LIKELIHOOD SPECIFICATION IN SIMULTANEOUS EQUATION MODELS FOR DISCRETE DATA

IVAN JELIAZKOV AND ANGELA VOSSMEYER

Abstract. The likelihood function of simultaneous equation models for discrete data is derived as the invariant distribution of a suitably defined Markov chain. The formulation dispenses with controversial recursivity requirements or the need to augment the model with *ad hoc* indeterminacy rules.

Keywords: Discrete data; Invariant distribution; Markov chain; Simultaneity.

JEL Codes: C35, C50.

1. Introduction

Simultaneous equation models represent a key contribution of econometrics to statistical science and are of central importance in many applications involving the joint determination of multiple outcomes. Typical examples include supply and demand analysis, factor demand, or models of strategic behavior. Despite its relative complexity, the continuous data case has been studied extensively and is well understood; various estimation approaches are also available in this setting.

Discrete data contexts, on the other hand, have presented puzzling paradoxes, such as potential outcome indeterminacy (incompleteness), the prospect that outcome probabilities may not sum to 1 (incoherence), and the breakdown of basic analogies with the continuous case. Attempts to salvage the model have revolved around introducing additional structure and imposing parameter constraints that essentially remove simultaneity and replace it with recursive endogeneity. Such constraints have been controversial not only because they may be implausible in applications or unfounded in economic theory, but also because they represent a troubling existential problem at the core of this class of models (Schmidt, 1981; Manski and McFadden, 1981).

In this article, we confront these earlier conclusions directly. We start with an examination of the model and provide a derivation of its likelihood function. A major complication that has been overlooked in this literature is that integration of the latent variables underlying the observed discrete outcomes cannot be performed in traditional ways. Once the problem is recognized, an alternative approach based on the theory of Markov chains is employed to marginalize the latent data and produce the likelihood function. The derivation resolves earlier complications and paradoxes, reveals conceptual linkages among models for discrete, censored, or continuous outcomes, and provides new insights on modeling and identification.

Date: May 2020.

Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697. E-mail: ivan@uci.edu. Robert Day School of Economics and Finance, Claremont McKenna College, 500 E. Ninth Street, Claremont, CA 91711. Email: angela.vossmeyer@cmc.edu.

The validity of the procedure is established under mild and generally tenable conditions that do not affront the sensibility of the model or require controversial restrictions or *ad hoc* rules. The framework is also easily generalizable to specifications involving different data types or higher dimensions.

2. Model and Likelihood Function

We focus on the bivariate simultaneous equation probit model

$$\begin{aligned} z_{i1} = & \quad x_{i1}'\beta_1 + y_{i2}\gamma_1 + \varepsilon_{i1} \\ z_{i2} = & \quad x_{i2}'\beta_2 + y_{i1}\gamma_2 + \varepsilon_{i2} \end{aligned}, \quad i = 1, \dots, n,$$

where $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})' \sim N_2(0, \Omega)$ and Ω is in correlation form. In matrix notation, the model becomes

$$z_i = X_i\beta + \Gamma y_i + \varepsilon_i \tag{1}$$

with

$$z_i = \begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix}, \qquad X_i = \begin{pmatrix} x_{i1}' & 0 \\ 0 & x_{i2}' \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix}.$$

Equation (1) specifies the conditional distribution of the latent z_i that can also be viewed as a set of reaction functions, or a model of "intentions" given others' "actions" (Maddala, 1983). The observed binary outcomes $y_i \equiv (y_{i1}, y_{i2})'$ relate to the latent z_i through the indicator functions $y_{ij} = 1\{z_{ij} > 0\}$, $j = 1, 2$, which define the region $\mathcal{B}_i = \mathcal{B}_{i1} \times \mathcal{B}_{i2}$ where the latent z_i are consistent with y_i, i.e.,

$$\mathcal{B}_{ij} = \begin{cases} (-\infty, 0] & \text{if } y_{ij} = 0 \\ (0, \infty) & \text{if } y_{ij} = 1 \end{cases}.$$

To draw broader connections in our derivations, we proceed in steps. We briefly examine versions of the model that have well-known likelihoods. Contrasts with the baseline model emerge and reveal the need for an alternative approach, which is then presented. Extensions are examined in Section 4.

To simplify notation, let θ denote the collection of all model parameters and $f(\cdot)$ represent a generic probability density or mass function, where possible dependence on the covariates is assumed but suppressed in the conditioning set.

If y_i is absent on the right-hand side of (1), the familiar multivariate probit (MVP) model is obtained

$$z_i = X_i\beta + \varepsilon_i\,, \qquad \varepsilon_i \sim N_2(0, \Omega)\,, \tag{2}$$

which defines $f(z_i|\theta)$, and as before, $y_{ij} = 1\{z_{ij} > 0\}$, $j = 1, 2$. Based on the recognition that $f(y_i|z_i, \theta) = f(y_i|z_i) = \Pr(y_i|z_i) = 1\{z_i \in \mathcal{B}_i\}$, the probability of observing y_i is given by

$$\begin{aligned} \Pr(y_i|\theta) &= \int f(y_i|z_i, \theta) f(z_i|\theta) dz_i \\ &= \int 1\{z_i \in \mathcal{B}_i\} f_N(z_i|X_i\beta, \Omega) dz_i \\ &= \int_{\mathcal{B}_{i2}} \int_{\mathcal{B}_{i1}} f_N(z_i|X_i\beta, \Omega) dz_{i1} dz_{i2}\,, \end{aligned} \tag{3}$$

and the log-likelihood is $\ln f(y|\theta) = \sum_{i=1}^n \ln \Pr(y_i|\theta)$.

There are no conceptual difficulties in extending (3) to accommodate a model with latent simultaneity (e.g., Blundell and Smith, 1989),

$$z_i = X_i\beta + \Gamma z_i + \varepsilon_i\,, \qquad \varepsilon_i \sim N_2(0,\Omega)\,, \tag{4}$$

because upon solving for z_i, the likelihood contribution easily can be expressed as

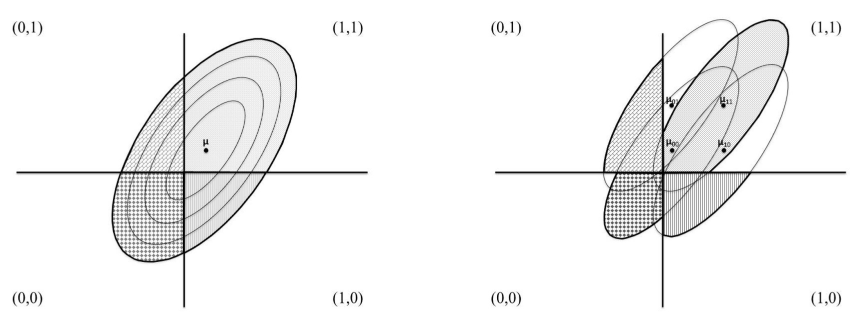
$$\Pr(y_i|\theta) = \int_{\mathcal{B}_{i2}}\int_{\mathcal{B}_{i1}} f_N(z_i|\mu_i,\Sigma)dz_{i1}dz_{i2}\,, \tag{5}$$

with $\mu_i = (I-\Gamma)^{-1}X_i\beta$ and $\Sigma = (I-\Gamma)^{-1}\Omega(I-\Gamma)^{-1\prime}$. Parameter identification can be established as long as the mapping from (4) to (5) is invertible, e.g., by requisite rank and order conditions or covariance constraints, and there is no perfect classification. Most common are exclusion restrictions, whereby each equation contains one or more relevant exogenous covariates not present in the other equation.

The main difficulty arises when dependence involves y_i as in (1). In this case, a heuristic adaptation of the integral (3) has appeared in the literature (e.g., Maddala, 1983; Tamer, 2003; Lewbel, 2007) as a representation for the likelihood contribution

$$\int_{\mathcal{B}_{i2}}\int_{\mathcal{B}_{i1}} f_N(z_i|X_i\beta+\Gamma y_i,\Omega)dz_{i1}dz_{i2}\,, \tag{6}$$

which has raised alarm about model incoherence. Whereas both (3) and (5) integrate to 1 over all possible y_i (Figure 1A), the same can not be guaranteed for (6), where both the mean and region of integration $\mathcal{B}_i$ depend on y_i, leading to 4 distinct distributions, each integrated over a different orthant (Figure 1B). While the probabilities in (6) will sum to 1 under the restriction $\gamma_1\gamma_2 = 0$, the constraint rules out simultaneity in favor of recursive endogeneity or no endogeneity at all. Moreover, substituting $y_i = 1\{z_i \in \mathcal{B}_i\}$ in (1) and backing out y_i from there, instead of the integral in (6), may not be possible without further structure, leading to concerns about model incompleteness. For example, Bajari et al. (2010) consider equilibrium selection rules together with estimation of the model parameters and Narayanan (2013) pursues the approach in a Bayesian setting. Much has been made of the paradoxical nature of this model relative to its continuous counterpart.



(A) MVP model. (B) Simultaneous equation model.

FIGURE 1. Contour plots of implied latent data densities.

Our work is motivated by a simple observation: at the heart of the difficulties highlighted in the literature lies the simple fact that neither (1) nor (6) represents the likelihood contribution $f(y_i|\theta) = \Pr(y_i|\beta, \gamma, \Omega)$. Key to a proper formalization of $f(y_i|\theta)$ is the recognition that the model is defined by two conditional probability functions, $f(z_i|y_i, \theta)$ from (1) and $f(y_i|z_i, \theta) = 1\{z_i \in \mathcal{B}_i\}$, whose product does not give the joint $f(y_i, z_i|\theta)$. As a consequence,

$$
\begin{aligned}
\Pr(y_i|\beta,\gamma,\Omega) &= \int f(y_i, z_i|\theta) dz_i \\
&\neq \int f(y_i|z_i,\theta) f(z_i|y_i,\theta) dz_i \\
&= \int 1\{z_i \in \mathcal{B}_i\} f_N(z_i|X_i\beta + \Gamma y_i, \Omega) dz_i \\
&= \int_{\mathcal{B}_{i2}} \int_{\mathcal{B}_{i1}} f_N(z_i|X_i\beta + \Gamma y_i, \Omega) dz_{i1} dz_{i2}\,,
\end{aligned}
\tag{7}
$$

which confirms that (6) does not represent $f(y_i|\theta)$, and as a consequence, the aforementioned complications based on (6) are irrelevant.

We therefore seek to obtain $f(y_i|\theta)$ without relying on a marginal-conditional hierarchy, $f(z_i|\theta)$ and $f(y_i|z_i, \theta)$ as in (3) or (5). Instead, we employ the conditionals $f(y_i|z_i, \theta)$ and $f(z_i|y_i, \theta)$ that define our simultaneous equation model and lean on Markov chain theory to develop results for the unique invariant distribution from which the ergodic probabilities $f(y_i|\theta)$ can be recovered. Conditional modeling is complementary to other approaches for specifying a joint distribution including direct modeling, the marginal-conditional decompositions used in hierarchical models, or the marginal-marginal specifications underlying copula analysis (Trivedi and Zimmer, 2005). Conditionally specified distributions are discussed in Besag (1974), Gelman and Speed (1993), and Arnold et al. (1992, 2001). The popular Gibbs sampler, formed by sequential sampling from the set of full-conditionals, was introduced in Gelfand and Smith (1990). Key results in Markov chain theory are reviewed in Meyn and Tweedie (1993), Tierney (1994), Chib and Greenberg (1996) and Chib (2001).

In general, existence of full conditional densities is not sufficient for the existence of a proper stationary distribution: integrability can fail even in simple cases, e.g., for the random walk $s|t \sim N(t, \sigma_s^2)$, $t|s \sim N(s, \sigma_t^2)$. Therefore, our discussion considers existence and non-degeneracy in addition to providing a practical way for obtaining the stationary distribution.

Proposition 1. *Let $f(z_i|y_i, \theta) = f_N(z_i|X_i\beta + \Gamma y_i, \Omega)$ and $f(y_i|z_i, \theta) = 1\{z_i \in \mathcal{B}_i\}$ be the conditionals defining the simultaneous equation model. Then, there exists a unique and proper marginal $f(y_i|\theta)$, the desired likelihood contribution, that is consistent with those conditionals.*

Proof. A Gibbs sampler with full-conditionals $f(z_i|y_i, \theta)$ and $f(y_i|z_i, \theta)$ gives rise to a Markov chain $\{z_i^{(m)}, y_i^{(m)}\}_{m=0}^{\infty}$ with proper stationary distribution $f(y_i, z_i|\theta)$, whose finite integrability is a consequence of y_i taking a finite number of possible values. Integrating z_i gives rise to a Markov chain $\{y_i^{(m)}\}_{m=0}^{\infty}$ with transition kernel $f(y_i^{(m+1)}|y_i^{(m)}, \theta)$ and invariant distribution $f(y_i|\theta)$. Since $f(z_i|y_i, \theta)$ is nowhere vanishing, all Markov transition probabilities are bounded away from zero, establishing irreducibility and aperiodicity; standard Markov theory then implies uniqueness of the invariant distribution and convergence from any starting point. □

3. Evaluating the Likelihood Function

Working with the joint or marginal Markov chain draws $\{z_i^{(m)}, y_i^{(m)}\}$ or $\{y_i^{(m)}\}$ offers a simple frequency procedure for evaluating $f(y_i|\theta) = \Pr(y_i|\theta)$ as

$$\hat{f}(y_i|\theta) = G^{-1} \sum_{g=1}^{G} 1\{y_i^{(g)} = y_i\}, \tag{8}$$

where the G draws are collected following the chain's burn-in period. The frequency estimator in (8) is computationally inexpensive as it does not require numerical integration over $\{z_i\}$; it is also versatile as the chain can easily be adjusted for variations in the conditionals $f(z_i|y_i, \theta)$ and $f(y_i|z_i, \theta)$. However, for a finite number of draws, estimates are not strictly bounded between 0 and 1, and boundary values can be especially problematic in evaluating the likelihood function. Another pitfall is the lack of differentiability of (8) with respect to θ because the frequency estimator has the form of a step function. These features impede its use in numerical optimization and complicate the asymptotics of estimators that rely on it.

The difficulties of the frequency estimator can be circumvented, at the cost of additional computation, by working with the Markov transition matrix implied by $f(y_i^{(m+1)}|y_i^{(m)}, \theta)$. In particular, assign the possible outcomes for y_i to a state vector $s = (s_1, \ldots, s_4)'$, for instance,

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

with associated Markov transition matrix P with entries $p_{kh} = \Pr(s_h^{(m+1)}|s_k^{(m)})$

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

The invariant distribution is the left eigenvector of P, which satisfies $\pi' = \pi' P$, i.e., π' is the characteristic vector of P corresponding to the characteristic root of 1 (Meyn and Tweedie, 1993; Hamilton, 1994; Greenberg, 2008). Given P, the vector of ergodic probabilities π is easily obtained by exploiting the properties $\pi' P = \pi'$ and $\pi' j = 1$, where j is a vector of ones (Hamilton, 1994, p. 684). The vector π contains the likelihood values

$$\pi = (\pi_1, \pi_2, \pi_3, \pi_4)' = (f(s_1|\theta), f(s_2|\theta), f(s_3|\theta), f(s_4|\theta))',$$

and the quantity we seek to evaluate, $f(y_i|\theta)$, is the element π_k where $y_i = s_k$.

Practical implementation of our approach requires evaluation of the entries in P, which can be achieved by numerical integration in bivariate cases, or by a variety of differentiable approaches from the literature on simulated likelihood estimation (see, e.g., Train, 2009; Jeliazkov and Lee, 2010) in higher dimensions.

4. Discussion

A number of important remarks are in order. First and foremost, our results suggest that if the specification of $f(z_i|y_i, \theta)$ and $f(y_i|z_i, \theta)$ in the simultaneous equation model is accepted, then there is a unique and proper $f(y_i|\theta)$, derived in Section 3, that is implied by, and consistent with, those conditionals. Deviations from this likelihood must therefore stem from implicit differences in the modeling assumptions, conditional specification, or problems in the derivation.

Second, earlier work has remarked on a puzzling dichotomy between discrete and continuous data settings. We note that the formulation employed here can be instrumental in bridging that gap. To see the linkages, consider the model

$$y_i = BX_i + \Gamma y_i + \varepsilon_i, \tag{9}$$

where the likelihood contribution $f(y_i|\theta)$ is based on the reduced form solution

$$y_i = (I - \Gamma)^{-1}(BX_i + \varepsilon_i). \tag{10}$$

Importantly, since $(I - \Gamma)^{-1} = I + \sum_{j=1}^{\infty} \Gamma^j$, we can obtain y_i iteratively (Fisher and Hughes Hallett, 1988; Young, 1981) as the fixed point in the dynamic system

$$y_i^{(m+1)} = BX_i + \Gamma y_i^{(m)} + \varepsilon_i, \tag{11}$$

which can be thought of as a sequence of instantaneous partial adjustments, a sequential feedback mechanism conceptually similar to impulse responses, that brings the system to equilibrium.[1] Because (11) defines a transition kernel $f(y_i^{(m+1)}|y_i^{(m)}, \theta)$ and y_i is the implied steady state, it is easy to see that the Markovian perspective unifies the data generation mechanisms in the continuous and discrete cases.[2]

Third, it should be clear from the derivations that the outcome probabilities π sum to one, and hence the model is fully coherent. Revisiting Figure 1B, we can see that the integration of each density over the 4 quadrants produces the row entries in the Markov transition matrix P (which naturally sum to 1). Incorrect interpretation of the conditional probabilities on the main diagonal of P as the marginal probabilities π has motivated earlier claims of model incoherence and the proposed solutions requiring the main diagonal of P to sum to 1 – a condition that is neither required nor particularly sensible for a Markov transition matrix.

Fourth, our derivations imply that the model is complete. Because of the functional equivalence between the likelihood function and the data generating process, key to any properly specified likelihood is that it lays out a clear recipe for data generation. While difficulties with simulating data from the simultaneous equation model have been broadly acknowledged, the literature has sought to question the model, rather than the likelihood derivation. Our aim has been to remedy the situation by providing clarity on the likelihood function and insights on data generation in this setting, which can often be neglected when inference is not likelihood-based. We also emphasize that Markov dynamics are only used for the purpose of presenting $f(y_i|\theta)$, but the model in (1) itself could be either static or dynamic.

Fifth, it should be clear that no restrictions on γ_1 or γ_2 are necessary since $f(z_i|y_i, \theta)$ is nowhere vanishing, and the entries in P will be strictly positive. Therefore, the model with full simultaneity is entirely sensible, and need not be restricted to one with recursive endogeneity. We note, however, that even though our specification rules out reducible or periodic P, some parameter settings, e.g., ones that

---

[1] Convergence from any $y_i^{(0)}$ is guaranteed if the spectral radius of Γ is less than 1; otherwise, the series is divergent and equality holds only when (11) is initialized at the value in (10) – a knife-edge solution that is incompatible with a stable equilibrium. This is not a concern in discrete data models, but checking Γ to rule out instability may be appropriate in continuous data settings.

[2] Yet another way to link continuous and discrete data models theoretically is to approximate a continuous y_i on a fine grid and rewrite (9) as a suitably defined ordinal model for the gridpoints. Progressively coarser grids will, in the limit, lead to the binary model considered here.

make certain transitions so remote as to be nearly impossible, can produce P that is *nearly* reducible or periodic. If this is due to the values of the true parameters, the problem will manifest itself similarly to perfect classification. For example, one outcome in y_i, possibly in combination with other covariates, may be a perfect classifier for the other. Importantly, perfect classification is a data-related, not model-driven, identification issue which affects the broader class of discrete data models. If, on the other hand, issues with P occur merely due to the choice of starting value or during the iterations of an estimation algorithm, then careful monitoring of the evolution of the estimation algorithm may be required.

Sixth, extensions of the approach to higher-dimensional systems, different distributional assumptions, or other discretization mechanisms, are conceptually straightforward, although to be practically viable, the number of states s should be kept manageable. This is also true if the integral equation $\pi' = \pi' P$, which is at the center of our methodology, is to be exploited in the construction of other estimators such as the method of simulated moments (McFadden, 1989; McFadden and Train, 2000), or various matching estimators. The proliferation of states is likely to be the main obstacle to possible extensions and implementation in other contexts.

Finally, we performed an extensive simulation study that included data generation and parameter estimation across a wide spectrum of settings. Both maximum likelihood and Bayesian estimation methods such as Metropolis-Hastings (MH) and Accept-Reject Metropolis-Hastings (ARMH) (Tierney, 1994; Chib and Greenberg, 1995), were implemented. The results provided empirical validation of our method and main conclusions. The estimation algorithms performed well with ARMH producing a nearly iid posterior sample and MH following closely behind. We now turn to a brief application in banking employing this methodology.

5. Empirical Application

We study bank behavior using data on the Reconstruction Finance Corporation (RFC), a government-sponsored rescue program designed to assist banks during the Great Depression. The dataset includes information on 1,794 banks and their balance sheets, interbank networks and local market conditions, and is described in detail in Vossmeyer (2014).

The RFC initially supported banks by providing emergency liquidity via collateralized loans; these loans were potential items in otherwise complex bank portfolios of assets and liabilities that affected, and were affected by, other bank lending and decision making. This interconnectedness, coupled with the presence of adverse selection, moral hazard, and dependence on various unobservable factors, suggest that the relationship between bank lending and RFC liquidity may be suitably viewed through the lens of a simultaneous equation model.

In our specification, y_{i1} equals 1 if bank i engaged in aggressive lending, i.e., above-average loan-to-deposit ratio measured in 1932, and y_{i2} equals 1 if the bank applied for assistance from the RFC in the same period. Table 1 provides posterior means and standard deviations from an ARMH estimation algorithm, which are based on 10,000 MCMC draws with a burn-in of 1,000 draws. The table also provides the MLE and standard errors as the proposal density in the ARMH algorithm was based on the mode and modal dispersion matrix of the log-likelihood.

The results exhibit strong simultaneity in the relationship, with large estimates of γ_1 and γ_2 (the respective coefficients on RFC Application and Aggressive Lending),

Table 1. Simultaneous equation probit model of aggressive bank lending and RFC applications.

Variable	Eq. 1: Aggressive Lending		Eq. 2: RFC Application	
	ARMH	MLE	ARMH	MLE
Intercept	2.362 (0.127)	2.382 (0.165)	-3.053 (0.315)	-3.109 (0.446)
Deposits/Assets	-4.049 (0.190)	-4.090 (0.351)	2.877 (0.349)	2.944 (0.497)
Cash/Assets			-4.333 (0.404)	-4.361 (0.604)
ln(Paid-Up Capital)	-0.876 (0.181)	-0.890 (0.199)	0.799 (0.173)	0.808 (0.214)
BankAssets/TownAssets	0.060 (0.064)	0.061 (0.078)		
BankAssets/CountyAssets			0.398 (0.113)	0.409 (0.147)
National Charter	-0.057 (0.089)	-0.052 (0.121)	-0.520 (0.112)	-0.531 (0.147)
Banking Assn. Member			0.265 (0.059)	0.263 (0.081)
# Correspondent Banks	-0.554 (0.083)	-0.550 (0.117)	0.488 (0.107)	0.488 (0.156)
ln(Bank Age)	0.007 (0.026)	0.006 (0.036)		
Nearby Bank Failure			-0.019 (0.058)	-0.016 (0.078)
Acres of Cropland ($\times 10^5$)	-0.076 (0.030)	-0.076 (0.037)		
Aggressive Lending (y_1)			2.609 (0.198)	2.639 (0.275)
RFC Application (y_2)	1.596 (0.177)	1.616 (0.355)		
ω_{ARMH}	0.148 (0.049)			
ω_{MLE}	0.145 (0.064)			

and a positive correlation parameter $\omega = \Omega[1, 2]$. Banks with aggressive lending positions were more likely to apply for RFC assistance as they lacked sufficient liquidity to meet withdrawal demands. Applying for RFC assistance was positively associated with bank lending: while the RFC encouraged lending, many banks viewed the RFC as a government backstop, leading to more aggressive positions without building liquidity buffers (Anbil and Vossmeyer, 2019).

This application illustrates that the methodology can be practical and useful and that the parameters can be estimated without difficulty in the absence of recrusivity restrictions. Consequently, the methods should find application in many fields including labor, industrial organization, transportation, finance, and others.

6. Concluding Remarks

Simultaneous equation models for discrete data have posed an open problem for decades. We show that the likelihood can be derived from the main ingredients of the model using results from the theory of Markov chains. The derivation resolves puzzling paradoxes highlighted earlier, dispenses with unnecessary parameter constraints, offers simple intuitive linkages to the analysis of continuous outcomes, and shows that such models are theoretically coherent, complete, and generalizable.

The approach also introduces the modeling through conditional distributions to economics, which, to the best of our knowledge, is entirely absent from the field despite its suitability in many economic settings such as demand estimation, the modeling of heterogeneity, time series analysis, and many latent data contexts.

References

Anbil, S. and Vossmeyer, A. (2019). The quality of banks at stigmatized lending facilities. *AEA Papers and Proceedings*, 109:506–510.

Arnold, B. C., Castillo, E., and Sarabia, J.-M. (1992). *Conditionally Specified Distributions*. Springer-Verlag, Berlin.

Arnold, B. C., Castillo, E., and Sarabia, J.-M. (2001). Conditionally specified distributions: An introduction. *Statistical Science*, 16(3):249–274.

Bajari, P., Hong, H., and Ryan, S. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

Besag, J. E. (1974). Spatial interactions and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236.

Blundell, R. and Smith, R. (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies*, 56(1):37–57.

Chib, S. (2001). Markov chain Monte Carlo: Computation and inference. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, pages 3569–3649.

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

Chib, S. and Greenberg, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12:409–431.

Fisher, P. and Hughes Hallett, A. (1988). Iterative techniques for solving simultaneous equation systems: A view from the economics literature. *Journal of Computational and Applied Mathematics*, 24(1-2):241–255.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gelman, A. and Speed, T. P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society, Series B*, 55(1):185–188.

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

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.

Jeliazkov, I. and Lee, E. H. (2010). MCMC perspectives on simulated likelihood estimation. *Advances in Econometrics: Maximum Simulated Likelihood*, 26:3–39.

Lewbel, A. (2007). Coherency and completeness of structural models containing a dummy endogenous variable. *International Economic Review*, 48(4):1379–1392.

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

Manski, C. F. and McFadden, D. L. (1981). *Structural Analysis of Discrete Data and Econometric Applications*. MIT Press, Cambridge.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometica*, 57(5):995–1026.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15:447–470.

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

Narayanan, S. (2013). Bayesian estimation of discrete games of complete information. *Quantitative Marketing and Economics*, 11(1):39–81.

Schmidt, P. (1981). Constraints on the parameters in simultaneous tobit and probit models. In Manski, C. F. and McFadden, D. L., editors, *Structural Analysis of Discrete Data and Econometric Applications*, pages 422–434.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70(1):147–165.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1761.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge Univeristy Press, Cambridge, 2 edition.

Trivedi, P. K. and Zimmer, D. M. (2005). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111.

Vossmeyer, A. (2014). Determining the proper specification for endogenous covariates in discrete data settings. *Advances in Econometrics*, 34:223–247.

Young, N. (1981). The rate of convergence of a power matrix series. *Linear Algebra and its Applications*, 35:261–278.