Multivariate ordinal regression for multiple repeated measurements

Laura Vana-Gür

Institute of Statistics and Mathematical Methods in Economics
TU Wien, Vienna, Austria

April 4th, 2024

TECHNISCHE UNIVERSITÄT WIEN

Motivation

- Panel data: Multiple ordinal outcomes correlated both in the cross-section and in time arise in a variety of applications: medicine, finance, ...
- R package **mvord** (Hirk, Hornik, and Vana, 2020) implements multivariate ordinal regression models where the outcomes can be correlated in the cross-section or in time (using different error structures such as the stationary distribution of an AR(1) process or general correlation matrices).
- In this work, we extend the framework to incorporate both longitudinal and cross-sectional dependence.

Outline

1. Model
2. Simulation
3. Empirical analysis
4. Wrap-up

Model: General setup

General setup:

- $y_{i,t}^{j}$ an ordinal observation
- $i \in \{1, \dots, n\}$ – subject index
- $t \in \{1, 2, \dots, T\}$ – time index among all equidistant time points,
- $j \in \{1, \dots q\}$ – outcome index

Latent variable representation:

$$y_{i,t}^{j} = r \Leftrightarrow \theta_{t,r-1}^{j} < \tilde{y}_{i,t}^{j} \leq \theta_{t,r}^{j}, \quad r \in \{1, \dots, K_j\},$$

where $-\infty \equiv \theta_{t,0}^{j} < \theta_{t,1}^{j} < \dots < \theta_{t,K_j-1}^{j} \equiv \infty$.

Model: Underlying regression model

$$\tilde{y}^j_{i,t} = (\boldsymbol{x}^j_{i,t})^\top \beta^j_t + \epsilon^j_{i,t}$$

- $\boldsymbol{x}^j_{i,t}$ is p-dimensional vector of time- and outcome-specific covariates
- β^j_t is a time- and outcome-specific vector of coefficients
- $\epsilon^j_{i,t}$ is an error term of subject i for outcome j in time t.

Model: Structure of the errors

We consider an auto-regressive structure on the q-dimensional error terms $\boldsymbol{\epsilon}_{i,t}$:

$$\boldsymbol{\epsilon}_{i,t} = \Psi \boldsymbol{\epsilon}_{i,t-1} + \Sigma_t^{1/2} \boldsymbol{u}_{i,t}, \qquad u_{i,t}^{j} \overset{iid}{\sim} F, \tag{1}$$

where

- $\Psi = \text{diag}(\psi_1, \psi_2, \ldots, \psi_q)$ is a diagonal matrix of persistence parameters for each outcome j with $|\psi_j| < 1$ (to ensure stationarity).
- $\Sigma_t = \Sigma_t^{1/2}(\Sigma_t^{1/2})^\top$ captures the cross-sectional correlation among the different outcomes at time t conditional on $\boldsymbol{\epsilon}_{i,t-1}$.

Model: Multivariate probit link

- A common approach is to assume that $u^j_{i,t} \sim N(0,1)$ and $\Sigma_t = \Sigma$. We assume Σ is a general correlation matrix.
- Then, the stationary distribution of $\boldsymbol{\epsilon}_i^* = (\boldsymbol{\epsilon}_{i,1}^\top, \boldsymbol{\epsilon}_{i,2}^\top, \dots, \boldsymbol{\epsilon}_{i,T}^\top)^\top$ is

$$\boldsymbol{\epsilon}_i^* = MVN_{qT}(\mathbf{0}, \Sigma^*), \quad \Sigma^* = \begin{pmatrix} \tilde{\Sigma} & (\Psi\tilde{\Sigma})^\top & (\Psi^2\tilde{\Sigma})^\top & \dots & (\Psi^{T-1}\tilde{\Sigma})^\top \\ \Psi\tilde{\Sigma} & \tilde{\Sigma} & (\Psi\tilde{\Sigma})^\top & \dots & (\Psi^{T-2}\tilde{\Sigma})^\top \\ \Psi^2\tilde{\Sigma} & \Psi\tilde{\Sigma} & \tilde{\Sigma} & \dots & (\Psi^{T-3}\tilde{\Sigma})^\top \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Psi^{T-1}\tilde{\Sigma} & \dots & \dots & \Psi\tilde{\Sigma} & \tilde{\Sigma} \end{pmatrix}.$$

- Here $\tilde{\Sigma}$ is the unconditional variance of the multivariate AR(1) process in Equation (1): $vec(\tilde{\Sigma}) = (I - \Psi \otimes \Psi)^{-1} vec(\Sigma)$.

Model: Multivariate logit link

- Specify a model where the $\boldsymbol{\epsilon}_i^* \sim \mathcal{L}_{qT}(\mathbf{0}, \Sigma^*)$ (multivariate logistic distribution of O'Brien and Dunson, 2004).
- Given that $\mathcal{L}_{qT}(\mathbf{0}, \Sigma^*)$ is well approximated by $MVT_{qT}(\mathbf{0}, \gamma\Sigma^*, \nu)$ for $\gamma = \pi^2(\nu - 2)/(3\nu)$ and $\nu \approx 8$ (O'Brien and Dunson, 2004), we specify:

$$\boldsymbol{\epsilon}_i^* \sim MVT_{qT}(\mathbf{0}, \gamma\Sigma^*, \nu), \qquad \nu = 8.$$

- Using results in Virolainen, 2021, this implies

$$u_{i,t}^j \sim t(0, 1, \nu + q),\ \boldsymbol{\epsilon}_{i,0} \sim MVT_q(\mathbf{0}, \gamma\tilde{\Sigma}, \nu),\ \Sigma_t = \frac{\nu - 2 + \boldsymbol{\epsilon}_{i,t-1}^\top (\gamma\tilde{\Sigma})^{-1} \boldsymbol{\epsilon}_{i,t-1}}{\nu - 2 + q}\gamma\Sigma.$$

Model: Likelihood function

In the complete case, the likelihood is the product of n $q \times T$-dimensional integrals:

$$L(\delta; Y, X) = \prod_{i=1}^{n} P\Bigg(\bigcap_{\substack{j \in 1,\dots,q \\ t \in \{1,\dots,T\}}} \{y^j_{i,t} = r^j_{i,t}\} | X_i \Bigg) = \int_{D_i} f_{qT}(\epsilon^*_i; \boldsymbol{\delta}, X^*_i) d^{qT} \epsilon^*_i. \qquad (2)$$

where $D_i = \prod_{t \in \{1,\dots,T\}} \prod_{j \in 1,\dots,q} (\theta^j_{t, r^j_{i,t}-1}, \theta^j_{t, r^j_{i,t}})$ is a Cartesian product (here $r^j_{i,t}$ denotes the observed ordinal class for subject i, time t and outcome j) and f_{qT} is the multivariate density of the error terms.

Model: Pairwise likelihood estimation

We approximate $L(\boldsymbol{\delta}; Y, X)$ by a pairwise likelihood, given by the product of the bivariate probabilities corresponding to all pairs in $\boldsymbol{y}_i^* = (\boldsymbol{y}_{i,1}^\top, \boldsymbol{y}_{i,2}^\top, \ldots, \boldsymbol{y}_{i,T}^\top)^\top$:

$$PL(\boldsymbol{\delta}; Y, X) = \prod_{i=1}^{n} \prod_{k=1}^{(q\cdot T)-1} \prod_{l=k+1}^{q\cdot T} PL_i^{(k,l)}(\boldsymbol{\delta}; Y, X),$$

$$PL_i^{(k,l)}(\boldsymbol{\delta}; Y_i, X_i) = P\Big((\boldsymbol{y}_i^*)_k = (\boldsymbol{r}_i)_k, (\boldsymbol{y}_i^*)_l = (\boldsymbol{r}_i)_l | X_i^*\Big).$$

One option is to only consider (more informative) pairs which lie close in time:

$$PL_i^{(k,l)}(\boldsymbol{\delta}, c; Y_i, X_i), = \Big[P\Big((\boldsymbol{y}_i^*)_k = (\boldsymbol{r}_i)_k, (\boldsymbol{y}_i^*)_l = (\boldsymbol{r}_i)_l | X_i\Big)\Big]^{1(l_t - k_t \leq c)}$$

Model: Inference

- Under regularity conditions, for a fixed value of c,

$$\sqrt{n}(\hat{\boldsymbol{\delta}}_{PL}(c) - \boldsymbol{\delta}) \sim MVN(\mathbf{0}, H(\boldsymbol{\delta}, c)^{-1} V(\boldsymbol{\delta}, c) H(\boldsymbol{\delta}, c)^{-1}), n \to \infty$$

  where the following estimators can be used:

$$\hat{H}(\boldsymbol{\delta}, c) = \sum_{i=1}^{n} \sum_{k=1}^{(q\cdot T)-1} \sum_{l=k+1}^{q\cdot T} \left(\frac{\partial \log PL_i^{(k,l)}(\boldsymbol{\delta}, c; Y_i, X_i)}{\partial \boldsymbol{\delta}} \right) \left(\frac{\partial \log PL_i^{(k,l)}(\boldsymbol{\delta}, c; Y_i, X_i)}{\partial \boldsymbol{\delta}} \right)^\top$$

$$\hat{V}(\boldsymbol{\delta}, c) = \sum_{i=1}^{n} \left(\sum_{k=1}^{(q\cdot T)-1} \sum_{l=k+1}^{q\cdot T} \frac{\partial \log PL_i^{(k,l)}(\boldsymbol{\delta}, c; Y_i, X_i)}{\partial \boldsymbol{\delta}} \right) \left(\sum_{k=1}^{(q\cdot T)-1} \sum_{l=k+1}^{q\cdot T} \frac{\partial \log PL_i^{(k,l)}(\boldsymbol{\delta}, c; Y_i, X_i)}{\partial \boldsymbol{\delta}} \right)^\top .$$

- Model comparison can be performed using information criteria such as the composite likelihood AIC or BIC (Varin and Vidoni, 2005)

Simulation 1: Different correlation settings

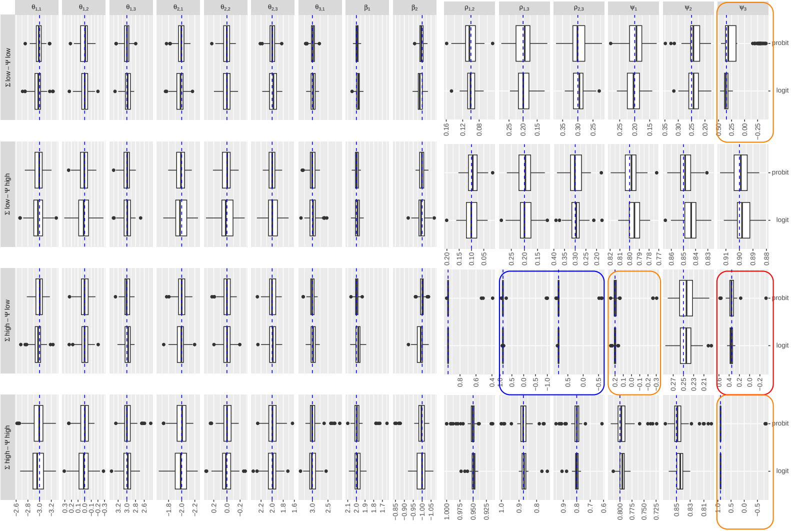
- moderate $n = 1000$ subjects,
- $q = 3$ multiple outcomes with $K_1 = 4$, $K_2 = 4$, $K_3 = 2$ and $T = 10$ time points,
- $\boldsymbol{\theta}^1 = (-\infty, -3, 0, 3, \infty)$, $\boldsymbol{\theta}^2 = (-\infty, -2, 0, 2, \infty)$, $\boldsymbol{\theta}^3 = (-\infty, 3, \infty)$
- $p = 2$ covariates from $N(0, 1)$ and $\boldsymbol{\beta} = (2, -1)^\top$.
- Error structure:

$$\Sigma_{\text{low}} = \begin{pmatrix} 1.000 & 0.100 & 0.200 \\ 0.100 & 1.000 & 0.300 \\ 0.200 & 0.300 & 1.000 \end{pmatrix}, \qquad \Sigma_{\text{high}} = \begin{pmatrix} 1.000 & 0.950 & 0.875 \\ 0.950 & 1.000 & 0.800 \\ 0.875 & 0.800 & 1.000 \end{pmatrix}$$

$$\Psi_{\text{low}} = \begin{pmatrix} 0.200 & 0 & 0 \\ 0 & 0.250 & 0 \\ 0 & 0 & 0.350 \end{pmatrix}, \qquad \Psi_{\text{high}} = \begin{pmatrix} 0.800 & 0 & 0 \\ 0 & 0.850 & 0 \\ 0 & 0 & 0.900 \end{pmatrix}$$

- $M = 100$ repetitions.

Simulation 1: Different correlation settings



Simulation 2: Different values of lag parameter c

- Similar setting as before, only for **probit**, $M = 100$ repetitions, $n = 200$ subjects

$$\Sigma = \begin{pmatrix} 1.000 & 0.100 & 0.500 \\ 0.100 & 1.000 & 0.900 \\ 0.500 & 0.900 & 1.000 \end{pmatrix}, \qquad \Psi = \begin{pmatrix} 0.800 & 0 & 0 \\ 0 & 0.500 & 0 \\ 0 & 0 & 0.200 \end{pmatrix}.$$

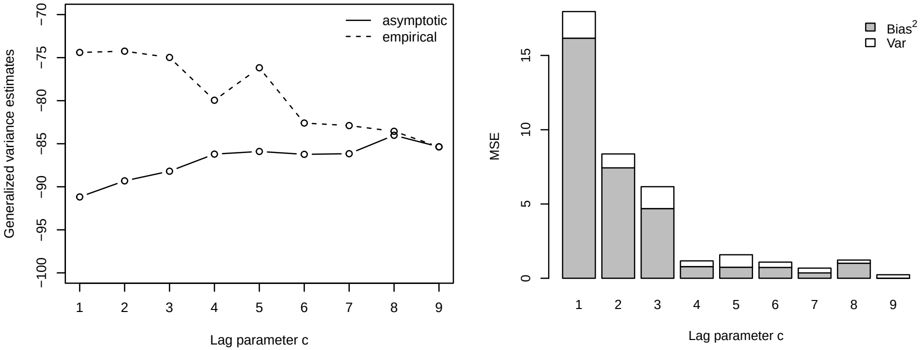
- Compute log generalized asymptotic estimate variance for $c = 1, \ldots, T-1$

$$\hat{g}(c) = \frac{1}{M}\sum_{m=1}^{M} \log\det\left(\hat{H}(\hat{\boldsymbol{\delta}}_{PL}^{(m)}(c), c)^{-1}\hat{V}(\hat{\boldsymbol{\delta}}_{PL}^{(m)}(c), c)\hat{H}(\hat{\boldsymbol{\delta}}_{PL}^{(m)}(c), c)^{-1}\right)$$

- Compute MSE for $c = 1, \ldots, T-1$, $\bar{\boldsymbol{\delta}}_{PL}(c) = \sum_{m=1}^{M} \hat{\boldsymbol{\delta}}_{PL}^{(m)}(c)/M$

$$MSE(c) = \underbrace{(\bar{\boldsymbol{\delta}}_{PL}(c) - \boldsymbol{\delta})^\top(\bar{\boldsymbol{\delta}}_{PL}(c) - \boldsymbol{\delta})}_{\text{bias}^2} + \underbrace{\frac{1}{M}\sum_{m=1}^{M}(\hat{\boldsymbol{\delta}}_{PL}^{(m)}(c) - \bar{\boldsymbol{\delta}}_{PL}(c))^\top(\hat{\boldsymbol{\delta}}_{PL}^{(m)}(c) - \bar{\boldsymbol{\delta}}_{PL}(c))}_{\text{variance}}.$$

Simulation 2: Different values of lag parameter c

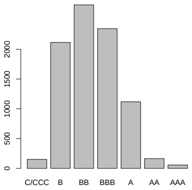


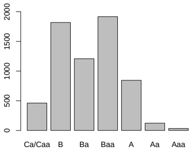
Empirical analysis: Data

- Collection of exchange-listed North American firms observed over the period 2003–2013 ($T = 11$).
- Issuer credit ratings from S&P and Moody's
- Default information obtained from UCLA-LoPucki Bankruptcy Research Database and the Mergent issuer default file
- $p = 7$ firm-level and market variables are built from the Compustat/CRSP databases.

Empirical analysis: Descriptives and model fit

- $n = 1519$ firms
- Imbalanced panel of **ratings** (77% S&P, 57% Moody's ratings) and **years** (around 1/3 of firms observed for all time points).
- Default rate: $< 1\%$.





- We estimate outcome-specific coefficients and thresholds.
- We up-sample the defaulted companies to have 50% failed vs 50% non-failed companies in the sample. After up-sampling, we achieve an overall default rate of 11.98%.

Model comparison and take-aways

	CLAIC	CLBIC
Proposed model	979045.65	993631.26
Longitudinal model	1007943.90	1022511.19
Cross-sectional model	1031555.63	1046438.48
Model with iid errors	1038281.48	1053091.83

Main takeaways from proposed model

- High persistence in ratings and default process
- High correlation among the ratings, moderate correlation between ratings and default.
- Lower thresholds estimated for S&P compared to Moody's in the speculative grades $\Rightarrow$ Moody's more conservative in the speculative grade regions.

Conclusion

- We propose a multivariate ordinal regression model which accounts for dependence between **repeated** and **multiple** ordinal measurements by ...
- ... a **multivariate AR(1)** structure on the errors of underlying process.
- Simulation study take-aways:
    - Pairwise likelihood approach recovers parameters well (provided the unbalancedness in the classes is not extreme)
    - A bootstrapping exercise should be employed for selection of lag parameter c.
- We provide an implementation of the model as an R package **mvordflex** (Hirk and Vana, 2024), which is an extension to the existing R package **mvord**.
- Future work: allow a full matrix Ψ (currently experimental feature) and incorporate in **mvord**.

Some references

- Hirk, Rainer, Kurt Hornik, and Laura Vana (2020). "**mvord**: An R Package for Fitting Multivariate Ordinal Regression Models". In: *Journal of Statistical Software* 93.4, pp. 1–41.
- Hirk, Rainer and Laura Vana (2024). *mvordflex: Multivariate Ordinal Regression Models – Extension to Three Dimensional Panel Data*. R package version 0.0.2.
- Li, Haocheng and Y Yi Grace (2013). "A pairwise likelihood approach for longitudinal data with missing observations in both response and covariates". In: *Computational Statistics & Data Analysis* 68, pp. 66–81.
- Liu, Dungang and Heping Zhang (2018). "Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach". In: *Journal of the American Statistical Association* 113.522, pp. 845–854.
- O'Brien, Sean M and David B Dunson (2004). "Bayesian Multivariate Logistic Regression". In: *Biometrics* 60.3, pp. 739–746.
- Varin, Cristiano and Paolo Vidoni (2005). "A Note on Composite Likelihood Inference and Model Selection". In: *Biometrika* 92.3, pp. 519–528.
- Virolainen, Savi (2021). "Gaussian and Student's t Mixture Vector Autoregressive Model With Application to the Asymmetric Effects of Monetary Policy Shocks in the Euro Area". In: *arXiv preprint arXiv:2109.13648*.

Thank you!

Laura Vana-Gür
CSTAT Computational Statistics (E105-6)
TU Wien

laura.vana.guer@tuwien.ac.at
https://github.com/lauravana/mvordflex

This research was supported by funds of the Österreichische Nationalbank (Austrian Central Bank), Anniversary Fund Project number: 18482 "*Multivariate ordinal regression models for enhanced credit risk modeling*".

Model: Missing values

- **Missing values**: construct the pairwise likelihood only from the bivariate probabilities corresponding to all pairs of observed responses. This approach assumes that the missing value mechanism is **completely at random**.
- Approaches to model the missing data mechanism jointly with the observations in longitudinal models can be found in e.g., Li and Grace, 2013.

Model: Constraints on parameters I

Constraints on coefficients and threshold parameters can be set by specifying appropriate constraint matrices. Let

$$\boldsymbol{\eta}_i^{\text{upper}} = B_i^{\text{upper}}\boldsymbol{\theta}^* - X_i^*\boldsymbol{\beta}^* = Z_i^{\text{upper}}\boldsymbol{\kappa}^*, \quad \boldsymbol{\eta}_i^{\text{lower}} = B_i^{\text{lower}}\boldsymbol{\theta}^* - X_i^*\boldsymbol{\beta}^* = Z_i^{\text{lower}}\boldsymbol{\kappa}^*,$$
$$Z_i = (B_i, -X_i^*), \quad \boldsymbol{\kappa}^* = ((\boldsymbol{\theta}^*)^\top, (\boldsymbol{\beta}^*)^\top)^\top,$$

where $\boldsymbol{\theta}^* = ((\boldsymbol{\theta}_1^1)^\top, \dots, (\boldsymbol{\theta}_1^q)^\top, \dots, (\boldsymbol{\theta}_T^1)^\top, \dots, (\boldsymbol{\theta}_T^q)^\top)^\top$ and the matrices B_i^{lower} and B_i^{upper} are $(q \times T) \times (T\sum_{j=1}^q (K_j - 1))$ block diagonal binary matrices

$$B_i^{\text{upper}} = \text{diag}((\boldsymbol{b}_{i,1}^{1,\text{upper}})^\top, \dots, (\boldsymbol{b}_{i,1}^{q,\text{upper}})^\top, \dots, (\boldsymbol{b}_{i,T}^{1,\text{upper}})^\top, \dots, (\boldsymbol{b}_{i,T}^{q,\text{upper}})^\top)$$
$$B_i^{\text{lower}} = \text{diag}((\boldsymbol{b}_{i,1}^{1,\text{lower}})^\top, \dots, (\boldsymbol{b}_{i,1}^{q,\text{lower}})^\top, \dots, (\boldsymbol{b}_{i,T}^{1,\text{lower}})^\top, \dots, (\boldsymbol{b}_{i,T}^{q,\text{lower}})^\top)$$

where the vector $\boldsymbol{b}_{i,t}^{j,\text{upper}}$ has length $K_j - 1$ and contains a one in the $r_{i,t}^j$-th position if $r_{i,t}^j \in \{1, \dots, K_j - 1\}$, else zero; the vector $\boldsymbol{b}_{i,t}^{j,\text{lower}}$ has length $K_j - 1$ and contains a one in the $(r_{i,t}^j - 1)$-th position if $r_{i,t}^j \in \{2, \dots, K_j\}$, else zero.

Model: Constraints on parameters II

The probabilities in the likelihood function can then be expressed as:

$$P\Bigg(\bigcap_{\substack{j \in 1,\dots,q \\ t \in \{1,\dots,T\}}} \{y_{i,t}^{j} = r_{i,t}^{j}\} \Bigg) = F_{qT}(Z_i^{\text{upper}} \boldsymbol{\kappa}^* | \Sigma^*, \dots) - F_{qT}(Z_i^{\text{lower}} \boldsymbol{\kappa}^* | \Sigma^*, \dots).$$

Assuming that $\tilde{\boldsymbol{\kappa}} = (\tilde{\boldsymbol{\theta}}^\top, \tilde{\boldsymbol{\beta}}^\top)^\top$ is the reduced $(h \times 1)$ vector of thresholds and coefficients to be estimated, the linear predictors can be rewritten as:

$$\boldsymbol{\eta}_i = Z_i C \tilde{\boldsymbol{\kappa}}$$

where C is a contrast matrix of dimension $(T \times \sum_{j=1}^{q}(K_j - 1) + qTp) \times h$.

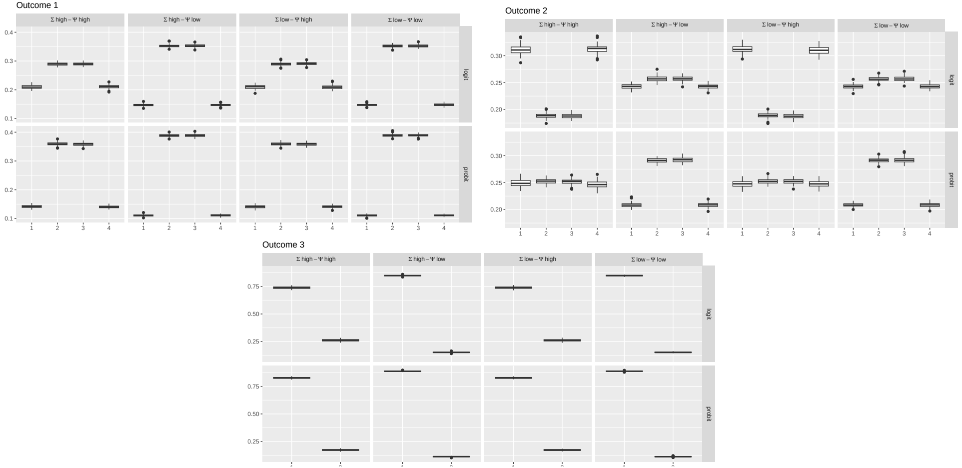
Model: Constraints on parameters III

For example, the C matrix for a model where all thresholds should be constant over time and one set of regression coefficients should be employed for all t and j would be of dimension $(T\sum_{j=1}^{q}(K_j-1)+qTp)\times(\sum_{j=1}^{q}(K_j-1)+p)$:

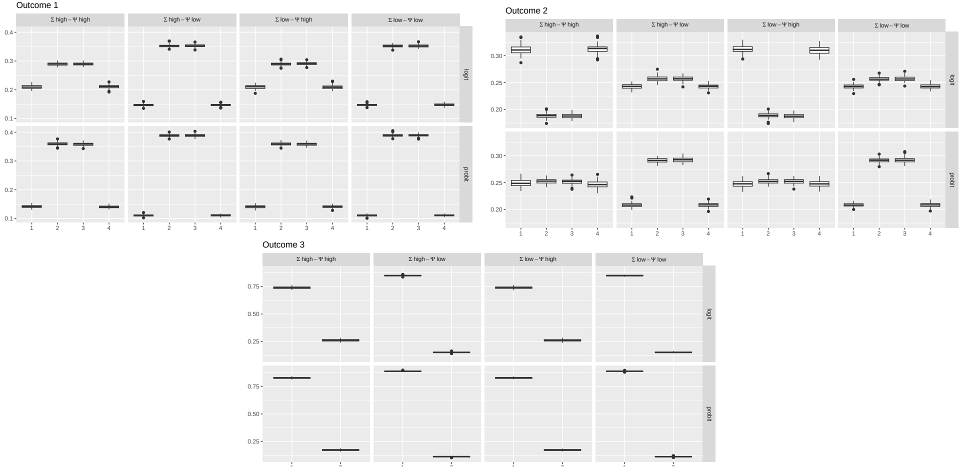
$$C=\begin{pmatrix} \underbrace{(1,\dots,1)^\top}_{T\text{ times}} \otimes I_{\sum_{j=1}^{q}(K_j-1)} & \mathbf{0}_{T\sum_{j=1}^{q}(K_j-1)\times p} \\ \mathbf{0}_{(q\cdot T\cdot p)\times p} & \underbrace{(1,\dots,1)^\top}_{q\cdot T\text{ times}} \otimes I_p \end{pmatrix},$$

where $I_.$ denotes the identity matrix, $\mathbf{0}$ is the zero matrix and $\otimes$ denotes the Kronecker product.

Simulation 1: Outcome distribution



Simulation 1: Outcome distribution



Simulation 1: Setting Σ high and Ψ high

Table: This table presents simulation results based on 100 repetitions, $n = 1000$, $T = 10$, $q = 3$ for the correlation setting Σ high and Ψ high.

		Multivariate probit link				Multivariate logit link			
	True	Mean Est	APB	Mean Asym SE	SD Sample	Mean Est	APB	Mean Asym SE	SD Sample
$\theta_{1,1}$	−3.0000	−2.9767	0.78%	0.0632	0.1076	−2.9719	0.94%	0.0771	0.1278
$\theta_{1,2}$	0.0000	0.0005	-	0.0430	0.0734	0.0130	-	0.0425	0.1098
$\theta_{1,3}$	3.0000	2.9792	0.69%	0.0624	0.1032	2.9956	0.15%	0.0776	0.1164
$\theta_{2,1}$	−2.0000	−1.9907	0.46%	0.0597	0.0949	−1.9807	0.96%	0.0666	0.1219
$\theta_{2,2}$	0.0000	−0.0025	-	0.0516	0.0738	0.0096	-	0.0501	0.1127
$\theta_{2,3}$	2.0000	1.9867	0.66%	0.0595	0.0934	1.9998	0.01%	0.0672	0.1091
$\theta_{3,1}$	3.0000	2.9719	0.94%	0.0846	0.1379	2.9979	0.07%	0.0959	0.1520
β_1	2.0000	1.9866	0.67%	0.0338	0.0437	1.9878	0.61%	0.0442	0.0431
β_2	−1.0000	−0.9932	0.68%	0.0219	0.0254	−0.9938	0.62%	0.0265	0.0304
$\rho_{1,2}$	0.9500	0.9523	0.24%	0.0044	0.0086	0.9495	0.05%	0.0055	0.0048
$\rho_{1,3}$	0.8750	0.8758	0.09%	0.0178	0.0229	0.8705	0.52%	0.0215	0.0217
$\rho_{2,3}$	0.8000	0.8027	0.34%	0.0195	0.0275	0.8003	0.04%	0.0228	0.0190
ψ_1	0.8000	0.7955	0.57%	0.0087	0.0111	0.7955	0.57%	0.0114	0.0078
ψ_2	0.8500	0.8470	0.36%	0.0069	0.0081	0.8451	0.58%	0.0088	0.0071
ψ_3	0.9000	0.8454	6.07%	0.0065	0.2467	0.8975	0.28%	0.0087	0.0063

Simulation 1: Setting Σ high and Ψ low

Table: This table presents simulation results based on 100 repetitions, $n = 1000$, $T = 10$, $q = 3$ for the correlation setting Σ high and Ψ low.

		Multivariate probit link				Multivariate logit link			
	True	Mean Est	APB	Mean Asym SE	SD Sample	Mean Est	APB	Mean Asym SE	SD Sample
$\theta_{1,1}$	−3.0000	−2.9928	0.24%	0.0369	0.0721	−2.9696	1.01%	0.0469	0.0915
$\theta_{1,2}$	0.0000	0.0018	-	0.0201	0.0655	0.0028	-	0.0210	0.0809
$\theta_{1,3}$	3.0000	2.9966	0.11%	0.0365	0.0685	2.9804	0.65%	0.0467	0.0867
$\theta_{2,1}$	−2.0000	−1.9939	0.31%	0.0292	0.0652	−1.9818	0.91%	0.0355	0.0825
$\theta_{2,2}$	0.0000	0.0013	-	0.0205	0.0634	0.0017	-	0.0205	0.0802
$\theta_{2,3}$	2.0000	1.9979	0.10%	0.0294	0.0706	1.9922	0.39%	0.0360	0.0858
$\theta_{3,1}$	3.0000	3.0006	0.02%	0.0399	0.0738	2.9835	0.55%	0.0490	0.0931
β_1	2.0000	1.9959	0.21%	0.0225	0.0174	1.9797	1.01%	0.0302	0.0317
β_2	−1.0000	−0.9993	0.07%	0.0152	0.0120	−0.9892	1.08%	0.0191	0.0222
$\rho_{1,2}$	0.9500	0.9458	0.44%	0.0054	0.0039	0.9495	0.05%	0.0053	0.0036
$\rho_{1,3}$	0.8750	0.8580	1.95%	0.0158	0.0192	0.8768	0.21%	0.0178	0.0107
$\rho_{2,3}$	0.8000	0.7888	1.39%	0.0207	0.0204	0.8007	0.09%	0.0307	0.0139
ψ_1	0.2000	0.1917	4.15%	0.0221	0.0219	0.2000	0.02%	0.0284	0.0171
ψ_2	0.2500	0.2448	2.06%	0.0219	0.0165	0.2449	2.03%	0.0276	0.0162
ψ_3	0.3500	0.3529	0.82%	0.0484	0.0467	0.3583	2.37%	0.0632	0.0374

Simulation 1: Setting Σ low and Ψ high

Table: This table presents simulation results based on 100 repetitions, $n = 1000$, $T = 10$, $q = 3$ for the correlation setting Σ low and Ψ high.

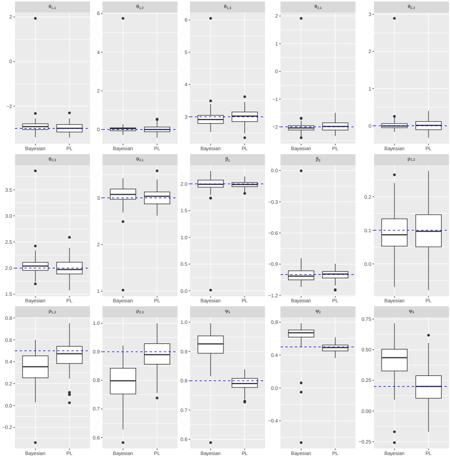
		Multivariate probit link				Multivariate logit link			
	True	Mean Est	APB	Mean Asym SE	SD Sample	Mean Est	APB	Mean Asym SE	SD Sample
$\theta_{1,1}$	−3.0000	−2.9846	0.51%	0.0582	0.0847	−2.9710	0.97%	0.0673	0.1217
$\theta_{1,2}$	0.0000	0.0053	-	0.0438	0.0729	0.0133	-	0.0413	0.1086
$\theta_{1,3}$	3.0000	3.0014	0.05%	0.0583	0.0789	2.9959	0.14%	0.0671	0.1163
$\theta_{2,1}$	−2.0000	−1.9852	0.74%	0.0570	0.0745	−1.9901	0.50%	0.0604	0.1112
$\theta_{2,2}$	0.0000	0.0090	-	0.0521	0.0751	0.0016	-	0.0487	0.1134
$\theta_{2,3}$	2.0000	2.0022	0.11%	0.0571	0.0749	1.9861	0.69%	0.0603	0.1126
$\theta_{3,1}$	3.0000	2.9981	0.06%	0.0865	0.1015	2.9820	0.60%	0.0875	0.1527
β_1	2.0000	1.9953	0.23%	0.0260	0.0228	1.9867	0.67%	0.0325	0.0317
β_2	−1.0000	−0.9976	0.24%	0.0175	0.0150	−0.9956	0.44%	0.0204	0.0252
$\rho_{1,2}$	0.1000	0.0938	6.21%	0.0297	0.0231	0.1003	0.27%	0.0354	0.0301
$\rho_{1,3}$	0.2000	0.1993	0.37%	0.0369	0.0308	0.1959	2.07%	0.0501	0.0334
$\rho_{2,3}$	0.3000	0.2991	0.29%	0.0348	0.0332	0.2973	0.89%	0.0474	0.0319
ψ_1	0.8000	0.7986	0.17%	0.0079	0.0066	0.7951	0.62%	0.0098	0.0075
ψ_2	0.8500	0.8482	0.21%	0.0062	0.0055	0.8442	0.68%	0.0075	0.0068
ψ_3	0.9000	0.8986	0.15%	0.0065	0.0061	0.8973	0.31%	0.0085	0.0064

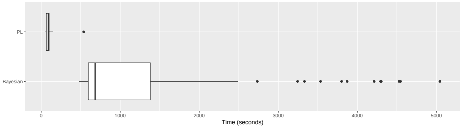
Simulation 1: Setting Σ low and Ψ low

Table: This table presents simulation results based on 100 repetitions, $n = 1000$, $T = 10$, $q = 3$ for the correlation setting Σ low and Ψ low.

		Multivariate probit link				Multivariate logit link			
	True	Mean Est	APB	Mean Asym SE	SD Sample	Mean Est	APB	Mean Asym SE	SD Sample
$\theta_{1,1}$	−3.0000	−2.9849	0.50%	0.0353	0.0648	−2.9704	0.99%	0.0446	0.0933
$\theta_{1,2}$	0.0000	0.0029	-	0.0202	0.0678	0.0026	-	0.0211	0.0803
$\theta_{1,3}$	3.0000	2.9992	0.03%	0.0351	0.0702	2.9822	0.59%	0.0444	0.0877
$\theta_{2,1}$	−2.0000	−1.9884	0.58%	0.0276	0.0645	−1.9776	1.12%	0.0324	0.0807
$\theta_{2,2}$	0.0000	0.0042	-	0.0206	0.0632	0.0027	-	0.0205	0.0763
$\theta_{2,3}$	2.0000	1.9954	0.23%	0.0276	0.0662	1.9852	0.74%	0.0326	0.0841
$\theta_{3,1}$	3.0000	2.9886	0.38%	0.0403	0.0758	2.9787	0.71%	0.0486	0.0884
β_1	2.0000	1.9933	0.34%	0.0187	0.0169	1.9808	0.96%	0.0254	0.0258
β_2	−1.0000	−0.9959	0.41%	0.0125	0.0109	−0.9896	1.04%	0.0156	0.0171
$\rho_{1,2}$	0.1000	0.1020	2.03%	0.0178	0.0155	0.0994	0.63%	0.0229	0.0141
$\rho_{1,3}$	0.2000	0.1991	0.47%	0.0346	0.0295	0.1954	2.30%	0.0427	0.0274
$\rho_{2,3}$	0.3000	0.2987	0.45%	0.0345	0.0300	0.2958	1.40%	0.0445	0.0241
ψ_1	0.2000	0.1979	1.06%	0.0328	0.0249	0.2039	1.94%	0.0407	0.0261
ψ_2	0.2500	0.2398	4.09%	0.0293	0.0245	0.2425	3.01%	0.0361	0.0295
ψ_3	0.3500	0.1889	46.03%	0.0709	0.2614	0.3555	1.56%	0.0859	0.0480

Simulation 3: Comparison with a Bayesian approach





Empirical application: Choice of AR(1)

We motivate the choice of AR(1) instead of AR(p) by:

- Short length of the time series.
- We extract the surrogate residuals (Liu and Zhang, 2018) from separate models with iid errors. For each firm we run ARIMA model on the residuals for each firm and each response and select the optimal autoregressive lag based on AIC. For the large majority of the firms chosen lag is 0 or 1.

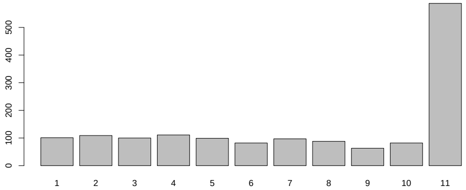


Figure: Distribution of the number of time points observed per firm in the sample.