UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Analysis of Discrete Data Models with Endogeneity, Simultaneity, and Missing Outcomes

Permalink
https://escholarship.org/uc/item/3x74w3rh

Author
Vossmeyer, Angela Marie

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

,

UNIVERSITY OF CALIFORNIA,
IRVINE

Analysis of Discrete Data Models with Endogeneity, Simultaneity, and Missing Outcomes

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Angela Vossmeyer

Dissertation Committee:
Professor Ivan Jeliazkov, Chair
Professor David Brownstone
Professor Gary Richardson

2015

Chapter 1 © 2015 Taylor & Francis Group
Chapter 2 © 2014 Emerald Publishing Limited
All other materials © 2015 Angela Vossmeyer

DEDICATION

To my parents, John and Cheryl Vossmeyer, for their endless love, support, and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my committee members, Professor Ivan Jeliazkov, Professor David Brownstone, and Professor Gary Richardson, for their invaluable guidance on my dissertation and assistance throughout my graduate studies. Specifically, I would like to thank Ivan Jeliazkov for mentoring me since my undergraduate studies, inspiring my interest in economics and statistics, and guiding me through my first three publications. He is dedicated to my academic and professional development, for which I am forever grateful. He will always be my role model.

I am thankful to Gary Richardson for motivating me to study banking and providing me with numerous opportunities to present at conferences and Federal Reserve banks. His guidance positively influenced the direction of my research and will remain instrumental throughout my academic career. I cannot thank him enough for all of the doors he opened for me. Additionally, I would like to acknowledge David Brownstone for his invaluable insights, which improved the quality of all aspects of my papers. It was an honor to serve as his teaching assistant because I never stopped learning from him. Finally, I am grateful to Dale Poirier who constantly supported my success in graduate school and always pushed the limits of my abilities.

I would also like to thank Sean Dowsing, not only for his love and support, but also for brainstorming research ideas with me, revising my papers, and providing me with preliminary banking data. He is my champion and my foundation.

I am grateful to my parents (John and Cheryl Vossmeyer), my sister (Erin Hillis), and brother-in-law (Jamie Hillis). I appreciate them for encouraging me and ensuring that I am always having fun. Thank you to all my classmates for the many thoughtful discussions.

Funding provided by the Kassouf Fellowship, Department of Economics, School of Social Sciences, University of California, Irvine, and All-UC Group in Economic History are acknowledged.

The copyright for Chapter 1 belongs to Taylor & Francis Group. It is forthcoming as “Sample Selection and Treatment Effect Estimation of Lender of Last Resort Policies” in the *Journal of Business and Economic Statistics*. The copyright for Chapter 2 belongs to Emerald Publishing Limited. It is published as “Determining the Proper Specification for Endogenous Covariates in Discrete Data Settings” in *Advances in Econometrics*, Vol. 34, pp. 223-247 (2014).

CURRICULUM VITAE

Angela Vossmeyer

EDUCATION

Doctor of Philosophy in Economics (2010 - 2015)
University of California, Irvine

Master of Arts in Economics (2010 - 2012)
University of California, Irvine

Bachelor of Arts in Quantitative Economics (2006 - 2010)
University of California, Irvine

RESEARCH INTERESTS

Econometrics; Markov chain Monte Carlo; Simulation-based inference; Computation and data analysis; Program and policy evaluation; Money and banking; Financial intermediation; Industrial organization.

PUBLICATIONS

"**Treatment Effects and Informative Missingness with an Application to Bank Recapitalization Programs**," *American Economic Review* (Papers and Proceedings), 104, 5, 212-217, 2014.

"**Determining the Proper Specification for Endogenous Covariates in Discrete Data Settings**," *Advances in Econometrics*, 34, 223-247, 2014.

"**Sample Selection and Treatment Effect Estimation of Lender of Last Resort Policies**," *Journal of Business & Economic Statistics*, forthcoming, 2015.

WORK EXPERIENCE

Teaching Assistant, Department of Economics, University of California, Irvine (2010 - 2015)

Graduate	Probability Camp, Statistics and Econometrics II, Statistics and Econometrics III, Statistics and Econometrics IV.
Undergraduate	Econometrics II, Applied Econometrics I, Business Forecasting, Microeconomics, International Business.

Add-in Programmer, EViews/IHS Global (2013 - 2014)

Research Assistant, Professor David Brownstone, Institute for Transportation Studies (2013)

Business Analyst, Synergy Global Entertainment (2007 - 2010)

PRESENTATIONS

Robert Day School of Economics and Finance, Claremont McKenna College, (November, 2014).

Department of Economics, Rutgers University, (October, 2014).

Department of Economics, University of California, Irvine, (October, 2014).

California Econometrics Conference, Stanford University, (September, 2014).

Brown Bag Seminar, Federal Reserve Bank of Richmond, (August, 2014).

Workshop on Monetary and Financial History, Federal Reserve Bank of Atlanta, (May, 2014).

NBER-NSF Seminar of Bayesian Inference in Econometrics and Statistics, University of Chicago, (May, 2014).

Central Banking in Historical Perspective Conference, Federal Reserve Bank of San Francisco, (March, 2014).

Conference on Bayesian Model Comparison, Advances in Econometrics, (February, 2014).

American Economic Association Annual Meeting, Philadelphia, PA, (January, 2014).

Economic History Association Annual Meeting, Washington D.C., (September, 2014).

HONORS & AWARDS

Sheen T. Kassouf Fellowship for Best Graduate Student (2014)
Department of Economics, University of California, Irvine

Best Econometrics Paper Award (2013)
Department of Economics, University of California, Irvine

Best Teaching Assistant Award (2013)
Department of Economics, University of California, Irvine

Art DeVany Prize for Best Presentation (2013)

Department of Economics, University of California, Irvine

Research Fellow (2013)
All UC-Group in Economic History

Summer Research Fellowship (2012, 2013, 2014)
Department of Economics, University of California, Irvine

REFEREE SERVICE

Advances in Econometrics; *Bayesian Inference in the Social Sciences*; *Economic Inquiry*; *International Journal of Mathematical Modelling and Numerical Optimization*; *Journal of Econometrics.*

AFFILIATIONS

American Economic Association; American Statistical Association; Econometric Society; Economic History Association; International Society for Bayesian Analysis.

ABSTRACT OF THE DISSERTATION

Analysis of Discrete Data Models with Endogeneity, Simultaneity, and Missing Outcomes

By

Angela Vossmeyer

Doctor of Philosophy in Economics

University of California, Irvine, 2015

Professor Ivan Jeliazkov, Chair

This thesis is concerned with specifying and estimating multivariate models in discrete data settings. The models are applied to several empirical applications with an emphasis in banking and monetary history. The approaches presented here are of central importance in model evaluation, policy analysis, and prediction.

The first chapter develops a framework for estimating multivariate treatment effect models in the presence of sample selection. The methodology deals with several important issues prevalent in program evaluation, including non-random treatment assignment, endogeneity, and discrete outcomes. The framework is applied to evaluate the effectiveness of bank recapitalization programs and their ability to resuscitate the financial system. This paper presents a novel bank-level data set and employs the new methodology to jointly model a bank's decision to apply for assistance, the central bank's decision to approve or decline the assistance, and the bank's performance. The article offers practical estimation tools to unveil new answers to important regulatory and government intervention questions.

The second chapter examines an important but often overlooked obstacle in multivariate discrete data models which is the proper specification of endogenous covariates.

Endogeneity can be modeled as latent or observed, representing competing hypotheses about the outcomes of interest. This paper highlights the use of existing Bayesian model comparison techniques to understand the nature of endogeneity. Consideration of both observed and latent modeling approaches is emphasized in two empirical applications. The first application examines linkages for banking contagion and the second application evaluates the impact of education on socioeconomic outcomes.

The third chapter, which is joint work with Professor Ivan Jeliazkov, studies the formulation of the likelihood function for simultaneous equation models for discrete data. The approach rests on casting the required distribution as the invariant distribution of a suitably defined Markov chain. The derivation resolves puzzling paradoxes highlighted in earlier work, shows that such models are theoretically coherent, and offers simple and intuitive linkages to the better understood analysis of continuous outcomes. The new methodology is employed in two applications involving simultaneous equation models of (i) female labor supply and family financial stability, and (ii) the interactions between health and wealth.

Chapter 1

Sample Selection and Treatment Effect Estimation of Lender of Last Resort Policies

1.1 Introduction

Many policies and programs that are the foci of economic research operate within a specific decision structure. This is commonly recognized as an application stage and an approval stage, where interested parties have to apply to be considered for a treatment and a governing body reviews the submitted applications and assigns treatment. The econometric complexities involved in dealing with multiple selection mechanisms and a larger set of treatment response outcomes have moved attention away from this decision structure toward simple, potentially misspecified models. This paper offers a general methodological framework for these application-approval settings, and extends conventional treatment models to allow for non-randomly missing data, endogeneity, and discrete outcomes. The model is applied to study the effectiveness of lender of last resort (LOLR) policies and their ability to resuscitate the banking system.

Existing treatment models consider two subgroups in the data, the treated group and the untreated (control) group. When this structure is applied to an LOLR study, it divides the sample into banks that receive assistance from the LOLR and banks that do not receive assistance. Complications arise because the initial selection mechanism in which banks choose to apply for assistance is ignored. Overlooking the application stage erroneously groups banks that do not apply for assistance with those that are declined assistance. Thus, the untreated group comprises the most and least healthy banks, leading to a fundamental misspecification. Motivated by these difficulties, this paper develops and implements a multivariate treatment effect model in the presence of sample selection to offer a framework for evaluating the impact of LOLR programs and banking policies.

The empirical focus of this paper is on the Reconstruction Finance Corporation (RFC) as the LOLR during the Great Depression. The RFC was a government-sponsored rescue program created to provide assistance to weak banks to reduce the incidence of bank failure and alleviate banking crises. It is one of the largest recapitalization programs ever implemented and continues to be an important case study for understanding and preventing financial panics, especially with regard to the recent financial crisis. The analysis of LOLR policies poses a number of challenges because regulator data are generally not publicly available and modeling must accommodate a difficult decision structure, endogenous treatments, correlation between outcomes, and non-random selection into policies and programs. This paper contributes to the existing literature on emergency financial regulation by constructing a novel bank-level data set and employing the new methodology to jointly model a bank's decision to apply for assistance from the LOLR, the LOLR's decision to approve the assistance, and the bank's performance following the disbursements. It should be noted that the model is not limited to banking contexts and other literatures face these obstacles as well, e.g., labor supply decisions, college admittance studies, credit approval decisions, health outcomes and drug treatments, welfare or housing program evaluation, and many others. Any scenario that features an application-approval decision structure can utilize the model and estimation strategy presented in this paper.

Figure 1.1 displays a graphical presentation of the multivariate treatment effect model in the presence of sample selection. The initial selection mechanism represents the application decision and is observed for every unit in the sample. The selected sample then enters a selected treatment (approval) stage where the program determines treatment assignment. Note that this stage is not observed for the non-selected sample. Following these 2 selection mechanisms are 3 potential outcomes for the non-selected (non-applicant) sample, selected untreated (applied-declined) sample, and selected treated (applied-approved) sample. The model considered here differs from existing work on treatment assignment by acknowledging whether an observation opts into or out of treatment to disentangle the information content in not selecting. Modeling this additional selection mechanism offers critical features of the data because without it, it is impossible to separately identify non-selected and selected untreated observations. In the LOLR study, the initial selection mechanism, deciding whether or not to apply for assistance from the central bank, is a very revealing step in the bailout process. Recognizing and modeling the decisions for the applicant and non-applicant groups allow for a broader understanding of the outcomes of interest.

This methodology incorporates and jointly models both sample selection and potential outcomes. Individually, these models are used frequently in economics. Sample selection (i.e., incidental truncation or informative missingness) arises when a depen-

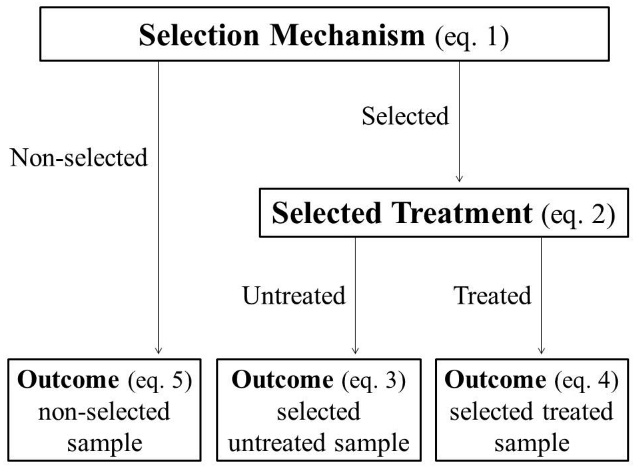

Figure 1.1: Multivariate treatment effect model in the presence of sample selection.

dent variable of interest is non-randomly missing for a subset of the sample. The factors that determine whether or not data are missing for an observation are often correlated with those that determine an outcome. Ignoring sample selection causes a researcher to base inference on a sample that does not represent the population of interest, which leads to specification errors. Classical estimation methods of sample selection models are developed in Gronau (1974) and Heckman (1976, 1979), and are further discussed in Wooldridge (2002). Bayesian developments in these models can be found in Greenberg (2008), Chib et al. (2009), van Hasselt (2011), and are discussed in van Hasselt (2014). The framework and estimation strategy for multiple selection mechanisms are discussed in Yen (2005), Li (2011), and Vossmeyer (2014).

Treatment models, also referred to as models of potential outcomes, are employed to compare responses of individuals that belong either to a treatment or control group (Roy, 1951; Rubin, 1978; Heckman and Honoré, 1990; Heckman and Vytlacil, 2007). These models feature two potential responses; however, only one is ever observed, the other is the counterfactual. Bayesian approaches motivated by the missingness of the counterfactual are given in Vijverberg (1993) and Poirier and Tobias (2003), who formulate an analysis for this model with the joint distribution of the potential outcomes. This involves placing a prior on the non-identified elements in the variance-covariance matrix, and simulating a posterior of the parameters and the counterfactuals. Alternatively, Chib (2007) provides a Bayesian analysis without the joint distribution of the potential outcomes. Chib's (2007) approach does not involve simulating the counterfactuals and as a result, is simpler in terms of prior inputs, computational burden, and improves the mixing properties of the Markov chain. For a discussion of Bayesian approaches to treatment models, see Li and Tobias (2014).

This paper contributes to the vast literature on sample selection and treatment effect models by extending the techniques in Chib (2007) and Chib et al. (2009), and developing a Bayesian framework for treatment effect modeling while dealing with the missing data that occur from sample selection. Furthermore, this paper designs a computationally efficient estimation algorithm that does not require simulation of the missing outcomes or the joint distribution of the potential outcomes and offers techniques for model comparison and treatment effect calculations. The rest of the paper is organized as follows: Section 1.2 presents the model, Section 1.3 presents estimation methods to fit the model, and Section 1.4 reports the performance of these techniques in a simulation study. The new methodology is applied to study bank recapitalization in Section 1.5. Section 1.6 contains additional considerations, including model comparison and sensitivity analysis, and finally, Section 1.7 offers concluding remarks.

1.2 The Model

The model stemming from Figure 1.1 contains 5 equations of interest: 1 selection mechanism, 1 selected treatment, and 3 treatment response outcomes for the different subsets of the sample (non-selected, selected untreated, and selected treated).

Note that the model is generalizable to larger systems of equations and additional endogenous regressors. For specificity, the model considered here will contain 5 equations. In detail, below are the equations for subjects $i = 1, \ldots, n$:

$$\textit{Selection Mechanism: } y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} \text{ (always observed)} \tag{1.1}$$

$$\textit{Selected Treatment: } y_{i2}^* = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2} \text{ (observed for selected sample)} \tag{1.2}$$

Potential Outcomes − Treatment Responses (only one is observed)

$$\text{Selected untreated sample: } y_{i3}^* = (\mathbf{x}_{i3}' \; y_{i1})\boldsymbol{\beta}_3 + \varepsilon_{i3} \tag{1.3}$$

$$\text{Selected treated sample: } y_{i4}^* = (\mathbf{x}_{i4}' \; y_{i1} \; y_{i2})\boldsymbol{\beta}_4 + \varepsilon_{i4} \tag{1.4}$$

$$\text{Non-selected sample: } y_{i5}^* = \mathbf{x}_{i5}'\boldsymbol{\beta}_5 + \varepsilon_{i5} \tag{1.5}$$

The model is characterized by 5 dependent variables of interest where $\mathbf{y}_i^* \equiv (y_{i1}^*, y_{i2}^*, y_{i3}^*, y_{i4}^*, y_{i5}^*)'$ are the continuous latent data and $\mathbf{y}_i \equiv (y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5})'$ are the corresponding discrete observed data. In the application, the latent variables relate to the observed censored outcomes by $y_{ij} = y_{ij}^* \cdot 1\{y_{ij}^* > 0\}$ for equations $j = 1, \ldots, 5$, which is the basis for the model throughout the paper (Tobin, 1958). However, the general system can take outcome variables that are continuous, binary, censored, or ordered. The continuous setting occurs when $y_{ij}^* = y_{ij}$, the binary setting is when $y_{ij} = 1\{y_{ij}^* > 0\}$, and the ordered setting is when $y_{ij} = \sum_{h=1}^{H} 1\{y_{ij}^* > \delta_{h-1}\}$ for H ordered alternatives where δ_h is a cutpoint between the categories. Note that y_{i1} and y_{i2} enter potential outcome equations (1.3) and (1.4) as endogenous regressors for the selected sample. This can be understood as the requested and approved treatments entering the performance equations. Treatment effects are calculated from these regressors and are discussed in Section 1.5.2.b.

Data missingness restricts the model to systems of 2 or 3 equations depending on the subsample to which the observation belongs and highlights the presence of non-identified parameters that will be examined shortly. If $y_{i1} = 0$, the observation is in the non-selected sample – y_{i1} and y_{i5} are observed, and y_{i2}, y_{i3}, and y_{i4} are not observed. If $y_{i1} > 0$ and $y_{i2} = 0$, the observation is in the selected untreated sample – y_{i1}, y_{i2} and y_{i3} are observed, and y_{i4} and y_{i5} are not observed. If $y_{i1} > 0$ and $y_{i2} > 0$, the observation is in the selected treated sample – y_{i1}, y_{i2} and y_{i4} are observed, and y_{i3} and y_{i5} are not observed.

The exogenous covariates $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3}, \mathbf{x}_{i4}, \mathbf{x}_{i5})$ are needed only when their corresponding equations are observed. For identification reasons, assume that the covariates in $\mathbf{x}_{i2}$ contain at least one more variable than those included in the other equations. This variable is regarded as the instrumental variable used in treatment models that is correlated with the treatment and not the errors (Chib, 2007; Greenberg, 2008). Although identification in models with incidental truncation does not require exclusions, they are typically empolyed so the resulting inference does not solely depend on distributional assumptions. Finally, the model assumes that the errors $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4}, \varepsilon_{i5})'$ have a multivariate normal distribution $\mathcal{N}_5(0, \mathbf{\Omega})$, where $\mathbf{\Omega}$ is an unrestricted symmetric positive definite matrix. Restrictions placed on this matrix can occur due to model variants, such as a probit selection mechanism. Algorithm adjustments due to these restrictions can be found in Chib et al. (2009). It is possible to explore other distributional forms for this joint model, but the normality assumption provides the groundwork for more flexible distributions, including finite mixtures, dirichlet processes, and scale mixtures.

1.2.1 The Likelihood Function

For the i-th observation, define the following vectors and matrices,

$$\mathbf{y}_{iC}^* = (y_{i1}^*, y_{i5}^*)', \qquad \mathbf{y}_{iD}^* = (y_{i1}^*, y_{i2}^*, y_{i3}^*)', \qquad \mathbf{y}_{iA}^* = (y_{i1}^*, y_{i2}^*, y_{i4}^*)',$$

$$\mathbf{X}_{iC} = \begin{pmatrix} \mathbf{x}_{i1}' & 0 \\ 0 & \mathbf{x}_{i5}' \end{pmatrix}, \quad \mathbf{X}_{iD} = \begin{pmatrix} \mathbf{x}_{i1}' & 0 & 0 \\ 0 & \mathbf{x}_{i2}' & 0 \\ 0 & 0 & (\mathbf{x}_{i3}' \; y_{i1}) \end{pmatrix}, \quad \mathbf{X}_{iA} = \begin{pmatrix} \mathbf{x}_{i1}' & 0 & 0 \\ 0 & \mathbf{x}_{i2}' & 0 \\ 0 & 0 & (\mathbf{x}_{i4}' \; y_{i1} \; y_{i2}) \end{pmatrix}.$$

Let $N_1 = \{i : y_{i1} = 0\}$ be the n_1 observations in the non-selected sample and $N_2 = \{i : y_{i1} > 0 \textit{ and } y_{i2} = 0\}$ be the n_2 observations in the selected untreated sample. Set $N_3 = \{i : y_{i1} > 0 \textit{ and } y_{i2} > 0\}$ to be the n_3 observations in the selected treated sample. Let $\boldsymbol{\theta}$ be the set of all model parameters.

Upon defining $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', \boldsymbol{\beta}_5')'$ and

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} & \Omega_{15} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} & \Omega_{25} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} & \Omega_{35} \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} & \Omega_{45} \\ \Omega_{51} & \Omega_{52} & \Omega_{53} & \Omega_{54} & \Omega_{55} \end{pmatrix},$$

note that in $\boldsymbol{\Omega}$, the elements Ω_{25}, Ω_{35}, Ω_{45}, and Ω_{34} are not identified because their corresponding equations cannot be observed at the same time. Thus, there are 11 unique estimable elements in $\boldsymbol{\Omega}$, whereas the remaining ones are non-identified parameters due to the missing outcomes. The covariance matrix of interest is,

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} & \Omega_{15} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} & \cdot \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \cdot & \cdot \\ \Omega_{41} & \Omega_{42} & \cdot & \Omega_{44} & \cdot \\ \Omega_{51} & \cdot & \cdot & \cdot & \Omega_{55} \end{pmatrix}.$$

In order to isolate the observed vectors and matrices that correspond to the 3 different subsets of the sample, define

$$\mathbf{J}_C = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{pmatrix}_{(k_1+k_5)\times K}, \ \mathbf{J}_D = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I} & 0 & 0 \end{pmatrix}_{(k_1+k_2+k_3)\times K},$$

$$\mathbf{J}_A = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I} & 0 \end{pmatrix}_{(k_1+k_2+k_4)\times K}$$

where $K = k_1 + k_2 + k_3 + k_4 + k_5$, which represents the number of covariates in each equation, so

$$\mathbf{J}_C\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_5')', \quad \mathbf{J}_D\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')', \quad \mathbf{J}_A\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_4')'$$

$$\mathbf{\Omega}_C = \begin{pmatrix} \Omega_{11} & \Omega_{15} \\ \Omega_{51} & \Omega_{55} \end{pmatrix}, \ \mathbf{\Omega}_D = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}, \ \mathbf{\Omega}_A = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{24} \\ \Omega_{41} & \Omega_{42} & \Omega_{44} \end{pmatrix}. \tag{1.6}$$

For $i \in N_1$ (non-selected sample),

$$f(\mathbf{y}^*_{iC}|\boldsymbol{\theta}) \propto |\boldsymbol{\Omega}_C|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}^*_{iC} - \mathbf{X}_{iC}\mathbf{J}_C\boldsymbol{\beta})'\boldsymbol{\Omega}_C^{-1}(\mathbf{y}^*_{iC} - \mathbf{X}_{iC}\mathbf{J}_C\boldsymbol{\beta})\right\}$$

$$\boldsymbol{\eta}^*_{iC} = \mathbf{y}^*_{iC} - \mathbf{X}_{iC}\mathbf{J}_C\boldsymbol{\beta},$$

for $i \in N_2$ (selected untreated sample),

$$f(\mathbf{y}^*_{iD}|\boldsymbol{\theta}) \propto |\boldsymbol{\Omega}_D|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}^*_{iD} - \mathbf{X}_{iD}\mathbf{J}_D\boldsymbol{\beta})'\boldsymbol{\Omega}_D^{-1}(\mathbf{y}^*_{iD} - \mathbf{X}_{iD}\mathbf{J}_D\boldsymbol{\beta})\right\}$$

$$\boldsymbol{\eta}^*_{iD} = \mathbf{y}^*_{iD} - \mathbf{X}_{iD}\mathbf{J}_D\boldsymbol{\beta},$$

and for $i \in N_3$ (selected treated sample),

$$f(\mathbf{y}^*_{iA}|\boldsymbol{\theta}) \propto |\boldsymbol{\Omega}_A|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}^*_{iA} - \mathbf{X}_{iA}\mathbf{J}_A\boldsymbol{\beta})'\boldsymbol{\Omega}_A^{-1}(\mathbf{y}^*_{iA} - \mathbf{X}_{iA}\mathbf{J}_A\boldsymbol{\beta})\right\}$$

$$\boldsymbol{\eta}^*_{iA} = \mathbf{y}^*_{iA} - \mathbf{X}_{iA}\mathbf{J}_A\boldsymbol{\beta},$$

which provide the terms for the complete-data likelihood, $f(\mathbf{y}, \mathbf{y}^*|\boldsymbol{\theta})$. This is given by,

$$\begin{aligned}\left[\prod_{i\in N_1} f(\boldsymbol{y}^*_{iC}|\boldsymbol{\theta})\right]\left[\prod_{i\in N_2} f(\boldsymbol{y}^*_{iD}|\boldsymbol{\theta})\right]\left[\prod_{i\in N_3} f(\boldsymbol{y}^*_{iA}|\boldsymbol{\theta})\right] &\propto |\boldsymbol{\Omega}_C|^{-n_1/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_1} \boldsymbol{\eta}^{*\prime}_{iC}\boldsymbol{\Omega}_C^{-1}\boldsymbol{\eta}^*_{iC}\right\} \\ \times |\boldsymbol{\Omega}_D|^{-n_2/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_2} \boldsymbol{\eta}^{*\prime}_{iD}\boldsymbol{\Omega}_D^{-1}\boldsymbol{\eta}^*_{iD}\right\} &\times |\boldsymbol{\Omega}_A|^{-n_3/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_3} \boldsymbol{\eta}^{*\prime}_{iA}\boldsymbol{\Omega}_A^{-1}\boldsymbol{\eta}^*_{iA}\right\}. \end{aligned} \tag{1.7}$$

The likelihood is defined in terms of the 3 subsets of the sample, without components for the non-identified parameters, which follows from Chib (2007). This decomposition is the basis for the computationally efficient estimation algorithm. Chib's (2007) approach is employed here, as opposed to modeling the non-identified parameters as in Vijverberg (1993) and Poirier and Tobias (2003), mainly for computational ease. The censoring of multiple outcome variables renders this likelihood analytically intractable and hence estimation relies on simulation-based techniques.

1.2.2 Prior Distributions

The model is completed by specifying the prior distributions for the parameters. It is assumed that $\boldsymbol{\beta}$ has a joint normal distribution with mean $\boldsymbol{\beta}_0$ and variance $\mathbf{B}_0$ and (independently) that the covariance matrix $\boldsymbol{\Omega}$ has an inverse Wishart distribution with parameters υ and $\mathbf{Q}$, so that the prior density is given by,

$$\pi(\boldsymbol{\beta}, \boldsymbol{\Omega}) = f_{\mathcal{N}}(\boldsymbol{\beta}\ |\boldsymbol{\beta}_0,\ \mathbf{B}_0) f_{\mathcal{IW}}(\boldsymbol{\Omega}|\upsilon, \mathbf{Q}). \tag{1.8}$$

Note that the prior on $\boldsymbol{\Omega}$ implies a distribution on functions of the elements in $\boldsymbol{\Omega}$ that will be used in sampling and marginal likelihood computations, which only involve the identified components. Further details on this approach are provided in the updating formulas for $\boldsymbol{\Omega}$ in Section 1.3.2. While Vijverberg's (1993) and Poirier and Tobias's (2003) approach is not adopted in this paper, specifying the prior in such a way makes this model easily generalizable to their approach of learning about the non-identified parameters.

1.3 Estimation

Combining the complete-data likelihood in (1.7) and the priors in (1.8) leads to a posterior distribution for $\boldsymbol{\theta}$ and $\mathbf{y}^*$,

$$\pi(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) \propto \left[\prod_{i\in N_1} f(\mathbf{y}^*_{iC}|\boldsymbol{\theta})\right] \left[\prod_{i\in N_2} f(\mathbf{y}^*_{iD}|\boldsymbol{\theta})\right] \left[\prod_{i\in N_3} f(\mathbf{y}^*_{iA}|\boldsymbol{\theta})\right] \times \pi(\boldsymbol{\beta}, \boldsymbol{\Omega}), \tag{1.9}$$

which is simulated by Markov chain Monte Carlo (MCMC) methods. The sampling algorithm is summarized as follows, where the notation "\" represents "except", e.g., $\mathbf{y}^*\backslash\ \mathbf{y}^*_1$ says all elements in $\mathbf{y}^*$ except $\mathbf{y}^*_1$:

Algorithm 1.1. *MCMC Estimation Algorithm for Censored Outcomes*

1. *Sample* $\boldsymbol{\beta}$ *from the distribution* $\boldsymbol{\beta}\ |\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\beta}$.

2. *Sample* $\boldsymbol{\Omega}$ *from the distribution* $\boldsymbol{\Omega}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\Omega}$ *in a 1-block, multi-step procedure.*

3. *For* $i \in N_1$, *sample* y^*_{i1} *from the distribution* $y^*_{i1}|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus\ \mathbf{y}^*_1$.

4. *For* $i \in N_2$, *sample* y^*_{i2} *from the distribution* $y^*_{i2}|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus\ \mathbf{y}^*_2$.

5. *For* $i : y_{i3} = 0$, *sample* y^*_{i3} *from the distribution* $y^*_{i3}|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus\ \mathbf{y}^*_3$.

6. *For* $i : y_{i4} = 0$, *sample* y^*_{i4} *from the distribution* $y^*_{i4}|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus\ \mathbf{y}^*_4$.

7. *For* $i : y_{i5} = 0$, *sample* y^*_{i5} *from the distribution* $y^*_{i5}|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus\ \mathbf{y}^*_5$.

Following Chib et al. (2009), the algorithm does not augment the selected sample with data from the potential sample. Not augmenting the full sample and employing a collapsed Gibbs sampler eases computational and storage demands, and provides improved simulation performance (Liu, 1994; Liu et al., 1994; Chib et al., 2009; Li, 2011). Furthermore, simulation of the posterior distribution is of the parameters and not the counterfactuals to simplify the prior inputs and improve the mixing properties of the Markov chain (Chib, 2007). Details of the sampler are in the following subsections.

1.3.1 Sampling $\boldsymbol{\beta}$

The posterior distribution displayed in (1.9) implies $\boldsymbol{\beta}\,|\mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}, \mathbf{B})$, where

$$\begin{aligned}\mathbf{b} =& \mathbf{B}(\mathbf{B}_0^{-1}\mathbf{b}_0 + \sum_{i\in N_1} \mathbf{J}_C'\mathbf{X}_{iC}'\boldsymbol{\Omega}_C^{-1}\mathbf{y}_{iC}^* + \sum_{i\in N_2} \mathbf{J}_D'\mathbf{X}_{iD}'\boldsymbol{\Omega}_D^{-1}\mathbf{y}_{iD}^* + \\ & \sum_{i\in N_3} \mathbf{J}_A'\mathbf{X}_{iA}'\boldsymbol{\Omega}_A^{-1}\mathbf{y}_{iA}^*), \\ \mathbf{B} =& (\mathbf{B}_0^{-1} + \sum_{i\in N_1} \mathbf{J}_C'\mathbf{X}_{iC}'\boldsymbol{\Omega}_C^{-1}\mathbf{X}_{iC}\mathbf{J}_C + \sum_{i\in N_2} \mathbf{J}_D'\mathbf{X}_{iD}'\boldsymbol{\Omega}_D^{-1}\mathbf{X}_{iD}\mathbf{J}_D + \\ & \sum_{i\in N_3} \mathbf{J}_A'\mathbf{X}_{iA}'\boldsymbol{\Omega}_A^{-1}\mathbf{X}_{iA}\mathbf{J}_A)^{-1}.\end{aligned}$$

Computations for $\boldsymbol{\beta}$ are done efficiently by updating observations in the N_1, N_2, and N_3 subsamples separately using the $\mathbf{J}$ matrices to select the relevant covariates. Thereby, the observed parts of the model are isolated and estimation proceeds without sampling conditional on unobserved components.

1.3.2 Sampling $\boldsymbol{\Omega}$

The sampling of $\boldsymbol{\Omega}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\Omega}$ is complicated because $\boldsymbol{\Omega}$ never enters the complete-data likelihood in (1.7) in its full form, and instead, enters as in (1.6) where $\boldsymbol{\Omega}$ is different for the 3 subgroups, N_1, N_2, and N_3. Thus, sampling must be done in multiple steps, or layers, for the different indices, as opposed to a single inverse Wishart step. Utilizing the sampling techniques in Chib et al. (2009), this paper initially samples from $\Omega_{11}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \Omega_{11}$ followed by sampling from the conditionals: $\boldsymbol{\Omega}_{tt\cdot l}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\Omega}_{tt\cdot l}$ and $\mathbf{B}_{lt}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\Omega}_{tt\cdot l}$, where

$$\begin{aligned}\boldsymbol{\Omega}_{tt\cdot l} &= \boldsymbol{\Omega}_{tt} - \boldsymbol{\Omega}_{tl}\boldsymbol{\Omega}_{ll}^{-1}\boldsymbol{\Omega}_{lt} \\ \mathbf{B}_{lt} &= \boldsymbol{\Omega}_{ll}^{-1}\boldsymbol{\Omega}_{lt},\end{aligned} \tag{1.10}$$

and t and l are indices for the elements in $\mathbf{\Omega}$ being updated.

Partition the hyperparameter $\mathbf{Q}$ from the inverse Wishart prior

$$\mathbf{Q} = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} & Q_{14} & Q_{15} \\ Q_{21} & Q_{22} & Q_{23} & Q_{24} & \cdot \\ Q_{31} & Q_{32} & Q_{33} & \cdot & \cdot \\ Q_{41} & Q_{42} & \cdot & Q_{44} & \cdot \\ Q_{51} & \cdot & \cdot & \cdot & Q_{55} \end{pmatrix},$$

where $\mathbf{Q}$ conforms suitably with $\mathbf{\Omega}$ for the 3 subgroups. To sample from the set of conditionals in (1.10), a change of variable technique is employed and the resulting density is proportional to a set of inverse Wisharts and matrix-variate normals. Thus, $\mathbf{\Omega}$ updates in a 1-block, multi-step procedure that samples in layers and conditions only on the identified parts of the model. Similar techniques are used in Li (2011), where he shows the computational efficiency of this approach. The step-by-step algorithm is described here.

From step 2 of Algorithm 1.1, sample $\mathbf{\Omega}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega}$ in a 1-block, nine-step procedure by drawing Ω_{11} , $\mathbf{\Omega}_{tt\cdot l} = \mathbf{\Omega}_{tt} - \mathbf{\Omega}_{tl}\mathbf{\Omega}_{ll}^{-1}\mathbf{\Omega}_{lt}$, and $\mathbf{B}_{lt} = \mathbf{\Omega}_{ll}^{-1}\mathbf{\Omega}_{lt}$, and then reconstructing $\mathbf{\Omega}$ from these quantities

2. (a) $\Omega_{11}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega} \sim \mathcal{IW}(\nu - 1 + n,\ Q_{11} + \sum_{N_1,N_2,N_3} \boldsymbol{\eta}_{i1}^* \boldsymbol{\eta}_{i1}^{*'})$
   i. $\boldsymbol{\eta}_{i1}^* = y_{i1}^* - \mathbf{x}_{i1}\mathbf{J}_1\boldsymbol{\beta},$ where $\mathbf{J}_1 = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \end{bmatrix}_{k_1 \times K}$
   (b) $\Omega_{22\cdot 1}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega} \sim \mathcal{IW}(\nu + n_2 + n_3, R_{22\cdot 1})$
   (c) $B_{12}|\mathbf{y}, \mathbf{y}^*, \Omega_{22\cdot 1} \sim \mathcal{MN}(R_{11}^{-1}R_{21}, \Omega_{22\cdot 1} \otimes R_{11}^{-1})$
   (d) Define $\mathbf{\Omega}_u = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$
   (e) $\Omega_{55\cdot 1}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega} \sim \mathcal{IW}(\nu + n_1, R_{55\cdot 1})$

(f) $B_{15}|\mathbf{y}, \mathbf{y}^*, \Omega_{55\cdot 1} \sim \mathcal{MN}(R_{11}^{-1}R_{51}, \Omega_{55\cdot 1} \otimes R_{11}^{-1})$

(g) $\mathbf{\Omega}_{33\cdot u}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega} \sim \mathcal{IW}(\nu + n_2, \mathbf{R}_{33\cdot u})$

(h) $\mathbf{B}_{u3}|\mathbf{y}, \mathbf{y}^*, \mathbf{\Omega}_{33\cdot u} \sim \mathcal{MN}(\mathbf{R}_u^{-1}\mathbf{R}_{3u}, \mathbf{\Omega}_{33\cdot u} \otimes \mathbf{R}_u^{-1})$

(i) $\mathbf{\Omega}_{44\cdot u}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \mathbf{\Omega} \sim \mathcal{IW}(\nu + n_3, \mathbf{R}_{44\cdot u})$

(j) $\mathbf{B}_{u4}|\mathbf{y}, \mathbf{y}^*, \mathbf{\Omega}_{44\cdot u} \sim \mathcal{MN}(\mathbf{R}_u^{-1}\mathbf{R}_{4u}, \mathbf{\Omega}_{44\cdot u} \otimes \mathbf{R}_u^{-1})$

where $\mathbf{R} = \mathbf{Q} + \sum \boldsymbol{\eta}_i^* \boldsymbol{\eta}_i^{*\prime}$, *and the following subsections are obtained by partitioning* $\mathbf{R}$ *to conform to* $\mathbf{Q}$, *and* $\mathbf{R}_{tt\cdot l} = \mathbf{R}_{tt} - \mathbf{R}_{tl}\mathbf{R}_{ll}^{-1}\mathbf{R}_{lt}$. *From these sampling densities,* $\mathbf{\Omega}$ *can be recovered.*

1.3.3 Sampling $\mathbf{y}^*$

For censored outcomes, $\mathbf{y}^*$ is sampled following Chib (1992) from a truncated normal with the usual conditional mean and conditional variance,

$$
\begin{aligned}
y_{i1}^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_1^* &\sim \mathcal{TN}_{(-\infty,0)}(\mathbf{x}_{i1}'\boldsymbol{\beta}_1 + E(\varepsilon_{i1}|\varepsilon_{i\setminus 1}), var(\varepsilon_{i1}|\varepsilon_{i\setminus 1})), \quad i \in N_1,\\
y_{i2}^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_2^* &\sim \mathcal{TN}_{(-\infty,0)}(\mathbf{x}_{i2}'\boldsymbol{\beta}_2 + E(\varepsilon_{i2}|\varepsilon_{i\setminus 2}), var(\varepsilon_{i2}|\varepsilon_{i\setminus 2})), \quad i \in N_2,\\
y_{i3}^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_3^* &\sim \mathcal{TN}_{(-\infty,0)}((\mathbf{x}_{i3}'\ y_{i1})\boldsymbol{\beta}_3 + E(\varepsilon_{i3}|\varepsilon_{i\setminus 3}), var(\varepsilon_{i3}|\varepsilon_{i\setminus 3})), \quad i : y_{i3} = 0,\\
y_{i4}^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_4^* &\sim \mathcal{TN}_{(-\infty,0)}((\mathbf{x}_{i4}'\ y_{i1}\ y_{i2})\boldsymbol{\beta}_4 + E(\varepsilon_{i4}|\varepsilon_{i\setminus 4}), var(\varepsilon_{i4}|\varepsilon_{i\setminus 4})), \quad i : y_{i4} = 0,\\
y_{i5}^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_5^* &\sim \mathcal{TN}_{(-\infty,0)}(\mathbf{x}_{i5}'\boldsymbol{\beta}_5 + E(\varepsilon_{i5}|\varepsilon_{i\setminus 5}), var(\varepsilon_{i5}|\varepsilon_{i\setminus 5})), \quad i : y_{i5} = 0.
\end{aligned}
$$

1.4 Simulation Study

This section evaluates the performance of the algorithm from Section 1.3 using simulated data. The model being considered is the system of 5 equations from (1.1)–(1.5), which is motivated by the subsequent application to bank recapitalization in Section 1.5. The simulated data contains 4 explanatory variables in the first equation, 5 in the second, 5 in the third (including $\mathbf{y}_1$), 6 in the fourth (including $\mathbf{y}_1$ and $\mathbf{y}_2$), and 4 in the fifth equation. Therefore, $\boldsymbol{\beta}$ is a 24×1 vector. Similarly, $vech(\boldsymbol{\Omega})$ has 11 unique estimable elements. Generalizations to more variables do not require conceptual changes to the estimation algorithm.

For the study, 36% of the observations are in the N_1 subsample, 16% are in the N_2 subset, and 48% are in the N_3 subset for $n = 1000$. The priors set for the equations are: $\boldsymbol{\beta} \sim \mathcal{N}(0, 5 \times I)$ and $\boldsymbol{\Omega} \sim \mathcal{IW}(9, 1.2 \times I_5)$. These priors and subsamples were selected to mimic the following bailout application. Note that various combinations of priors, observations, censoring, and explanatory variables were considered but the results are not presented as they did not vary much from the performance pattern in this base case. Posterior mean estimates are based on 11,000 MCMC draws with a burn-in of 1,000. The total run time of the algorithm is about 1 minute and 8 seconds. The true values are uncovered accurately and quickly. To exemplify the correct sampling, posterior means for $vech(\boldsymbol{\Omega}) = (0.24, 0.10, 0.25, 0.08, 0.12, 0.24, 0.09, 0.10, 0.26, 0.07, 0.23)'$ and the corresponding true values are $(0.25, 0.10, 0.25, 0.10, 0.10, 0.25, 0.10, 0.10, 0.25, 0.10, 0.25)'$. For reference, the posterior standard deviations are $(0.01, 0.01, 0.01, 0.02, 0.03, 0.03, 0.01, 0.01, 0.02, 0.03, 0.02)'$. Benefits of this sampling approach include the computational speed and the low storage costs, which arise from collapsing the Gibbs sampler.

Further evaluations of the sampler are studied with inefficiency factors over 25 Monte Carlo repetitions. Inefficiency factors are a "measure of the extent of mixing of the Markov chain output" (Chib, 2007). The inefficiency factor of the k-th parameter is defined as $1 + 2\sum_{l=1}^{L} \rho_k(l)(\frac{L-l}{L})$, where $\rho_k(l)$ is the sample autocorrelation at the l-th lag and L is the lag in which the autocorrelations taper off (Chib et al., 2009). Small values (near 1) imply that the output is mixing well. Boxplots of the inefficiency factors are displayed in Figure 1.2.

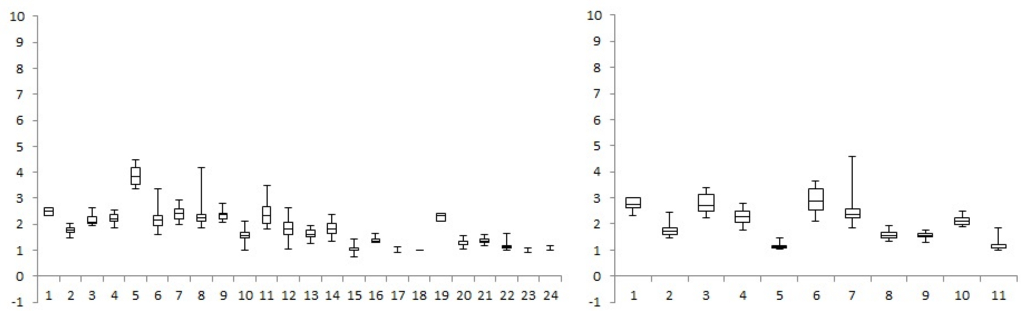

Figure 1.2: Boxplots of inefficiency factors for $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$. The left panel is $\boldsymbol{\beta}$, the right panel is $vech(\boldsymbol{\Omega})$: $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', \boldsymbol{\beta}_5')'$, $vech(\boldsymbol{\Omega}) = (\Omega_{11}, \Omega_{21}, \Omega_{22}, \Omega_{31}, \Omega_{32}, \Omega_{33}, \Omega_{41}, \Omega_{42}, \Omega_{44}, \Omega_{51}, \Omega_{55})'$.

The plots suggest these parameters are sampled efficiently as the values for all of the parameters are low and near one. The fifth β has the highest inefficiency factor. This variable is the constant for equation 2, and poor mixing of the constant is also found in Li (2011). In addition, the inefficiency factors for $\boldsymbol{\Omega}$ vary between 1 and 4, with the higher values occurring where $\mathbf{y}^*$ is being sampled more frequently. Given that every dependent variable in this system is censored, the inefficiency factors are promising. The lack of augmentation in the sampler, which arises from not simulating the outcomes that are missing due to the selection mechanism and non-identified parameters in the treatment outcomes, shows decreased storage costs while still maintaining tractability in the sampling densities. Chib et al. (2009) and Li (2011) compare samplers with less augmentation to samplers that simulate missing outcomes and both studies find improved sampler performance in the former case. Following their results, the algorithm developed here does not augment the outcomes that are missing or the counterfactuals, and the results show excellent sampler performance.

1.5 Application

This section applies the multivariate treatment effect model with sample selection to study LOLR polices and bank recapitalization. Bailout programs and the central bank's LOLR function emphasize a trade-off between liberal and strict lending policies. Research in favor of these programs finds that LOLR policies play a positive role in reducing bank failures and improving monetary conditions (Butkiewicz, 1995; Richardson and Troost, 2009). Loose lending policies can prevent the spread of contagion, bank runs, and mass liquidation. Other studies find that rescue programs can be harmful either by requiring banks to hypothecate their best collateral or by creating moral hazard incentives for banks to take on excessive risk (Mason, 2001; Freixas and Rochet, 2008). These issues have been deliberated since the concept of LOLR was described in the 19th-century by Bagehot (1873). Bagehot established modern LOLR theory which states that monetary authorities, in the face of panic, should lend unsparingly at a penalty rate to illiquid but solvent banks. This mechanism should prevent struggling healthy banks from falling victim to undue deposit losses, bank runs, and insolvency.

This paper focuses on the Reconstruction Finance Corporation (RFC) as the LOLR during the Great Depression.[1] The RFC was established in early 1932 during the Hoover administration with the primary objective of providing liquidity to the banking system, and later became part of the New Deal under President Roosevelt. It was created as a government-sponsored agency of the Executive Branch of the United States. The RFC operated under two regimes. From February 1932–March 1933, the RFC made loans collateralized by banks' best assets. The Emergency Banking Act of 1933 liberalized the powers of the RFC and allowed it to recapitalize banks through

[1] At its conception, the RFC was an LOLR and closely followed Bagehot's theory. Over time, the power of the RFC expanded beyond the usual LOLR duties. However, the LOLR terminology is used here because the paper focuses on the RFC's role in the banking system.

preferred stock purchases. Despite the changes in the program, the application process for banks remained similar across both periods. Therefore, this paper looks at the effectiveness of the RFC across both regimes.[2] For a thorough explanation of the RFC and its operations, see Butkiewicz (1995) and Mason (2001, 2003).

Previous work on the LOLR function of the RFC includes Butkiewicz (1995), Mason (2001), Calomiris et al. (2013), and Vossmeyer (2014). Examining the RFC presents limitations because data are not readily available and need to be hand-coded from record books. As a result, many of these previous papers either look at a time series of RFC lending (Butkiewicz, 1995) or bank-level data restricted to Federal Reserve member banks (Mason, 2001; Calomiris et al., 2013). In addition, dealing with sample selection is difficult because the quarterly and monthly *Reports of Activities of the Reconstruction Finance Corporation* do not report applied or declined assistance. This paper overcomes these limitations and contributes to this literature by constructing a novel bank-level data set from the original applications to the RFC. With this more detailed data, the new methodology can be employed to jointly model a bank's decision to apply for assistance from the RFC, the RFC's decision to approve the assistance, and the bank's performance following the disbursements. This modeling structure can be seen in Figure 1.3 (which is the applied version of Figure 1.1).

Figure 1.3 displays a system of 5 equations as in (1.1)–(1.5). The initial selection mechanism (equation 1.1) represents a bank's decision to apply for assistance from the RFC and is observed for every bank in the sample. The selected sample of banks that apply for assistance enters the selected treatment stage, whereas, the non-applicant sample is missing in the next stage. The selected treatment stage (equation 1.2) represents the RFC's decision to approve or decline the submitted application. Following

[2] Although the two regimes differ by the responsibilities imposed on banks and risk taken by the RFC, examining each period individually is complicated because, for much of the sample, banks that received assistance in the first regime, also received assistance in the second regime.

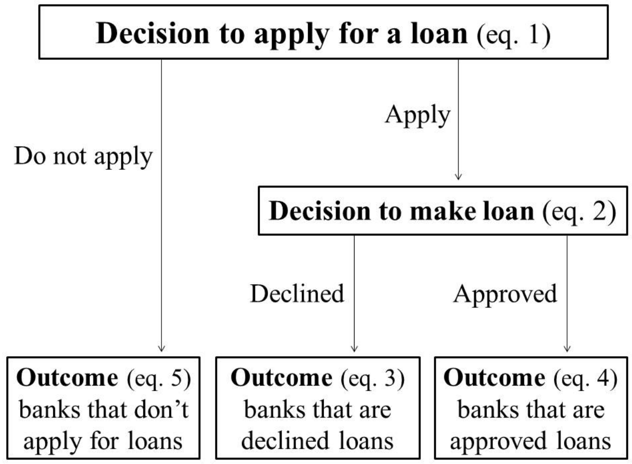

Figure 1.3: Model of the application-approval decision structure for RFC assistance.

these 2 equations are 3 potential outcomes, or treatment responses (equations 1.3–1.5). The treatment responses represent bank performance following the assistance. Banks that apply and are granted assistance comprise the selected treated sample. Banks that apply and are denied assistance comprise the selected untreated sample, and banks that do not apply comprise the non-selected sample. Estimating separate treatment response equations for each group is important because selection into these groups is non-random, thus it allows for the coefficients on the estimated parameters and the error variance to differ across equations.

The additional dimension modeled here that has lacked in the literature is the mechanism in which banks select into or opt out of treatment – a critical step in this emergency bailout process. Ignoring this stage makes disentangling non-applicant and declined banks impossible, which is problematic because these banks are fundamentally different. Modeling the full bailout process offers a more complete framework for understanding the impacts of the RFC.

1.5.1 Data

This paper employs two bank-level data sets: RFC data and bank balance sheet data. The RFC data set is constructed from the *RFC Card Index to Loans Made to Banks and Railroads, 1932-1957*, acquired from the National Archives. The cards report the name and address of the borrower, date, request and amount of the loan, whether the loan was approved or declined, and loan renewals. Further information is obtained from the *Paid Loan Files* and *Declined Loan Files*, which include the exact information regulators had on the banks from the applications and the original examiners' reports on the decisions. This data set is merged with a separate data set constructed from the *Rand McNally Bankers' Directory.* This directory describes balance sheets, charters, correspondent relationships, and other characteristics for all banks (Federal Reserve members and nonmembers) in a given state for a given year. This information identifies the non-selected, or non-applicant sample. Additional data are gathered from the 1930 U.S. census of agriculture, manufacturing and population, describing the characteristics of the county and a bank's business environment. Census covariates include the number of wholesale retailers, number of manufacturing facilities, acres of cropland, and percent of votes which were Democratic.

The data are applied to the 5 equation model as follows: the outcome variable for equation (1.1), y_{i1}, is the total amount of RFC assistance requested by each bank by December 1933.[3] This outcome is censored with point mass at zero for banks that do not apply for assistance and a continuous distribution for the different loan requests. The outcome variable for equation (1.2), y_{i2}, is the total amount of RFC assistance approved. This outcome is also censored with point mass at zero for banks that are declined assistance and a continuous distribution for the different loan approvals. The

[3]The RFC started in early 1932, so by the end of 1933 many banks submitted multiple applications. Thus, this outcome is the sum of the RFC assistance requested from each bank.

RFC's decision to lend was based on the solvency of the banks. However, after going through the examiners' reports, it is apparent that the RFC also considered banks' importance to their local market and features of that market. In addition, there is a debate in the existing literature on whether or not there was political influence in the decision-making process (Kroszner, 1994; Mason, 2003). To control for these past findings, the covariates that enter equation (1.2) include information from the RFC loan applications and political indicators.

Finally, the outcome variable for equations (1.3)–(1.5) is the total amount of "loans and discounts" (hereafter, referred to as LD) for each bank taken from its January 1935 balance sheet. The year 1935 is selected because the intervening time allowed banks to utilize their relief funds.[4] The outcome for the treatment responses is again censored with point mass at zero for banks that failed since the time of the loan application period in 1932–1933 and a continuous distribution with LD representing a bank's health and the state of the local economy. LD is chosen to measure a bank's performance following the literature on the credit crunch and its relation to economic activity (Bernanke, 1983; Calomiris and Mason, 2003a).

Descriptive Statistics

The data set includes all banks operating in 1932 in Alabama, Arkansas, Michigan, Mississippi, and Tennessee. Unlike previous studies on the RFC, this includes non-Federal Reserve member banks, as well as member institutions. Solely looking at members may misrepresent the banking population because these institutions were often healthier and had additional outlets for relief funds through the discount window. The sample consists of 1,794 banks, of which 908 banks applied for RFC assistance

[4]Later years are not considered because the FDIC was established in 1934 and its operations increased greatly over time, thus disentangling this financial restructuring becomes more difficult.

and 800 of those were approved while 108 were declined assistance. Roughly half of the banks in each state applied for assistance from the RFC. From the applicant pool, about 88% of the submitted applications were approved. Table 1.1 presents descriptive statistics on the RFC applications and approvals organized by state.

Table 1.1: RFC applications and approvals by state

	AL	**AR**	**MI**	**MS**	**TN**
No. banks applied for RFC funds	120	157	332	142	157
% of banks applied	48	56	52	60	40
No. banks approved RFC funds	102	141	288	130	139
No. banks declined RFC funds	18	16	44	12	18
% of applications approved	85	90	87	92	89
Total RFC assistance approved ($ millions)	18.7	16.7	155.6	21.5	55.4

Table 1.2 presents descriptive statistics on bank balance sheets, charters, memberships, correspondent networks, departments, and market shares. In addition, the table includes average county characteristics for each state. These 5 states are studied because many relief efforts were focused in these areas, and they provide variation across bank and county characteristics, sizes, and Federal Reserve districts. Federal Reserve district variation is necessary because RFC lending was concurrent with lending through the discount window. Federal Reserve policies differed across districts and hence impacted the rate at which banks failed (Richardson and Troost, 2009). Richardson and Troost (2009) find that the loose lending policies in the 6th district reduced bank failures, relative to the strict policies in the 8th district. The lending capabilities of the RFC were larger than that of the Federal Reserve's discount lending because the RFC could assist nonmember banks, which is 81% of this sample. County characteristics are important because previous work finds that bank distress is a continuation of agricultural distress (Calomiris and Mason, 2003b; Richardson, 2007). In addition, apparent from the applications and examiners' reports, the RFC considered some county characteristics in its approval process.

Table 1.2: Financial characteristics of the banks in each state in 1932 and county characteristics.

Variable	**Alabama**	**Arkansas**	**Michigan**	**Mississippi**	**Tennessee**
No. Banks	250	278	638	235	393
Average Age	24	22	30	25	25
Fed District(s)	6	8	7, 9	6, 8	6, 8
Financial Ratios (averages)					
Cash / Assets	0.160	0.235	0.115	0.161	0.143
Deposits / Liabilities	0.612	0.750	0.751	0.730	0.675
Financial Characteristics (averages - $1000)					
Total Assets	1027	558	2309	695	1121
Loans & Discounts (LD)	582	273	1203	356	649
Bonds & Securities	228	131	611	187	178
Cash & Exchanges	138	121	280	110	177
Paid-Up Capital	106	53	157	55	101
Deposits	716	437	1750	528	765
Surplus & Profits	92	41	126	47	80
Charters, Memberships, Depts. (counts)					
State Bank	166	222	438	208	308
National Bank	82	44	102	26	83
Amer Bk Ass'n (ABA)	160	188	370	171	185
State Bank Ass'n	215	252	542	219	374
Safe Deposit Dept.	123	121	451	94	145
Bond Dept.	23	25	113	29	35
Savings Dept.	74	126	557	152	189
Trust Dept.	63	47	118	52	133
Correspondents (averages)					
Total Correspondents	2.6	2.4	2.8	2.9	2.4
Out of State Corres.	1.5	1.4	1.5	2.5	1
Market Shares (averages)					
Liab. / County Liab.	0.27	0.26	0.13	0.33	0.24
Liab. / Town Liab.	0.68	0.75	0.17	0.74	0.69
HHI	0.66	0.60	0.29	0.70	0.54
County Characteristics (averages)					
No. Wholesale Retailers	31.3	22.4	45.3	15.4	25.3
% Voted Democratic	79.8	81.6	47.2	85.2	71.1
No. Manufact. Est.	41.4	22.5	44.5	33.8	30.4
Cropland ($\times$1000 acres)	115.6	96.6	122.9	81.9	78.1
Town Pop. ($\times$1000)	14.6	4.7	49.6	4.5	14.2

Table 1.2 also describes banks' correspondent relationships, which are considered pathways for financial contagion. Correspondent banks were designated in reserve cities of the Federal Reserve system and often provided smaller, local banks with liquidity (Richardson and Troost, 2009). Correspondent relationships between bigger and smaller banks built a structure for the Federal Reserve to influence nonmember institutions. However, they also created linkages for contagion to spread, thus controlling for them is important.

Table 1.3: Characteristics of the banks in each subgroup in 1932 and county characteristics.

Variable	**Non-Applicant**	**Declined**	**Approved**
No. Banks	886	108	800
Average Age	25	25	27
Financial Ratios (averages)			
Cash / Assets	0.17	0.11	0.13
Deposits / Liabilities	0.71	0.70	0.72
Cash / Deposits	0.29	0.17	0.19
Equity	0.21	0.18	0.20
Charters and Memberships (counts)			
State Bank	609	73	660
National Bank	198	23	116
ABA Member	487	63	518
Correspondents (averages)			
Total Correspondents	2.5	2.7	2.7
Out of State Corres.	1.4	1.6	1.5
Market Shares (averages)			
Liab. / County Liab.	0.21	0.20	0.23
Liab. / Town Liab.	0.71	0.66	0.73
County Characteristics (averages)			
No. Wholesale Retailers	27	33	28
% Vote Democratic	67	65	67
No. Manufact. Est.	34	44	37
Cropland ($\times$1000 acres)	100	116	102

Table 1.3 reports similar statistics, but separated by the three subgroups: banks that do not apply for assistance, banks that apply and are declined, and banks that apply and are approved. The table displays some fundamental differences between these

groups of banks, particularly in the financial ratios. Banks that apply for assistance hold less cash, with the declined sample holding the lowest amount, relative to banks that do not apply for assistance. Approved banks tend to be more important to their local market, coinciding with information in the examiners' files. Additionally, declined banks appear to operate in areas with more manufacturing and agricultural, relative to the other subsamples. Before the fall of 1930, the decrease in agricultural prices concentrated bank failures in farming areas, explaining the harsh economic condition of these regions (Richardson, 2007). These fundamental differences across the banks' balance sheets and locations motivate the joint model employed in this paper.

1.5.2 Results

Table 1.4 displays the results for the multivariate treatment effect model with sample selection. The following discussion reviews the results for each equation, Section 1.5.2.a describes the results for the variance-covariance matrix, and Section 1.5.2.b presents covariate and treatment effects. The results presented in Table 1.4 are from the model with the highest marginal likelihood and posterior model probability, which is referred to as the benchmark model. Marginal likelihood computations and discussions are in Section 1.6.1. The priors on $\boldsymbol{\beta}$ in the benchmark model are centered at 0 with a variance of 5, and the priors on $\boldsymbol{\Omega}$ imply that $E(\boldsymbol{\Omega}) = .4 \times I$ and $SD(\text{diag}(\boldsymbol{\Omega})) = 0.57 \times I$. Section 1.6.2 reports the sensitivity of the results to the prior specification.

Column 2 of Table 1.4 presents the results for the application step of the recapitalization process, equation (1.1). The results indicate that banks with high cash to asset ratios are less likely to apply for loans, and banks with high deposit to liability ratios

Table 1.4: Posterior means and standard deviations are based on 11,000 MCMC draws with a burn-in of 1,000. Columns 2-6 display the results for equations 1-5, respectively. In the raw data, financial characteristics and RFC requested and approved amounts are divided by 1,000,000.

Variable	1) Application	2) RFC Decision	3) Declined	4) Approved	5) Non-applicant
Intercept	-0.796 (0.172)	-0.748 (0.208)	-0.947 (0.043)	0.369 (0.218)	-0.206 (0.084)
Bank Age	0.002 (0.001)		-0.003 (0.006)	-0.001 (0.001)	-0.001 (0.001)
Financial Characteristics					
Paid-Up Capital	1.478 (0.167)	1.447 (0.191)	8.458 (2.172)	3.436 (0.190)	1.694 (0.171)
Loans & Discounts	0.265 (0.021)	0.325 (0.023)	-0.737 (0.175)	-0.086 (0.18)	0.060 (0.019)
Bonds & Securities	-0.497 (0.044)	-0.549 (0.052)			
Cash / Assets	-1.74 (0.296)	-1.608 (0.353)	-0.262 (1.045)	0.913(0.290)	-0.121 (0.112)
Deposit / Liab.	0.273 (0.073)	0.197 (0.078)			
Total Assets			0.227 (0.110)	0.184 (0.007)	0.261 (0.011)
Correspondents					
No. Corres.			-0.107 (0.107)	0.060 (0.021)	0.023 (0.017)
Corres. Out State			0.154 (0.114)	0.014 (0.021)	0.009 (0.017)
Charters, Memberships, and Depts.					
Bond Dept.	0.115 (0.035)				
Savings Dept.	-0.046 (0.026)				
Trust Dept.	0.042 (0.029)				
ABA Member		-0.028 (0.026)			
National Bank		-0.056 (0.037)	0.707 (0.375)	-0.077 (0.076)	-0.048 (0.049)
State Bank			0.435 (0.338)	-0.044 (0.070)	-0.010 (0.043)
County Characteristics					
Wholesale Retail			0.005 (0.003)	0.003 (0.000)	0.000 (0.000)
% Vote Demo.		0.000 (0.000)			
Manufact. Est.		0.000 (0.000)	-0.004 (0.003)	-0.001 (0.000)	0.000 (0.000)
Acres Cropland		-0.310 (0.170)	-0.340 (1.115)	0.572 (0.280)	0.289 (0.193)
Town Pop. 1932	-0.950 (0.184)	-1.494 (0.208)			
Town Pop. 1935			-0.511 (0.143)	-3.560 (0.720)	-5.412 (0.677)
Market Shares					
Liab./County Liab.		0.116 (0.050)			
Liab./Town Liab.	0.178 (0.074)		0.442 (0.266)	0.000 (0.056)	-0.019 (0.038)
Dummies					
Fed Dist. 6	0.138 (0.168)	0.068 (0.208)	0.241 (0.450)	-0.050 (0.193)	0.155 (0.060)
Fed Dist. 7	0.341 (0.067)	0.244 (0.206)	0.271 (0.431)	-0.282 (0.194)	0.002 (0.061)
Fed Dist. 8	0.337 (0.171)	0.329 (0.212)	0.124 (0.463)	-0.277 (0.197)	0.157 (0.001)
RFC Request (y_1)			1.155 (0.660)	-1.028 (0.217)	
RFC Approve (y_2)				1.460 (0.192)	

are more likely to apply. These results are intuitive because a high cash to asset ratio implies a bank is liquid and does not need assistance from the RFC. Whereas, a high deposit to liability ratio implies a bank may be at risk for a run on its deposits. Bank runs were the main mechanism that caused bank failure in the 1930s (Bernanke, 1983) and were a credible threat to many depository institutions. A bank's vulnerability to a run is also apparent in the result for market share. The higher the share of liability at all banks in the town held at an individual bank, the more likely the bank is to apply for assistance. With financial panics spreading through many towns, a bank that is responsible for a large share of depositors is more likely to protect itself from liquidation by applying for RFC assistance.

The results also indicate that banks with more paid-up capital are more likely to apply for relief funds. This finding accords well with the existing literature, which states that the wealth of insider shareholders increases firms' reliance on outside funds (Calomiris, 1993). Furthermore, banks in the 7th and 8th Federal Reserve districts are more likely to apply for RFC assistance, coinciding with Richardson and Troost's (2009) results. The authors show that the loose discount lending policies practiced in the 6th district were more effective at reducing bank failures, relative to the 8th district. Thus, banks in the 8th district sought additional assistance from the RFC since their regional Federal Reserve office practiced strict lending policies.

Column 3 of Table 1.4 presents the results for the approval step of the recapitalization process, equation (1.2). The results indicate that paid-up capital and LD have a positive impact on loan approval, whereas, bonds and securities have a negative impact. During this time, there was an "increased desire by banks for very liquid or rediscountable assets" (Bernanke, 1983). In order for a bank to receive liquid assets from the RFC, they must offer their illiquid assets as collateral, such as LD. The RFC accepted LD as collateral and rarely accepted bonds.[5] As a result, LD positively affects the receipt of assistance, and bonds and securities negatively affect the receipt of funds. Furthermore, Calomiris and Wilson (2004) find that a low deposit default risk is achieved by sufficient capital and limited asset risk. Thus, the RFC preferred to lend to banks with more capital and lower deposit default risk.

Also in equation (1.2), town population and acres of cropland have a negative impact on loan approval. These results coincide with the information reported on the applications to the RFC and in the examiners' reports. The examiners often commented on balance sheet characteristics, agriculture, town population, and county market size. They rarely commented on manufacturing, which explains why it is not supported in this equation. Cropland is negatively related to RFC approval because farmers were experiencing more difficulty than homeowners and nearly half were delinquent in loan payments (Bernanke, 1983). As a result, banks in these areas were unstable and the RFC may have been reluctant to offer relief. Also, a bank's county liability ratio has a positive impact on loan approval. This finding agrees with the comments in the examiners' reports and Mason's (2001) finding that "banks' importance to their local market has a significant positive effect on whether banks receive loans."

The results for the RFC's approval decision do not present any political, charter, or regional bias. The membership dummies (national bank[6], American Banking Association (ABA) member, and Federal Reserve districts) and the Democratic votes variable have 95% credibility intervals that include zero, which are computed using quantiles. These results accord well with Mason's (2003) finding that the distribution of RFC funds is not associated with political measures. It should be noted that although this paper controls for county-level political measures, it does not control

[5] Defaults on bonds from 1930 to 1939 were nearly triple the number from the previous decade (Calomiris, 1993).

[6] National banks are members of the Federal Reserve.

for politicians' personal appeals and relationships. Jesse Jones, the director of the RFC from 1933 to 1939, documents a number of personal solicitations for RFC funds from business acquaintances and members of the Executive Branch (Jones, 1951). Although he mentions he did not give in to these proposals, it would be difficult for a statistical study to control for the personal relationships of the RFC board members. However, this paper does show that there is no general geographic or membership influence.

The results for the treatment responses, equations (1.3)–(1.5), correspond to different subsets of the sample depending on which equation is observed. The outcome variable for each equation is LD in 1935. The results for equation (1.3), column 4 of Table 1.4, correspond to banks that apply for RFC assistance and are declined – the selected untreated sample. The endogenous covariate, RFC requested amount, has a positive impact on bank lending. This result is likely picking up an effect of the bank's health during the application process because in order to apply for an RFC loan, a bank had to make a collateral offering. The more collateral a bank was offering, the more RFC assistance they requested. Thus, application amounts played a positive role because healthier banks offered more collateral. Lagged LD from 1932 has a negative effect on LD in 1935. This result is intuitive given that it is in the class of declined banks. Credit extended in 1932 may have been defaulted on, reducing banks' health and making fewer funds available for LD in 1935. The national bank indicator is positive, suggesting that Federal Reserve member banks are more stable than nonmember institutions in the selected untreated sample. Federal Reserve member banks operate within a specific regulatory structure that restricts their ability to take on risk, so in a panic, they perform better. Furthermore, these banks have an additional outlet for relief funds beyond the RFC because they can receive assistance through the discount window.

The results for equation (1.4), column 5 in Table 1.4, correspond to banks that apply for RFC funds and are approved (i.e., the selected treated sample). The endogenous covariate for RFC lending y_{i2} is positive, advocating that RFC assistance increases bank lending and demonstrating the benefits of recapitalization and LOLR policies. Further examinations and interpretations of this result are in Section 1.5.2.b. Lagged LD does not appear to be supported in this subgroup. This result is especially interesting when compared to the estimates for lagged LD in the other subsamples, or treatment responses, in which cases it is statistically different from 0. This is suggestive of a "resetting" effect. The RFC accepted LD in 1932 as collateral in exchange for liquidity. Therefore, lagged LD has no impact on these banks because they were provided a fresh platform for lending.

Unlike the other treatment responses, several of the county characteristics in equation (1.4) (approved banks) affect bank lending. The number of wholesale retailers and acres of cropland in a county positively impact bank lending. Once a bank receives RFC funds, they use these funds to stimulate the local economy and promote confidence in their bank, which is clear from these results. Alternatively, the number of manufacturing facilities in a county negatively affects bank lending. In general, heavy manufacturing areas were in worse conditions during the Depression, so their recovery may have been prolonged (Rosenbloom and Sundstrom, 1999).

The indicators for national bank, state bank, and Federal Reserve districts have credibility intervals overlapping zero, displaying a sense of homogeneity across banks in different regions and of different charters, all of which receive RFC assistance. A bank's correspondent relationships in 1932 positively impacts LD in 1935. This result aligns with intuition because correspondent banks provided local banks with liquidity, often by discounting short-term commercial paper, and urged smaller institutions to extend credit. Thus, the more correspondent relationships a bank has, the more access it has to liquidity. The selected treated, or approved bank, sample is the only subgroup where the number of correspondent relationships is statistically different from 0. This finding illustrates the positive impact the RFC has on a network of banks and the banking system as a whole. With support from the RFC, correspondent relationships remain healthy and positive for individual banks, which increases lending to local communities and stabilizes the financial system.

The results for equation (1.5), the last column in Table 1.4, correspond to banks that do not apply for RFC assistance (i.e., the non-selected sample). Banks' LD in 1932 positively affects their LD in 1935. This result differs from the other two subsamples where it was negative for declined banks and not supported for approved banks. Banks that were stable enough to not apply for RFC assistance, continued their stability through 1935. Unlike the results for the other treatment response equations where the Federal Reserve district indicators are not supported, the indicators for the Federal Reserve districts 6 and 8 in the non-applicant sample have a positive effect on bank lending. For banks that do not apply for RFC funds, access to the discount window may have provided assistance, positively impacting lending in these regions.

The importance of recognizing the multi-step bailout process is emphasized in the results for the treatment response equations. Modeling each subsample separately is important because the parameter estimates are vastly different across equations. Discrepancies include: the sign and credibility interval for estimates of lagged LD are different in equations (1.3)–(1.5), county characteristics and correspondent relationships are only supported in the approved subsample, Federal Reserve membership is only impactful for the declined subsample, and Federal Reserve districts only matter for the non-applicant subsample. Banks are non-randomly placed into these subgroups by properly modeling the selection mechanisms and the application-approval decision structure, which stresses the value of the multivariate setup.

1.5.2.a Results for Omega

This section presents the estimates for the variance-covariance matrix, $\mathbf{\Omega}$. The results display the importance of joint modeling because there is a high degree of correlation between the selection equations and outcomes.

Table 1.5: Posterior means, standard deviations, and implied correlation form for $\mathbf{\Omega}$.

$\mathbf{\Omega}$	Ω_{11}	Ω_{12}	Ω_{22}	Ω_{13}	Ω_{23}	Ω_{33}	Ω_{14}	Ω_{24}	Ω_{44}	Ω_{15}	Ω_{55}
Mean	1.17	1.33	1.64	-0.83	-0.97	0.93	-1.05	-1.28	1.20	-0.04	0.11
SD	0.05	0.06	0.08	0.42	0.50	0.50	0.13	-0.16	0.25	0.02	0.01
Correlation	1	0.95	1	-0.79	-0.78	1	-0.88	-0.91	1	-0.10	1

Estimates of $\mathbf{\Omega}$ capture aspects of the model that are unobserved, e.g., loss of confidence, panic, corporate governance, and unmeasurable elements of a bank's risk, decisions, or health. Results for the third equation display a negative relationship between applying for RFC assistance and bank lending. Recall that the endogenous covariate, y_{i1} in equation (1.3), is positive for this relationship. The characteristics a researcher cannot observe or control for present a negative correlation between applications and lending.

Similarly, the results for Ω_{24} display a negative correlation between the unobserved factors of RFC loan approval and bank lending. Again, the direct effect of RFC lending on bank lending is positive; however, unobservables not captured within the model are negatively related to bank success. This result is coherent because the selected sample of banks that apply for RFC loans are banks fearing a run on their deposits or liquidation. By non-randomly selecting into the applicant class of banks, they are revealing worse health or vulnerability to the market. Ignoring correlation in the outcomes and joint modeling can lead to specification errors and biases, which are considered in detail in the next section.

1.5.2.b Covariate and Treatment Effects

Interpretation of the resulting parameter estimates presented in Table 1.4 is complicated by the censoring of the outcome variables. Analysis up to this point has been based on sign and 95% credibility intervals. Further interpretation is afforded using covariate and treatment effect calculations, which are important for understanding the model and for determining the impact of a change in one or more of the covariates. This section considers the magnitude of the parameter estimates and discusses two treatment effects.

A key estimate of interest is β_{RFC}, which is the coefficient on the endogenous covariate y_{i2} in equation (1.4). After controlling for a bank's health, business environment, and contagion channels, β_{RFC} reflects the impact of RFC assistance on bank lending. To calculate how a change in RFC lending transfers to bank lending, the covariate effect is averaged over both observations and MCMC draws from the posterior distribution. The marginal effect of the RFC is 0.571, which can be interpreted as \$10,000 of RFC assistance translates to \$5,710 of LD in 1935. This is a strong, positive result and it accords well with the loan-to-deposit ratios during the 1930s and banking panics, in general. In normal economic times, this would be considered low; however, in adverse macroeconomic conditions this ratio is standard.[7] RFC assistance is effectively pushed beyond banks, trickling into local economies through lending, thus promoting and restoring confidence in the financial system.

To stress the importance of this new methodology, which properly considers the decision structure and composition of treatment and control groups, several erroneous models are estimated. A model that ignores correlation in the outcomes and non-random selection ("Ignore Joint Modeling") is considered, as well as a model that not

[7] Banks curtail their lending during the crises – loan-to-deposit ratios fell from 0.85 in 1929 to 0.58 in 1933 (Calomiris, 1993).

only ignores joint modeling but also ignores the censoring in the outcome variables ("OLS"). The results are presented in Table 1.6. The marginal effect of the RFC in the first erroneous model is 0.41. This result is nearly 30% downward biased, severely underestimating the effectiveness of LOLR policies. This error arises as a result of selection bias. If a researcher interprets this result, it would show that liquidity injections from the RFC have a lower conversion to LD than a standard deposit. Thus, banks would be hoarding their relief funds and not extending credit to the local communities, mitigating the RFC's impact on economic activity. The second erroneous model employs ordinary least squares and displays a near 1-to-1 conversion of RFC funds to bank lending. This result demonstrates the attenuation bias that can occur if censoring in the outcome variable is ignored. These fundamental misspecifications can have detrimental effects on the results and interpretations, which highlights the importance of the multivariate treatment effect model developed in this paper.

Table 1.6: Differing results from erroneous models.

Model	$\boldsymbol{\beta_{RFC}}$	**Marg. Eff.**
Benchmark Model	1.46 (0.19)	0.57
Ignore Joint Modeling	1.07 (0.09)	0.41
OLS (eq. 4)	0.94 (0.08)	·

Up to this point, this paper has shown how RFC lending impacts bank lending. Another aspect to consider is how RFC lending affects the probability of bank failure, which can be done with treatment effects. To illustrate the main ideas of the treatment effects, suppose that one is interested in the average difference in the implied probabilities between the cases when $x_i^\dagger$ is set to the value $x_i^\ddagger$, representing a change in a covariate. Given the values of the other covariates $\mathbf{z}_i$, and those of the model parameters $\boldsymbol{\theta}$, one can obtain the probabilities $\Pr(y_i = 0|x_i^\dagger, \mathbf{z}_i, \boldsymbol{\theta})$ and $\Pr(y_i = 0|x_i^\ddagger, \mathbf{z}_i, \boldsymbol{\theta})$. Interest centers upon the predictive distribution $\{\Pr(y_i = 0|x_i^\dagger) - \Pr(y_i = 0|x_i^\ddagger)\}$, which is marginalized over $\{\mathbf{z}_i\}$ and $\boldsymbol{\theta}$ (Jeliazkov et al., 2008). Formally, the objective is to obtain a sample of draws and evaluate

$$\{\Pr(y_i = 0|x_i^{\dagger}) - \Pr(y_i = 0|x_i^{\ddagger})\} = \int \{\Pr(y_i = 0|x_i^{\dagger}, \mathbf{z}_i, \boldsymbol{\theta}) - \Pr(y_i = 0|x_i^{\ddagger}, \mathbf{z}_i, \boldsymbol{\theta})\}\pi(\mathbf{z}_i)\pi(\boldsymbol{\theta}|y)d\mathbf{z}_i d\boldsymbol{\theta},$$

where y_i is set to zero to motivate the application to bank failure. The mean of this predictive distribution gives the expected difference in the computed pointwise probabilities as $x_i^{\dagger}$ is changed to $x_i^{\ddagger}$. Computation of these probabilities is afforded by employing the CRT method, developed in Jeliazkov and Lee (2010).

This method is employed to calculate 2 treatment effects. The first case to consider is the difference in the probability of bank failure if the RFC did not offer any assistance. To see how removing the treatment from the treated banks affects bank success, two probabilities need to be considered, $\Pr(y_{i4} = 0|\mathbf{x}_{i4}, y_{i2}^{\ddagger}, \boldsymbol{\theta})$ and $\Pr(y_{i4} = 0|\mathbf{x}_{i4}, y_{i2}^{\dagger}, \boldsymbol{\theta})$, where $y_{i2}^{\ddagger}$ represents zero RFC assistance and $y_{i2}^{\dagger}$ represents the original treatment. Thus, interest lies in how a bank's probability of failure changes if the RFC never approved any loans. The mean of the predictive distribution $\{\Pr(y_{i4} = 0|y_{i2}^{\ddagger}) - \Pr(y_{i4} = 0|y_{i2}^{\dagger})\}$ is 0.126. In other words, if the RFC did not offer any assistance, the probability of bank failure for the selected treated sample (approved banks) increases by 12.6 percentage points. To shed some light on the result, this is about 101 banks or about 6% of the entire sample of 1,794 banks. Nearly 40 percent (10,000 of the approximately 25,000 banks) of all banks in existence in the United States in 1929 were suspended by 1933 and were closed during the intervening period of economic hardship (Mitchener, 2005). Generally speaking, without the RFC, thousands of additional banks could have suspended operations. Therefore, this paper finds major evidence of how RFC assistance resuscitated the banking system.

Thus far, the RFC has been shown in a positive light. However, one might wonder if the RFC could have done more to mitigate banking panics. To address this, the second treatment effect looks at how RFC assistance could have changed the outcomes for banks that were declined loans. For this scenario, the RFC approved loans are equated to the amounts requested on declined banks' applications. Two probabilities to consider are, $\Pr(y_{i3} = 0|\mathbf{x}_{i3}, y_{i2}^{\ddagger}, \boldsymbol{\theta})$ and $\Pr(y_{i3} = 0|\mathbf{x}_{i3}, y_{i2}^{\dagger}, \boldsymbol{\theta})$, where $y_{i2}^{\ddagger}$ represents declined loans (the original case) and $y_{i2}^{\dagger}$ represents the case where the RFC approved the full requested amounts. This situation displays the difference in the probability of bank failure if the RFC approved applications for the selected untreated sample (declined banks). The mean of the predictive distribution $\{\Pr(y_{i3} = 0|y_{i2}^{\ddagger}) - \Pr(y_{i3} = 0|y_{i2}^{\dagger})\}$ is 0.025. If the RFC assisted banks that were declined loans, the probability of failure for the selected untreated sample decreases by 2.5 percentage points. This is much different than the 12.6 finding for approved banks, which demonstrates the importance of the selection process. RFC assistance is almost 5 times more effective in the approved bank subsample. The banks the RFC declined to assist were helpless because full assistance from the RFC would not have had a major impact on their ability to survive and thrive in the economy.

The results of the two scenarios are clear. LOLR policies and bank recapitalization aided a bank's survival if the bank was healthy enough to receive a loan. Once non-randomly appointed to the treated group, banks that received RFC loans converted a majority of their relief funds to LD, supporting local economies. The results also indicate that the selection procedures adopted by the RFC were successful. Assistance to all struggling banks would have been wasteful because most of the untreated banks were not healthy enough to have benefitted from an influx of funds. The focal point of these results is that proper consideration of the decision structure and composition of treatment and control groups are of fundamental importance because the results vary for different subgroups of banks. The only way to identify the differences in these subgroups is by modeling the preceding selection mechanisms and the overall decision structure, which offers a more complete evaluation of the effectiveness of LOLR programs.

1.6 Additional Considerations

1.6.1 Model Comparison

An issue in the analysis of LOLR policies is model formulation since the appropriate specification is subject to uncertainty. Uncertainty due to variable selection is especially prevalent here because there are a number of financial ratios and characteristics that represent a bank's health and risk, as well as county characteristics that represent a bank's business environment, which can be included in the model. However, including all of these measures may lead to overfitting. Existing techniques in Bayesian model comparison can be employed to discover which set of covariates selected to explain the relationships in the model is best supported by the data.

Uncertainty also lies in the exclusion restrictions useful for inference. Arguing for the inclusion of a variable in the selected treatment equation and exclusion from the treatment responses is difficult and presents dubious constraints. Complications lie in disentangling aspects that influence the LOLR's decision to approve assistance and other factors that affect bank profitability. This paper not only utilizes Bayesian model comparison methods to address issues of variable selection, but also to address restriction uncertainty.

For model comparison, given the data $\mathbf{y}$, interest centers upon a collection of models $\{\mathcal{M}_1, ..., \mathcal{M}_L\}$, each characterized by a model-specific parameter vector $\boldsymbol{\theta}_l$ and sampling density $f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)$. Bayesian model selection proceeds by comparing the models through their posterior odds ratio which is written as,

$$\frac{\Pr(\mathcal{M}_i|\mathbf{y})}{\Pr(\mathcal{M}_j|\mathbf{y})} = \frac{\Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j)} \times \frac{m(\mathbf{y}|\mathcal{M}_i)}{m(\mathbf{y}|\mathcal{M}_j)}$$

for models $\mathcal{M}_i$ and $\mathcal{M}_j$. Chib (1995) recognized the *basic marginal likelihood identity (BMI)* in which the marginal likelihood for model $\mathcal{M}_l$ can be expressed as

$$m(\mathbf{y}|\mathcal{M}_l) = \frac{f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi(\boldsymbol{\theta}_l|\mathcal{M}_l)}{\pi(\boldsymbol{\theta}_l|\mathbf{y}, \mathcal{M}_l)}.$$

Calculation of the marginal likelihood is then reduced to finding an estimate of the posterior ordinate, typically taken as the posterior mean or mode. Evaluation of the likelihood is done by employing the CRT method from Jeliazkov and Lee (2010).

This paper compares 12 models that differ by variable and restriction selection. Motivation for the exclusion restrictions is based on information provided in the applications to the RFC. Information the RFC requested and commented on for each bank includes: all balance sheet information, charters, county agriculture, and county market density. However, there are several other characteristics that are not on the applications or mentioned in the examiners' reports, implying the RFC did not have this information on hand when making their decisions. This includes bank age, town information, departments, and correspondent networks. These natural exclusions motivate the restrictions considered in the competing model specifications.

Results for the model comparison are displayed in Table 1.7. The first column of the table describes deviations from the benchmark model, e.g., Model 1 includes additional variables for asset ratios, HHI, and state effects in each equation that do not appear in the benchmark model. The second column lists the unique variables for each covariate vector in the selection equation ($\mathbf{x}_1$), selected treatment ($\mathbf{x}_2$), and treatment responses ($\mathbf{x}_3 - \mathbf{x}_5$), respectively, that are excluded from 1 or more stages.

The third column presents the log-marginal likelihood estimate and the forth column reports the numerical standard error.

Results for the variable selection show that state effects are unnecessary. The marginal likelihood decreases by 25 on the log scale when state indicators are included. In general, the data support more parsimonious models without including numerous financial characteristics. When multiple financial measures representing similar aspects of a bank's health are included, the marginal likelihood decreases, thus overfitting the model. In addition, the data support specifications that have both financial ratios and levels, which grasp both bank profitability and size. Models that include mostly ratios are not only less supported by the data, but also have poor mixing properties of the Markov chain.[8]

Competing hypotheses about the direct effect of RFC assistance are tested. Model 7 is nearly identical to the benchmark model, however, without the endogenous covariates y_{i1} and y_{i2} (RFC requested and approved amounts) entering equations (1.3) and (1.4). This model hypothesizes that there is no direct effect of RFC assistance, just an indirect effect through the correlation in the unobservables. When this model is compared to the benchmark model, the marginal likelihood decreases by 454 points on the log scale. The data overwhelmingly support the benchmark model, where the actions of the RFC have a direct impact on bank success. Furthermore, the benchmark model supports the notion that the size of the loan matters. If the data allow these endogenous covariates to enter the treatment response equations, as a result of not using binary selection, they can have immense explanatory power. Model 12 is considered to detect potential nonlinearities in the endogenous regressors. Table 1.7 displays a lower marginal likelihood. Additionally, the point estimates and credibility intervals did not support the polynomial functions of the endogenous regressors.

[8]Poor mixing properties are also found when log-transformations are considered.

Table 1.7: Results of the model comparison. The first column describes deviations from the benchmark model, the second column identifies unique variables in each covariate vector for the selection equation ($\mathbf{x}_1$), selected treatment ($\mathbf{x}_2$), and treatment responses ($\mathbf{x}_3 - \mathbf{x}_5$), respectively, that are excluded from 1 or more stages. The third column shows the log-marginal likelihood estimate and the forth column reports the numerical standard error.

Model	Unique Vars.	Marg-lik.	NSE
1) State effects, asset ratios, HHI	Safe deposit dept. Manufacturing, cropland Town market shares	-8,100.8	0.85
2) Town shares, equity ratios	Bank age % vote Democratic State bank association member	-8,063.9	0.50
3) Total assets, loan/dep ratios	All departments HHI Correspondents	-8,055.2	0.56
4) Correspondents, cropland	Bank age % vote Democratic Departments	-8,031.4	0.73
5) Additional financial vars.	Bank age, departments Cropland, county shares, % Democratic Wholesale, town shares, corresp.	-7,982.7	0.47
6) Benchmark	**Bank age, departments** **Cropland, county shares, % Demo.** **Wholesale, town shares, corresp., 1935 pop.**	**-7,978.6**	**0.43**
7) No endogenous covarites	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp., 1935 pop.	-8,433.2	0.46
8) No y_{i1} endogeneity	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp., 1935 pop.	-8,013.4	0.42
9) All financial ratios (no levels)	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp., 1935 pop.	-8,504.3	0.91
10) State effects	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp., 1935 pop.	-8,003.0	0.71
11) Non-app. restrictions	Bonds & securities Surplus & profits Equity	-8,227.0	0.68
12) Quadratic endogenous vars.	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp., 1935 pop.	-7,992.5	0.48

Therefore, there is no evidence of nonlinearity in this case, however, it is an important issue to explore in empirical applications.

Results of the model comparison with different restrictions strongly favor the natural exclusions from the applications. Unique variables excluded from the selected treatment stage include bank age, departments, town market shares, and correspondent relationships – all of which are not included in a bank's application to the RFC. Although the RFC did not request this information, these features do explain a bank's decision to apply for assistance and its performance. Consider Model 11 where the exclusions are not based on information in the applications, but instead, other financial characteristics. The marginal likelihood falls by 248.4 points on the log scale, giving it a posterior model probability of approximately 0. Understandably, exclusions are a tough argument, however, with the use of Bayesian model comparison techniques, the data show support for the application-based exclusions over others.

Across the models in Table 1.7, the impact of the RFC remains positive and statistically different from 0.[9] The results are robust to the competing specifications and hypotheses about the RFC. Most importantly, the results highlighted in this paper arise from the model that is best supported by the data and that has the highest posterior model probability.

1.6.2 Sensitivity Analysis

The priors for the benchmark model appear at the beginning of Section 1.5.2. Prior selection generally involves some degree of uncertainty and this section evaluates how

[9]The marginal effect of the RFC on bank lending varies from .38–.74, with the lowest and highest values coming from models 8 and 5, respectively. When specification 8 is considered in the erroneous model "Ignore Joint Modeling" from Table 1.6, the marginal effect drops to 0.02, which again stresses the importance of modeling the non-random selection mechanisms and allowing for correlation in the outcomes and endogeneity.

sensitive the results are to the assumptions about the prior distribution.

The key coefficient of interest, β_{RFC}, is the estimate on the endogenous variable y_{i2} in equation (1.4). The coefficient reported in Table 1.4 shows $\beta_{RFC} = 1.460$, which implies that RFC assistance has a positive impact on bank lending. To check the sensitivity of this result to the prior specification, Table 1.8 reports the coefficient β_{RFC} for different hyperparameters.

Table 1.8: β_{RFC} as a function of the hyperparameters. The priors for $\boldsymbol{\beta}$ in the benchmark model are centered at zero with a variance of 5.

	SD(β_{RFC})		
Mean(β_{RFC})	1.5	4.4	14.14
-1	1.418	1.481	1.491
0	1.436	1.483	1.491
1	1.453	1.486	1.491

The results indicate nearly no sensitivity around the benchmark result of 1.460. This finding holds true for all of the parameter estimates. Skeptics of bank recapitalization who would place strong negative priors on their economic benefit would be overridden by the data. The data speak loudly for the benchmark results and the overall findings. Note that the model rankings in Section 1.6.1 are also not sensitive to the different prior specifications.

1.7 Concluding Remarks

This paper presents a methodological framework for multivariate treatment effect models in the presence of sample selection and discrete data, and provides a general modeling framework for application-approval decision structures. The model is not limited to banking contexts and is applicable to a multitude of problems prevalent in economics, including modeling the effectiveness of job training and housing programs, health treatments, education policies, credit approval decisions, and many others. The algorithm designed in this paper is computationally efficient, modular, and easily extendable. Furthermore, the paper offers straightforward techniques for treatment effects and model comparison.

The methods established in this paper are applied to the analysis of LOLR regulation. The results indicate that bank recapitalization is effective at decreasing the probability of bank failure, stimulating bank lending, and resuscitating a struggling economy. Rescue programs not only keep individual banks healthy, but they also promote positive relationships with correspondent networks and counties in which the banks operate. The importance of the multivariate treatment effect model and accommodating the RFC's selection procedure is highlighted in the findings because the results vary for the different subgroups of banks and there are strong correlations between the equations. Although RFC assistance is beneficial for the treated group, it would have been minimally helpful for banks that are declined assistance because their economic condition is too severe.

Studying the RFC is an important and relevant topic because it was used as a reference for the current program, the Troubled Asset Relief Program (TARP), employed during the recent crisis. Further research on LOLR policies should focus on the multi-step decision mechanisms that place banks into different policy treatments to answer questions, such as whether and to what extent these programs stabilize the economy or simply privatize the gains and nationalize the losses. Overall, this model offers practical estimation tools to unveil new answers to questions involving sample selection and treatment response data as in the application and loan approval settings.

Chapter 2

Determining the Proper Specification for Endogenous Covariates in Discrete Data Settings

2.1 Introduction

In empirical applications, endogenous regressors are generally the key variables of interest. Treatment models, triangular systems with recursive endogeneity, and sequential decision-making all feature endogenous covariates that often represent the main components of the study. In continuous data settings, modeling endogeneity is simple and interpretation is straightforward. In discrete data settings, modeling endogeneity is complicated because it can take several forms based on latent or observed data. This is not a limitation of the system. Instead, this feature increases the flexibility of such models because unobservables can be captured as explanatory variables separately from variables that are observable to the econometrician. This is only beneficial if a formal model comparison can be performed to decipher which depicted relationship is best supported by the data and to resolve competing hypotheses about the type of endogeneity. Investigating both approaches strengthens the eventual results by increasing a researcher's understanding of the relationships and the dependence structure being modeled. Model testing is easily afforded by existing Bayesian model comparison techniques.

Initial latent data modeling innovations occurred in psychometrics, where the traditional usage of latent variables focused on measurement error and hypothetical constructs (Muthen, 2002). In econometrics, latent data analysis advances discrete choice methods in which choice outcomes are linked to latent utility. For a review, see Jeliazkov and Rahman (2012). Bayesian econometrics further benefited by the association between data augmentation and latent variables (Tanner and Wong, 1987). Although latent modeling approaches have captured a variety of statistical concepts, including random coefficients, missing data, discrete choice, and finite mixture modeling, latent variables are rarely employed as regressors since an investigator has not or can not measure or observe them. Macroeconomics has moved in this direction with factor models; however, latent covariates remain unexplored in applied microeconomic research. Furthermore, across all fields, little attention has been applied to formally compare such models. In a recent marketing paper, Mintz et al. (2013) look at both specifications and find that a latent measure of information processing pattern better explains an individual's propensity to buy. Overlooking the consideration of both approaches can lead to specification errors and misrepresent the relationships being examined.

This paper employs Bayesian model selection methods for comparing latent and observed endogeneity models in two empirical applications. Each application features competing hypotheses discussed in the literature and a formal motivation for using observed or latent endogeneity. The first application examines banking contagion and the relative influence or spread of contagion from both regional and network linkages. The second application considers the impact of education on adult socioeconomic status. These applications highlight a key aspect of this research topic. While an applied researcher may have *a priori* expectations of the "correct" model, in most cases and especially in these examples, arguments for both approaches are easily formed, making it difficult for a researcher to completely rule out a specification without performing model comparison.

The rest of this paper is organized as follows: Section 2.2.1 reviews each specification in a simple bivariate system of equations and Section 2.2.2 discusses existing techniques for model selection. Section 2.3 considers the application to financial contagion and Section 2.4 considers the application to education. Finally, Section 2.5 offers concluding remarks.

2.2 Methodology

2.2.1 Model

To exemplify the approaches discussed in this paper, consider a bivariate model with recursive endogeneity, where latent data are referred to as $\mathbf{y}_i^*$ and observed data are referred to as $\mathbf{y}_i$. The two different modeling techniques, latent and observed, are shown as:

Observed Endogeneity

$$y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1}, \tag{2.1}$$

$$y_{i2}^* = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + y_{i1}\gamma_2 + \varepsilon_{i2},$$

Latent Endogeneity

$$y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1}, \tag{2.2}$$

$$y_{i2}^* = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + y_{i1}^*\gamma_2 + \varepsilon_{i2}.$$

The latent data are related to the observed outcomes by a link function depending on the values y_{ik} can take, for equations $k = 1, 2$. The binary setting occurs when $y_{ik} = 1\{y_{ik}^* > 0\}$, and the censored setting occurs when $y_{ik} = y_{ik}^* \cdot 1\{y_{ik}^* > 0\}$. The link function for ordered data is $y_{ik} = \sum_{j=1}^{J} 1\{y_{ik}^* > \alpha_{k,j-1}\}$ for J ordered alternatives, where α_{kj} is a cut-point between the categories. In this context, the observed endogeneity system differs from (2.1) and instead is:

$$y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1},$$

$$y_{i2}^* = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + 1\{y_{i1} = 2\}\gamma_{22} + 1\{y_{i1} = 3\}\gamma_{23} + \ldots + 1\{y_{i1} = J\}\gamma_{2J} + \varepsilon_{i2}, \tag{2.3}$$

where there is a set of endogenous indicator variables for $J-1$ categories, as opposed to a single endogenous regressor as in (2.1). This case is explored in the second application to education in Section 2.4. For simplicity, assume $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})' \sim N_2(0, \boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$ in (2.1), (2.2), and (2.3). Models with endogeneity have been difficult to estimate when the response variables of interest are not continuous because standard two-stage estimators are inapplicable in this context. Therefore, estimation in this paper relies on Markov chain Monte Carlo (MCMC) methods, which are discussed in detail in each application.

The latent data have the customary random utility interpretation underlying the theory on discrete choice analysis in econometrics. Therefore, even though the observed data can only take certain values, the latent variables that determine those outcomes are unrestricted. The fact that the latent utilities can be changing without necessarily inducing a corresponding change in the observed variable is a key distinction between these models (Mintz et al., 2013). Furthermore, the observed and latent specifications pose different relationships between the variables of interest because latent data measure intentions and observed data measure actual actions or outcomes. In (2.2), latent endogeneity says that intentions about y_{i1} determine intentions about y_{i2}. In (2.1), observed endogeneity says that actions about y_{i1} determine intentions about y_{i2} (Maddala, 1983). Despite the clear interpretation of each model, in most cases, it is difficult for a researcher to decipher which specification is correct. Generally, convincing hypotheses or arguments can be made in support of either modeling approach, hence motivating the need for model comparison.

It is important to note that these considerations extend to larger systems of equations, models for sample selection, potential outcomes, simultaneous equations, and more. For a review of some of these models, see van Hasselt (2014) and Li and Tobias (2014). Any multivariate discrete outcome model should not overlook this problem. The bivariate system is considered here to stress the importance of the issue, highlight the ease of considering both models, and offer a more complete understanding of the relationships in each application.

2.2.2 Model Comparison

Model comparison techniques are often employed to deal with issues of model uncertainty and variable selection. This paper utilizes these same approaches to determine the nature of endogeneity. The methods used in this paper are from Chib (1995) and Chib and Jeliazkov (2001), which are computationally convenient and do not require much additional coding. The applications in this paper span a number of discrete choice models, including ordered probit, Tobit, and binary probit. Therefore, the model comparison methods discussed here are general across these classes of models.

Given the data $\mathbf{y}$, interest centers upon the models $\{\mathcal{M}_l, \mathcal{M}_o\}$ where $\mathcal{M}_l$ represents the latent endogeneity model and $\mathcal{M}_o$ represents the observed endogeneity model. Each model is characterized by a sampling density $\{f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l), f(\mathbf{y}|\mathcal{M}_o, \boldsymbol{\theta}_o)\}$ where $\{\boldsymbol{\theta}_l, \boldsymbol{\theta}_o\}$ are model-specific parameter vectors. Bayesian model selection proceeds by comparing the models through their posterior odds ratio

$$\frac{\Pr(\mathcal{M}_l|\mathbf{y})}{\Pr(\mathcal{M}_o|\mathbf{y})} = \frac{\Pr(\mathcal{M}_l)}{\Pr(\mathcal{M}_o)} \times \frac{m(\mathbf{y}|\mathcal{M}_l)}{m(\mathbf{y}|\mathcal{M}_o)}. \tag{2.4}$$

Chib (1995) recognized the *basic marginal likelihood identity* in which the marginal likelihood for model $\mathcal{M}_l$ can be expressed as

$$m(\mathbf{y}|\mathcal{M}_l) = \frac{f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi(\boldsymbol{\theta}_l|\mathcal{M}_l)}{\pi(\boldsymbol{\theta}_l|\mathbf{y}, \mathcal{M}_l)}. \tag{2.5}$$

Calculation of the marginal likelihood is then reduced to finding an estimate of the posterior ordinate $\pi(\boldsymbol{\theta}_l^*|\mathbf{y}, \mathcal{M}_l)$ at a single point $\boldsymbol{\theta}_l^*$, which is often taken as the posterior mean or mode. Since the topics in this paper involve multivariate discrete data, sampling densities are often analytically intractable. A straightforward approach for evaluating the likelihood function employed in this paper is the Chib-Ritter-Tanner

(CRT) method, which was developed in Jeliazkov and Lee (2010).

Decomposition of the posterior ordinate varies across the examples in this paper, with the most difficult being the multivariate ordered probit model used in the education application. This case is a bit more complex due to the additional cut-point parameters $\boldsymbol{\delta}$.[1] This section outlines the decomposition used for the ordered probit model, which follows from Jeliazkov et al. (2008). It should be noted that the decomposition for the binary probit and Tobit models are simplified versions of the ordered probit without computations for the $\boldsymbol{\delta}$ parameter vector. Let $\boldsymbol{\lambda} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ be a parameter vector for all endogenous and exogenous covariates. Estimation of the posterior ordinate can be facilitated using the decomposition

$$\pi(\boldsymbol{\lambda}^*, \boldsymbol{\Omega}^*, \boldsymbol{\delta}^* | \mathbf{y}) = \pi(\boldsymbol{\lambda}^* | \mathbf{y}) \pi(\boldsymbol{\Omega}^* | \mathbf{y}, \boldsymbol{\lambda}^*) \pi(\boldsymbol{\delta}^* | \mathbf{y}, \boldsymbol{\Omega}^*, \boldsymbol{\lambda}^*).$$

Estimation of $\pi(\boldsymbol{\lambda}^* | \mathbf{y})$ is done by averaging the full conditional density with draws $\left\{ \mathbf{y}^{*(g)}, \boldsymbol{\Omega}^{(g)} \right\} \sim \pi(\mathbf{y}^*, \boldsymbol{\Omega} | \mathbf{y})$ from the main MCMC run for $g = 1, \ldots, G$,

$$\pi(\boldsymbol{\lambda}^* | \mathbf{y}) \approx G^{-1} \sum_{g=1}^{G} \pi(\boldsymbol{\lambda}^* | \mathbf{y}, \mathbf{y}^{*(g)}, \boldsymbol{\Omega}^{(g)}).$$

The next ordinate, $\pi(\boldsymbol{\Omega}^* | \mathbf{y}, \boldsymbol{\lambda}^*)$, can be estimated using a reduced run to obtain

$$\pi(\boldsymbol{\Omega}^* | \mathbf{y}, \boldsymbol{\lambda}^*) \approx G^{-1} \sum_{g=1}^{G} \pi(\boldsymbol{\Omega}^* | \mathbf{y}, \boldsymbol{\lambda}^*, \mathbf{y}^{*(g)}).$$

The last ordinate, $\pi(\boldsymbol{\delta}^* | \mathbf{y}, \boldsymbol{\Omega}^*, \boldsymbol{\lambda}^*)$, which is unique to the ordered probit setting, is estimated using the methods in Chib and Jeliazkov (2001).

The benefits of Bayesian model comparison go beyond dealing with issues of vari-

[1] The cut-points α_{kj}, which are defined in the link function for ordinal data, are transformed such that $\delta_{kj} = \ln(\alpha_{kj} - \alpha_{k,j-1})$. A discussion of this transformation is offered in Section 2.4.1.

able selection and model uncertainty. In this context, additional benefits include understanding the type of endogeneity and the dependence structure between a set of outcome variables. This allows a researcher to distinguish between several competing specifications and further investigate the relationships of interest. Emphasis is placed on Bayesian techniques because considering both specifications is a straightforward extension of the methodology. Gibbs sampling methods employed in discrete data models already generate the latent $\mathbf{y}^*$s in the data augmentation part of a sampler. Employing these draws as data involves minimal additional coding as discussed in the next section.

2.3 Bank Contagion

The first application addresses financial contagion in two ways. First, this application examines both latent and observed measures of a bank failure and how these determine nearby bank performance. The impact of a regional bank failure on bank health remains unclear in the existing literature. Calomiris and Mason (2003b) find that a nearby failure decreases the probability of survival for the remaining banks. However, there is also research on efficient bank runs (Freixas and Rochet, 2008), which finds positive effects stemming from nearby bank failures due to market competition. If an inefficient bank fails, its customers can go to the remaining banks in the market for deposits and lending, thereby benefiting the existing depository institutions. Second, the application evaluates different linkages for the spread of contagion. Channels for contagion have been found in both regional and correspondent[2] networks (Aharony and Swary, 1996; Kaufman, 1994; Richardson, 2007; Richardson and Troost, 2009). While both linkages may be present, financial regulators need to determine which

[2]Correspondents were banks with ongoing relationships facilitated by deposits of funds (Richardson, 2007). These networks linked banks across the country and indicated the extent to which a bank was important within the national network of banking (Calomiris et al., 2013).

channel is stronger in order to implement policies for restricting the spread of contagion and preventing bank runs. This paper addresses these issues and examines the consequences of bank failures during the 1930s by looking at town-wide failure rates and changes to correspondent networks.

The models for banks $i = 1, \ldots, n$, are those in (2.1) and (2.2) where $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})' \sim N_2(0, \mathbf{\Omega})$ and $\mathbf{\Omega} = \begin{pmatrix} 1 & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$. The models are characterized with two dependent variables in which $\mathbf{y}_i^* \equiv (y_{i1}^*, y_{i2}^*)'$ are the continuous latent data and $\mathbf{y}_i \equiv (y_{i1}, y_{i2})'$ are the corresponding discrete observed data. For the first outcome, the latent variables relate to the observed binary outcomes by $y_{i1} = 1\{y_{i1}^* > 0\}$ and

$$y_{i1} = \begin{cases} 0 & \text{No bank failure occurred nearby between 1929 - 1932} \\ 1 & \text{Bank failure occurred nearby between 1929 - 1932.} \end{cases}$$

The first outcome y_{i1} indicates whether or not a bank failure occurred between 1929-1932 in the town the subject bank does business. $\mathbf{\Omega}$ incorporates the usual unit variance restriction in probit models, which is a normalization for identification. The second outcome (y_{i2}) measures a bank's performance in 1933, where there is point mass at 0 for banks that were suspended since 1932 and a continuous distribution with "loans and discounts" (hereafter referred to as LD) representing bank health. LD is chosen to measure a bank's performance following the literature on the credit crunch and its relation to economic activity (Bernanke, 1983; Calomiris and Mason, 2003a). The latent variables $\{y_{i2}^*\}$ relate to the observed censored outcomes by $y_{i2} = y_{i2}^* \cdot 1\{y_{i2}^* > 0\}$.

The endogenous covariate, y_{i1} in (2.1) or y_{i1}^* in (2.2), displays the impact of a nearby or regional bank failure on a bank's health and lending. Both the latent and observed specifications are easily motivated by hypotheses discussed in the existing literature.

The latent counterpart of a regional bank failure can reveal unobserved factors that affect bank distress and profitability, such as corporate governance, risk behavior, and loss of confidence. Although an econometrician can observe whether a bank fails, additional information bankers have on the performance of their bank and business environment remains unobservable. In addition, existing evidence suggests that bank runs, which were the main mechanisms that caused bank failure (Bernanke, 1983), were often facilitated through "word-of-mouth" or information-based contagion (Park, 1991). For instance, if an individual's neighbor speculated about a pending bank failure, the individual is likely to withdraw deposits from his bank to avoid undue losses. Depositors lack financial information, resulting in withdrawal decisions based on the condition of the banking system as a whole (Park, 1991). Although researchers cannot measure the speculative nature of banking panics, the latent specification can act as a proxy for these factors. These arguments formally motivate the latent endogeneity model.

On the other hand, the literature notes that bank failures trigger panic (Chen, 1999). Despite speculation about bank health, depositors do not react until an indicator for failure is triggered. The literature also notes that publicizing the names of failed banks worsened remaining bank health. Additionally, the publication of the names of banks receiving financial assistance mitigated lender of last resort relief efforts (Butkiewicz, 1995). These observed outcomes, or triggers, support the observed endogeneity model. As mentioned previously, although a researcher may have an *a priori* expectation of the correct specification, it is hard to completely rule out the opposing approach. Therefore, a formal model comparison is necessary to better understand how a nearby failure affects bank performance and to ensure the employed specification accurately captures the interactions and decisions of banks during financial crises.

The data collected for this application are from the *Rand McNally Bankers' Directory.*

This directory details balance sheets, correspondent relationships, and characteristics for all banks in a given state. Additional data are gathered from the 1930 U.S. census of agriculture, manufacturing and population, which describe the characteristics of the county and banks' business environment. The sample includes all banks operating in 1932 in Alabama, Arkansas, Michigan, Mississippi, and Tennessee for a total of 1,794 banks. These 5 states are considered because they provide variation across bank characteristics, size, Federal Reserve districts, and county characteristics. Table 2.1 presents descriptive statistics on the banks and average county characteristics for each state. For further information on the data set, see Vossmeyer (2014).

Table 2.1: Financial characteristics of the banks in each state in 1932 and county characteristics.

Variable	**Alabama**	**Arkansas**	**Michigan**	**Mississippi**	**Tennessee**
Number of Banks	250	278	638	235	393
Average Age	24	22	30	25	25
Federal Reserve District	6	8	7, 9	6, 8	6, 8
Financial Characteristics (avg.-\$1000)					
Deposits	716	437	1750	528	765
Deposits/Liab. (*ratio*)	0.612	0.750	0.751	0.730	0.675
Loans & Discounts (LD)	582	273	1203	356	649
Charters and Memberships (counts)					
State Bank	166	222	438	208	308
National Bank	82	44	102	26	83
Amer Bk Ass'n (ABA)	160	188	370	171	185
Correspondents (averages)					
Total Correspondents	2.6	2.4	2.8	2.9	2.4
Out of State Corres.	1.5	1.4	1.5	2.5	1
$\triangle$ Correspondents	0.19	0.31	0.22	0.28	0.40
Market Shares (averages)					
Liab./County Liab.	0.27	0.26	0.13	0.33	0.24
Liab./Town Liab.	0.68	0.75	0.17	0.74	0.69
Herfindahl index (HHI)	0.66	0.60	0.29	0.70	0.54
County Characteristics (averages)					
No. Wholesale Retailers	31.3	22.4	45.3	15.4	25.3
Cropland ($\times$1000 acres)	115.6	96.6	122.9	81.9	78.1
Town Pop. ($\times$1000)	14.6	4.7	49.6	4.5	14.2

A key covariate of interest listed in Table 2.1 is $\triangle$ *Correspondents* – an indicator variable that takes the value 1 if a correspondent was removed from a bank's network and 0 otherwise. Recall that there are two linkages for contagion, regional which the nearby bank failure variable (y_{i1}) captures, and correspondent networks which this variable captures. Correspondent banks are usually designated in reserve cities of the Federal Reserve system and often provided smaller, local banks with liquidity (Richardson and Troost, 2009). Correspondent relationships between bigger and smaller banks built a structure for the Federal Reserve to influence nonmember institutions. However, this structure created pathways for contagion to spread. Therefore, controlling for it in the model is important and of interest to policy-makers in order to mitigate contagion through its many channels.

2.3.1 Estimation

The model is completed by specifying the prior distributions. For $\boldsymbol{\lambda} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, $\pi(\boldsymbol{\lambda}) = \mathcal{N}(\boldsymbol{\lambda}|\mathbf{d}_0, \mathbf{D}_0)$ and $\pi(\boldsymbol{\Omega}) \propto \mathcal{IW}(\nu_0, \mathbf{R}_0) 1\{\omega_{11} = 1\}$, where the prior on $\boldsymbol{\Omega}$ (inverse Wishart) is on the derived quantities that appear in Algorithm 2.1. The hyperparameters for the priors are selected using a training sample of 100 banks. A thorough sensitivity analysis is provided in Section 2.3.2. Algorithm 2.1 presents the Gibbs sampling and data augmentation methods to simulate the posterior distribution for the observed endogeneity specification.

Algorithm 2.1. *MCMC Estimation Algorithm - Observed Specification*

1. *Sample* $[\boldsymbol{\lambda}|\boldsymbol{y}^*, \boldsymbol{\Omega}] \sim N\left(\hat{\boldsymbol{d}}, \hat{\boldsymbol{D}}\right)$, *where* $\hat{\boldsymbol{d}}$ *and* $\hat{\boldsymbol{D}}$ *are given by* $\hat{\boldsymbol{d}} = \hat{\boldsymbol{D}}\left(\boldsymbol{D}_0^{-1}\boldsymbol{d}_0 + \sum_{i=1}^n \boldsymbol{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{y}_i^*\right)$ *and* $\hat{\boldsymbol{D}} = \left(\boldsymbol{D}_0^{-1} + \sum_{i=1}^n \boldsymbol{W}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{W}_i\right)^{-1}$ *where*
$$\boldsymbol{W}_i = \begin{pmatrix} \mathbf{x}_{i1}' & 0 & 0 \\ 0 & \mathbf{x}_{i2}' & y_{i1} \end{pmatrix}.$$

2. *Sample* $\boldsymbol{\Omega}$ *in a one-block, two-step procedure by drawing* $\omega_{22\cdot1} \equiv \omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12}$ *and* ω_{21}, *then reconstructing* $\boldsymbol{\Omega}$ *from these quantities,*

(a) $\omega_{22\cdot1} \sim \mathcal{IW}(\nu_0 + n, Q_{22})$

(b) $\omega_{21} \sim \mathcal{N}(Q_{11}^{-1}Q_{12}, \omega_{22\cdot1}Q_{11}^{-1})$, *where*

$$\mathbf{Q} = \mathbf{R}_0 + \sum_{i=1}^{n}(\mathbf{y}_i^* - \mathbf{W}_i\boldsymbol{\lambda})(\mathbf{y}_i^* - \mathbf{W}_i\boldsymbol{\lambda})',$$

and $\boldsymbol{Q}$ *is partitioned conformably with* $\boldsymbol{\Omega}$, *i.e.,*

$$\mathbf{Q} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}.$$

3. *For* $i = 1, \ldots, n$, *sample* $y_{i1}^*|y_{i2}^*, y_{i1}, \boldsymbol{\lambda}, \boldsymbol{\Omega} \sim \mathcal{TN}_{\mathcal{A}_i}(\mu_{1|2}, V_{1|2})$, *where* $\mathcal{TN}$ *is the truncated normal distribution and* $\mu_{1|2}$ *and* $V_{1|2}$ *are the usual conditional mean and conditional variance, respectively. If* $y_{i1} = 0$, $\mathcal{A}_i$ *is* $(-\infty, 0)$, *and if* $y_{i1} = 1$, $\mathcal{A}_i$ *is* $(0, \infty)$.

4. *For* $i : y_{i2} = 0$, *sample* $y_{i2}^*|y_{i1}^*, y_{i2}, \boldsymbol{\lambda}, \boldsymbol{\Omega} \sim \mathcal{TN}_{\mathcal{A}_i}(\mu_{2|1}, V_{2|1})$, *where the region* $\mathcal{A}_i$ *is* $(-\infty, 0)$ *implied by the censoring of* y_{i2}.

This sampler can be easily adapted to handle the latent specification. The most convenient approach is to move to the reduced-form where the system in (2.2) can be re-written as,

$$\begin{aligned}
\begin{pmatrix} 1 & 0 \\ -\gamma_2 & 1 \end{pmatrix} \begin{pmatrix} y_{i1}^* \\ y_{i2}^* \end{pmatrix} &= \begin{pmatrix} \mathbf{x}_{i1}' & 0 \\ 0 & \mathbf{x}_{i2}' \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \\
\mathbf{A}\mathbf{y}_i^* &= \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \\
\Leftrightarrow \mathbf{y}_i^* &= \mathbf{A}^{-1}X_i\boldsymbol{\beta} + \mathbf{A}_i^{-1}\varepsilon_i \\
\Leftrightarrow \mathbf{y}_i^* &= \boldsymbol{\mu}_i + \nu_i \\
\nu_i &\sim \mathcal{N}(0, \mathbf{A}^{-1}\boldsymbol{\Omega}\mathbf{A}^{-1\prime}).
\end{aligned} \tag{2.6}$$

Simply change the data augmentation steps of the sampler (steps 3-4) to use the conditional mean and conditional variance from the reduced form, avoiding any additional computational burden brought on by considering both specifications. This is a straightforward adjustment because the $\mathbf{y}^*$s are already being generated. In the latent specification, they are generated and passed through the sampler as data, so $\boldsymbol{W}_i = \begin{pmatrix} \mathbf{x}'_{i1} & 0 & 0 \\ 0 & \mathbf{x}'_{i2} & y^*_{i1} \end{pmatrix}$ in steps 1 and 2 of Algorithm 2.1.

2.3.2 Results

The results for the application are based on 11,000 MCMC draws with a burn-in of 1,000. The inefficiency factors for the parameters remain low with slightly higher values occurring for the parameters on the endogenous covariates and variables common to both equations. The point estimates for the exogenous covariates are similar across both specifications except in cases where the variable has a 95% credibility interval that includes 0. The following discussion covers the basic results for each equation, followed by the model comparison and sensitivity analysis.

Table 2.2 presents the posterior means and standard deviations for both specifications. The results indicate that a nearby bank failure (y_{i1}) positively affects lending for existing banks, which disagrees with some of the literature. However, unlike other studies, this paper is looking at a longer window of impact. Previous papers examine the immediate impact of a nearby failure, which is distress. Whereas, this paper finds that the long-run impact is positive, corroborating the research on efficient bank runs (Freixas and Rochet, 2008). This result supports the market competition hypothesis where failing banks leave additional depositors in the market as customers for the remaining banks. As a result, the failure of a nearby bank strengthens the balance sheets for surviving depository institutions. The results also illustrate that the

Table 2.2: Banking application – posterior means and standard deviations for the bivariate system of equations.

	Observed Endogeneity		Latent Endogeneity	
	Nearby Fail	**Lending**	**Nearby Fail**	**Lending**
Intercept	-5.023 (0.260)	-0.480 (0.137)	-6.514 (0.349)	-0.122 (0.138)
Financial Characteristics				
Bank Age		0.009 (0.002)		0.013 (0.002)
Lagged LD		0.134 (0.016)		0.107 (0.017)
National Bank		0.367 (0.093)		0.488 (0.102)
Deposits/Liabilities		0.115 (0.094)		0.143 (0.111)
△ Correspondents		-0.176 (0.081)		-0.138 (0.087)
Total Correspondents	0.041 (0.029)	-0.014 (0.029)	-0.021 (0.039)	-0.032 (0.031)
County Characteristics				
Town Population	1.262 (1.008)	0.193 (0.327)	-0.658 (1.046)	-2.627 (0.540)
Wholesale Retailers	0.001 (0.001)		0.001 (0.001)	
Acres of Cropland	0.218 (0.481)		-0.025 (0.570)	
Liab./Town Liab.	2.273 (0.180)	-0.025 (0.119)	3.106 (0.238)	-0.152 (0.119)
Herfindahl index (HHI)	0.441 (0.111)		0.548 (0.131)	
No. Banks in town	1.000 (0.051)		1.426 (0.074)	
Fed. Dist 7	0.123 (0.110)	-0.619 (0.097)	0.181 (0.132)	-0.715 (0.097)
Fed. Dist 8	0.265 (0.101)	-0.027 (0.093)	0.280 (0.113)	-0.012 (0.092)
Fed. Dist 9	0.069 (0.207)	-0.239 (0.206)	0.021 (0.238)	-0.394 (0.213)
y_1		1.314 (0.090)		
y_1^*				0.122 (0.017)

negative impacts of regional contagion diminish over time. Although the initial panic is not captured here, the contagion channel disappears and presents market benefits. The increased lending by the remaining banks in the town increases economic activity and restores confidence in the financial system.

The second channel for the spread of contagion is represented by the variable △ *Correspondents*. The result for this covariate demonstrates that a reduction in a bank's correspondent network has a negative impact on bank lending. This result aligns with intuition because correspondent banks often provided short-term commercial paper to smaller banks and urged them to extend credit (Richardson and Troost, 2009). When a bank was removed or a failure occurred in these networks, it constrained the liquid assets available and, as a result, banks extended less credit. The marginal effect of a correspondent removal, averaged over both observations and MCMC draws, is approximately -0.037 and -0.051 for the latent and observed specifications, respectively. In other words, a reduction in a bank's correspondent network decreases lending by about $\$370 - \$510$ for every $\$10,000$. This contagion channel presents moderate long-term negative effects. Policy-makers providing *ex post* liquidity following a downturn may want to focus relief efforts on banks experiencing failures across their correspondent networks as the negative effects may still be lingering.

Other results for the first equation display that the higher share of liability held at an individual bank (*Liab./Town Liab.*) and the higher the market concentration in a town (*HHI*), the more likely that town was to experience a bank failure between 1929-1932. In addition, relative to the 6th Federal Reserve district, the 8th district is more likely to experience bank failures, which accords well with Richardson and Troost's (2009) paper. The results for the second equation show that older banks and national banks have higher lending in 1933 relative to younger and nonmember institutions.

Model comparison results are presented in Table 2.3 and reveal that the data strongly support the latent endogeneity specification, where the full magnitude of y^*_{i1}, even values whose extent is driven by observed covariates and unobserved factors outside $\{0,1\}$, is relevant for bank lending. The marginal likelihood is nearly 50 points higher on the log scale, giving the observed specification a posterior model probability of approximately 0. The latent measure of a nearby bank failure captures unobserved aspects of bank profitability and the loss of confidence cultivating through these struggling local economies, which better explain bank lending. This result highlights an interesting point from Calomiris and Mason's (2003b) paper where they state, "Indicator variables are uninformative about the particular mechanism through which illiquidity and contagion produces a bank failure." Calomiris and Mason (2003b) further urge researchers to interpret indicators with caution when examining contagion

Table 2.3: Banking application – results for the variance-covariance matrix and marginal likelihood estimates.

	Observed Endogeneity	**Latent Endogeneity**
ω_{12}	-1.078 (0.053)	-0.305 (0.130)
ω_{22}	2.125 (0.087)	2.044 (0.088)
$\mathbf{\Omega}_{corr}$	$\begin{pmatrix} 1 & -0.740 \\ -0.740 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.214 \\ -0.214 & 1 \end{pmatrix}$
Log-Marginal Lik.	-1713.1	-1666.3
Numerical S.E.	(0.423)	(0.245)

as there may be evidence of missing fundamentals and loss of financial confidence. The issues these authors refer to can be mitigated by employing a latent measure for a nearby bank failure. The authors' analysis explains the strong support from the data for the latent specification and further corroborates the underlying hypothesis.

Without considering both modeling approaches, a researcher using observed endogeneity could misinterpret the relationship between regional failures and bank lending, and fundamentally mispecify financial panics. The nonlinear dichotomizing mechanism in which regional failures determine bank performance is inadequate. The latent variable approach encompasses the speculative nature of bank runs. One of the most documented and well understood features of banking crises is asymmetric information, which dates back to the original research by Bagehot (1873). Depositors, bankers, and central bankers lack credible information (Diamond and Dybvig, 1983; Gorton, 1985), resulting in speculative and fundamental bank runs. When an econometrician is examining the interactions involved in banking panics, asymmetric information should be apparent and the models being considered should reflect this feature.

Table 2.3 also displays the results for the variance-covariance matrix $\mathbf{\Omega}$. There is a negative correlation between the equations for a nearby bank failure and bank performance. After controlling for a number of balance sheet and county characteristics, the errors present a negative relationship, which implies there are harmful effects of a nearby bank failure not controlled for in the model. This result holds for both model specifications but there is a clear magnitude difference between the two. The observed specification may overstate the relationship and the "true incidence of panic, since relevant fundamentals are likely omitted" from the model (Calomiris and Mason, 2003b).

The results discussed thus far employ a training sample prior of 100 banks. The sensitivity of the results to the hyperparameters is displayed in Table 2.4. The model rankings do not change across different training sample sizes. This indicates that the data speak loudly for the results and support the latent endogeneity model. Researchers interested in modeling and understanding decisions and relationships of banks in adverse macroeconomic conditions can employ latent variables to accommodate asymmetric information and to better capture the interactions between the outcomes of interest.

Table 2.4: Banking application – sensitivity analysis for different training sample sizes.

Training Sample Size	Observed Log-Marg. Lik.	Latent Log-Marg. Lik.
No sample (priors centered at zero)	-1837.2	-1778.9
50	-1769.5	-1736.1
100	-1713.1	-1666.3
150	-1666.1	-1621.1
200	-1627.5	-1556.8

2.4 Education

The second application considers the impact of education on adult socioeconomic status. The return on schooling has been an ongoing area of research for empirical economists. Despite the great deal of attention focused on accommodating issues with survey responses, latent endogenous variables are lacking in this literature.

Education's impact on adult socioeconomic status is of particular interest because there are convincing arguments for both observed and latent endogeneity. Observed endogeneity is motivated by societal evaluations of education, which are generally looked at by crossing particular achievement thresholds or cut-points, e.g., high school degree and college degree. Without a college degree, individuals cannot apply for many jobs notwithstanding 15 years of schooling. Therefore, labor market outcomes are determined by these observed degree or threshold crossing indicators. Alternatively, a common issue in the analysis of returns to schooling is the inadequacy of measures for motivation and ability, which also explain socioeconomic status. Observed measures for ability have been proposed, such as standardized test scores. However, these data are often not available. Therefore, the latent representation of education can act as a proxy for unobservable characteristics linking education and socioeconomic outcomes. While the hypotheses for each approach are convincing, there have been no attempts to compare these models, which is necessary to ensure the most accurate specification is employed. Furthermore, knowing which approach is best supported by the data advances our understanding of how education determines socioeconomic status.

Studying education provides a unique opportunity for model comparison. Most education and socioeconomic outcomes are discretized by ordered categories. For instance, education categories can be defined by degree level and socioeconomic status can be categorized by income brackets. This ordinal data setting makes the observed specification in (2.1) invalid, and instead, the system in (2.3) can be employed. The specification for observed endogeneity for 5 education categories (defined shortly) and individuals $i = 1, \ldots, n$ is:

$$\begin{aligned} y_{i1}^* =& \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1}, \\ y_{i2}^* =& \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + 1\{y_{i1} = 2\}\gamma_{22} + 1\{y_{i1} = 3\}\gamma_{23} + 1\{y_{i1} = 4\}\gamma_{24} + 1\{y_{i1} = 5\}\gamma_{25} + \varepsilon_{i2}, \end{aligned} \tag{2.7}$$

where $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})' \sim N_2(0, \boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$. This specification differs from (2.1) because the endogenous covariate enters as a set of dummy variables for each category, whereas previously it entered as a single endogenous regressor.[3] The latent specification remains identical to the system in (2.2), which is now a more parsimonious model relative to the observed approach. It is important to note that the elements in $\boldsymbol{\Omega}$ are left free. Location and scaling restrictions are accommodated by fixing two cut-points. The different approaches for identification in multivariate ordered probit models are discussed in Jeliazkov et al. (2008).

The model is characterized by two dependent variables, where $\mathbf{y}_i^* \equiv (y_{i1}^*, y_{i2}^*)'$ are the continuous latent data and $\mathbf{y}_i \equiv (y_{i1}, y_{i2})'$ are the corresponding discrete observed data. For equations $k = 1, 2$, the latent variables relate to the observed ordered outcomes by $y_{ik} = \sum_{j=1}^{J} 1\{y_{ik}^* > \alpha_{k,j-1}\}$ for J ordered alternatives where α_{kj} is a cut-point between the categories given by,

$$y_{i1} = \begin{cases} 1 & \text{Less than high school} \\ 2 & \text{High school degree} \\ 3 & \text{Some college} \\ 4 & \text{College degree} \\ 5 & \text{Graduate education} \end{cases}, \qquad y_{i2} = \begin{cases} 1 & \text{Poverty line and below} \\ 2 & \text{Lower - middle class} \\ 3 & \text{Middle class and up.} \end{cases}$$

[3] If y_{i1} entered directly, this would lead to a cardinal interpretation of the categories, which is incorrect.

The outcome y_{i1} represents the amount of education an individual completes and y_{i2} represents an individual's socioeconomic status. The second outcome is measured by an income-to-needs ratio. Income is measured using the actual amount of total income, which is the sum of taxable income and transfer income. Needs is measured as the poverty threshold taken from the Census Bureau. These thresholds are based on family size and age of the household. An income-to-needs ratio below 1.3 indicates the poverty line and below, between 1.3-3 indicates lower-middle class, and above 3 represents the middle class and up.[4] The endogenous covariate y_{i1} represents the impact of education on adult socioeconomic status.

The data collected for this application are from the Panel Study of Income Dynamics (PSID). The sample includes $2,779$ respondents from the 1999 survey. The data set contains information on childhood health, parental socioeconomic status, parental education, adult socioeconomic status, adult health, and educational attainment for individuals between the ages of 30-50. The year 1999 is selected because it features retrospective reports on childhood health. Table 2.5 offers descriptive statistics on the data and details the discretization for a number of variables.

2.4.1 Estimation

The model is completed by specifying the prior distributions,

$$\boldsymbol{\lambda} \sim \mathcal{N}(\boldsymbol{d}_0, \mathbf{D}_0),$$
$$\boldsymbol{\Omega} \sim \mathcal{IW}(\nu_0, \mathbf{R}_0).$$

The hyperparameters are selected using a training sample of 200 individuals. Algorithm 2.2 presents the sampling methods to simulate the posterior distribution for the

[4]The cut-point 1.3 is selected because it is the threshold for food stamps.

Table 2.5: Descriptive statistics for the sample of 2,779 respondents from the PSID.

Variable	Sample Proportion	Variable	Sample Proportion
Respondent Education (Educ)		**Mother's Education** (Meduc)	
< High School Degree	0.17	< High School Degree	0.33
High School Degree	0.32	High School Degree	0.47
Some College	0.25	Some College	0.09
College Degree	0.17	College Degree	0.08
Graduate School	0.09	Graduate School	0.03
Father's Education (Feduc)		**Childhood Health**	
< High School Degree	0.39	Poor	0.16
High School Degree	0.38	Average	0.38
Some College	0.07	Excellent	0.46
College Degree	0.10	**Adult Health**	
Graduate School	0.06	Poor	0.10
Adult Socioeconomic Status		Average	0.62
Low	0.13	Excellent	0.28
Medium	0.27	**Marital Status**	
High	0.60	Single	0.16
Parental Socioeconomic Status (pSES)		Divorced	0.24
Low	0.25	Married	0.60
Medium	0.45	**Race**	
High	0.30	White / Asian	0.63
Debt		Non-white	0.37
Debt	0.55	**Employment**	
No Debt	0.45	Employed	0.88
Sex		Unemployed	0.12
Male	0.76	**Age (average)**	40
Female	0.24		

observed endogeneity specification, which follow from Jeliazkov et al. (2008). Note, the cut-point parameters α_{kj} are transformed to ensure the ordering constraints, so

$$\delta_{kj} = \ln(\alpha_{kj} - \alpha_{k,j-1}), \text{ where } 2 \leq j \leq J-1 \text{ for equations } k = 1, 2.$$

Algorithm 2.2. *MCMC Estimation Algorithm - Observed Specification*

1. *For each equation k, sample $\boldsymbol{\delta}_k, \mathbf{y}_k^* | \mathbf{y}, \boldsymbol{\lambda}, \mathbf{y}_{\setminus k}^*$ as follows:*

   (a) *Sample $\boldsymbol{\delta}_k | \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\Omega}, \mathbf{y}_{\setminus k}^*$ using the Metropolis-Hastings algorithm*

   (b) *Sample $y_{ik}^* | \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\Omega}, \mathbf{y}_{\setminus k}^* \sim \mathcal{TN}_{(\alpha_{k,j-1}, \alpha_{kj})}(\mu_{k|\setminus k}, V_{k|\setminus k})$ for $i = 1, \ldots, n$.*

2. *Sample $[\boldsymbol{\lambda} | \boldsymbol{y}^*, \boldsymbol{\Omega}] \sim N\left(\hat{\boldsymbol{d}}, \hat{\boldsymbol{D}}\right)$, where $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{D}}$ are given by*

$$\hat{\boldsymbol{d}} = \hat{\boldsymbol{D}} \left(\boldsymbol{D}_0^{-1} \boldsymbol{d}_0 + \sum_{i=1}^n \boldsymbol{W}_i' \boldsymbol{\Omega}^{-1} \mathbf{y}_i^* \right) \text{ and } \hat{\boldsymbol{D}} = \left(\boldsymbol{D}_0^{-1} + \sum_{i=1}^n \boldsymbol{W}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{W}_i \right)^{-1}.$$

3. *Sample* $\boldsymbol{\Omega} \sim \mathcal{IW}(\nu_0 + n, \mathbf{R}_0 + (\mathbf{y}^* - \mathbf{W}\boldsymbol{\lambda})'(\mathbf{y}^* - \mathbf{W}\boldsymbol{\lambda}))$.

As a matter of notation, "$\backslash k$" is used to represent all elements in a set except the kth one. Estimation of the latent specification follows these steps closely, however, employs the reduced-form trick discussed in (2.6) for the data augmentation step in $1(b)$.

2.4.2 Results

The results are based on 11,000 MCMC draws with a burn-in of 1,000. Analysis of the sensitivity of the results to the training sample size is conducted as in the first application. The results again show no sensitivity to the training sample size and model rankings do not change for different hyperparameters. The inefficiency factors for the parameters remain low with the highest values (≈ 20) occurring for the parameters on the endogenous covariates in both specifications. The following discussion reviews the basic results for each equation, then the model comparison results.

Table 2.6 presents the posterior means and standard deviations for both specifications. The results from the first equation accord well with the existing literature on the determinants of educational attainment (Haveman and Wolfe, 1995). The results show that parental education and parental socioeconomic status play a positive role in educational attainment. Parents with more income are able to invest in their children's education with schooling supplies, tutors, and financial assistance in college. Furthermore, parents who themselves achieve a higher level of education often motivate their children to do the same. The results also indicate that whites complete

Table 2.6: Education application – posterior means and standard deviations for the bivariate system of equations.

	Observed Endogeneity		**Latent Endogeneity**	
	Education	**SES**	**Education**	**SES**
Intercept	-0.683 (0.593)	-1.706 (0.193)	-0.696 (0.597)	-1.353 (0.199)
Child Health-Exc	-0.269 (0.586)		-0.265 (0.586)	
Child Health-Avg	-0.429 (0.587)		-0.425 (0.586)	
Feduc-High School	0.188 (0.058)		0.189 (0.059)	
Feduc-Some College	0.616 (0.097)		0.613 (0.095)	
Feduc-College	0.840 (0.094)		0.835 (0.093)	
Feduc-Graduate	1.033 (0.114)		1.028 (0.115)	
Meduc-High School	0.197 (0.060)		0.204 (0.060)	
Meduc-Some College	0.499 (0.093)		0.498 (0.092)	
Meduc-College	0.534 (0.103)		0.531 (0.102)	
Meduc-Graduate	0.780 (0.147)		0.775 (0.147)	
pSES-High	0.339 (0.067)	0.073 (0.065)	0.351 (0.064)	0.035 (0.067)
pSES-Med	0.344 (0.053)	0.079 (0.058)	0.351 (0.060)	0.049 (0.059)
Single		-0.003 (0.064)		0.005 (0.064)
Married		0.154 (0.062)		0.168 (0.062)
Employed		0.634 (0.067)		0.637 (0.067)
Debt		0.202 (0.042)		0.196 (0.042)
Adult Health-Exc		0.359 (0.080)		0.352 (0.080)
Adult Health-Avg		0.244 (0.069)		0.245 (0.070)
Educ-High School		0.553 (0.086)		
Educ-Some College		0.813 (0.127)		
Educ-College		1.312 (0.172)		
Educ-Graduate		1.702 (0.241)		
Latent Educ (y_1^*)				0.424 (0.055)
White	0.344 (0.053)	0.416 (0.056)	0.347 (0.053)	0.389 (0.056)
Age	0.030 (0.004)	0.015 (0.004)	0.030 (0.004)	0.014 (0.004)
Male	0.101 (0.055)	0.359 (0.064)	0.101 (0.054)	0.358 (0.063)

more education relative to non-whites, and males complete more schooling relative to females.

The results of the second equation coincide well with intuition and what is often found in the literature on socioeconomic achievement. Parental socioeconomic status positively affects adult socioeconomic status and individuals who are married have a higher income-to-needs ratio relative to divorced individuals. An interesting result is the positive coefficient on the debt variable. The debt variable is measured by

Table 2.7: Education application – results for the variance-covariance matrix and marginal likelihood estimates.

	Observed Endogeneity	**Latent Endogeneity**
ω_{11}	1.258 (0.043)	1.263 (0.043)
ω_{12}	-0.168 (0.074)	-0.195 (0.072)
ω_{22}	0.739 (0.049)	0.744 (0.057)
$\mathbf{\Omega}_{corr}$	$\begin{pmatrix} 1 & -0.174 \\ -0.174 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -0.200 \\ -0.200 & 1 \end{pmatrix}$
Log-Marginal Lik.	-7864.0	-7862.9
Numerical S.E.	(0.053)	(0.044)
$\Pr(\mathcal{M}_k \mid y)$	0.333	0.667

summing all debt excluding debt from the purchase of a house. However, debt does not necessarily indicate financial distress, as wealthier individuals have more access to debt and may finance other purchases, including vehicles, businesses, and school loans. The results also indicate that health has a positive relationship with wealth. Healthy individuals are less likely to miss work due to illness or disability and are likely to be more productive.

The covariate of interest, education, positively impacts adult socioeconomic status in both specifications. Relative to no high school degree, the coefficient on each discrete category for additional schooling gets incrementally larger in the observed specification. The results for $\mathbf{\Omega}$, presented in Table 2.7, demonstrate that after controlling for family background, health, and other demographics, there is a negative correlation between education and adult socioeconomic status. Although the direct effect of education on wealth is positive, there is a negative relationship between the errors. This result has been noted in the literature (Becker and Chiswich, 1966; Griliches, 1977), and a simple explanation for this is luck. The basic explanatory variables do not control for instances of good fortune.

Table 2.7 also presents the model comparison results. The marginal likelihoods and posterior model probabilities are very close for both specifications with a slight preference toward the latent endogeneity model. The latent specification is more parsimonious, with three less covariates than the observed endogeneity model. Further investigation into this relationship is necessary due to the lack of model preference. An approach for understanding these interactions is to separate the sample by age cohort. The intuition for this comes from recognizing that individuals' degree level may only be influential in obtaining their first few jobs. As the amount of time an individual is in the labor force increases, the individual accumulates work history, skills, and references, which mitigate the importance of degree level. Consider job postings for entry-level positions, the salary level is often listed as "competitive", whereas for more senior positions, "depending on experience" is a common listing. Following this intuition, both models are compared for two separate age cohorts. The results are displayed in Table 2.8.

The results align with the age group hypotheses. The data support the observed specification for the 30-35 age cohort. Alternatively, the data support the latent specification for the 40-50 age cohort. This is a major result because it displays how the dependence structure between educational attainment and adult socioeconomic status changes with age. For younger individuals, the primary way to signal intelligence and ability is through degree-level. As a result, society evaluates the amount of education completed by observed threshold crossing indicators, thus resulting in labor market outcomes for these individuals. On the other hand, for older individuals, the data support the underlying latent specification, capturing ability, work history, and other factors unobserved by the econometrician. The importance of education diminishes as other elements become more prominent and, eventually, better explain labor market outcomes and wealth. Additionally, the latent approach offers a unique opportunity for studies examining outcomes of older cohorts. Surveys often do not contain a comprehensive work history of the respondents, therefore, if these data

Table 2.8: Education application – model comparison and marginal likelihood estimates for different age cohorts.

	Ages 30–35		Ages 40–50	
	Observed	**Latent**	**Observed**	**Latent**
Observations	843	843	1489	1489
Log-Marginal Lik.	-2450.6	-2455.3	-4196.1	-4192.0
$\Pr(\mathcal{M}_k \mid y)$	0.991	0.009	0.017	0.983

are unavailable, the latent measure of education can offer some insight into these characteristics.

These results truly stress the importance of considering and comparing both specifications and contribute to the literature on returns to schooling by demonstrating the evolution in the dependence structure between education and socioeconomic status. Pathway models employed in labor economics attempt to capture life-cycle interactions and intergenerational transmissions of education, health, and wealth. The complexity of these models increases because these outcomes are dynamic and change over time. Future research can employ both observed and latent measures of these outcomes to better explain these relationships pertaining to age cohorts.

2.5 Concluding Remarks

This paper addresses an important but often overlooked issue, which is the proper specification of endogenous covariates. In multivariate discrete data models, endogeneity can be based on latent or observed data. Bayesian model comparison techniques can be employed to determine which approach is best supported by the data, thus increasing the understanding of the nature of endogeneity and the dependence structure between the relationships being modeled.

While most applied researchers have an *a priori* expectation of the correct model, it is extremely difficult to rule out hypotheses in support of the alternative approach, which is apparent from the two empirical applications considered in this paper. Therefore, model selection provides important insights that resolve competing hypotheses about the interactions of interest. The results from the first application show that the latent representation of a bank failure better explains regional financial contagion, relative to conditioning on its observed counterpart. The latent measure captures the speculative nature of banking panics and accommodates asymmetric information issues discussed in the literature, thus providing a more accurate model of banking crises. The results for the second application show that the observed representation of education better explains socioeconomic outcomes for younger cohorts. This dependence structure changes as individuals age and the latent measure of education becomes more meaningful. This paper employs a bivariate system of equations in the applications, however, these approaches are generalizable to a number of methodologies.

These results stress the importance of employing model selection techniques to distinguish between competing specifications. The issues discussed here are present in any multivariate discrete data setting and should be addressed in a number of applied literatures. Ignoring these techniques can cause a researcher to misinterpret the depicted relationships and misunderstand the nature of endogeneity.

Chapter 3

The Likelihood Function for Discrete Simultaneous Equations

3.1 Introduction

Simultaneous equation models constitute one of the main contributions of economics to statistical science and are of central importance in the analysis of supply and demand systems, strategic games, interactions, and multivariate decision making, among others. Despite the relative complexity of the setting, the analysis of simultaneous equation models for continuous data is now well understood. In contrast, matters are quite different when it comes to models for discrete outcomes. Earlier work has pointed out a variety of complications where basic analogies with the continuous case appear to break down and, even more fundamentally, call into question the very sensibility of specifying simultaneous equation models for discrete data. Specifically, in the case of binary data, an important paradox is that outcome probabilities need not sum up to 1 unless parameter constraints are imposed which essentially remove the simultaneity and replace it with recursive endogeneity. Such constraints have been viewed as questionable and entirely unfounded in economic theory, and even though the problem has perplexed a long line of researchers, it has remained unsolved for over three decades.

This paper provides results that confirm the skeptics' suspicions about the dubious necessity of such constraints. As we show, the problem is not with values that certain parameters can take, but with the proper formulation of the likelihood function. Not only does the derivation of the likelihood function point out that such constraints are not necessary, but it also provides insights on the heretofore tenuous link between continuous and discrete simultaneous equation models. The derivation also clarifies the set of constraints that must be met in order for the model to be identified, yet those are of a much more familiar (perfect classification) type that is sample (not model) specific, and therefore does not call into question the sensibility of the specification.

The formulation, coherency, and estimation of simultaneous equation models with discrete dependent variables were initially considered in Amemiya (1974, 1978), Heckman (1978), Nelson and Olson (1978), and Maddala (1983) and has been an on-going research topic for econometricians. In these papers, a coherency condition is required to guarantee the existence of a unique reduced form, which as mentioned previously, puts dubious constraints on the parameter values in the model. This condition is also known as "principle assumption" or condition for "logical consistency". In addition to the coherency condition, the standard identifiability conditions hold, namely that each equation has a unique element in its matrix of covariates that appears in no other equations (Judge et al., 1985). For a discussion of the appropriate coherency conditions and identification for each model, see Blundell and Smith (1994).

More recently, the industrial organization literature has discussed these models in the form of discrete games with complete information (Bresnahan and Reiss, 1991;

Berry, 1992). This literature formulates the absence of a one-to-one mapping between parameters and outcomes as a problem of multiplicity of equilibria and has proposed several ways to deal with this multiplicity, including obtaining the identified set estimates rather than unidentified point estimates of parameters (Ciliberto and Tamer, 2009) and estimating an equilibrium selection rule (Bajari et al., 2010; Narayanan, 2013).

The goal of this paper is to provide a new derivation of the likelihood function for discrete simultaneous equation models, which is based on Markov chain theory, and to resolve these coherency paradoxes highlighted in earlier work. Furthermore, the paper suggests a new methodological approach for econometric modeling, where models can be built from a set of seemingly circular conditional relations. The methods in this paper employ a conditional-conditional decomposition, which is a new alternative to those used in other econometric models. The outline of the paper is as follows: Section 3.2 provides an overview of the problems that arise in discrete simultaneous equation models, Section 3.3 offers the derivation of the likelihood function, and Section 3.4 provides both classical and Bayesian estimation methods, discusses new identification conditions, and presents a simulation study. Two real-data applications are considered in Section 3.5 and Section 3.6 offers concluding remarks.

3.2 Overview of the Problem

In the interest of clarity and simplicity, we begin with a simple bivariate probit model for binary data in the absence of simultaneity and gradually increase the generality of the setting. Once the basic ideas are presented, this task turns out to be relatively straightforward.

Consider the basic bivariate probit with no simultaneity where,

$$\begin{aligned} z_{i1} &= x_{i1}'\beta_1 + \varepsilon_{i1} \\ z_{i2} &= x_{i2}'\beta_2 + \varepsilon_{i2} \end{aligned} \tag{3.1}$$

for $i = 1, \ldots, n$ and $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}) \sim N_2(0, \Sigma)$, where $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The unit variance on the diagonal of Σ is the usual normalization for identification in the probit model. The observed choice $y_i \equiv (y_{i1}, y_{i2})'$, is related to the latent data through

$$y_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \leq 0 \end{cases} . \tag{3.2}$$

Figure 3.1 provides a graphical representation of the model, where the latent $z_i \equiv (z_{i1}, z_{i2})'$ comes from a bivariate normal density with mean $\mu_i \equiv (x_{i1}'\beta_1, x_{i2}'\beta_2)'$ and the contours represent the correlation implied by Σ. The four possible outcome probabilities, $\Pr(y_{i1} = 1, y_{i2} = 1)$, $\Pr(y_{i1} = 0, y_{i2} = 1)$, $\Pr(y_{i1} = 0, y_{i2} = 0)$, and $\Pr(y_{i1} = 1, y_{i2} = 0)$, are captured in quadrants 1 through 4, respectively. In this case, it is easy to see that these probabilities sum to 1.

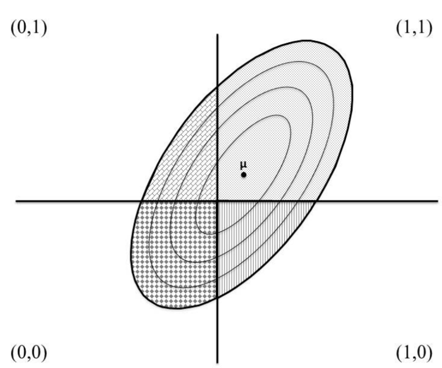

Figure 3.1: The simple bivariate probit model. The shaded regions of the 4 possible outcome probabilities sum to 1.

To illustrate the apparent logical inconsistency that appears in models with simultaneity, consider a simple example in Maddala (1983), where the model is specified as

$$
\begin{aligned}
z_{i1} &= x_{i1}'\beta_1 + y_{i2}\gamma_1 + \varepsilon_{i1} \\
z_{i2} &= x_{i2}'\beta_2 + y_{i1}\gamma_2 + \varepsilon_{i2}
\end{aligned}
\tag{3.3}
$$

and the relationship between y_i and z_i is given in (3.2). The system generalizes trivially to models with q equations by writing

$$
z_i = BX_i + \Gamma y_i + \varepsilon_i, \text{ where } B = \begin{pmatrix} \beta_1' & 0 & \cdots & 0 \\ 0 & \beta_2' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta_q' \end{pmatrix}, \qquad X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iq} \end{pmatrix}, \tag{3.4}
$$

and Γ is a suitably defined matrix with zeros on the main diagonal. In the bivariate case, because of the simultaneity in the model, the mean of z_i shifts for each possible outcome y_i, and hence there are 4 separate distributions as shown in Figure 3.2. Given a specific value of y_i, the distribution with the corresponding mean is integrated over the relevant quadrant implied by that outcome. However, because the integrals are computed with respect to different distributions (the shaded regions in Figure 3.2), they need not sum to 1 unless certain coherence constraints are imposed.

Whereas Figure 3.2 captures the general case, Maddala (1983) shows that even when the errors are independent and there are no exogenous covariates, the probabilities

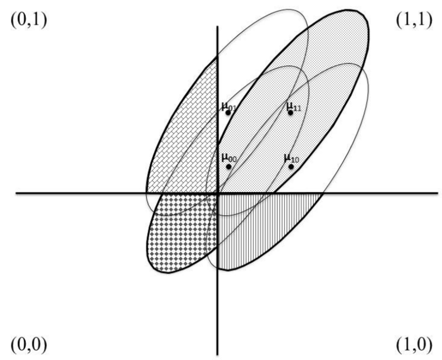

Figure 3.2: The simultaneous system. The shaded regions of the 4 possible outcome probabilities no longer sum to 1.

need not sum to 1. In particular, given these simplifying assumptions, we have

$$
\begin{aligned}
\Pr(y_1 = 1, y_2 = 1) =& F_1(\gamma_1)F_2(\gamma_2)\\
\Pr(y_1 = 1, y_2 = 0) =& F_1(0)[1 - F_2(\gamma_2)]\\
\Pr(y_1 = 0, y_2 = 1) =& [1 - F_1(\gamma_1)]F_2(0)\\
\Pr(y_1 = 0, y_2 = 0) =& [1 - F_1(0)][1 - F_2(0)],
\end{aligned}
$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are the distribution functions of ε_{i1} and ε_{i2}, which sum to

$$1 + F_1(0)F_2(0) - F_1(\gamma_1)F_2(0) - F_1(0)F_2(\gamma_2) + F_1(\gamma_1)F_2(\gamma_2).$$

Unless $\gamma_1 = 0$ or $\gamma_2 = 0$, these probabilities do not sum to 1. The case where $\gamma_1 = 0$ is depicted in Figure 3.3, where because of the constraint, the sum of the two shaded regions equals 1. The implication of such constraints is rather strong – they rule out full simultaneity and only allow recursive endogeneity. Moreover, these kinds of constraints will typically have no economic justification and will often be challenged by applied researchers. In the next section, we show that such constraints are not required and that discrete data models with full simultaneity are indeed quite

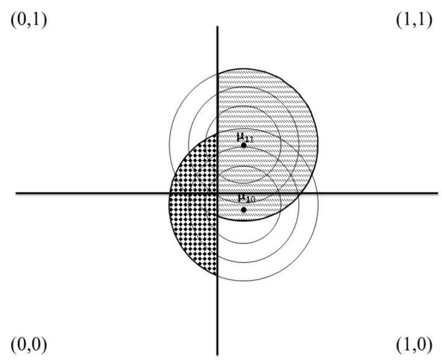

Figure 3.3: Example of coherency constraint, forcing the shaded regions to sum to 1.

reasonable. As we point out, the problem lies with the proper formulation of the likelihood function and not with particular parameter restrictions.

Before we continue, however, we briefly digress to mention that this problem does not arise in systems of simultaneous equations that are specified in terms of the latent data, e.g.,

$$\begin{aligned} z_{i1} =& x'_{i1}\beta_1 + z_{i2}\gamma_1 + \varepsilon_{i1} \\ z_{i2} =& x'_{i2}\beta_2 + z_{i1}\gamma_2 + \varepsilon_{i2}. \end{aligned} \tag{3.5}$$

In this specification, the structural equations do not involve the observed values y_i but their unobserved latent counterparts z_i. This greatly simplifies estimation because the reduced form system can be easily obtained from equation (3.5) and inference can proceed as in the usual multivariate probit model. To see this, note that in matrix notation, the model can be written as

$$z_i = BX_i + \Gamma z_i + \varepsilon_i, \text{ where } B = \begin{pmatrix} \beta'_1 & 0 \\ 0 & \beta'_2 \end{pmatrix}, \qquad X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix}, \tag{3.6}$$

resulting in the reduced form

$$\begin{aligned} z_i =& (1-\Gamma)^{-1} B X_i + (1-\Gamma)^{-1}\varepsilon_i \\ =& \Pi X_i + \nu_i, \end{aligned}$$

where $\Pi = (1-\Gamma)^{-1}B$ is a matrix of reduced form coefficients. A variety of existing methods, e.g., maximum simulated likelihood or Bayesian Markov chain Monte Carlo (MCMC) simulation, can be implemented to estimate the model and recover the structural parameters, subject to proper identification. Early reviews of different estimators for the structural parameters in the latent specification are offered in Blundell and Smith (1989). A priori, however, the fact that the latent specification does not suffer from the same difficulties as the one based on observed outcomes, does not mean that we should not consider models where outcomes depend on the observed y_i. As Maddala (1983) has argued, the appropriate model could often be one where the "intentions about y_{i1} depend on actual actions on y_{i2} and intentions about y_{i2} depend on actual actions on y_{i1}", which justifies formulation (3.3).

3.3 Derivation of the Likelihood Function

3.3.1 Continuous Outcomes

To motivate the derivation of the discrete data likelihood, we first revisit the continuous data case

$$y_i = BX_i + \Gamma y_i + \varepsilon_i,$$

where the matrices B, X_i and Γ are defined as in (3.6). The value of y_i that solves this system

$$y_i = (1-\Gamma)^{-1}(BX_i + \varepsilon_i)$$

produces the likelihood contribution $f(y_i|\theta)$, for $i = 1, \ldots, n$, where θ represents the model parameters β_1, β_2, γ_1, γ_2, and the unique free elements of Σ. For a simple two equation case, the reduced-form is often found by substituting or plugging one equation into the other, e.g.,

$$\begin{aligned} y_{i1} =& x'_{i1}\beta_1 + \gamma_1(x'_{i2}\beta_2 + y_{i1}\gamma_2 + \varepsilon_{i2}) + \varepsilon_{i1} \\ y_{i1} =& \frac{1}{1-\gamma_1\gamma_2}\left(x'_{i1}\beta_1 + \gamma_1 x'_{i2}\beta_2 + \gamma_1\varepsilon_{i2} + \varepsilon_{i1}\right). \end{aligned} \tag{3.7}$$

The solution for y_i can also be obtained as the fixed point in the dynamic system

$$\begin{aligned} y_i^{(t+1)} =& (BX_i + \varepsilon_i) + \Gamma y_i^{(t)} \\ y_i^{(t)} =& y_i^{(t+1)}, \end{aligned} \tag{3.8}$$

which is initialized with some starting value $y_i^{(0)}$.

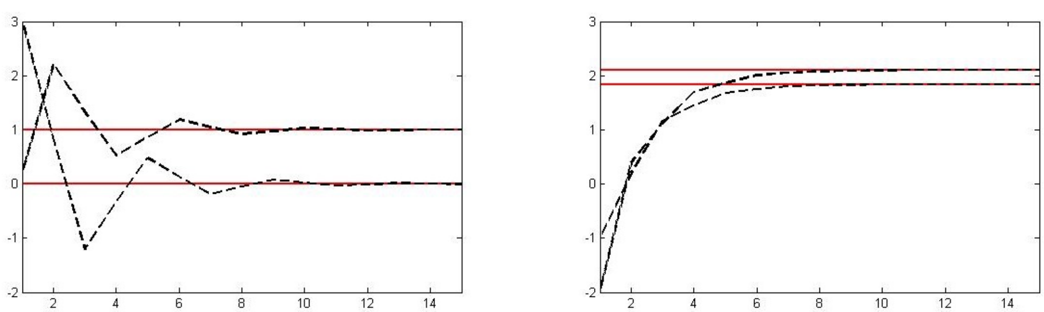

Figure 3.4: The dashed line represents the iterative solution and the solid line represents the analytical solution.

Figure 3.4 shows examples of how the iterative solution emerging from (3.8) reaches

$(1-\Gamma)^{-1}(BX_i + \varepsilon_i)$.[1] The representation in (3.8) also implies a useful way of thinking about simultaneous equation systems as models of sequential feedback and partial adjustment, conceptually similar to the more familiar impulse response analysis in dynamic systems, where values of one variable affect, and are affected by, other variables until the system reaches steady state. The reason that this alternative way of obtaining $f(y_i|\theta)$ is discussed here is that it provides an analogy with the derivation of $f(y_i|\theta)$ in the discrete case, as discussed next.

3.3.2 Discrete Outcomes

To understand the difficulty of obtaining the likelihood function in the discrete simultaneous equation model in (3.3), first consider the multivariate probit model in equations (3.1) and (3.2). These two equations identify $f(z_i|\theta)$ and $f(y_i|z_i, \theta)$, respectively, whereby $f(y_i|\theta)$ is obtained as

$$f(y_i|\theta) = \int f(y_i|z_i, \theta) f(z_i|\theta) dz_i \tag{3.9}$$

using simulated maximum likelihood or MCMC methods. This marginal-conditional decomposition is not available in the simultaneous case. Instead, we have two conditional distributions, $f(z_i|y_i, \theta)$ and $f(y_i|z_i, \theta)$, obtained from equations (3.3) and (3.2), respectively. The problem of obtaining $f(y_i|\theta)$ cannot be approached as in (3.9) because the joint distribution $f(y_i, z_i|\theta)$ is not the product of the 2 ingredient distributions defining the model. This is the fundamental complication motivating our analysis and has hampered earlier theoretical approaches and empirical applications for some time.

[1] Recall that $(I-\Gamma)^{-1} = I + \Gamma + \Gamma^2 + \Gamma^3 + \ldots$. For some parameter settings, the set of starting values for which the iterative solution converges to $(1-\Gamma)^{-1}(BX_i + \varepsilon_i)$ is a proper subset of $\Re^d$, i.e., there could be starting values which lead to divergent sequences $\{y_i^{(t)}\}$. This problem does not arise in discrete data models—our main object of interest—and hence we do not dwell on it here.

Further complications arise because the substitution approach to finding the reduced form for continuous simultaneous equations in (3.7) only obfuscates matters for discrete outcomes, where substituting leads to

$$z_{i1} = x_{i1}'\beta_1 + \gamma_1 1\{x_{i2}'\beta_2 + \gamma_2 1\{z_{i1} > 0\} + \varepsilon_{i2} > 0\} + \varepsilon_{i1}.$$

This argument is circuitous because $\Pr(z_{i1} > 0) = \Pr(y_{i1} = 1)$ depends on $1\{z_{i1} > 0\}$. This is a tautology because if $1\{z_{i1} > 0\} = 1$ then $\Pr(z_{i1} > 0) = 1$. Looking at this differently, it is clear that

$$\begin{aligned} y_{i1} =& 1\{x_{i1}'\beta_1 + y_{i2}\gamma_1 + \varepsilon_{i1} > 0\} \\ y_{i2} =& 1\{x_{i2}'\beta_2 + y_{i1}\gamma_2 + \varepsilon_{i2} > 0\} \end{aligned} \tag{3.10}$$

is not the likelihood or data generating process. It tells you what y_i would be if you had y_i, and therefore it cannot be used to generate data. Later in this section we will come back to (3.10) and discuss it in terms of the actual likelihood function.

Here, we derive the likelihood contribution $f(y_i|\theta)$ from the theory of Markov processes and use it to show that many of the aforementioned paradoxes vanish. The derivation rests on the observation that the two conditional densities, $f(y_i|z_i, \theta)$ and $f(z_i|y_i, \theta)$, can form a Markov chain (densities in a Gibbs sampler), which can be iterated to yield the joint distribution $f(y_i, z_i|\theta)$. Ignoring z_i yields draws from the distribution $f(y_i|\theta)$. This approach presents an important avenue for future modeling – employing conditional distributions – which is a new alternative to the existing conditional-marginal and marginal-marginal decompositions. Conditional-marginal modeling is popular and well understood in econometrics, as is demonstrated in the multivariate probit example in (3.9). Marginal-marginal decompositions are employed in copula modeling frameworks which "couple" marginal distributions (Trivedi and Zimmer, 2005). An important consideration with the conditional-conditional approach is that, in general, the existence of full conditionals is not sufficient for the existence of a joint distribution. The existence of $f(y_i, z_i|\theta)$ in this case is guaranteed under the mild condition that y_i takes only a finite number of possible values. Because we are dealing with discrete data, the existence of $f(y_i, z_i|\theta)$ is not an issue; however, in continuous data settings, the existence would require more attention. By iterating $f(y_i|z_i, \theta)$ and $f(z_i|y_i, \theta)$ to yield $f(y_i, z_i|\theta)$, we can obtain $f(y_i|\theta)$, which is our original object of interest.

A simple approach to obtaining $f(y_i|\theta)$ is to use the draws $y_i^{(g)}$ available from $\{z_i^{(g)}, y_i^{(g)}\} \sim f(z_i, y_i|\theta)$ for $g = 1, \ldots, G$. The draws can then be used to evaluate $f(y_i|\theta)$ through the frequency estimator

$$\hat{f}(y_i|\theta) = \frac{1}{G}\sum_{g=1}^{G} 1\left\{y_i^{(g)} = y_i\right\}.$$

This estimator is general and applicable to a variety of settings, however, its practical appeal is limited because it is not bounded away from 0 or 1, it is not differentiable, and it is not suitable for evaluating low probability events. To deal with these deficiencies, we present another way of evaluating $f(y_i|\theta)$, based on recognizing that Markov chains have a unique invariant distribution that can be obtained by considering the Markov transition matrix between states. To clarify, consider the setup in (3.3) where the possible outcomes can be assigned to the following state vector $s = (s_1, \ldots, s_p)'$ (in this case $p = 4$)

$$y_{s_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad y_{s_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad y_{s_3} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad y_{s_4} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The Markov transition matrix P where

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix},$$

can be used to obtain the steady state as the left eigenvector of P which satisfies $\pi' = \pi' P$, where π' is the characteristic vector of P corresponding to characteristic root equal to 1 or

$$\pi_j = \sum_i \pi_i p_{ij}, \quad j = 1, \ldots, S,$$

(Meyn and Tweedie, 1993; Greenberg, 2008, Ch. 6). The reason we are interested in the vector π is because it contains the likelihood values $f(y_i|\theta)$ for the possible states, i.e.,

$$\pi = (\pi_1, \pi_2, \ldots, \pi_s)' = (f(y_{s_1}|\theta), \ldots, f(y_S|\theta))'.$$

In other words, the quantity we seek to evaluate – the likelihood contribution $f(y_i|\theta)$ – is the element of π where $y_{s_j} = y_i$, i.e. $f(y_i|\theta) = \pi_{ij}$ where π_i is the eigenvector corresponding to P and y_i is the index of the state in the state vector where $y_{s_j} = y_i$.

Looking back at (3.10), we can now view a similar equation as a transition equation in a Markov chain whose iteration yields $f(y_i|\theta)$,

$$\begin{aligned} y_{i1}^{(t+1)} &= 1\{x_{i1}'\beta_1 + y_{i2}^{(t)}\gamma_1 + \varepsilon_{i1} > 0\} \\ y_{i2}^{(t+1)} &= 1\{x_{i2}'\beta_2 + y_{i1}^{(t)}\gamma_2 + \varepsilon_{i2} > 0\}. \end{aligned} \tag{3.11}$$

The solution is a long sequence of instantaneous adjustments and the model is completed by requiring the distribution of y_i on the right side of (3.11) to be that of y_i on the left side. Equation (3.11) bears striking similarity to the continuous case in (3.8).

Identification here is obtained via the usual conditions in Markov chain theory. In order for the Markov chain defined by P to converge to a suitable invariant distribution π, it must be irreducible and aperiodic. For example, P should not be block diagonal and the chain should not cycle through the states at regularly-spaced intervals. It is unlikely that P can be reducible or periodic, but it can be nearly reducible or periodic. This problem will manifest itself similarly to perfect classification, and therefore is sample (not model) specific. The contentious parameter restrictions that were previously necessary for identification are no longer required and the sensibility of these models is no longer questionable. These models are coherent and can be estimated in several ways, which will be discussed next.

3.4 Estimation

In this section we discuss estimation of the simultaneous equation model presented earlier. Classical maximum likelihood estimation proceeds as usual under appropriate identification restrictions, in particular, rank and order conditions as well as the restrictions presented in Section 3.3. As usual, the maximum likelihood estimator is defined as

$$\hat{\theta} = \arg\max_{\theta} \ln f(y|\theta), \tag{3.12}$$

where $f(y|\theta) = \prod_i f(y_i|\theta)$, and in our context $f(y_i|\theta)$ is set equal to the element of π where $y_{s_j} = y_i$. To obtain π, we need estimates of the elements of the Markov transition matrix P. Numerical integration may be suitable for bivariate models, but higher dimensions call for simulation-based integration, i.e. Geweke-Hajivassiliou-Keane (GHK), Stern, CRB, CRT, ARK, and ASK methods. For a review of these simulation-based algorithms, see Jeliazkov and Lee (2010). Under typical regularity conditions for maximum likelihood estimation, asymptotically $\hat{\theta} \sim N(\theta_0, \Theta)$ with $\Theta = -E[\partial^2 \ln f(y_i|\theta)/\partial\theta\partial\theta']^{-1}$.

The model can also be estimated using MCMC algorithms. Gibbs sampling à la Albert and Chib (1993) is not available because the likelihood function is now quite complicated compared to that of the multivariate probit and standard data augmentation techniques do not apply. Therefore, this paper presents the accept-reject Metropolis-Hastings (ARMH) algorithm (Tierney, 1994). For a review of this algorithm, see Chib and Greenberg (1995) and Chib and Jeliazkov (2005). With Bayesian methods, we are interested in the posterior density as the target density

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta),$$

where $f(y|\theta)$ is the likelihood obtained from the Markov transition matrix. Here, we will describe the general ARMH algorithm. Let $h(\theta|y)$ denote a source density and $\mathcal{D} = \{\theta : f(y|\theta)\pi(\theta) \leq ch(\theta|y)\}$, where c is a constant and $\mathcal{D}^c$ is the complement of $\mathcal{D}$. Then the ARMH algorithm is defined as follows.

Algorithm 3.1. *ARMH Algorithm*

1. *A-R Step: Generate a draw $\theta' \sim h(\theta|y)$. Accept the draw with probability*

$$\alpha_{AR}(\theta'|y) = min\left\{1, \frac{f(y|\theta')\pi(\theta')}{ch(\theta'|y)}\right\}$$

and repeat the process until the draw is accepted.

2. *M-H step: Given the current value θ and the proposed value θ'*

   (a) *If $\theta \in \mathcal{D}$, set $\alpha_{MH}(\theta, \theta'|y) = 1$*

   (b) *If $\theta \in \mathcal{D}^c$ and $\theta' \in \mathcal{D}$, set $\alpha_{MH}(\theta, \theta'|y) = \frac{ch(\theta|y)}{f(y|\theta)\pi(\theta)}$*

   (c) *If $\theta \in \mathcal{D}^c$ and $\theta' \in \mathcal{D}^c$, set $\alpha_{MH}(\theta, \theta'|y) = min\left\{1, \frac{ch(\theta|y)}{f(y|\theta)\pi(\theta)}\right\}$*

   Return θ' with probability $\alpha_{MH}(\theta, \theta'|y)$, otherwise return θ.

Both estimation techniques, maximum likelihood and ARMH, are tested on artificially generated data. The model considered for the simulation study is a bivariate probit where,

$$
\begin{aligned}
z_{i1} =& x_{i11}\beta_{11} + x_{i12}\beta_{12} + y_{i2}\gamma_1 + \varepsilon_{i1} \\
z_{i2} =& x_{i21}\beta_{21} + x_{i22}\beta_{22} + y_{i1}\gamma_2 + \varepsilon_{i2},
\end{aligned}
$$

for $i = 1, \ldots, 2000$. Table 3.1 reports maximum likelihood estimates and standard errors as well as posterior means and standard deviations. Both estimation techniques recovered the true parameters accurately and quickly.

Table 3.1: Artificial data illustration. The table reports posterior means and standard deviations, which are based on 5,000 MCMC draws, and maximum likelihood estimates.

		MLE		**ARMH**	
Parameters	True	$\hat{\beta}_{MLE}$	s.e.	mean	s.d.
β_{11}	0.40	0.354	(0.035)	0.355	(0.035)
β_{12}	0.50	0.471	(0.038)	0.472	(0.038)
γ_1	1.30	1.300	(0.059)	1.301	(0.065)
β_{21}	0.20	0.199	(0.037)	0.200	(0.036)
β_{22}	0.60	0.599	(0.046)	0.602	(0.045)
γ_2	1.50	1.526	(0.077)	1.530	(0.081)
ρ	0.20	0.219	(0.062)	0.222	(0.058)

To further evaluate the performance of the ARMH algorithm, inefficiency factors are studied over 20 Monte Carlo repetitions. Inefficiency factors are a measure of the extent of mixing of the Markov chain output (Chib, 2001, 2007; Chib et al., 2009). The inefficiency factor of the k-th parameter is defined as $1 + 2\sum_{l=1}^{L} \psi_k(l)(\frac{L-l}{L})$, where $\psi_k(l)$ is the sample autocorrelation at the l-th lag and L is the lag in which the autocorrelations taper off. Small values (near 1) imply that the output is mixing well. Boxplots of the inefficiency factors are displayed in Figure 3.5. The plots suggest these parameters are sampled efficiently as the values for all of the parameters are low and near one. Poor mixing properties are often found for endogenous covariates, however in this case, we see that all parameters are sampled well.

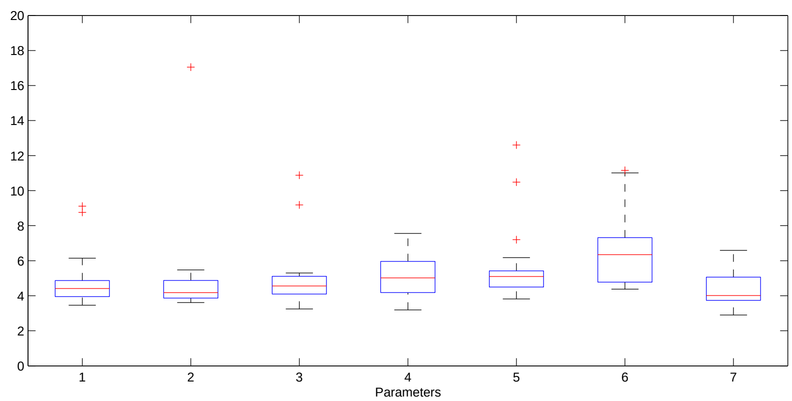

Figure 3.5: Boxplots of the inefficiency factors for $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \gamma_1, \gamma_2, \rho)$, respectively, over 20 Monte Carlo repetitions.

3.5 Application

Data involving simultaneous relationships are very common in the economics literature. In addition to basic supply and demand relationships that involve simultaneity, empirical studies have explored a number of other relationships, including wages and fringe benefits (Vella, 1993), female labor supply and household income (Blundell and Smith, 1989), wages and discrimination (Heckman, 1978), husband and wife fertility decisions (Sobel and Arminger, 1992), travel demand analysis of modal choice (van

Wissen and Golob, 1990), along with many more. Another empirical literature that employs simultaneous equations is that of entry decisions in airline markets (Berry, 1992; Ciliberto and Tamer, 2009). Ciliberto and Tamer (2009) obtain identified set estimates for their application. Their setup is more game-theoretic, so it is not directly comparable to the one in this paper. However, the results reported for the present study are identified point estimates.

This paper considers two canonical examples of simultaneous relationships in labor economics involving data from the Panel Study of Income Dynamics (PSID). The first application is motivated by Blundell and Smith (1994) who study a joint model of women's labor force participation and household income. The second application considers the relationship between health and wealth outcomes for a sample of 30–60 year old individuals. Much attention has been devoted to determining the causal direction of this relationship, however, many studies have concluded that the relationship is reciprocal (Luo and Waite, 2005; Currie and Madrian, 1999), making simultaneous equation modeling appropriate for this context.

The data set for the first application consists of variables for wife labor force status and family financial stability for a sample of 2,920 families from the 1999 PSID survey. The model we consider is that in equations (3.3) and (3.2) where $y_{i1} = 1$ if the wife is currently in the labor force – employed or currently looking for a job – and is zero otherwise; $y_{i2} = 1$ if the family is financially stable and zero otherwise. Financial stability is defined as an income-to-needs ratio greater than 2. In the sample, nearly 74% of married women are in the labor force, and 32% of families encounter financial difficulties. The covariates x_{i1} that enter the labor force participation equation include the wife's age, the presence of a young child (child under 5), and the wife's years of education. The covariates x_{i2} that enter the financial stability equation include husband's age, race, education, employment, and number of children. Whereas the covariates follow closely the specification in Blundell and Smith (1994), the main distinguishing feature of our specification is that it includes two discrete outcome variables, unlike Blundell and Smith (1994) where income is measured as a continuous variable and only labor force participation is discrete.

Table 3.2: ARMH and MLE results. Posterior means and standard deviations are based on 5,000 MCMC draws.

Variable	**Eq. 1: Wife Employment**		**Eq. 2: Financially Stable**	
	ARMH	MLE	ARMH	MLE
Intercept	-1.990 (0.354)	-1.952 (0.337)	-5.292 (0.975)	-5.242 (0.908)
Age (f)	-0.010 (0.173)	-0.015 (0.147)	0.442 (0.438)	0.457 (0.398)
Educ (f)	2.443 (0.325)	2.413 (0.334)		
Educ (m)			5.129 (0.757)	5.103 (0.731)
Child under 5	-0.260 (0.004)	-0.268 (0.044)		
No. Child			-2.870 (0.291)	-2.919 (0.334)
White (f)	-0.217 (0.064)	-0.218 (0.065)		
White (m)			0.955 (0.154)	0.930 (0.170)
Emp (m)			0.936 (0.166)	0.917 (0.162)
Emp (f) (y_1)			2.651 (0.253)	2.573 (0.368)
Income (y_2)	1.220 (0.180)	1.224 (0.190)		
ρ_{ARMH}		0.407 (0.073)		
ρ_{MLE}		0.400 (0.075)		

Table 3.2 presents the MLE and ARMH results of the bivariate system. The results support the simultaneous relationship between married women's labor force participation and household financial situation as both endogenous variables are statistically different from zero. A married woman is more likely to participate in the labor force if her family is financially stable. Similarly, if a married woman is working in the labor force, it positively impacts her family's financial stability. The correlation between the two equations is about 0.40, implying a strong relationship between these two discrete outcomes.

The second application also employs data from the PSID. The data set consists of 4,405 survey respondents, where $y_{i1} = 1$ if a respondent reports better than "good" health and zero otherwise; $y_{i2} = 1$ if a respondent reports an income-to-needs ratio greater than 2 and zero otherwise. This represents wealth categories higher than the "middle-lower" class. The covariates included in each equation and the results are presented in Table 3.3.

Table 3.3: ARMH and MLE results. Posterior means and standard deviations are based on 5,000 MCMC draws.

Variable	**Eq. 1: Health**		**Eq. 2: Wealth**	
	ARMH	MLE	ARMH	MLE
Intercept	-2.453 (0.626)	-2.060 (0.485)	0.493 (0.050)	0.490 (0.053)
Fedu_HS			0.183 (0.043)	0.176 (0.043)
Fedu_Coll			0.154 (0.073)	0.145 (0.077)
Fedu_Grad			0.201 (0.102)	0.191 (0.106)
Medu_HS			0.251 (0.045)	0.254 (0.044)
Medu_Coll			0.420 (0.092)	0.407 (0.088)
Medu_Grad			0.248 (0.114)	0.261 (0.112)
Edu_HS	0.104 (0.066)	0.119 (0.062)	0.352 (0.043)	0.353 (0.047)
Edu_Coll	0.309 (0.062)	0.314 (0.071)	0.798 (0.066)	0.803 (0.071)
Edu_Grad	0.242 (0.076)	0.239 (0.085)	0.930 (0.104)	0.942 (0.102)
ChildHealth_GD	-1.255 (0.040)	-1.252 (0.045)		
ChildHealth_PR	-0.989 (0.112)	-0.976 (0.110)		
Race_Black	-0.009 (0.061)	-0.022 (0.057)	-0.443 (0.055)	0.430 (0.046)
Female	-0.031 (0.055)	-0.032 (0.053)	-0.070 (-0.068)	0.062 (0.045)
Single	0.360 (0.088)	0.343 (0.010)	-0.744 (0.069)	-0.739 (0.731)
Divorce	-0.146 (0.056)	-0.148 (0.059)	-0.596 (0.044)	-0.597 (0.046)
Single_Black	0.369 (0.181)	0.336 (0.217)	-0.417 (0.098)	-0.426 (0.106)
Health (y_1)			0.411 (0.155)	0.377 (0.145)
Wealth (y_2)	2.332 (0.639)	1.930 (0.499)		
ρ_{ARMH}		0.081 (0.029)		
ρ_{MLE}		0.080 (0.032)		

The results again display the importance of the simultaneous model. Both endogenous covariates are statistically different from zero and the correlation between the two equations is 0.08. These results accord well with the literature and intuition. Healthier individuals are able to be more productive in their work leading to increased wealth. Similarly, wealthy individuals have more access to health and preventative care, positively impacting their health.

These 2 illustrations display the importance of simultaneous equation models in discrete data settings. Previously, models for these applications would not be estimable because the values the parameters take violate the coherency condition. Our derivation of the likelihood function and estimation methods flexibly allow these applications to be considered and help us learn about two important relationships – female labor supply and family income, and health and wealth – which are of interest to many policy-makers.

3.6 Conclusion

This article provides a formulation of the likelihood function for simultaneous equations with discrete data, which employs Markov chain theory to cast the required distribution as the invariant distribution of a Markov chain. The framework is applied to simultaneous probit equations, but it extends nicely to other discrete data, e.g., ordinal, censored, etc. Additionally, the methods can be extended to higher dimensions with more discrete outcomes and endogenous covariates. Both classical and Bayesian estimation techniques are considered and both perform well in an artificial data setting and real data applications.

The approach presented in this paper overcomes several challenges that have long hindered these models. Namely, they required a coherency condition that put dubious constraints on the values the parameters can take, calling into question the sensibility of the model. The methodology presented here does not require a coherency condition and allows for these models to be estimated under minor identification restrictions. Moreover, we employ a conditional-conditional modeling approach which can be generalized to several frameworks, including random coefficients, hierarchical models, and state space models.

Bibliography

Aharony, J. and Swary, I. (1996), "Additional Evidence on the Information-based Contagion Effects of Bank Failures," *Journal of Banking and Finance*, 20, 57–69.

Albert, J. and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models with the Dependent Variables are Truncated Normal," *Econometrica*, 42, 999–1012.

Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalized Probit Model," *Econometrica*, 46, 1193–1205.

Bagehot, W. (1873), *Lombard Street: A Description of the Money Market*, London, NY: H.S. King.

Bajari, P., Hang, H., and Ryan, S. (2010), "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78, 1529–1568.

Becker, G. and Chiswich, B. (1966), "Education and the Distribution of Earnings," *American Economic Review*, 56, 358–369.

Bernanke, B. S. (1983), "Nonmonetary Effects of the Financial Crisis in Propagation of the Great Depression," *American Economic Review*, 73, 257–276.

Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889–918.

Blundell, R. and Smith, R. (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56, 37–57.

Blundell, R. and Smith, R. (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, 64, 355–373.

Bresnahan, T. and Reiss, P. (1991), "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57–81.

Butkiewicz, J. (1995), "The Impact of Lender of Last Resort during the Great Depression: The Case of the Reconstruction Finance Corporation," *Explorations in Economic History*, 32, 197–216.

Calomiris, C. (1993), "Financial Factors in the Great Depression," *Journal of Economic Perspectives*, 7, 61–85.

Calomiris, C. and Mason, J. (2003a), "Consequences of Bank Distress during the Great Depression," *American Economic Review*, 93, 937–947.

Calomiris, C. and Mason, J. (2003b), "Fundamentals, Panics, and Bank Distress during the Depression," *American Economic Review*, 93, 1615–1647.

Calomiris, C. and Wilson, B. (2004), "Bank Capital and Portfolio Management: The 1930s "Capital Crunch" and the Scramble to Shed Risk," *Journal of Business*, 77, 421–456.

Calomiris, C., Mason, J., Weidenmier, M., and Bobroff, K. (2013), "The Effects of the Reconstruction Finance Corporation Assistance on Michigan's Banks' Survival in the 1930s," *Explorations in Economic History*, 50, 525–547.

Chen, Y. (1999), "The Role of the First-Come, First-Served Rule and Information Externalities," *Journal of Political Economy*, 107, 946–968.

Chib, S. (1992), "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79–99.

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. (2001), "Markov Chain Monte Carlo: Computation and Inference," in *Handbook of Econometrics*, eds. J. Heckman and E. Leamer, vol. 5, pp. 3569–3649.

Chib, S. (2007), "Analysis of Treatment Response Data without the Joint Distribution of Potential Outcomes," *Journal of Econometrics*, 140, 401–412.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335.

Chib, S. and Jeliazkov, I. (2001), "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

Chib, S. and Jeliazkov, I. (2005), "Accept-Reject Metropolis-Hastings Sampling and Marginal Likelihood Estimation," *Statistica Neerlandica*, 59, 30–44.

Chib, S., Greenberg, E., and Jeliazkov, I. (2009), "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Ciliberto, F. and Tamer, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Currie, J. and Madrian, B. (1999), "Health, Health Insurance and the Labor Market," in *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card, vol. 3, pp. 3309–3416.

Diamond, D. and Dybvig, P. (1983), "Bank Runs, Deposit Insurance and Liquidity," *Journal of Political Economy*, 91, 401–419.

Freixas, X. and Rochet, J. C. (2008), *Microeconomics of Banking*, MIT Press, Cambridge, MA.

Gorton, G. (1985), "Bank Suspension of Convertibility," *Journal of Monetary Economics*, 15, 177–193.

Greenberg, E. (2008), *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1–22.

Gronau, R. (1974), "Wage Comparisons-A Selectivity Bias," *Journal of Political Economy*, 82, 1118–1143.

Haveman, R. and Wolfe, B. (1995), "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33, 1829–1878.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables," *Annals of Economic and Social Measurement*, 15, 475–492.

Heckman, J. (1978), "Dummy Endogeneous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Heckman, J. and Honoré, B. E. (1990), "The Empirical Context of the Roy Model," *Econometrica*, 58, 1121–1149.

Heckman, J. and Vytlacil, E. (2007), "Econometric Evaluation of Social Programs, Part I: Causal Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, eds. J. Heckman and E. Leamer, 6B, pp. 4779–4874, Elsevier.

Jeliazkov, I. and Lee, E. H. (2010), "MCMC Perspectives on Simulated Likelihood Estimation," *Advances in Econometrics*, 26, 3–39.

Jeliazkov, I. and Rahman, M. A. (2012), "Binary and Ordinal Data Analysis in Economics: Modeling and Estimation," in *Mathematical Modelling with Multidisciplinary Applications*, ed. X.-S. Yang, pp. 123–150, Wiley.

Jeliazkov, I., Graves, J., and Kutzbach, M. (2008), "Fitting and Comparison of Models for Multivariate Ordinal Outcomes," *Advances in Econometrics*, 23, 115–156.

Jones, J. H. (1951), *Fifty Billion Dollars: My Thirteen Years with the RFC, 1932-1945*, New York: Macmillan Co.

Judge, G. C., Griffiths, W., Hill, R. C., Lutkepohl, H., and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, Wiley Series in Probability and Mathematical Statistics., New York, 2 edn.

Kaufman, G. G. (1994), "Banking Contagion: A Review of the Theory and Evidence," *Journal of Financial Services Research*, pp. 123–150.

Kroszner, R. (1994), "The Political Economy of the Reconstruction Finance Corporation's Bailout of the U.S. Banking System During the Great Depression," *Mimeo.*

Li, M. and Tobias, J. (2014), "Bayesian Analysis of Treatment Effect Models," in *Bayesian Inference in the Social Sciences*, eds. I. Jeliazkov and X.-S. Yang, pp. 63–90, Wiley.

Li, P. (2011), "Estimation of Sample Selection Models with Two Selection Mechanisms," *Computational Statistics and Data Analysis*, 55, 1099–1108.

Liu, J. S. (1994), "The Collapsed Gibbs Sampler in Bayesian Computations With Applications to a Gene Regulation Problem," *Journal of the American Statistical Association*, 89, 958–966.

Liu, J. S., Wong, W. H., and Kong, A. (1994), "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27–40.

Luo, Y. and Waite, L. J. (2005), "The Impact of Childhood and Adult SES on Physical, Mental, and Cognitive Well-Being in Later Life," *Journals of Gerontology: Series B*, 60, 93–101.

Maddala, G. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Mason, J. (2001), "Do Lender of Last Resort Policies Matter? The Effects of Reconstruction Finance Corporation Assistance to Banks During the Great Depression," *Journal of Financial Services Research*, 20, 77–95.

Mason, J. (2003), "The Political Economy of Reconstruction Finance Corporation Assistance During the Great Depression," *Explorations in Economic History*, 40, 101–121.

Meyn, S. P. and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

Mintz, O., Currim, I. S., and Jeliazkov, I. (2013), "Information Processing Pattern and Propensity to Buy: An Investigation of Online Point-of-Purchase Behavior," *Marketing Science*, 32, 716–732.

Mitchener, K. (2005), "Bank Supervision, Regulation, and Instability During the Great Depression," *Journal of Economic History*, 65, 152–185.

Muthen, B. O. (2002), "Beyond SEM: General Latent Variable Modeling," *Behaviormetrika*, 29, 81–117.

Narayanan, S. (2013), "Bayesian Estimation of Discrete Games of Complete Information," *Quantitative Marketing and Economics*, 11, 39–81.

Nelson, F. and Olson, L. (1978), "Specification and Estimation of Simultaneous-Equation Model with Limited Dependent Variables," *International Economic Review*, 19, 695–709.

Park, S. (1991), "Bank Failure Contagion in Historical Perspective," *Journal of Monetary Economics*, 28, 271–286.

Poirier, D. and Tobias, J. (2003), "On the Predictive Distributions of Outcome Gains in the Presence of an Unidentified Parameter," *Journal of Business and Economic Statistics*, 21, 258–268.

Richardson, G. (2007), "Categories and Causes of Bank Distress During the Great Depression, 1929-1933: The illiquidity Versus Insolvency Debate Revisited," *Explorations in Economic History*, 44, 588–607.

Richardson, G. and Troost, W. (2009), "Monetary Intervention Mitigated Banking Panics During the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Border 1929-1933," *Journal of Political Economy*, 117, 1031–1073.

Rosenbloom, J. and Sundstrom, W. (1999), "The Sources of Regional Variation in the Severity of the Great Depression: Evidence from U.S. Manufacturing, 1919-37," *Journal of Economic History*, 59, 714–747.

Roy, A. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58.

Sobel, M. E. and Arminger, G. (1992), "Modeling Household Fertility Decisions: A Nonlinear Simultaneous Probit Model," *Journal of the American Statistical Association*, 87, 38–47.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701–1761.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.

Trivedi, P. K. and Zimmer, D. M. (2005), "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, 1, 1–111.

van Hasselt, M. (2011), "Bayesian Inference in a Sample Selection Model," *Journal of Econometrics*, 165, 221–232.

van Hasselt, M. (2014), "Bayesian Analysis of Sample Selection Models," in *Bayesian Inference in the Social Sciences*, eds. I. Jeliazkov and X.-S. Yang, pp. 91–113, Wiley.

van Wissen, L. and Golob, T. (1990), "Simultaneous equation systems involving binary choice variables," *Geographical Analysis*, 22, 224–243.

Vella, F. (1993), "A Simple Estimator for Simultaneous Models with Censored Endogeneous Regressors," *International Economic Review*, 34, 441–457.

Vijverberg, W. (1993), "Measuring the Unidentified Parameter on the Extended Roy Model of Selectivity," *Journal of Econometrics*, 57, 69–89.

Vossmeyer, A. (2014), "Treatment Effects and Informative Missingness with an Application to Bank Recapitalization Programs," *American Economic Review*, 104, 212–217.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

Yen, S. T. (2005), "A Multivariate Sample-Selection model: Estimating Cigarette and Alcohol Demands with Zero Observations," *American Journal of Agricultural Economics*, 87, 453–466.