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

Angela Vossmeyer
University of California, Irvine*

February 13, 2014

Abstract

This article develops a Bayesian framework for estimating multivariate treatment effect models in the presence of sample selection. The methodology is applied to a banking study that evaluates the effectiveness of bank recapitalization programs and their ability to resuscitate the financial system. The analysis of lender of last resort (LOLR) policies presents several challenges because regulator data is generally not publicly available and modeling involves a difficult decision structure which is at the intersection of treatment effect models, sample selection, endogeneity, and discrete data modeling. Motivated by these difficulties, this paper implements the new methodology and employs a novel bank-level data set to jointly model a bank's decision to apply for a loan, the LOLR's decision to approve or decline the loan, and the bank's performance a few years after the disbursements. This analysis addresses questions regarding LOLR regulation including whether and to what extent these programs stabilize the economy or simply privatize the gains and nationalize the loses. Overall, the methodology offers practical estimation tools to unveil new answers to important regulatory and policy questions.

Keywords: Bayesian estimation; Markov chain Monte Carlo (MCMC); Sample selection; Treatment responses; Data augmentation; Bank recapitalization; Lender of last resort.

1 Introduction

Models for sample selection and treatment effects represent two of the greatest challenges the field of economics, and even more broadly the social sciences, have overcome. Difficulties include dealing with non-random missing data, and non-random treatment assignment. Unlike the other sciences, economists need to model these features of the data because controlled experiments or randomized trials are often not available. The importance of these models to the field is immeasurable and this

*Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100; e-mail: angela.vossmeyer@uci.edu. The author is greatly indebted to Ivan Jeliazkov for his insightful guidance. The author also thanks Gary Richardson, Dale Poirier, David Brownstone, Jiawei Chen, Sean Dowsing, Jennifer Muz and Timothy Wong for their helpful comments. Funding is acknowledged from the All-UC Group in Economic History.

paper adds to the literature on this methodology by extending conventional treatment models to allow for non-random missing data.

The focus of this paper is not only on a new methodology but also on an empirical application. The methodological component of this paper develops a multivariate treatment effect model that allows for sample selection, endogeneity and discrete outcomes. This framework is applied to a study that evaluates the effectiveness of bank recapitalization programs and their ability to resuscitate the financial system, which is the empirical component of this paper. Banking regulations, as with many policies and programs that are the foci of economic research, operate under selection mechanisms or decision structures that must be accommodated by econometric models to ensure proper inference. However, the complexities involved in modeling selection equations and treatment response data restrict attention to single equation models. The dangers of improper modeling are bias and misrepresentation of the population of interest. These issues are prevalent in the analysis of lender of last resort (LOLR) policies, which underlies the application in this paper.

Existing treatment models consider two subgroups in the data, the treated group and the control, or untreated, group. This formulation, when applied to an LOLR study, divides the sample into banks that receive loans from the LOLR and banks that do not receive loans. Complications arise because an initial selection mechanism, the application step of the recapitalization process, is ignored. Avoiding the question of whether or not the bank applied for assistance from the LOLR 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 for non-randomly selected data to offer a more complete framework for evaluating the impact of LOLR programs. It should be noted that other literatures face these obstacles as well, e.g., labor supply decisions, college admittance studies, credit approval decisions, health outcomes and drug treatment research, welfare or housing program evaluation, and many others. Any scenario that consists of an application step and an approval step can utilize this model and estimation strategy.

Sample selection (also known as incidental truncation or informative missingness) arises when a dependent 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. The selected sample is defined as a subset of observational units in which all variables are observed. The non-selected sample is a subset that has missing outcomes due to incidental truncation, and the potential sample is all observational units that can be observed. Ignoring sample selection causes a researcher to base estimation on a sample that does not represent the population of interest, which leads to specification errors. The basic structure for sample selection models is featured in Figure 1.

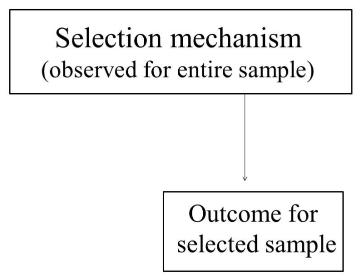

Figure 1: Standard sample selection model.

Classical estimation methods of sample selection models are developed in Gronau (1974) and Heckman (1976, 1979), and are further discussed in Wooldridge (2002). Gronau's model was motivated by incidences of wage offers and labor force participation. His interest was in estimating the expected hourly wage given a set of covariates for a randomly drawn individual. However, wages are not observed for everyone in the working age population. The sample selection problem arises because wages are only observed for people in the labor force. Heckman pointed out that the presence of selection bias can be viewed as an omitted variable problem and devised a two-step estimation procedure for the sample selection model. Bayesian developments in these models can be found in Chib et al. (2009), Greenberg (2008) and Van Hasselt (2011). The framework and estimation strategy for multiple selection mechanisms are discussed in Li (2011) and Yen (2005).

Treatment models, also referred to as models of potential outcomes or the Roy model (Roy, 1951; Heckman and Honoré, 1990), are employed to compare responses of individuals who belong either to a treatment or a control group. These models feature two potential responses; however, only one is ever observed, the other is the counterfactual. In most economic applications, observations are non-randomly assigned to the treatment or control groups which leads to correlation between the treatment assignment and the response variable of interest. Hence, selection into these groups must be accounted for in an econometric model. The structure for treatment models is featured in Figure 2.

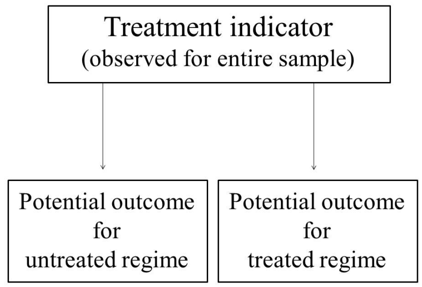

Figure 2: Standard potential outcomes treatment model.

Estimation techniques for treatment models featured in Lee (1978) and Amemiya (1985) use a modified least squares procedure that makes use of Inverse Mills Ratios to get around difficulties involved with evaluating complicated likelihood and distribution functions. Bayesian approaches motivated by the missingness of the counterfactual in Vijverberg (1993) and Poirier and Tobias (2003) formulate an analysis for this model with the joint distribution of the potential outcomes. Chib (2007), on the other hand, provides a Bayesian analysis without the involvement of a joint model of the potential outcomes. Causal inference without the inclusion of the counterfactual was initially developed in Dawid (2000). Modeling the joint distribution involves placing a prior on the non-identified elements in the variance-covariance matrix, and simulating a posterior of the parameters and the counterfactuals. 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.

Although models for sample selection and treatment effects are used frequently on their own, techniques that incorporate and jointly model both are lacking in the literature. To address this deficiency, this paper employs a Bayesian framework for treatment effect modeling while dealing with the missing data that occur from sample selection. This structure can be seen in Figure 3, which presents the multi-step data generation process leading to the observed data. This framework can be recognized as the decision structure employed in many programs consisting of an application step and an approval step. The initial selection mechanism is observed for the entire sample. The selected sample then enters a selected treatment stage followed by a set of potential

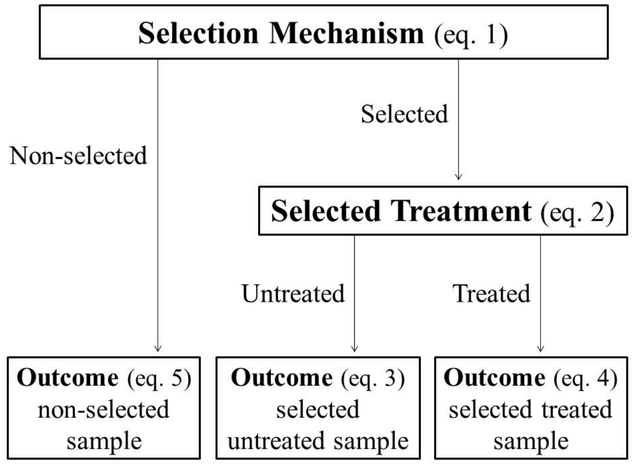

Figure 3: Treatment model with incidental truncation. The bolding represents selection equations and treatment response equations.

outcomes or treatment responses. The model considered here differs from the existing literature 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 revealing features of the data. As mentioned earlier, in the bank recapitalization study, the initial selection mechanism, deciding whether or not to apply for a loan, is a very affirming step in the recapitalization process. Recognizing and modeling the decisions for the non-selected or non-applicant group allows for a broader understanding of the outcomes of interest.

The rest of the paper is organized as follows: Section 2 presents the multivariate treatment effect model with sample selection, Section 3 presents estimation methods to fit the model, and Section 4 reports the performance of these techniques in a simulation study. The new methodology is applied to a study of bank recapitalization in Section 5. Section 6 contains additional considerations including prior sensitivity, model comparison and treatment effects, and finally, Section 7 offers concluding remarks.

2 The Model

The model stemming from Figure 3 contains 5 equations of interest – 1 selection mechanism, 1 selected treatment and 3 treatment response outcomes for the different subsets of the sample: the non-selected sample, the selected untreated sample, and the selected treated sample. It is important to note that this system is not limited to 5 equations. The new methodology is easily extendable to larger systems of equations as well as endogenous regressors. If another equation observed for the entire sample and correlated with the errors of the selection mechanisms or treatment responses presented itself, this model could properly incorporate that endogeneity by allowing it to enter as a covariate in the other equations. In detail, below are the equations for subjects $i = 1, \ldots, n$:

$$\begin{aligned}
\textit{Selection Equation} &: \quad y_{i1}^* = \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} \quad (\textit{always observed}) &(1)\\
\textit{Selected Treatment} &: \quad y_{i2}^* = \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2} \quad (\textit{observed for selected sample}) &(2)\\
\textit{OUTCOMES} &: \quad \text{Treatment Responses } (\textit{only one is observed}) &\\
\textit{For Selected Untreated Sample} &: \quad y_{i3}^* = (\mathbf{x}_{i3}' \; y_{i1})\boldsymbol{\beta}_3 + \varepsilon_{i3} &(3)\\
\textit{For Selected Treated Sample} &: \quad y_{i4}^* = (\mathbf{x}_{i4}' \; y_{i1} \; y_{i2})\boldsymbol{\beta}_4 + \varepsilon_{i4} &(4)\\
\textit{For Non-Selected Sample} &: \quad y_{i5}^* = \mathbf{x}_{i5}'\boldsymbol{\beta}_5 + \varepsilon_{i5} &(5)
\end{aligned}$$

The model is characterized by 5 dependent variables of interest where $\mathbf{y}_i^* = (y_{i1}^*, y_{i2}^*, y_{i3}^*, y_{i4}^*, y_{i5}^*)$ are the continuous latent data and $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5})$ are the corresponding discrete observed data. In the application, the latent variables relate to the observed censored outcomes by $y_{i1} = y_{i1}^* I(y_{i1}^* > 0)$, $y_{i2} = y_{i2}^* I(y_{i2}^* > 0)$, $y_{i3} = y_{i3}^* I(y_{i3}^* > 0)$, $y_{i4} = y_{i4}^* I(y_{i4}^* > 0)$, and $y_{i5} = y_{i5}^* I(y_{i5}^* > 0)$ (Tobin, 1958), which is the basis for the model throughout the paper. However, the general system can take outcome variables that are continuous, binary, censored or ordered. The continuous setting occurs when $y_{ij}^* = y_{ij}$ for equations $j = 1, \ldots, 5$. For the binary setting, $y_{ij} = I(y_{ij}^* > 0)$, and for the ordered setting $y_{ij} = \sum_{h=1}^{H} 1\{y_{ij}^* > \delta_{h-1}\}$ for H ordered alternatives where δ_h is a cutpoint between the categories.

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 so 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 so 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 so 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 covariates in $\mathbf{x}_2$ 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). 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).

2.1 The Likelihood Function

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

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

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

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. N_1 is the set of indices for which y_{i1} and y_{i5} are observed, N_2 is the set for which y_{i1}, y_{i2} and y_{i3} are observed, and N_3 is the set for which y_{i1}, y_{i2} and y_{i4} are observed. 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

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

note that in $\mathbf{\Omega}$, the elements Ω_{35}, Ω_{45},and Ω_{34} are not identified because their corresponding equations cannot be observed at the same time, and Ω_{25} is missing due to the incidental truncation. Thus, there are 11 unique elements in $\mathbf{\Omega}$ that can be estimated, 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 vectors and matrices that correspond to the 3 different subsets of the sample, define

$$\mathbf{J}_C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{2\times K}, \; \mathbf{J}_D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}_{3\times K} \text{ and } \mathbf{J}_A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{3\times K}$$

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

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

$$\boldsymbol{\Omega}_C = \begin{pmatrix} \Omega_{11} & \Omega_{15} \\ \Omega_{51} & \Omega_{55} \end{pmatrix}, \;\; \boldsymbol{\Omega}_D = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}, \;\; \boldsymbol{\Omega}_A = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{24} \\ \Omega_{41} & \Omega_{42} & \Omega_{44} \end{pmatrix}. \tag{6}$$

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

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

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

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

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

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

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

$$\begin{aligned} \prod_{i\in N_1} f(\boldsymbol{\eta}_{iC}^*|\boldsymbol{\theta}) \prod_{i\in N_2} f(\boldsymbol{\eta}_{iD}^*|\boldsymbol{\theta}) \prod_{i\in N_3} f(\boldsymbol{\eta}_{iA}^*|\boldsymbol{\theta}) \;\; \propto \;\; & |\boldsymbol{\Omega}_C|^{-n_1/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_1} \boldsymbol{\eta}_{iC}^{*\prime}\boldsymbol{\Omega}_C^{-1}\boldsymbol{\eta}_{iC}^*\right\} \\ & \times|\boldsymbol{\Omega}_D|^{-n_2/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_2} \boldsymbol{\eta}_{iD}^{*\prime}\boldsymbol{\Omega}_D^{-1}\boldsymbol{\eta}_{iD}^*\right\} \\ & \times|\boldsymbol{\Omega}_A|^{-n_3/2} \exp\left\{-\frac{1}{2}\sum_{i\in N_3} \boldsymbol{\eta}_{iA}^{*\prime}\boldsymbol{\Omega}_A^{-1}\boldsymbol{\eta}_{iA}^*\right\}. \end{aligned} \tag{7}$$

The decomposition of the complete-data likelihood into the 3 subsets of the sample is the basis for the computationally efficient estimation algorithm.

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 inverted Wishart distribution with parameters υ and $\mathbf{Q}$,

$$\pi(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \mathcal{N}(\boldsymbol{\beta}\ |\boldsymbol{\beta}_0, \mathbf{B}_0)\mathcal{IW}(\boldsymbol{\Omega}|\upsilon, \mathbf{Q}). \tag{8}$$

3 Estimation

Combining the complete-data likelihood in (7) and priors in (8) leads to a posterior distribution for $\boldsymbol{\theta}$ and $\mathbf{y}^*$,

$$\pi(\boldsymbol{\theta}, \boldsymbol{y^*}|\boldsymbol{y}) \propto \left[\prod_{i\in N_1} f(\mathbf{y}^*_{iC}|\boldsymbol{\theta})\right]\left[\prod_{i\in N_2} f(\mathbf{y}^*_{iD}|\boldsymbol{\theta})\right]\left[\prod_{i\in N_3} f(\mathbf{y}^*_{iA}|\boldsymbol{\theta})\right] \times \mathcal{N}(\boldsymbol{\beta}\ |\boldsymbol{\beta}_0, \mathbf{B}_0) \times \mathcal{IW}(\boldsymbol{\Omega}|\upsilon, \mathbf{Q}), \tag{9}$$

which is simulated by Markov chain Monte Carlo (MCMC) methods. The discreteness of multiple outcome variables render this likelihood analytically intractable and hence estimation relies on simulation-based techniques. The sampling algorithm is summarized as follows:

Algorithm 1 *MCMC Estimation Algorithm for Censored Outcomes*

1. Sample $\boldsymbol{\beta}$ from the distribution $\boldsymbol{\beta}\ |\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\beta}$.[1]
2. Sample $\boldsymbol{\Omega}$ from the distribution $\boldsymbol{\Omega}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\Omega}$ in a one block, multi-step procedure.
3. For $i \in N_1,$ sample $\mathbf{y}_1^*$ from the distribution $\mathbf{y}_1^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_1^*$.
4. For $i \in N_2,$ sample $\mathbf{y}_2^*$ from the distribution $\mathbf{y}_2^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_2^*$.
5. For $i \in N_{2o},$ sample $\mathbf{y}_3^*$ from the distribution $\mathbf{y}_3^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_3^*$.
6. For $i \in N_{3o},$ sample $\mathbf{y}_4^*$ from the distribution $\mathbf{y}_4^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_4^*$.
7. For $i \in N_{1o},$ sample $\mathbf{y}_5^*$ from the distribution $\mathbf{y}_5^*|\mathbf{y}, \boldsymbol{\theta}, \mathbf{y}^* \setminus \mathbf{y}_5^*$.

N_{2o} is defined as the truncated portion of the N_2 sample in $\mathbf{y}_3,$ N_{3o} is defined as the truncated region of the N_3 sample in $\mathbf{y}_4,$ and N_{1o} is defined as the discrete part of the N_1 sample in $\mathbf{y}_5$ since all the equations are Tobit equations with censored dependent variables. Following Chib et al. (2009), it is important to note that the algorithm does not augment the selected sample with data

[1] The notation "\" represents "except", e.g., $\mathbf{y}^* \setminus \mathbf{y}_1^*$ says all elements in $\mathbf{y}^*$ except $\mathbf{y}_1^*$.

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, simulating the posterior is of the parameters and not the counterfactuals to simplify the prior inputs and improve the mixing of the Markov chain (Chib, 2007). Details of the sampler are in the following subsections.

3.1 Sampling $\boldsymbol{\beta}$

The posterior distribution displayed in (9) implies $\boldsymbol{\beta}\,|\mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}, \mathbf{B})$, where

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

Computations for $\boldsymbol{\beta}$ are done efficiently by updating observations in the N_1, N_2 and N_3 subsamples separately. Thereby, isolating the observed parts of the model and proceeding without computations for the unobserved components.

3.2 Sampling $\boldsymbol{\Omega}$

The sampling of $\boldsymbol{\Omega}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \boldsymbol{\Omega}$ is complicated because the form of $\boldsymbol{\Omega}$ in the complete-data likelihood given in (6) and (7) is different for observations in N_1, N_2 and N_3. Thus, sampling must be done in multiple steps, or layers, for the different subsets of the sample, as opposed to a single step. Utilizing the sampling techniques for $\boldsymbol{\Omega}$ in Chib et al. (2009), this paper initially samples from $\Omega_{11}|\mathbf{y}, \mathbf{y}^*, \boldsymbol{\theta} \setminus \Omega_{11}$ because it is observed for the entire sample, followed by sampling from a set of marginal and conditional distributions.

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

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

the posterior full-conditional distribution of $\mathbf{\Omega}$ is defined as

$$\begin{aligned}\pi(\mathbf{\Omega}|\mathbf{y},\mathbf{y}^*,\boldsymbol{\theta}\setminus\mathbf{\Omega}) &\propto |\mathbf{\Omega}_C|^{-n_1/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Omega}_C^{-1}\sum_{i\in N_1}\boldsymbol{\eta}_{iC}^*\boldsymbol{\eta}_{iC}^{*'}\right)\right]\\ &\times|\mathbf{\Omega}_D|^{-n_2/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Omega}_D^{-1}\sum_{i\in N_2}\boldsymbol{\eta}_{iD}^*\boldsymbol{\eta}_{iD}^{*'}\right)\right]\\ &\times|\mathbf{\Omega}_A|^{-n_3/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Omega}_A^{-1}\sum_{i\in N_3}\boldsymbol{\eta}_{iA}^*\boldsymbol{\eta}_{iA}^{*'}\right)\right].\end{aligned}$$

To sample from the set of conditional and marginal distributions, a change of variable technique is employed and the resulting density is proportional to a set of inverse Wisharts and a matric-variate normals. Thus, $\mathbf{\Omega}$ updates in a 1-block, multi-step procedure which samples in layers and conditions only on the identified parts of the model. These techniques are also 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, sample $\Omega|\mathbf{y},\mathbf{y}^*,\boldsymbol{\theta}\setminus\mathbf{\Omega}$ in a one block, nine-step procedure by first drawing Ω_{11} , $\Omega_{tt\cdot l}=\Omega_{tt}-\Omega_{tl}\Omega_{ll}^{-1}\Omega_{lt}$, and $B_{lt}=\Omega_{ll}^{-1}\Omega_{lt}$, and then reconstructing $\mathbf{\Omega}$ from these quantities

2. (a) $\Omega_{11}|\mathbf{y},\mathbf{y}^*,\boldsymbol{\theta}\setminus\mathbf{\Omega}\sim\mathcal{IW}(\nu-1+n,\ \mathbf{Q}_{11}+\sum_{N_1,N_2,N_3}\boldsymbol{\eta}_{i1}^*\boldsymbol{\eta}_{i1}^{*'})$
      i. $\boldsymbol{\eta}_{i1}^*=y_{i1}^*-\mathbf{x}_{i1}\mathbf{J}_1\boldsymbol{\beta}$, where $\mathbf{J}_1=\begin{bmatrix}1 & 0 & 0 & 0 & 0\end{bmatrix}_{1\times K}$
   (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 $\mathbf{R}_{ll\cdot t}=\mathbf{R}_{ll}-\mathbf{R}_{lt}\mathbf{R}_{tt}^{-1}\mathbf{R}_{tl}$. From these sampling densities, $\mathbf{\Omega}$ can be recovered.

3.3 Sampling $\mathbf{y}^*$

$\mathbf{y}^*$ is sampled following Chib (1992) from a truncated normal,

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

4 Simulation Study

This section evaluates the performance of the algorithm from Section 3 using simulated data. The model being considered is the system of 5 equations from (1) – (5) which is motivated by the subsequent application to bank recapitalization considered in Section 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 forth (including $\mathbf{y}_1$ and $\mathbf{y}_2$) and 4 in the fifth equation, so $\boldsymbol{\beta}$ is a 24×1 vector. Similarly, $vech(\boldsymbol{\Omega})$ has 11 unique estimable elements. Variants and generalizations to more variables do not require conceptual changes to the estimation algorithm.

For the study, $n = 2000$ observations with 36% of the sample in the N_1 subset, 16% in the N_2 subset, and 48% in the N_3 subset. The percent of the sample that is censored in each equation is: 36%, 25%, 25%, 1%, and 6%. Note that various combinations of 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. The priors set for the equations are: $\boldsymbol{\beta} \sim \mathcal{N}(0, 5 \times I)$ and $\boldsymbol{\Omega} \sim \mathcal{IW}(9, 1.2 \times I_5)$. Posterior mean estimates are based on 10,000 MCMC draws with a burn-in of 1,000. The total time of the algorithm is about 1 minute and 8 seconds. A major benefit of this sampler is the computational speed and low storage cost.

The true values are uncovered well and quickly. Further evaluations of the performance 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 $1+2\sum_{l=1}^{L}\rho_k(l)(1-\frac{1}{l})$, where $\rho_k(l)$ is the sample autocorrelation at the l-th lag for the k-th parameter and L is the lag in which the autocorrelations taper off (Chib et al., 2009). Li (2011) states that this "quantity measures the efficiency loss when using correlated MCMC samples instead of independent samples." Small values (near 1) imply that the output is mixing well.

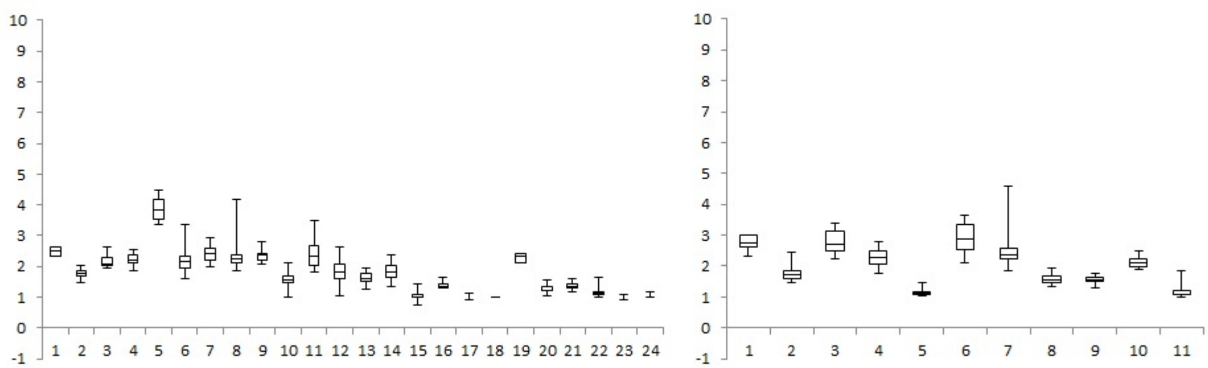

Figure 4: Boxplots of inefficiency factors for $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$. The left panel is $\boldsymbol{\beta}$, the right panel is $vech(\boldsymbol{\Omega})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \boldsymbol{\beta}_4', \boldsymbol{\beta}_5')$, $vech(\boldsymbol{\Omega}) = (\Omega_{11}, \Omega_{21}, \Omega_{22}, \Omega_{31}, \Omega_{32}, \Omega_{33}, \Omega_{41}, \Omega_{42}, \Omega_{44}, \Omega_{51}, \Omega_{55})'$.

Boxplots of the inefficiency factors from the algorithm are displayed in Figure 4. The plots suggest these parameters are sampled efficiently. The inefficiency factors for all 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 frequently. Given that every dependent variable in this system is discrete, the inefficiency plots are promising. The lack of augmentation in the sampler, which arises by not simulating the outcomes that are missing due to the selection mechanism or that are not identified 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 outcomes that are not identified, and the results show excellent sampler performance.

5 Application

This section applies the multivariate treatment effect model with sample selection to a study of LOLR polices and bank recapitalization. Financial relief programs and the central bank's LOLR function emphasize a trade-off between liberal and strict lending policies. Research in favor of recapitalization 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 LOLR policies 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; Mishkin, 2006). These issues have been deliberated since the concept of LOLR was proposed 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. The RFC is one of the largest recapitalization programs ever implemented. It was established during the Hoover administration with the primary objective of providing liquidity to, and restoring confidence in the banking system. The RFC was created in January 1932 and later became part of the New Deal under Roosevelt's administration. Activities of the RFC did not change the monetary base as the RFC had no power to print or create money. Instead, financing for RFC loans was provided from the Treasury. For a thorough explanation of the RFC and its operations, see Mason (2003).

The empirical analysis of LOLR programs poses a number of challenges because regulator data is generally not publicly available, and modeling must accommodate a difficult decision structure, endogeneity of policies, correlation between policies and outcomes, and non-random selection into policies and programs. This paper adds to the existing literature by employing a novel bank-level data set and developing a new methodology to jointly model a bank's decision to apply for a loan from the LOLR, the LOLR's decision to approve the loan, and the bank's performance a few years after the disbursements. This modeling structure can be seen in Figure 5 (which closely resembles Figure 3).

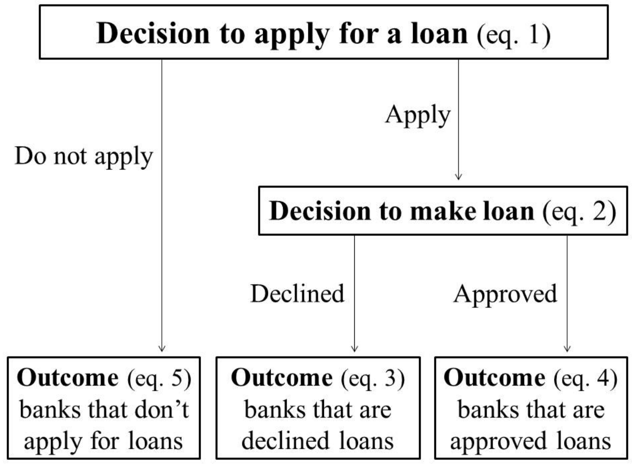

Figure 5: Process of applying for and receiving assistance from the LOLR.

Figure 5 represents a system of 5 equations as in (1) – (5). The initial selection mechanism (equation 1), is observed for every bank, and represents a bank's decision to apply for assistance from the RFC. Banks that apply for loans, the selected sample, enter the selected treatment stage, while those that do not apply, the non-selected sample, are not observed in the next equation. The selected treatment stage (equation 2) represents the RFC's decision to approve or decline the submitted loan application. Following these 2 equations are 3 potential outcomes or treatment responses (equations 3–5). The treatment responses represent bank performance reported a few years after the loans were disbursed. 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 (non-selected, selected untreated, selected treated) is important because selection into these groups is non-random, so it allows for the coefficients on the estimated parameters and the error variance to differ across equations.

To reiterate from the introduction, extending standard treatment models to allow for an initial selection mechanism is immensely important to the analysis of LOLR policies. Modeling whether a bank is selecting into or opting out of treatment is revealing of a bank's health. Ignoring this step in the recapitalization process groups declined banks with non-applicant banks in an untreated sample, which misrepresents the population of interest leading to a fundamental misspecification.

5.1 Data

This paper employs two novel 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 in College Park, Maryland. The cards report the name and address of borrower, date, request and amount of loan, and whether the loan was approved or declined. Further information is obtained from the "Paid Loan Files" and "Declined Loan Files" which include the exact information the regulators had on each bank and the original examiner's report on each decision. This data set is merged with a separate data set constructed from the *Rand McNally Banker's Directory*. This directory describes balance sheets, correspondent relationships and characteristics for all banks in a given state for a given year. This information identifies the non-selected, or non-applicant sample. In addition, the *Rand McNally Banker's Directory* indicates whether or not a bank is a member of the American Banking Association (ABA) and what departments are at each bank (e.g., bond, safe deposit, trust and savings). Additional data are gathered from the 1930 U.S. census of agriculture, manufacturing and population which describe 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, y_{i1}, is the total amount of RFC assistance requested by December 1933.[2] 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 2, y_{i2}, is the total amount of RFC money approved.[3] 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 RFC's decision to lend depended on whether or not they thought the bank was solvent. The economic state of the bank was the sole determinant of loan decisions, however, some of the existing literature recognizes political influence in the decision-making process. Kroszner (1994) argues that the RFC was politically biased, and the executive branch had a significant impact on

[2]The RFC started in early 1932, so by the end of 1933 some banks had submitted multiple applications. Thus, this outcome is the sum of all RFC loan requests for each bank.

[3]The data do not speak for banks that received both declined and approved loans. In this sample, if a bank had multiple applications, it had the same decision from the RFC. In addition, most subsequent approvals were for loan renewals.

the distribution of the RFC's funds between the years of 1933 – 1935. On the contrary, Mason (2003) argues that notwithstanding political pressure, "there is scant empirical evidence to suggest that RFC assistance was distributed on the basis of anything more than economic conditions." Thus, the covariates that enter equation 2 include information from the RFC loan applications and political indicators.

Finally, the outcome variable for equations 3 – 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 years allowed banks to utilize their relief funds. Later years are not considered because the FDIC was established in 1934 and its operations increased greatly in 1935, so the effects of the RFC become blended with that of the FDIC after 1935. The outcome for equations 3 – 5 is again censored with point mass at zero for banks that failed since the time of the loan applications 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).

5.1.1 Descriptive Statistics

The data set includes all banks operating in 1932 in Alabama, Arkansas, Mississippi, Michigan and Tennessee. The sample consists of 1,794 banks, of which 908 banks applied for RFC assistance and 800 of those were approved while 108 were declined assistance. Table 1 presents descriptive statistics on the RFC applications and approvals organized 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 $ Requested ($ millions)	21.8	18.3	178.5	22.7	57.9
Total RFC $ Approved ($ millions)	18.7	16.7	155.6	21.5	55.4

Table 1: RFC applications and approvals.

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. However, applications were often approved for less money. When a bank applied for an RFC loan, it had to make a collateral offering. Banks offered their illiquid assets as collateral in exchange for liquid assets. Differences in the requested and approved amounts of loans were likely the result of the collateral offering. If a bank was not willing to put up enough collateral, the RFC did not approve the full loan; they approved a lesser amount.[4]

Table 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 they are representative of the banking system and provide variation across bank characteristics, size, Federal Reserve districts and county characteristics. Federal Reserve district variation is necessary because RFC lending was concurrent with lending through the discount window. Federal Reserve policies differed across districts which 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 open market operations because the RFC could lend to non-Fed member banks, which is 81% of this sample. County characteristics, such as farm prices and manufacturing output, are important because previous research finds that they play a role in a bank's success or failure (Butkiewicz, 1995; Calomiris and Mason, 2003b). In addition, the RFC considered some county characteristics in its loan approval process which is apparent from the loan applications and the RFC examiners' reports.

Finally, in modeling the elements of financial panics, it is important to acknowledge contagion. Linkages for contagion are found in banks' correspondent networks during the Great Depression which are also reported in Table 2. Correspondent banks 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. These linkages created paths for contagion to spread, and as a result, their inclusion into the model is important.

[4] The RFC initially required banks to pledge their best assets as collateral for assistance. However, the RFC's lending power was expanded in the 1933 Emergency Banking Act allowing the RFC to purchase bank preferred stock to make loans. Although the RFC participated in both LOLR policies and bank recapitalization, this paper is looking at the aggregate effectiveness of the program, so no distinction is made.

Variable	**Alabama**	**Arkansas**	**Michigan**	**Mississippi**	**Tennessee**
No. Banks	250	278	638	235	393
Age (avg.)	24	22	30	25	25
Fed District(s)	6	8	7, 9	6, 8	6, 8
Financial Ratios (avg.)					
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	(avg. - $1000)				
Total Assets	1027	558	2309	695	1121
Loans and Discounts	582	273	1203	356	649
Bonds and Securities	228	131	611	187	178
Cash and Exchanges	138	121	280	110	177
Paid-Up Capital	106	53	157	55	101
Deposits	716	437	1750	528	765
Surplus and Profits	92	41	126	47	80
Charters, Memshps, Depts.	(counts)				
State Bank	166	222	438	208	308
Federal Reserve Mem.	82	44	102	26	83
ABA Member	160	188	370	171	185
State Bank Ass'n Mem.	215	252	542	219	374
Safe Deposit Dept.	123	121	451	94	145
Bond Dept.	23	25	113	29	35
Savings Dept.	74	126	557	152	189
Trust Dept.	63	47	118	52	133
Correspondents (avg.)					
Total Correspondents	2.6	2.4	2.8	2.9	2.4
Out of State Corres.	1.5	1.4	1.5	2.5	1
Market Shares (avg.)					
Liab. / County Liab.	0.27	0.26	0.13	0.33	0.24
Liab. / Town Liab.	0.68	0.75	0.17	0.74	0.69
HHI	0.66	0.60	0.29	0.70	0.54
County characteristics (avg.)					
No. Wholesale Retailers	31.3	22.4	45.3	15.4	25.3
% Voted Democratic	79.8	81.6	47.2	85.2	71.1
No. Manuf. Establish.	41.4	22.5	44.5	33.8	30.4
Cropland ($\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 2: Financial characteristics of the banks in each state in 1932 and county characteristics.

5.2 Results

Table 3 displays the results of the multivariate treatment effect model in the presence of sample selection. The following discussion reviews the results from each equation and Section 5.2.1 presents the results for the variance-covariance matrix. Sensitivity analysis, model comparison and treatment effects are considered in Section 6. The results presented in Table 3 are from the model with the highest marginal likelihood, the benchmark model. Marginal likelihood computations and discussions are in Section 6.2. The priors for $\boldsymbol{\beta}$ in the benchmark model are centered at 0 with a standard deviation of 2.24 and the priors for $\boldsymbol{\Omega}$ imply that $E(\boldsymbol{\Omega}) = .4 \times I$ and $SD(\text{diag}(\boldsymbol{\Omega})) = 0.57 \times I$. Section 6.1 reports the sensitivity of the results to the prior specification.

Column 2 of Table 3 presents the results for the application step of the recapitalization process, the first equation. 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 are more likely to apply for loans. These results are intuitive because a high cash to asset ratio implies a bank is liquid and does not need a loan. Whereas, a high deposit to liability ratio implies a bank is at risk for a run which was the main mechanism that caused bank failure in the 1930s (Bernanke, 1983) and was a credible threat to many depository institutions. A bank's vulnerability to a run is also apparent in the results for market shares. The higher share of liability at all banks in the county held at an individual bank, the more likely the bank is to apply for a loan. With contagious runs spreading through many counties, a bank that is responsible for a large share of depositors is more likely to protect itself from liquidation by applying for assistance from the RFC.

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 district were more likely to apply for RFC assistance which coincides with Richardson and Troost's (2009) results. Noting that the 6th Federal Reserve district had looser lending policies than the 8th district, Richardson and Troost (2009) find that the 6th district was more effective than the 8th district at reducing bank failures. 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 3 presents the results for the approval step of the recapitalization process,

Variable	Application	RFC Decision	Declined	Approved	Unapplied
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)
Finan. Character. (1932)					
Paid-Up Capital	1.478 (0.167)	1.447 (0.191)			
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)			
Deposit / Liabilities	0.273 (0.073)	0.197 (0.078)			
No. Correspondents			-0.107 (0.107)	0.060 (0.021)	0.023 (0.017)
Corres. Out of State			0.154 (0.114)	0.014 (0.021)	0.009 (0.017)
Charters, Membs., 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)			
Fed. Reserve Mem.		-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 Retailers			0.005 (0.003)	0.001 (0.000)	0.000 (0.000)
% Vote Demo.		0.000 (0.000)			
Manufact. Establish.		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 (1.143)	-3.560 (0.720)	-5.412 (0.677)
Finan. Character. (1935)					
Paid-Up Capital			8.458 (2.172)	3.436 (0.190)	1.694 (0.171)
Total Assets			0.227 (0.140)	0.184 (0.007)	0.261 (0.011)
Cash / Assets			-0.262 (1.045)	0.913 (0.290)	-0.121 (0.112)
Market Shares (1932)					
Liab. / County Liab.	0.178 (0.074)	0.116 (0.050)			
Liab. / Town Liab.	-0.026 (0.036)		0.442 (0.266)	0.000 (0.056)	-0.019 (0.038)
Dummies					
Fed Dist. 6	0.318 (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 (0.094)	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 App. Amt. (y_1)			1.155 (0.660)	-1.028 (0.217)	
RFC Approv. Amt. (y_2)				1.460 (0.192)	

Table 3: Posterior means and standard deviations. Results are based on 10,000 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 100,000.

equation 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 would have to offer the their illiquid assets, such as LD. The RFC accepted LD as collateral and rarely accepted bonds.[5] As a result, LD positively affected the probability of receiving a loan, and bonds and securities negatively affected the receipt of funds. Furthermore, Calomiris and Wilson (2004) found 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.

Most of the county characteristics in equation 2 have credibility intervals that include zero except for town population and acres of cropland – both of which have a negative impact on loan approval. Farming areas may have been less likely to receive RFC funds because the farming relief program, the Agricultural Adjustment Act, was operating at the same time as the RFC and allocating subsidies to farmers. Bernanke (1983) states that in 1933, farmers were experiencing more difficulty than homeowners and nearly half were delinquent in payments. The RFC may have been reluctant to lend to areas where the notoriously delinquent resided. Also, a bank's county liability ratio has a positive impact on loan approval. This finding agrees with Mason (2001) who also finds 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 (Federal Reserve member, ABA member and Federal Reserve districts) and the Democratic votes variable have credibility intervals that include zero. 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 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.

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

The results for the treatment responses, equations 3 – 5, correspond to different subsets of the sample depending which equation is observed. The outcome variable for each equation is LD in 1935. The results for equation 3, column 4 of Table 3, correspond to banks that apply for RFC funds and are declined, the selected untreated sample. The endogenous covariate, y_{i1}, has a positive impact on bank lending. This result implies that, although the loan was declined, the amount of RFC funds requested in a bank's application positively impacted a bank's LD in 1935. 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 it is in the class of banks that are declined funds. Credit extended in 1932 was likely defaulted on reducing the health of the bank and making fewer funds available for LD in 1935. The dummy for state charted banks has a credibility interval that overlaps zero, however, the Federal Reserve member dummy is positive suggesting that Federal Reserve member banks were more stable than nonmember institutions in the selected untreated sample. Federal Reserve member banks operate within a specific regulatory structure which restricts their ability to take on risk so in a panic they perform better. Furthermore, these banks have an additional facet for relief funds beyond the RFC as they can also receive assistance through the discount window.

The results for equation 4 (column 5 in Table 3) correspond to banks that apply for RFC funds and are approved, the selected treated sample. The endogenous covariate, y_{i2}, which is the key covariate of interest, has a positive impact on bank lending. This implies that RFC assistance increases bank lending, which demonstrates the benefits of recapitalization and LOLR policies. Further examinations and interpretations of this result appear in Section 6. Lagged LD has a credibility interval that includes zero. This result is especially interesting when compared to the coefficients for lagged LD in the other subsamples, or treatment responses, which is statistically different from 0. This is suggestive of a "resetting" effect, where the RFC accepted illiquid collateral in exchange for a cash injection providing banks with a fresh platform for lending.

Unlike the other treatment responses, a number of the county characteristics in equation 4 (approved banks) significantly affect bank lending. The number of wholesale retailers and acres of cropland in a county positively affect bank lending. Once a bank received RFC funds, they used these funds to stimulate the local economy and promote confidence in their bank, which is clear from these results. This is in alignment with Butkiewicz (1995) who found that agriculture greatly affects the banking system. 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 dummies for Federal Reserve member, Federal Reserve districts, and state charter 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. The number of correspondent relationships in 1932 positively impacted LD in 1935. This result is intuitive because correspondent banks provided local banks with liquidity often by discounting short-term commercial paper. Thus, the more correspondent relationships a bank had, the more access it had to liquidity. In addition, correspondents often encouraged smaller institutions to extend credit. The selected treated or approved bank sample is the only subgroup where the number of correspondent relationships is significant. This finding supports the positive impact the RFC had on the banking system. With support from the RFC, correspondent relationships remained healthy and positive for individual banks, thus, increasing bank lending and stabilizing the financial system.

The results for equation 5, the last column in Table 3, correspond to banks that do not apply for RFC assistance, the non-selected sample. Banks' LD from 1932 positively affects their LD in 1935. This result differs from the other two subsamples where it was negative for declined banks and insignificant for approved banks. Banks that were stable enough not to apply for RFC assistance, continued their stability through 1935. Unlike the results for the other treatment response equations where the Federal Reserve district dummies are insignificant, the dummies 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 which positively impacted lending.

The importance of recognizing the multi-step recapitalization process is emphasized in the results for the treatment response equations. Modeling each subsample as a separate equation is important because the parameter estimates are vastly different across equations. Discrepancies between equations include: the sign and credibility interval for estimates of lagged LD is different in each equation, most of the county characteristics only matter for the approved subsample, the Federal Reserve member dummy 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 decision structures which stresses the value of the multivariate setup.

5.2.1 Results for Omega

This section presents the estimates for the variance-covariance matrix, $\mathbf{\Omega}$. The estimates display the importance of joint modeling because all of the coefficients are significantly different from 0 except for Ω_{15} which has a credibility interval that contains zero,

$$\mathbf{\Omega} = \begin{pmatrix} 1.175 & 1.330 & -0.834 & -1.056 & -0.041 \\ 1.330 & 1.646 & -0.974 & -1.288 & . \\ -0.834 & -0.974 & 0.930 & . & . \\ -1.056 & -1.288 & . & 1.201 & . \\ -0.041 & . & . & . & 0.119 \end{pmatrix}.$$

Estimates for $\mathbf{\Omega}$ capture aspects of the model that are unobserved, i.e., managerial relationships, corporate governance, and characteristics of a bank's risk, decisions or health that are not controlled for or cannot be measured. 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 (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_{42} 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, unobserved factors not captured within the model are negatively related to bank lending. This result accords well with intuition because the selected sample of banks that applied for RFC loans were banks fearing a deposit run or liquidation. By non-randomly selecting into the "applied" class of banks, they are revealing worse health or vulnerability. Another interesting finding is the insignificance of the Ω_{15}. There is no relationship between the decision to not apply and bank lending. These results demonstrate the importance of modeling selection mechanisms as there are strong correlations between many of these outcomes. Ignoring correlation in the outcomes and joint modeling can lead to specification errors, which are considered in detail in Section 6.3.

6 Additional Considerations

6.1 Sensitivity Analysis

The priors for the benchmark model appear at the beginning of Section 5.2. Prior selection generally involves some degree of uncertainty and this section evaluates how 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 4, which is the amount of RFC assistance in the subsample of approved banks. The coefficient reported in Table 3 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 4 reports the coefficient β_{RFC} for different hyperparameters.

	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

Table 4: β_{RFC} as a function of hyperparameters. The priors for $\boldsymbol{\beta}$ in the benchmark model are centered at zero with a standard deviation of 2.24.

The results indicate nearly no sensitivity around the benchmark result of 1.460. This finding holds true for the results of the entire model and all of its parameters as well. Even the most skeptical critics of bank recapitalization who would place strong negative priors on their economic benefit would be overridden by the data. This implies that the data speak loudly for the benchmark results, which strengthens the overall findings.

6.2 Model Comparison

An issue in the analysis of LOLR policies is model formulation since the appropriate specification is subject to uncertainty. A number of confounding events are occurring in a weakened economy including local shocks, policy responsiveness, and contagious bank runs. Thus, model uncertainty is due to the problem of variable selection. There are a number of financial ratios and characteristics that represent a bank's health and risk, and county characteristics that represent a bank's business environment that can be included in the model. However, including all of these aspects 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, or instruments, necessary for model identification. Arguing for the exclusion of a variable in the selection equation and inclusion in the selected treatment equation is difficult and presents dubious constraints. Difficulties lie in disentangling aspects that influence a bank's decision to apply for assistance from the LOLR and the LOLR's decision to approve or decline the submitted application. This paper utilizes Bayesian model comparison methods to address issues of instrument uncertainty.

For model comparison, given the data $\mathbf{y}$, interest centers upon a collection of models $\{\mathcal{M}_1, ..., \mathcal{M}_L\}$ representing competing hypothesis about $\mathbf{y}$. Each model $\mathcal{M}_l$ is 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 in $\{\mathcal{M}_L\}$ through their posterior odds ratio, which for any two models $\mathcal{M}_i$ and $\mathcal{M}_j$ is written as,

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

where $m(\mathbf{y}|\mathcal{M}_l) = \int f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi_l(\boldsymbol{\theta}_l|\mathcal{M}_l)d\theta_l$ is the marginal likelihood of $\mathcal{M}_l$. Chib (1995) provides a method for calculating the marginal likelihood based on the recognition that the marginal likelihood can be re-expressed as

$$m(\mathbf{y}|\mathcal{M}_1) = \frac{f(\mathbf{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi(\boldsymbol{\theta}_l|\mathcal{M}_l)}{\pi(\boldsymbol{\theta}_l|\mathbf{y}, \mathcal{M}_l)}.$$

Existing techniques are available for estimating the marginal likelihood for multivariate censored data. Methods used in this paper were proposed and developed in Jeliazkov and Lee (2010). Specifically, this paper employs the CRT (Chib-Ritter-Tanner) method to evaluate the likelihood. This paper compares 8 models that differ by variable and instrument selection. Results from the model comparison are displayed in Table 5. 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 fixed effects in each equation that do not appear in the benchmark model. The second column lists the instruments 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. The last column presents the marginal likelihood estimate.

Results for the variable selection show that state fixed effects are unnecessary. The marginal

Model	**Instruments**	**Marg Lik**
1) State fixed effects, asset ratios, HHI	Safe deposit dept. Manufacturing, farming Town market shares	-9,655.1
2) Only town market shares	Bank age % vote Democratic Gov't securities	-9,643.9
3) Total assets, liability ratios	All departments HHI Asset ratios, correspondents	-9,468.2
4) Departments in treatment responses	Bank age % vote Democratic Departments	-9,465.1
5) Correspondents in every equation	Bank age, departments Cropland, county shares, % Democratic Wholesale, town shares	-9,395.4
6) Benchmark	**Bank age, departments** **Cropland, county shares, % Demo.** **Wholesale, town shares, corresp.**	**-9,390.8**
7) No endogenous covarites	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp.	-9,680.8
8) State fixed effects	Bank age, departments Cropland, county shares, % Demo. Wholesale, town shares, corresp.	-9,590.7

Table 5: Results of the model comparison. The first column describes deviations from the benchmark model, the second column identifies instruments in each equation for the selection equation ($\mathbf{x}_1$), selected treatment ($\mathbf{x}_2$) and treatment responses ($\mathbf{x}_3 - \mathbf{x}_5$), respectively. The third column shows the marginal likelihood estimate.

likelihood estimate decreases by roughly 200 on the log scale when state dummies are included. In general, the data support more parsimonious models without including dummies for departments and additional balance sheet ratios. When multiple ratios or financial characteristics representing similar aspects of a bank's health are included, the marginal likelihood decreases representing overfitting of the model.

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} entering equations 3 and 4. This model signifies that there is no direct effect of RFC assistance, just an indirect effect through the correlation in the unobservables. When this model is compared against the benchmark model, the marginal likelihood decreases by 290 points on the log scale. Thus, the data heavily support the benchmark model, where the actions of the RFC have a direct effect on bank success. Further, this model supports the notion that the size of the loan matters. If the data allow these endogenous covariates to enter, as a result of not using 0 or 1 selection, they can have large explanatory power.

Results of the instrument comparison display preference toward multiple instruments or exclusions. Valid instruments for the selection mechanism include the age of the bank and its departments. These two features are not included in a bank's application to the RFC. Thus, their inclusion in the application step and not in the approval step is coherent as the RFC did not have this information on-hand when evaluating each bank. Instruments for the selected treatment equation include variables for cropland (excluded from equation 1), county market shares (excluded from equations 3–5), 1932 town population (excluded from equations 3–5) and democratic votes (excluded from equations 1, 3–5). In the examiners' report on application decisions, the amount of farming in a county was reported and commented on as a factor of their decision. In addition, examiners commented only on county market shares, not town market shares. Identification of the treatment responses is easier because it features data from 1935, as opposed to data solely from 1932. Instruments include wholesale retailers, correspondent relationships, and town market shares. Wholesale information and town market shares are never mentioned in the RFC examiners' reports. Although these features are not included in the approval process, they do explain the performance of a bank. Understandably, these exclusions are a tough argument, however, with the use of Bayesian model comparison techniques, the data show preference toward some exclusions more than others.

6.3 Covariate and Treatment Effects

Interpretation of the resulting parameter estimates presented in Table 3 is complicated by the discreteness of the outcome variables. Analysis up until this point has been based on sign and credibility intervals. However, 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 parameter estimates and discusses two treatment effects.

The key estimate of interest is β_{RFC}, which reflects the impact of RFC assistance on bank lending in equation 4. To calculate how a change in RFC lending transfers to bank lending, the marginal effect is averaged over both observations and MCMC draws. The marginal effect for β_{RFC} is 0.571 which can be interpreted as, \$10,000 of RFC assistance translates to \$5,710 of LD in 1935. This result accords well with the deposit-to-loan ratios during the 1930s and during banking panics, in general.[6] In normal economic times, this would be considered low; however, in adverse macroeconomic conditions this number is standard. The money that was not converted to loans was likely kept in cash reserves to prepare for a bank run. RFC assistance was effectively pushed beyond banks trickling into local economies, thus promoting and restoring confidence in the financial system.

To stress the importance of this new methodology where there is proper consideration of the decision structure and the composition of treatment and control groups, a model that ignores correlation in the outcomes and non-random selection is considered. The marginal effect of β_{RFC} in this erroneous model is 0.41. This result is nearly 30% downward biased, which underestimates the effectiveness of LOLR policies. These errors arise as a result of selection bias and overstating the health of banks in the control group. If a researcher interpreted this result, it would show that cash injections from the RFC had a lower conversion to LD than a standard deposit. Thus, the RFC would not effectively be achieving its goal of restoring health in the banking system because banks would be hoarding cash and not extending credit to the local communities. This fundamental misspecification can have detrimental effects on the results which emphasizes the importance of

[6]Banks curtail their lending during the crises, e.g. deposit-to-loan ratios fell from 0.85 in 1929 to 0.58 in 1933 (Calomiris, 1993). Deposit-to-loan ratios fell to even lower levels during the recent recession.

joint modeling.

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 $\Pr(y_i = 0|x_i^\ddagger, \mathbf{z}_i, \boldsymbol{\theta})$. Formally, the objective is to obtain a sample of draws and evaluate

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

The result gives the expected difference in the computed pointwise probabilities as $x_i^\dagger$ is changed to $x_i^\ddagger$ (Jeliazkov et al., 2008). Computation of these probabilities is afforded by employing the CRT method, developed in Jeliazkov and Lee (2010).

This method is employed to calculate 2 different 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 computed, $\Pr(y_{i4} = 0|\mathbf{x}_{i4}, y_{i2}^\ddagger, \boldsymbol{\theta})$ 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 results indicate,

$$\{\Pr(y_{i4} = 0|y_{i2}^\ddagger) - \Pr(y_{i4} = 0|y_{i2}^\dagger)\} = 0.126.$$

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. This is major evidence of how RFC funds stabilized the economy.

The second treatment effect to consider is 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 scenario where the RFC approved the full requested amounts. This situation displays the difference in the probability of failure if the RFC approved applications for the selected untreated sample (declined banks). The results show,

$$\{\Pr(y_{i3} = 0|y_{i2}^\ddagger) - \Pr(y_{i3} = 0|y_{i2}^\dagger)\} = 0.025.$$

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 from the 12.6 percentage points finding for approved banks which demonstrates the importance of the selection process. RFC loans are 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 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. Therefore, proper consideration of the decision structure and composition of treatment and control groups are of fundamental importance to evaluating the effectiveness of LOLR programs.

7 Concluding Remarks

This paper presents a methodological framework for multivariate treatment effect models in the presence of sample selection and discrete data. The model 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. On the technical side, the methodology developed here is computationally efficient, and has low storage costs.

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. Recapitalization funds not only keep individual banks healthy, but they also promote positive relationships with correspondent networks and counties in which the banks operate. The use of the multivariate treatment effect model is extremely important to the findings in this paper because the results vary for the different subgroups of banks, there are strong correlations between the equations, and the selection process is proven successful. Although RFC assistance was beneficial for the treated group, it would have been minimally helpful for banks that were declined assistance because their economic condition was too severe.

Studying the RFC is an important and relevant topic because the RFC was used as a model for the current program, the Troubled Asset Relief Program (TARP). Further research on LOLR policies should focus on the multi-step decision mechanisms that place banks into different policies and programs to answer questions including 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.

References

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Boston.

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

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

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. W. 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.

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

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.

Dawid, A. P. (2000), "Causal Inference without Counterfactuals," *Journal of the American Statistical Association*, 95, 407–424.

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

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

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

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

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

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

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

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

Lee, L. F. (1978), "Unionism and wage rates: a simultaneous equations model with qualitative and limited dependent variables," *International Economic Review*, 19, 415–433.

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.

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

Mishkin, F. (2006), "How Big a Problem is Too Big to Fail? A Review of Gary Stern and Ron Feldman's Too Big to Fail: The Hazards of Bank Bailouts," *Journal of Economic Literature*, 44, 988–1004.

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. and Troost, W. (2009), "Monetary Intervention Mitigated Banking Panics During the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Boarder 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, 1919-19," *Journal of Economic History*, 59, 714–747.

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

Tobin, J. (1958), "Estimation of relationships for limited dependent variables," *Econometrica*, 26, 24–36.

Van Hasselt, M. (2011), "Bayesian Inference in a sample selection model," *Journal of Econometrics*, 165, 221–232.

Vijverberg, W. (1993), "Measuring the unidentified parameter on the extended Roy model of selectivity," *Journal of Econometrics*, 57, 69–89.

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.