Dynamic Discrete Data Models with Correlated Errors

Preliminary Draft (March 2025)

Shubham Karnawat

Department of Economics, University of California Irvine

Abstract

The paper presents a general approach to modeling and estimation of time series models with discrete outcomes where errors are autoregressive and lagged dependence may be in observed discrete outcomes or latent dependent variable. With such a general framework, estimation is challenging due to large dimension of the latent variable and high correlation in MCMC draws. To deal with these estimation issues, the paper presents efficient Markov chain Monte Carlo algorithms that with a novel blocking technique in sampling the latent variable. The importance of modeling autoregressive errors is demonstrated by comparing the models with its modeling counterparts having independent errors. The performance of the proposed algorithms is demonstrated in multiple simulation studies and the benefits of the proposed models are illustrated in a study of the US business cycles.

Keywords: Probit Model, Discrete Data, Time Series, Gibbs Sampling, Markov chain Monte Carlo, Multivariate Truncated Normal

1. Introduction

Discrete data arise in a variety of time series contexts. Important applications in economics pertain to forecasting business cycle expansions and recessions (Kauppi and Saikkonen (2008) and Chauvet and Potter (2005)), changes in macroeconomic policy stance (see Eichengreen et al. (1985) and Hu and Phillips (2004)), dynamic discrete choice modeling of labor decision outcomes, changing opinions over time, modeling win probabilities in sports, etc. Logistic regression (or logit) and probit regression models are popular methods of analyzing such data in both classical econometrics as well as Bayesian statistics. Theoretical foundation

Email address: shubham.karnawat@uci.edu (Shubham Karnawat)

This is a preliminary version of the paper. We welcome any comments, suggestions, or feedback that could help improve the work.

of both classical and Bayesian inference of discrete data has their roots in random utility representation of discrete outcomes. According to this framework, the response variable take discrete values based on an underlying latent continuous variable, which, in turn can be derived from the individual's utility (or preference) from the theory of choice modeling. Bayesian methods enable incorporation of latent process dynamics into observed data models via hierarchical parameter estimation. This feature has two advantages. First, it enables explicit modeling of the latent process through latent data augmentation into likelihood of the data. Second, it provides alternative model specification by allowing explicit dependence on past latent variables where dependence on observed discrete response is a matter of interest. This paper seeks to answer a key question arising from this observation. Does lagged-dependence on latent process (hereby latent-dependent) provide better model fit on the real data examples as compared to lagged-dependence on observed discrete outcomes (state-dependent).

Modeling dynamic dependence on the observed discrete outcomes has received some attention in the classical statistics (see Jacobs and Lewis (1978),). However, classical models present several challenges usually absent in the analysis of continuous time series models. Firstly, classical inference involves popular likelihood based methods such as logit and probit models where computation of likelihood is trivial. These methods become intractable when exact computation of likelihood is infeasible. Secondly, models do not allow for simultaneous dynamic dependence in the model through both state-dependence and autocorrelated random error components. For example, Kauppi and Saikkonen (2008) considered both state-dependence and latent-dependence specifications but imposes independence assumption on the errors to utilize maximum likelihood estimator. Though, several appealing alternatives to maximum likelihood have been proposed in the literature. For example, Park and Phillips (2000) develop a non-stationary discrete data model allowing for the discrete process to be integrated, however, assume that explanatory variables are exogenous, which can be restrictive in some applications. In a recent paper, De Jong and Woutersen (2011) developed a smooth maximum score estimator for the binary time series model allowing for the lagged dependence on the past realizations of response variable.

In the Bayesian literature, research on modeling endogenous dependence on the latent process remains limited, despite the aforementioned advantages. In a recent paper, Mintz et al. (2013) develop a multivariate discrete choice model with endogenous dependence on latent variable to study the relationship between information processing patterns and propensity to buy using cross-sectional data on 895 online shoppers. They find that the latent measures of information processing provides a better fit to the data. In contrast, Vossmeyer

(2014) find observed representation of education better explains socioeconomic outcomes for younger cohorts. Both these studies highlight the importance of considering both the model specifications in the absence of apriorinformation on proper specification. In the time series context, Chauvet and Potter (2005) extended the standard Bayesian probit model to allow for latent dependence to predict the US business cycles. However, their model allows latent dependence only up to one lag and no serial correlation in the errors.

This paper explores Bayesian modeling of binary response data with alternative dynamic dependence specifications, state-dependent vs latent-dependent under general error structures. The paper proposes simple and efficient Gibbs sampling procedure to estimate the four model specifications, a state-dependence model with independent errors, state-dependence model with serially correlated errors, latent-dependence model with independent errors, and latent-dependence model with serially correlated errors. The remainder of the paper is organized as follows. Section 2 and Section 3 introduce state-dependent models and latent-dependent models respectively and present the Gibbs sampling algorithms for the proposed models. Section 4 illustrates the performance of the model in multiple simulation studies. Section 5 implements the algorithms in an emprirical application from finance. Section briefly discusses possible future works and extensions. Section 6 concludes with some final remarks.

2. State-dependence Model

This section presents the discrete data model with state dependence under two different correlation structures, independent errors and autoregressive errors. In all the model specifications discussed in the paper, y_t represents the binary response variable of interest, z_t denotes the underlying latent process driving the response variable, $X's$ are the exogenous covariates, and finally ε_t represents the stochastic error term.

2.1. Model 1: State-dependence with independent errors

The first model specification considered in the paper imposes independence assumption on error terms and allow for lagged dependence on the state variables. The proposed model specification can be studied through binary probit model example with the response y_t taking binary values. We utilize $y_{t-1}, \ldots, y_{t-J}$ as the covariates to capture the state-dependence dynamics. Intuitively, we allow the probability of response at any period t to depend on the past responses. For example, winning the current rally ($y_t = 1$, if wins) in tennis may affect the probability of winning the next rally, or winning a hand in poker might affect the chances of winning the next hand. We adopt the latent variable approach, where z is the continuous latent random variable expressed as a function of covariates and errors as

$$\begin{aligned}
z_t &= x_t'\beta + \sum_{j=1}^{J} \phi_j y_{t-j} + \varepsilon_t, \qquad t = 1, \cdots, T \\
y_t &= \mathbb{1}\{z_t > 0\} \\
\varepsilon_t &\overset{\text{iid}}{\sim} \mathcal{N}(0,1)
\end{aligned} \tag{1}$$

The model does not suffer from problems of stationarity as lagged dependence on past states only shifts the intercept of the latent variable z. Thus, we do not need to impose any parameter restrictions on the coefficient of lagged dependence terms and we can treat lagged dependent terms as categorical covariates. Rewriting the model by combining the exogenous covariates and lagged y_t simplifies the model further,

$$\begin{aligned}
z_t &= w_t'\gamma + \varepsilon_t, \qquad t = 1, \cdots, T, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \\
y_t &= \mathbb{1}\{z_t > 0\}
\end{aligned} \tag{2}$$

Stacking the observations over time, we get

$$y = \mathbb{1}\{W\gamma + \varepsilon \geq \mathbf{0}_T\}, \qquad \varepsilon \sim \mathcal{N}(0, \mathbf{I}_T) \tag{3}$$

We assume Gaussian priors on the model parameters $\gamma = (\beta, \phi)$ and joint posterior distribution of the latent variable and model parameter is given by

$$\pi(z, \gamma | y) \propto \exp\left\{-\frac{(z - W\gamma)'(z - W'\gamma)}{2}\right\} \times \exp\left\{-\frac{(\gamma - \gamma_0)'\Gamma_0^{-1}(\gamma - \gamma_0)}{2}\right\}$$

where the first term on the right-hand side follows directly from the normality of exogenous errors and (γ_0, Γ_0) are the prior hyperparameters. The normality assumption on errors and priors provides a two-step Gibbs sampling procedure to estimate the model parameters. The details of the Gibbs sampling procedure are provided in the Algorithm 1.

Note that, the consideration of binary probit only serves as an example for this specification. The model specification is readily generalizable to other limited dependent variable settings such as ordinal, small count data, or censored data models. For example, replacing the second line in the equation 2 by $y_t = \mathbb{1}\{z_t > 0\} z_t$ leads to Tobit or censored (with lower truncation) data models. To motivate Tobit model specification, consider an application from macroeconomics, econometric estimation of the Taylor rule. Taylor rule is a simple monetary policy rule that serves as a useful benchmark on the interest rates for analyzing monetary

Algorithm 1 (Gibbs sampling in state-dependence with independent errors)

---

(1) Sample $\gamma|y, z \sim \mathcal{N}(\hat{\gamma}, \hat{\Gamma})$, where,

$$\hat{\Gamma} = (\Gamma_0^{-1} + W'W)^{-1} \qquad \text{and} \qquad \hat{\gamma} = \hat{\Gamma}(\Gamma_0^{-1}\gamma_0 + W'z)x$$

(2) Sample the vector z conditionally on (y_t, γ) by drawing $z_t|z_{-t}, y, \gamma \sim TN_{\mathcal{B}_t}(w_t'\gamma, 1)$ for each $t = 1, \cdots, T$, where

$$\mathcal{B}_t = \begin{cases} (-\infty, 0] & \text{if } \; y_t = 0 \\ (0, \infty) & \text{if } \; y_t = 1 \end{cases}$$

---

policy stance[2]. Since, its introduction, the Taylor rule has been ubiquitously employed in the macroeconomics models to study the dynamics of the system. In the DSGE models, the optimal monetary policy question is also studied under discretion and commitment to a policy rule by assuming that the interest rates follow a simple Taylor-like rule. In the literature, Taylor rule is often estimated as unrestricted OLS with interest rate smoothing as

$$i_t = \rho i_{t-1} + \gamma + \theta(x_t - \bar{x}) + \phi(\pi_t - \bar{\pi}) + \varepsilon_t^i \tag{4}$$

where, i_t is the nominal interest rate or policy rate, x_t is the output gap, $\bar{x}$ is the targeted output gap, π_t denotes the inflation and $\bar{\pi}$ denotes the targeted inflation rate, typically set to 2%. The central bank in the US often relies on the Taylor rule specified above in equation 4 in their large macroeconomic DSGE models for forecasting key macroeconomic variables. However, estimating the Taylor rule with a simple OLS may lead to biased estimates of the Taylor rule coefficients (θ, ϕ) due to truncation of interest rates from below, often known as Zero Lower Bound (or ZLB) on interest rates. However, New Keynesian models are rarely modeled with this truncation because of the complexity and non-linearities arising due to ZLB (Fernández-Villaverde et al. (2015)). To explicitly estimate the Taylor rule with ZLB constraints, one can modeify estimate equation 2 as

$$\begin{aligned} z_t &= x_t'\beta + \sum_{j=1}^{J} \phi_j i_{t-j} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \\ i_t &= \mathbb{1}\{z_t > 0\}\, z_t \end{aligned} \tag{5}$$

[2]See Federal Reserve Board Monetary Policy Report July 2021

2.2. Model 2: State-dependence with correlated errors

Model 1 specification addresses the dynamics of the response variable only through lagged dependence of past states. In this section, We extend the model in section 1 by introducing another source of dynamic dependence by relaxing the independence assumption on errors. Specifically, We assume that errors follow an AR process this providing dynamic dependency through error correlation structure. The specified model can be written as

$$
\begin{aligned}
z_t &= x_t'\beta + \sum_{j=1}^{J}\phi_j y_{t-j} + \varepsilon_t, \qquad t = J+1, \cdots, T \\
y_t &= \mathbb{1}\{z_t > 0\} \\
\varepsilon_t &= \sum_{i=1}^{p}\theta_i \varepsilon_{t-i} + \nu_t, \qquad \nu_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)
\end{aligned}
\tag{6}
$$

where y_t denotes the observed binary outcome at time t, x_t' is $1 \times k$ vector of explanatory variables or covariates that may or may not be stationary, β is $k \times 1$ vector of parameters, z_t is the unobserved latent variable that determines the binary outcome y_t through the indicator function. Finally, ε_t is the serially correlated error term which is assumed to follow a stationary auto-regressive process of order p. For what follows next, it is convenient to rewrite the model by stacking the response data y_t by time.

$$
\begin{aligned}
z &= X\beta + L\phi + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0_{T\times 1}, \Omega) \\
z &= W\gamma + \varepsilon \\
y &= \mathbb{1}\{z > 0_{T\times 1}\}
\end{aligned}
\tag{7}
$$

where $X = (x_1', \ldots, x_T')'$, $L = (y_{t-1}, \ldots, y_1)$, $W = [X, L]$ and $\gamma = (\beta', \phi')'$, $\phi = (\phi_1, \ldots, \phi_J)'$. The paramater $\theta = (\theta_1, \ldots, \theta_p)'$ are the lag coefficients for the autoregressive error process. The joint normality of the error term ε follows directly from the normality of ε_t and the variance covariance matrix Ω is a symmetric Toeplitz matrix implied by the auto regressive process with entries $[\Omega_{i,j}] = Cov(\varepsilon_i, \varepsilon_j) \equiv \rho_{|i-j|}$ determined by the Yule-Walker equations

$$
\rho_k = \theta_1\rho_{k-1} +, \ldots \theta_p\rho_{k-p}, \qquad k = p, \ldots, T-1 \tag{8}
$$

The first p values (ρ_0, ρ_{p-1}) are given by the first p columns of matrix $[I - F \otimes F]^{-1}$, where $\otimes$ denotes the Kronecker product and matrix F is given by

$$
F = \begin{bmatrix} & \theta' & \\ \mathbf{I}_{p-1} & & \mathbf{0}_{(p-1)\times 1} \end{bmatrix} \tag{9}
$$

It is important to note that the inverse of the covariance matrix (Ω^{-1}) is banded, which greatly simplifies the computation since typically $p \ll T$. Given the normality of the errors, the joint data (y) and latent data (z) density of the model is given by

$$\begin{aligned} f(y, z|\gamma, \Omega) &= f(y|z, \gamma, \Omega) f(z|\gamma, \Omega) \\ &= \prod_{t=1}^{T} f(y_t|z_t, \gamma, \Omega) \mathcal{N}(z|W\gamma, \Omega) \end{aligned} \tag{10}$$

We make following assumptions on the choice of prior densities of the model parameters

$$\pi(\beta) \sim \mathcal{N}(\beta_0, B_0), \qquad \pi(\phi) \sim \mathcal{N}(\phi_0, \Phi_0), \qquad \pi(\theta) \sim \mathcal{N}(\theta_0, \Theta_0)\mathcal{I}_\theta \tag{11}$$

where $\mathcal{I}_\theta$ implies that the prior specification of θ satisfies the stationarity condition that roots of polynomial $\Theta(L) = 1 - \theta_1 L - \cdots - \theta_p L^J$ lie outside the unit circle. The normality assumption of the error term implies that prior follows a multivariate truncated normal distribution with the truncation region satisfies $\mathcal{I}_\theta = 1$.
The complete data density, thus is given by

$$\begin{aligned} f(y, z, \gamma, \Omega) &= [\prod_{t=1}^{T} f(y_t|z_t, \gamma, \Omega)] \mathcal{N}_T(z|W\gamma, \Omega)\pi(\beta)\pi(\phi)\pi(\theta)\mathcal{I}_\theta \\ &= [\prod_{t=1}^{T} \mathbb{1}\{z_t \in \mathcal{B}_t\}] \mathcal{N}_T(z|W\gamma, \Omega)\pi(\gamma)\pi(\theta)\mathcal{I}_{\mathcal{S}_\theta} \end{aligned} \tag{12}$$

We propose an efficient MCMC procedure to sample the parameters of interest (z, γ, θ) from their respective posterior conditional distributions (see Appendix 1 for the derivation). The posterior conditional distribution of the latent data z follows a truncated multivariate normal distribution. Sampling from a potentially large dimensional truncated normal distribution is a recurring problem in Bayesian discrete data literature (see Gelfand et al. (1992), Geweke (1996) and Yu and Tian (2011) for common applications). The traditional rejection-based method is challenging for two reasons. First, identifying the importance density or proposal density blanketing the target multivariate truncated normal is difficult. Secondly, the acceptance probability is quite low leading to the rejection of a lot of draws before accepting one draw, making it computationally inefficient. Geweke (1991) and Robert (1995) proposed a Gibbs sampling procedure to simulate truncated multivariate normal by sampling one at a time from univariate truncated normal. However, their procedure suffers from a high correlation between the subsequent draws leading to slow convergence. We propose a novel block partition sampling procedure to sample from a multivariate truncated normal. We utilize, partitioning of multivariate normal $X \sim N(\mu, \Sigma)$ to lower dimensional vector as

$$
\begin{aligned}
& X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \\
& X_1|X_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(X_2 - \mu_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \\
& X_2|X_1 \sim N\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(X_1 - \mu_1\right), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)
\end{aligned} \tag{13}
$$

The partition and be easily generalized to a B-block partitioning $X = (X_1, \cdots, X_B)$ and conditional distribution of $X_b|X_{\setminus b}$ can be obtained similar to equation 13. The linear constraints that transforms the multivariate normal to a multivariate truncated normal remains the same as each variable enters in only one constraint.

$$
\begin{pmatrix} l_1 \\ \vdots \\ l_B \end{pmatrix} \leq \begin{pmatrix} X_1 \\ \vdots \\ X_B \end{pmatrix} \leq \begin{pmatrix} u_1 \\ \vdots \\ u_B \end{pmatrix}
$$

To sample from the full truncated normal distribution we utilize the Gibbs sampling procedure in the spirit of Robert (1995) by sampling each block at a time. However, we still have to sample from T/B dimensional truncated normal to sample each partitioned block. To sample from T/B truncated distribution, we utilize the exponential minimax tilting approach proposed by Botev (2017). The method employs an accept-reject method using exponential tilting to sample from multivariate normal distribution under linear restriction and can generate exact independent and identically distributed samples from the distribution. The drawback of the procedure is it becomes highly inefficient as the dimension is increased (greater than 100 in practice) and there is a high correlation between the partitioned block. This is a curse of dimensionality as acceptance probability becomes too small and draws are inaccurate even with an exponential tilting approach. The key to our approach is to carefully design the block partitioning of multivariate truncated normal to utilize the desired properties of the exponential tilting approach. We study the block sampling procedure in detail in simulation exercise 2. Algorithm 2 provides the detailed approach to our block-partition sampling. Sampling of the rest of the parameters is straightforward. The posterior conditional of $\gamma = (\beta', \phi')'$ follows multivariate normal . Simulation of θ is done by the independence chain Metropolis-Hastings algorithm following Chib and Greenberg (1994).

Algorithm 2 (MCMC sampling in state-dependence and correlated errors)

---

(1) Sample $\gamma|y, z, \rho \sim N\left(\hat{\gamma}, \hat{\Gamma}\right)$, where,

$$\hat{\Gamma} = \left(\Gamma_0^{-1} + W'\Omega^{-1}W\right)^{-1} \quad \text{and} \quad \hat{\gamma} = \hat{\Gamma}\left(\Gamma_0\gamma_0^{-1} + W'\Omega^{-1}z\right)$$

(2) Sample the vector z conditionally on (y, γ, ρ) by drawing $z_b|z_{-b}y, \gamma, \rho \sim TN_{\mathcal{B}_b}(\mu_b, \Omega_b)$ for each $b = 1, \cdots, B$, where

$$\mu_b = E\left(z_b|z_{-b}, y, \gamma, \theta\right) \qquad \Omega_b = var\left(z_b|z_{-b}, y, \gamma, \beta\right)$$

$$\mathcal{B}_b = \mathcal{B}_{\frac{T*(b-1)}{B}+1} \cdots \times \mathcal{B}_{\frac{T*b}{B}} \quad \text{and} \quad \mathcal{B}_t = \begin{cases} (-\infty, 0] & \text{if } y_t = 0 \\ (0, \infty) & \text{if } y_t = 1 \end{cases}$$

(3) Sample $\theta|y, \gamma, z$ using Metropolis Hastings algorithm, with the proposal $\theta' \sim N\left(\hat{\theta}, \hat{\Theta}\right) I_{s_\theta}$ and acceptance probability α_{MH} where

$$\hat{\Theta} = \left(\Theta_0^{-1} + E'E\right)^{-1} \quad \text{and} \quad \hat{\theta} = \hat{\Theta}\left(\Theta_0^{-1}\theta_0 + E'e\right)$$

$$e_1 = (\varepsilon_1, \ldots, \varepsilon_J), \qquad e = (\varepsilon_1, \ldots, \varepsilon_T)' \qquad \text{and} \qquad \varepsilon_t = z_t - w_t\gamma$$

and E is a $((T-p) \times p)$ matrix, with t-th row given by

$$(e_{t-1}, \cdots, e_{t-p}), \qquad t \geq p+1$$

$$\alpha_{MH} = \min\left\{\frac{\Psi(\theta')}{\Psi(\theta)}, 1\right\} \text{ and } \Psi(\theta) = |\Omega_\theta|^{-\frac{T}{2}} exp\left\{-\frac{1}{2}\left(e_1'\Omega_\theta^{-1}e_1\right)\right\}$$

---

3. Latent-dependence Model

In this section, We present the models that explore the dynamics in the response variable and the latent variable that determine the response through lagged dependence on the latent variable itself. This is a key distinction from the models presented in section 2 that latent variables can take any values in the real line and hence are unrestricted. Latent variable modeling originated from psychology literature as a factor analysis of test scores to deconstruct the mental attributes associated with the IQ level. In Economics, latent variable modeling has its theoretical roots in the random utility model framework, which argues that latent variables can be interpreted as the differences in the utility of choosing an alternative against a set of discrete outcomes. Since then, latent variable formulations have been widely applied across various models in discrete choice literature. The latent variable representation is particularly useful in Bayesian discrete choice models with the emergence of Gibbs sampling procedures and data augmentation (see Jeliazkov and Rahman (2012)).

The idea of latent variables as covariates is not completely new either, Particularly in a time series context. Geweke (1977) in his seminal work proposed dynamic factor models for economic time series data. The work has gained huge traction in empirical macroeconometrics and finance literature (see Stock and Watson (2006) for a comprehensive review of the literature). However, this idea has been largely unexplored in the discrete choice models. In a recent paper, Mintz et al. (2013) utilizes both state-dependent and latent dependent specifications to study consumer propensity to buy using latent measures of information processing patterns. Vossmeyer (2014) in another paper, explored latently and observed endogeneity specifications in multivariate discrete choice settings. She employed Bayesian model comparison methods to address the proper specification of the endogenous covariates. However, both these papers are applied in the cross-sectional context. This paper explores the dynamics of the latent dependent model in time series binary response models.

3.1. Model 3: Latent-dependence model with independent errors

The first model specification We explore in this set up imposes i.i.d. restrictions on the error terms. Thus, the correlation structure of the model is same as the Model 1 in section 2. The latent dependent model with i.i.d. assumptions on errors is specified as

$$z_t = x_t'\beta + \sum_{j=1}^{J} \phi_j z_{t-j} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1), \qquad t = 1, \cdots, T \tag{14}$$

$$y_t = \mathbb{1}\{z_t > 0\} \tag{15}$$

where $y_t, z_t, x_t, \beta, \phi$ and ε_t follow the same notation as in the state-dependent model. The difference with the state-dependent model is that the continuous latent data z_t depends on its lagged values and errors are assumed to be independent and identically distributed. We assume that z is a stationary process. This is equivalent to setting initial values of z's to zeros. The joint data and latent data density can then be written as

$$\begin{aligned} f(y, z|\beta, \phi) &= \prod_{t=1}^{T} f(y_t|z_t, \beta, \phi)\, f(z_t|z_{t-1}, \cdots, z_{t-J}, \beta, \phi) \\ &= f(y|z, \beta, \phi) \prod_{t=1}^{J} f(z_t|\beta, \phi) \prod_{t=J+1}^{T} f(z_t|z_{t-1}, \cdots, z_{t-J}, \beta, \phi) \end{aligned} \tag{16}$$

where the second line follows directly from the law of total probability. To obtain the joint density, stack the latent data over time to get the unconditional distribution of vector z

Algorithm 3 (MCMC sampling in latent-dependence and independent errors)

---

(1) Sample $\beta|y,z,\phi \sim N\left(\hat{\beta},\hat{B}\right)$, where,

$$\hat{B} = \left(B_0^{-1} + X'\Omega^{-1}X\right)^{-1} \text{ and } \hat{\gamma} = \hat{B}\left(B_0^{-1}\beta_0 + X'\Omega^{-1}\left(z - L\phi\right)\right)$$

(2) Sample the vector z conditionally on (y,γ,ρ) by drawing $z_b|z_{-b}y,\gamma,\rho \sim TN_{\mathcal{B}_b}\left(\mu_b,\Omega_b\right)$ for each $b = 1,\cdots,B$, where

$$\mu_b = E\left(z_b|z_{-b},y,\gamma,\theta\right) \qquad \Omega_b = var\left(z_b|z_{-b},y,\gamma,\beta\right)$$

$$\mathcal{B}_b = \mathcal{B}_{\frac{T*(b-1)}{B}+1}\cdots\times\mathcal{B}_{\frac{T*b}{B}} \qquad \text{and} \qquad \mathcal{B}_t = \begin{cases}(-\infty,0] & \text{if } y_t = 0\\ (0,\infty) & \text{if } y_t = 1\end{cases}$$

(3) Sample ϕ conditional on (y,z,β), using MH with a proposal from truncated normal $\phi \sim N\left(\hat{\phi},\hat{\Phi}\right)I_{s_\phi}$, where,

$$\hat{\Phi} = \left(\Phi_0^{-1} + L'\Omega^{-1}L\right)^{-1} \text{ and } \hat{\phi} = \hat{\Phi}\left(\Phi_0^{-1}\phi_0 + L'\Omega^{-1}\left(z - X\beta\right)\right)$$

$$L = \left(z_{-1},\cdots,z_{-p}\right), \qquad z_{-k} \text{ is the kth lag of z}$$

$$\alpha_{MH_\phi} = \min\left\{\frac{\chi\left(\phi'\right)}{\Psi\left(\phi\right)},1\right\} \text{ and}$$

$$\chi\left(\phi\right) = |A^{-1}A'^{-1}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\left(Az_1 - X_1\beta\right)'\left(Az_1 - X_1\beta\right)\right)\right\}$$

$$\begin{bmatrix} z_1\\ z_2\\ \vdots\\ z_T\end{bmatrix} = \begin{bmatrix} x_1'\beta\\ x_2'\beta\\ \vdots\\ x_T'\beta\end{bmatrix} + \begin{bmatrix} z_0 & z_{-1} & \dots & z_{-p+1}\\ z_1 & z_0 & \dots & z_{-p+2}\\ \vdots & & & \\ z_{t-1} & z_{t-2} & \dots & z_{t-p}\end{bmatrix}\begin{bmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_p\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_T\end{bmatrix} \tag{17}$$

Rewrite the equation 17 by taking the terms involving $z_1,\ldots,z_{T-1}$ to left hand side

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0\\ -\phi_1 & 1 & 0 & \dots & 0\\ \vdots & \vdots & & & \vdots\\ -\phi_p & \dots & 1 & \dots & 0\\ \vdots & \vdots & & & \vdots\\ 0 & \cdots & -\phi_p & \dots & 1\end{bmatrix}}_{A}\underbrace{\begin{bmatrix} z_1\\ z_2\\ \vdots\\ z_{p+1}\\ \vdots\\ z_T\end{bmatrix}}_{z} = \underbrace{\begin{bmatrix} x_1'\beta\\ x_2'\beta\\ \vdots\\ x_{p+1}'\beta\\ \vdots\\ x_T'\beta\end{bmatrix}}_{X\beta} + \underbrace{\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_{p+1}\\ \vdots\\ \varepsilon_T\end{bmatrix}}_{\varepsilon}$$

This implies that $z \sim \mathcal{N}\left(A^{-1}X\beta, A^{-1}A'^{-1}\right)$. The complete data density, thus is given by

$$f\left(y, z|\beta, \phi\right) = \mathbb{1}\left\{z \in \mathcal{B}\right\} \frac{\exp\left\{-\frac{1}{2}\left(z - A^{-1}X\beta\right)'\left(A^{-1}A'^{-1}\right)^{-1}\left(z - A^{-1}X\beta\right)\right\}}{\left|A^{-1}A'^{-1}\right|^{-\frac{1}{2}}} \tag{18}$$

On the priors of β and ϕ, We make similar assumptions as with state-dependent model. However, note that no restrictive assumptions were made on coefficient of lagged variables ϕ for the state-dependent model.

$$\pi\left(\beta\right) \sim \mathcal{N}\left(\beta_0, B_0\right), \qquad \pi\left(\phi\right) \sim \mathcal{N}\left(\phi_0, \Phi_0\right)\mathcal{I}_{\mathcal{S}_\phi} \tag{19}$$

3.2. Model 4: Latent-dependence with correlated errors

In this section, We generalize model 3 by imposing an autoregressive correlation structure on error terms. Assuming the errors to follow an AR process contributes to the dynamics of the latent variable to depend on both past utilities as well as past shocks. The model is specified as follows

$$\begin{aligned} z_t &= x_t'\beta + \sum_{j=1}^{J}\phi_j z_{t-j} + \varepsilon_t, \qquad t = 1, \cdots, T \\ y_t &= \mathbb{1}\left\{z_t > 0\right\} \\ \varepsilon_t &= \sum_{i=1}^{p}\theta_i\varepsilon_{t-i} + \nu_t, \qquad \nu_t \overset{\text{iid}}{\sim} N\left(0, 1\right) \end{aligned} \tag{20}$$

Similar to latent dependence model with independent errors, stacking the the latent data over time to get the unconditional distribution

$$\begin{aligned} Az &= X\beta + \varepsilon \\ z &= A^{-1}X\beta + A^{-1}\varepsilon \end{aligned}$$

This implies that $z \sim \mathcal{N}\left(A^{-1}X\beta, A^{-1}\Omega A'^{-1}\right)$. The complete data density, thus is given by

$$f\left(y, z|\beta, \phi, \theta\right) = \mathbb{1}\left\{z \in \mathcal{B}\right\} \frac{\exp\left\{-\frac{1}{2}\left(z - A^{-1}X\beta\right)'\left(A^{-1}\Omega A'^{-1}\right)^{-1}\left(z - A^{-1}X\beta\right)\right\}}{\left|A^{-1}\Omega A'^{-1}\right|^{-\frac{1}{2}}} \tag{21}$$

We make usual assumptions for the priors

$$\pi\left(\beta\right) \sim \mathcal{N}\left(\beta_0, B_0\right), \qquad \pi\left(\phi\right) \sim \mathcal{N}\left(\phi_0, \Phi_0\right)\mathcal{I}_\phi, \qquad \pi\left(\theta\right) \sim \mathcal{N}\left(\theta_0, \Theta_0\right)\mathcal{I}_\theta \tag{22}$$

Algorithm 4 (MCMC sampling in latent-dependence and correlated errors)

---

(1) Sample the vector z conditionally on (y, γ, ρ) by drawing $z_b | z_{-b} y, \gamma, \rho \sim TN_{\mathcal{B}_b}(\mu_b, \Omega_b)$ for each $b = 1, \cdots, B$, where

$$\mu_b = E(z_b | z_{-b}, y, \gamma, \theta) \qquad \Omega_b = var(z_b | z_{-b}, y, \gamma, \beta)$$

$$\mathcal{B}_b = \mathcal{B}_{\frac{T*(b-1)}{B}+1} \cdots \times \mathcal{B}_{\frac{T*b}{B}} \quad \text{and} \quad \mathcal{B}_t = \begin{cases} (-\infty, 0] & \text{if } y_t = 0 \\ (0, \infty) & \text{if } y_t = 1 \end{cases}$$

(2) Sample $\beta | y, z, \theta, \phi \sim N\left(\hat{\beta}, \hat{B}\right)$, where,

$$\hat{B} = \left(B_0^{-1} + X'\Omega^1 X^{-1}\right)^{-1} \quad \text{and} \quad \hat{\beta} = \hat{B}\left(B_0^{-1}\beta_0 + X'\Omega^{-1}Az\right)$$

(3) Sample ϕ conditional on y, β, z, θ using Metropolis Hastings algorithm, with the proposal $\phi' \sim N\left(\hat{\phi}, \hat{\Phi}\right) I_{s_\phi}$ and acceptance probability α_{MH_ϕ} where

$$\hat{\Phi} = \left(\Phi_0^{-1} + L'\Omega^{-1}L\right)^{-1} \quad \text{and} \quad \hat{\phi} = \hat{\Phi}\left(\Phi_0^{-1}\phi_0 + L'\Omega^{-1}(z - X\beta)\right)$$

$$\alpha_{MH_\phi} = ming\left\{\frac{\Psi(\phi')}{\Psi(\phi)}, 1g\right\} \text{ and}$$

$$\Psi(\phi) = |A^{-1}\Omega_\theta A'^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\left(z_1 - A^{-1}X_1\beta\right)'\left(A^{-1}\Omega_\theta A'^{-1}\right)\left(z_1 - A^{-1}X_1\beta\right)\right)\right\}$$

(4) Sample θ conditional on y, β, z, ϕ using Metropolis Hastings algorithm, with the proposal $\theta' \sim N\left(\hat{\theta}, \hat{\Theta}\right) I_{s_\theta}$ and acceptance probability α_{MH_θ} where

$$\hat{\Theta} = \left(\Theta_0^{-1} + E'E\right)^{-1} \quad \text{and} \quad \hat{\theta} = \hat{\Theta}\left(\Theta_0^{-1}\theta_0 + E'e_2\right)$$

$$e_1 = (\varepsilon_1, \ldots, \varepsilon_J), \qquad e = (\varepsilon_1, \ldots, \varepsilon_T)' \qquad \text{and} \qquad \varepsilon_t = z_t - X_t\beta - z_{t-1}\phi$$

and E is a $((T-p) \times p)$ matrix, with t-th row given by

$$(e_{t-1}, \cdots, e_{t-p}), \qquad t \geq p + 1$$

$$\alpha_{MH} = min\left\{\frac{\Psi(\theta')}{\Psi(\theta)}, 1\right\} \text{ and } \Psi(\theta) = |\Omega_\theta|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\left(e_1'\Omega_\theta^{-1}e_1\right)\right\}$$

---

4. Simulation Studies

To illustrate the performance of the proposed MCMC algorithm for the four model specifications, We carry out numerous simulation studies. The results of the simulation study are reported in subsequent sub sections below.

4.1. Simulation Study 1: State and Latent Dependent Model Estimation

In this subsection, We present the four examples, one each for the different model specifications proposed in the paper. The details of the data generating process for each of the cases is described below

Example 1: State-Dependent Model with IID errors

$$
\begin{aligned}
&y_t = \mathbb{1}\left\{\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t\right\}, \qquad t = 1, \cdots, 500 \\
&x_{2t} \sim \mathcal{N}(2,1), x_{3t} \sim \mathcal{N}(-1,1), \varepsilon_t \sim \mathcal{N}(0,1) \\
&\beta = (-1, 2, 3)', \theta = (0.8, -0.5)
\end{aligned}
$$

Example 2: State-Dependent Model with AR(1) errors

$$
\begin{aligned}
&y_t = \mathbb{1}\left\{\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta y_{t-1}\varepsilon_t\right\}, \qquad t = 1, \cdots, 250 \\
&\varepsilon_t = \theta\varepsilon_{t-1} + \nu_t \\
&x_{2t} \sim \mathcal{N}(0,1), x_{3t} \sim \mathcal{N}(0,1), \nu_t \sim \mathcal{N}(0,1) \\
&\beta = (-1, -2, 1)', \theta = -0.8, \theta = 0.9
\end{aligned}
$$

Example 3: Latent-Dependent Model with IID errors

$$
\begin{aligned}
&y_t = \mathbb{1}\left\{\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \theta z_{t-1}\varepsilon_t\right\}, \qquad t = 1, \cdots, 250 \\
&x_{2t} \sim \mathcal{N}(0,1), x_{3t} \sim \mathcal{N}(0,1), \varepsilon_t \sim \mathcal{N}(0,1) \\
&\beta = (-1, -1, 1)', \theta = -0.8
\end{aligned}
$$

Example 4: Latent-Dependent Model with AR(1) errors

$$
\begin{aligned}
&y_t = \mathbb{1}\left\{\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t\right\}, \qquad t = 1, \cdots, 300 \\
&\varepsilon_t = \theta\varepsilon_{t-1} + \nu_t \\
&x_{2t} \sim \mathcal{N}(2,1), x_{3t} \sim \mathcal{N}(-1,1), \varepsilon_t \sim \mathcal{N}(0,1) \\
&\beta = (1, -2, -1)', \theta = -0.6, \theta = 0.8
\end{aligned}
$$

We do not assume any specific prior information on any of the parameters, the prior means were set to zero and prior variance set at 100 implying a very diffuse prior. The results are obtained from 10000 MCMC draws with a burn-in sample of 2000 draws. Table 1 reports the posterior mean and standard deviation of the model parameters from the 10000 MCMC draws

Model 1: State-Dependent with IID Errors				
Parameter	True Value	Mean	Std	IF
β_1	-1	-0.94	0.28	8.85
β_2	2	2.05	0.21	52.16
β_3	3	3.21	0.31	57.08
θ_1	0.8	0.88	0.22	7.89
θ_2	-0.5	-0.48	0.2	6.51

Model 2: State Dependent with AR(1) Errors				
Parameter	True Value	Mean	Std	IF
β_1	-1	-0.83	0.59	1.55
β_2	-2	-1.82	0.24	37.05
β_3	1	0.87	0.16	21.91
θ	-0.8	-0.98	0.29	8.32
θ	0.9	0.85	0.05	9.39

Model 4: Latent Dependent with IID Errors				
Parameter	True Value	Mean	Std	IF
β_1	-1	-0.77	0.2	4.28
β_2	-1	-0.85	0.18	4.23
β_3	1	0.69	0.15	3.64
θ_1	-0.8	-0.65	0.06	2.78

Model 4: Latent Dependent with AR(1) Errors				
Parameter	True Value	Mean	Std	IF
β_1	1	0.63	0.25	4.14
β_2	-2	-1.85	0.2	30.83
β_3	-1	-1.01	0.14	21.29
θ	-0.6	-0.61	0.02	46.02
θ	0.8	0.61	0.07	37.19

Table 1: Posterior mean (MEAN), standard deviation (STD) and inefficiency factor (IF) of the model parameters in Simulation Study 1

along with the Inefficiency factors. Inefficiency factor is computed as $\text{IF} = 1+2\sum_{l=1}^{L}\rho(l)\frac{(L-l)}{L}$, $\rho(l)$ is the sample autocorrelation at lag l and summation is truncated at values L at which the correlation becomes smaller than 0.1. In all the simulation examples, the algorithms were able to recover the true parameter values. The trace plots and Inefficiency factors show good convergence of the MCMC chain. One thing to note here is that the persistence of the latent z on its lag is attenuated towards zero. However, high inefficiency factors for all the parameters and instability of the MCMC algorithm revealed that Model 4 might suffer from the identification of parameters. We investigate the issue of identification by performing additional simulation examples covered in section 4.3.

4.2. Simulation Study 2: State-dependent model with AR Errors and Block Sampling

As discussed in section 2.2, sampling from a multivariate truncated normal distribution is difficult, especially when the dimension is high. In this simulation study, We utilize the block sampling approach provided in Algorithm 2 to compare the efficiency gains from blocking the latent variable. The data was simulated from the model 7 with one state-dependent lags and AR(1) errors ($J = 1$, $p = 1$), which we label as AR(1)-AR(1) case. We generate time series data of sample size 250 with $\beta = (-1, -2, 1)$, $\theta = 0.60$, $\theta = 0.8$. We assume diffuse priors of the model coefficient with $\beta \sim \mathcal{N}_3(\mathbf{0}, 100I)$, $\theta \sim \mathcal{N}(0, 100)$ and $\theta \sim \mathcal{N}(0, 100)\mathcal{I}_{\mathcal{S}_\theta}$. In total, We draw 12500 MCMC samples including a burn-in sample of 2500 draws. The results from the simulation study are reported in the table below.

Table 2 reports the result from simulation study 1 for the state-dependent model. For this

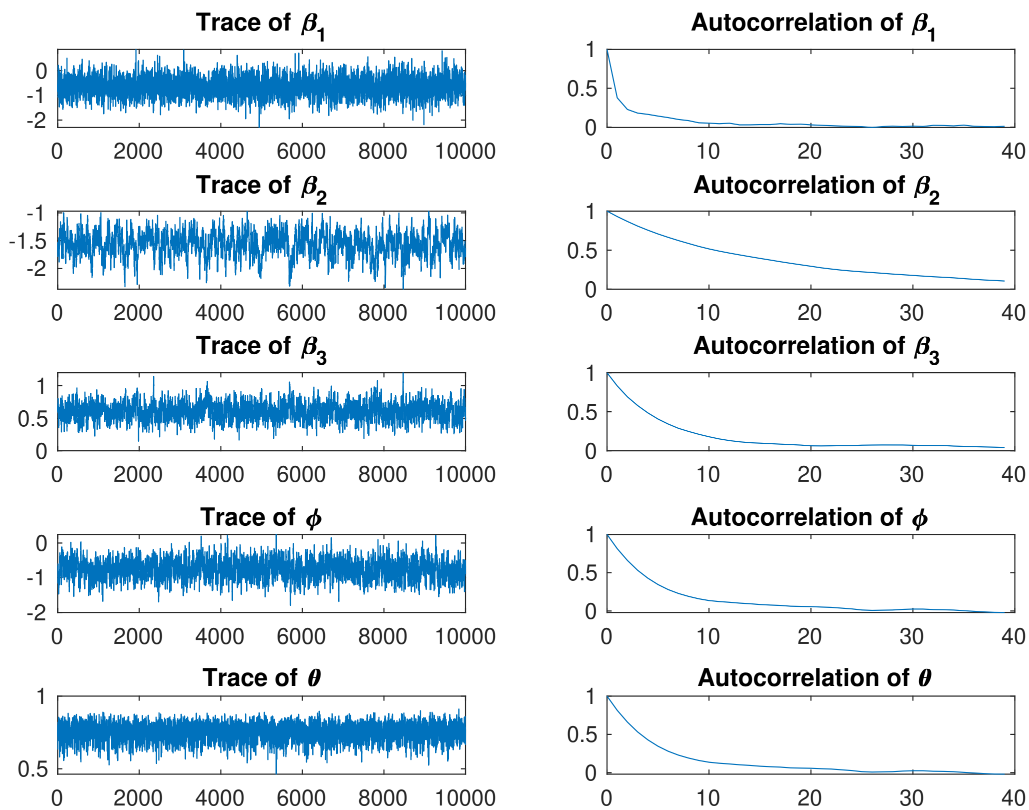

Figure 1: Trace plots and MCMC sample autocorrelation of parameters in Simulation Study 2-Sampling in a Block of 10

model, We implement the block sampling of latent z with a block size of 1, 10, 25, 50, 125, and 250. For all the cases except the sampling in a block of 25's, the MCMC algorithm recovers the true model parameters (β, θ). The posterior mean of θ is remarkably close to the true parameter of the error-generating process, highlighting the success of the algorithm. For all the examples except block sampling in 25, the algorithm shows good convergence. For brevity, We report the trace plot of the MCMC draws for sampling in a block of 50. As shown in figure 1, the trace plots of the MCMC draws show good convergence. To monitor the performance of the MCMC chain, We also compute the sample inefficiency factor as in Chib and Jeliazkov (2006) Inefficiency factor is computed as $\text{IF} = 1 + 2\sum_{l=1}^{L} \rho(l)\frac{(L-l)}{L}$, $\rho(l)$ is the sample autocorrelation at lag l and summation is truncated at values L at which the correlation becomes smaller than 0.1. As can be seen from the results table, sampling the latent data in blocks can improve the convergence of overall MCMC chains (Inefficiency factors are better for a block of 10s, the 50s, 125s, and 250s). However, we see poor convergence for

Geweke's Method					Sampling in a block of 50's				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	-1	-0.83	0.59	1.55	β_1	-1	-0.94	0.65	1.78
β_2	-2	-1.82	0.24	37.05	β_2	-2	-2.65	0.56	125.13
β_3	1	0.87	0.16	21.91	β_3	1	0.93	0.19	48.88
θ	-0.8	-0.98	0.29	8.32	θ	-0.8	-1.1	0.31	12.64
θ	0.9	0.85	0.05	9.39	θ	0.9	0.87	0.04	30.82
Sampling in a block of 10's					**Sampling in a block of 125's**				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	-1	-0.65	0.34	2.59	β_1	-1	-0.83	0.61	1.4
β_2	-2	-1.57	0.21	20.28	β_2	-2	-1.89	0.25	29.46
β_3	1	0.61	0.13	7.57	β_3	1	0.9	0.17	22.68
θ	-0.8	-0.75	0.27	6.73	θ	-0.8	-1.06	0.31	10.05
θ	0.9	0.75	0.05	2.42	θ	0.9	0.86	0.05	6.54
Sampling in a block of 25's					**Sampling in a block of 250's**				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	-1	-1.21	0.66	37.56	β_1	-1	-0.84	0.6	1.7
β_2	-2	-4.69	1.47	177.81	β_2	-2	-1.86	0.27	38.63
β_3	1	0.79	0.17	14.93	β_3	1	0.89	0.17	25.81
θ	-0.8	-0.84	0.33	19.49	θ	-0.8	-0.99	0.3	10.89
θ	0.9	0.87	0.04	54.13	θ	0.9	0.86	0.05	9.91

Table 2: Posterior mean (MEAN), standard deviation (STD) and inefficiency factor (IF) of the State-dependent model parameters in Simulation Study 2

the block of the 25's and 50's. Indeed, for the block sampling in a block of 25, the posterior mean of β_2 is far off from the truth, due to poor convergence of the algorithm.

4.3. Simulation Study 3: Latent-dependent model with AR errors

In this simulation study, We explore the issues with model parameter identification in the latent-dependent model with AR errors. We generate 300 observation from model 15 with $\beta = (1, 1, -1)$ and various combinations of *theta* and θ. We assume diffuse priors of the model coefficient with $\beta \sim \mathcal{N}_3(\mathbf{0}, 100I)$, $\theta \sim \mathcal{N}(0, 100)\mathcal{I}_{\mathcal{S}_\theta}$. Similar to simulation study 1, We draw 12500 MCMC samples including a burn-in of 2500 samples, which are dropped from the computation of posterior moments.

The results from the simulation study are presented in table 3 below. Although, the algorithm recovers the true parameter values in simulation examples. With real data sets, the inefficiency factors could be high due to potential interaction of the two AR coefficients θ

Model 4: Simulation example 1					Model 4: Simulation example 2				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	1	1.08	0.15	6.89	β_1	1	1.12	0.14	7.52
β_2	1	0.99	0.15	9.73	β_2	1	0.93	0.13	8
β_3	-1	-1.03	0.14	9.49	β_3	-1	-0.98	0.13	8.82
ϕ	0.5	0.51	0.06	4.33	ϕ	0.65	0.6	0.04	3.36
θ	0.3	0.17	0.12	4.93	θ	-0.3	-0.33	0.14	7.68
Model 4: Simulation example 3					Model 4: Simulation example 4				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	1	0.87	0.14	5.01	β_1	1	1.25	0.15	7.96
β_2	1	0.81	0.14	6.15	β_2	1	0.96	0.13	7.81
β_3	-1	-0.8	0.13	6.17	β_3	-1	-1.02	0.13	8.28
ϕ	0.7	0.67	0.07	8.95	ϕ	0.7	0.63	0.04	3.29
θ	0.5	0.09	0.07	8.41	θ	0.4	-0.38	0.13	6.07
Model 4: Simulation example 5					Model 4: Simulation example 6				
Parameter	True Value	Mean	Std	IF	Parameter	True Value	Mean	Std	IF
β_1	1	1.34	0.17	6.84	β_1	1	1.05	0.14	6.45
β_2	1	1	0.13	8.56	β_2	1	0.96	0.15	11.26
β_3	-1	-1.13	0.14	9.07	β_3	-1	-0.94	0.13	10.36
ϕ	0	0.12	0.07	6.55	ϕ	0.7	0.69	0.04	2.73
θ	0.4	0.18	0.15	8.45	θ	0	-0.12	0.15	11.46

Table 3: Posterior mean (MEAN), standard deviation (STD) and inefficiency factor (IF) of the State-dependent model parameters in Simulation Study 3

and θ on past latent variable z_{t-1} and error terms ε_t. The problem is both z_{t-1} and ε are not observed and affect z_t additively. To further explore the problem, We derive the analytical expression of unconditional variance of the latent variable z (see appendix for an expression of 5 by 5 covariance matrix). Interchanging θ and θ does retain some of the covariance structure. One possible way to overcome this challenge is to introduce more variability in the response data. In section 6, We discuss possible extensions of the model other discrete choice data models.

5. Empirical Application

Macroeconomic fluctuations or business cycles are characterized by periods of economic expansion and contraction. These fluctuations affect the household's investment and consumption decisions; businesses' production and risk management; central bank's and government's policy decisions. Studying business cycles and accurately forecasting business cycles are, thus, of significant importance to economists. Unsurprisingly, there exists a vast literature on the history of business cycles and models for predicting these fluctuations. The earliest study of business cycles dates back to the mid-eighteenth century to French statistician

Clement Juglur (cite literature on history and forecasting). However, despite rich forecasting methods and relevant data, the predictive accuracy of the proposed model has arguably been subpar (Out sample prediction pseudo R-squared less than 0.3). One plausible explanation could be that these models have not accounted for persistence in unanticipated shocks and economic activities driving the business cycles. Chauvet and Potter (2005) proposed a dynamic probit model with lagged latent variables and heteroskedastic variances, however, their model does not incorporate serial correlations. Kauppi and Saikkonen (2008) built on the former to incorporate past probabilities of recessions but imposed independent error structures. The methodology proposed in this paper bridges this gap by allowing dynamics in the business cycle through dependence on past states, dynamics in leading financial indicators, and dynamics through serially correlated error structures. Note, the models nest the existing dynamic probit model as special cases.

In this paper, We utilize the monthly NBER-based indicator of recessionary and expansionary periods. The indicator is a binary variable, with a value 1 indicating a recession following the peak through the trough, and 0 indicating business cycle expansion. The predictive ability of financial variables such as yield curve (or term-spread) forecasting business cycles has been well established in the literature (see Estrella and Mishkin (1998)). In addition, We include the GZ credit spread measure developed by Gilchrist and Zakrajsek credit spread measure (GZ spread) as the leading indicator of business cycles. They show that the predictive ability of the credit spread is due to excess bond premium (EBP), which is a measure of the marginal effect of GZ spread on default risk. As in (Gilchrist and Zakrajšek, 2012), We redefine the binary response variable as $y_{t,t+12} = 1$ if there is an NBER-dated recession at some point between month t and month t+12, capturing recessionary periods in the one-year rolling window. Hence, $Prob(y_{t,t+12} = 1)$ captures the likelihood of a recession during the next year at time t. The preliminary regression result shown in the table below compares the predictive ability of excess bond spreads versus the GZ spread in the simple state-dependent model presented in 2.1.

The regression equations for the proposed models are :

$$
\begin{aligned}
y_t &= \mathbb{1}\left\{\beta_1 + \beta_2 EBP_t + \beta_3 TS_t + \theta_1 y_{t-1} + \varepsilon_t\right\}, \qquad \varepsilon_t \sim \mathcal{N}(0,1) \\
y_t &= \mathbb{1}\left\{\beta_1 + \beta_2 EBP_t + \beta_3 TS_t + \theta_1 y_{t-1} + \varepsilon_t\right\}, \qquad \varepsilon_t \sim AR(1) \\
y_t &= \mathbb{1}\left\{\beta_1 + \beta_2 EBP_t + \beta_3 TS_t + \theta z_{t-1} + \varepsilon_t\right\}, \qquad \varepsilon_t \sim \mathcal{N}(0,1) \\
y_t &= \mathbb{1}\left\{\beta_1 + \beta_2 EBP_t + \beta_3 TS_t + \theta z_{t-1} + \varepsilon_t\right\}, \qquad \varepsilon_t \sim AR(1)
\end{aligned}
$$

State Dependent and Uncorrelated error- GZ Spread					
	Mean	Std	95 % CI		IF
Constant	-1.86	0.29	-2.46	-1.31	11.93
GZ Spread	0.26	0.15	-0.02	0.58	14.76
Term Spread	-0.73	0.16	-1.05	-0.43	58.25
y_{t-1}	4.1	0.37	3.42	4.85	56.87
Log Likelihood =-48.57					
State Dependent and Uncorrelated error- EB Premium					
	Mean	Std	95 % CI		IF
Constant	-1.4	0.2	-1.79	-1.03	11.33
Excess Bond Spread	0.8	0.32	0.22	1.47	21.33
Term Spread	-0.75	0.13	-1.02	-0.5	30.71
y_{t-1}	3.89	0.33	3.26	4.54	29.01
Log Likelihood =-46.18					

Table 4: Dynamic Probit Regression model with GZ spread vs Excess Bond Premium.

The table 5 reports the posterior moments and inefficiency factors from the abovementioned regression equations. The estimated parameter values have expected signs, an increase excess bond spread increases the default risk and hence probability of a recession. Similarly, the slope of the yield curve negatively affects the probability of a recession. The posterior mean of ϕ are positive and highly significant in the state dependent models. This implies if the economy is currently in recession, it increases the likelihood of a recession in next period, as expected. Unsurprisingly, the AR coefficient in the latent dependent model are also positive and highly significant, implying that past probabilities of recession affects the future probabilities, confirming the results from past studies. Our latent dependent model are closely related to Chauvet and Potter (2005), hence it is useful to compare our results to their findings. Posterior means for AR coefficient on latent dependent model in 5 are very close to their estimated value of 0.8, confirming their findings.We also find that, AR coefficients on the error terms are also positive and significant for the latent dependent model, suggesting that it is important to account for the serial correlation in errors. This result is also supported by the fact that lthe latent dependent with correlated errors have the lowest log likelihoof among the estimated models. It should be noted that the due to presence of serial correlation, exact likelihood of the model can not be computed and hence needs to be appoximated. Jeliazkov and Lee (2010) provides several mcmc methods to compute the approximate likelihood. However, in this paper, we utlize the quasi Monte carlo approach proposed in Botev (2017).

State Dependent and Uncorrelated errors						Latent Dependent and Uncorrelated errors					
	Mean	Std	Lower	Upper	IF		Mean	Std	Lower	Upper	IF
Constant	-1.4	0.2	-1.79	-1.03	11.33	Constant	0.24	0.13	0.01	0.5	14.24
Excess Bond Premium	0.8	0.32	0.22	1.47	21.33	Excess Bond Premium	0.81	0.21	0.43	1.26	17.46
Term Spread	-0.75	0.13	-1.02	-0.5	30.71	Term Spread	-0.54	0.12	-0.8	-0.33	36.65
y_{t-1}	3.89	0.33	3.26	4.54	29.01	z_{t-1}	0.82	0.04	0.74	0.87	27.78
Log Likelihood =- -46.1754						Log Likelihood =- -46.1754					
State Dependent and Correlated error						Latent Dependent and Correlated error					
	Mean	Std	Lower	Upper	IF	Post	Mean	Std	Lower	Upper	IF
Constant	-2.4	0.46	-3.68	-1.83	76.43	Constant	0.7	0.32	0.15	1.4	28.05
Excess Bond Premium	1.06	0.43	0.4	2.09	45.35	Excess Bond Premium	1.61	0.55	0.7	2.82	73.27
Term Spread	-0.36	0.13	-0.62	-0.09	10.5	Term Spread	-1.42	0.36	-2.15	-0.78	104.56
y_{t-1}	3.57	0.46	2.63	4.46	13.41	z_{t-1}	0.87	0.03	0.81	0.92	35.53
θ	0.35	0.33	-0.35	0.86	120.76	θ	0.65	0.15	0.28	0.86	100.12
Log Likelihood =-50.48						Log Likelihood =-39.55					

Table 5: Posterior moments and confidence intervals computed using 10000 MCMC samples with a burnin of 2000 samples. The Log likelihood values are approximated using Quasi-Monte Carlo methods from Botev (2017)

6. Future works and extensions

As discussed in section 4.3, adding more variability to the response data can improve the performance of the proposed algorithm for the latent dependent model with correlated error structures. Censored (Tobit) data adds more variability to the observed responses as uncensored component of the data are observed and continuous. Hence applying the proposed methodology to Tobit type data is a natural extension of the paper. Dueker (2005) proposed a VAR apporach to jointly model the dynamics in observed discrete responses and continuous data. However, they impose independent error structures in the model. The state dependent and latent dependent model equivalent for the censored data can be written as

- Observed dependence

$$\begin{aligned} z_t &= x_t'\beta + \phi y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \theta\varepsilon_{t-1} + \nu_t \\ y_t &= \mathbb{1}\left\{z_t\right\} z_t \end{aligned} \tag{23}$$

- Latent dependence

$$\begin{aligned} z_t &= x_t'\beta + \phi z_{t-1} + \varepsilon_t \\ \varepsilon_t &= \theta\varepsilon_{t-1} + \nu_t \\ y_t &= \mathbb{1}\left\{z_t\right\} z_t \end{aligned} \tag{24}$$

6.1. Future Works: Marginal Likelihood

Computing the marginal likelihood for state and latent dependent models will provide a way to formally compare the predictive performance of the proposed state and latent dependent model specifications. Marginal likelihood for any general model M_i given the parameter space (Θ) can be obtained using the Bayes‘ theorem as

$$m(y|M_i) = \frac{f(y|\Theta, M_i)\pi(\Theta|M_i)}{\pi(\Theta|y, M_i)} \tag{25}$$

where the numerator is the product of the likelihood and the prior, whereas the denominator is the posterior distribution of the parameters θ. The marginal density provides a way to compare the model across the model space M. In order to compare the model we calculate the marginal density at a particular ordinate say θ^*, chosen to be at a high density point under the posterior density. The marginal likelihood estimated at θ^* can be written as

$$\hat{m}(y|M_i) = \frac{f(y|\Theta^*, M_i)\hat{\pi}(\Theta^*|M_i)}{\hat{\pi}(\Theta^*|y, M_i)} \tag{26}$$

Taking the log and suppressing the conditionality on M_i we get,

$$\log(\hat{m}(y)) = \log(f(y|\Theta^*)) + \log(\hat{\pi}(\Theta^*)) - \log(\hat{\pi}(\Theta^*|y)) \tag{27}$$

The first term on the right side of the equation (27) correspond to the likelihood of the data. For the model specification in section 2.1, the likelihood is readily available, however, evaluating likelihood of the other specifications require multivariate integral of the latent data density over regions corresponding to observed response data. Exact computation of these integrals are not always feasible and hence needs to be approximated. Jeliazkov and Lee (2010) provides a comprehensive review of several simulated likelihood approaches that can be utilized to approximate the likelihood. The second term is the prior density and is readily available for the Bayesian model. Finally, citetChib-1995 provides a framework to estimate the posterior ordinate term $\pi(\Theta^*|y)$ from the MCMC simulations.

7. Conclusion

In this paper, we propose a Bayesian estimation procedure for lagged state and latent data-dependent dynamic probit models with correlated error structures. The dynamics are controlled by both lagged dependence and dynamics in errors. The paper utilizes latent data augmentation to derive posterior conditionals and propose Gibbs sampling for estimation of the models. To sample from the truncated multivariate normal distribution, propose an efficient block sampling procedure that reduces the correlation of draws. Truncated multivariate distribution in the individual blocks is sampled using exponential minimax tilting methods that produce exact identical and independent draws. The performance of the proposed algorithms are evident through multiple Monte Carlo simulation studies. The algorithms recover the true parameters underlying the dynamics in the observed data for both the state and latent-dependent models. The empirical application on foecasting business cycle highlights the utility of the model. We find that the accounting for serial correlation in the latent dependent model significantly improves the model fit and hence the predictive performance.

Appendix .1. ***Conditional Densities in State Dependent model***

In this appendix, We derive the conditional posteriors of the state dependent model parameters. The conditional posteriors of z, $\gamma = (\beta', \theta')'$ are tractable and follow standard distributions. However, θ is sampled using the independence chain Metropolis-Hastings algorithm. The derivations below follow the ordering as presented in Algorithm 2.

(1) Full conditional density of latent data z is proportional to $\prod_{t=1}^{T} \mathbb{1}\left\{z_t \in \mathcal{B}_t\right\} \mathcal{N}_T(z|W\gamma, \Omega)$. Hence, conditionally on (y, γ, θ), z follows multivariate truncated normal distribution. However, since sampling from high dimensional truncated normal is inefficient, z is sampled in a block. Consider a $\frac{T}{B}$ dimensional b$^{\text{th}}$ block of z, z_b

$$\pi(z_b|y, z_{\backslash b}, \gamma, \theta) \sim \mathcal{TN}_{T/B}(z|\mu_b, \Omega_b)$$

where $\mu_b = W_b\gamma + \Omega'_{b\neg b}\Omega^{-1}_{\neg b\neg b}(z_{\backslash b} - W_{\neg b}\gamma)$ and $\Omega = \Omega_{bb} - \Omega'_{b\neg b}\Omega^{-1}_{\neg b\neg b}\Omega_{b\neg b}$.
$\Omega^{-1}_{\neg b\neg b}$ is a $\frac{T-B}{B}$ by $\frac{T-B}{B}$ matrix obtained from Ω by eliminating rows and columns of b$^{\text{th}}$ block and $\Omega_{b\neg b}$ is obtained by eliminating the rows of b$^{\text{th}}$ block.

(2) The conditional posterior $\pi(\gamma|y, z, \theta)$ is proportional to $\pi(\gamma) \times f(z|\gamma, \theta)$

$$\begin{aligned}
\pi(\gamma|y, z, \theta) &\propto \exp\left\{\frac{-(\gamma - \gamma_0)\Gamma_0^{-1}(\gamma - \gamma_0)}{2}g\right\} \exp\left\{\frac{-(z - W\gamma)'(z - W\gamma)}{2}\right\} \\
&\propto \exp\left\{\frac{-\gamma\Gamma_0^{-1}\gamma - \gamma W'W\gamma + \gamma\Gamma_0^{-1}\gamma_0 + \gamma W'z}{2}\right\} \\
&\propto \exp\left\{\frac{-(\gamma - \hat{\gamma})\hat{\Gamma}^{-1}(\gamma - \hat{\gamma})}{2}\right\}
\end{aligned}$$

where, $\hat{\Gamma} = (\Gamma_0^{-1} + W'W)$ and $\hat{\gamma} = \hat{\Gamma}(\Gamma_0^{-1}\gamma_0 + W'z)$

(3) The conditional posterior for the autoregressive coefficients of the error term θ is proportional to $f(z|\gamma, \theta)\pi(\theta)\mathcal{I}_{s_\theta}$. In order to derive the full conditional, We utilize the alternatively representation of $f(z|\gamma, \theta)$ as

$$\begin{aligned}
\pi(\theta|y, z, \gamma) &\propto f(z_1, \ldots, z_p|\gamma, \theta) \prod_{t=p+1}^{T} f(z_t|z_{t-1}, \ldots, z_{t-p}, \gamma, \theta)\pi(\theta)\mathcal{I}_{s_\theta} \\
\pi(\theta|y, z, \gamma) &\propto \exp\left\{\frac{-(e_1\Omega_\theta^{-1}e_1)}{2}\right\} \prod_{t=p+1}^{T} \exp\left\{\frac{-(\varepsilon_t - \sum_{i=1}^{p}\theta_i\varepsilon_{t-i})^2)}{2}\right\} \pi(\theta)\mathcal{I}_{s_\theta} \\
\pi(\theta|y, z, \gamma) &\propto \exp\left\{\frac{-(e_1\Omega_\theta^{-1}e_1)}{2}\right\} \exp\left\{\frac{-(e - E\theta)'(e - E\theta)}{2}\right\} \pi(\theta)\mathcal{I}_{s_\theta}
\end{aligned}$$

$$\pi(\theta|y,z,\gamma) \propto \exp\left\{\frac{-(e_1\Omega_\theta^{-1}e_1)}{2}\right\}\exp\left\{\frac{-\theta' E' E\theta - \theta\Theta_0^{-1}\theta + \theta' E e + \theta\Theta_0^{-1}\theta_0}{2}\right\}\mathcal{I}_{s_\theta}$$

$$\pi(\theta|y,z,\gamma) \propto \exp\left\{\frac{-(e_1\Omega_\theta^{-1}e_1)}{2}\right\}\exp\left\{\frac{-(\theta-\hat{\theta})\hat{\Theta}^{-1}(\theta-\hat{\theta})}{2}\right\}\mathcal{I}_{s_\theta}$$

where $\hat{\Theta} = (\Theta^{-1} + E'E)^{-1}$ and $\hat{\theta} = \hat{\Theta}(\Theta_0^{-1}\theta_0 + E'e)$

Appendix .2. ***Conditional densities in latent dependent model: IID Errors***

In this appendix, We derive the posterior conditional distribution presented in the 3.

(1) The full conditional density of latent data z, $\pi(z|y,\beta,\phi,\theta \propto f(y|z)f(z|\beta,\phi,\theta)$.

$$\pi(z_t|y,\beta,\phi) \sim \mathcal{TN}(A^{-1}X\beta, (A'A)^{-1})$$

where the matrix A depends on ϕ and is given by equation 3.1.

(2) The conditional posterior $\pi(\beta|y,z,\theta)$ is proportional to $\pi(\beta)\times f(z|\beta,\theta)$

$$\begin{aligned}
\pi(\beta|y,z,\theta) &\propto \exp\left\{\frac{-(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)}{2}\right\}\exp\left\{\frac{-(z-A^{-1}X\beta)'\,(A'A)\,(z-A^{-1}X\beta)}{2}\right\}\\
&\propto \exp\left\{-\frac{\beta' B_0^{-1}\beta + \beta' X' X\beta - \beta' B_0^{-1}\beta_0 - \beta' X'(Az)}{2}\right\}\\
&\propto \exp\left\{-\frac{(\beta-\hat{\beta})\hat{B}^{-1}(\beta-\hat{\beta})}{2}\right\}
\end{aligned}$$

where, $\hat{B} = (B_0^{-1} + X'X)$ and $\hat{\beta} = \hat{B}(B_0^{-1}\beta_0 + X'(Az))$

(3) The conditional posterior for the autoregressive coefficients of the error term θ is proportional to $f(z|\beta,\theta)\pi(\theta)\mathcal{I}_{s_\theta}$.

$$\begin{aligned}
&\pi(\theta|y,z,\gamma) \propto f(z|\beta,\theta)\pi(\theta)\mathcal{I}_{s_\theta}\\
&\pi(\theta|y,z,\gamma) \propto \exp\left\{-\frac{(z-X\beta-L\theta)'(z-X\beta-L\theta)}{2}\right\}\exp\left\{-\frac{(\theta-\theta_0)'\theta_0^{-1}(\theta-\theta_0)}{2}\right\}\\
&\pi(\theta|y,z,\gamma) \propto \exp\left\{\frac{-\theta' L' L\theta - \theta\theta_0^{-1}\theta + \theta' L(z-X\beta) + \theta\theta_0^{-1}\theta_0}{2}\right\}\mathcal{I}_{s_\theta}\\
&\pi(\theta|y,z,\gamma) \propto \exp\left\{\frac{-(\theta-\hat{\theta})\hat{\Theta}^{-1}(\theta-\hat{\theta})}{2}\right\}\mathcal{I}_{s_\theta}
\end{aligned}$$

where $\hat{\theta} = (\theta_0^{-1} + L'L)^{-1}$ and $\hat{\theta} = \hat{\theta}(\theta_0^{-1}\theta_0 + L'(z-X\beta))$

References

Botev, Z. (2017), "The Normal Law Under Linear Restrictions: Simulation and Estimation via Minimax Tilting," *Journal of the Royal Statistical Society – Series B*, 79, 125–148.

Chauvet, M. and Potter, S. (2005), "Forecasting Recessions Using the Yield Curve," *Journal of forecasting*, 24, 77–103.

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

Chib, S. and Jeliazkov, I. (2006), "Inference in semiparametric dynamic models for binary longitudinal data," *Journal of the American Statistical Association*, 101, 685–700.

De Jong, R. M. and Woutersen, T. (2011), "Dynamic Time Series Binary Choice," *Econometric Theory*, 27, 673–702.

Dueker, M. (2005), "Dynamic Forecasts of Qualitative Variables: a Qual VAR Model of US Recessions," *Journal of Business & Economic Statistics*, 23, 96–104.

Eichengreen, B., Watson, M. W., and Grossman, R. S. (1985), "Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model," *The Economic Journal*, 95, 725–745.

Estrella, A. and Mishkin, F. S. (1998), "Predicting US Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, 80, 45–61.

Fernández-Villaverde, J., Gordon, G., Guerrón-Quintana, P., and Rubio-Ramírez, J. F. (2015), "Nonlinear Adventures at the Zero Lower Bound," *Journal of Economic Dynamics and Control*, 57, 182–204.

Gelfand, A. E., Smith, A. F., and Lee, T.-M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems using Gibbs Sampling," *Journal of the American Statistical Association*, 87, 523–532.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series Models," in *Latent Variables in Socioeconomic Models*, eds. A. S. Aigner and D. J. Goldberger, pp. 365–383, North Holland, Amsterdam.

Geweke, J. (1991), "Efficient Simulation from the Multivariate Normal and Student-t Sistributions subject to Linear Constraints and the Evaluation of Constraint Probabilities," in *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pp. 571–578, Citeseer.

Geweke, J. (1996), "Bayesian inference for linear models subject to linear inequality constraints," *Modelling and Prediction Honoring Seymour Geisser*, pp. 248–263.

Gilchrist, S. and Zakrajšek, E. (2012), "Credit Spreads and Business Cycle Fluctuations," *American economic review*, 102, 1692–1720.

Hu, L. and Phillips, P. C. (2004), "Dynamics of the Federal Funds Target Rate: a Nonstationary Discrete Choice Approach," *Journal of Applied Econometrics*, 19, 851–867.

Jacobs, P. A. and Lewis, P. A. (1978), "Discrete Time Series Generated by Mixtures. I: Correlational and runs properties," *Journal of the Royal Statistical Society: Series B (Methodological)*, 40, 94–105.

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

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

Kauppi, H. and Saikkonen, P. (2008), "Predicting US Recessions with Dynamic Binary Response Models," *The Review of Economics and Statistics*, 90, 777–791.

Mintz, O., Currim, I. S., and Jeliazkov, I. (2013), "Information processing pattern and propensity to buy: An investigation of online point-of-purchase behavior," *Marketing Science*, 32, 716–732.

Park, J. Y. and Phillips, P. C. (2000), "Nonstationary binary choice," *Econometrica*, 68, 1249–1280.

Robert, C. P. (1995), "Simulation of Truncated Normal Variables," *Statistics and computing*, 5, 121–125.

Stock, J. H. and Watson, M. W. (2006), "Forecasting with Many Predictors," *Handbook of economic forecasting*, 1, 515–554.

Vossmeyer, A. (2014), "Determining the Proper Specification for Endogenous Covariates in Discrete Data Settings," in *Bayesian Model Comparison,Advances in Econometrics*, pp. 223–247, Emerald Group Publishing Limited.

Yu, J. W. and Tian, G. L. (2011), "Efficient Algorithms for Generating Truncated Multivariate Normal Distributions," *Acta Mathematicae Applicatae Sinica, English Series*, 27, 601.