	1,123,460,250.00	1,190,311,000.00
yintper	1,000.62	996.76	945.26	1,027.24	1,105.57	1,166.16	1,233.76	1,294.11	1,375.32	1,476.98	1,581.30
YN	7,980,950,000.00	8,195,816,750.00	8,367,377,750.00	8,855,561,750.00	9,360,219,750.00	9,838,224,000.00	10,359,012,500.00	10,920,795,000.00	11,531,127,500.00	12,228,810,000.00	13,017,470,000.00
YOL	1,218,571,250.00	1,267,955,400.00	1,319,688,925.00	1,401,416,675.00	1,466,793,925.00	1,540,168,425.00	1,624,301,375.00	1,712,309,425.00	1,806,281,275.00	1,904,171,825.00	2,014,163,400.00
YOLY	534,182,000.00	552,965,800.00	562,719,675.00	587,893,125.00	612,609,925.00	639,809,175.00	669,933,025.00	702,908,225.00	740,404,775.00	780,711,575.00	823,852,400.00
YP	8,319,161,750.00	8,724,835,250.00	8,964,547,750.00	9,476,389,750.00	9,999,263,250.00	10,500,722,500.00	11,051,980,000.00	11,637,125,000.00	12,294,567,500.00	13,021,402,500.00	13,828,360,000.00
YP96C	7,737,289,350.00	7,963,097,301.00	8,064,155,500.00	8,337,956,750.00	8,585,252,000.00	8,798,906,500.00	9,030,734,750.00	9,261,382,500.00	9,522,409,500.00	9,814,444,750.00	10,135,977,500.00
YPROP	1,582,769,750.00	1,623,670,175.00	1,601,333,225.00	1,707,198,350.00	1,803,177,600.00	1,884,250,325.00	1,982,148,400.00	2,079,584,700.00	2,200,686,225.00	2,344,733,900.00	2,499,672,075.00
YRENT	202,550,000.00	210,303,375.00	213,796,950.00	220,962,250.00	227,751,175.00	236,359,075.00	246,422,700.00	257,223,500.00	267,333,850.00	276,000,675.00	286,844,075.00

APPENDIX VII: COMPLETE MODEL FORECAST

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
AC_G	0.9602	0.9210	0.8920	0.8974	0.9121	0.9165	0.9154	0.9024	0.8813	0.8673	0.8530
AC_N	0.9824	0.9807	0.9759	0.9720	0.9706	0.9691	0.9677	0.9668	0.9661	0.9660	0.9665
C_G	15,151,024.00919	15,530,957.44288	15,909,860.18126	16,970,425.60284	18,049,847.51850	19,116,676.86825	20,263,978.00074	21,519,432.47756	22,915,759.86228	24,484,901.24629	26,254,308.86937
C_N	56,223,414.90524	57,633,296.84232	59,039,354.00751	62,974,969.82436	66,980,559.55775	70,939,419.97061	75,196,900.35966	79,855,723.28160	85,037,306.64173	90,860,179.53958	97,426,213.54946
cp	0.9887	0.9876	0.9845	0.9820	0.9811	0.9802	0.9792	0.9787	0.9782	0.9781	0.9785
cpi	1.7033	1.7519	1.7775	1.8148	1.8614	1.9096	1.9603	2.0147	2.0716	2.1299	2.1922
epv_G	0.0147	0.0138	0.0128	0.0123	0.0120	0.0116	0.0111	0.0105	0.0098	0.0092	0.0087
epv_N	0.0239	0.0234	0.0230	0.0218	0.0208	0.0199	0.0191	0.0182	0.0175	0.0167	0.0160
GF_G	1,150,195.22	1,194,708.21	1,279,243.93	1,353,073.83	1,424,731.67	1,489,853.36	1,551,953.28	1,608,445.23	1,663,712.77	1,725,212.92	1,797,491.00
GF_N	6,482,489.02	6,733,363.85	7,209,806.33	7,625,910.90	8,029,773.78	8,396,799.04	8,746,793.58	9,065,181.63	9,376,668.98	9,723,283.19	10,130,641.74
GSL_G	1,229,568.75	1,306,090.10	1,360,864.04	1,419,588.94	1,490,122.04	1,562,642.83	1,633,889.99	1,708,454.69	1,787,564.11	1,870,977.70	1,959,653.61
GSL_N	12,689,255.48	13,478,962.42	14,044,234.21	14,650,280.22	15,378,187.89	16,126,608.70	16,861,885.57	17,631,399.63	18,447,815.66	19,308,651.16	20,223,794.14
I_G	9,328,022.60	8,549,780.94	8,452,648.97	9,241,324.76	10,101,983.68	10,778,256.99	11,537,641.25	12,490,066.53	13,654,645.78	14,937,542.49	16,441,995.56
I_N	8,371,685.18	7,588,926.42	7,491,858.88	8,237,856.07	9,048,260.11	9,669,840.34	10,375,134.56	11,264,172.40	12,357,926.33	13,567,764.77	14,971,830.21
INT_G	22,793,367.94	23,162,973.66	23,473,742.88	25,067,586.78	26,747,855.84	28,384,665.37	30,184,337.54	32,279,431.39	34,750,140.05	37,425,991.65	40,415,760.28
INT_N	36,119,304.95	36,748,043.65	37,432,921.15	39,960,919.53	42,630,507.49	45,208,860.38	47,997,818.20	51,155,411.76	54,791,313.92	58,776,208.42	63,237,976.33
INTT_G	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74	60,414,418.74
INTT_N	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55	122,831,311.55
L	2,056,202.73	2,040,289.49	2,040,402.62	2,083,574.40	2,141,935.52	2,195,834.50	2,248,131.83	2,306,783.54	2,374,100.55	2,450,193.81	2,538,349.04
L_G	354,710.34	332,987.95	307,688.29	309,201.82	316,426.74	318,711.94	320,281.46	319,506.64	316,801.16	316,003.77	315,537.72
L_N	1,701,492.39	1,707,301.54	1,732,714.32	1,774,372.58	1,825,508.79	1,877,122.55	1,927,850.37	1,987,276.90	2,057,299.39	2,134,190.04	2,222,811.32
lamVA_G	0.3547	0.3484	0.3423	0.3363	0.3305	0.3249	0.3194	0.3140	0.3088	0.3037	0.2988
lamVA_N	0.6311	0.6369	0.6431	0.6500	0.6573	0.6652	0.6736	0.6826	0.6921	0.7022	0.7128
LD_G	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52	60,475,662.52
LD_N	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92	122,829,782.92
N	3,460,537.23	3,489,772.37	3,519,217.15	3,548,822.33	3,578,547.71	3,608,361.08	3,638,239.53	3,668,167.21	3,698,132.10	3,728,125.70	3,758,141.77
NTAX	-1,880,255.75	-2,943,327.92	-4,181,283.30	-4,373,991.80	-4,583,230.72	-4,957,256.90	-5,019,396.89	-5,113,615.32	-5,592,033.73	-6,050,626.37	-6,423,918.96
REO	0.9723	0.9620	0.9657	0.9757	0.9878	1.0003	1.0141	1.0311	1.0505	1.0713	1.0943
RF_G	0.9778	1.0000	1.0173	1.0140	1.0052	1.0026	1.0033	1.0110	1.0240	1.0328	1.0421
RF_N	0.8536	0.8543	0.8560	0.8575	0.8580	0.8586	0.8591	0.8595	0.8597	0.8598	0.8596
RLC_G	0.8723	0.8366	0.8103	0.8152	0.8286	0.8326	0.8316	0.8198	0.8006	0.7879	0.7749
RLC_N	0.7358	0.7345	0.7309	0.7280	0.7269	0.7258	0.7247	0.7241	0.7236	0.7235	0.7239
rpc_G	0.6487	0.6521	0.6553	0.6578	0.6600	0.6625	0.6651	0.6680	0.6710	0.6739	0.6768
rpc_N	0.8676	0.8684	0.8692	0.8700	0.8708	0.8716	0.8725	0.8733	0.8741	0.8749	0.8757
RWR	0.7727	0.7633	0.7539	0.7513	0.7518	0.7504	0.7482	0.7445	0.7398	0.7364	0.7333
RYD	43,005,301.51	43,590,765.46	44,315,792.58	46,164,931.86	47,912,210.03	49,465,249.62	50,985,752.26	52,577,474.38	54,568,713.52	56,858,986.60	59,399,822.20
sp_G	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
sp_N	0.9824	0.9807	0.9759	0.9720	0.9706	0.9691	0.9677	0.9668	0.9661	0.9660	0.9665
TAX	10,671,725.28	10,551,929.79	10,305,081.43	10,827,131.90	11,446,476.76	12,051,972.62	13,057,421.49	14,172,006.07	15,078,858.41	16,089,880.59	17,268,881.19
TRAN	12,551,981.03	13,495,257.71	14,486,364.73	15,201,123.70	16,029,707.48	17,009,229.52	18,076,818.38	19,285,621.39	20,670,892.14	22,140,506.96	23,692,800.15
TWPER	3,347,487.89	3,408,659.56	3,442,238.71	3,641,454.19	3,886,547.94	4,108,925.34	4,348,703.01	4,614,000.92	4,909,769.13	5,254,436.58	5,657,115.82
VA	95,279,179.43	97,184,078.88	99,406,388.51	106,520,342.95	114,150,217.43	121,673,277.74	129,896,533.18	139,231,732.98	149,991,911.72	161,972,232.12	175,538,624.48
VA_G	24,189,297.99	24,130,236.17	23,946,608.25	25,133,078.19	26,358,107.19	27,503,786.80	28,780,847.03	30,326,881.73	32,209,140.34	34,204,896.63	36,422,145.45
VA_N	71,089,881.44	73,053,842.71	75,459,780.27	81,387,264.75	87,792,110.24	94,169,490.94	101,115,686.15	108,904,851.25	117,782,771.39	127,767,335.49	139,116,479.02
W	22,184.58	23,008.81	23,379.50	24,294.04	25,299.47	26,166.68	27,104.20	28,068.18	29,048.61	30,142.62	31,347.78
W_G	32,251.56	33,558.50	35,075.20	36,580.95	38,000.58	39,401.97	40,825.59	42,375.26	44,150.68	46,125.05	48,337.80
W_N	20,085.92	20,951.22	21,302.62	22,152.93	23,097.91	23,919.50	24,824.61	25,767.94	26,723.06	27,776.14	28,935.97
wsd	45,616,001.79	46,944,626.84	47,703,593.07	50,618,443.16	54,189,836.19	57,457,705.85	60,933,819.31	64,747,215.92	68,964,315.86	73,855,255.37	79,571,602.77
wsd_g	11,439,960.01	11,174,576.93	10,792,229.79	11,310,895.25	12,024,399.83	12,557,877.87	13,075,680.45	13,539,178.13	13,986,987.41	14,575,689.10	15,252,400.15
wsd_n	34,176,041.78	35,770,049.90	36,911,363.28	39,307,547.91	42,165,436.36	44,899,827.98	47,858,138.86	51,208,037.79	54,977,328.45	59,279,566.27	64,319,202.61
X_G	25,723,697.82	26,384,405.97	26,341,566.29	27,917,101.09	29,578,788.62	31,283,055.61	33,209,898.02	35,553,449.03	38,440,125.74	41,464,573.78	44,766,848.88
X_N	2,899,986.01	2,759,802.99	2,488,459.34	2,713,764.81	3,024,647.79	3,294,866.28	3,569,559.04	3,845,941.71	4,152,900.07	4,494,843.88	4,837,427.97
XT_G	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26	22,977,886.26
XT_N	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18	2,604,767.18
YD	73,251,073.96	76,368,813.37	78,773,015.64	83,778,509.56	89,181,476.36	94,458,944.10	99,949,677.34	105,925,801.69	113,042,619.34	121,105,553.99	130,213,516.59
YOL	14,422,838.78	14,839,431.53	15,495,684.90	16,619,156.55	17,603,422.28	18,714,856.14	20,004,547.49	21,439,425.69	23,038,476.55	24,766,342.66	26,756,525.51
YP	71,370,818.22	73,425,485.45	74,591,732.34	79,404,517.76	84,598,245.64	89,501,687.20	94,930,280.45	100,812,186.36	107,450,585.61	115,054,927.62	123,789,597.63
YPROP	14,679,465.54	15,050,086.65	14,834,693.07	15,808,372.24	16,691,535.11	17,438,050.54	18,340,616.65	19,239,545.67	20,357,562.34	21,687,766.17	23,118,585.18

Table 5.3: Projection of Key Indicators

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
L	2,056,203	2,040,289	2,040,403	2,083,574	2,141,936	2,195,834	2,248,132	2,306,784	2,374,101	2,450,194	2,538,349
L_G	354,710	332,988	307,688	309,202	316,427	318,712	320,281	319,507	316,801	316,004	315,538
L_N	1,701,492	1,707,302	1,732,714	1,774,373	1,825,509	1,877,123	1,927,850	1,987,277	2,057,299	2,134,190	2,222,811
VA ($Mil)	95,279	97,184	99,406	106,520	114,150	121,673	129,897	139,232	149,992	161,972	175,539
VA_G ($Mil)	24,189	24,130	23,947	25,133	26,358	27,504	28,781	30,327	32,209	34,205	36,422
VA_N ($Mil)	71,090	73,054	75,460	81,387	87,792	94,169	101,116	108,905	117,783	127,767	139,116
W	22,185	23,009	23,380	24,294	25,299	26,167	27,104	28,068	29,049	30,143	31,348
W_G	32,252	33,559	35,075	36,581	38,001	39,402	40,826	42,375	44,151	46,125	48,338
W_N	20,086	20,951	21,303	22,153	23,098	23,919	24,825	25,768	26,723	27,776	28,936

Table 5.4 :Regionalized Components of Final Demands and Output

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Q	154,744	157,256	160,236	171,115	182,711	193,938	206,078	219,754	235,396	252,560	271,793
Q_G	48,231	48,760	49,228	52,635	56,333	59,980	64,031	68,764	74,328	80,324	87,028
Q_N	106,513	108,496	111,007	118,480	126,378	133,957	142,047	150,990	161,068	172,235	184,765
INT	46,121	47,016	47,920	51,256	54,778	58,211	61,952	66,235	71,211	76,646	82,733
INT_G	14,785	15,105	15,383	16,489	17,654	18,805	20,075	21,561	23,317	25,221	27,353
INT_N	31,336	31,911	32,537	34,767	37,124	39,406	41,877	44,673	47,893	51,425	55,380
VA	108,623	110,240	112,316	119,859	127,933	135,726	144,126	153,520	164,186	175,914	189,060
VA_G	33,446	33,655	33,845	36,146	38,679	41,176	43,955	47,203	51,011	55,103	59,675
VA_N	75,178	76,584	78,471	83,713	89,254	94,551	100,171	106,317	113,175	120,811	129,385
C	58,605	60,175	61,743	65,952	70,242	74,499	79,084	84,111	89,708	95,996	103,088
C_G	9,828	10,128	10,426	11,163	11,913	12,665	13,477	14,374	15,377	16,500	17,768
C_N	48,777	50,047	51,317	54,789	58,329	61,834	65,607	69,737	74,331	79,496	85,320
I	13,314	12,166	12,051	13,246	14,547	15,569	16,726	18,180	19,964	21,937	24,239
I_G	6,051	5,575	5,539	6,079	6,668	7,141	7,674	8,343	9,162	10,066	11,128
I_N	7,263	6,590	6,512	7,167	7,880	8,429	9,052	9,837	10,802	11,871	13,111
GSL	11,806	12,557	13,099	13,680	14,375	15,092	15,798	16,538	17,325	18,154	19,037
GSL_G	798	852	892	934	984	1,035	1,087	1,141	1,199	1,261	1,326
GSL_N	11,009	11,705	12,207	12,746	13,392	14,057	14,711	15,397	16,125	16,894	17,711
GF	6,370	6,626	7,105	7,525	7,933	8,306	8,663	8,991	9,313	9,670	10,088
GF_G	746	779	838	890	940	987	1,032	1,074	1,116	1,163	1,217
GF_N	5,624	5,847	6,267	6,635	6,993	7,319	7,631	7,916	8,196	8,507	8,872
X	18,528	18,716	18,317	19,457	20,835	22,260	23,854	25,700	27,876	30,157	32,607
X_G	16,023	16,321	16,149	17,080	18,174	19,348	20,685	22,270	24,156	26,114	28,236
X_N	2,505	2,395	2,168	2,376	2,661	2,912	3,169	3,430	3,720	4,044	4,371
rpc_G	0.649	0.652	0.655	0.658	0.660	0.662	0.665	0.668	0.671	0.674	0.677
rpc_N	0.868	0.868	0.869	0.870	0.871	0.872	0.872	0.873	0.874	0.875	0.876

Table 5.6: Production Block

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Epv_G	0.0147	0.0138	0.0128	0.0123	0.0120	0.0116	0.0111	0.0105	0.0098	0.0092	0.0087
Epv_N	0.0239	0.0234	0.0230	0.0218	0.0208	0.0199	0.0191	0.0182	0.0175	0.0167	0.0160
AC_G	0.9602	0.9210	0.8920	0.8974	0.9121	0.9165	0.9154	0.9024	0.8813	0.8673	0.8530
AC_N	0.9824	0.9807	0.9759	0.9720	0.9706	0.9691	0.9677	0.9668	0.9661	0.9660	0.9665
RF_G	0.9778	1.0000	1.0173	1.0140	1.0052	1.0026	1.0033	1.0110	1.0240	1.0328	1.0421
RF_N	0.8536	0.8543	0.8560	0.8575	0.8580	0.8586	0.8591	0.8595	0.8597	0.8598	0.8596
RLC_G	0.8723	0.8366	0.8103	0.8152	0.8286	0.8326	0.8316	0.8198	0.8006	0.7879	0.7749
RLC_N	0.7358	0.7345	0.7309	0.7280	0.7269	0.7258	0.7247	0.7241	0.7236	0.7235	0.7239

Table 5.7: Income Components ($1000000)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
WSD	45,616	46,945	47,704	50,618	54,190	57,458	60,934	64,747	68,964	73,855	79,572
+ YOL	14,423	14,839	15,496	16,619	17,603	18,715	20,005	21,439	23,038	24,766	26,757
+ YPROP	14,679	15,050	14,835	15,808	16,692	17,438	18,341	19,240	20,358	21,688	23,119
- TWPER	3,347	3,409	3,442	3,641	3,887	4,109	4,349	4,614	4,910	5,254	5,657
= YP	71,371	73,425	74,592	79,405	84,598	89,502	94,930	100,812	107,451	115,055	123,790
- TAX	10,672	10,552	10,305	10,827	11,446	12,052	13,057	14,172	15,079	16,090	17,269
+ TRAN	12,552	13,495	14,486	15,201	16,030	17,009	18,077	19,286	20,671	22,141	23,693
= YD	73,251	76,369	78,773	83,779	89,181	94,459	99,950	105,926	113,043	121,106	130,214
∏ Cpi	1.70	1.75	1.78	1.81	1.86	1.91	1.96	2.01	2.07	2.13	2.19
= RYD	43,005	43,591	44,316	46,165	47,912	49,465	50,986	52,577	54,569	56,859	59,400

VITA

Steven Ray Miller

Candidate for the Degree of

Doctor of Philosophy

Dissertation: A REGIONAL POLICY SIMULATION AND FORECAST MODEL FOR THE STATE OF OKLAHOMA: A MAXIMUM ENTROPY APPROACH

Major Field: Urban/Regional Economics

EDUCATIONAL BACKGROUND

B.A., Economics, Oklahoma State University, May 1995
Ph.D., Economics, Oklahoma State University, December 2005

RESEARCH FIELDS

Urban/Regional Economics, Econometrics, Monetary Economics.

AWARDS AND HONORS

Outstanding Manuscript: SWBS 2005.

Duck Woo Nam Distinguished Graduate Fellowship: 2004.

ARTICLES IN TRADE JOURNALS

"Systems-Estimation Using Generalized Cross-Entropy", Forthcoming in the 2005 *Proceedings of the Southwest Business Symposium*.

A Data Envelopment Analysis Approach to MSAs' Efficiency, with Susanne Rassouli-Currier. Submitted

"Oklahoma: If We Aren't High-Tech, Where Are Our Competitive Advantages?" With Tim Ireland and Mark C. Snead. *Oklahoma Business Bulletin*, Forthcoming.

"An Evaluation of Alternative Strategies for Incorporating Inter-Industry Relationships into a Regional Employment Forecasting Model," with Dan S. Rickman, 2002. *The Review of Regional Studies*. 32 (1):133-147.

PRESENTATIONS

Systems Estimation Using Generalized Cross Entropy. Presented at the *Southwest Business Symposium*, March 2005.

The Choice of Estimation Techniques and the Effect on MSA's Efficiency. Presented by Rassouli-Currier at the *Southwest Business Symposium*, March 2005.

Systems Estimation Using Generalized Cross Entropy. Presented at the *Missouri Valley Economic Association*, October 2004.

A Data Envelopment Analysis Approach to MSAs' Efficiency, with Susanne Rassouli-Currier. Presented by Rassouli-Currier at the *Missouri Valley Economic Association*, October 2004.

Regional Aggregate Forecasting Models. Presented at the *Southwest Business Symposium*, April 2004.

Regional Bayesian Forecasting and Forecast Aggregation Bias. Presented at the *Missouri Valley Economic Association*, February 2004.

Regional Stochastic Frontier Models: Further Insights Of Regional Agglomeration Effects, with Susanne Rassouli-Currier. Presented by Rassouli-Currier at the *Southwest Business Symposium*, April 2003.

Regional Stochastic Frontier Models: Further Insights Of Regional Agglomeration Effects, with Susanne Rassouli-Currier. Presented at the *Missouri Valley Economic Association*, February 2003.

Calculating the Dynamic Relationship Between Basic and Non-Basic Employment, with Russell W. McKenzie. Presented at the *Missouri Valley Economic Association*, February 2002.

An Evaluation of Alternative Strategies for Incorporating Inter-Industry Relationships into a Regional Employment Forecasting Model. Presented at the *Missouri Valley Economic Association*, February 2001.

WORKING PAPERS AND MONOGRAPHS

Recovering Suppressed Data Using Generalized Cross-Entropy.

"Oklahoma City Metro Area Forecast: 2005 Economic Outlook" The Oklahoma Economy, 2005. pp. 10-13.

Systems Estimation Using Generalized Cross Entropy.

A Data Envelopment Analysis Approach to MSAs' Efficiency, with Susanne Rassouli-Currier.

Regional Bayesian Forecasting and Forecast Aggregation Bias.

Regional Stochastic Frontier Models: Further Insights Of Regional Agglomeration Effects, with Susanne Rassouli-Currier.

Calculating the Dynamic Relationship between Basic and Non-Basic Employment, with Russell W. McKenzie.

RESEARCH EXPERIENCE

Research Associate: July 2005
Center for Applied Economic Research: Oklahoma State University

Assist in the formulation of an ongoing local-area forecasting model including recovering employment suppressions and formulating programming environment.

Research Associate: June 2005
Center for Applied Economic Research: Oklahoma State University

Assist in the development of a state general equilibrium model for the analysis of policy impacts of child-care subsidies.

Document Editor: August 2003 – May 2004
Dan S. Rickman, Ron L. Moomaw, *Review of Regional Science.*

Provide proofing, editing, and formatting of peer reviewed journal articles.

Research Assistant: September 1998 – May 2004
Dan S. Rickman: Oklahoma State University

Program and maintain regional econometric model, create output tables and forecasts, and collect input data on an ongoing basis.

Document Editor: August 2003 – May 2004
Dan S. Rickman, Ron L. Moomaw, *Review of Regional Science.*

Provide proofing, editing, and formatting of peer reviewed journal articles.

Research Assistant: August 2003 – December 2003
Mark Snead: Oklahoma State University
Funded Research Project: Oklahoma Department of Human Services

Author report outlining the characteristics of licensed countywide child care facilities for all counties in Oklahoma.

Research Assistant: May 2003 – August 2003
Dan S. Rickman: Oklahoma State University
Mark D. Partridge: Cloud State University

Develop macros and programs to analyze and process regional, MSA, and county data. Develop programs to model and test spatial autocorrelation.

Research Assistant: June 2002 – December 2002
Dan S. Rickman and Mark C. Snead: Oklahoma State University
Funded Research Project: Oklahoma Department of Human Services

State comparison of state implemented IV-D programs for child support collections. Author analysis of common characteristics of successful statewide programs.

Research Assistant: September 2001 - December 2001
Mark C. Snead and Tim L. Krehbiel, Oklahoma State University
Funded Research Project: Oklahoma Department of Treasury

Assist in development of a dynamic portfolio optimization method and write programs for analysis and application by Oklahoma state teacher retirement fund administrators.

TEACHING EXPERIENCE

Instructor: Oklahoma State University-Tulsa, Spring 2005
Money and Banking,
Business and Government.
Instructor: Oklahoma State University-Tulsa, Fall 2004
Money and Banking,
International Economic Relations.
Instructor: Oklahoma State University, Fall 2003
Money and Banking,
Half semester of Managerial Economics.
Instructor: Oklahoma State University, Spring 2003
Intermediate Macroeconomics.
Instructor: Oklahoma State University, Fall 2002
Intermediate Macroeconomics.
Principles of Microeconomics.
Instructor: Oklahoma State University, Spring 2002
Principles of Microeconomics.
Instructor: Oklahoma State University, Fall 2001
Principles of Macroeconomics.
Instructor: Oklahoma State University, Summer 2001
Principles of Microeconomics.
Instructor: Oklahoma State University, Fall 2000 – Spring 2001
Economics of Social Issues.

OKLAHOMA STATE UNIVERSITY: SERVICES

CBA: Student Technology Fee Committee Representative, 2003-2004.

CBA: Student Technology Fee Committee Representative, 2000-2001.

University: Student Technology Fee Committee Representative, 2001.

Dept. Of Economics and Legal Studies: Developed Student Tracking Databank and interface, 1999.

Chairperson: Parking Taskforce, Off-Campus Student Association 1991-1992

Student Fee Allocation Committee, Student Government Association, 1991-1992

Student Representative: Off-Campus Student Association 1990-1992

REVIEWER

Journal of Economics

Name: Steven Ray Miller Date of Degree: December, 2005

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: A REGIONAL POLICY SIMULATION AND FORECAST MODEL FOR THE STATE OF OKLAHOMA: A MAXIMUM ENTROPY APPROACH

Pages in Study: 213 Candidate for the Degree of Doctor of Philosophy

Major Field: Urban/Regional Economics

Scope and Method of Study: A variant of Bayesian estimation called entropy is employed to dynamically calibrate a regional general equilibrium-type model to maximize the fit to historical observations for the state of Oklahoma. It is postulated that such dynamically calibrated policy simulation model will not only be useful in policy analysis, but also for forecasting applications granting policy simulations with a time-path response to facilitate timing. Generalized Cross-Entropy estimation is employed to dynamically calibrate the system as a whole advantaging from the full set of general equilibrium constraints. Such estimation transforms the traditional econometric estimation to a non-linear math programming problem in non-linear constraints.

Findings and Conclusions: Entropic estimation allows complete systems estimation consistent with the full set of general equilibrium constraints transcending criticisms of single equation estimation. Projections of the model create forecasts that compare favorably with more traditional econometric methods for forecasting. The addition of complete market structure in estimation extends the forecasting application to policy analysis allowing for a large breadth of policy applications that illustrate not only the overall impact implications but also the timing of those implications.

ADVISER'S APPROVAL: Dan S. Rickman