Journal of Econometrics 256 (2026) 106190

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Likelihood specification in simultaneous equation models for discrete data

Ivan Jeliazkov[a,*], Angela Vossmeyer[b,1]

[a] Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA, 92697, USA
[b] NBER and the Robert Day School of Economics and Finance, Claremont McKenna College, 500 E Ninth Street, Claremont, CA, 91711, USA

ARTICLE INFO

JEL classification:
Codes C35
C50

Keywords:
Bayesian inference
Conditional distribution modeling
Discrete data
Invariant distribution
Likelihood function
Markov chain
Simultaneity

ABSTRACT

In this article, we examine the foundations of a rich and diverse literature in economics and derive the likelihood function of simultaneous equation models for discrete data as the invariant distribution of a suitably specified Markov process. This formulation offers a well-defined reduced form of the model and dispenses with the need for controversial recursivity requirements and *ad hoc* indeterminacy rules. The derivation resolves puzzling paradoxes highlighted in earlier work and shows that the likelihood is unique, proper, coherent, complete, and theoretically grounded in conditional distribution modeling – a framework that has yet to be popularized in economics. We note possible extensions and relevant links with other models, comment on computational issues, and implement the methodology in three empirical applications involving female labor force participation, the interactions between health and wealth, and the interplay between banks' lending practices and their reliance on assistance from the lender of last resort during the Great Depression.

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 that 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 simultaneous equation models for continuous data. Approaches intended to deal with these issues have revolved around introducing additional structure or imposing parameter constraints that essentially rule out simultaneity and replace it with recursive endogeneity. Such constraints have been controversial not only because they may be implausible and counterintuitive in applications or unfounded in economic theory, but also because they represent a troubling existential problem at the core of this class of models (Maddala and Lee, 1976; Schmidt, 1981; Manski and McFadden, 1981; Maddala, 1983). These early results have unfortunately hindered subsequent empirical applications despite the presence of numerous discrete data settings where simultaneity is clearly appropriate.

* Corresponding author.
E-mail addresses: ivan@uci.edu (I. Jeliazkov), angela.vossmeyer@cmc.edu (A. Vossmeyer).

[1] The authors are grateful to two anonymous referees, an Associate Editor, the Co-Editors S. Kumbhakar and S. Mallick, and the Managing Editor M. Jansson, for their helpful input and valuable insights during the review process.

https://doi.org/10.1016/j.jeconom.2026.106190
Received 30 September 2024; Received in revised form 5 December 2025; Accepted 11 January 2026
Available online 21 January 2026

In this article, we confront these challenges directly. We begin by examining the model and its relevant background, before providing a derivation of the likelihood function. Our analysis reveals that a subtle, yet crucial, aspect of the model's conditional structure has been overlooked in the earlier literature, resulting in incorrect expressions for the outcome probabilities, which has consequently led to the mistaken perception of paradoxes and logical inconsistencies. Once the problem is recognized, an approach based on Markov process theory is employed to marginalize the latent data, obtain the outcome probabilities, and produce the likelihood function. This derivation not only resolves previous complications and paradoxes, but also uncovers conceptual linkages among models for discrete, censored, and continuous outcomes, and offers new insights into modeling and identification.

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 additional *ad hoc* rules. The framework is also generalizable to specifications involving different data types or higher dimensions.

2. Model and likelihood function

We focus on the bivariate simultaneous equation probit model

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

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

$$z_i = X_i\beta + \Gamma y_i + \varepsilon_i \tag{2}$$

with

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

Eq. (2) 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.,

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

The origins of this model and the broader literature on simultaneity in discrete, censored, and mixed data settings can be traced back to Amemiya (1974, 1978), Maddala and Lee (1976), Nelson and Olson (1978), Heckman (1978), Elliott (1980), Manski and McFadden (1981) and the references therein, Maddala (1983), Judge et al. (1985), and Rivers and Vuong (1988), among others.

To draw broader connections in subsequent derivations, we proceed in steps. We briefly examine versions of the model that have well-known likelihood functions. Contrasts with the baseline model emerge and reveal the need for an alternative methodology, which is then presented. Extensions are examined in Section 3.

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 x is assumed but suppressed in the conditioning set.

The familiar multivariate probit (MVP) model emerges if y_i is absent on the right-hand side of (2), i.e.,

$$z_i = X_i\beta + \varepsilon_i, \qquad \varepsilon_i \sim N_2(0, \Omega), \tag{3}$$

which defines $f(z_i|\theta)$, and as before, $y_{ij} = 1\{z_{ij} > 0\}$, $j = 1, 2$. Now, 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 can be obtained as

$$\begin{aligned} f(y_i|\theta) &= \int f(y_i|z_i, \theta) f(z_i|\theta) dz_i \\ &= \int 1\{z_i \in \mathcal{B}_i\} f_N(z_i|X_i\beta, \Omega) dz_i \\ &= \int_{\mathcal{B}_i} f_N(z_i|X_i\beta, \Omega) dz_i, \end{aligned} \tag{4}$$

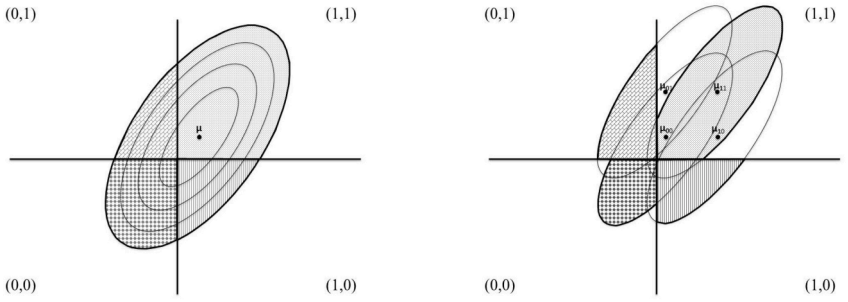
where $f_N(z_i|X_i\beta, \Omega)$ represents the normal density for z_i with mean $X_i\beta$ and covariance matrix Ω, and the log-likelihood becomes $\ln f(y|\theta) = \sum_{i=1}^n \ln f(y_i|\theta)$.

There are no conceptual difficulties in extending (4) to accommodate a model with latent simultaneity (e.g., Maddala and Lee, 1976; Maddala, 1983; Blundell and Smith, 1989; Dagenais, 1999),

$$z_i = X_i\beta + \Gamma z_i + \varepsilon_i, \qquad \varepsilon_i \sim N_2(0, \Omega), \tag{5}$$

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

$$f(y_i|\theta) = \int_{\mathcal{B}_i} f_N(z_i|\mu_i, \Sigma) dz_i, \tag{6}$$



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

Fig. 1. Contour plots of the latent data densities implied by the MVP and simultaneous equation models.

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 (5) to (6) 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. Latent model formulations under a variety of modeling and estimation approaches have been employed, for example, to address issues related to the measurement of happiness (Mamatzakis and Tsionas, 2021), prices (Atkinson et al., 2018), or the analysis of driving behavior (Sarwar et al., 2017).

The main difficulty with simultaneity arises when dependence involves y_i as in (2). In this case, the likelihood contribution $\Pr(y_i|\theta)$ has been represented by a heuristic adaptation of (4), namely

$$\int_{\mathcal{B}_i} f_N(z_i | X_i\beta + \Gamma y_i, \Omega) dz_i \,, \tag{7}$$

raising concerns about model incoherence (Maddala and Lee, 1976; Maddala, 1983; Tamer, 2003; Lewbel, 2007). Underlying the incoherence argument is the observation that whereas both (4) and (6) integrate to 1 over all possible y_i (Fig. 1(A)), the same can not be guaranteed in (7) because both the mean of z_i and region of integration $\mathcal{B}_i$ depend on y_i, leading to 4 distinct conditional distributions, each integrated over a different orthant (Fig. 1(B)). It has been noted that the summability problem of (7) can be resolved under the restriction $\gamma_1\gamma_2 = 0$, but this constraint has been highly contentious: it has no parallel in continuous data settings and has been viewed as arbitrary and problematic in both theory and applications. It rules out simultaneity in favor of recursive endogeneity, where one outcome is known before the other, or removes endogeneity altogether. Alternative estimation approaches for binary and censored simultaneous equation models in the context of incoherence constraints have been considered in Blundell and Smith (1989, 1994), Bresnahan and Reiss (1991), Vella (1993), and Dagenais (1999), among others.

Another common practice has been to substitute (2) into the indicator functions $y_{ij} = 1\{z_{ij} > 0\}$, $j = 1, 2$, treating the result as a data generation mechanism

$$\begin{aligned} y_{i1} &= 1\{x_{i1}'\beta_1 + y_{i2}\gamma_1 + \varepsilon_{i1} > 0\} \\ y_{i2} &= 1\{x_{i2}'\beta_2 + y_{i1}\gamma_2 + \varepsilon_{i2} > 0\} \end{aligned}, \quad i = 1, \ldots, n, \tag{8}$$

but efforts to circumvent the problems with the integral in (7) by instead backing out y_i from (8) have been equally troubling. Solutions to (8) may not always be unique since for some θ and ε_i, Eq. (8) can be qualified as a correspondence satisfied by multiple values of y_i, suggesting model incompleteness. Various methods, primarily in game theoretic contexts, have been suggested for resolving outcome multiplicity through further structure. These have included the agglomeration of multiple indeterminate outcomes into a single category (Bresnahan and Reiss, 1990, 1991), choosing unique solutions from the region of incompleteness with known probability (Kooreman, 1994), and the joint estimation of equilibrium selection rules and model parameters (Bajari et al., 2010; Narayanan, 2013). Tamer (2003) presents estimation techniques that employ probability bounds under multiplicity of equilibria.

Much has been made of the seemingly paradoxical nature of discrete simultaneous equation models relative to their continuous counterparts. The significant modeling and estimation innovations of previous research notwithstanding, our contribution to this class of models focuses on elucidating its statistical foundations. Earlier conclusions about the perceived contradictions and anomalies of this setting are critically assessed, showing that these problems can be negotiated once the modeling structure is properly understood and the likelihood function is correctly formulated.

Our first main result is that (7) does not represent $f(y_i|\theta) = \Pr(y_i|\beta, \gamma, \Omega)$, which gives the likelihood contribution (as a function of θ), or the data generating process and reduced form of the model (as a function of y_i). Key to a proper formalization of $f(y_i|\theta)$ is the recognition that the model is defined by two conditional probability functions, $f(z_i|y_i, \theta)$ implied by (2) 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,

$$\begin{aligned} f(y_i|\theta) &= \int f(y_i, z_i|\theta) dz_i \\ &= \int f(y_i|z_i, \theta) f(z_i|\theta) dz_i \\ &\neq \int f(y_i|z_i, \theta) f(z_i|y_i, \theta) dz_i \\ &= \int 1\{z_i \in \mathcal{B}_i\} f_N(z_i|X_i\beta + \Gamma y_i, \Omega) dz_i \\ &= \int_{\mathcal{B}_i} f_N(z_i|X_i\beta + \Gamma y_i, \Omega) dz_i \,, \end{aligned} \tag{9}$$

which confirms that (7) does not represent $f(y_i|\theta)$. This simple derivation clearly shows that incoherence arguments stemming from (7) are flawed and unfounded.

The same can be said about incompleteness conclusions based on (8) because that equation does not represent a data generating mechanism or reduced form of the model that is consistent with the conditionals $f(z_i|y_i, \theta)$ and $f(y_i|z_i, \theta)$ defining the simultaneous equation model. Whereas in the MVP case in (3), $y_i \sim f(y_i|\theta)$ can be obtained by the method of composition – drawing $z_i \sim f(z_i|\theta)$ marginally of y_i then setting $y_{ij} = 1\{z_{ij} > 0\}$ – it should be clear that discretizing $z_i \sim f(z_i|y_i, \theta)$ as in (8) can not provide a proper recipe for generating y_i in the simultaneous equation case as y_i already appears in the conditioning set, which is the root cause of the inequality in (9). The problems with this approach are further underscored by the contradictory distributional implications it creates between Eqs. (1) and (8), even though both are conditioned on y_i. Rather than being an intrinsic feature of the originally stated model, incompleteness in this case stems from the use of solution techniques that are unsuitable in the conditional framework as they do not lead to a *marginal* depiction of the data generating process from the given set of *conditional* relations.

We thus seek to obtain $f(y_i|\theta)$ in a manner that aligns with the conditional structure of the model. A key question is whether the distributions $f(y_i|z_i, \theta)$ and $f(z_i|y_i, \theta)$ are *compatible* and uniquely determine a proper joint distribution $f(y_i, z_i|\theta)$ with these conditionals and with marginals given by $f(y_i|\theta)$ and $f(z_i|\theta)$. To address this question, we draw upon the literature on conditionally specified distributions, notably the work of Besag (1974), Gelman and Speed (1993), and Arnold et al. (1992, 1999, 2001), which has seen multiple applications in statistics but has yet to be employed in econometrics. Conditional modeling is also at the core of the popular Gibbs sampler (Gelfand and Smith, 1990), a Markov chain simulation approach formed by sequential sampling from the set of full-conditionals (key results in Markov chain theory are reviewed in Meyn and Tweedie, 1993; Tierney, 1994; Chib and Greenberg, 1996, and Chib, 2001). This modeling approach is complementary to other methods for specifying joint distributions including direct modeling, hierarchical marginal-conditional decompositions, or marginal-marginal specifications employed in copula analysis (Trivedi and Zimmer, 2005) including extensions to handle endogenous covariates and vector autoregressive (VAR) settings (Tran and Tsionas, 2022; Tsionas et al., 2022).

Let $\mathcal{S} \equiv \{y_i : \Pr(y_i) > 0\}$ denote the sample space of y_i consisting of the sample points

$$s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad s_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad s_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad s_4 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{10}$$

If $f(y_i|z_i, \theta)$ and $f(z_i|y_i, \theta)$ are compatible and uniquely determine a joint distribution $f(y_i, z_i|\theta)$, we can build a factorization which represents a set of functional equations that identify the probabilities $\{f(y_i|\theta)\}_{y_i \in \mathcal{S}}$. Specifically, for every possible outcome $y_i \in \mathcal{S}$, we have

$$\begin{aligned} f(y_i|\theta) &= \int f(y_i, z_i|\theta) dz_i \\ &= \int f(y_i|z_i, \theta) f(z_i|\theta) dz_i \\ &= \int f(y_i|z_i, \theta) \sum_{s_j \in \mathcal{S}} f(z_i|s_j, \theta) f(s_j|\theta) \, dz_i \\ &= \sum_{s_j \in \mathcal{S}} f(s_j|\theta) \int f(y_i|z_i, \theta) f(z_i|s_j, \theta) \, dz_i \\ &= \sum_{s_j \in \mathcal{S}} f(s_j|\theta) \int_{\mathcal{B}_i} f(z_i|s_j, \theta) \, dz_i \,, \end{aligned} \tag{11}$$

which may be readily contrasted with the expressions in (7) and (9). A notable feature of the derivation in (11) is that the outcome probabilities appear on both sides of the equation, implicitly linking the probability of each element in the sample space to all others; a system of equations emerges when (11) is used to represent the outcome probabilities $f(y_i|\theta)$ for all possible $y_i \in \mathcal{S}$. To rigorously analyze and interpret these relationships, as well as to develop practical methods for likelihood evaluation, we draw upon Markov chain theory. The framework allows us to establish foundational results such as the existence, uniqueness, and non-degeneracy of the joint distribution $f(y_i, z_i|\theta)$ and, by extension, its marginals. These results are essential because the existence of full conditionals

alone does not guarantee the existence of a proper joint distribution.[1] Markov theory is also central to interpreting and solving the functional relations in (11), which also lead to practical approaches for computing the ergodic probabilities $\{f(y_i|\theta)\}_{y_i \in S}$.

Proposition 1 (Existence and Uniqueness). *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 joint distribution $f(y_i, z_i|\theta)$ that is consistent with these conditionals. Furthermore, the marginals $f(y_i|\theta)$ and $f(z_i|\theta)$, the former of which is the desired reduced form and likelihood contribution of the model, are also unique and proper.*

Proof. Consider a Gibbs sampler that iteratively samples from the full conditionals $f(z_i|y_i,\theta)$ and $f(y_i|z_i,\theta)$. This process generates a Markov chain $\{z_i^{(m)}, y_i^{(m)}\}_{m=0}^{\infty}$ with invariant distribution $f(y_i, z_i|\theta)$, whose finite integrability is a consequence of y_i taking a finite number of possible values. Since $f(z_i|y_i,\theta)$ is nowhere vanishing, all Markov transition probabilities are bounded away from zero. This establishes irreducibility and aperiodicity of the Markov chain, which implies uniqueness of the invariant distribution and guarantees convergence from any starting point. By extension, the marginal Markov chain draws $\{y_i^{(m)}\}_{m=0}^{\infty}$ and $\{z_i^{(m)}\}_{m=0}^{\infty}$ have invariant distributions $f(y_i|\theta)$ and $f(z_i|\theta)$, respectively, which are also unique and proper. □

Having shown uniqueness, integrability, and convergence from any starting point, we now consider the question of evaluating $f(y_i|\theta)$ in practice. One approach is to work with the joint or marginal Markov draws $\{z_i^{(m)}, y_i^{(m)}\}$ or $\{y_i^{(m)}\}$ discussed in Proposition 1, offering a simple frequency estimator of $f(y_i|\theta) = \Pr(y_i|\theta)$ as

$$\hat{f}(y_i|\theta) = G^{-1} \sum_{g=1}^{G} 1\{y_i^{(g)} = y_i\}, \tag{12}$$

where the set of G draws is collected following the chain's burn-in period. The frequency estimator in (12) 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 $f(y_i|z_i,\theta)$. Such variations, for instance, could include not only different distributional assumptions on z_i (e.g., logistic, Student's t, etc.), but also differences in the discretization rules that map z_i to y_i, thereby accommodating ordinal (e.g., Albert and Chib, 1993; Jeliazkov et al., 2008) or censored (e.g., Amemiya, 1974; Heckman, 1978) outcomes.

Despite these advantages, the frequency estimator in (12) has important limitations when the number of simulation draws G is finite. In particular, estimates are not guaranteed to lie strictly within the interval $(0, 1)$ – the accept-reject mechanism underlying the estimator can yield boundary values of zero for low-probability events, which is highly problematic when evaluating the likelihood function (specific examples can be found in Jeliazkov and Lee, 2010, Sec. 4). Another pitfall is the lack of differentiability of (12) 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. Additional discussion and references are available in (Train, 2009, Ch. 5).

The drawbacks of the frequency estimator can be circumvented, at the cost of additional computation, by working with the last line in Eq. (11). Central to understanding this expression is the insight that when the sample points in (10) are viewed as states in a Markov chain, the integrals in (11) yield the Markov transition probabilities of the chain $\{y_i^{(m)}\}$ marginally of $\{z_i^{(m)}\}$, i.e.,

$$p_{kh} \equiv \Pr(y_i^{(m+1)} = s_h | y_i^{(m)} = s_k, \theta) = \int_{\mathcal{B}_{s_h}} f(z_i|s_k,\theta) dz_i, \tag{13}$$

which define the probability of transitioning from state s_k to state s_h.[2] Collect all such transition probabilities into a Markov transition matrix P with entries

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

As discussed earlier, this is a simple and well behaved context with a finite dimensional state space and a transition matrix P with strictly positive entries, implying aperiodicity and irreducibility. A fundamental result from Markov process theory can now be invoked, enabling convenient computation of the invariant distribution

$$\pi = (\pi_1, \pi_2, \pi_3, \pi_4)' = (f(s_1|\theta), f(s_2|\theta), f(s_3|\theta), f(s_4|\theta))', \tag{14}$$

thereby providing $f(y_i|\theta)$ as the element π_k where $y_i = s_k$.

Proposition 2 (Invariance). *The invariant distribution π of the Markov chain $\{y_i^{(m)}\}$ governed by the transition matrix P is the solution to the equation*

$$\pi' = \pi' P,$$

i.e., π' is the left eigenvector of P corresponding to the characteristic root of 1, normalized so that $\pi'\mathbf{1} = 1$.

[1] Integrability must be verified explicitly, as it can fail even in simple settings, for example, the random walk defined by $s|t \sim N(t, \sigma_s^2)$ and $t|s \sim N(s, \sigma_t^2)$. Improper models of this sort are discussed in (Arnold et al., 1999, Ch. 6).

[2] The entries represent the orthant probabilities associated with each of the 4 densities in Fig. 1(B).

Proof. Since P is a matrix of transition probabilities, $P\mathbf{1} = \mathbf{1}$, and hence $\mathbf{1}$ is a right eigenvector of P corresponding to eigenvalue 1. Since all entries of P are positive, by the Perron-Frobenius theorem there is only one eigenvalue 1 and all others must be strictly inside the unit circle. Therefore, as $n \to \infty$, Λ^n in the spectral decomposition $P^n = T\Lambda^n T^{-1}$ will converge to a matrix of zeros except for a 1 in the location of the unity eigenvalue and P^n will converge to $\mathbf{1}\pi'$, where $\mathbf{1}$ and π' are the corresponding column and row of T and T^{-1}, respectively. The result then follows since $P^{n+1} = P^n P$ and both P^n and P^{n+1} share the same limit. □

This is a standard invariance result in Markov chain theory, see Billingsley (1986), Meyn and Tweedie (1993), Hamilton (1994, Ch. 22), or Greenberg (2008, Ch. 6), which states that the stationary distribution π is a vector that remains unaffected by the operation of the transition matrix P. For an aperiodic and irreducible Markov chain, π is unique and represents a vector of *unconditional* probabilities needed for determining $f(y_i|\theta)$.

It is important to note that the invariance condition in Proposition 2 is precisely what was derived in (11). Therefore, it provides a useful interpretation of our results within the theory of Markov chains. But this invariance condition is also useful for computing the vector of ergodic probabilities π. In particular, given P, the vector π is easily obtained by exploiting the properties $\pi' = \pi' P$ and $\pi'\mathbf{1} = 1$ [p. 684] (Hamilton, 1994) by defining

$$A = \begin{bmatrix} I - P' \\ \mathbf{1}' \end{bmatrix} \quad \text{and} \quad a = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix},$$

thereby yielding $\pi = (A'A)^{-1}A'a$ as the solution to the system $A\pi = a$.

To summarize, implementation of this approach requires evaluation of the entries in P, which can be achieved by numerical integration through quadrature methods in bivariate cases, or by a variety of differentiable techniques from the literature on simulated likelihood estimation (see, e.g., Train, 2009; Jeliazkov and Lee, 2010) in general settings. Once P is obtained, the vector of outcome probabilities π in (14) can be readily computed as detailed in the preceding paragraph. The likelihood contribution $f(y_i|\theta)$ is then given by the entry of π where the state corresponds to the observed y_i. Note that the framework developed in this section also implies a data-generating mechanism: observations may be simulated either by sampling y_i according to the probability vector π, or equivalently, by taking any draw (upon convergence) from the Markov chain in Proposition 1, which shares the same stationary distribution.

The likelihood function $f(y|\theta) = \prod_{i=1}^n f(y_i|\theta)$ can be used for both maximum likelihood estimation and Bayesian inference through algorithms such as Metropolis-Hastngs (MH) and Accept-Reject Metropolis-Hastings (ARMH) (Tierney, 1994; Chib and Greenberg, 1995; Chib and Jeliazkov, 2005). Appendix A provides computational details on the estimation of P and the ARMH sampler used in our applications.

3. 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 $f(y_i|\theta)$ that is defined by, and consistent with, those conditionals.

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 continuous data model

$$y_i = BX_i + \Gamma y_i + \varepsilon_i, \tag{15}$$

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

$$y_i = (I - \Gamma)^{-1}(BX_i + \varepsilon_i). \tag{16}$$

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

$$y_i^{(m+1)} = BX_i + \Gamma y_i^{(m)} + \varepsilon_i, \tag{17}$$

which can be thought of as a sequence of instantaneous partial adjustments, or a sequential feedback mechanism conceptually similar to impulse responses, that brings the system to equilibrium.[3] Because Eq. (17) 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.[4]

Third, it should be clear from the derivations that the outcome probabilities π sum to one, and hence the model is fully coherent. Revisiting Fig. 1(B), 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 in this context.

[3] 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 (17) is initialized at the value in (16) – 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.

[4] Yet another way to link continuous and discrete data models theoretically is to approximate a continuous y_i on a fine grid and rewrite (15) as a suitably defined ordinal model for the gridpoints. Progressively coarser grids will, in the limit, lead to the binary model considered here.

Fourth, our derivations imply that the model is complete as it admits a well defined reduced form. 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 and reduced form derivations. 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 (2) itself could be either static or dynamic, depending on whether lagged dependent variables are included in X_i. Because of the discrete nature of y_i, such lags would merely serve as intercept shifts.

Fifth, it should be clear that no restrictions on γ_1 or γ_2 are necessary because for all possible $y_i \in S$, the density $f(z_i|y_i, \theta)$ is nowhere vanishing, and therefore all entries in P are strictly positive, implying irreducibility and aperiodicity (Appendix A). 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 the specification rules out reducible or periodic P, some parameter settings, e.g., ones that 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 probability matching or weighting estimators. The proliferation of states is likely to be the main obstacle to possible extensions and implementation in other contexts.

Finally, in Appendix A, we offer illustrative examples on the computation of the Markov transition matrix P and the resulting outcome probabilities π, and examine how the estimation algorithm can be impacted by the size of the model and different values of the simultaneity parameters γ_1 and γ_2. While larger models and stronger simultaneity can impact estimation, the estimation algorithm performs well overall. We now turn our attention to three applications employing this methodology.

4. Empirical applications

Simultaneous relationships are very common in economics. In addition to basic supply and demand interactions that involve simultaneity, empirical studies have explored a number of other relationships including wages and fringe benefits (Vella, 1993), female labor supply and household income (Blundell and Smith, 1989), wages and discrimination (Heckman, 1978), husband and wife fertility decisions (Sobel and Arminger, 1992), and modal choice in travel demand analysis (van Wissen and Golob, 1990), among others. In addition, the industrial organization literature has investigated simultaneous systems related to entry in airline markets (Berry, 1992; Ciliberto and Tamer, 2009) and construction bidding (Bajari et al., 2010).

To demonstrate the utility of our methods, we consider three empirical illustrations: two classic examples of simultaneous relationships in labor economics involving data from the Panel Study of Income Dynamics (PSID) along with one example from the finance literature. The first application is motivated by Blundell and Smith (1994) who analyze a joint model of women's labor force participation and household income. The second application explores the interactions between health and wealth outcomes. Finally, the third application examines the dynamics of bank lending and emergency liquidity assistance during a financial crisis. In Appendix B, we present information about the performance of our estimation approaches in these real data applications.

4.1. Female labor force participation

The labor force participation decisions of married women have been extensively studied in the economics literature, where important determinants include race and socioeconomic status. However, disentangling the decision to participate in the workforce from household income is complex, as a woman's choice to work impacts the family's financial stability, while that stability also influences her decision to work. Without addressing simultaneity directly, applied researchers tackling this issue face a variety of specification concerns, often leading to a challenging search for plausibly exogenous conditions where participation and income would be affected differently – a nearly impossible endeavor. In contrast, using a simultaneous equation modeling framework allows us to present a structural approach where these outcomes can be jointly determined.

The data set for the this application consists of variables for married women's labor force status and family financial stability for a sample of 2920 families from the 1999 PSID survey. The model we consider is that in Eq. (1), where $y_{i1} = 1$ if the married woman is currently in the labor force – employed or currently looking for a job – and is zero otherwise; $y_{i2} = 1$ if the married woman's family is financially stable and zero otherwise. Financial stability is defined as an income-to-needs ratio greater than 2 to align with classifications of the middle-lower class and above. In the sample, nearly 74% of married women are in the labor force, and 32% of families encounter financial difficulties.

The covariates x_{i1} included in the labor force participation equation consist of the married woman's age, the presence of a young child (under 5 years old), and her years of education. Meanwhile, the covariates x_{i2} in the financial stability equation include the

Table 1
ARMH and MLE estimates in the labor force participation and financial stability application.

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

husband's age, race, education, employment status, and number of children. Our covariate selection and model specification closely follow those of Blundell and Smith (1994), with a key distinction being our inclusion of two discrete outcome variables, whereas Blundell and Smith (1994) measures income as a continuous variable and considers only labor force participation as discrete.[5] The simultaneous system is estimated by maximum likelihood and ARMH. The results from the two estimation methods are nearly identical and are presented in Table 1.

The results support the simultaneous specification for the relationship between married women's labor force participation and household financial stability. From Table 1, we see that the coefficients on both endogenous variables are far from zero, as is the error correlation parameter ρ. Specifically, a married woman is more likely to enter the labor force when her family is financially stable, and conversely, her employment positively affects her family's financial situation. The estimated correlation between the two equations is approximately 0.40, indicating a strong relationship between these two outcomes of interest.

Furthermore, the results for other covariates align with existing research and intuition: higher levels of education positively influence both employment decisions and financial stability, while the presence of children tends to have a negative impact. Additionally, being non-white increases the likelihood of a married woman participating in the labor force. Overall, the modeling approach presented here provides a robust empirical framework for labor economists to explore outcomes related to labor force participation.

4.2. Health and wealth

Economists have extensively studied the interactions between health and wealth outcomes through various methods, including life-cycle models, birth weight analyses, and twin studies, among others. A significant focus has been placed on determining the causal direction of this relationship. Many studies, however, conclude that the relationship is reciprocal (see, e.g., Luo and Waite, 2005; Currie and Madrian, 1999). Wealth can lead to better health by providing access to quality insurance and preventive care, while good health can enhance wealth by enabling individuals to work harder and more efficiently. Therefore, a simultaneous equation model is particularly suited for this context to mitigate simultaneity biases.

To investigate the interactions between health and wealth, we utilize data from the 1999 PSID survey, which includes 4405 respondents. In our analysis, $y_{i1} = 1$ indicates that a respondent reports health status better than "good" while $y_{i2} = 1$ signifies that a respondent's income-to-needs ratio exceeds 2, representing wealth categories above the "middle-lower" class. The covariate specification is derived from existing work, incorporating controls for education, parental education, childhood health, race, gender, and marital status, all of which are discretized. Education categories include less than high school (omitted), high school diploma, college degree, and graduate education. Childhood health categories are classified as excellent/very good (omitted), good/fair, and poor. Marital status categories include married (omitted), single, and divorced.

The bivariate simultaneous health-wealth model defined in Eq. (1) is estimated using maximum likelihood and ARMH. Once again, the results from both estimation techniques align closely and are presented in Table 2. The findings underscore the importance of the simultaneous model: both endogenous variables are statistically different from zero, and the correlation between the two equations is 0.08. We observe that health positively influences wealth, and wealth, in turn, positively impacts health. Healthier individuals tend to be more productive in their work, leading to increased wealth, while wealthy individuals gain greater access to healthcare and preventive services, which in turn enhances their health. Furthermore, the results for other covariates align well with existing literature and intuition, indicating that higher levels of education, good health in childhood, and parental education positively affect both health and wealth outcomes. While the literature on the relationships between health and wealth, education and wealth, and intergenerational transmissions of these factors is extensive, Ashenfelter and Krueger (1994), Case et al. (2005), and Currie (2009) provide a useful overview. In general, our findings agree with these studies and the economic intuition behind these relationships.

[5] For a related model with continuous and discrete data outcomes, see Sarwar et al. (2017).

Table 2
ARMH and MLE estimates in the health and wealth application.

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

4.3. Bank lending and emergency liquidity

Bank risk, financial stability, and regulatory intervention are all endogenously related (Tsionas, 2016; Delis et al., 2017). During financial crises, studying these relationships can be particularly challenging due to the rapid actions of financial intermediaries and confounding banking policy, providing a useful setting to employ our methodology. We investigate 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 1794 banks, encompassing their balance sheets, interbank networks, and local market conditions, as outlined by 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 activities 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, $y_{i1} = 1$ if bank i engaged in aggressive lending, i.e., above-average loan-to-deposit ratio measured in 1932, and $y_{i2} = 1$ if the bank applied for assistance from the RFC in the same period. Table 3 provides posterior means and standard deviations from an ARMH estimation algorithm as well as MLE with the associated standard errors.

The results indicate a strong simultaneous relationship, with large estimates of γ_1 and γ_2 (the respective coefficients on RFC Application and Aggressive Lending), and a positive correlation parameter ρ. Banks with aggressive lending positions were more likely to apply for RFC assistance as they lacked sufficient liquidity to meet withdrawal demands. This result aligns with Vossmeyer (2016) who reviews RFC application processes. Further, we find that 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).

The results also show that well-capitalized banks are less strongly associated with aggressive lending. These banks may be more risk adverse and carry more bonds and cash on their balance sheet. Banks with a high capital ratio and banks that hold a national charter are less likely to apply for RFC assistance. Nationally charted banks were in better health during the Great Depression and had access to the discount window, reducing their need to apply to the RFC. Mason (2001) and Mason (2003) also find similar results about banks selecting into the RFC program.

4.4. Summary and comparisons

These applications illustrate that the methodology is both practical and useful in a variety of settings. The ability to estimate parameters without difficulty while allowing for full simultaneity, rather than imposing restrictions or requiring additional structure, enhances the appeal of these techniques and our findings. In each of our applications, the key simultaneity parameters γ_1 and γ_2 differ from zero and are precisely estimated, as are the correlation parameters ρ. This lends strong support to the simultaneous specification, underscoring the value in capturing interdependent relationships between variables.

We document the performance of the ARMH estimation algorithm across the three applications by computing inefficiency factors for the sampled parameters. These factors measure the mixing efficiency of the Markov chain relative to a hypothetical *iid* sample

Table 3
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)		
ρ_{ARMH}		0.148 (0.049)		
ρ_{MLE}		0.145 (0.064)		

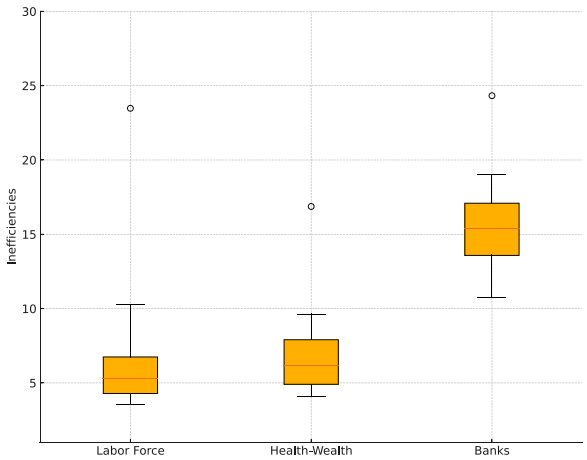


Fig. 2. Boxplots of the inefficiency factors for the three applications.

(Chib, 2001). The inefficiency factor for the kth parameter is defined as $1 + 2\sum_{l=1}^{L} \psi_k(l)(\frac{L-l}{L})$, where $\psi_k(l)$ is the sample autocorrelation at the lth lag and L is the lag in which the autocorrelations taper off. Values close to one indicate that the chain mixes as well as an *iid* sampler. Fig. 2 presents boxplots of the inefficiency factors. Overall, the results suggest good mixing, even in the context of multivariate binary data, where sampling challenges are more common. Notably, inefficiency factors are low across all parameters, including those associated with endogenous covariates, which are often known to mix poorly, indicating that ARMH is performing efficiently in these applications.

Following the model comparisons considered in Sarwar et al. (2017), we estimate three simpler specifications for the data in our applications and present the results in Appendix B. Specifically, the first alternative specification employs univariate (equation-by-equation) probit models, which ignores both the simultaneity structure and correlation in the errors. Second, we estimate the standard multivariate probit model in Eqs. (3)-(4), which does not account for simultaneity, but allows for correlations between the errors in each equation. Finally, we focus on a system with latent simultaneity as discussed in Eq. (5). The results of these alternative methodologies are presented in Appendix B, Tables B.1–B.3, which can be compared to Tables 1–3, respectively.

There are important differences between these approaches and our simultaneous specification. For instance, in the health-wealth application, father's college education is a statistically insignificant predictor of wealth when simultaneity is ignored, whereas its effect is positive and statistically different from zero in the fully simultaneous system. The insignificance does not align with results from the literature on intergenerational transmissions of wealth and education, and is a consequence of the misspecification arising from ignoring simultaneity and the correlation in the errors.

There are some striking differences in the banking application as well. The models that ignore simultaneity imply that capital does not significantly influence aggressive lending behavior. In addition, the comparison models (including the case of latent simulteneity) suggest that the deposit ratio does not influence applications to the RFC. In our fully simultaneous system, these results are statistically different from zero and align with the literature which shows that capital is negatively associated with aggressive lending behavior and the deposit ratio is positively associated with applications to the RFC (Vossmeyer, 2016; Anbil and Vossmeyer, 2019).

These results highlight the dangers of ignoring simultaneity in a system of equations for discrete data. But even if simultaneity is modeled, e.g., using the latent simultaneity model, we mention that the magnitude of the estimated error correlations ρ in those specifications were far larger and had the opposite sign relative to estimates from the observed simultaneity specifications. While this certainly presents an interesting area for future study, here we caution that the high correlations in the latent simultaneity specification also tend to lead to ill-behaved likelihoods for those models, requiring additional care in modeling and estimation.

The implications of our proposed methodology extend beyond the specific applications discussed. Consequently, the methods should find application and offer a robust framework for future empirical studies in many fields including labor economics, industrial organization, transportation, finance, environmental economics, trade and development, and others.

5. Concluding remarks

Simultaneous equation models for discrete data have posed an open challenge for decades. Our critique of earlier results on incompleteness and incoherence in this class of models rests on the recognition that the structure used to define the model is formalized in terms of conditional distributions. As a consequence, representations of the outcome probabilities needed for data generation and likelihood evaluation that are based on marginal-conditional modeling hierarchies are incorrect in this setting. In contrast, we employ Markov process theory to demonstrate that there exists a unique and proper $f(y_i|\theta)$ consistent with the model structure and show that these models are well defined and sensible.

The derivation resolves puzzling paradoxes highlighted in earlier work, dispenses with unnecessary parameter constraints, and establishes simple and intuitive linkages to the analysis of continuous outcomes. It further demonstrates that simultaneous equation models for discrete data are theoretically coherent, complete, and generalizable. The validity of the procedure is established under tenable conditions that are standard in empirical work. Building upon these results, we also provide tools for parameter estimation and employ them in applications to study female labor force participation, the interactions between health and wealth, and the interplay bank lending and lender-of-last-resort programs.

Modeling through conditional distributions has been absent from economics despite its suitability for enhancing core techniques in many areas of econometrics. Yet, one benefit of the methodology considered herein is its generality. Consequently, in addition to traditional areas of simultaneous equation applications in cross-sectional settings, the approach is also applicable to general modeling and latent data contexts in time series and panel data analysis. We intend to explore the effectiveness of such techniques in future research.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Estimation details

A.1. Orthant probability estimation

Estimating the entries of the Markov transition matrix P in Section 2 requires evaluating the integrals

$$p_{kh} = \int_{\mathcal{B}_{s_h}} f(z_i|s_k,\theta)dz_i = \int_{\mathcal{B}_{s_h}} f_N(z_i|X_i\beta+\Gamma s_k,\Omega)dz_i, \tag{A.1}$$

where $f_N(z_i|X_i\beta+\Gamma s_k,\Omega)$ denotes the normal density for z_i with mean $X_i\beta+\Gamma s_k$ and covariance Ω, and $\mathcal{B}_{s_h} = \mathcal{B}_{s_h[1]} \times \mathcal{B}_{s_h[2]}$ with

$$\mathcal{B}_{s_h[j]} = \begin{cases} (-\infty, 0] & \text{if } s_h[j] = 0 \\ (0, \infty) & \text{if } s_h[j] = 1 \end{cases}, \quad j = 1, 2.$$

The quantity p_{kh} represents the transition probability $\Pr(y_i^{(m+1)} = s_h | y_i^{(m)} = s_k, \theta)$ of moving from state s_k to state s_h in the Markov chain $\{y_i^{(m)}\}$ marginalized over the latent $\{z_i^{(m)}\}$. Note that for every k and h, $p_{kh} > 0$ because the integrand in (A.1) is strictly positive over any orthant $\mathcal{B}_{s_h}$, whereby P leads to an irreducible and aperiodic Markov chain implying a unique stationary π.

The integrals in (A.1) can be evaluated using bivariate quadrature methods available in standard statistical software. However, these techniques need to be applied with care because they may produce inaccurate approximations when the evaluation grid is not well positioned, such as in cases involving indefinite integrals, high correlations, extreme probabilities, or sharply varying integrands. Moreover, quadrature methods scale poorly in higher dimensions as the number of gridpoints grows exponentially. To mitigate the computational costs of achieving precise estimates in such instances, we also consider scalable and differentiable alternatives from the literature on simulated likelihood estimation.

The approach rests on samples from the truncated normal distribution of interest $z_i^{(g)} \sim TN_{\mathcal{B}_{s_h}}(z_i|X_i\beta+\Gamma s_k,\Omega)$, $g=1,\ldots,G$, which can be produced by iterative sampling of its conditionals $z_{ij}|z_{i,\setminus j} \sim f\left(z_{ij}|z_{i,\setminus j},s_h,s_k,\theta\right)=\mathcal{TN}_{\mathcal{B}_{s_h}}\left(\mu_{ij},\sigma_{ij}^2\right)$ as proposed in Geweke (1991), where μ_{ij} and σ_{ij}^2 are the conditional mean and variance of z_{ij} given $z_{i,\setminus j}$, which are obtained by the usual conditional updating formulas for a Gaussian density. With these draws, we can approximate (A.1) as

$$\hat{p}_{kh}=f_N(z_i^*|X_i\beta+\Gamma s_k,\Omega)-\hat{f}_{TN_{\mathcal{B}_{s_h}}}(z_i^*|X_i\beta+\Gamma s_k,\Omega),$$

where z_i^* is a fixed value, such as the mean of the draws $\{z_i^{(g)}\}$, and

$$\begin{aligned}\hat{f}_{TN_{\mathcal{B}_{s_h}}}(z_i^*|X_i\beta+\Gamma s_k,\Omega) &= \hat{f}\left(z_{i1}^*|s_h,s_k,\theta\right)f\left(z_{i2}^*|z_{i1}^*s_h,s_k,\theta\right)\\ &= \left[\frac{1}{G}\sum_{g=1}^{G}f(z_{i1}^*|z_{i2}^{(g)},s_h,s_k,\theta)\right]f\left(z_{i2}^*|z_{i1}^*s_h,s_k,\theta\right).\end{aligned}$$

The approach can be extended without additional simulation to $J>2$ dimensional cases by employing the Gibbs kernel average [Sec. 3.2] (Jeliazkov and Lee, 2010)

$$\hat{f}_{TN_{\mathcal{B}_{s_h}}}(z_i^*|X_i\beta+\Gamma s_k,\Omega)=\frac{1}{G}\sum_{g=1}^{G}\prod_{j=1}^{J}f\left(z_{ij}^*|\{z_{ik}^*\}_{k<j},\{z_{ik}^{(g)}\}_{k>j},s_h,s_k,\theta\right).$$

A.1.1. Examples

We present two illustrative examples that demonstrate the computations in specific cases and also relate to a simulation study we conduct in Appendix A.2.1. Specifically, we consider the integrals in Eq. (A.1), which are required to evaluate the entries in the Markov transition matrix P. Under the assumptions $x_{i1}'\beta_1=x_{i2}'\beta_2=0$, $\rho=0.2$, we examine two settings of the parameters (γ_1,γ_2): a low-simultaneity case of $(0.3,0.5)$ and a high-simultaneity case of $(1.3,1.5)$, allowing us to explore how the strength of simultaneity affects the outcome distribution. For the low setting, $(0.3,0.5)$, we obtain

$$P=\begin{pmatrix}0.4547 & 0.2368 & 0.1453 & 0.1633\\ 0.3396 & 0.1604 & 0.2217 & 0.2783\\ 0.2820 & 0.2180 & 0.2820 & 0.2180\\ 0.3740 & 0.3175 & 0.1825 & 0.1260\end{pmatrix},$$

which yields $\pi'=(0.3784,0.2311,0.1971,0.1934)$. A Markov chain of length $10,000$ generated according to Proposition 1 led to a frequency estimate from Eq. (12) of $(0.3812,0.2314,0.1942,0.1932)$, aligning with π'. In contrast, using the incorrect outcome probabilities from Eq. (7) yields $(0.4547,0.1604,0.2820,0.1260)$, which sums to 1.0232 and illustrates the incoherence arguments that have been central in this literature.

For the high-simultaneity case $(\gamma_1,\gamma_2)=(1.3,1.5)$, the estimated transition matrix becomes

$$P=\begin{pmatrix}0.8482 & 0.0850 & 0.0118 & 0.0550\\ 0.4652 & 0.0348 & 0.0620 & 0.4380\\ 0.2820 & 0.2180 & 0.2820 & 0.2180\\ 0.4768 & 0.4564 & 0.0436 & 0.0232\end{pmatrix},$$

leading to the highly skewed $\pi'=(0.7473,0.1209,0.0290,0.1028)$. This result is again confirmed by a frequency estimate of $(0.7500,0.1216,0.0280,0.1004)$. Once again, the estimate from Eq. (7), given by $(0.8482,0.0348,0.2820,0.0232)$, sums to 1.1882 – even further from 1 than before.

We see that relative to the first case, the outcome probabilities in the second case are much more unbalanced, which highlights two challenges. First, frequency estimators like (12) may become unreliable if the chain does not adequately visit low-probability states, potentially yielding zero estimates. Second, the imbalance can degrade parameter identification, adversely affecting estimation as discussed in Appendix A.2.1.

A.2. The accept-reject Metropolis-Hastings (ARMH) algorithm

Due to the form of the likelihood function in the simultaneous equations setting, standard Gibbs samplers for probit models (e.g., Albert and Chib, 1993; Chib and Greenberg, 1998) are not applicable. Consequently, we adopt the Accept-Reject Metropolis-Hastings (ARMH) algorithm (Tierney, 1994; Chib and Greenberg, 1995) as our primary sampling strategy. Marginal likelihood estimation, parameter blocking, and practical implementation issues are addressed in Chib and Jeliazkov (2005).

We are interested in sampling the posterior density

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

where $f(y|\theta)$ is the likelihood function derived in Section 2, and $\pi(\theta)$ denotes the user-defined prior on the parameters. Given a proposal density $h(\theta|y)$ and a positive constant c, let $\mathcal{D}=\{\theta: f(y|\theta)\pi(\theta)\le ch(\theta|y)\}$ represent the region of domination and let $\mathcal{D}^c$ be its complement. Then ARMH sampling proceeds as follows.

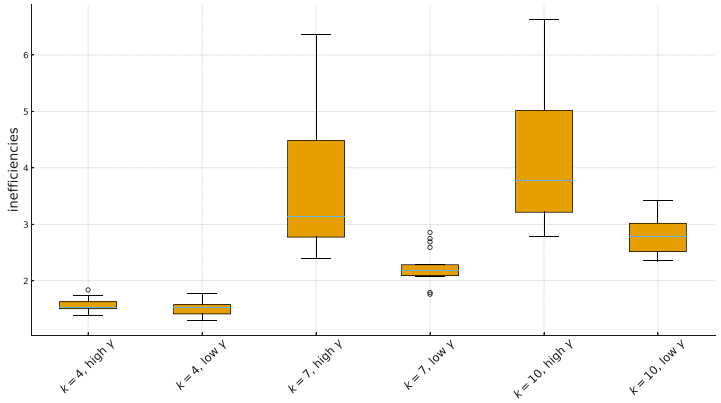


Fig. A.1. Inefficiency factors in the simulation study.

Algorithm 1. *ARMH simulation*

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

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

   repeat until a draw is accepted.
2. *M-H step: Given the current value* θ *and the proposed value* θ'
   (a) *if* $\theta \in \mathcal{D}$, *set* $\alpha_{MH}(\theta, \theta'|y) = 1$
   (b) *if* $\theta \in \mathcal{D}^c$ *and* $\theta' \in \mathcal{D}$, *set* $\alpha_{MH}(\theta, \theta'|y) = \frac{ch(\theta|y)}{f(y|\theta)\pi(\theta)}$
   (c) *if* $\theta \in \mathcal{D}^c$ *and* $\theta' \in \mathcal{D}^c$, *set* $\alpha_{MH}(\theta, \theta'|y) = \min\left\{1, \frac{f(y|\theta')\pi(\theta')h(\theta|y)}{f(y|\theta)\pi(\theta)h(\theta'|y)}\right\}$
   Return θ' *with probability* $\alpha_{MH}(\theta, \theta'|y)$, *else return* θ.

Note that with global domination, when $\theta \in \mathcal{D}$ for all θ in the support Θ of $\pi(\theta|y)$, the ARMH algorithm reduces to an accept-reject sampler and produces *iid* draws; when $\theta \in \mathcal{D}^c$ for all $\theta \in \Theta$, we have an independence chain MH sampler. In intermediate cases when both $\mathcal{D}$ and $\mathcal{D}^c$ are non-empty, ARMH tends to improve the quality of the Markov chain draws relative to M-H sampling at the cost of conducting the additional AR step.

In our applications, the ARMH proposal density was taken to be a Student's t density with ν degrees of freedom $h(\theta|y) = f_{T_\nu}(\theta|\hat{\theta}, V)$, where $\hat{\theta}$ and V were tailored on the basis of the maximum likelihood estimates $\hat{\theta} = \arg\max_\theta \ln f(y|\theta)$ and $V = -[\nabla^2 \ln f(y|\theta)]^{-1}\big|_{\theta=\hat{\theta}}$. To produce a heavy-tailed proposal, we used $\nu = 10$, and set c so that at the mode $f(y|\hat{\theta})\pi(\hat{\theta})/ch(\hat{\theta}) = 1.25$.

A.2.1. Performance of the algorithm

To examine how the estimation algorithm performs as model complexity and features of the data sample vary, we investigate two key dimensions: (i) the number of covariates and (ii) the magnitude of the simultaneity parameters γ_1 and γ_2. To that end, we generate k-dimensional covariate vectors x_{i1} and x_{i2} as independent standard normal variables, set $\beta_1 = \beta_2 = 0.5 \times \mathbf{1}_k$, and let $\rho = 0.2$. We consider three values of k, namely $k = 4, 7, 10$, leading to models with 11, 17, or 23 parameters, respectively. The sample size is set at $n = 2000$, to be roughly in line with the data sets in our applications. The parameters (γ_1, γ_2) are set to either $(0.3, 0.5)$ or $(1.3, 1.5)$, allowing us to compare algorithm performance under weak versus strong simultaneity.

The lower setting yields a relatively balanced distribution across the four possible outcomes. In contrast, the higher setting tends to concentrate mass on specific outcomes, while assigning negligible mass to others (see the example in Appendix A.1.1). This imbalance reduces the variability in observed outcomes, weakening identification and making inference more challenging.[6]

Results from these experiments are presented in Fig. A.1. As expected, smaller models achieve close to *iid* sampling, but as k grows, inefficiency factors grow as well. For a given model size, it is also clear that more poorly identified models with larger values of γ_1 and γ_2 tend to lead to higher inefficiencies. Overall, despite these caveats, the single-block ARMH algorithm still offers good mixing, making it a suitable candidate for estimation of these models. For models where the parameter vector θ is too large to be sampled in one block, the multi-block ARMH sampler may offer improved mixing (see, e.g., Chib and Jeliazkov, 2005).

[6] For simplicity, we consider only positive γ parameters, but because of symmetries in this setting, the qualitative implications of strong simultaneity also apply to other magnitudes and signs of γ_1 and γ_2.

Appendix B. Application: Comparisons

This appendix offers a comparison between our baseline simultaneous equation model, applied in the three empirical settings of Section 4, and the following alternative specifications: (i) univariate probit model, (ii) the multivariate probit (MVP) model defined in Eqs. (3)-(4), and (iii) a system incorporating latent simultaneity, as specified in Eq. (5). The corresponding results are reported in Tables B.1–B.3, which can be compared to the baseline results in Tables 1–3, respectively.

Table B.1
Alternative methods: labor force participation application.

Variable	Eq. 1: Wife Employment			Eq. 2: Financially Stable		
	Univariate	MVP	Sim Latent	Univariate	MVP	Sim Latent
Intercept	−1.189 (0.368)	−1.423 (0.360)	−0.780 (0.264)	−4.306 (0.484)	−4.083 (0.470)	−0.721 (0.348)
Age (f)	0.151 (0.164)	0.303 (0.161)	−0.031 (0.201)	0.578 (0.236)	0.647 (0.230)	0.043 (0.224)
Educ (f)	1.619 (0.325)	2.195 (0.353)	1.632 (0.240)			
Educ (m)				4.744 (0.423)	4.837 (0.412)	0.801 (0.176)
Child < 5	−0.192 (0.054)	−0.165 (0.052)	−0.167 (0.034)			
No. Child				−1.678 (0.181)	−1.619 (0.177)	−0.859 (0.139)
White (f)	−0.120 (0.061)	−0.012 (0.058)	−0.074 (0.038)			
White (m)				0.567 (0.069)	0.552 (0.067)	0.228 (0.061)
Emp (m)				0.749 (0.101)	0.728 (0.098)	0.259 (0.052)
Emp (f) (y_1)				0.568 (0.070)	–	0.939 (0.028)
Income (y_2)	0.639 (0.075)	–	0.437 (0.065)			
ρ_{MVP}			0.291 (0.041)			
ρ_{SimLat}			−0.93 (0.038)			

Table B.2
Alternative methods: health and wealth application.

Variable	Eq. 1: Health			Eq. 2: Wealth		
	Univariate	MVP	Sim Latent	Univariate	MVP	Sim Latent
Intercept	−0.539 (0.064)	−0.392 (0.050)	−0.977 (0.050)	0.576 (0.050)	0.606 (0.050)	0.698 (0.053)
Fedu_HS				0.104 (0.045)	0.108 (0.045)	0.157 (0.031)
Fedu_Coll				0.064 (0.080)	0.069 (0.080)	0.202 (0.048)
Fedu_Grad				0.105 (0.110)	0.111 (0.109)	0.205 (0.058)
Medu_HS				0.240 (0.046)	0.241 (0.046)	0.159 (0.028)
Medu_Coll				0.374 (0.095)	0.382 (0.095)	0.286 (0.055)
Medu_Grad				0.096 (0.123)	0.109 (0.123)	0.335 (0.068)
Edu_HS	0.224 (0.052)	0.248 (0.052)	−0.181 (0.058)	0.379 (0.046)	0.388 (0.045)	0.367 (0.053)
Edu_Coll	0.491 (0.061)	0.531 (0.059)	−0.451 (0.086)	0.880 (0.069)	0.907 (0.068)	0.843 (0.075)
Edu_Grad	0.425 (0.074)	0.469 (0.073)	−0.601 (0.113)	1.028 (0.101)	1.052 (0.101)	0.976 (0.097)
ChHlth_GD	−1.171 (0.037)	−1.252 (0.045)	−0.696 (0.061)			
ChHlth_PR	−0.846 (0.102)	−0.976 (0.110)	−0.509 (0.069)			
Race_Black	−0.137 (0.050)	−0.164 (0.049)	0.318 (0.052)	−0.443 (0.046)	−0.448 (0.046)	−0.440 (0.044)
Female	−0.036 (0.046)	−0.040 (0.046)	0.032 (0.045)	−0.077 (−0.045)	−0.078 (0.045)	−0.075 (0.041)
Single	0.191 (0.071)	0.161 (0.070)	0.655 (0.081)	−0.740 (0.073)	−0.734 (0.073)	−0.752 (0.074)
Divorce	−0.292 (0.049)	−0.323 (0.049)	0.323 (0.060)	−0.618 (0.044)	−0.631 (0.044)	−0.608 (0.052)
Single_Black	−0.152 (0.113)	−0.194 (0.113)	0.224 (0.123)	−0.409 (0.105)	−0.413 (0.105)	−0.396 (0.109)
Health (y_1)				0.181 (0.146)	–	0.103 (0.047)
Wealth (y_2)	0.191 (0.050)	–	0.814 (0.033)			
ρ_{MVP}			0.085 (0.028)			
ρ_{SimLat}			−0.801 (0.036)			

Table B.3
Alternative methods: banking application.

Variable	Eq. 1: Aggressive Lending			Eq. 2: RFC Application		
	Univariate	MVP	Sim Latent	Univariate	MVP	Sim Latent
Intercept	2.295 (0.209)	2.415 (0.207)	2.209 (0.233)	−0.035 (0.187)	−3.109 (0.446)	0.239 (0.416)
Dep/Assets	−3.308 (0.206)	−3.334 (0.205)	−2.684 (0.273)	0.168 (0.215)	−0.168 (0.199)	−0.038 (0.507)
Cash/Assets				−3.363 (0.361)	−3.334 (0.360)	−3.206 (0.367)
ln(Paid-Up Cap)	−0.480 (0.254)	−0.401 (0.251)	−0.689 (0.232)	0.551 (0.221)	0.527 (0.220)	0.520 (0.221)
Bank/Town Ass	0.089 (0.104)	0.088 (0.103)	0.094 (0.085)			
Bank/Cnty Ass				0.263 (0.134)	0.267 (0.133)	0.302 (0.133)
Nat. Charter	−0.190 (0.090)	−0.242 (0.089)	0.030 (0.100)	−0.494 (0.084)	−0.513 (0.084)	−0.476 (0.095)
Bank Assn Mem				0.291 (0.069)	0.288 (0.069)	0.203 (0.072)
# Corr. Banks	−0.508 (0.097)	−0.477 (0.097)	−0.550 (0.094)	0.134 (0.095)	0.090 (0.094)	0.148 (0.115)
ln(Bank Age)	0.062 (0.046)	0.076 (0.045)	0.020 (0.041)			
Near Bank Fail				−0.082 (0.077)	−0.084 (0.076)	−0.072 (0.061)
Acr. Crop. $\times 10^5$	−0.116 (0.048)	−0.109 (0.048)	−0.117 (0.044)			
Aggr. Lend (y_1)				0.279 (0.071)	–	0.060 (0.171)
RFC App. (y_2)	0.350 (0.066)	–	0.571 (0.009)			
ρ_{MVP}			0.173 (0.041)			
ρ_{SimLat}			−0.475 (0.141)			

References

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88, 669–679.
Amemiya, T., 1974. Multivariate regression and simultaneous equation models with the dependent variables are truncated normal. Econometrica 42 (6), 999–1012.
Amemiya, T., 1978. The estimation of a simultaneous equation generalized probit model. Econometrica 46 (5), 1193–1205.
Anbil, S., Vossmeyer, A., 2019. The quality of banks at stigmatized lending facilities. AEA Pap. Proc. 109, 506–510.
Arnold, B.C., Castillo, E., Sarabia, J.-M., 1992. Conditionally Specified Distributions. Springer-Verlag, Berlin.
Arnold, B.C., Castillo, E., Sarabia, J.-M., 1999. Conditional Specification of Statistical Models. Springer-Verlag, New York.
Arnold, B.C., Castillo, E., Sarabia, J.-M., 2001. Conditionally specified distributions: an introduction. Stat. Sci. 16 (3), 249–274.
Ashenfelter, O., Krueger, A., 1994. Estimates of the economic return to schooling from a new sample of twins. Am. Econ. Rev. 84 (5), 1157–1173.
Atkinson, S.E., Primont, D., Tsionas, M.G., 2018. Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions. J. Econ. 204 (2), 131–146.
Bajari, P., Hong, H., Ryan, S., 2010. Identification and estimation of a discrete game of complete information. Econometrica 78 (5), 1529–1568.
Berry, S., 1992. Estimation of a model of entry in the airline industry. Econometrica 60 (4), 889–918.
Besag, J.E., 1974. Spatial interactions and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 36, 192–236.
Billingsley, P., 1986. Probability and Measure. John Wiley & Sons, New York. 2nd edition.
Blundell, R., Smith, R., 1989. Estimation in a class of simultaneous equation limited dependent variable models. Rev. Econ. Stud. 56 (1), 37–57.
Blundell, R., Smith, R., 1994. Coherency and estimation in simultaneous models with censored or qualitative dependent variables. J. Econ. 64, 355–373.
Bresnahan, T., Reiss, P., 1990. Entry in monopoly markets. Rev. Econ. Stud. 57 (4), 531–553.
Bresnahan, T., Reiss, P., 1991. Empirical models of discrete games. J. Econ. 48 (1), 57–81.
Case, A., Fertig, A., Paxson, C., 2005. The lasting impact of childhood health and circumstance. J. Health Econ. 24 (2), 365–389.
Chib, S., 2001. Markov chain Monte Carlo: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics. Vol. 5, pp. 3569–3649.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. Am. Stat. 49, 327–335.
Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econ. Theory 12, 409–431.
Chib, S., Greenberg, E., 1998. Analysis of multivariate probit models. Biometrika 85, 347–361.
Chib, S., Jeliazkov, I., 2005. Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. Stat. Neerl. 59, 30–44.
Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77 (6), 1791–1828.
Currie, J., 2009. Healthy, wealthy and wise: socioeonomic status, poor health in childhood and human capital development. J. Econ. Lit. 47 (1), 87–122.
Currie, J., Madrian, B., 1999. Health, health insurance and the labor market. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Vol. 3, pp. 3309–3416.
Dagenais, M., 1999. A simultaneous probit model. Cah. Econ. Brux. 163 (3), 325–346.
Delis, M., Iosifidi, M., Tsionas, M.G., 2017. Endogenous bank risk and efficiency. Eur. J. Oper. Res. 260 (1), 376–387.
Elliott, D., 1980. Recursive systems containing qualitative endogenous variables representing nonstochastically dependent events. Econometrica 48 (3), 761–763.
Fisher, P.G., Hughes Hallett, A.J., 1988. Iterative techniques for solving simultaneous equation systems: a view from the economics literature. J. Comput. Appl. Math. 24 (1-2), 241–255.
Gelfand, A.E., Smith, A. F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.
Gelman, A., Speed, T.P., 1993. Characterizing a joint probability distribution by conditionals. J. R. Stat. Soc. Ser. B 55 (1), 185–188.
Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In: Keramidas, E.M. (Ed.), Computing Science and Statistics. Fairfax: Interface Foundation of North America, Inc., pp. 571–578.
Greenberg, E., 2008. Introduction to Bayesian Econometrics. Cambridge University Press, New York.
Hamilton, J., 1994. Time Series Analysis. Princeton University Press, New Jersey.
Heckman, J., 1978. Dummy endogeneous variables in a simultaneous equation system. Econometrica 46 (4), 931–959.
Jeliazkov, I., Graves, J., Kutzbach, M., 2008. Fitting and comparison of models for multivariate ordinal outcomes. Adv. Econ. Volume 23: Bayesian Econometrics, 115–156.
Jeliazkov, I., Lee, E.H., 2010. MCMC perspectives on simulated likelihood estimation. Adv. Econ. Maximum Simul. Likelihood 26, 3–39.
Judge, G.C., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.-C., 1985. The Theory and Practice of Econometrics. Wiley Series in Probability and Mathematical Statistics., New York. 2 ed.
Kooreman, P., 1994. Estimation of econometric models of some discrete games. J. Appl. Econ. 9 (3), 255–268.
Lewbel, A., 2007. Coherency and completeness of structural models containing a dummy endogenous variable. Int. Econ. Rev. 48 (4), 1379–1392.
Luo, Y., Waite, L.J., 2005. The impact of childhood and adult SES on physical, mental, and cognitive well-being in later life. J. Gerontol. Ser. B 60 (2), 93–101.
Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.
Maddala, G.S., Lee, L.-F., 1976. Recursive models with qualitative endogenous variables. Ann. Econ. Soc. Meas. 5 (4), 525–545.
Mamatzakis, E., Tsionas, M., 2021. Making inference of British household's happiness efficiency: a Bayesian latent model. Eur. J. Oper. Res. 294, 312–326.
Manski, C.F., McFadden, D.L., 1981. Structural Analysis of Discrete Data and Econometric Applications. MIT Press, Cambridge.

Mason, J.R., 2001. Do lender of last resort policies matter? The effects of reconstruction finance corporation assistance to banks during the great depression. J. Financ. Serv. Res. 20 (1), 77–95.
Mason, J.R., 2003. The political economy of reconstruction finance corporation assistance during the great depression. Explor. Econ. Hist. 40 (2), 101–121.
McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometica 57 (5), 995–1026.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econ. 15, 447–470.
Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer-Verlag, London.
Narayanan, S., 2013. Bayesian estimation of discrete games of complete information. Quantit. Mark. Econ. 11 (1), 39–81.
Nelson, F., Olson, L., 1978. Specification and estimation of simultaneous-equation model with limited dependent variables. Int. Econ. Rev. 19 (3), 695–709.
Rivers, D., Vuong, Q., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. J. Econ. 39, 347–366.
Sarwar, M.T., Fountas, G., Anastasopoulos, P.C., 2017. Simultaneous estimation of discrete outcome and continuous dependent variable equations: a bivariate random effects modeling approach with unrestricted instruments. Anal. Methods Accid. Res. 16, 23–34.
Schmidt, P., 1981. Constraints on the parameters in simultaneous tobit and probit models. In: Manski, C.F., McFadden, D.L. (Eds.), Structural Analysis of Discrete Data and Econometric Applications, pp. 422–434.
Sobel, M.E., Arminger, G., 1992. Modeling household fertility decisions: a nonlinear simultaneous probit model. J. Am. Stat. Assoc. 87 (417), 38–47.
Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria. Rev. Econ. Stud. 70 (1), 147–165.
Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Stat. 22, 1701–1761.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge Univeristy Press, Cambridge. 2 edition.
Tran, K., Tsionas, M., 2022. Efficient semiparametric copula estimation of regression models with endogeneity. Econ. Rev. 41 (5), 485–504.
Trivedi, P.K., Zimmer, D.M., 2005. Copula modeling: an introduction for practitioners. Found. Trends Econ. 1 (1), 1–111.
Tsionas, M., Izzeldin, M., Trapani, L., 2022. Estimation of large dimensional time varying VARs using copulas. Eur. Econ. Rev. 141, 103952.
Tsionas, M.G., 2016. Parameters measuring bank risk and their estimation. Eur. J. Oper. Res. 250 (1), 291–304.
Vella, F., 1993. A simple estimator for simultaneous models with censored endogeneous regressors. Int. Econ. Rev. 34 (2), 441–457.
Vossmeyer, A., 2014. Determining the proper specification for endogenous covariates in discrete data settings. Adv. Econ. 34, 223–247.
Vossmeyer, A., 2016. Sample selection and treatment effect estimation of lender of last resort policies. J. Bus. Econ. Stat. 34 (2), 197–212.
van Wissen, L.J., Golob, T.F., 1990. Simultaneous equation systems involving binary choice variables. Geogr. Anal. 22 (3), 224–243.
Young, N.J., 1981. The rate of convergence of a power matrix series. Linear Algebra Appl. 35, 261–278.