Composite Likelihood Estimation of AR-Probit Model: Application to Credit Ratings*

Kerem Tuzcuoglu †

April 21, 2017

Abstract

In this paper, persistent discrete data are modeled by Autoregressive Probit model and estimated by Composite Likelihood (CL) estimation. Autocorrelation in the latent variable results in an intractable likelihood function containing high dimensional integrals. CL approach offers a fast and reliable estimation compared to computationally demanding simulation methods. I provide consistency and asymptotic normality results of the CL estimator and use it to study the credit ratings. The ratings are modeled as imperfect measures of the latent and autocorrelated creditworthiness of firms explained by the balance sheet ratios and business cycle variables. The empirical results show evidence for rating assignment according to Through-the-cycle methodology, that is, the ratings do not respond to the short-term fluctuations in the financial situation of the firms. Moreover, I show that the ratings become more volatile over time, in particular after the crisis, as a reaction to the regulations and critics on credit rating agencies.

Keywords: Composite likelihood, autoregressive probit, autoregressive panel probit, stability of credit ratings, through-the-cycle methodology

JEL Classification: C23, C25, C58, G24, G31

*I would like to thank Serena Ng, Jushan Bai, Bernard Salanié, Sokbae Lee, Jón Steinsson, Aysun Alp, as well as seminar participants at Columbia University, for their comments and suggestions. All errors are, of course, my own.

†Ph.D. Candidate, Economics Department, Columbia University kerem.tuzcuoglu@columbia.edu.

1 Introduction

Persistent discrete variables are extensively used in both economics and finance literature. Credit ratings, changes in the Federal Funds Target Rate, NBER recession dates, unemployment status, and school grades are just a few important examples among many. These variables have a fair amount of persistence in them: credit ratings of companies change rarely; the policy rate is usually adjusted gradually by central banks; a recession (expansion) in a quarter tends to be followed by a recession (expansion) in the next quarter. To understand the nature of these variables, one needs to take care of discreteness and persistence at the same time.

But, modeling and estimating persistent discrete data can be challenging. Incorporating time series concepts (to capture the persistence) into the nonlinear nature of discrete data might need complex models that are hard to estimate. To deal with such complications, I borrow a method – composite likelihood estimation – from statistics literature and bring it to economics where the method is not widely known. Composite likelihood (CL) estimation is a likelihood-based method that uses the partial specification of full-likelihood. CL becomes very useful especially in cases where writing or computing the full-likelihood is infeasible, yet marginal or conditional likelihoods are easier to formulate. In particular, CL can offer a fast and robust estimation for models with complex likelihood function that can be written only in terms of a large dimensional integral, which renders implementation of the full-likelihood maximization approach impractical or computationally demanding.

An interesting model with such challenging likelihood is an autoregressive probit (AR-Probit) model, where discrete (binary or categorical) data are modeled as a nonlinear function of an underlying continuous autoregressive latent process. Mathematically, an AR-Probit model can be represented as

$$\begin{aligned} y_{it}^* &= \rho y_{i,t-1}^* + \beta' x_{it} + \varepsilon_{it} \\ y_{it} &= \mathbb{1}[y_{it}^* \geq 0] \end{aligned}$$

where i represents firms, t represents time, and $\mathbb{1}(\cdot)$ represents the indicator function. The discrete y_{it} can be considered as an imperfect measure of the latent process y_{it}^*. Hence, the autoregressive property of the latent process drives the persistence in the discrete variable. But, the nonlinear dynamic dependency between y_{it} and y_{it}^* results in an intractable likelihood function with a high dimensional integral that does not have an explicit solution. Although there are methods (e.g., simulated maximum likelihood, Bayesian estimation techniques) to compute/approximate likelihoods containing integrals, they are computationally demanding. More importantly, they might become unstable and even impractical if the dimension of the integral is large – in the empirical part, my model has 55 dimensional integral. In this paper, I extend the above model into various directions and use composite likelihood approach to estimate the complex likelihood of the AR-Probit model.

Lindsay [1988] defined composite likelihood as a likelihood-type object formed by multiplying together individual component likelihoods, each of which corresponds to a marginal or conditional event. The merit of CL is to reduce the computational complexity so that it is possible to deal with large datasets and complex dependencies, especially when the use of standard likelihood methods are not feasible. The formal definition of composite likelihood is as follows.

Definition 1. *Let $\{f(y;\theta),\ y \in \mathcal{Y},\ \theta \in \Theta\}$ be a parametric statistical model with $\mathcal{Y} \subseteq \mathbb{R}^T$, $\Theta \subseteq \mathbb{R}^d$, $T \geq 1$ and $d \geq 1$. Consider a set of events $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}, i \in \mathcal{I}\}$, where $\mathcal{I} \subseteq \mathbb{N}$ and $\mathcal{F}$ is some sigma algebra on $\mathcal{Y}$. A composite likelihood is defined as*

$$L_C(\theta; y) = \prod_{i \in \mathcal{I}} f(y \in \mathcal{A}_i; \theta)^{w_i},$$

where $f(y \in \mathcal{A}_i; \theta) = f(\{y_j \in \mathcal{Y} : y_j \in \mathcal{A}_i\}; \theta)$, with $y = (y_1, \ldots, y_T)$, while $\{w_i, i \in \mathcal{I}\}$ is a set of suitable weights. The associated log-likelihood is $\mathcal{L}_C(\theta; y) = \log L_C(\theta; y)$.

The definition of composite likelihood is very general, even encompassing the full-likelihood as a special case. Hence, the definition does not tell how to formulate composite likelihood in special cases; it just states that composite likelihood is a weighted collection of likelihoods. In practice, CL is chosen as a subset of the full-likelihood. For a T-dimensional data vector y, the most common choices are marginal composite likelihood $L_C(\theta; y) = \prod_t^T f(y_t|\theta)$ and pairwise composite likelihood $L_C(\theta; y) = \prod_{t=1}^T \prod_{s \neq t} f(y_t, y_s|\theta)$ or $L_C(\theta; y) = \prod_{t=1}^{T-J} \prod_{j=1}^J f(y_t, y_{t+j}|\theta)$. In this sense, CL is considered to be pseudo-likelihood, quasi-likelihood, and partial-likelihood by several authors (Besag [1974], Cox [1975]). Compared to the traditional maximum likelihood estimator, the CL method may be statistically less efficient, but consistency, asymptotic normality, and significantly faster computation are among the appealing properties of the CL estimator. Moreover, it can be more robust to model misspecification compared to ML estimation or simulation methods since one needs only correct sub-models in CL approach.

AR-Probit is clearly not the only model to estimate persistent discrete data – though it is more akin to standard time series models. One can consider replacing the lag of the unobserved variable y^*_{t-1} by the lag of the observed outcome y_{t-1}. In the literature, this model is called Dynamic Probit. This is a state-dependence model whereas AR-Probit is closer to habit-persistence models. Dynamic Probit models are useful when the discrete variable is an important policy variable since the past discrete observation y_{t-1} creates a jump in the continuous latent variable. On the other hand, AR-Probit models are useful when the discrete variable is an imperfect measure of the underlying dynamic state variable.

A good example where AR-Probit can be preferred would be credit ratings where the rating assigned to a firm is an imperfect measure of firm's underlying creditworthiness evaluated by a credit rating agency. A firm does receive AA rating not because it was assigned AA previously, but because the financial situation of the firm is persistent and yields a similar level of credit conditions as previously. Another example is NBER recession dates. Many papers (e.g., Dueker [1997], Kauppi and Saikkonen [2008]) use the past recession dummy variable to predict its future values. Here we should consider this question: "Is the economy in a recession because it was in a recession in the previous period, or is it because the underlying state of the economy is persistent and was in a bad state previously?". A case can be made that, the second argument explains the recessions better. From this point of view, the AR-Probit seems a better option to model persistent discrete data in some cases. Moreover, Beck et al. [2001] argued that AR-Probit yields often superior results than Dynamic Probit. Regarding estimation, maximum likelihood can easily be applied to Dynamic Probit model (de Jong and Woutersen [2011]) since the discrete data have Markovian property. However, in AR-Probit model, the discrete data are not Markovian anymore, and the likelihood contains integrals to be computed or approximated – which can be computationally challenging. With composite likelihood, in particular, with modeling only pairwise likelihoods, one bypasses the need for simulations and still achieve an estimator with desirable asymptotic properties.

This paper contributes to two strands of literature. First, it contributes to the composite likelihood literature by providing the consistency and asymptotic normality results of the CL estimator in the AR-Probit model. CL is gaining substantial attention in the statistics field but has relatively little coverage in econometrics and other related fields. To be precise, there have been just a handful of papers that used composite likelihood in the economics and finance literature. Varin and Vidoni [2008] showed how pairwise likelihood can be applied, from simple models, like AR(1) model with a dynamic latent process, to more complex ones, like AR-Tobit model. That paper can be considered an introduction of composite likelihood approach to econometrics literature. Afterwards, Engle et al. [2008] and Pakel et al. [2011] both utilized CL estimator in multivariate GARCH models to avoid inverting large-dimensional covariance matrices. Bhat et al. [2010] compared the performance of simulated maximum likelihood (SML) to CL in a Panel Probit model with autocorrelated error structure and found that CL needs much less computational time and provides more stable estimation (see Reusens and Croux [2016] for an application of this model to sovereign credit ratings). CL is attractive to estimate DSGE models, in particular with stochastic singularities (Qu [2015]) or misspecifications (Canova and Matthes [2016]). CL can also be employed to deal with high dimensional copulas (Oh and Patton [2016] and Heinen et al. [2014]). Finally, Bel et al. [2016] use CL in a multivariate logit model and show that CL has much smaller computation time with a small efficiency loss compared to MLE. In statistics literature, Varin and Vidoni [2006] show the applicability and usefulness of CL estimation in AR-Probit model. Standard asymptotic results for CL estimation under general theory have already been presented in the literature (see Lindsay [1988], Molenberghs and Verbeke [2005], Varin et al. [2011]). However, finding the required assumptions and proving the asymptotic results of CL estimator specifically in AR-Probit models, to the best of my knowledge, is a theoretical contribution. CL, as a general class of estimators, is known to be consistent and asymptotically normal, but, in this paper, I provide the required assumptions to achieve these asymptotic results in the AR-Probit model.

Second, this paper contributes to the corporate bond ratings literature by studying the stability of the ratings in a model with firm specific variables. It is known that there is a trade-off between accuracy and stability of credit ratings (Cantor and Mann [2006]). More accurate ratings require more volatility in rating assignments to capture the changes in the creditworthiness of companies in a timely fashion. This paper contributes by presenting a new methodology and findings in measuring stability. In particular, to the best of my knowledge, this is the first paper at the firm-level analysis, where the rating stability is measured by a single estimated coefficient – the persistence parameter ρ. Moreover, by using time-varying coefficients (ρ_t), the rating stability changes is estimated over time. But, why is the rating stability important? The rating stability has its benefits for investors, issuers, and credit rating agencies. Moreover, rating stability is desirable to prevent pro-cyclical effects in the economy – ratings that respond to temporary information might exacerbate the situation and contribute to the market volatility. For this reason, credit rating agencies promised to assign ratings according to *Through-the-cycle* (TTC) methodology[1], which means that the ratings do not reflect short-term fluctuations, but rather indicate the long-term trustworthiness of a firm (see Altman and Rijken [2006] for more details on TTC). The literature is divided on verifying TTC rating claim by rating agencies. A branch of literature found evidence for pro-cyclical ratings, thus argues that rating agencies uses *Point-in-time* (PIT)methodology instead of TTC (Nickell et al. [2000], Bangia et al. [2002], Amato and Furfine [2004], Feng et al. [2008], Topp and Perl [2010], Freitag [2015]). On the other hand, there are others showing that rating agencies can in fact see through the cycle (Carey and Hrycay [2001], Altman and Rijken [2006], Löffler [2004, 2013], and Kiff et al. [2013]). In this paper, I provide empirical evidence for TTC rating approach by showing that during the Great Recession, rating agencies actually tried to hold the ratings stable for the first 2-3 quarters of the recession before starting downgrading the firms. Only afterward, when rating agencies realized that the changes in the credit situation of the firms are not short-term, ratings are let to be more volatile.

[1] Standard and Poor's [2002, p.41]: "The ideal is to rate *through the cycle*. There is no point in assigning high ratings to a company enjoying peak prosperity if that performance level is expected to be only temporary. Similarly, there is no need to lower ratings to reflect poor performance as long as one can reliably anticipate that better times are just around the corner."

The rest of the paper proceeds as follows. Section 2 gives an overview of the composite likelihood approach and the advantages over other estimation techniques. Section 3 introduces the Panel AR-Probit model, explains how to construct the pairwise composite likelihood, and states the theoretical asymptotic results. The last large section is dedicated to the empirical application. In that section, extensions of the baseline model are provided together with the estimation results and robustness checks. All the mathematical proofs are left to the Technical Appendix.

2 Composite Likelihood Literature

The literature on composite likelihood goes back to late 1980s, but it became popular especially after the early 2000s. The papers using CL are mostly focused on statistics, computer science, and biology to handle the estimation of very complex systems. In the economics literature, and more so in finance, CL is relatively an unknown topic. The first paper that defines composite likelihood is Lindsay [1988]. CL has its roots in the pseudo-likelihood of Besag [1974] and the partial likelihood of Cox [1975]. Varin et al. [2011] gives a thorough overview of the topic.

CL has a wide variety of applications, but I will focus on the literature involving models that have dynamic latent variables; a feature that is present in AR-Probit model. The first example is Le Cessie and Van Houwelingen [1994], where correlated binary data, even though the underlying process is not explicitly modeled, is estimated by pairwise likelihoods. Varin and Vidoni [2006], mentioned above, is the first paper that applied CL to AR-Probit. Varin and Czado [2009] offered pairwise composite likelihood in panel probit model with autoregressive error structure. This model is also used later in Bhat et al. [2010]. Some theoretical results of CL estimator in a general class of models with a dynamic latent variable (e.g., stochastic volatility, AR-Poisson) is introduced in Ng et al. [2011]. Gao and Song [2011] applied EM algorithm to composite likelihood in hidden Markov models. A dynamic factor structure in a probit model was analyzed in Vasdekis et al. [2012].

Theoretical properties of composite likelihood are closely related to pseudo-likelihoods (see Molenberghs and Verbeke [2005] for some asymptotic results in general context). Because CL comprises either marginal or conditional likelihoods which are in fact parts of the full likelihood, some nice theoretical results directly follow from the properties of the full likelihood. For instance, CL satisfies the Kullback-Leibler information inequality since log-likelihood of each conditional or marginal event ℓ_i belongs to the full likelihood, thus

$$\mathbb{E}_{\theta_0}[\ell_i(\theta)] \leq \mathbb{E}_{\theta_0}[\ell_i(\theta_0)] \quad \text{for all } \theta.$$

Kullback-Leibler inequality together with some regular mild assumptions gives consistency of the CL estimator. However, being a "miss-specified likelihood", the asymptotic variance of the CL estimator is not the inverse of the information matrix. Instead, it is in the so-called sandwich form (it also goes by the name Godambe Information in the statistics literature due to Godambe [1960]). Regarding hypothesis testing, Wald and score test statistics are standard; however (composite) likelihood ratio test statistic does not have a χ^2 distribution asymptotically. It has a non-standard asymptotic distribution in the form of a weighted summation of independent χ^2 distributions where the weights are the eigenvalues of the multiplication of the inverse Hessian and the information matrix (see Kent [1982]). Adjustments to CL ratio statistics can also be made so that one obtains asymptotically a χ^2 distribution (see Chandler and Bate [2007] and Pace et al. [2011]). Model selection can be done according to information criteria such as AIC and BIC with composite likelihoods as shown in Varin and Vidoni [2005] and Lindsay et al. [2011]. The information criterion contains the composite likelihood and a penalty term that depends on the multiplication of the inverse Hessian and the information matrix.

Composite Likelihood provides computational ease and sometimes even computational possibility of the estimation. Moreover, it is more robust than full likelihood approach since only the likelihoods that are part of the composite likelihood must be correctly modeled instead of the correctly specified full likelihood. For instance, a pairwise composite likelihood in AR-Probit model requires the correct specification of the bivariate probabilities instead of the correct specification of all dependencies of the data. However, composite likelihood comes with a cost: efficiency loss. It is hard to establish a general efficiency result for composite likelihoods. Mardia et al. [2007] show that composite conditional estimators are fully efficient in exponential families that have a certain closure property under subsetting. For instance, AR(1) model falls into this category; it is easy to show that the conditional composite likelihood $\prod_{t=1}^{T} f(y_t|y_{t-1})$ actually is the (conditional) full likelihood in AR(1) model. Lindsay et al. [2011] have a theory on optimally weighting the composite likelihood to increase the efficiency. However, they stated: "*We conclude that the theory of optimally weighted estimating equations has limited usefulness for the efficiency problem we address.*". Similarly, Harden [2013] proposed a weighting scheme for composite likelihood but the simulations showed minimal improvements regarding efficiency. There are several studies for efficiency on specific examples. For instance, Davis and Yau [2011] analyzed the efficiency loss of the CL estimator in AR(FI)MA models where both the full-likelihood and the pairwise likelihood can be computed. They find that in AR models and long-memory processes with a small integration parameter, the efficiency loss is ignorable. However, CL might have substantial efficiency loss in MA models and long-memory processes if the order of integration is high. Hjort and Varin [2008] conjectured that CL can be seen as a penalized likelihood in general Markov chain models and find that efficiency loss of CL estimator compared to ML is negligible. Joe and Lee [2009] and Varin and Vidoni [2006] find evidence that, in time series context, including only nearly adjacent pairs in the composite likelihood $\prod_{t=1}^{T-J}\prod_{j=1}^{J} f(y_t, y_{t+j})$ might have advantages over all-pairs composite likelihood $\prod_{t\neq s} f(y_t, y_s)$. The idea follows from the fact that far apart observations bring almost no information but end up bringing more noise to the estimation.

Identification of the parameters in CL is the most tricky part. So far, the literature has not been able to provide conditions which guarantee identifiability. Since CL can contain very different components of the full likelihood, it is not always clear when identification can or cannot be achieved. A very simple example helps us understand the issue. Consider an AR(1) model $y_t = \rho y_{t-1} + \sigma e_t$. If we choose marginal distribution $f(y_t) = \mathcal{N}(0, \sigma^2/(1-\rho^2))$ as CL then we cannot identify the parameters (ρ, σ) separately. However, using conditional distribution $f(y_t|y_{t-1}) = \mathcal{N}(\rho y_{t-1}, \sigma^2)$ as CL enables us to identify the parameters. Even in such an easy example, the choice of composite likelihood matters dramatically in terms of identification. In more complex models, it is not clear, in general, which sub-likelihoods should be included in the CL so that one can identify all of the parameters. For now, the identification is checked case by case until a unified theory on identification in CL literature is developed.

Composite likelihood might be relatively new in the economics literature, but its underlying idea of modeling misspecified likelihood has been used for many years under different names like pseudo-likelihood, partial-likelihood or quasi-likelihood. For instance, asymptotic theory on pseudo maximum likelihood based on exponential families is analyzed by Gourieroux et al. [1984]. Fermanian and Salanié [2004] suggests estimating parts of the full-likelihood of an autoregressive Tobit model by nonparametric simulated maximum likelihood. Molenberghs and Verbeke [2005] has a chapter on pseudo-likelihoods with applications and theoretical results. In the finance literature, Lando and Skødeberg [2002] used partial-likelihood to estimate some of the parameters from only a particular part of the likelihood function. In Duan et al. [2012], to avoid intensive numerical estimations in a forward intensity model, pseudo-likelihood is constructed with overlapping data to utilize the available data to the fuller extent.

CL is not the only estimation technique to estimate complex models where the full-likelihood contains large dimensional integral. Simulated maximum likelihood (SML) and Bayesian techniques have been the most common choices in economics and finance literature to compute these integrals. In economics, Hajivassiliou and Ruud [1994], Gourieroux and Monfort [1996], Lee [1997], Fermanian and Salanié [2004](non-parametric SML); in finance, Gagliardini and Gouriéroux [2005], Feng et al. [2008], Koopman et al. [2009], and Koopman et al. [2012] (Monte Carlo ML) can be given as examples among many papers that used SML. One concern about SML in these models is the computational complexity. In fact, Feng et al. [2008] stated that "Practitioners might, however, find this method complicated and possibly time-consuming." and "Although the SML estimators are consistent and efficient for large number of simulations, practitioners may find the procedure quite difficult and time-consuming.". Thus they offered an auxiliary estimation where the model is estimated first without dynamics, then the dynamics of the factor are estimated in a second step. Bhat et al. [2010] compared the performances of SML (GHK simulator – one of the most frequently used SML techniques) and CL estimation in a Panel Probit with correlated errors model. The number of categories for the ordered outcome ($y_{it} \in \{1, \dots, S\}$) and the time dimension are the key factors for computation times for SML. Thus they were kept at low levels. In their simulations, $N = 1000$, $T = 5$, and $S = 5$. The results show that both estimation techniques recovered the true parameters successfully, and there is almost no difference in efficiency between CL and SML. This result is interesting since CL is supposed to be less efficient than full-likelihood approach. However, SML is efficient when the number of draws tends to infinity; otherwise, the simulation error in approximating the likelihood is not negligible. If one cannot simulate a large number of times – due to computational power and time restrictions, SML also ends up being inefficient. Hence, CL and SML provide comparable estimation results, but in terms of computation times, CL is approximately 40 times faster than SML. In terms of Bayesian techniques, Chauvet and Potter [2005], Dueker [2005], McNeil and Wendin [2007], and Stefanescu et al. [2009] use Gibbs sampling in latent dynamic probit models. However, Müller and Czado [2012] showed that in such models, Gibbs sampler exhibits bad convergence properties, therefore, suggested a more sophisticated group move multigrid Monte Carlo Gibbs sampler. Yet, this proposed technique was criticized by Varin and Vidoni [2006] and Bhat et al. [2010] for increasing the computational complexity. Finally, it is worth to mention that Gagliardini and Gourieroux [2014] proposed an efficient estimator that does not require any simulation. They used Taylor approximation of the likelihood to estimate, but their theory needs the following conditions in order the approximation error to become negligible: $N \to \infty$, $T \to \infty$, and $T^\nu/N = O(1)$ for $\nu > 1$ (or $\nu > 1.5$ for stronger results). However, in their simulations and applications, they used $N = 1000$ and $T = 20$, where T is actually not large.

A final word can be said on the similarity between CL and GMM estimation technique. In GMM, the researcher should choose the orthogonality conditions to estimate the parameters. However, selecting the most informative moments is not an easy task (see Andrews [1999] for some optimality conditions). In this regard, CL is similar to GMM since the researcher should choose the collection of likelihoods which will be included in the composite likelihood. Moreover, there is no theory that tells how to choose them optimally. CL is attractive when the model is very complicated; thus, most of the time, the researcher is already limited by the model complexity or computational burden. For instance, in an AR-Probit model, one can easily model bivariate and maybe trivariate probabilities, but computing quadruple probabilities becomes complicated and reduces the attractiveness of the CL. The composite likelihood (as well as maximum likelihood) estimator can be considered a subset of the method of moment estimators. In particular, one can always choose the orthogonality conditions for GMM estimation as the score functions derived from the (composite) likelihood. In this sense, it is hard to pin down the difference between GMM and CL estimator. However, in panel data applications with strictly exogenous regressors, it is well known that the orthogonality conditions are of order T^2. In an application like the one in this paper, where $N = 516$ and $T = 55$, the number of moment conditions is extremely high. Not all moments are informative, but choosing "the best ones" among them is a hard exercise. Moreover, computing the optimal weighting matrix and taking its inverse is practically impossible. This situation results in a noisy GMM estimation whereas there is not such an issue in CL estimation since one just adds the log-likelihoods for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Simulation results for a comparison of CL versus GMM are provided in the following section after introducing the pairwise composite likelihood estimation. The results clearly favors for the CL estimation in a setting similar to the empirical application of this paper. As a result, GMM can be considered a set of estimators that contains MLE and CLE as special cases, however, in some large scale applications, it might be beneficial to use CL over GMM.

3 Panel AR-Probit Model and Pairwise Composite Likelihood

In this section, I introduce the baseline Panel AR-Probit model and construct a pairwise composite likelihood. Moreover, I state the objective function to be maximized and the assumptions needed for consistency and asymptotic normality of the resulting composite likelihood estimator. The proofs are left to the appendix.

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, let i denotes the i^{th} firm and t denotes the time. I assume that the innovations are $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, 1)$ over both i and t. The choice of normal distribution is somewhat important: with the estimation approach that is explained below, the errors should belong to a family of probability distribution that is closed under convolution. More explanation will be given at the end of the section. The variance of the innovations is assumed to be 1 in order to identify other parameters, which is a typical assumption in any probit model. The $(K \times 1)$ dimensional explanatory variables are denoted by x_{it}, and are assumed to be strictly exogenous in the sense that $f(\varepsilon_\mathbf{i}|\mathbf{x_i}) = f(\varepsilon_\mathbf{i})$, where the notation $\mathbf{z_i}$ denotes the T-dimensional vector $(z_{i1}, \ldots, z_{iT})'$. Moreover, the regressors are independent and identically distributed on the cross-section. A univariate, continuous, latent, autoregressive process y^*_{it} is generated by its lag $y^*_{i,t-1}$, x_{it} and ε_{it} in a linear relationship. Depending on the level of y^*_{it}, the univariate discrete variable y_{it} is generated. The $((K + 1) \times 1)$ dimensional parameters to be estimated are $\theta \equiv (\rho, \beta')'$. Theoretically, $|\rho| < 1$ is not required for stationarity since T is fixed. However, when T is at least moderately large, $|\rho| < 1$ is needed for empirical stability of the estimator.

The continuous variable y_{it}^* is unobserved, however the binary variable $y_{it} \in \{0, 1\}$ is observed. Hence, an autoregressive panel probit model can be written as, for $t = 1, \dots, T$,

$$y_{it}^* = \rho y_{i,t-1}^* + \beta' x_{it} + \varepsilon_{it}, \tag{1}$$

$$y_{it} = \mathbb{1}[y_{it}^* \geq 0]. \tag{2}$$

The initial condition will be defined below. The generating process of the latent autoregressive y_{it}^* is Markov, however the same is not true for the discrete value y_{it}. The variable y_{it} depends nonlinearly on the autoregressive y_{it}^*, thus y_{it} does depend not only on $y_{i,t-1}$ but also on the whole history of y_{it}, i.e., on $\{y_{i,t-1}, \dots, y_{i1}\}$. In other words, y_{it} exhibits non-Markovian property because $y_{i,t-1}$ contains only partial information – interval information – concerning y_{it}^*. Therefore, the values $\{y_{i,t-2}, \dots, y_{i1}\}$ contain additional imperfect but useful information for y_{it}. Hence, the typical Markov property in linear time series models is not valid in AR-Probit model. For this reason, one needs to integrate out y_{it}^*, which results in a T dimensional integral in the likelihood function for each individual firm i that does not have an explicit analytical solution.

$$L_i(\mathbf{y_i}|\mathbf{x_i}; \theta) = \int \cdots \int f(\mathbf{y_i}|\mathbf{y_i}^*; \theta) f(\mathbf{y_i}^*|\mathbf{x_i}; \theta) d\mathbf{y_i}^*,$$

It is not feasible to maximize $\sum_{i=1}^N \log L_i(\mathbf{y_i}|\mathbf{x_i}; \theta)$ by maximum likelihood estimation unless T is fairly small (see Matyas and Sevestre [1996]). For a very small T, one can either compute T-variate probabilities – it gets exponentially complicated as T enlarges to compute the probability of the history $\{y_{i1}, \dots, y_{iT}\}$ – or one can approximate the integrals, say by Gauss–Hermite quadrature. However, all these solutions are feasible for very small T. One can use simulation-based techniques – as well as Bayesian – to compute large dimensional integrals, however as mentioned in the previous sections, these estimation techniques are computationally demanding and might have convergence issues. Hence, composite likelihood estimation is a good alternative for panel AR-Probit models with large N and not-so-small T. In particular, a composite likelihood consisting of only pairwise dependencies will be easy to estimate since the nonlinear dependencies are reduced to a level that is easy to handle. For instance, a composite likelihood of pairs with at most J-lag distant apart can be formed by

$$\ell_i(\mathbf{y_i}|\mathbf{x_i}; \theta) = \sum_{t=1}^{T-J} \sum_{j=1}^{J} \log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta). \tag{3}$$

One could write composite likelihood of each pairs $f(y_{it}, y_{is}|\mathbf{x_i}; \theta)$ for $s \neq t$ rather than $f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta)$. However, in a time series framework, the dependency between two observations becomes negligible as they get more distant. Thus, in practice, all-pairs-likelihood might be even inferior to J-pairs likelihood in terms of estimated variance (Varin and Vidoni [2006], Joe and Lee [2009]). Before computing the pairwise probabilities in (3), it will be constructive to compute marginal probabilities.

First, since $y^*_{i,t-1}$ is not observed, I use backward substitution on latent process, that is, the current laten variable becomes a weighted sum of the past observations and innovations, where the weights are decreasing at an exponential rate. Second, the initial value should be modeled. One might assume $y^*_{io} = 0$ or y^*_{io} is drawn from its unconditional distribution[2]. However, the former is too unrealistic and the latter requires modelling a process for x_{it}. Hence, I assume that y^*_{io} is drawn from its conditional distribution, i.e., $y^*_{io} = \beta' x_{io} + \frac{1}{\sqrt{1-\rho^2}}\varepsilon_{io}$.

$$\begin{aligned} y^*_{it} &= \rho^t y^*_{io} + \sum_{k=0}^{t-1} \rho^k \beta' x_{i,t-k} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k}, \\ &= \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} + \frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{io} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k},, \end{aligned} \tag{4}$$

which implies that

$$\begin{aligned} \mathbb{E}[y^*_{it}|\mathbf{x_i}] &= \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} \\ \mathrm{Var}(y^*_{it}|\mathbf{x_i}) &= \frac{1}{1-\rho^2} \end{aligned}$$

By using (4), one can compute the marginal probability of a realization y_{it} in the following way.

$$\begin{aligned} \mathbb{P}\left(y_{it} = 0 \mid \mathbf{x_i}; \theta\right) &= \mathbb{P}\left(y^*_{it} < 0 \mid \mathbf{x_i}; \theta\right) \\ &= \mathbb{P}\left(\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} + \frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{io} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k} < 0 \,\middle|\, \mathbf{x_i}; \theta\right) \\ &= \mathbb{P}\left(\frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{io} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k} < -\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} \,\middle|\, \mathbf{x_i}; \theta\right) \\ &= \mathbb{P}\left(\frac{\frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{io} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k}}{\sqrt{\frac{1}{1-\rho^2}}} < \frac{-\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k}}{\sqrt{\frac{1}{1-\rho^2}}} \,\middle|\, \mathbf{x_i}; \theta\right) \\ &= \mathbf{\Phi}\left(-\sqrt{1-\rho^2}\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k}\right) \\ &= \mathbf{\Phi}\left(m_t(\mathbf{x_i}, \theta)\right) \end{aligned}$$

where $m_t(\mathbf{x_i}, \theta) \equiv -\sqrt{1-\rho^2}\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k}$, which can be considered as the normalized conditional mean of the latent process. Note that, the second to last equation follows since $\rho^t \varepsilon_{io} +$

[2]Note that $\mathbb{E}[y^*_{it}] = \beta' \mathbb{E}[x_{it}]/(1-\rho)$ and $\mathrm{Var}[y^*_{it}] = (\beta' \mathrm{Var}[x_{it}]\beta + 1)/(1-\rho^2)$. Thus, $y^*_{it} \sim \mathcal{N}(\mathbb{E}[y^*_{it}], \mathrm{Var}[y^*_{it}])$.

$\sqrt{1-\rho^2}\sum_{k=0}^{t-1}\rho^k\varepsilon_{i,t-k}\sim\mathcal{N}(0,1)$. As mentioned at the beginning of the section, this approach cannot be applied to any type of error distribution; one needs the distribution of the weighted infinite sum of errors to be the same distribution as that a single error term. In other words, the error distribution should be a stable distribution [3]. While normal distribution is a stable distribution, logistic distribution is not. That is, the convolution of logistic distribution does not result in a logistic distribution [4].

Next, let's compute the bivariate probability of a realization $(y_{it}, y_{i,t+j}) = (0,0)$.

$$
\begin{aligned}
&\mathbb{P}\left(y_{it}=0, y_{i,t+j}=0 \mid \mathbf{x_i};\theta\right)\\
&=\mathbb{P}\left(y_{it}^*<0,\ y_{i,t+j}^*<0 \mid \mathbf{x_i};\theta\right)\\
&=\mathbb{P}\left(\sum_{k=0}^{t}\rho^k\beta' x_{i,t-k}+\frac{\rho^t\varepsilon_{io}}{\sqrt{1-\rho^2}}+\sum_{k=0}^{t-1}\rho^k\varepsilon_{i,t-k}<0,\right.\\
&\qquad\left.\sum_{k=0}^{t+j}\rho^k\beta' x_{i,t+j-k}+\frac{\rho^{t+j}\varepsilon_{io}}{\sqrt{1-\rho^2}}+\sum_{k=0}^{t+j-1}\rho^k\varepsilon_{i,t+j-k}<0\ \middle|\ \mathbf{x_i};\theta\right)\\
&=\mathbb{P}\left(\frac{\frac{\rho^t\varepsilon_{io}}{\sqrt{1-\rho^2}}+\sum_{k=0}^{t-1}\rho^k\varepsilon_{i,t-k}}{\sqrt{\frac{1}{1-\rho^2}}}<m_t(\mathbf{x_i},\theta),\ \frac{\frac{\rho^{t+j}\varepsilon_{io}}{\sqrt{1-\rho^2}}+\sum_{k=0}^{t+j-1}\rho^k\varepsilon_{i,t+j-k}}{\sqrt{\frac{1}{1-\rho^2}}}<m_{t+j}(\mathbf{x_i},\theta)\ \middle|\ \mathbf{x_i};\theta\right)\\
&=\mathbb{P}\left(Z_1\le m_t(\mathbf{x_i},\theta),\ Z_2\le m_{t+j}(\mathbf{x_i},\theta)\ \middle|\ \mathbf{x_i};\theta\right),
\end{aligned}
\tag{5}
$$

where (Z_1, Z_2) are bivariate standard normally distributed with the correlation coefficient $r=\rho^j$.

[3](Feller [1971], page 169) Let $X, X_1, X_2, \dots$ be independent and identically distributed. The distribution is called *stable* if $\forall\ n\ \exists\ c_n>0$ and $\gamma\in\mathbb{R}$ such that $(X_1+\cdots+X_n)$ has the same distribution as $c_nX+\gamma$. The well-known stable distributions are Gaussian, Cauchy, and Lévy distributions. Note that the latter two distributions do not have even a well-defined mean. If the stable distributions are in general unknown, can we at least characterize them? The answer is yes.
(Hall et al. [2002], page 5) A random variable Z has a stable distribution with shape, scale, skewness, and location parameters $(\alpha,\sigma,\beta,\mu)$, denoted by $Z\sim S(\alpha,\sigma,\beta,\mu)$, if its log characteristics function has the form

$$
\log\mathbf{E}[e^{iuZ}]=\begin{cases} \mathrm{i}\mu u-\sigma^\alpha|u|^\alpha[1-\mathrm{i}\beta\,\mathrm{sgn}(u)\tan(\pi\alpha/2)] & \text{if } \alpha\neq 1\\ \mathrm{i}\mu u-\sigma|u|[1-\mathrm{i}\beta\,\mathrm{sgn}(u)(2/\pi)\log(u)] & \text{if } \alpha=1\end{cases}
$$

where $\alpha\in(0,2]$, $\sigma>0$, $\mu\in(-\infty,\infty)$, and $\beta\in[-1,1]$. It is easy to see that when $\alpha=2$, Z is a normal random variable with mean μ and variance $2\sigma^2$. When $\alpha=1$ and $\beta=0$, Z has a Cauchy distribution. Thus, with this characterization, one can compute the cumulative probabilities of any stable distribution at a given point; hence normal distribution is not the only option. However, computationally, the analysis will be very cumbersome.

[4](Ojo [2003]) Let $X_1, X_2, \dots, X_n$ be n iid logistic random variables so that their distribution is $F(x)=e^x/(1+e^x)$. Let $S_n=X_1+\cdots+X_n$, then, the distribution of the partial sum S_n is found to be

$$
F_n(S)=\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\frac{(-1)^{n+1}}{(n-1)!}\binom{n-1}{k}\frac{j!}{(j+k+1-n)!}\sum_{r=0}^{\infty}(-1)^{nr}r^{j+k+1-n}e^{(r+1)S}\sum_{m=0}^{k}\frac{(-1)^m k! S^{k-m}}{(k-m)!}
$$

By using the rectangle property of a bivariate distribution [5], we conclude that

$$\begin{aligned}
\mathbb{P}\left(y_{it}=0, y_{i,t+j}=0 \mid \mathbf{x_i};\theta\right) &= \mathbf{\Phi}_2\left(m_{it}(\theta), m_{i,t+j}(\theta) \,\middle|\, r(\theta)\right)\\
\mathbb{P}\left(y_{it}=1, y_{i,t+j}=0 \mid \mathbf{x_i};\theta\right) &= \mathbf{\Phi}(m_{i,t+j}(\theta)) - \mathbf{\Phi}_2\left(m_{it}(\theta), m_{i,t+j}(\theta) \,\middle|\, r(\theta)\right)\\
\mathbb{P}\left(y_{it}=0, y_{i,t+j}=1 \mid \mathbf{x_i};\theta\right) &= \mathbf{\Phi}(m_{it}(\theta)) - \mathbf{\Phi}_2\left(m_{it}(\theta), m_{i,t+j}(\theta) \,\middle|\, r(\theta)\right)\\
\mathbb{P}\left(y_{it}=1, y_{i,t+j}=1 \mid \mathbf{x_i};\theta\right) &= 1 - \mathbf{\Phi}(m_{it}(\theta)) - \mathbf{\Phi}(m_{i,t+j}(\theta)) + \mathbf{\Phi}_2\left(m_{it}(\theta), m_{i,t+j}(\theta) \,\middle|\, r(\theta)\right)
\end{aligned}$$

where $r(\theta)=\rho^j$, $m_{it}(\theta)=m_t(\mathbf{x_i},\theta)$, and $\mathbf{\Phi}_2(\cdot,\cdot|r)$ denotes the bivariate standard normal distribution with the correlation coefficient r.

3.1 Pairwise Composite Likelihood Estimator

In this subsection, the objective function, the associated estimator and the assumptions for consistency and asymptotic normality are introduced. Having found the bivariate probabilities, the objective function – pairwise composite log-likelihood– can be defined as

$$\begin{aligned}
\mathcal{L}_c(\theta|\boldsymbol{y},\boldsymbol{x}) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-J}\sum_{j=1}^{J}\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta) && (6)\\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-J}\sum_{j=1}^{J}\sum_{s_1=0}^{1}\sum_{s_2=0}^{1}\mathbb{1}(y_{it}=s_1, y_{i,t+j}=s_2)\log\mathbb{P}\left(y_{it}=s_1, y_{i,t+j}=s_2 \mid \mathbf{x_i};\theta\right), && (7)
\end{aligned}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function, $\boldsymbol{y}=(\mathbf{y_1},\ldots,\mathbf{y_N})$, and $\mathbf{y_i}=(y_{i1},\ldots,y_{iT})$. The notation is similar for $\boldsymbol{x}$. The composite likelihood estimator is found by maximizing the objective function, where Θ is the parameter space,

$$\hat{\theta}_N = \arg\max_{\theta\in\Theta}\mathcal{L}_c(\theta|\boldsymbol{y},\boldsymbol{x}). \tag{8}$$

For consistency of the estimator, the following assumptions are needed – some of them have already been mentioned in the text.

Assumption 1. *The true parameter value $\theta_0 \in \Theta \subseteq \mathbb{R}^K$, Θ is compact.*

Assumption 2. *The innovations are independent and identically distributed over i and t, that is, $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1)$.*

Assumption 3. *The covariates $\mathbf{x_i}$ are independent and identically distributed over i.*

Assumption 4. *The covariates $\mathbf{x_i}$ are strictly exogenous. Moreover, $\mathbf{E}(\mathbf{x_i}\mathbf{x_i}')$ is invertible.*

[5] For any two random variables X and Y with the bivariate cumulative distribution function G, one can write $\mathbb{P}(x_1 \le X \le x_2,\ y_1 \le Y \le y_2) = G(x_2,y_2) - G(x_1,y_2) - G(x_2,y_1) + G(x_1,y_1)$

Note that, the compactness assumption requires some prior knowledge by the econometrician about the region where the true parameter might be. Assumptions 2 and 3 are typical in panel probit models. The first part of Assumption 4 is stringent; it is not always easy to find strictly exogenous regressors, in particular in time series. For the sake of theoretical part, I will keep this assumption. One can allow for the endogeneity of the regressors if the model is transformed into a VAR-Probit model where (y_{it}^*, x_{it}) is modeled endogenously by their past values. It is an interesting model, but it is left as a future work for now. The continuity and the measurability of the objective function are easy to prove since bivariate Gaussian cumulative distribution function Φ_2 and $m_{it}(\theta)$ are all continuous and measurable functions. Thus, $\log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta)$ is continuous in θ for a given $(y_{it}, y_{i,t+j}, \mathbf{x_i})$, and is a measurable function of $(y_{it}, y_{i,t+j}|\mathbf{x_i})$ for a given θ. Also note that, since y_{it} is a measurable function of y_{it}^*, its stationarity is implied by the stationarity of y_{it}^*.

Theorem 1. *Under the assumptions (1) through (4), the composite likelihood estimator defined in (8) is consistent, i.e., $\hat{\theta}_N \to_p \theta_0$, as $N \to \infty$ and $T < \infty$.*

Since each piece of the full likelihood satisfies the Kullback-Leibler inequality, so will the chosen pieces for the composite likelihood. This property helps the estimation procedure to discriminate the true parameter value from other possible parameters. $\mathbb{E}\left[\log f(\theta_o)\right] \geq \mathbb{E}\left[\log f(\theta)\right]$ since $\mathbb{E}\left[\log \frac{f(\theta)}{f(\theta_o)}\right] \leq \log \mathbb{E}\left[\frac{f(\theta)}{f(\theta_o)}\right] = 0$. The proof for $\mathbb{E}\left[\log f(\theta_o)\right] \neq \mathbb{E}\left[\log f(\theta)\right]$, which implies that θ_o is the unique maximizer, is left to the appendix.

For asymptotic normality of the estimator, the following assumptions are needed. Note that, $\log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta)$ is twice continuously differentiable since both the univariate and bivariate cumulative normal distribution is in fact infinitely differentiable. Assumption 5 is necessary since if the true parameter is on the boundary then the resulting distribution will not be Gaussian.

Assumption 5. *The true parameter value is in the interior of the parameter space, i.e., $\theta_0 \in \mathring{\Theta}$.*

Assumption 6. $\mathbb{E}\|\mathbf{x_i}\|^4 < \infty$

The finiteness of the fourth order moment of the covariates is needed for the finiteness of the variance of the score function.

Theorem 2. *Under the assumptions (1) through (6), the composite likelihood estimator defined in (8) is asymptotically normal. The asymptotic covariance matrix is in the sandwich-form as defined below. As $N \to \infty$,*

$$\sqrt{T}(\hat{\theta} - \theta_o) \to_d \mathcal{N}\left(0, H(\theta_o)^{-1} G(\theta_o) H(\theta_o)^{-1}\right)$$

where $H(\theta) = \mathbb{E}\left[\frac{\partial^2 \ell_i(\theta)}{\partial\theta\partial\theta'}\right]$, $G(\theta) = \mathbb{E}\left[\frac{\partial \ell_i(\theta)}{\partial\theta}\frac{\partial \ell_i(\theta)}{\partial\theta'}\right]$, and $\ell_i(\theta) = \sum_{t=1}^{T-J}\sum_{j=1}^{J} \log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta)$.

The asymptotic theory on the CL estimator in AR-Probit model, conceptually, is not different than the asymptotic theory on pseudo-likelihoods (or quasi-likelihoods). However, the difficulty arises due to the nonlinearity in the parameters. The cumulative distribution function Φ is not the only source of the nonlinearity; the function m_{it} is also nonlinear in parameters – especially in ρ. This 'double' nonlinearity result in complicated derivative functions of the composite likelihood. Hence, computing the derivatives and finding bounds for them become nontrivial. Despite this extra nonlinearity, the moment conditions on the process x_t is not different than those in static model, thanks to the assumption $|\rho| < 1$. For instance, the finiteness of $|\sum_{k=0}^{\infty} \rho^k \beta' x_{t-k}| \leq \sum_{k=0}^{\infty} |\rho|^k \|\beta\| \|x_{t-k}\|$ in expectation is simply implied by the finiteness of $\|x_t\|$ in expectation since $|\rho| < 1$. The complications and the nonlinearity of the model disappear when $\rho = 0$. Thus, at any point in the proof, one can recover the conditions for static probit by imposing $\rho = 0$.

Finally, in order to compute consistent estimator of the asymptotic covariance matrix, I introduce consistent estimators for $H(\theta_0)$ and $G(\theta_0)$. They are

$$\widehat{H}(\hat{\theta}_N) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T-J} \sum_{j=1}^{J} \frac{\partial^2 \log f(y_{it}, y_{i,t+j} | \mathbf{x_i}; \hat{\theta}_N)}{\partial \theta \partial \theta'}$$

$$\widehat{G}(\hat{\theta}_N) = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T-J} \sum_{j=1}^{J} \frac{\partial \log f(y_{it}, y_{i,t+j} | \mathbf{x_i}; \hat{\theta}_N)}{\partial \theta} \right) \left(\sum_{t=1}^{T-J} \sum_{j=1}^{J} \frac{\partial \log f(y_{it}, y_{i,t+j} | \mathbf{x_i}; \hat{\theta}_N)}{\partial \theta} \right)'$$

where the derivatives of the likelihood function are

$$\frac{\partial \log f(y_{it}, y_{i,t+j} | \mathbf{x_i}; \hat{\theta}_N)}{\partial \theta} = \sum_{s_1=0}^{1} \sum_{s_2=0}^{1} \mathbb{1}_{s_1,s_2} \frac{\frac{\partial}{\partial \theta} \mathbb{P}_{s_1,s_2}(\hat{\theta}_N)}{\mathbb{P}_{s_1,s_2}(\hat{\theta}_N)}$$

$$\frac{\partial^2 \log f(y_{it}, y_{i,t+j} | \mathbf{x_i}; \hat{\theta}_N)}{\partial \theta \partial \theta'} = \sum_{s_1=0}^{1} \sum_{s_2=0}^{1} \frac{\mathbb{1}_{s_1,s_2}}{\mathbb{P}_{s_1,s_2}(\hat{\theta}_N)} \left[\frac{\partial^2 \mathbb{P}_{s_1,s_2}(\hat{\theta}_N)}{\partial \theta \partial \theta'} - \frac{1}{\mathbb{P}_{s_1,s_2}(\hat{\theta}_N)} \frac{\partial \mathbb{P}_{s_1,s_2}(\hat{\theta}_N)}{\partial \theta} \frac{\partial \mathbb{P}_{s_1,s_2}(\hat{\theta}_N)}{\partial \theta'} \right]$$

Here the notation is simplified and the dependencies on (i, t, j) are suppressed. Clearly, $\mathbb{1}_{s_1,s_2}$ denotes $\mathbb{1}(y_{it} = s_1, y_{i,t+j} = s_2)$, and $\mathbb{P}_{s_1,s_2}(\theta)$ denotes $\mathbb{P}\left(y_{it} = s_1, y_{i,t+j} = s_2 \mid \mathbf{x_i}; \theta\right)$. More details on the derivatives of the probability functions are given in the appendix.

A small note on choosing the lag length is worth to mention. As in the MLE case, one can use AIC/BIC type of criteria to choose the lag length J in an optimal way. The criteria are in their usual forms as in the pseudo-likelihood or quasi-likelihood estimation cases: $AIC(\hat{\theta}_N) = -2\mathcal{L}_c(\hat{\theta}_N | x, y) + 2 \operatorname{tr}\{G(\hat{\theta}_N) H(\hat{\theta}_N)^{-1}\}$ and $BIC(\hat{\theta}_N) = -2\mathcal{L}_c(\hat{\theta}_N | x, y) + \log(N) \operatorname{tr}\{G(\hat{\theta}_N) H(\hat{\theta}_N)^{-1}\}$. In theory, the larger is the lag length J the more efficient is the estimator; however, in practice with finite N and T, sometimes larger J might bring less efficiency after some point due to the fact that there might not be any useful information left after a certain J, and including these terms in the composite likelihood might create extra noise (see Varin and Vidoni [2006] for a simulation exercise in a time series setting). The same is true with the pairwise $f(y_{it}, y_{i,t+j})$ vs the triplet $f(y_{it}, y_{i,t+j}, y_{i,t+j+k})$ composite likelihood. The triplet composite likelihood is, in theory, more efficient than the pairwise likelihood. However, in practice with finite data application, the $(j+k)$th lag might just bring noise instead of useful information; moreover, it increases computational burden exponentially.

There is nothing particular about large N fixed T setup of the composite likelihood in this paper. Composite likelihood approach to AR-Probit model can also be used in a univariate time series setting with $N = 1$ and large T, as well as in a large N large T panel setting. The identification conditions and the derivatives of the bivariate probabilities will not be affected by any of these changes. Certain moment conditions should be adjusted to provide the finiteness of the composite likelihood and the hessian. Extra attention should be paid to the variance of $G(\theta_0)$ matrix since the terms in the score function will be correlated. Therefore, one needs to compute the long-run variance when computing $G(\theta_0)$. Thus, its estimator should utilize Newey-West type of long-run variance estimator.

3.1.1 Comparison of CLE to GMM

As mentioned in the introduction, CLE and GMM resemble each other in the sense that pieces of likelihoods are chosen for CLE whereas moments are chosen for GMM. They both require a choice by the researcher. Theoretically, GMM is a more general estimation technique since it assumes CLE as a special case where one can choose the moments as the score of the CLE. In this case, CLE will be identical to GMM. In this section, I will compare pairwise CLE and GMM where the most obvious and common moments are chosen. The simulation setup will mimic the setup of the empirical part of this paper. In particular, I will argue that in a large N and moderate T panel setting, GMM is inferior to CLE in terms of estimation performance as well as computation time. The problem with GMM is that there are too many moments when T is not small, which makes the computation of the efficient GMM infeasible.

Before showing the simulation result, let's first analyze the moments for the GMM.

$$\begin{aligned}
\mathbb{E}[y_{it}|\mathbf{x_i}] &= \mathbb{P}(y_{it} = 1|\mathbf{x_i}) = \mathbf{\Phi}(-m_{it}) \\
\text{Var}[y_{it}|\mathbf{x_i}] &= \mathbf{\Phi}(-m_{it})\mathbf{\Phi}(m_{it}) \\
\mathbb{E}[y_{it}y_{i,t+j}|\mathbf{x_i}] &= \mathbb{P}(y_{it} = 1, y_{i,t+j} = 1|\mathbf{x_i}) = 1 - \mathbf{\Phi}(m_{it}) - \mathbf{\Phi}(m_{i,t+j}) + \mathbf{\Phi}_2\left(m_{it}, m_{i,t+j} \mid \rho^j\right)
\end{aligned}$$

Let's count the number of moments implied in each moment condition. For a K-dimensional covariate vector x_{it}, the condition $\mathbb{E}\left[\left\{y_{it} - \mathbf{\Phi}(-m_{it})\right\} x_{it}\right] = 0$ for all t gives TK-many moments. The moment condition for the variance $\mathbb{E}\left[\left\{[y_{it} - \mathbf{\Phi}(-m_{it})]^2 - \mathbf{\Phi}(-m_{it})\mathbf{\Phi}(m_{it})\right\} x_{it}\right] = 0$ for all t implies also TK-many moments. Finally, $\mathbb{E}\left[\left\{y_{it}y_{i,t+j} - 1 + \mathbf{\Phi}(m_{it}) + \mathbf{\Phi}(m_{i,t+j}) - \mathbf{\Phi}_2\left(m_{it}, m_{i,t+j} \mid \rho^j\right)\right\} x_{it}\right] = 0$ for all t and all j yields $\sum_{j=1}^{J}(T-j)K$-many moments. All in all, just by using the mean, variance, and covariance moments, we end up with $TK + TK + \sum_{j=1}^{J}(T-j)K$-many moments. In the empirical study of this paper, $N = 516$, $T = 55$, $J = 8$, and $K = 11$. The number of moments in such a setup would be 5654. To compute the efficient GMM, one needs to invert (5654×5654)-dimensional weighting matrix, which is practically impossible. Thus, the second best thing one can do is to invert the diagonal (5654×5654)-dimensional weighting matrix, which weights each moment inversely according to its noise level. For the simulation purposes, I choose a smaller setup than the empirical study of the paper. In the simulations, $(N, T, J, K) = (500, 50, 4, 3)$ which implies 870 moments. Table 1 shows the simulation results. In this table, GMM indicates the first step GMM with the identity matrix as the weighting matrix. Hence, it can be considered nonlinear least squares estimation. E-GMM represents the second step GMM where the weighting matrix is the inverse of the variances of each moment computed in the first step. Thus, it can be considered weighted nonlinear least squares. Note that, the efficient GMM is not feasible since one needs to invert a (870×870)-dimensional matrix. The simulation study consists of 200 simulations.

The simulations clearly indicate that composite likelihood estimator outperforms GMM estimator in this setup. The estimation of the autocorrelation parameter ρ is very accurate with CLE. Moreover, the root mean squared errors (RMSE) are smaller for CLE except for one parameter. In terms of computation times, CLE is around 100 times faster than GMM in a personal computer. Naturally, this comparison can change from computer to computer, or might depend on the coding; but, the computational attractiveness of CLE will stay intact.

[Figure 1 here]

4 Large N Moderate T Application: Credit Ratings

4.1 Introduction

Credit ratings reflect the creditworthiness of a borrower or obligor. Hence, they constitute an essential part of investors' decision of buying a company's bonds – even its stocks. The accuracy and timeliness of the ratings are important for the financial markets and the economy. Inaccurate or miss-timed ratings can aggravate a crisis (Ferri et al. [1999]). Therefore it is important to correctly model credit ratings. A part of credit risk modeling literature goes back to Altman [1968] where corporate failure is analyzed by discriminant analysis based on accounting ratios (Altman Z-score models). Since then, there have been numerous papers on credit scoring. However, these firm-level models have been criticized for being static and missing the dynamic nature of the ratings. In this paper, I propose a novel model for the credit rating literature: a panel autoregressive ordered probit which takes into account the persistence of credit ratings, firm-specific variables, and the business cycle. In this model, the observed discrete variable y_{it} will represent the rating that firm i at time t receives. The latent variable y_{it}^* can be seen in different ways: y_{it}^* is the continuous rating that the firm gets during the creditworthiness assessment. Since the rating agency cannot publish a continuous rating, they discretize these ratings and publishes letter grades. Another interpretation can be the creditworthiness of the firm estimated by the rating agency. If this estimate falls between certain threshold, then the firm is assigned a letter grade according to the interval it belongs to. All in all, the latent variable reflects the view of the rating agency regarding the firm. In the model, the persistence of the ratings is driven by the autocorrelation in the unobserved credit quality of the firm as seen from the credit rating agency's perspective, which depends on the firm's financial ratios and the state of the economy. Since the credit quality of a firm, in general, changes slowly, the assigned ratings change slowly as a result. Moreover, because of the complicated nature of the model, maximum likelihood estimation is impractical; thus, I use composite likelihood estimation to estimate the model. The results show a small improvement over the static probit model for in-sample predictions, but large improvements for pseudo out-of-sample predictions and the estimated transition matrices. The dynamic nature of the model helps estimate the transition much more successfully than the static model.

Investors might want to keep the risk level of their portfolio at a certain level. Since re-balancing a portfolio is costly, investors prefer rating stability rather than ratings to reflect temporary changes in companies' financial conditions. At the same time, ratings should reflect an accurate estimate of the borrower's condition. Hence, timeliness of ratings is also important. In this regard, credit rating agencies face a trade-off between rating stability and accuracy (Cantor and Mann [2006]). There are two conceptual approaches for assigning ratings: Through-the-cycle (TTC) and Point-in-time (PIT). TTC (also called cycle-neutral) methodology focuses more on the permanent component of default risk rather than short-term fluctuations due to business cycles. Hence, TTC approach renders ratings more stable and rating migration more prudent. However, with this methodology, the timeliness of the ratings can be at question since ratings might lag the actual default risk of a company. In PIT method, on the other hand, current conditions of a company has a big effect on its credit rating. With this approach, ratings predominantly reflect the current condition of the borrower. The question of whether credit rating agencies assign ratings according to TTC or PIT has attracted a lot of attention from researchers. There are several papers providing evidence for the pro-cyclical behavior of ratings. See, for instance, Nickell et al. [2000], Bangia et al. [2002], Amato and Furfine [2004], Koopman and Lucas [2005], Feng et al. [2008], Topp and Perl [2010], Auh [2015], and Freitag [2015]. Some of these papers conclude that this is evidence for PIT ratings. However, rating actions being pro-cyclical does not necessarily imply that the ratings reflect short-term fluctuations. For instance, during a recession, there might be a significant change in the long-term credit quality of a firm. Therefore, lower ratings during a recession do not mean that the rating agencies cannot see through the cycle. On the TTC side of the literature, Löffler [2004, 2013] and Kiff et al. [2013] show that the credit ratings have predictive power on the long-term component on default probabilities. The slow and delayed reaction by credit rating agencies are empirically documented by Altman and Rijken [2006] and Löffler [2005]. The empirical results in this paper support this phenomenon, thus provide evidence for TTC ratings. In particular, the results show that at the beginning of Great Recession, the rating agencies waited 2-3 quarters before reflecting the changes in the credit quality of the firms on their ratings. This finding is quantitatively in line with the estimates in Altman and Rijken [2006]. They find that the TTC methodology delays the timing of rating migrations by 0.70 year on average.

[Figure 1 here]

Cheng and Neamtiu [2009] described the rating quality by three properties: accuracy, stability (or volatility), and timeliness. In this paper, I focus on the stability of the ratings. TTC methodology is designed to induce rating stability (Carey and Hrycay [2001]). However, little research is done on quantifying the stability of the credit ratings. One way of inferring rating stability is through the frequency of rating transitions. Smaller transition frequencies mean higher stability. Figure 1 displays percentage of issuers for which their ratings stayed unchanged from one year to another. For instance, the highest stability in the ratings occurred in the year 2004 – more than 75% of the firms stayed in the same rating class. On the other hand, the lowest stability occurred during the Great Recession. Motivated by this observation, the rating stability can be measured through the diagonal (and near diagonal) elements of a transition matrix. The diagonal element of a transition matrix shows the percentage of firms that stayed in the same rating class in a given year. Jafry and Schuermann [2004] and Amato et al. [2013] generate "mobility indices" by analyzing the singular values and the eigenvalues of the yearly rating transition matrices. These indices capture the time-varying stability of ratings over the years. But, these values are obtained without controlling for any firm or macroeconomic conditions. In this regard, these indices represent the unconditional stability of the ratings. There are also portfolio models that use transition matrices and control for business cycle effects. However, these models are "cohort-style", that is, all the firms within a rating class are identical. Hence, the rating classes exhibit persistence and correlations among themselves where individual firms actually do not matter. But, the research by Lando and Skødeberg [2002] and Frydman and Schuermann [2008] shows that two firms with identical current credit ratings can have substantially different transition probabilities. This shows that ratings exhibit non-Markovian property. In this regard, AR-Probit model seems a better model to analyze the ratings since the ordered outcome variable in this model is non-Markovian. Moreover, the analysis is done at the firm level by controlling for firm balance sheet ratios and business cycle variables. Finally, the rating stability is inferred by the autocorrelation parameter (ρ) of the underlying continuous latent variable. Hence, the stability is induced from the fundamentals of a firm. The higher is ρ the higher is the stability of ratings. Moreover, the model can be extended to a time-varying parameter model, where the rating stability can be estimated at a quarterly level.

In Dynamic Probit model, last period's credit rating $y_{i,t-1}$ is used to predict the current one y_{it}. Mizen and Tsoukas [2009] use this model for forecasting purposes. As they noted, the knowledge of the previous rating helps the forecast but the coefficients of $y_{i,t-1}$ do not say anything about persistence, instead, it just shifts the latent credit quality of the firm by constant term. Note that, the right-hand-side variable $y_{i,t-1}$ actually represents a dummy variable for each rating class. Consider a firm with a high rating, say AAA. There is no upper level ratings that a AAA firm can go to; such firms can face only downgrades. Therefore, a AAA rating in the previous period will necessarily render the coefficient of $y_{i,t-1}$ to be negative. A similar argument will be valid for AA firms as well since it is more likely for a AA firm to face a downgrade than an upgrade. From the opposite angle, firms that are below investment grade will be more likely to have upgrades – especially if the data do not include default observations. Thus, the most likely direction for a speculative grade is upwards, which renders the coefficient of previous BB/B/CCC rating to be positive. All in all, the coefficients of the previous rating artificially captures the convergence to the middle rating BBB instead of capturing the stability. The assigned ratings should be a result of the unobserved creditworthiness of a firm, not a determinant of it.

The empirical analysis focuses on three parts. The first part is a comparison of static probit model and AR-Probit model. The estimation results show a significant and economically large persistent parameter. In AR-Probit, one can compute short-term and long-term effects of the explanatory variables. In static model, however, the effects last only one term. In comparison, I find that the estimates from the static model are indeed an average between the short-term and long-term cumulative effects computed from AR-Probit estimates. In particular, static model estimates are close to the 2-to-4 quarter cumulative effects derived from AR-Probit model. Regarding in-sample prediction performances, both model performs similarly – though AR-Probit model has a better prediction for infrequent rating classes. Regarding transition matrix estimates, AR-Probit model shows 5 times better performance than the static model since it accounts for persistence in the ratings. In the second part, time-varying parameter model is utilized to estimate changes in the rating stability over time. The results show that the stability declines over time. But, the decline is more prominent after the crisis. It means that the rating agencies try to assign the ratings more timely. One reason for that can be the critics on rating agencies during the recession. Another one is the new regulations enforced on the agencies to increase their liability. Trying to increase the rating quality after critics and regulations are in line with the findings in the literature (see Cheng and Neamtiu [2009]). Facing widespread criticism, the credit rating agencies might be concerned for their reputation and start assigning the ratings more conservatively. In the third part, I show the evidence for sluggish rating adjustment during the recent crisis. Based on the time-varying parameter model, the results show that the credit rating agencies increase the weights on the past information – to keep the ratings stable – when they face changes in the financial situation of the firm.

4.2 Data

I use Financial Ratios Suite by WRDS database for quarterly balance sheet financial ratios of companies. The measure of credit rating is the S&P Long-Term Issuer Level rating obtained from Compustat in WRDS database. Ratings are available in monthly frequency; I convert them to quarterly frequency by taking the last rating within each quarter assuming that the most up-to-date information on a firm is contained in the most recent rating. Firm-level data are between 2002Q1-2015Q3. I use the quarterly macro data set of McCracken and Ng [2015] to extract business cycle factors. The factors are estimated by principal component analysis from a large panel of macro variables that include real sector, employment, housing, prices, interest rates, money and credit, exchange rates, and financial data. In the data set, there are in total 218 variables between 1971Q2–2015Q3. After extracting two principal components with the largest eigenvalues, I took the corresponding dates of the factors that match with the data range of credit ratings. Hence, the estimated factors capture the business cycle of the economy. Finally, NBER recession dates are obtained from FRED.

Even though the term "credit rating of a firm" is frequently used, the corporate bond that is issued by the obligor receives a rating, rather than the obligor itself. An obligor can issue several bonds, and each issue might have a different rating. However, senior unsecured long-term bonds' rating are close to the issuer rating since the debt defaults only when the issuer defaults. Therefore, Long-Term Issuer Level ratings are de-facto the creditworthiness of the obligor. I convert letter ratings into ordinal numbers from 1 to 7 corresponding to the grades {CCC, B, BB, BBB, A, AA, AAA}, respectively, that is CCC=1 and AAA=7. Note that, I grouped ratings without considering notches +/−, e.g., AA−, AA and AA+ belong to a single category denoted as AA. The CCC category contains all the ratings including any C letter, i.e., CCC+, CCC, CCC−, CC, and C. Observations with D, SD (Suspended) or NM (Not meaningful) ratings are excluded. In the robustness analyses, defaulted firms will also be included in the dataset.

Using financial ratios for credit rating determination in discrete choice models is common in the literature. I do not claim that this is the exact method the credit rating agencies do follow to generate the ratings. Yet, as described in Van Gestel et al. [2007], the real rating process may be well approximated by such models with financial ratios as determinants. Moreover, Standard and Poor's [2013] gave a list of *Key Financial Ratios* that are used in rating adjustment process. One thing to note is that some financial ratios are highly correlated with each other, thus one needs to take this into account while choosing the variables. Another important thing is that some ratios are not available at a quarterly frequency. Based on these criteria, I use the following set of financial ratios that capture solvency, financial soundness, profitability, and valuation of a firm: total debt leverage (debt/assets), long-term debt to total debt ratio (ltd/debt), return on assets (roa), coverage ratio (cash/debt), net profit margin, and valuation ratio (price/sales). The detailed definitions and descriptive statistics of the variables are given in Table 2 and Table 3, respectively.

[Table 2 here]

[Table 3 here]

As a solvency measure, the debt-to-asset ratio is used, which reflects the leverage level of the firm. In general, the higher is this ratio the riskier is the company in terms of meeting its debt payments. Financial soundness is implied by the ratios of total long-term debt to total debt [6] and operating cash flow to debt. The former ratio shows the capital structure of the firm and is negatively related to credit ratings whereas the latter one is a coverage ratio showing the ability to carry the debt of the company and is positively related to credit ratings. Profitability is another important aspect showing how easy a firm can generate income. It is captured by return on assets and net profit margin [7] , which are positively correlated to credit ratings. Finally, as a valuation ratio, I use market value to sales ratio[8].

[6] In the literature, many papers use the ratios debt-to-assets and long term debt to assets together in regressions. However, these variables are highly correlated ($\sim 75\%$). To avoid multicollinearity, I prefer using long term debt to total debt ratio to capture the debt structure instead of long term debt to assets.

[7] Operating profit margin (opm) is more frequently used than net profit margin (npm) in the literature. However, Corr$(roa, opm) = 0.84$, but Corr$(roa, npm) = 0.64$. It means that $\{roa, opm\}$ is likely to create multicollinearity problem whereas $\{roa, opm\}$ is not. Moreover, given the fact that Corr$(opm, npm) = 0.76$ and that opm and npm have very similar definitions, I choose nmp over opm for the analysis.

[8] Many papers use price-to-book ratio instead of price-to-sales. However, these papers have annual data. But, at the quarterly frequency the data for p/b have a lot of missing values. Another famous choice is price-to-earnings ratio. However, especially during the crises many firms suffer losses, i.e., they do not make any earnings, which renders p/e

To control for the state of the economy, the literature uses various choices of business cycle variables. The NBER recession dummy seems the most common choice, but choice of macro fundamental variables differ from paper to paper. While some papers use GDP growth rate (Feng et al. [2008], Koopman et al. [2009], and Alp [2013]), others create their business cycle indicator (Amato and Furfine [2004], Freitag [2015]). Hence, it is not clear which business cycle variable should be used. For this reason, I prefer using estimated factors from a large macroeconomic data set. The first two principal components (called Factor1 and Factor2) explain more than 20% of the total variation in 218 business cycle variables. They are especially related to the real economy sector. For instance, they explain around 70% of the variation in real variables such as output, exports, imports, personal income, private investment, and housing starts. These estimated factors appear to be positively correlated with the ratings whereas the ratings are lower during the NBER recession dates, as expected.

For the empirical baseline results, a balanced panel data set is used. In this data set, there are 516 firms over 55 quarters (2002Q1–2015Q3) with no missing data. Hence, they are the firms that 'survived' throughout the data period. The frequency of the ratings in this data set is given on the left side of Table 4. Since more than 70% of the ratings are in the investment grade category, we can consider this data set as investment-grade firms. As a robustness check, estimation results based on an unbalanced panel data set are also presented. The reason why this data set is not the baseline is two folds. First, it is not clear how to model D rated firms. Should the D ratings be excluded from the analysis or do these observations indeed contain useful information in terms of rating dynamics? Note that, a D rating does not necessarily mean that the firm is out of the market. There are firms that have consecutive D ratings for a few quarters, but then continue being rated without any interruption in their rating history[9]. Second, due to the autocorrelation in the latent variable, modeling the missing data in the middle of a firm's history is not straightforward; it will result in a complex formulation. As a result, I include data for the firms that do not have any missing data once they entered the market until they leave. A firm is allowed to enter the data set after the initial date 2002Q1 and to leave it before 2015Q3. Since the firms with an extremely short span of data are not representative and exhibit large variations, I excluded firms that have less than 5 years of quarterly data. Moreover, D ratings are also included in the data set as long as the balance sheet data are also available. Finally, there are 1406 firms with an average of 38 quarters in this data set. The frequency of the ratings for the unbalanced panel data is given on the right side

ratio meaningless for a crucial period in the dataset. For these reasons, I prefer using price-to-sales ratio over p/b and p/e.

[9]For instance, Xerium Technologies Inc. filed bankruptcy for 2010Q1–2010Q2, but was rated B in 2010Q3 and continued being in the market. As long as we can observe how the defaulted firms' balance sheet data evolve, coming from bankruptcy back to business, in fact, contains useful information. Such cases are obviously rare (most firms' data ends once they default), and omitting them will not affect the estimates.

of Table 4. In this data set, ratings are more evenly distributed. In particular, we have a relatively higher representation of sub-investment firms compared to the balanced panel. This difference will allow us to highlight characteristic differences between investment and non-investment firms.

[Table 4 here]

4.3 Extensions of the Model

In this subsection, I extend the baseline model into various directions. All these models will be used in the empirical part to address different aspects of the credit rating data. The first extension is changing the binary response variable into an ordered one. This model will be the working model of the empirical part. Another extension is allowing for random effects to control for firm heterogeneity. Another interesting extension is allowing for time-varying parameters, in particular, time-varying autocorrelation coefficient. Finally, I will analyze unbalanced panel probit model.

4.3.1 Panel AR Ordered Probit Model

For $i = 1, \dots, N$ and $t = 1, \dots, T$, let i denotes the i^{th} firm and t denotes time. I assume that the innovations are $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, the $(K \times 1)$ dimensional explanatory variables are denoted as x_{it}, which are assumed to be strictly exogenous. The time dependent continuous variable y_{it}^* is unobserved, however the ordinal variable $y_{it} \in \{1, \dots, S\}$ is observed. The levels for y_{it} is merely for classification; the mathematical distance between two ordinal values is meaningless. Hence, a Panel Autoregressive Ordered Probit model can be written as

$$y_{it}^* = \rho y_{i,t-1}^* + \beta' x_{it} + \varepsilon_{it}, \tag{9}$$

$$y_{it} = s \quad \text{if} \quad \tau_{s-1} < y_{it}^* \leq \tau_s, \tag{10}$$

where $s = 1, \dots, S$ and the threshold coefficients are $\tau_0 = -\infty < \tau_1 = 0 < \tau_2 < \cdots < \tau_{S-1} < \tau_S = \infty$. For $S = 2$, that is when y_{it} is binary, the model is a simple probit model; for $S > 2$, it is called ordered probit model. One can relax the assumption of $\tau_1 = 0$ if y_{it}^* does not contain any constant term and is a mean-zero process. The calculation of the bivariate probabilities of j period distant observations will be done in a similar way as in (5). Note that,

$$
\begin{aligned}
\mathbb{E}[y_{it}^*|\mathbf{x_i}] &= \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} \\
\mathrm{Var}[y_{it}^*|\mathbf{x_i}] &= \frac{1}{1-\rho^2} \\
\mathrm{Cov}[y_{it}^*, y_{i,t+j}^*|\mathbf{x_i}] &= \frac{\rho^j}{1-\rho^2},
\end{aligned}
$$

yield the following bivariate probabilities.

$$
\begin{aligned}
&\mathbb{P}\left(y_{it} = s_1, y_{i,t+j} = s_2 \mid \mathbf{x_i}; \theta\right) \\
&\quad = \mathbb{P}\left(\tau_{s_1} < y_{it}^* \le \tau_{s_1+1},\ \tau_{s_2} < y_{i,t+j}^* \le \tau_{s_2+1} \mid \mathbf{x_i}; \theta\right) \\
&\quad = \mathbb{P}\left(m_{s_1,t}(\mathbf{x_i}, \theta) < Z_1 \le m_{s_1+1,t}(\mathbf{x_i}, \theta),\ m_{s_2,t+j}(\mathbf{x_i}, \theta) < Z_2 \le m_{s_2+1,t+j}(\mathbf{x_i}, \theta) \,\Big|\, \mathbf{x_i}; \theta\right) \\
&\quad = \Phi_2\left(m_{s_1+1}(\mathbf{x_i}, \theta), m_{s_2+1}(\mathbf{x_i}, \theta) \,\Big|\, r(\theta)\right) - \Phi_2\left(m_{s_1+1}(\mathbf{x_i}, \theta), m_{s_2}(\mathbf{x_i}, \theta) \,\Big|\, r(\theta)\right) \\
&\qquad - \Phi_2\left(m_{s_1}(\mathbf{x_i}, \theta), m_{s_2+1}(\mathbf{x_i}, \theta) \,\Big|\, r(\theta)\right) + \Phi_2\left(m_{s_1}(\mathbf{x_i}, \theta), m_{s_2}(\mathbf{x_i}, \theta) \,\Big|\, r(\theta)\right), \qquad (11)
\end{aligned}
$$

where $m_{s,t}(\mathbf{x_i}, \theta) \equiv \sqrt{1-\rho^2}\left(\tau_s - \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k}\right)$, Z_1 and Z_2 are jointly standard normally distributed with the correlation coefficient $r = \rho^j$. Based on these bivariate probabilities, one can write down the pairwise composite log-likelihood of the Panel AR Ordered Probit model.

$$
\begin{aligned}
\mathcal{L}_c(\theta|y, x) &= \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T-J} \sum_{j=1}^{J} \log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta) \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T-J} \sum_{j=1}^{J} \sum_{s_1=1}^{S} \sum_{s_2=1}^{S} \mathbb{1}(y_{it} = s_1, y_{i,t+j} = s_2) \log \mathbb{P}\left(y_{it} = s_1, y_{i,t+j} = s_2 \mid \mathbf{x_i}; \theta\right), \quad (12)
\end{aligned}
$$

where the probabilities are given in (11).

4.3.2 Panel AR Ordered Probit with Random Effects

In this model, I assume y_{it}^* depends also on firm-specific random effect α_i which is allowed to depend on observed $\mathbf{x_i} = (x_{i1}, \ldots, x_{iT})$ values, but is independent over i conditional on $\mathbf{x_i}$. Hence, a Random Effects Panel Autoregressive Ordered Probit model can be written as

$$
\begin{aligned}
&y_{it}^* = \rho y_{i,t-1}^* + \beta' x_{it} + \alpha_i + \varepsilon_{it}, \\
&y_{it} = s \ \text{ if } \ \tau_{s-1} < y_{it}^* \le \tau_s,
\end{aligned}
$$

The firm-specific random effects will capture the effects that change over firms but not over time. For instance, it is likely to capture the location of the firm, the sector the firm is operating in, managerial activities, etc. as long as these variables do not change over time. Otherwise, they will be captured by idiosyncratic shocks $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. One cannot treat the unobserved firm heterogeneity as fixed effects and estimate them as parameters due to incidental parameter problem; it will yield biased and inconsistent estimates. There are a couple of ways to deal with the unobserved random α_i: it can be integrated out or replaced by a reduced form equation. In either way, one needs to make a distributional assumption for α_i, say $\alpha_i \overset{iid}{\sim} \mathcal{N}(\gamma_o + \gamma_1' \bar{x}_i, \sigma_\alpha^2)$ where $\bar{x}_i$ is the average of x_{it} over t. For identification purposes, we need to put a restriction on the variances of the innovations. The following will be assumed for the random effects model $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 1$. Note that, the parameter θ in this model also contains the extra parameters arising due to random effects, that is, $(\gamma_o, \gamma_1', \sigma_\alpha^2, \sigma_\varepsilon^2)$. The first approach will yield a likelihood function containing a one-dimensional integral, which can be approximated as accurately as desired by Gauss–Hermite quadrature [10].

$$
\begin{aligned}
L\theta|\boldsymbol{y}, \boldsymbol{x}) &= \prod_{i=1}^{N} \int f(\mathbf{y_i}|\mathbf{x_i}, \alpha_i) \phi\left(\alpha_i | \gamma_o + \gamma_1' \bar{x}_i, \sigma_\alpha^2\right) d\alpha_i \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{\pi}} \sum_{g=1}^{G} w_g f\left(\mathbf{y_i} \,\middle|\, \mathbf{x_i}, (\sqrt{2}\sigma_\alpha H_g + \gamma_o + \gamma_1' \bar{x}_i)\right)
\end{aligned}
$$

where $\phi(z|\mu, \sigma^2)$ denotes the normal probability density with mean μ and variance σ^2 evaluated at z, the nodes H_g are the zeros of g^{th} order Hermite polynomial, and w_g are the corresponding weights. Hence, $\sum_{g=1}^{G}$ approximates the integral by evaluating the function f at specific nodes and then weighting them. The pairwise likelihood of $f(\mathbf{y_i}|\mathbf{x_i}, \alpha_i)$ can be computed in the exactly same way as in (12), where α_i will be replaced by normalized nodes $\sqrt{2}\sigma_\alpha H_g + \gamma_o + \gamma_1' \bar{x}_i$.

In the other approach, α_i is replaced by a reduced form equation. Let's assume that $y_{i0}^* = \beta' x_{i0} + \frac{1}{1-\rho}\alpha_i + \frac{1}{\sqrt{1-\rho^2}}\varepsilon_{i0}$ where $\varepsilon_{i0} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\alpha_i = \mu + \gamma' \bar{x}_i + \eta_i$ where $\eta_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ such

[10] Gauss–Hermite quadrature is used for numerical integration; it approximates a specific type of integral in the following way.

$$
\int_{-\infty}^{\infty} h(x) \exp(-x^2) dx \cong \sum_{k=1}^{K} w_k h(x_k),
$$

where the nodes x_k are zeros of k^{th} order Hermite polynomial and w_k are corresponding weights. A table for the nodes and the weights can be found in Abramowitz et al. [1972], page 924. If one has a normal density instead of $\exp(-x^2)$, a standardization will be needed.

$$
\int_{-\infty}^{\infty} h(x) \exp(-0.5(x-\mu)^2/\sigma^2) dx = \int_{-\infty}^{\infty} h(\sqrt{2}\sigma x + \mu) \exp(-x^2) dx \cong \sum_{k=1}^{K} w_k h(\sqrt{2}\sigma x_k + \mu).
$$

that $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 1$.

$$
\begin{aligned}
y_{it}^* &= \rho y_{i,t-1}^* + \beta' x_{it} + \alpha_i + \varepsilon_{it} \\
&= \rho^t y_{i0}^* + \sum_{k=0}^{t-1} \rho^k \beta' x_{i,t-k} + \frac{1-\rho^t}{1-\rho}\alpha_i + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k} \\
&= \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} + \frac{1}{1-\rho}\alpha_i + \frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{i0} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k} \\
&= \frac{\mu}{1-\rho} + \frac{\gamma'}{1-\rho}\bar{x}_i + \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} + \frac{1}{1-\rho}\eta_i + \frac{\rho^t}{\sqrt{1-\rho^2}}\varepsilon_{i0} + \sum_{k=0}^{t-1} \rho^k \varepsilon_{i,t-k}
\end{aligned}
$$

The presence of η_i generates autocorrelation in the composite error term $\eta_i + \varepsilon_{it}$, which needs to be taken into account in calculating bivariate probabilities and the correlation coefficients. The following statistics will be used in the bivariate probabilities.

$$
\begin{aligned}
\mathbb{E}[y_{it}^*|x] &= \frac{\mu}{1-\rho} + \frac{\gamma'}{1-\rho}\bar{x}_i + \sum_{k=0}^{t} \rho^k \beta' x_{i,t-k} \\
\text{Var}[y_{it}^*|x] &= \frac{\sigma_\alpha^2}{(1-\rho)^2} + \frac{\sigma_\varepsilon^2}{1-\rho^2} \qquad (13) \\
\text{Cov}[y_{it}^*, y_{i,t+j}^*|x] &= \frac{\sigma_\alpha^2}{(1-\rho)^2} + \frac{\rho^j \sigma_\varepsilon^2}{1-\rho^2} \\
\text{Corr}[y_{it}^*, y_{i,t+j}^*|x] &= \frac{\frac{\sigma_\alpha^2}{(1-\rho)^2} + \frac{\rho^j \sigma_\varepsilon^2}{1-\rho^2}}{\frac{\sigma_\alpha^2}{(1-\rho)^2} + \frac{\sigma_\varepsilon^2}{1-\rho^2}} \qquad (14)
\end{aligned}
$$

This correlation formula looks complicated, however, note that if there was no firm heterogeneity, i.e., if $\sigma_\alpha^2 = 0$, then the correlation becomes ρ^j, as it was the case in Panel AR-Probit model in (11). If there was no idiosyncratic variation, i.e., $\sigma_\varepsilon^2 = 0$, then the correlation becomes 1 since the persistence in the composite error does not diminish due to the presence of η_i for each t. Hence, the correlation formula above represents the weighted average of these two extreme case correlations. Based on this correlation coefficient, one can say that in Panel AR-Probit models the interpretation of ρ is different depending on the presence of random effects. In models without random effects, ρ solely controls the autocorrelation of the latent state variable y_{it}^*. In contrast, when the firm-specific random effects exist, ρ is not the only source of persistence because α_i also persists over time thereby adds to the persistence of the latent process.

Denoting $\tilde{x}_{it} = (x_{it}', 1, \bar{x}_i')'$, $\delta = (\beta', \mu, \gamma_1')'$, and using backwards substitution on y_{it}^* gives let us compute the bivariate probabilities as follows.

$$\begin{aligned}
&\mathbb{P}\left(y_{it}=s_1, y_{i,t+j}=s_2 \mid \mathbf{x_i};\theta\right) \\
&=\mathbb{P}\left(\tau_{s_1}<y_{it}^*\leq \tau_{s_1+1},\ \tau_{s_2}<y_{i,t+j}^*\leq \tau_{s_2+1} \mid \mathbf{x_i};\theta\right) \\
&=\mathbb{P}\left(m_{s_1,t}(\mathbf{x_i},\theta)<Z_1\leq m_{s_1+1,t}(\mathbf{x_i},\theta),\ m_{s_2,t+j}(\mathbf{x_i},\theta)<Z_2\leq m_{s_2+1,t+j}(\mathbf{x_i},\theta) \,\Big|\, \mathbf{x_i};\theta\right), \qquad (15)
\end{aligned}$$

where $m_{s,t}(\mathbf{x_i},\theta) \equiv \sqrt{\frac{\sigma_\alpha^2}{(1-\rho)^2}+\frac{\sigma_\varepsilon^2}{1-\rho^2}}\left(\tau_s-\frac{\mu}{1-\rho}-\frac{\gamma'}{1-\rho}\bar{x}_i-\sum_{k=0}^{t}\rho^k\beta' x_{i,t-k}\right)$, Z_1 and Z_2 are jointly standard normally distributed with the correlation coefficient given in (14). As a result, this approach resembles the basic model defined in (9)-(10) except that now the errors are allowed to be autocorrelated with a special structure and that the regressors contain time averages. Therefore, the pairwise log-likelihood of this model will resemble that in (12) with a more complicated bivariate probability structure taking into account the extra autocorrelation in the errors.

4.3.3 Panel AR Ordered Probit with Time-Varying Coefficients

In this section, I allow the parameters to change over time – in particular the persistence parameter ρ. Especially in credit ratings application, time-varying persistence will be important to see how rating stability changes over time. This will give us some idea whether credit rating agencies assign their ratings according to the business cycle. In other words, it will indicate whether credit rating agencies use TIC or PIT approach. Moreover, we can also see whether policy changes, like Dodd–Frank Act, have any effect on rating stability. The model with time-varying persistence can be written as

$$y_{it}^*=\rho_t y_{i,t-1}^*+\beta' x_{it}+\varepsilon_{it}, \qquad (16)$$

$$y_{it}=s \ \text{ if } \ \tau_{s-1}<y_{it}^*\leq\tau_s, \qquad (17)$$

Let's assume that $y_{i0}^*=\beta' x_{i0}+\varepsilon_{i0}$ where $\varepsilon_{i0}\overset{iid}{\sim}\mathcal{N}(0,1)$ (note that, if $y_{i0}^*=0$ is assumed then ρ_1 cannot be estimated). Let's denote the new parameter vector as $\theta=(\rho_1,\ldots,\rho_T,\beta',\tau')'$. A sufficient condition for a non-explosive behavior of y_{it}^* is $|\rho_t|<1$ for all t. Time-varying parameters will be identified from the cross-section variation. In other words, information on ρ_t will be accumulated by the rating changes of the firms between periods $(t-1)$ and t. In this model, again backward substitution will be used, but the time-varying parameters will complicate the bivariate probabilities in terms of parameters.

$$
\begin{aligned}
y_{i1}^* &= \rho_1 y_{i0}^* + \beta' x_{i1} + \varepsilon_{i1} \\
y_{i2}^* &= \rho_2 y_{i1}^* + \beta' x_{i2} + \varepsilon_{i2} = \rho_2\rho_1 y_{i0}^* + \beta' x_{i2} + \rho_2 \beta' x_{i1} + \varepsilon_{i2} + \rho_2 x \varepsilon_{i1} \\
&\vdots \\
y_{it}^* &= \left(\prod_{k=1}^{t} \rho_k\right) y_{i0}^* + \beta' x_{it} + \sum_{k=2}^{t} \left(\prod_{s=k}^{t} \rho_s\right) \beta' x_{i,k-1} + \varepsilon_{it} + \sum_{k=2}^{t} \left(\prod_{s=k}^{t} \rho_s\right) \varepsilon_{i,k-1} \\
&= \beta' x_{it} + \sum_{k=1}^{t} \left(\prod_{s=k}^{t} \rho_s\right) \beta' x_{i,k-1} + \varepsilon_{it} + \sum_{k=1}^{t} \left(\prod_{s=k}^{t} \rho_s\right) \varepsilon_{i,k-1}
\end{aligned}
$$

The following statistics will be used in computing the bivariate probabilities.

$$
\begin{aligned}
\mathbb{E}[y_{it}^*|x] &= \beta' x_{it} + \sum_{k=1}^{t} \left(\prod_{s=k}^{t} \rho_s\right) \beta' x_{i,k-1} \\
\mathrm{Var}[y_{it}^*|x] &= 1 + \sum_{k=1}^{t} \prod_{s=k}^{t} \rho_s^2 \\
\mathrm{Cov}[y_{it}^*, y_{i,t+j}^*|x] &= \left(\prod_{s=t+1}^{t+j} \rho_s\right) \left[1 + \sum_{k=1}^{t} \prod_{s=k}^{t} \rho_s^2\right] \\
\mathrm{Corr}[y_{it}^*, y_{i,t+j}^*|x] &= \left(\prod_{s=t+1}^{t+j} \rho_s\right) \sqrt{\frac{1 + \sum_{k=1}^{t} \prod_{s=k}^{t} \rho_s^2}{1 + \sum_{k=1}^{t+j} \prod_{s=k}^{t+j} \rho_s^2}}
\end{aligned}
\tag{18}
$$

As t gets larger, the term inside the square root will be very close to 1. Thus, the correlation between accumulated discounted errors can be approximated by $\prod_{s=t}^{t+j-1} \rho_{s+1}$. This is also the correlation between y_{it}^* and $y_{i,t+j}^*$ conditional on $\mathbf{x_i}$. If all the persistence parameters are the same $\rho_t = \rho$, then the correlation coefficient simply becomes ρ^j as t gets larger, as in the baseline Panel AR-Probit case. The bivariate probabilities of the pair $(y_{it}, y_{i,t+j})$ in this model can be computed in the following way.

$$
\begin{aligned}
&\mathbb{P}\left(y_{it} = s_1, y_{i,t+j} = s_2 \mid \mathbf{x_i}; \theta\right) \\
&= \mathbb{P}\left(\tau_{s_1} < y_{it}^* \leq \tau_{s_1+1},\ \tau_{s_2} < y_{i,t+j}^* \leq \tau_{s_2+1} \mid \mathbf{x_i}; \theta\right) \\
&= \mathbb{P}\Bigg(\tau_{s_1} < \beta' x_{it} + \sum_{k=1}^{t-1} \left(\prod_{s=k}^{t-1} \rho_{s+1}\right) \beta' x_{ik} + \varepsilon_{it} + \sum_{k=1}^{t-1} \left(\prod_{s=k}^{t-1} \rho_{s+1}\right) \varepsilon_{ik} \leq \tau_{s_1+1}, \\
&\qquad \tau_{s_2} < \beta' x_{i,t+j} + \sum_{k=1}^{t+j-1} \left(\prod_{s=k}^{t+j-1} \rho_{s+1}\right) \beta' x_{ik} + \varepsilon_{i,t+j} + \sum_{k=1}^{t+j-1} \left(\prod_{s=k}^{t+j-1} \rho_{s+1}\right) \varepsilon_{ik} \leq \tau_{s_1+1} \leq \tau_{s_2+1} \,\Bigg|\, \mathbf{x_i}; \theta\Bigg) \\
&= \mathbb{P}\left(m_{s_1,t}(\mathbf{x_i}, \theta) < Z_{it} \leq m_{s_1+1,t}(\mathbf{x_i}, \theta),\ m_{s_2,t+j}(\mathbf{x_i}, \theta) < Z_{i,t+j} \leq m_{s_2+1,t+j}(\mathbf{x_i}, \theta) \,\middle|\, \mathbf{x_i}; \theta\right),
\end{aligned}
\tag{19}
$$

where

$$m_{s,t}(\mathbf{x_i},\theta) \equiv \frac{\tau_s - \beta' x_{it} - \sum_{k=1}^{t-1}\left(\prod_{s=k}^{t-1}\rho_{s+1}\right)\beta' x_{ik}}{\sqrt{1+\sum_{k=1}^{t-1}\prod_{s=k}^{t-1}\rho_{s+1}^2}} \quad \text{and} \quad Z_{it} \equiv \frac{\varepsilon_{it} + \sum_{k=1}^{t-1}\left(\prod_{s=k}^{t-1}\rho_{s+1}\right)\varepsilon_{ik}}{\sqrt{1+\sum_{k=1}^{t-1}\prod_{s=k}^{t-1}\rho_{s+1}^2}},$$

Z_{it} and $Z_{i,t+j}$ are jointly normally distributed with zero mean, unit variance, with the correlation coefficient given in the equation (18). Hence, the pairwise log-likelihood is the same as in (12) except that the parameter vector now contains $\{\rho_t\}_{t=2}^T$ and the probabilities are given in (19).

4.3.4 Unbalanced Panel AR-Probit

I allow missing values. CL allows any part of the likelihood. Hence, the likelihood of only observed data will be included. That is, the likelihood of a firm with available data between t_i and T_i will contribute to the composite likelihood. For simplicity, I assume that there is no missing observation once the data for a firm started. However, a firm may enter the market after the initial date of data or may leave the market before the end date. Hence, $1 \leq t_i < T_i \leq T$ for all i. The backwards substitution (4) of the state variable for firm i is done for the available observations, i.e., for $t = t_i, \ldots, T_i$. Similarly for the bivariate probabilities. Hence, the pairwise log-likelihood of the data including missing values is

$$\begin{aligned}
\mathcal{L}_c(\theta|\boldsymbol{y},\boldsymbol{x}) &= \frac{1}{N}\sum_{i=1}^{N}\log f(\mathbf{y_i}|\mathbf{x_i};\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=t_i}^{T_i-J}\sum_{j=1}^{J}\log f(y_{it},y_{i,t+j}|\mathbf{x_i};\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=t_i}^{T_i-J}\sum_{j=1}^{J}\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}\mathbb{1}(y_{it}=s_1,y_{i,t+j}=s_2)\log\mathbb{P}\left(y_{it}=s_1,y_{i,t+j}=s_2 \mid \mathbf{x_i};\theta\right),
\end{aligned} \tag{20}$$

where t_i is the initial date with available observation for firm i, and T_i is the date when the available observations for firm i ends. I assume between t_i and T_i, there is no missing values. In theory, one can allow for missing observations even between t_i and T_i, however, modeling the persistence in an autoregressive setting with missing values in-between the observations would be extremely complicated.

4.4 Empirical Results

In this section, I compare the estimation results of static probit and AR-Probit models estimated by maximum likelihood and composite likelihood, respectively. Table 5 shows the estimates of the models. In the table, for both static and autoregressive model, the signs of financial ratios and business cycle variables are the same. Moreover, the signs are as expected a priori, except for cash/debt. The sign of cash related variables is estimated to be negative in many other papers in the literature, see for example Alp [2013]. This puzzling phenomenon is explained by Acharya et al. [2012] by observing that firms tend to have high cash holdings for precautionary motive. In-sample prediction accuracy seems slightly higher for static model compared to AR model (47% vs. 44% correctly estimated ratings). However, this measure of predictive performance can be greatly affected when there is a dominant outcome in the data. The correct prediction of each rating class reveals that this higher prediction accuracy of the static model comes from predicting accurately only the BBB rating class, which contains the largest number of observations. On the other hand, AR-Probit model yields more balanced correct prediction over rating classes. To avoid this seemingly successful predictability, I also report the average of the correct predictions over each rating class. This measure is in line with the technique proposed by Merton [1981] to distinguish between true predictability and seemingly successful predictability of a 'stopped-clock.' Based on the average correct prediction, AR-Probit model has a better prediction power overall (26% vs 22%).

The most important difference between the static and autoregressive models is, clearly, a significant and economically large persistence parameter ($\hat{\rho} = 0.62$), which indicates that there is, in fact, a need for a time series component in analyzing credit ratings. A significant ρ shows that the soundness of a company evaluated by the rating agency, represented by latent y^*_{it}, measured by its financial ratios and the current economic conditions exhibits significant persistence. Any unexpected change in the rating agency's view of the soundness of a firm has a half-life more than a quarter. Even after a year, $\hat{\rho}^4 = 15\%$ of the shock is affecting the continuous rating of the firm. In the static model, on the contrary, there is no difference between the short-term effect and the long-term effect. Any change affects the current ratings and then disappears. Therefore, the estimates of the static model reflect the "mid-term effect" of a shock. In particular, the estimates in the static model is the closest to the 6-month cumulative effects derived from the AR-Probit estimates (see Table 5).

[Table 5 here]

An important question is how a change in the unobserved state of the firm will be reflected in its ratings. Due to nonlinearity in the model, it is hard to interpret the coefficients. The impact of a shock on ratings depends on the current conditions of the firm. Considering the estimated thresholds, $(\tau_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4, \hat{\tau}_5, \hat{\tau}_6) = (0,\ 1.2,\ 2.3,\ 3.6,\ 4.8,\ 5.8)$, the estimated average distance between the rating categories is $(\hat{\tau}_6 - \tau_1)/5 = 1.17$. Note that τ_1 is set to be zero for identification purposes, in particular, to identify the constant term in the measurement equation. One should also note that this average distance is valid between the rating categories B–AA because the region for CCC is $(-\infty, 0]$, and the region for AAA is $(5.8, \infty)$. Hence, focusing on the middle rating classes, on average, a 1.17 units increase in the latent variable will result in a rating upgrade in the current quarter. In reality, a letter grade change within a quarter is an extreme case, but it is still a useful exercise to understand the dynamics of the model. Because of the persistence, the shock will also have an effect on the state of the firm in the j^{th} quarter after the shock by a magnitude of $1.17\hat{\rho}^j$. In other words, in each quarter over a year, the effect of a shock with a magnitude 1.17 will be $(1.17,\ 0.73,\ 0.45,\ 0.28,\ 0.18)$, or in cumulative terms the effect will be $(1.17,\ 1.9,\ 2.35,\ 2.63,\ 2.81)$. It means that contemporaneously there is a rating upgrade by one letter, and one more increase by the end of 6 months, and there are no upgrades afterward. However, note that I am using only letter grades instead of notches; most ratings have +/- notches as an addition. Let's assume that the notches are evenly distributed between two thresholds. As a conclusion, this analysis shows that, on average, a shock that increases a company's credit rating contemporaneously by one letter will result in a further increase in its rating by two notches in the following quarter, and another notch increase in the second quarter, and one last notch increase in the fourth quarter, which results in a 2 letter and a notch increase by the end of the year. Hence, a full letter upgrade in a quarter will have a momentum effect with another letter and a notch upgrade gradually within a year. The shock that increases a firm's rating by two letters and a notch within a year must be a very large shock, which is a rare event clearly. But based on this exercise, more interpretable and reasonable economic examples will be given below by using shocks to financial ratios and business cycle variables.

Next, I will analyze the effect of a 1-year long recession on credit ratings. In each quarter during the recession, the latent variable is down by the estimated coefficient of the recession $\hat{\beta}_{\text{rec}} = -0.27$. Moreover, due to the persistence in y^*_{it}, the negative effects of the previous periods are carried over with exponentially decaying weights. If we divide this cumulative effect by the average rating distance, we obtain the normalized cumulative effect of a recession on letter grades. The cumulative effect at the j^{th} quarter is calculated simply by the formula $\beta_{\text{rec}}(1 + \rho + \cdots + \rho^j)/(\tau_6/5)$. The estimated effect of a recession on credit ratings for each quarter is $(-0.23, -0.61, -1.07, -1.59)$. The results imply that there is one letter and two notches decrease in the credit ratings by the end of the year during a recession. If we apply the same idea to the financial ratios to see their long-run effect after a one-time change by a magnitude of their one standard deviation, we obtain the following results. In the long run, a one standard deviation change in the ratio of long-term debt to liabilities has the highest impact on credit ratings, which is a downgrade by almost three notches. The second highest impact occurs when the ratio of operating income before depreciation to assets, which reflects profitability, increases by one standard deviation. Its long-run effect is an upgrade by two notches. The valuation variable price/sales appears to affect the rating by one notch increase in the long-run. The effects of other variables are below one notch. These findings verify that financial soundness and profitability are the most important factors for credit ratings. Amato and Furfine [2004] and Alp [2013] also find that the Long-term debt ratio is one of the most influential variable on credit ratings.

Taking into account the persistence of the credit ratings, AR-Probit model results in predicting the rating transitions much more successfully than the static model. Transition matrices are very useful in credit risk models to measure future credit loss. Transition matrices are usually estimated at a one year or five year horizon. At any given horizon, AR-Probit yields a more accurate estimated transition matrix. However, for the sake of brevity, I present rating transition matrix for the entire sample in Table 7. The first matrix, say M_d, is the transition matrix obtained from the data. For instance, there are 289 instances where AAA rating is followed by the same rating, and eight downgrades from AAA to AA occurred between 2002Q1–2015Q3. The second matrix, say M_s, in the table is obtained from the predictions of static probit model, and the last one, say M_a, is obtained from AR-Probit estimates. While static model over-predicts rating transitions involving BBB and under-predicts other transition AR-Probit estimates the transitions much more accurately. To quantify the accuracy comparison, I compute the ratio of matrix norms of the difference, that is, $\|M_d - M_s\|/\|M_d - M_a\|$. A ratio larger than one means AR-Probit model yields closer estimates to true data transition matrix than the static model does. This ratio is found to be 5.3, which favors clearly for AR-Probit. A similar performance superiority can be found in pseudo-out-of-sample predictions results. In this table, I estimate both Static Probit and AR-Probit models by using the data until 2012Q2. Then, I forecast 2012Q3-2015Q3 data. The Table 8 clearly shows that the future transitions are captured much better by AR-Probit model compared to Static Probit model. By a similar comparison, the transition forecast by the latter is 3.4 times better than the former.

[Table 6 here]

[Table 7 here]

[Table 8 here]

4.4.1 Rating Stability Over Time and Through-the-cycle Ratings

This section analyzes the rating stability, implied by the persistence parameter, over time. The model with a time-varying persistence that is explained in (16)-(17) is employed for this analysis. A higher autoregressive coefficient means higher likelihood for ratings to stay where they are. Hence, more persistence in the latent variable implies more stability (thus less volatility) in the credit ratings. The estimation results are given in Table 9. The signs and magnitudes of the estimates are very close to corresponding estimates in the Panel AR Ordered Probit model in Table 5. The geometric mean of the time-varying parameters is 0.67 which also close to the estimate in constant parameter model where $\hat{\rho}$ was 0.62. Hence, on average, the estimates and the correct prediction performances of constant-parameter and time-varying parameter models are very similar. However, there are two important findings from the estimation of time-varying persistence model: 1) the stability of the ratings, in general, is declining over time, and even more so after the crisis. Hence, ratings become more timely and volatile after the crisis. 2) S&P assigns the credit ratings according to through-the-cycle approach as they claimed it (Standard and Poor's [2003]). Each of these results addresses an important aspect of the credit ratings that has attracted attention in the literature.

[Table 9 here]

The first result is about less persistent and more timely ratings, especially after the recent crisis. The yearly persistence estimates together with a simple time trend are plotted in Figure 2. The rating stability is slightly decreasing in the pre-crisis period with a steeper decline in the post-crisis period. This could be a reaction to the heavy critics directed to credit rating agencies due to their poor performance right before and during the crisis. This might also be the result of Dodd-Frank Wall Street Reform and Consumer Protection Act passed on July 2010 that aims to increase the responsibility of credit rating agencies for their poor performance, thus improving the accuracy of the ratings. However, the revisions and implementations of Dodd-Frank Act are dispersed over time. For instance, U.S. Securities and Exchange Commission made final rule-making provisions on credit rating agencies even as late as July 2014. For this reason, it is hard to tell the underlying reason of the decline in rating stabilities. But, rating agencies tend to increase the timeliness of their ratings when they face reputation concerns or new legislative restrictions (Cheng and Neamtiu [2009]).

[Figure 2 here]

[Figure 3 here]

The second result is about whether rating agencies do rate through the cycle. They try to adjust the ratings slowly to changes in corporate creditworthiness to achieve rating stability (Altman and Rijken [2006]). Figure 3 (as well as Figure 2) clearly shows this delayed adjustment during the recession. In 2008Q1, rating agencies saw some deterioration in financial ratios of the companies. Thus they face a challenging situation to make a decision. They could consider this a transitory situation and keep the ratings as they are or decrease the ratings because the creditworthiness of the firms is lower now. Through-the-cycle approach means that agencies do not react to changes in firms' financial situations right away. Hence, they wait until they believe that the change in the firm's situation is not only a short-term fluctuation. Thus, while rating a firm in 2008Q1, rating agencies put more weight on the previous level of creditworthiness than the most recent data. This means that the persistence parameter ρ_{2008Q1} should increase since it represents the weight on past creditworthiness. In fact, the estimated persistence parameter went from $\hat{\rho}_{2007Q4} = 0.656$ to $\hat{\rho}_{2008Q1} = 0.744$. In 2008Q2, balance sheet ratios deteriorated further. Thus, it is now harder to keep the ratings stable, which requires a higher ρ_{2008Q2} parameter, and its estimate increases to $\hat{\rho}_{2008Q2} = 0.751$. It may not be a significant increase, but conditional on the business cycle situation and the fact that the recession started affecting most of the economy, keeping the stability of ratings even at the same level was a tough decision. Starting 2008Q3, it appeared to the rating agencies that the changes in the firm's financial situations are more long-term than just a short-lived shock. Hence the agencies began reducing the stability of the ratings in fact with a very sharp drop to $\hat{\rho}_{2008Q4} = 0.615$. This delayed reaction of the credit rating agencies is a clear indication of the fact that they follow through-the-cycle approach for their rating assignments. This 2-3 quarter sluggish response from the rating agencies coincide with the estimates of Altman and Rijken [2006]. They show that the TTC method delays the timing of rating migrations by 0.70 year on average, which corresponds to 2.8 quarters.

As mentioned previously, the persistence of the ratings are further reduced even after the crisis; this might be a structural change in the rating assignment. That is, rating agencies might rate *less* through-the-cycle and more point-in-time to reduce the critics of missing timely rating changes and to comply with the new legislations. Whether the agencies changed their rating methodology after the crisis is an interesting topic, but I will not discuss it further in this paper.

4.5 Robustness Checks and Other Estimation Results

4.5.1 Unbalanced Panel AR-Probit Estimations

In this section, I enlarge the cross-section of the data by including firms that left the data set earlier than 2015Q3 (due to default or any other reason) and that entered the data set after 2002Q1. I do not treat the default case as an absorbing state since firms might continue doing business while and after bankruptcy.

The main findings about declining stability and rating through-the-cycle are even more supported by the unbalanced panel. Figure 4 compares two estimated series for $\hat{\rho}_t$ between 2002Q1–2015Q2 from balanced and unbalanced panel data sets. Before the crisis, the series follow each other very closely. However, once the crisis started, assuming that lower rating firms are affected more from the crises than the higher rating ones, unbalanced panel data yields higher persistence estimates since with more low-level rating firms it requires more weights from the past observations to keep the ratings stable. In other words, achieving rating stability during the crisis with higher a representation of junk bonds needs higher persistence parameter. The dashed line emphasizes more that the rating agencies rate through-the cycle. Moreover, after the crises, the dashed line is under the solid line, which indicates that the rating agencies try to rate the firms with lower ratings more timely.

[Figure 4 here]

4.5.2 Panel AR-Probit with Random Effects Estimations

Firm-specific effects might play a significant role in the creditworthiness of the firms. But one cannot estimate firm-specific fixed effects due to incidental parameter problem. Hence, random effects model is used, following the Mundlak approach. The details of the model are given in the Section 4.3.2. In the model with the random effects, it is not clear how the persistence of the creditworthiness should be calculated. The source of the persistence is not only $\rho y^*_{i,t-1}$ anymore. The random effects η_i generate persistence in the latent variable since the firm-specific effects are carried over time. In Table 10, I present random effects estimation results for both balanced and unbalanced data sets, and compare them with the panel AR-Probit model without random effects.

[Table 10 here]

4.5.3 Effect of Dodd-Frank Act on Rating Stability

The quality of ratings issued by credit rating agencies was intensely questioned, especially during and after the Great Recession. Rating agencies have been blamed for inflated ratings being that were part of the reason of 2007-2009 crisis. To incentivize the credit rating agencies to provide more accurate corporate bond ratings and to increase the authority supervision on rating agencies, Congress passed the Dodd-Frank Wall Street Reform and Consumer Protection Act in July 2010. This financial reform bill increases the accountability for inaccurate ratings and make the agencies more liable for misconduct. To assess the effects of Dodd-Frank Act on the persistence of the ratings, I use a restricted version of time-varying parameter model. The values of ρ_t before 2010Q3 are restricted to be the same, call $\rho_{\text{pre-DF}}$. Similarly, the ones after 2010Q4 are set to be the same, call $\rho_{\text{post-DF}}$. One can consider it a threshold in the persistence parameter at time 2010Q3. The results show that there is a significant difference in rating stability in pre and post Dodd-Frank Act periods. However, the autocorrelation coefficient dropping significantly from 0.673 to 0.635 may not create any significant economic difference. Assuming that stability is inversely related to accuracy, insignificant decrease in the stability is an indication of insignificant improvement in accuracy. On a similar basis, Dimitrov et al. [2015] find no significant evidence that Dodd-Frank Act disciplines rating agencies to provide more accurate and informative credit ratings.

[Table 11 here]

5 Conclusion

In this paper, I borrow a method – composite likelihood estimation – from statistics literature and bring it to economics where the method is not widely known. Composite likelihood allows me to solve complex problems in a tractable fashion. One interesting example is Autoregressive Probit model, where the discrete outcome is a non-linear function of an autocorrelated latent process. The likelihood function of this model contains high dimensional integrals, which is intractable. Composite likelihood offers a fast, reliable, and accurate estimation. I prove the consistency and asymptotic normality of the CL estimator in a Panel AR-Probit model.

In the empirical part, I apply the model to credit ratings. First of all, I compare the AR-Probit model to Static Probit model, which has been widely employed in the credit rating literature. The in-sample and out-of-sample prediction performances clearly favors for AR-Probit over the static version. Then, I extent the AR-Probit model into time-varying parameter setup to understand the rating process and see whether it has changed over time. Since credit ratings enter into many risk models, it might be useful to know how quickly the ratings are responding to new information and whether their mobility has changed over time. Time-varying autocorrelation parameter help measure these aspects. I see evidence for 2-3 quarters of delayed rating adjustment during the recent crisis. Moreover, I show that the ratings become more responsive over time possible due to the new regulations (e.g., Dodd-Frank Act) and critics on the credit rating agencies.

References

M. Abramowitz, I. A. Stegun, et al. Handbook of mathematical functions with formulas, graphs, and mathematical tables. *National Bureau of Standards Applied Mathematics Series*, 55, 1972.

V. Acharya, S. A. Davydenko, and I. A. Strebulaev. Cash holdings and credit risk. *Review of Financial Studies*, 25(12):3572–3609, 2012.

A. Alp. Structural shifts in credit rating standards. *The Journal of Finance*, 68(6):2435–2470, 2013.

E. I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4):589–609, 1968.

E. I. Altman and H. A. Rijken. A point-in-time perspective on through-the-cycle ratings. *Financial Analysts Journal*, 62(1):54–70, 2006.

J. D. Amato and C. H. Furfine. Are credit ratings procyclical? *Journal of Banking & Finance*, 28 (11):2641–2677, 2004.

P. Amato, G. Ferri, P. Lacitignola, and N. Sinisi. Searching for consistent indices of rating mobility through the crisis. 2013.

D. W. Andrews. Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67(3):543–563, 1999.

J. K. Auh. Procyclical credit rating policy. *Georgetown McDonough School of Business Research Paper*, (2581738), 2015.

A. Bangia, F. X. Diebold, A. Kronimus, C. Schagen, and T. Schuermann. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445–474, 2002.

N. Beck, D. Epstein, S. Jackman, and S. Ohalloran. Alternative models of dynamics in binary time-series-cross-section models: The example of state failure. In *annual meeting of the Society for Political Methodology, Emory University*, 2001.

K. Bel, D. Fok, and R. Paap. Parameter estimation in multivariate logit models with many binary choices. *Econometric Reviews*, pages 1–17, 2016.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

C. R. Bhat, C. Varin, and N. Ferdous. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. *Advances in Econometrics*, 26:65, 2010.

F. Canova and C. Matthes. A composite likelihood approach for dynamic structural models. Technical report, 2016.

R. Cantor and C. Mann. Analyzing the tradeoff between ratings accuracy and stability. *Journal of Fixed Income, September*, 2006.

M. Carey and M. Hrycay. Parameterizing credit risk models with rating data. *Journal of banking & finance*, 25(1):197–270, 2001.

R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183, 2007.

M. Chauvet and S. Potter. Forecasting recessions using the yield curve. *Journal of Forecasting*, 24 (2):77–103, 2005.

M. Cheng and M. Neamtiu. An empirical analysis of changes in credit rating properties: Timeliness, accuracy and volatility. *Journal of Accounting and Economics*, 47(1):108–130, 2009.

D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

R. A. Davis and C. Y. Yau. Comments on pairwise likelihood in time series models. *Statistica Sinica*, pages 255–277, 2011.

R. M. de Jong and T. Woutersen. Dynamic time series binary choice. *Econometric Theory*, 27(04): 673–702, 2011.

V. Dimitrov, D. Palia, and L. Tang. Impact of the dodd-frank act on credit ratings. *Journal of Financial Economics*, 115(3):505–520, 2015.

J.-C. Duan, J. Sun, and T. Wang. Multiperiod corporate default predictiona forward intensity approach. *Journal of Econometrics*, 170(1):191–209, 2012.

M. Dueker. Dynamic forecasts of qualitative variables: a qual var model of us recessions. *Journal of Business & Economic Statistics*, 23(1):96–104, 2005.

M. J. Dueker. Strengthening the case for the yield curve as a predictor of us recessions. *Review*, 79, 1997.

R. F. Engle, N. Shephard, and K. Sheppard. Fitting vast dimensional time-varying covariance models. 2008.

W. Feller. An introduction to probability and its applications, vol. ii. *Wiley, New York*, 1971.

D. Feng, C. Gouriéroux, and J. Jasiak. The ordered qualitative model for credit rating transitions. *Journal of Empirical Finance*, 15(1):111–130, 2008.

J.-D. Fermanian and B. Salanié. A nonparametric simulated maximum likelihood estimation method. *Econometric Theory*, 20(04):701–734, 2004.

G. Ferri, L.-G. Liu, and J. E. Stiglitz. The procyclical role of rating agencies: Evidence from the east asian crisis. *Economic Notes*, 28(3):335–355, 1999.

L. Freitag. Procyclicality and path dependence of sovereign credit ratings: The example of europe. *Economic Notes*, 44(2):309–332, 2015.

H. Frydman and T. Schuermann. Credit rating dynamics and markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075, 2008.

P. Gagliardini and C. Gouriéroux. Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics*, 3(2):188–226, 2005.

P. Gagliardini and C. Gourieroux. Efficiency in large dynamic panel models with common factors. *Econometric Theory*, 30(05):961–1020, 2014.

X. Gao and P. X.-K. Song. Composite likelihood em algorithm with applications to multivariate hidden markov model. *Statistica Sinica*, pages 165–185, 2011.

V. P. Godambe. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211, 1960.

C. Gourieroux and A. Monfort. *Simulation-based econometric methods*. Oxford university press, 1996.

C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, pages 681–700, 1984.

V. A. Hajivassiliou and P. A. Ruud. Classical estimation methods for ldv models using simulation. *Handbook of econometrics*, 4:2383–2441, 1994.

P. Hall, L. Peng, and Q. Yao. Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series Analysis*, 23(3):313–331, 2002.

S. Harden. *Weighted composite likelihoods*. PhD thesis, UCL (University College London), 2013.

A. Heinen, J. Kau, D. C. Keenan, M. L. Kim, and C. Slawson. Contagion in subprime mortgage defaults: a composite likelihood approach. *Contagion*, 2014.

N. L. Hjort and C. Varin. Ml, pl, ql in markov chain models. *Scandinavian Journal of Statistics*, 35(1):64–82, 2008.

Y. Jafry and T. Schuermann. Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance*, 28(11):2603–2639, 2004.

H. Joe and Y. Lee. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100(4):670–685, 2009.

H. Kauppi and P. Saikkonen. Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791, 2008.

J. T. Kent. Robust properties of likelihood ratio tests. *Biometrika*, 69(1):19–27, 1982.

J. Kiff, M. Kisser, and M. L. Schumacher. *Rating through-the-cycle: What does the concept imply for rating stability and accuracy?* Number 13-64. International Monetary Fund, 2013.

S. J. Koopman and A. Lucas. Business and default cycles for credit risk. *Journal of Applied Econometrics*, 20(2):311–323, 2005.

S. J. Koopman, R. Kräussl, A. Lucas, and A. B. Monteiro. Credit cycles and macro fundamentals. *Journal of Empirical Finance*, 16(1):42–54, 2009.

S. J. Koopman, A. Lucas, and B. Schwaab. Dynamic factor models with macro, frailty, and industry effects for us default counts: the credit crisis of 2008. *Journal of Business & Economic Statistics*, 30(4):521–532, 2012.

D. Lando and T. M. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *journal of banking & finance*, 26(2):423–444, 2002.

S. Le Cessie and J. Van Houwelingen. Logistic regression for correlated binary data. *Applied Statistics*, pages 95–108, 1994.

L.-F. Lee. Simulation estimation of dynamic switching regression and dynamic disequilibrium modelssome monte carlo results. *Journal of Econometrics*, 78(1):179–204, 1997.

B. G. Lindsay. Composite likelihood methods. *Contemporary mathematics*, 80(1):221–39, 1988.

B. G. Lindsay, G. Y. Yi, and J. Sun. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, pages 71–105, 2011.

G. Löffler. An anatomy of rating through the cycle. *Journal of Banking & Finance*, 28(3):695–720, 2004.

G. Löffler. Avoiding the rating bounce: Why rating agencies are slow to react to new information. *Journal of Economic Behavior & Organization*, 56(3):365–381, 2005.

G. Löffler. Can rating agencies look through the cycle? *Review of Quantitative Finance and Accounting*, 40(4):623–646, 2013.

K. V. Mardia, G. Hughes, and C. C. Taylor. Efficiency of the pseudolikelihood for multivariate normal and von mises distributions. *Preprint*, 2007.

L. Matyas and P. Sevestre. The econometrics of panel data. *A Handbook of the Theory with Applications*, 1996.

M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, (just-accepted), 2015.

A. J. McNeil and J. P. Wendin. Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance*, 14(2):131–149, 2007.

R. C. Merton. On market timing and investment performance. i. an equilibrium theory of value for market forecasts. *Journal of business*, pages 363–406, 1981.

P. Mizen and S. Tsoukas. Modelling the persistence of credit ratings when firms face financial constraints, recessions and credit crunches. Technical report, 2009.

G. Molenberghs and G. Verbeke. Models for discrete longitudinal data. Technical report, Springer,, 2005.

G. Müller and C. Czado. An autoregressive ordered probit model with application to high-frequency financial data. *Journal of Computational and Graphical Statistics*, 2012.

C. T. Ng, H. Joe, D. Karlis, and J. Liu. Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica*, pages 279–305, 2011.

P. Nickell, W. Perraudin, and S. Varotto. Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227, 2000.

D. H. Oh and A. J. Patton. High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 2016.

M. O. Ojo. A remark on the convolution of the generalized logistic random variables. *ASSET serves A*, 1(2), 2003.

L. Pace, A. Salvan, and N. Sartori. Adjusting composite likelihood ratio statistics. *Statistica Sinica*, pages 129–148, 2011.

C. Pakel, N. Shephard, and K. Sheppard. Nuisance parameters, composite likelihoods and a panel of garch models. *Statistica Sinica*, pages 307–329, 2011.

Z. Qu. A composite likelihood framework for analyzing singular dsge models. Technical report, Boston University-Department of Economics, 2015.

P. Reusens and C. Croux. Sovereign credit rating determinants: the impact of the european debt crisis. *Available at SSRN 2777491*, 2016.

Standard and Poor's. Corporate ratings criteria. *RatingsDirect*, 2002.

Standard and Poor's. Corporate ratings criteria. *RatingsDirect*, 2003.

Standard and Poor's. Corporate methodology: Ratios and adjustments. *RatingsDirect*, 2013.

C. Stefanescu, R. Tunaru, and S. Turnbull. The credit rating process and estimation of transition probabilities: A bayesian approach. *Journal of Empirical Finance*, 16(2):216–234, 2009.

T. M. Stoker. Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society*, pages 1461–1481, 1986.

R. Topp and R. Perl. Through-the-cycle ratings versus point-in-time ratings and implications of the mapping between both rating types. *Financial Markets, Institutions & Instruments*, 19(1): 47–61, 2010.

T. Van Gestel, D. Martens, B. Baesens, D. Feremans, J. Huysmans, and J. Vanthienen. Forecasting and analyzing insurance companies' ratings. *International Journal of Forecasting*, 23(3):513–529, 2007.

C. Varin and C. Czado. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, page kxp042, 2009.

C. Varin and P. Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.

C. Varin and P. Vidoni. Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics & Data Analysis*, 51(4):2365–2373, 2006.

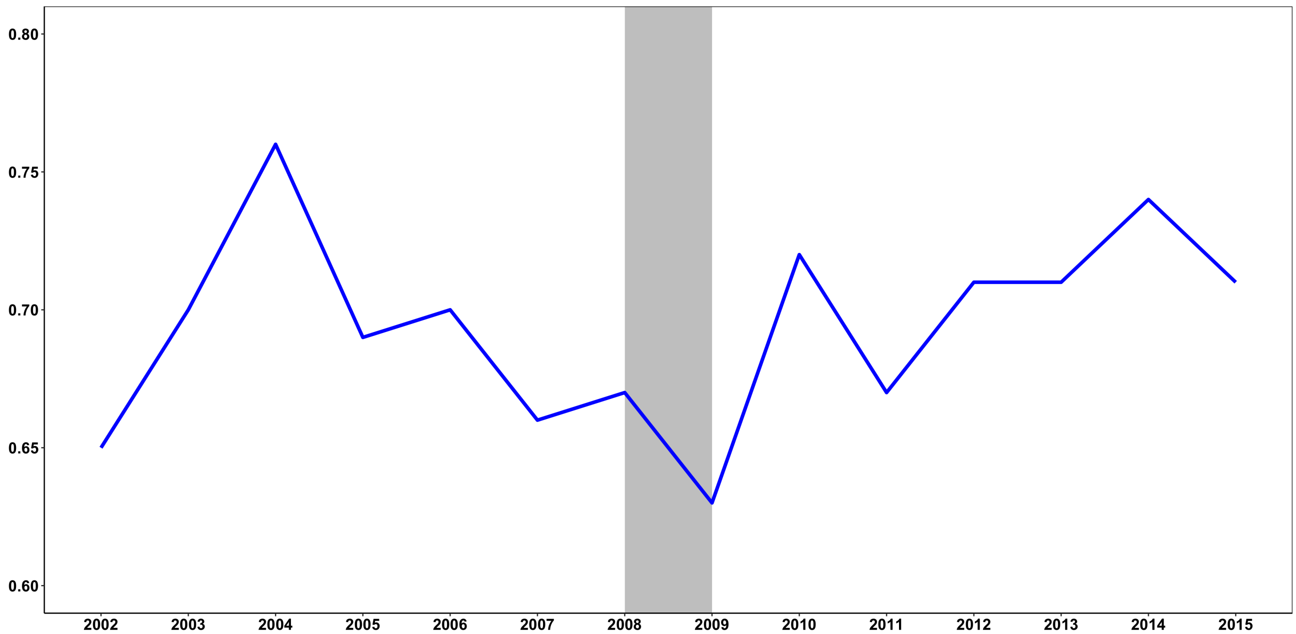
C. Varin and P. Vidoni. Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28(1-3):170–185, 2008.

C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42, 2011.

V. G. Vasdekis, S. Cagnone, and I. Moustaki. A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika*, 77(3):425–441, 2012.

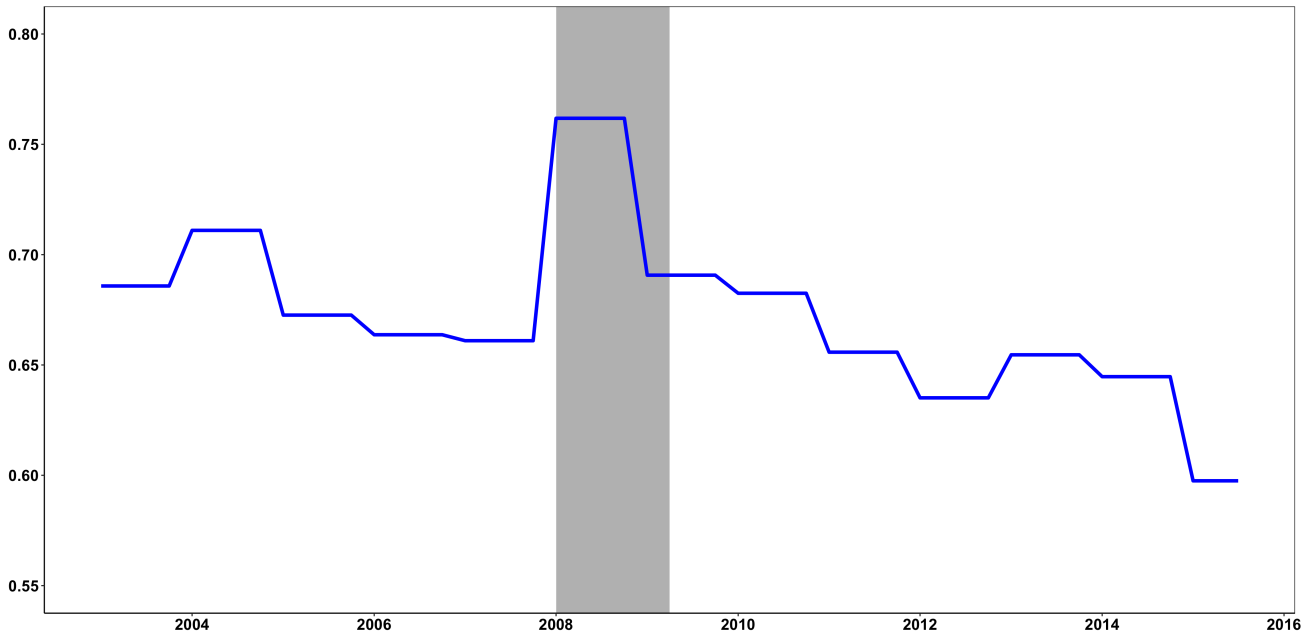
6 Figures

Figure 1: Percentage of issuers with unchanged ratings (year-to-year)



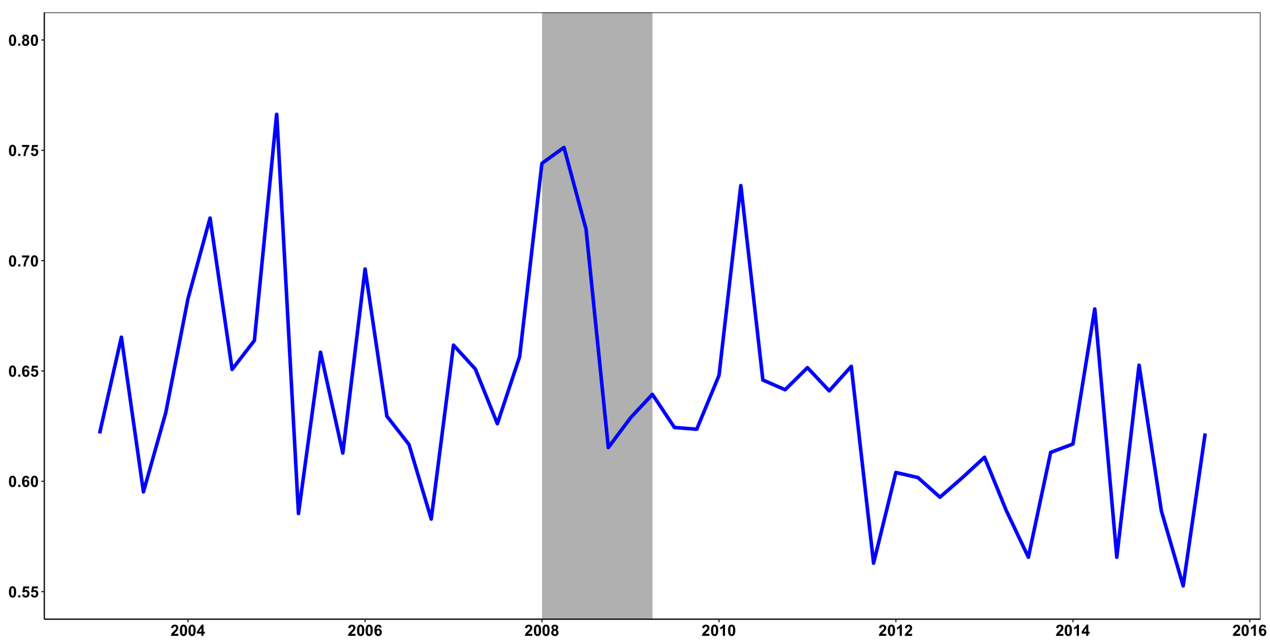
Note: This figure shows the percentage of issuers whose ratings remain unchanged from year to year. These data are taken directly from annual public S&P reports (see 2015 Annual Global Corporate Default Study And Rating Transitions)

FIGURE 2: TIME-VARYING PERSISTENCE ρ_t (YEARLY)



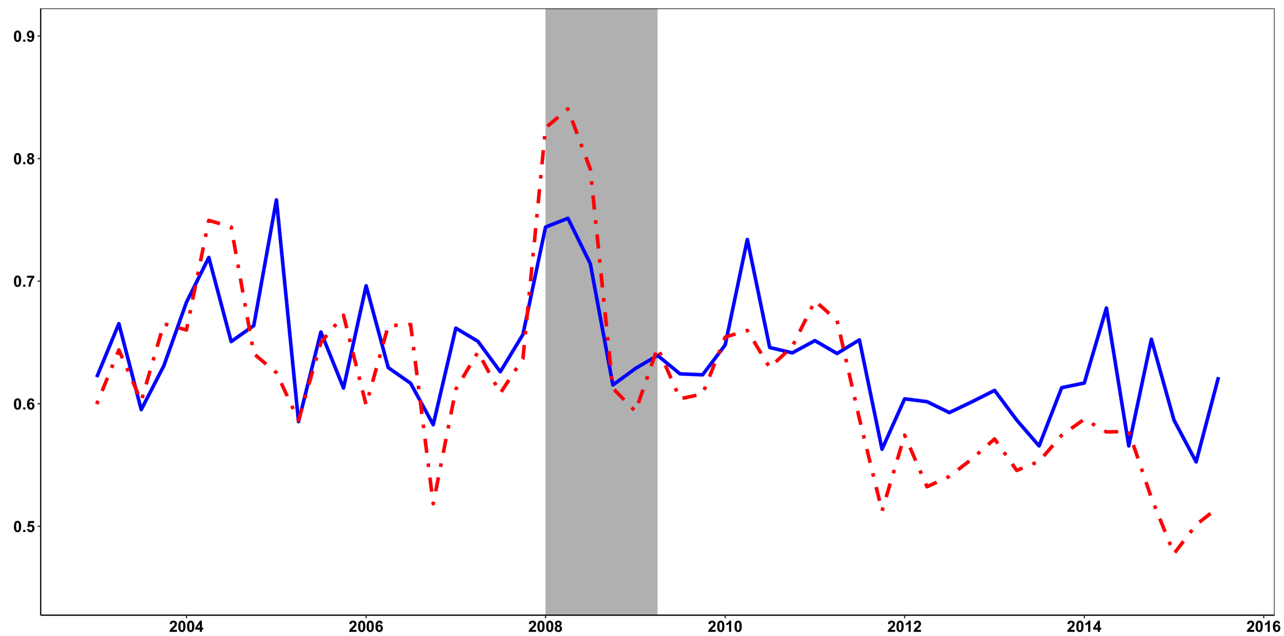
Note: This figure shows how the stability of the ratings change over time. Time-varying persistence parameter ρ_t is assumed to be constant within a year

FIGURE 3: TIME-VARYING PERSISTENCE ρ_t (QUARTERLY)



Note: This figure shows how the stability of the ratings change over time. Stability is measured by the time-varying parameter ρ_t.

Figure 4: Time-varying persistence ρ_t – Balanced vs Unbalanced Panel



Note: This figure compares the time-varying persistence parameter estimates $\hat{\rho}_t$ from balanced panel (solid) versus unbalanced panel data (dashed). The correlation between the two series is 78%.

7 Tables

TABLE 1: $(N, T, J, K) = (500, 50, 4, 3)$, #Sim=200, (#moments=870)

	ESTIMATES				RMSE		
	TRUE	CLE	E-GMM	GMM	CLE	E-GMM	GMM
ρ	0.8	0.796	0.769	0.771	**0.005**	0.034	0.033
β_0	0.8	0.856	0.968	0.945	**0.059**	0.170	0.146
β_1	0.0	-0.000	0.001	0.000	**0.006**	0.012	0.011
β_2	-0.8	-0.850	-0.788	-0.783	0.051	**0.029**	0.042

TABLE 2: Financial ratios affecting the credit ratings of corporate firms

Class	**Variable**	**Description**
Solvency	debt/assets	Total debt as a fraction of total assets
Fin. Soundness	ltd/debt	Long-term debt as a fraction of total debt
Fin. Soundness	cash/debt	Operating cash flow as a fraction of total debt
Profitability	return on assets	Operating income before depreciation as a fraction of total assets
Profitability	profit margin	Net income (income after interest and taxes) as a fraction of sales
Valuation	price/sales	Market capitalization as a fraction of total sales

Table 3: Descriptive statistics of explanatory variables by rating category

		Mean	25%	Median	75%	Std. Dev.			Mean	25%	Median	75%	Std. Dev.
	CCC	0.43	0.23	0.45	0.60	0.21		CCC	0.35	0.07	0.20	0.41	0.43
	B	0.56	0.40	0.60	0.74	0.22		B	0.86	0.26	0.50	1.13	1.00
	BB	0.48	0.35	0.49	0.62	0.19		BB	1.11	0.43	0.78	1.32	1.18
ltd/debt	BBB	0.36	0.23	0.38	0.48	0.17	price/sales	BBB	1.40	0.68	1.12	1.79	1.11
	A	0.29	0.16	0.30	0.42	0.16		A	1.84	0.98	1.54	2.41	1.25
	AA	0.26	0.16	0.28	0.35	0.13		AA	2.11	1.07	2.06	2.84	1.18
	AAA	0.14	0.06	0.12	0.20	0.11		AAA	2.53	1.27	2.43	3.30	1.36
	CCC	0.00	-0.05	0.03	0.07	0.10		CCC	-0.35	-0.63	-0.10	0.15	1.04
	B	0.10	0.06	0.09	0.13	0.08		B	-0.02	-0.43	-0.01	0.45	0.82
	BB	0.13	0.09	0.12	0.16	0.09		BB	0.02	-0.38	-0.01	0.53	0.81
return on assets	BBB	0.13	0.08	0.12	0.17	0.08	factor1	BBB	0.00	-0.43	-0.01	0.53	0.81
	A	0.14	0.09	0.14	0.19	0.08		A	0.02	-0.43	-0.01	0.53	0.80
	AA	0.16	0.07	0.18	0.23	0.09		AA	0.02	-0.38	0.00	0.59	0.83
	AAA	0.16	0.07	0.18	0.23	0.09		AAA	0.05	-0.29	0.00	0.59	0.84
	CCC	0.02	-0.04	0.03	0.07	0.13		CCC	-0.31	-0.87	-0.45	0.28	0.71
	B	0.10	0.03	0.07	0.13	0.20		B	-0.25	-0.87	-0.32	0.25	0.90
	BB	0.14	0.07	0.12	0.19	0.13		BB	-0.18	-0.79	-0.29	0.28	0.90
cash/debt	BBB	0.16	0.08	0.14	0.21	0.12	factor2	BBB	-0.21	-0.82	-0.30	0.25	0.90
	A	0.20	0.09	0.17	0.27	0.17		A	-0.20	-0.82	-0.30	0.28	0.89
	AA	0.21	0.06	0.21	0.32	0.16		AA	-0.10	-0.68	-0.19	0.32	0.91
	AAA	0.25	0.08	0.28	0.38	0.17		AAA	-0.01	-0.49	-0.13	0.35	0.85
	CCC	-0.27	-0.38	-0.12	-0.04	0.46		CCC	0.19	0.00	0.00	0.00	0.40
	B	-0.06	-0.05	0.00	0.04	0.41		B	0.11	0.00	0.00	0.00	0.32
	BB	0.03	0.01	0.04	0.07	0.14		BB	0.11	0.00	0.00	0.00	0.31
profit margin	BBB	0.07	0.03	0.06	0.10	0.24	recession	BBB	0.11	0.00	0.00	0.00	0.31
	A	0.09	0.06	0.09	0.14	0.58		A	0.10	0.00	0.00	0.00	0.30
	AA	0.12	0.07	0.12	0.15	0.07		AA	0.12	0.00	0.00	0.00	0.33
	AAA	0.13	0.10	0.12	0.17	0.05		AAA	0.11	0.00	0.00	0.00	0.32
	CCC	0.87	0.77	0.83	0.99	0.22							
	B	0.78	0.61	0.76	0.92	0.28							
	BB	0.67	0.55	0.64	0.77	0.21							
debt/assets	BBB	0.65	0.54	0.64	0.74	0.16							
	A	0.64	0.50	0.63	0.77	0.19							
	AA	0.64	0.52	0.61	0.82	0.19							
	AAA	0.59	0.47	0.52	0.79	0.16							

Table 4: Frequency of ratings (2002Q1-2015Q3)

2002Q1-2015Q3 Surviving Firms			2002Q1-2015Q3 All Firms		
Rating	# of Obs	Percentage	Rating	# of Obs	Percentage
D	(—)	(—)	D	389	0.4%
CCC	41	0.6%	CCC	1863	1.8%
B	613	8.5%	B	18190	17.3%
BB	1473	20.3%	BB	25355	24.1%
BBB	2963	40.8%	BBB	31201	29.7%
A	1796	24.8%	A	21996	20.9%
AA	292	4.0%	AA	5059	4.8%
AAA	77	1.1%	AAA	1144	1.1%
N	T	NT	N	Avg T_i	NT
516	55	28380	1406	38	53999

TABLE 5: Estimation Results: Static vs AR-Probit

	Static	AR-Probit	6-month Cum.Effect $\beta(1+\rho)$	Long-run Effect $\beta/(1-\rho)$
$\hat{\rho}$	—	0.62		
debt/assets	-0.77	-0.37	-0.60	-0.98
ltd/debt	-3.04	-1.62	-2.63	-4.30
return on assets	3.70	2.57	4.17	6.82
cash/debt	-0.57	-0.53	-0.86	-1.41
profit margin	0.40	0.21	0.34	0.56
price/sales	0.15	0.11	0.18	0.29
factor1	-0.03	-0.04		
factor2	0.04	0.08		
recession	-0.11	-0.27		
const	4.56	1.75		
$\hat{\tau}_2$	1.64	1.20		
$\hat{\tau}_3$	2.64	2.33		
$\hat{\tau}_4$	3.94	3.57		
$\hat{\tau}_5$	5.26	4.78		
$\hat{\tau}_6$	6.06	5.83		
% of correct prediction	47%	44%		
CCC	2%	7%		
B	12%	30%		
BB	32%	41%		
BBB	78%	56%		
A	29%	39%		
AA	0%	7%		
AAA	0%	3%		
Average of correct prediction	22%	26%		

Table 6: Actual vs Predicted Ratings

		$\widehat{\text{CCC}}$	$\widehat{\text{B}}$	$\widehat{\text{BB}}$	$\widehat{\text{BBB}}$	$\widehat{\text{A}}$	$\widehat{\text{AA}}$	$\widehat{\text{AAA}}$	CorPred (44%)
AR-Probit	CCC	10	53	65	18	0	0	0	7%
	B	84	719	993	480	87	16	1	30%
	BB	37	455	2342	2293	576	59	15	41%
	BBB	9	230	1950	6495	2607	289	32	56%
	A	4	64	476	3196	2773	444	77	39%
	AA	0	6	67	254	729	74	2	7%
	AAA	0	0	10	43	114	122	10	3%
									Avg = 26%

		$\widehat{\text{CCC}}$	$\widehat{\text{B}}$	$\widehat{\text{BB}}$	$\widehat{\text{BBB}}$	$\widehat{\text{A}}$	$\widehat{\text{AA}}$	$\widehat{\text{AAA}}$	CorPred (47%)
Static Probit	CCC	3	14	82	47	0	0	0	2%
	B	2	286	1269	747	73	3	0	12%
	BB	0	102	1860	3485	320	10	0	32%
	BBB	1	11	871	9039	1671	15	4	78%
	A	4	1	101	4808	2061	57	2	29%
	AA	0	0	13	556	560	3	0	0%
	AAA	0	0	0	52	238	9	0	0%
									Avg = 22%

Table 7: Rating Transition Matrix (entire sample)

		CCC	B	BB	BBB	A	AA	AAA
Data	CCC	121	22	0	0	0	0	0
	B	19	2244	68	0	0	0	0
	BB	4	80	5510	84	0	0	0
	BBB	0	2	95	11233	63	0	0
	A	0	0	0	106	6789	12	0
	AA	0	0	0	0	31	1084	0
	AAA	0	0	0	0	0	8	289
		CCC	B	BB	BBB	A	AA	AAA
AR-Probit	CCC	122	21	0	0	0	0	0
	B	20	1088	390	0	0	0	0
	BB	0	101	5021	671	1	0	0
	BBB	0	1	298	11752	482	0	0
	A	1	0	1	350	6308	104	0
	AA	0	0	0	0	95	873	27
	AAA	0	0	0	0	0	27	110
		CCC	B	BB	BBB	A	AA	AAA
Static Probit	CCC	5	3	1	0	0	0	0
	B	3	340	59	3	0	0	0
	BB	0	64	3712	341	0	0	0
	BBB	2	1	326	17677	382	0	0
	A	0	1	1	386	4433	21	0
	AA	0	0	0	0	27	69	1
	AAA	0	0	0	0	0	3	3

Table 8: 3-Year Forecast (2012Q3-2015Q3) – Rating Transition Matrix

		CCC	B	BB	BBB	A	AA	AAA
Data	CCC	23	3	0	0	0	0	0
	B	3	537	12	0	0	0	0
	BB	0	8	1058	11	0	0	0
	BBB	0	0	10	2394	20	0	0
	A	0	0	0	10	1390	2	0
	AA	0	0	0	0	0	167	0
	AAA	0	0	0	0	0	1	27

		CCC	B	BB	BBB	A	AA	AAA
AR-Probit	CCC	12	2	0	0	0	0	0
	B	1	333	255	0	0	0	0
	BB	0	12	1476	405	0	0	0
	BBB	0	0	29	2128	152	0	0
	A	0	0	0	23	802	9	0
	AA	0	0	0	0	0	37	0
	AAA	0	0	0	0	0	0	0

		CCC	B	BB	BBB	A	AA	AAA
Static Probit	CCC	0	0	0	0	0	0	0
	B	1	89	9	0	0	0	0
	BB	0	14	728	59	0	0	0
	BBB	0	0	77	3745	71	0	0
	A	0	0	0	78	805	0	0
	AA	0	0	0	0	0	0	0
	AAA	0	0	0	0	0	0	0

Table 9: Time-varying Persistence Parameter

	AR-Probit (ρ_t)
$\hat{\rho}_{2003}$	0.686
$\hat{\rho}_{2004}$	0.711
$\hat{\rho}_{2005}$	0.673
$\hat{\rho}_{2006}$	0.664
$\hat{\rho}_{2007}$	0.661
$\hat{\rho}_{2008}$	0.762
$\hat{\rho}_{2009}$	0.691
$\hat{\rho}_{2010}$	0.683
$\hat{\rho}_{2011}$	0.656
$\hat{\rho}_{2012}$	0.635
$\hat{\rho}_{2013}$	0.655
$\hat{\rho}_{2014}$	0.645
$\hat{\rho}_{2015}$	0.598
debt/assets	-0.347
ltd/debt	-1.602
return on assets	2.681
cash/debt	-0.676
profit margin	0.202
price/sales	0.097
factor1	-0.077
factor2	-0.010
recession	-0.400
const	1.710
$\hat{\tau}_2$	1.432
$\hat{\tau}_3$	2.744
$\hat{\tau}_4$	4.136
$\hat{\tau}_5$	5.624
$\hat{\tau}_6$	6.750
% of correct prediction	44%
CCC	8%
B	29%
BB	40%
BBB	58%
A	39%
AA	6%
AAA	6%
Average correct prediction	26%
$N \times T$	28380

TABLE 10: Estimation Results for AR-Probit and AR-Probit with Random Effects

	Balanced Panel		Unbalanced Panel	
	AR-Probit	AR-Probit RE	AR-Probit	AR-Probit RE
$\hat{\rho}$	0.623	0.500	0.782	0.484
debt/assets	-0.368	-0.950	-0.282	-0.412
ltd/debt	-1.622	-0.160	-1.122	-0.143
return on assets	2.571	1.921	2.169	1.757
cash/debt	-0.535	0.304	-0.256	2.770
profit margin	0.207	0.016	0.000	0.030
price/sales	0.113	0.017	0.002	0.097
$\overline{\text{d/a}}$	—	0.071	—	-0.036
$\overline{\text{ltd/d}}$	—	-2.711	—	-2.171
$\overline{\text{roa}}$	—	1.858	—	1.042
$\overline{\text{c/d}}$	—	-1.771	—	0.049
$\overline{\text{npm}}$	—	2.074	—	-0.254
$\overline{\text{p/s}}$	—	0.119	—	-0.118
factor1	-0.035	-0.049	-0.141	0.089
factor2	0.080	-0.011	0.161	0.333
recession	-0.266	-0.275	-0.661	-0.727
const	1.753	3.472	1.524	4.564
$\hat{\sigma}_{\varepsilon}$	—	0.587	—	0.565
$\hat{\tau}_2$	1.200	1.775	0.578	2.352
$\hat{\tau}_3$	2.325	3.312	2.082	2.910
$\hat{\tau}_4$	3.574	5.288	3.430	3.827
$\hat{\tau}_5$	4.784	7.228	4.893	5.416
$\hat{\tau}_6$	5.831	9.326	6.373	7.397
$\hat{\tau}_7$	—	—	7.477	8.450

Table 11: Dodd-Frank Threshold in Persistence Parameter ρ

	AR-Probit
$\hat{\rho}_{\text{pre-DF}}$	0.673
$\hat{\rho}_{\text{post-DF}}$	0.635
debt/assets	-0.388
ltd/debt	-1.542
return on assets	2.346
cash/debt	-0.492
profit margin	0.190
price/sales	0.087
factor1	-0.104
factor2	0.035
recession	-0.349
const	1.741
$\hat{\tau}_2$	1.311
$\hat{\tau}_3$	2.517
$\hat{\tau}_4$	3.872
$\hat{\tau}_5$	5.183
$\hat{\tau}_6$	6.209
% of correct prediction	44%
CCC	7%
B	30%
BB	41%
BBB	59%
A	35%
AA	5%
AAA	3%
Average correct prediction	26%

A Technical Appendix

Let's re-write the likelihood and the associated probabilities.

$$\mathcal{L}_c(\theta|y,x) = \frac{1}{N}\sum_{i=1}^{N} \ell_i(\theta|\mathbf{y_i},\mathbf{x_i}),$$

$$\ell_i(\theta|\mathbf{y_i},\mathbf{x_i}) = \sum_{t=1}^{T-J}\sum_{j=1}^{J} \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)$$

$$\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta) = \mathbb{1}_{00}\log \mathbb{P}_{00} + \cdots \mathbb{1}_{11}\log \mathbb{P}_{11}$$

$$\begin{aligned}
P_{00} &\equiv \mathbb{P}\left(y_{it}=0, y_{i,t+j}=0|\mathbf{x_i};\theta\right) = \mathbf{\Phi}_2\left(m_{i,t}(\theta), m_{i,t+j}(\theta)|r(\theta)\right)\\
P_{10} &\equiv \mathbb{P}\left(y_{it}=1, y_{i,t+j}=0|\mathbf{x_i};\theta\right) = \mathbf{\Phi}(m_{i,t+j}(\theta)) - P_{00}\\
P_{01} &\equiv \mathbb{P}\left(y_{it}=0, y_{i,t+j}=1|\mathbf{x_i};\theta\right) = \mathbf{\Phi}(m_{i,t}(\theta)) - P_{00}\\
P_{11} &\equiv \mathbb{P}\left(y_{it}=1, y_{i,t+j}=1|\mathbf{x_i};\theta\right) = 1-\mathbf{\Phi}(m_{i,t}(\theta)) - \mathbf{\Phi}(m_{i,t+j}(\theta)) + P_{00}
\end{aligned}$$

A.1 The Consistency Proof

Lemma 1. $\mathbb{E}\left[\sup_{\theta\in\Theta}|\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)|\right] < \infty$

Proof of Lemma 1. First, I find an upper bound for $|\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)|$ in terms of bivariate probabilities. Next, I will show that each of the log bivariate probabilities are bounded by the same term. To keep the notation short, I will suppress the parameters and the data dependencies in the probabilities whenever explicit dependency is not needed. For example, $\mathbb{P}_{kl}$ will denote $\mathbb{P}\left(y_{it}=k, y_{i,t+j}=l \mid \mathbf{x_i};\theta\right)$ for $\{k,l\}\in\{0,1\}$. Similarly, $\mathbb{1}_{kl}$ denotes $\mathbb{1}(y_{it}=k, y_{i,t+j}=l)$.

$$\begin{aligned}
\left|\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)\right| &= \left|\mathbb{1}_{00}\log\mathbb{P}_{00} + \mathbb{1}_{10}\log\mathbb{P}_{10} + \mathbb{1}_{01}\log\mathbb{P}_{01} + \mathbb{1}_{11}\log\mathbb{P}_{11}\right|\\
&\le \left|\log\mathbb{P}_{00}\right| + \cdots + \left|\log\mathbb{P}_{11}\right|
\end{aligned}$$

Let's begin analyzing the probability $\log P_{00}$ by using the mean value expansion. For a $\bar\theta$ between θ and 0,

$$\left|\log\mathbb{P}_{00}(\theta)\right| = \left|\log\mathbb{P}_{00}(0) + \theta'\frac{\partial\log\mathbb{P}_{00}(\bar\theta)}{\partial\theta}\right| = \left|\log\mathbf{\Phi}_2(0,0|0) + \theta'\frac{\frac{\partial\mathbb{P}_{00}(\bar\theta)}{\partial\theta}}{P_{00}(\bar\theta)}\right| \le \frac{1}{4} + \|\theta\|\left\|\frac{\frac{\partial\mathbb{P}_{00}(\bar\theta)}{\partial\theta}}{P_{00}(\bar\theta)}\right\|$$

Let's focus on the last norm. Using the equation (28) in the section B, which provides the first derivative of P_{00}, yields

$$\begin{aligned}\left\|\frac{\frac{\partial}{\partial\theta}\mathbb{P}_{00}}{P_{00}}\right\| = \left\|\frac{\frac{\partial}{\partial\theta}\mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}{\mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right\| \leq \|m'_{i,t}\|&\left|\frac{\phi(m_{i,t})\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right| \\ + \|m'_{i,t+j}\|&\left|\frac{\phi(m_{i,t+j})\mathbf{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right| \\ + \frac{\|r'\|}{\sqrt{1-r^2}}&\left|\frac{\phi(m_{i,t})\phi\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right| \end{aligned} \tag{21}$$

We need to find upperbounds for the terms in the absolute value. The idea of these upperbounds can be understood when one considers the special case of no autocorrelation, i.e., when $\rho = 0$. Hence, if $r = \rho^j = 0$, then the ratios in (21) become

$$\frac{\phi\left(m_{i,t}\right)\mathbf{\Phi}(m_{i,t+j})}{\mathbf{\Phi}(m_{i,t})\mathbf{\Phi}(m_{i,t+j})} \leq c(1+|m_{i,t}|) \text{ and } \frac{\phi\left(m_{i,t}\right)\phi(m_{i,t+j})}{\mathbf{\Phi}(m_{i,t})\mathbf{\Phi}(m_{i,t+j})} \leq c(1+\max\{m_{i,t}^2, m_{i,t+j}^2\})$$

A non-zero autocorrelation does not change the limiting behavior of these ratios. The details are given in the subsection B.2, in particular in the equations (39) and (41).

$$\begin{aligned}\left\|\frac{\frac{\partial}{\partial\theta}\mathbb{P}_{00}}{P_{00}}\right\| &\leq c\|m'_{i,t}\|\left(1+\max\{|m_{i,t}|,|m_{i,t+j}|\}\right) \\ &+ c\|m'_{i,t+j}\|\left(1+\max\{|m_{i,t}|,|m_{i,t+j}|\}\right) \\ &+ c\frac{\|r'\|}{\sqrt{1-r^2}}\left(1+\max\{m_{i,t}^2, m_{i,t+j}^2\}\right)\end{aligned}$$

The same upperbounds are found for other probabilities. For instance, consider the first ratio in $\|\mathbb{P}'_{10}/\mathbb{P}_{10}\|$ when $r = 0$.

$$\frac{\phi(m_{i,t})\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}(m_{i,t+j}) - \mathbf{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \overset{(r=0)}{=} \frac{\phi(m_{i,t})\mathbf{\Phi}(m_{i,t+j})}{\mathbf{\Phi}(m_{i,t+j}) - \mathbf{\Phi}(m_{i,t})\mathbf{\Phi}(m_{i,t+j})} = \frac{\phi(-m_{i,t})}{\mathbf{\Phi}(-m_{i,t})} \leq c(1+|m_{i,t}|)$$

Thus, the limiting behavior of the ratios involving normal pdf and cdf's is common for each probability. Therefore, if we combine all four bivariate probabilities and the associated upperbpunds, we obtain an upper bound for the log likelihood in terms of m and its derivative m'. Without loss of generality, let's assume that $|m_{i,t+j}| \leq |m_{i,t}|$ and $\|m'_{i,t+j}\| \leq \|m'_{i,t}\|$.

$$\begin{aligned}\left|\log f(y_{it},y_{i,t+j}|\mathbf{x_i};\theta)\right| \leq \left|\log\mathbb{P}_{00}\right| + \cdots + \left|\log\mathbb{P}_{11}\right| &\leq 1 + \|\theta\|\left(\left\|\frac{\frac{\partial}{\partial\theta}\mathbb{P}_{00}(\bar\theta)}{P_{00}(\bar\theta)}\right\| + \cdots\left\|\frac{\frac{\partial}{\partial\theta}\mathbb{P}_{11}(\bar\theta)}{P_{11}(\bar\theta)}\right\|\right) \\ &\leq 1 + c\|\theta\|\left(\|m'_{i,t}\|(1+|m_{i,t}|) + \frac{\|r'\|}{\sqrt{1-r^2}}\left(1+m_{i,t}^2\right)\right)\end{aligned}$$

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)\right|\right] \\
&\leq \mathbb{E}\left[\sup_{\theta\in\Theta}\left|\log \mathbb{P}_{00}(\theta)\right| + \cdots + \left|\log \mathbb{P}_{11}(\theta)\right|\right] \\
&\leq 1 + \mathbb{E}\left[\sup_{\theta\in\Theta} c\|\theta\| \left(\|m'_{i,t}(\theta)\|(1+|m_{i,t}(\theta)|) + \frac{\|r'(\theta)\|}{\sqrt{1-r^2(\theta)}}\left(1+m^2_{i,t}(\theta)\right)\right)\right] \\
&\leq 1 + C\mathbb{E}\left(\sup_{\theta\in\Theta}\|m'_{i,t}(\theta)\|\left(1+|m_{i,t}(\theta)|\right)\right) + C\mathbb{E}\left(\sup_{\theta\in\Theta} m^2_{i,t}(\theta)\right) \\
&\leq C\left\{\mathbb{E}\left[\sup_{\theta\in\Theta}\left(\sum_{k=0}^{t}\left(\frac{\rho^2}{1-\rho^2}+k\right)|\rho|^{k-1}\|\beta\|\|x_{i,t-k}\| + \sum_{k=0}^{t}|\rho|^k\|x_{i,t-k}\|\right)\left(1+\sum_{k=0}^{t}|\rho|^k\|\beta\|\|x_{i,t-k}\|\right)\right]\right. \\
&\qquad \left. + \mathbb{E}\left[\sup_{\theta\in\Theta}\sum_{k=0}^{t}\sum_{l=0}^{t}|\rho|^{k+l}\|\beta\|^2\|x_{i,t-k}\|\|x_{i,t-l}\|\right]\right\} \\
&\leq \bar{C}\left\{\mathbb{E}\left[\left(\sum_{k=0}^{t}(1+k)\bar{\rho}^k\|x_{i,t-k}\|\right)\left(1+\sum_{k=0}^{t}\bar{\rho}^k\|x_{i,t-k}\|\right)\right] + \mathbb{E}\left[\sum_{k=0}^{t}\sum_{l=0}^{t}\bar{\rho}^{k+l}\|x_{i,t-k}\|\|x_{i,t-l}\|\right]\right\} \\
&\leq \bar{C}\left\{\sum_{k=0}^{t}(1+k)\bar{\rho}^k\mathbb{E}\|x_{i,t-k}\| + \sum_{k=0}^{t}\sum_{l=0}^{t}(1+k)\bar{\rho}^{(k+l-1)}\mathbb{E}\|x_{i,t-k}\|\|x_{i,t-l}\|\right\},
\end{aligned}
$$

where $1 > \bar{\rho}$ is an upperbound for ρ. The details for each upperbound regarding $|m_{i,t}|$, $\|m'_{i,t}\|$ and m^2_{it} can be found in B.3. Thus, it follows that $\mathbb{E}\left[\sup_{\theta\in\Theta}|\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)|\right] < \infty$ provided that $\mathbb{E}(\mathbf{x_i}\mathbf{x_i}')$ is nonsingular, which implies $\mathbb{E}\|\mathbf{x_i}\|^2 < \infty$. □

Lemma 2. *$\theta \neq \theta_0$ implies $f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta) \neq f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)$.*

Proof of Lemma 2. I already showed that $\mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)] \geq \mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)]$ for all $\theta \in \Theta$. This lemma shows that θ_0 is the unique maximizer. It is easier to work with univariate case as opposed to bivariate probabilities. Remember that

$$
\mathbb{P}(y_{it}=0|\mathbf{x_i};\theta) = \mathbf{\Phi}\left(-\sqrt{1-\rho^2}\sum_{k=0}^{t}\rho^k\beta' x_{i,t-k}\right) = \mathbf{\Phi}\left(\sum_{k=0}^{t}\rho^k\widetilde{\beta}' x_{i,t-k}\right),
$$

where $\widetilde{\beta} = -\sqrt{1-\rho^2}\beta$. This transformation is one-to-one. Thus, identifying $(\rho, \widetilde{\beta})$ would be the same as identifying (ρ, β). Let's denote $\mathbf{x}_{i,0:t} = (x'_{i0}, \ldots, x'_{it})'$ and $r_t(\theta) = (\rho^t\beta', \ldots, \rho\beta', \beta')'$, which are both $((t+1)K \times 1)$ dimensional vectors. Note that $B'_t(\theta)\mathbf{x}_{i,0:t} = \sum_{k=0}^{t}\rho^k\beta' x_{i,t-k}$.

$$
\begin{aligned}
\theta \neq \theta_o \quad &\implies \quad B_t(\theta) \neq B_t(\theta_0) \\
&\implies \quad B_t'(\theta)\mathbf{x}_{i,0:t} \neq B_t'(\theta_0)\mathbf{x}_{i,0:t} \quad \text{with positive probability if} \quad \mathbb{E}[\mathbf{x}_{i,0:t}\mathbf{x}_{i,0:t}'] > 0 \\
&\implies \quad \mathbf{\Phi}\left(B_t'(\theta)\mathbf{x}_{i,0:t}\right) \neq \mathbf{\Phi}\left(B_t'(\theta_0)\mathbf{x}_{i,0:t}\right) \quad \text{with positive probability} \\
&\implies \quad \mathbb{P}(y_{it} = 0|\mathbf{x_i};\theta) \neq \mathbb{P}(y_{it} = 0|\mathbf{x_i};\theta_0) \quad \text{with positive probability} \\
&\implies \quad f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta) \neq f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0) \quad \text{with positive probability}
\end{aligned}
$$

We need the positive definiteness of $\mathbb{E}[\mathbf{x}_{i,0:t}\mathbf{x}_{i,0:t}']$ for each t. Hence, the required condition for identification is $\mathbb{E}[\mathbf{x_i}\mathbf{x_i}']> 0$. If the parameters are identified in the univariate model, then they will also be identified in the bivariate model since using more information will only help the identification in estimation.

An alternative way for identification in the bivariate case would be using an approach that is commonly used in *average derivative estimation* techniques (see, for instance, Stoker [1986]). Assume that there is at least one continuous regressor, and let the first regressor, denoted by $x_{it}^{(1)}$, be one of the continuous regressors.

$$
\begin{aligned}
\frac{\partial \mathbb{E}\left[\mathbb{1}\left(y_{it} = 0, y_{i,t+j} = 0\right)|\mathbf{x_i};\theta\right]}{\partial x_{it}^{(1)}} &= \frac{\partial \mathbf{\Phi}_2\left(\sum_{k=0}^{t} \rho^k \beta' x_{i,t-k}, \sum_{k=0}^{t+j} \rho^k \beta' x_{i,t+j-k} \,\middle|\, r = \rho^j\right)}{\partial x_{it}^{(1)}} \\
&= \beta_1 \mathbf{\Phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) + \rho^j \beta_1 \mathbf{\Phi}\left(\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\frac{\partial \mathbb{E}\left[\mathbb{1}\left(y_{it} = 0, y_{i,t+j} = 0\right)|\mathbf{x_i};\theta\right]}{\partial x_{i,t-1}^{(1)}} &= \rho\beta_1 \mathbf{\Phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) + \rho^{j+1}\beta_1 \mathbf{\Phi}\left(\frac{m_t - r m_{t+j}}{\sqrt{1-r^2}}\right) \\
&= \rho\left[\beta_1 \mathbf{\Phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) + \rho^j \beta_1 \mathbf{\Phi}\left(\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right)\right] \\
&= \rho \frac{\partial \mathbb{E}\left[\mathbb{1}\left(y_{it} = 0, y_{i,t+j} = 0\right)|\mathbf{x_i};\theta\right]}{\partial x_{it}^{(1)}}
\end{aligned}
$$

Thus, taking their ratio would isolate ρ. After identifying ρ, we can proceed to identify β in a similar way as in the univarite case.

$$
\begin{aligned}
&\beta \neq \beta_o \\
&\Longrightarrow B_t(\rho,\beta) \neq B_t(\rho,\beta_0) \quad \text{for all } t \\
&\Longrightarrow B_t'(\rho,\beta)\mathbf{x}_{i,0:t} \neq B_t'(\rho,\beta_0)\mathbf{x}_{i,0:t} \quad \text{for all } t \text{ with positive probability if } \mathbb{E}[\mathbf{x}_{i,0:t}\mathbf{x}_{i,0:t}'] > 0 \\
&\Longrightarrow \mathbf{\Phi}_2\left(B_t'(\rho,\beta)\mathbf{x}_{i,0:t}, B_{t+j}'(\rho,\beta)x_{i,0:t+j}|\rho^j\right) \neq \mathbf{\Phi}_2\left(B_t'(\rho,\beta_0)\mathbf{x}_{i,0:t}, B_{t+j}'(\rho,\beta_0)x_{i,0:t+j}|\rho^j\right) \quad \text{with pos. prob.} \\
&\Longrightarrow \mathbb{P}(y_{it}=0|\mathbf{x_i};\rho,\beta) \neq \mathbb{P}(y_{it}=0|\mathbf{x_i};\rho,\beta_0) \quad \text{with positive probability} \\
&\Longrightarrow f(y_{it}, y_{i,t+j}|\mathbf{x_i};\rho,\beta) \neq f(y_{it}, y_{i,t+j}|\mathbf{x_i};\rho,\beta_0) \quad \text{with positive probability}
\end{aligned}
$$

The third step needs some explanation since for constants $(a, b, \bar{a}, \bar{b})$ such that $a > \bar{a}$ and $b < \bar{b}$ one can still have $\mathbf{\Phi}_2(a, b|r) = \mathbf{\Phi}_2(\bar{a}, \bar{b}|r)$. But, this is not the case with probability one as long as x_{it} is a well-behaving random variable with a full support. Consider a simple case where $\beta \neq \beta_o$ but $\mathbf{\Phi}_2(\beta' x_{io},\ \beta' x_{i1} + \rho\beta' x_{io}|\rho) = \mathbf{\Phi}_2(\beta_o' x_{io},\ \beta_o' x_{i1} + \rho\beta_o' x_{io}|\rho)$. The equality holds if $\beta' x_{io} > \beta_o' x_{io}$ and $\beta' x_{i1} + \rho\beta' x_{io} < \beta_o' x_{i1} + \rho\beta_o' x_{io}$ (note that the inequalities between arguments can flip without loss of generality.) Moreover, for given β, β_o, and x_{io}; x_{i1} is determined uniquely since $\mathbf{\Phi}_2$ is strictly increasing in its both arguments. Hence, any other x_{i1} that can be picked with a positive probability will render $\beta = \beta_o$, thus, will provide the identification. The assumption $\mathbb{E}[\mathbf{x_i}\mathbf{x_i}'] > 0$ ensures that we can pick another x_{i1} with a positive probability. Since $\mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)] \geq \mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)]$ is already shown, we can conclude that $\mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)] > \mathbb{E}[\log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)]$ for all $\theta \in \Theta$. Hence, $\mathbb{E}[\ell_i(\theta_0)] > \mathbb{E}[\ell_i(\theta)]$ for all $\theta \in \Theta$. □

Lemma 1 and Lemma 2 constitute the consistency proof for $\hat{\theta}$.

A.2 The Asymptotic Normality Proof

In this section, I analyze the asymptotic distribution of $\hat{\theta}$. In particular, I will prove that

$$\sqrt{N}(\hat{\theta} - \theta_0) \longrightarrow_d \mathcal{N}\left(0, H(\theta_0)^{-1}G(\theta_0)H(\theta_0)^{-1}\right),$$

where

$$H(\theta) = \mathbb{E}\left[\frac{\partial^2 \ell_i(\theta)}{\partial\theta\partial\theta'}\right] = \mathbb{E}\left[\sum_{t=1}^{T-J}\sum_{j=1}^{J}\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'}\right]$$

$$G(\theta) = \mathbb{E}\left[\frac{\partial \ell_i(\theta)}{\partial\theta}\frac{\partial \ell_i(\theta)}{\partial\theta'}\right] = \mathbb{E}\left[\left(\sum_{t=1}^{T-J}\sum_{j=1}^{J}\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}\right)\left(\sum_{t=1}^{T-J}\sum_{j=1}^{J}\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}\right)'\right]$$

Note that since composite likelihood does not use the full information as the full likelihood does, we have $G(\theta_0) \neq H(\theta_0)$ in this case. More details will be give in the following subsections. As typical in asymptotic normality proofs, I utilize the Mean Value Expansion of the composite likelihood evaluated at the CLE around the true parameter. That is, for a mean value $\widetilde{\theta}$ that lies between $\hat{\theta}$ and θ_0, we have

$$0 = \frac{\partial \mathcal{L}_c(\hat{\theta})}{\partial \theta} = \frac{\partial \mathcal{L}_c(\theta_0)}{\partial \theta} + \frac{\partial^2 \mathcal{L}_c(\widetilde{\theta})}{\partial \theta \partial \theta'}(\hat{\theta} - \theta_0) = \frac{1}{N}\sum_{i=1}^{N} s_i(\theta_0|\mathbf{y_i}, \mathbf{x_i}) + \left[\frac{1}{N}\sum_{i=1}^{N} h_i(\widetilde{\theta}|\mathbf{y_i}, \mathbf{x_i})\right](\hat{\theta} - \theta_0).$$

Arranging the terms, using the uniform convergence property of the hessian and the asymptotic normality of the score function, we obtain the desired result.

$$\sqrt{N}(\hat{\theta} - \theta_0) = \left[\frac{1}{N}\sum_{i=1}^{N} h_i(\widetilde{\theta})\right]^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} s_i(\theta_0) \longrightarrow_d \mathcal{N}\left(0, \mathbb{E}[h_i(\theta_0)]^{-1}\mathbb{E}[s_i(\theta_0)s_i(\theta_0)']\mathbb{E}[h_i(\theta_0)]^{-1}\right)$$

The details are provided in the following subsections.

A.2.1 The Score

The score of the individual composite likelihood is

$$s_i(\theta|\mathbf{y_i}, \mathbf{x_i}) = \frac{\partial \ell_i(\theta|\mathbf{y_i}, \mathbf{x_i})}{\partial \theta} = \sum_{t=1}^{T-J}\sum_{j=1}^{J} \frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i}; \theta)}{\partial \theta} = \sum_{t=1}^{T-J}\sum_{j=1}^{J}\left[\frac{\mathbb{1}_{00}}{\mathbb{P}_{00}}\frac{\partial \mathbb{P}_{00}}{\partial \theta} + \cdots + \frac{\mathbb{1}_{00}}{\mathbb{P}_{00}}\frac{\partial \mathbb{P}_{11}}{\partial \theta}\right]$$

Rearranging the terms in the parenthesis by using the following probabilities

$$\begin{aligned}
\mathbb{P}'_{00} &= \frac{\partial}{\partial \theta}\mathbf{\Phi}_2\left(m_{i,t}(\theta), m_{i,t+j}(\theta)|r(\theta)\ \right)\\
\mathbb{P}'_{10} &= m'_{i,t+j}\boldsymbol{\phi}\left(m_{i,t+j}\right) - \mathbb{P}'_{00}\\
\mathbb{P}'_{01} &= m'_{i,t}\boldsymbol{\phi}\left(m_{i,t}\right) - \mathbb{P}'_{00}\\
\mathbb{P}'_{11} &= -m'_{i,t}\boldsymbol{\phi}\left(m_{i,t}\right) - m'_{i,t+j}\boldsymbol{\phi}\left(m_{i,t+j}\right) + \mathbb{P}'_{00},
\end{aligned}$$

we can write the derivative of the individual likelihood as

$$\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}$$
$$= \mathbb{P}'_{00}\left(\frac{\mathbb{1}_{00}}{\mathbb{P}_{00}} - \frac{\mathbb{1}_{10}}{\mathbb{P}_{10}} - \frac{\mathbb{1}_{01}}{\mathbb{P}_{01}} + \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right) + m'_{i,t}\boldsymbol{\phi}\left(m_{i,t}\right)\left(\frac{\mathbb{1}_{01}}{\mathbb{P}_{01}} - \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right) + m'_{i,t+j}\boldsymbol{\phi}\left(m_{i,t+j}\right)\left(\frac{\mathbb{1}_{10}}{\mathbb{P}_{10}} - \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right)$$

$$\begin{aligned}
&= m'_{i,t}\boldsymbol{\phi}\left(m_{i,t}\right)\left\{\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)\left[\frac{\mathbb{1}_{00}}{\mathbb{P}_{00}} - \frac{\mathbb{1}_{10}}{\mathbb{P}_{10}}\right] + \boldsymbol{\Phi}\left(-\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)\left[\frac{\mathbb{1}_{01}}{\mathbb{P}_{01}} - \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right]\right\} \\
&+ m'_{i,t+j}\boldsymbol{\phi}\left(m_{i,t+j}\right)\left\{\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)\left[\frac{\mathbb{1}_{00}}{\mathbb{P}_{00}} - \frac{\mathbb{1}_{01}}{\mathbb{P}_{01}}\right] + \boldsymbol{\Phi}\left(-\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)\left[\frac{\mathbb{1}_{10}}{\mathbb{P}_{10}} - \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right]\right\} \\
&+ \frac{r'}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(m_{i,t+j}\right)\left[\frac{\mathbb{1}_{00}}{\mathbb{P}_{00}} - \frac{\mathbb{1}_{10}}{\mathbb{P}_{10}} - \frac{\mathbb{1}_{01}}{\mathbb{P}_{01}} + \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\right]
\end{aligned} \tag{22}$$

Note that since $\mathbb{E}[\mathbb{1}_{kl}|\mathbf{x_i}] = \mathbb{P}_{kl}$ for all $\{k,l\} \in \{0,1\}$, we have $\mathbb{E}[s_i(\theta_0)|\mathbf{x_i}] = 0$, thus $\mathbb{E}[s_i(\theta_0)] = 0$. Moreover, $s_i(\theta)$ is iid over i by Assumptions 2 and 3. Remember that the probabilities in the score vector contain $\varepsilon_\mathbf{i}$ and $\mathbf{x_i}$ only, which are assumed to be iid over i. For instance,

$$P_{00} = \mathbb{P}\left(\rho^t\varepsilon_{io} + \sqrt{1-\rho^2}\sum_{k=0}^{t-1}\rho^k\varepsilon_{i,t-k} \leq m_{i,t}(\mathbf{x_i},\theta);\quad \rho^{t+j}\varepsilon_{io} + \sqrt{1-\rho^2}\sum_{k=0}^{t+j-1}\rho^k\varepsilon_{i,t+j-k} \leq m_{i,t+j}(\mathbf{x_i},\theta)\,\middle|\,\mathbf{x_i};\theta\right)$$

Hence, since s_i is iid with a finite variance $G(\theta_0)$, we can use Lindeberg-Lévy central limit theorem to obtain $\frac{1}{\sqrt{N}}\sum_{i=1}^N s_i(\theta_0) \longrightarrow_d \mathcal{N}\left(0, G(\theta_0)\right)$, where

$$G(\theta_0) = \mathbb{E}[s_i(\theta_0)s_i(\theta_0)'] = \mathbb{E}\left[\left(\sum_{t=1}^{T-J}\sum_{j=1}^{J}\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}\right)\left(\sum_{t=1}^{T-J}\sum_{j=1}^{J}\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}\right)'\right]$$

The variance $G(\theta_0)$ is finite if $\sum_{t=1}^{T-J}\sum_{j=1}^{J}\mathbb{E}\left[\frac{\partial \log f(y_{it},y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta}\frac{\partial \log f(y_{it},y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta'}\right]$ is finite – due to Cauchy-Schwarz inequality. With the help of the finite fourth order moment assumption, the finiteness of the expected cross-product is shown in the next session where the hessian is analyzed.

It is always informative to compare AR-Probit terms with the static probit case. In particular, we can set $r = \rho^j = 0$ and see that the score of AR-Probit in this case coincides with that of static Probit case. When $\rho = 0$, we have $m_{i,t} = -\beta' x_{it}$, $m'_{i,t} = -x_{it}$, $\mathbb{P}(y_{i,t+j} = 0|\mathbf{x_i}) = \boldsymbol{\Phi}(m_{i,t+j}) = \mathbb{P}_{t+j,0}$, $\mathbb{P}_{00} = \mathbb{P}_{t,0}\mathbb{P}_{t+j,0}$, and $\mathbb{1}_{00} = \mathbb{1}_{t,0}\mathbb{1}_{t+j,0}$, etc. Hence, after putting $\rho = 0$, the first line of (22) becomes

$$
\begin{aligned}
&- x_{it}\boldsymbol{\phi}(\beta' x_{it})\left\{P_{t+j,0}\left[\frac{\mathbb{1}_{t,0}\mathbb{1}_{t+j,0}}{\mathbb{P}_{t,0}\mathbb{P}_{t+j,0}} - \frac{\mathbb{1}_{t,1}\mathbb{1}_{t+j,0}}{\mathbb{P}_{t,1}\mathbb{P}_{t+j,0}}\right] + P_{t+j,1}\left[\frac{\mathbb{1}_{t,0}\mathbb{1}_{t+j,1}}{\mathbb{P}_{t,0}\mathbb{P}_{t+j,1}} - \frac{\mathbb{1}_{t,1}\mathbb{1}_{t+j,1}}{\mathbb{P}_{t,1}\mathbb{P}_{t+j,1}}\right]\right\} \\
&= -x_{it}\boldsymbol{\phi}(\beta' x_{it})\left[\frac{\mathbb{1}_{t,0}}{\mathbb{P}_{t,0}} - \frac{\mathbb{1}_{t,1}}{\mathbb{P}_{t,1}}\right] \\
&= -x_{it}\boldsymbol{\phi}(\beta' x_{it})\left[\frac{1-y_{it}}{\mathbb{P}_{t,0}} - \frac{y_{it}}{\mathbb{P}_{t,1}}\right] \\
&= x_{it}\boldsymbol{\phi}(\beta' x_{it})\frac{y_{it} - \boldsymbol{\Phi}(\beta' x_{it})}{\boldsymbol{\Phi}(\beta' x_{it})\boldsymbol{\Phi}(-\beta' x_{it})}
\end{aligned}
$$

which is exactly the score function of the Static Probit model.

A.2.2 The Hessian

In this subsection, I compute the hessian of the composite likelihood function and show that it is uniformly bounded. The hessian is found to be

$$h(\theta|\mathbf{y_i}, \mathbf{x_i}) = \frac{\partial^2 \ell_i(\theta|\mathbf{y_i}, \mathbf{x_i})}{\partial\theta\partial\theta'} = \sum_{t=1}^{T-J}\sum_{j=1}^{J} \frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'}$$

where

$$\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'} = \frac{\mathbb{1}_{00}}{\mathbb{P}_{00}}\left(\frac{\partial^2\mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}}\frac{\partial\mathbb{P}_{00}}{\partial\theta}\frac{\partial\mathbb{P}_{00}}{\partial\theta'}\right) + \cdots + \frac{\mathbb{1}_{11}}{\mathbb{P}_{11}}\left(\frac{\partial^2\mathbb{P}_{11}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{11}}\frac{\partial\mathbb{P}_{11}}{\partial\theta}\frac{\partial\mathbb{P}_{11}}{\partial\theta'}\right)$$

An upperbound for the individual hessian $h(\theta|\mathbf{y_i}, \mathbf{x_i})$ will depend on an upperbound for the second derivative of log-likelihood.

$$\left\|\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'}\right\| \leq \left\|\frac{1}{\mathbb{P}_{00}}\frac{\partial^2\mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}^2}\frac{\partial\mathbb{P}_{00}}{\partial\theta}\frac{\partial\mathbb{P}_{00}}{\partial\theta'}\right\| + \cdots + \left\|\frac{1}{\mathbb{P}_{11}}\frac{\partial^2\mathbb{P}_{11}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{11}^2}\frac{\partial\mathbb{P}_{11}}{\partial\theta}\frac{\partial\mathbb{P}_{11}}{\partial\theta'}\right\|$$

The norms are analyzed in the section B.2. For instance, the first norm – as well as other three norms – are bounded by

$$
\begin{aligned}
&\left\|\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'}\right\| && (23)\\
&\leq c(1+|m_{i,t}|)\left(\left\|\frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'}\right\| + \left\|\frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'}\right\| + \left\|\frac{\partial^2 r}{\partial\theta\partial\theta'}\right\|\right) && (24)\\
&+ c(1+m_{i,t}^2)\left(\left\|\frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\right\| + \left\|\frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}\right\| + \left\|\frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}\right\| + \left\|\frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\right\|\right) && (25)\\
&+ c(1+|m_{i,t}|^3)\left(\left\|\frac{\partial m_{i,t}}{\partial\theta}\frac{\partial r}{\partial\theta'}\right\| + \left\|\frac{\partial r}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\right\| + \left\|\frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial r}{\partial\theta'}\right\| + \left\|\frac{\partial r}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}\right\|\right) && (26)\\
&+ c(1+m_{i,t}^4)\left\|\frac{\partial r}{\partial\theta}\frac{\partial r}{\partial\theta'}\right\| && (27)
\end{aligned}
$$

Note that, the norms in (25) contain norms of second order moments of $\mathbf{x_i}$; the norms in (26) contain norms of first order moments of $\mathbf{x_i}$; and $\left\|\frac{\partial r}{\partial \theta}\frac{\partial r}{\partial \theta'}\right\|$ in (27) is finite. Also note that, except the line (24), every other line contains norms of the fourth order moment of $\mathbf{x_i}$. Thus, if $\mathbb{E}\|\mathbf{x_i}\|^4 < \infty$, then

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|h(\theta|\mathbf{y_i},\mathbf{x_i})\|\right] \leq \sum_{t=1}^{T-J}\sum_{j=1}^{J}\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta)}{\partial\theta\partial\theta'}\right\|\right] < \infty$$

This condition gives the uniform convergence of $\frac{1}{N}\sum_{i=1}^{N} h(\widetilde{\theta}|\mathbf{y_i},\mathbf{x_i})$ for any consistent estimator $\widetilde{\theta}$. Hence,

$$\frac{1}{N}\sum_{i=1}^{N} h(\widetilde{\theta}|\mathbf{y_i},\mathbf{x_i}) \longrightarrow_p H(\theta_0)$$

$$H(\theta_0) = \mathbb{E}\left[h(\theta_0|\mathbf{y_i},\mathbf{x_i})\right] = \sum_{t=1}^{T-J}\sum_{j=1}^{J}\mathbb{E}\left[\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)}{\partial\theta\partial\theta'}\right]$$

We need $H(\theta_0)$ to be nonsingular. It is usually hard to prove negative definiteness of the hessian matrix in non-linear models. However, with composite likelihood we can utilize its nice features that it borrows from the full likelihood. In particular, note that even thought the Bartlett equality does not hold for the composite likelihood, in general, it still holds for each piece of the composite likelihood. That is, $\mathbb{E}[\frac{\partial\ell_i(\theta_o)}{\partial\theta}\frac{\partial\ell_i(\theta_o)}{\partial\theta'}] \neq -\mathbb{E}[\frac{\partial^2\ell_i(\theta_o)}{\partial\theta\partial\theta'}]$. However,

$$\mathbb{E}\left[\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)}{\partial\theta}\frac{\partial \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)}{\partial\theta'}\right] = -\mathbb{E}\left[\frac{\partial^2 \log f(y_{it}, y_{i,t+j}|\mathbf{x_i};\theta_0)}{\partial\theta\partial\theta'}\right] < 0$$

Hence, $H(\theta_0)$ is invertible. Therefore, we can conclude that, for any consistent estimator $\widetilde{\theta}$,

$$\left[\frac{1}{N}\sum_{i=1}^{N} h(\widetilde{\theta}|\mathbf{y_i},\mathbf{x_i})\right]^{-1} \longrightarrow_p H(\theta_0)^{-1}$$

B Mathematical Details

This section analyzes mathematical properties of functions of normal density and normal cumulative distribution, especially the ones that are need throughout the analysis in this paper.

B.1 Derivatives of a Bivariate Normal Distribution

The derivative of bivariate normal distribution with respect to the mean and variance parameters are analyzed in this subsection. To facilitate the algebra, I use the change of variables in the following way: $\Sigma^{-1/2}[x,y]' = [z_1, z_2]'$ where $\Sigma = \begin{bmatrix} 1 & r(\theta) \\ r(\theta) & 1 \end{bmatrix}$. Hence, the bivariate normal distribution can be written as

$$\mathbf{\Phi}_2\left(m_t(\theta), m_{t+j}(\theta) \mid r(\theta)\right) = \int_{-\infty}^{m_t(\theta)} \int_{-\infty}^{m_{t+j}(\theta)} \frac{1}{2\pi}|\Sigma|^{-0.5} \exp\left\{-\frac{1}{2}[x,y]\Sigma^{-1}[x,y]'\right\} dydx$$

$$= \int_{-\infty}^{m_t(\theta)} \int_{-\infty}^{\frac{-r(\theta)z_1 + m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(z_1^2 + z_2^2)\right\} dz_2 dz_1$$

$$= \int_{-\infty}^{m_t(\theta)} \int_{-\infty}^{\frac{-r(\theta)z_1 + m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}} \boldsymbol{\phi}(z_1)\boldsymbol{\phi}(z_2) dz_2 dz_1$$

Now, it is easier to take the derivative by using the second fundamental theorem of calculus.

$$\frac{\partial \mathbf{\Phi}_2\left(m_t(\theta), m_{t+j}(\theta) \mid r(\theta)\right)}{\partial \theta}$$
$$= m_t'(\theta)\frac{\partial \mathbf{\Phi}_2\left(m_t, m_{t+j} \mid r\right)}{\partial m_t} + m_{t+j}'(\theta)\frac{\partial \mathbf{\Phi}_2\left(m_t, m_{t+j} \mid r\right)}{\partial m_{t+j}} + r'(\theta)\frac{\partial \mathbf{\Phi}_2\left(m_t, m_{t+j} \mid r\right)}{\partial r}$$
$$= m_t'(\theta)\boldsymbol{\phi}(m_t(\theta))\mathbf{\Phi}\left(\frac{-r(\theta)m_t(\theta) + m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) + m_{t+j}'(\theta)\boldsymbol{\phi}(m_{t+j}(\theta))\mathbf{\Phi}\left(\frac{m_t(\theta) - r(\theta)m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right)$$
$$+ r'(\theta)\int_{-\infty}^{m_t(\theta)} \frac{\partial\left(\frac{-rz_1 + m_{t+j}(\theta)}{\sqrt{1-r^2}}\right)}{\partial r}\boldsymbol{\phi}(z_1)\boldsymbol{\phi}\left(\frac{-r(\theta)z_1 + m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) dz_1$$
$$= m_t'(\theta)\boldsymbol{\phi}(m_t(\theta))\mathbf{\Phi}\left(\frac{-r(\theta)m_t(\theta) + m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) + m_{t+j}'(\theta)\boldsymbol{\phi}(m_{t+j}(\theta))\mathbf{\Phi}\left(\frac{m_t(\theta) - r(\theta)m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right)$$
$$+ \frac{r'(\theta)}{\sqrt{1-r(\theta)^2}}\boldsymbol{\phi}(m_{t+j}(\theta))\int_{-\infty}^{\frac{m_t(\theta) - r(\theta)m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}} z\boldsymbol{\phi}(z)dz$$

The last integral is similar to the expectation of a truncated normal variable up to a constant. In particular, density and the expectation of a truncated standard normal variable with the truncation interval (a, b) is

$$f_{TN}(z) = \frac{\boldsymbol{\phi}(z)}{\mathbf{\Phi}(b) - \mathbf{\Phi}(a)} \quad \text{and} \quad \mathbb{E}[z] = \int_a^b z f_{TN}(z)dz = -\frac{\boldsymbol{\phi}(b) - \boldsymbol{\phi}(a)}{\mathbf{\Phi}(b) - \mathbf{\Phi}(a)}$$

In the case above, $a = -\infty$ and $b = \frac{m_t(\theta) - r(\theta)m_l(\theta)}{\sqrt{1-r(\theta)^2}}$. Moreover, I need to divide and multiply $\boldsymbol{\phi}(z)$ by $\Phi(b)$ to transform it into truncated standard normal density. Hence, the derivative of the bivariate normal distribution can be written as

$$\begin{aligned}\frac{\partial \boldsymbol{\Phi}_2\left(m_t(\theta), m_{t+j}(\theta) \mid r(\theta)\right)}{\partial \theta} &= m_t'(\theta)\boldsymbol{\phi}(m_t(\theta))\boldsymbol{\Phi}\left(\frac{-r(\theta)m_t(\theta)+m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) \\ &+ m_{t+j}'(\theta)\boldsymbol{\phi}(m_{t+j}(\theta))\boldsymbol{\Phi}\left(\frac{m_t(\theta)-r(\theta)m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) \\ &+ \frac{r'(\theta)}{\sqrt{1-r(\theta)^2}}\boldsymbol{\phi}(m_t(\theta))\boldsymbol{\phi}\left(\frac{-r(\theta)m_t(\theta)+m_{t+j}(\theta)}{\sqrt{1-r(\theta)^2}}\right) \qquad (28)\end{aligned}$$

Note that $\mathbb{P}_{00} \equiv \boldsymbol{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)$. Hence, the cross-product of the first derivative is, after suppressing the notation for θ,

$$\begin{aligned}&\frac{\partial P_{00}}{\partial \theta}\frac{\partial P_{00}}{\partial \theta'} = \\ &= \frac{\partial m_{i,t}}{\partial \theta}\frac{\partial m_{i,t}}{\partial \theta'}\boldsymbol{\phi}\left(m_{i,t}\right)^2\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)^2 \\ &+ \frac{\partial m_{i,t+j}}{\partial \theta}\frac{\partial m_{i,t+j}}{\partial \theta'}\boldsymbol{\phi}\left(m_{i,t+j}\right)^2\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)^2 \\ &+ \frac{\partial r}{\partial \theta}\frac{\partial r}{\partial \theta'}\frac{1}{1-r^2}\boldsymbol{\phi}\left(m_{i,t}\right)^2\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)^2 \\ &+ \left(\frac{\partial m_{i,t}}{\partial \theta}\frac{\partial m_{i,t+j}}{\partial \theta'}+\frac{\partial m_{i,t+j}}{\partial \theta}\frac{\partial m_{i,t}}{\partial \theta'}\right)\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right) \\ &+ \left(\frac{\partial m_{i,t}}{\partial \theta}\frac{\partial r}{\partial \theta'}+\frac{\partial r}{\partial \theta}\frac{\partial m_{i,t}}{\partial \theta'}\right)\frac{1}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_{i,t}\right)^2\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right) \\ &+ \left(\frac{\partial m_{i,t+j}}{\partial \theta}\frac{\partial r}{\partial \theta'}+\frac{\partial r}{\partial \theta}\frac{\partial m_{i,t+j}}{\partial \theta'}\right)\frac{1}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_{i,t+j}\right)^2\boldsymbol{\phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)\end{aligned}$$

The second derivative would be

$$
\begin{aligned}
\frac{\partial^2 P_{00}}{\partial\theta\partial\theta'} = & \frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'}\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t + m_{t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}(-m_{i,t})\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial}{\partial\theta'}\left[\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right]\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'}\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}(-m_{i,t+j})\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial}{\partial\theta'}\left[\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right]\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\phi}\left(\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial}{\partial\theta'}\left[\frac{r'}{\sqrt{1-r^2}}\right]\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial r}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\frac{1}{\sqrt{1-r^2}}(-m_{i,t})\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
& + \frac{\partial r}{\partial\theta}\frac{\partial}{\partial\theta'}\left[\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right]\frac{1}{\sqrt{1-r^2}}\left(-\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right)\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right)
\end{aligned}
$$

The following terms will be replaced in the above equation.

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\left[\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right] &= -\frac{r}{\sqrt{1-r^2}}\frac{\partial m_{i,t}}{\partial\theta} + \frac{1}{\sqrt{1-r^2}}\frac{\partial m_{i,t+j}}{\partial\theta} - \frac{\frac{\partial r}{\partial\theta}}{\sqrt{1-r^2}}\left(\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right) \\
\frac{\partial}{\partial\theta}\left[\frac{m_{i,t} - rm_{i,t+j}}{\sqrt{1-r^2}}\right] &= \frac{1}{\sqrt{1-r^2}}\frac{\partial m_{i,t}}{\partial\theta} - \frac{r}{\sqrt{1-r^2}}\frac{\partial m_{i,t+j}}{\partial\theta} - \frac{\frac{\partial r}{\partial\theta}}{\sqrt{1-r^2}}\left(\frac{-rm_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
\frac{\partial}{\partial\theta'}\left[\frac{r'}{\sqrt{1-r^2}}\right] &= \frac{\frac{\partial^2 r}{\partial\theta\partial\theta'}}{\sqrt{1-r^2}} + \frac{\frac{\partial r}{\partial\theta}\frac{\partial r}{\partial\theta'}}{1-r^2}\frac{r}{\sqrt{1-r^2}}
\end{aligned}
$$

Hence, the second derivative of P_{00} is found to be

$$
\begin{aligned}
&\frac{\partial^2 \mathbb{P}_{00}}{\partial\theta\partial\theta'} \\
&= \frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'} \boldsymbol{\phi}\left(m_{i,t}\right) \mathbf{\Phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
&+ \frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'} \boldsymbol{\phi}\left(m_{i,t+j}\right) \mathbf{\Phi}\left(\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
&+ \frac{\partial^2 r}{\partial\theta\partial\theta'} \frac{1}{\sqrt{1-r^2}} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
&+ \frac{\partial m_{i,t}}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'} \left[-m_{i,t} \boldsymbol{\phi}\left(m_{i,t}\right) \mathbf{\Phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) - \frac{r}{\sqrt{1-r^2}} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right)\right] \\
&+ \frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'} \left[-m_{i,t+j} \boldsymbol{\phi}\left(m_{i,t+j}\right) \mathbf{\Phi}\left(\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right) - \frac{r}{\sqrt{1-r^2}} \boldsymbol{\phi}\left(m_{i,t+j}\right) \boldsymbol{\phi}\left(\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right)\right] \\
&+ \frac{\partial r}{\partial\theta} \frac{\partial r}{\partial\theta'} \frac{1}{(1-r^2)^{3/2}} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \left[r + \frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}} \frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right] \\
&+ \left(\frac{\partial m_{i,t}}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'} + \frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'}\right) \frac{1}{\sqrt{1-r^2}} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \\
&+ \left(\frac{\partial m_{i,t}}{\partial\theta} \frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'}\right) \frac{1}{1-r^2} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \left[-\frac{m_{i,t} - r m_{i,t+j}}{\sqrt{1-r^2}}\right] \\
&+ \left(\frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'}\right) \frac{1}{1-r^2} \boldsymbol{\phi}\left(m_{i,t}\right) \boldsymbol{\phi}\left(\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right) \left[-\frac{-r m_{i,t} + m_{i,t+j}}{\sqrt{1-r^2}}\right]
\end{aligned}
$$

The following will be useful in computing the hessian.

$$
\frac{1}{\mathbb{P}_{00}}\frac{\partial^2\mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}^2}\frac{\partial\mathbb{P}_{00}}{\partial\theta}\frac{\partial\mathbb{P}_{00}}{\partial\theta'} \tag{29}
$$

$$
= \frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \tag{30}
$$

$$
+ \frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'}\frac{\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \tag{31}
$$

$$
+ \frac{\partial^2 r}{\partial\theta\partial\theta'}\frac{1}{\sqrt{1-r^2}}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \tag{32}
$$

$$
\begin{aligned}
&+ \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \\
&\quad\times\left[-m_{i,t} - \frac{r}{\sqrt{1-r^2}}\frac{\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)} - \frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
&+ \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}\frac{\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \\
&\quad\times\left[-m_{i,t+j} - \frac{r}{\sqrt{1-r^2}}\frac{\boldsymbol{\phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)} - \frac{\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{34}
$$

$$
\begin{aligned}
&+ \frac{\partial r}{\partial\theta}\frac{\partial r}{\partial\theta'}\frac{1}{(1-r^2)^{3/2}}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \\
&\quad\times\left[r + \frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
&+ \left(\frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'} + \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\right)\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t,m_{t+j}|r\right)} \\
&\quad\times\left[\frac{1}{\sqrt{1-r^2}} - \frac{\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}\frac{\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{36}
$$

$$
\begin{aligned}
&+ \left(\frac{\partial m_{i,t}}{\partial\theta}\frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'}\right)\frac{1}{1-r^2}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \\
&\quad\times\left[-\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{37}
$$

$$
\begin{aligned}
&+ \left(\frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'}\right)\frac{1}{1-r^2}\frac{\boldsymbol{\phi}\left(m_{i,t}\right)\boldsymbol{\phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)} \\
&\quad\times\left[-\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\boldsymbol{\phi}\left(m_{i,t+j}\right)\boldsymbol{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{72\,\boldsymbol{\Phi}_2\left(m_{i,t},m_{i,t+j}|r\right)}\right]
\end{aligned} \tag{38}
$$

The limiting behavior of $\left\| \frac{1}{\mathbb{P}_{00}} \frac{\partial^2 \mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}^2} \frac{\partial \mathbb{P}_{00}}{\partial\theta} \frac{\partial \mathbb{P}_{00}}{\partial\theta'} \right\|$ will be analyzed in Section B.2

B.2 Limits on Functions of Univariate and Bivariate Normal Distribution

The ratio $\phi(x)/\Phi(x)$ is known to be bounded by $C(1+|x|)$. Note that the ratio approaches 0 as x approaches positive infinity, and approaches to the negative 45 degree line as x approaches negative infinity. One can show it by taking the limit and using L'Hôpital rule, i.e., $\lim_{x\to-\infty} \phi(x)/\Phi(x) = \lim_{x\to-\infty} -x\phi(x)/\phi(x)$. Hence, the ratio goes to ∞ with a linear rate. A slightly more general case will be needed for the analysis. Let $a > 0$ and $c > 0$, where it is easy to calculate for other cases too.

$$\lim_{x\to-\infty} \frac{\phi(ax+b)}{\Phi(cx+d)} = \lim_{x\to-\infty} \frac{-a(ax+b)\phi(ax+b)}{c\phi(cx+d)} = \begin{cases} 0 & \text{if } \ a > c \\ 0 & \text{if } \ a = c \text{ and } b > d \\ \infty & \text{if } \ a < c \\ \infty & \text{if } \ a = c \text{ and } b < d \\ \lim_{x\to-\infty} -(ax+b) & \text{if } \ (a,b) = (c,d) \end{cases}$$

Depending on the parameters, ratio of two normal densities is also a normal density up to a constant multiplicative term. For constants (a,b,c,d) with $|a| > |c|$, a straightforward calculation yields

$$\frac{\phi(ax+b)}{\phi(cx+d)} = \frac{\phi\left(\sqrt{a^2-c^2}x + \frac{ab-cd}{\sqrt{a^2-c^2}}\right)}{\phi\left(\frac{ad-bc}{\sqrt{a^2-c^2}}\right)}$$

If $|a| < |c|$, then we can consider the reciprocal of the ratio, which will give again a normal density. Next, I will analyze the limiting behavior of two ratios that appear throughout the analysis.

Claim 1. *The following ratio diverges to ∞ at a linear rate. In particular,*

$$\frac{\phi(m_t)\,\Phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi_2(m_t, m_{t+j}|r)} \le c\left(1 + \max\{|m_t|, |m_{t+j}|\}\right) \tag{39}$$

I will prove the claim by looking at different cases on m_t and m_{t+j}.

Proof.

Case 1. *m_t: fixed, $m_{t+j} \to \infty$*
The ratio in (39) is finite since

$$\frac{\phi(m_t)\,\Phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi_2(m_t, m_{t+j}|r)} \longrightarrow \frac{\phi(m_t)}{\Phi(m_t)} < \infty$$

Case 2. *m_t: fixed, $m_{t+j} \to -\infty$*

The ratio in (39) converges to 0, if $r > 0$. Otherwise, (39) diverges to ∞ with a linear rate, that is, (39) is bounded by a multiple of $|m_{t+j}|$. We'll use L'Hôpital rule due to $0/0$ indeterminacy.

$$\overset{\frac{\partial(\cdot)}{\partial m_{t+j}}}{\Longrightarrow} \frac{\frac{1}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\phi}\left(m_{t+j}\right)\boldsymbol{\Phi}\left(\frac{m_t-rm_{t+j}}{\sqrt{1-r^2}}\right)} = c_1 \frac{\boldsymbol{\phi}\left(\frac{r}{\sqrt{1-r^2}}(m_{t+j}+c_2)\right)}{\boldsymbol{\Phi}\left(\frac{m_t-rm_{t+j}}{\sqrt{1-r^2}}\right)} \begin{cases} \longrightarrow \frac{0}{1} = 0 & \text{if } r > 0 \\ \leq c(1+|m_{t+j}|) & \text{if } r < 0 \end{cases} \tag{40}$$

Case 3. *$m_t \to \infty$, m_{t+j}: fixed*

The ratio in (39) converges to 0 since

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} \longrightarrow \frac{0}{\boldsymbol{\Phi}\left(m_{t+j}\right)} = 0$$

Case 4. *$m_t \to -\infty$, m_{t+j}: fixed*

The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow} \frac{-m_t\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right) - \frac{r}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)} = -m_t - \frac{r}{\sqrt{1-r^2}}\frac{\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}$$

$$\overset{A}{\approx} -m_t + \frac{r}{\sqrt{1-r^2}}\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}} = \frac{m_t - rm_{t+j}}{\sqrt{1-r^2}} \leq c\left(1+|m_t|\right)$$

Case 5. *$m_t \to \infty$, $m_{t+j} \to \infty$*

The ratio in (39) converges to 0.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} \longrightarrow \frac{0}{1} = 0$$

Case 6. *$m_t \to \infty$, $m_{t+j} = km_t$ with $k < 0$ and $k < r$*

The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} = \frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}\left(m_t, km_t|r\right)} \longrightarrow \frac{0 \times 0}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow} \frac{-m_t\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+\frac{k-r}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\boldsymbol{\phi}(km_t)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)} = \frac{-m_t\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+\frac{k-r}{\sqrt{1-r^2}}}{\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+k\frac{\boldsymbol{\phi}(km_t)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}(m_t)\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}}$$

$$= \frac{-m_t\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+\frac{k-r}{\sqrt{1-r^2}}}{\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+k\frac{\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}} \overset{A}{\approx} \begin{cases} \frac{-m_t\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}-\frac{k-r}{\sqrt{1-r^2}}}{\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}+k\frac{-\sqrt{1-r^2}}{(1-kr)}\frac{1}{m_t}} \leq c\left(1+|m_t|\right) & \textit{if} \;\; 1-kr<0 \\ \frac{-m_t\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}-\frac{k-r}{\sqrt{1-r^2}}}{\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}+k\frac{\boldsymbol{\Phi}(0)}{\boldsymbol{\phi}(0)}} \longrightarrow c<\infty & \textit{if} \;\; 1-kr=0 \\ \frac{-m_t\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}-\frac{k-r}{\sqrt{1-r^2}}}{\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}+k\exp\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)} \longrightarrow 0 & \textit{if} \;\; 1-kr>0 \end{cases}$$

Case 7. *$m_t \to \infty$, $m_{t+j} = km_t$ with $k = r < 0$.*
The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} = \frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}(0)}{\boldsymbol{\Phi}\left(m_t, rm_t|r\right)} \longrightarrow \frac{0}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow} \frac{-m_t\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}(0)}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}(0)+r\boldsymbol{\phi}(rm_t)\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)} = \frac{-m_t}{\frac{1}{2}+r\frac{\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)}{\boldsymbol{\phi}(\sqrt{1-r^2}m_t)}} \overset{A}{\approx} \frac{-m_t}{\frac{1}{2}-\frac{r}{\sqrt{1-r^2}}\frac{1}{m_t}} \leq c\left(1+|m_t|\right)$$

Case 8. *$m_t \to \infty$, $m_{t+j} = km_t$ with $-1 < r < k < 0$.*
The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} = \frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}\left(m_t, km_t|r\right)} \longrightarrow \frac{0\times 1}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow} \frac{-m_t\boldsymbol{\phi}\left(m_t\right)}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\boldsymbol{\phi}(km_t)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)} = \frac{-m_t}{\frac{1}{2}+r\frac{\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)}{\boldsymbol{\phi}(\sqrt{1-k^2}m_t)}} \overset{A}{\approx} \frac{-m_t}{\frac{1}{2}-r\times 0} \leq c\left(1+|m_t|\right)$$

Note that $\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)/\boldsymbol{\phi}(\sqrt{1-k^2}m_t) \to 0$ exponentially fast since $\sqrt{1-r^2} > \sqrt{1-k^2}$.

Case 9. *$m_t \to -\infty$, $m_{t+j} = km_t$ with $k < r$*

The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2(m_t,m_{t+j}|r)}=\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}(m_t,km_t|r)}\longrightarrow\frac{0\times 1}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow}\frac{-m_t\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\boldsymbol{\phi}(km_t)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}=\frac{-m_t\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\frac{\boldsymbol{\phi}(km_t)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}(m_t)}}$$

$$=\begin{cases}\dfrac{-m_t\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\frac{\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(\sqrt{1-k^2}m_t\right)}}\overset{A}{\approx}\frac{-m_t}{1+k\times 0}\le c\,(1+|m_t|) & \textit{if}\quad -1\le k<r\\[2em] \dfrac{-m_t\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\boldsymbol{\phi}\left(\sqrt{k^2-1}m_t\right)\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}\overset{A}{\approx}\frac{-m_t}{1+k\times 0}\le c\,(1+|m_t|) & \textit{if}\quad k<-1\end{cases}$$

Note that $\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)/\boldsymbol{\phi}\left(\sqrt{1-k^2}m_t\right)\to 0$ *exponentially since* $\frac{1-kr}{\sqrt{1-r^2}}>\sqrt{1-k^2}$.

Case 10. $m_t\to-\infty$, $m_{t+j}=km_t$ *where* $k=r$

The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2(m_t,m_{t+j}|r)}=\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}(0)}{\boldsymbol{\Phi}(m_t,rm_t|r)}\longrightarrow\frac{0\times\frac{1}{2}}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow}\frac{-m_t\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}(0)}{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}(0)+r\boldsymbol{\phi}(rm_t)\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)}=\frac{-m_t}{\frac{1}{2}+r\frac{\boldsymbol{\Phi}(\sqrt{1-r^2}m_t)}{\boldsymbol{\phi}(\sqrt{1-r^2}m_t)}}\overset{A}{\approx}\frac{-m_t}{\frac{1}{2}-\frac{r}{\sqrt{1-r^2}}\frac{1}{m_t}}\le c\,(1+|m_t|)$$

Case 11. $m_t\to-\infty$, $m_{t+j}=km_t$ *where* $r<k$

The ratio in (39) is bounded by a multiple of $|m_t|$.

$$\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2(m_t,m_{t+j}|r)}=\frac{\boldsymbol{\phi}(m_t)\,\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\Phi}(m_t,km_t|r)}\longrightarrow\frac{0\times 0}{0}$$

$$\overset{\frac{\partial(\cdot)}{\partial m_t}}{\Longrightarrow} \frac{-m_t\phi(m_t)\,\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+\frac{k-r}{\sqrt{1-r^2}}\phi(m_t)\,\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi(m_t)\,\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)+k\phi(km_t)\Phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)} = \frac{-m_t\frac{\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+\frac{k-r}{\sqrt{1-r^2}}}{\frac{\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+k\frac{\phi(km_t)\Phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\phi(m_t)\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}}$$

$$= \frac{-m_t\frac{\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+\frac{k-r}{\sqrt{1-r^2}}}{\frac{\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+k\frac{\Phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}} \overset{A}{\approx} \frac{-m_t\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}-\frac{k-r}{\sqrt{1-r^2}}}{\frac{-\sqrt{1-r^2}}{(k-r)}\frac{1}{m_t}+k\frac{-\sqrt{1-r^2}}{(1-kr)}\frac{1}{m_t}} \leq c\,(1+|m_t|)$$

Note that $\frac{\phi(m_t)\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi(km_t)} = \phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)$.

□

Claim 2. *The following ratio diverges to* ∞ *at a quadratic rate. In particular,*

$$\frac{\phi(m_t)\,\phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi_2(m_t,m_{t+j}|r)} \leq c\left(1+\max\{m_t^2,m_{t+j}^2\}\right) \tag{41}$$

I will prove the claim by looking at different cases on m_t *and* m_{t+j}.

Proof.

Case 1. m_t*: fixed,* $m_{t+j}\to\infty$
The ratio in (41) *is finite since*

$$\frac{\phi(m_t)\,\phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi_2(m_t,m_{t+j}|r)} \longrightarrow \frac{0}{\Phi(m_t)}=0$$

Case 2. m_t*: fixed,* $m_{t+j}\to-\infty$
The ratio in (41) *converges to 0, if* $r>0$*. Otherwise,* (41) *diverges to* ∞ *with a quadratic rate, that is,* (41) *is bounded by a multiple of* $|m_{t+j}|^2$*. We'll use the results in* (40).

$$\overset{\frac{\partial(\cdot)}{\partial m_{t+j}}}{\Longrightarrow} -\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\frac{\frac{1}{\sqrt{1-r^2}}\phi(m_t)\,\phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\phi(m_{t+j})\,\Phi\left(\frac{m_t-rm_{t+j}}{\sqrt{1-r^2}}\right)} \begin{cases}\longrightarrow 0 & \text{if } r>0\\ \leq c(1+|m_{t+j}|^2) & \text{if } r<0\end{cases}$$

Case 3. *$m_t \to \infty$, m_{t+j}: fixed*

The ratio in (39) converges to 0 since

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} \longrightarrow \frac{0}{\boldsymbol{\Phi}\left(m_{t+j}\right)} = 0$$

Case 4. *$m_t \to -\infty$, m_{t+j}: fixed*

The ratio in (41) is bounded by a multiple of $|m_t|^2$.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} \longrightarrow \frac{0}{0}$$

$$\xRightarrow{\frac{\partial(\cdot)}{\partial m_t}} \frac{-m_t\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right) + \frac{r}{\sqrt{1-r^2}}\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}$$

$$= \frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)\left[-m_t + \frac{r}{\sqrt{1-r^2}}\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right]}{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)} = -\frac{\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}\frac{m_t - rm_{t+j}}{\sqrt{1-r^2}}$$

$$\overset{A}{\approx} -\frac{r}{\sqrt{1-r^2}}\frac{-rm_t + m_{t+j}}{\sqrt{1-r^2}}\frac{m_t - rm_{t+j}}{\sqrt{1-r^2}} \leq c(1 + |m_{t+j}|^2)$$

Case 5. *$m_t \to \infty$, $m_{t+j} \to \infty$*

The ratio in (41) converges to 0.

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} \longrightarrow \frac{0}{1} = 0$$

For the following cases when $m_{t+j} = km_t$, we will transform the ratio (41) into

$$\frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\phi}\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j}|r\right)} = \frac{\boldsymbol{\phi}\left(m_t\right)\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(m_t, km_t|r\right)} = \frac{\boldsymbol{\phi}\left(\sqrt{1+\frac{(k-r)^2}{1-r^2}}m_t\right)}{\boldsymbol{\phi}\left(m_t, km_t|r\right)}$$

Moreover, the derivative of the numerator and denumerator yields

$$\frac{-\left[1+\frac{(k-r)^2}{1-r^2}\right] m_t \boldsymbol{\phi}\left(\sqrt{1+\frac{(k-r)^2}{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(m_t\right) \boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)+k \boldsymbol{\phi}(k m_t) \boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}=\frac{-\left[1+\frac{(k-r)^2}{1-r^2}\right] m_t}{\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)}+k \frac{\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}} \tag{42}$$

Case 6. *$m_t \rightarrow \infty$, $m_{t+j}=k m_t$ with $k<0$*
The ratio in (41) is bounded by a multiple of $|m_t^2|$.

$$\frac{\boldsymbol{\phi}\left(m_t\right) \boldsymbol{\phi}\left(\frac{-r m_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j} | r\right)} \longrightarrow \frac{0}{0}$$

Note that

$$\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)} \stackrel{A}{\approx}\begin{cases}-\frac{\sqrt{1-r^2}}{k-r} \frac{1}{m_t} \rightarrow 0 \text{ at a linear rate} & \text{if } k<r \\ \frac{\boldsymbol{\phi}(0)}{\boldsymbol{\Phi}(0)}<\infty & \text{if } k=r \\ \exp\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right) \rightarrow \infty \text{ exponentially} & \text{if } k>r\end{cases}$$

$$\frac{\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)} \stackrel{A}{\approx}\begin{cases}-\frac{\sqrt{1-r^2}}{1-kr} \frac{1}{m_t} \rightarrow 0 \text{ at a linear rate} & \text{if } 1<kr \\ \frac{\boldsymbol{\phi}(0)}{\boldsymbol{\Phi}(0)}<\infty & \text{if } 1=kr \\ \exp\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right) \rightarrow \infty \text{ exponentially} & \text{if } 1>kr\end{cases}$$

Hence, we have three possible limits for (42).

$$\frac{-\left[1+\frac{(k-r)^2}{1-r^2}\right] m_t}{\frac{\boldsymbol{\Phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{k-r}{\sqrt{1-r^2}} m_t\right)}+k \frac{\boldsymbol{\Phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}{\boldsymbol{\phi}\left(\frac{1-kr}{\sqrt{1-r^2}} m_t\right)}}\begin{cases}\longrightarrow 0 \text{ exponentially} \\ \stackrel{A}{\approx} c(1+|m_t|) \\ \stackrel{A}{\approx} c(1+|m_t|^2)\end{cases}$$

Case 7. *$m_t \rightarrow -\infty$, $m_{t+j}=k m_t$*
The ratio in (41) is bounded by a multiple of $|m_t^2|$.

$$\frac{\boldsymbol{\phi}\left(m_t\right) \boldsymbol{\phi}\left(\frac{-r m_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\boldsymbol{\Phi}_2\left(m_t, m_{t+j} | r\right)} \longrightarrow \frac{0}{0}$$

By using the same arguments as above, we can conclude that

$$\frac{-\left[1+\frac{(k-r)^2}{1-r^2}\right]m_t}{\frac{\Phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{k-r}{\sqrt{1-r^2}}m_t\right)}+k\frac{\Phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}{\phi\left(\frac{1-kr}{\sqrt{1-r^2}}m_t\right)}}\begin{cases}\longrightarrow 0 \textit{ exponentially}\\ \overset{A}{\approx} c(1+|m_t|)\\ \overset{A}{\approx} c(1+|m_t|^2)\end{cases}$$

□

A similar analysis shows that same limits are obtained for other similar ratios involving bivariate normal distribution. In particular, the following ratios that occur in the bivariate probabilities $\{P_{00}, P_{10}, P_{01}, P_{11}\}$ have the same limiting behavior.

$$\frac{\phi(m_t)\,\Phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi(m_{t+j})-\Phi_2(m_t,m_{t+j}|r)}\leq c\left(1+\max\{|m_t|,|m_{t+j}|\}\right)$$

$$\frac{\phi(m_t)\,\Phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{1-\Phi(m_t)-\Phi(m_{t+j})+\Phi_2(m_t,m_{t+j}|r)}\leq c\left(1+\max\{|m_t|,|m_{t+j}|\}\right)$$

$$\frac{\phi(m_t)\,\phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{\Phi(m_{t+j})-\Phi_2(m_t,m_{t+j}|r)}\leq c\left(1+\max\{m_t^2,m_{t+j}^2\}\right)$$

$$\frac{\phi(m_t)\,\phi\left(\frac{-rm_t+m_{t+j}}{\sqrt{1-r^2}}\right)}{1-\Phi(m_t)-\Phi(m_{t+j})+\Phi_2(m_t,m_{t+j}|r)}\leq c\left(1+\max\{m_t^2,m_{t+j}^2\}\right)$$

For instance, note that $\frac{\partial\Phi_2(m_{i,t},m_{i,t+j}|r)}{\partial m_{i,t}}=-\frac{\partial[\Phi(m_{i,t+j})-\Phi_2(m_{i,t},m_{i,t+j}|r)]}{\partial m_{i,t}}$, and that $\frac{\partial\Phi_2(m_{i,t},m_{i,t+j}|r)}{\partial m_{i,t+j}}=\phi(m_{i,t+j})\,\Phi\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)$ whereas $\frac{\partial[\Phi(m_{i,t+j})-\Phi_2(m_{i,t},m_{i,t+j}|r)]}{\partial m_{i,t+j}}=\phi(m_{i,t+j})\,\Phi\left(-\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)$. Thus, once the L'Hôpital rule is used, there is no difference between $\Phi_2(m_{i,t},m_{i,t+j}|r)$, $\Phi(m_{i,t})-\Phi_2(m_{i,t},m_{i,t+j}|r)$, $\Phi(m_{i,t+j})-\Phi_2(m_{i,t},m_{i,t+j}|r)$, and $1-\Phi(m_{i,t})-\Phi(m_{i,t+j})-\Phi_2(m_{i,t},m_{i,t+j}|r)$ in terms of limiting behavior.

Next, we find an upperbound for $\left\|\frac{1}{\mathbb{P}_{00}}\frac{\partial^2\mathbb{P}_{00}}{\partial\theta\partial\theta'}-\frac{1}{\mathbb{P}_{00}^2}\frac{\partial\mathbb{P}_{00}}{\partial\theta}\frac{\partial\mathbb{P}_{00}}{\partial\theta'}\right\|$ by using the Claims 1 and 2. The terms inside the square brackets in (33), (34), (37), and (38) are bounded by $|m_{i,t}|$ linearly - again, I assume without loss of generality, $|m_{i,t}| \geq |m_{i,t+j}|$ and $m_{i,t}^2 \geq m_{i,t+j}^2$.

$$\left| -m_{i,t} - \frac{r}{\sqrt{1-r^2}} \frac{\phi\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)} - \frac{\phi\left(m_{i,t}\right)\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| \leq c(1+|m_{i,t}|)$$

$$\left| -m_{i,t+j} - \frac{r}{\sqrt{1-r^2}} \frac{\phi\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)} - \frac{\phi\left(m_{i,t+j}\right)\mathbf{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| \leq c(1+|m_{i,t}|)$$

$$\left| -\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\phi\left(m_{i,t}\right)\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| \leq c(1+|m_{i,t}|)$$

$$\left| -\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\phi\left(m_{i,t+j}\right)\mathbf{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| \leq c(1+|m_{i,t}|)$$

The term inside the square brackets in (35) is bounded by $|m_{i,t}|$ quadratically

$$\left| r + \frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}} - \sqrt{1-r^2}\frac{\phi\left(m_{i,t}\right)\phi\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| \leq c(1+m_{i,t}^2)$$

The term inside the square brackets in (36) is finite

$$\left| \frac{1}{\sqrt{1-r^2}} - \frac{\mathbf{\Phi}\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}{\phi\left(\frac{-rm_{i,t}+m_{i,t+j}}{\sqrt{1-r^2}}\right)}\frac{\phi\left(m_{i,t+j}\right)\mathbf{\Phi}\left(\frac{m_{i,t}-rm_{i,t+j}}{\sqrt{1-r^2}}\right)}{\mathbf{\Phi}_2\left(m_{i,t}, m_{i,t+j}|r\right)} \right| < \infty$$

Therefore, we can conclude that the norm of (29) is bounded by

$$\begin{aligned}
\left\| \frac{1}{\mathbb{P}_{00}}\frac{\partial^2 \mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}^2}\frac{\partial \mathbb{P}_{00}}{\partial\theta}\frac{\partial \mathbb{P}_{00}}{\partial\theta'} \right\| \leq & \left\| \frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'} \right\| c(1+|m_{i,t}|) \\
& + \left\| \frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'} \right\| c(1+|m_{i,t}|) \\
& + \left\| \frac{\partial^2 r}{\partial\theta\partial\theta'} \right\| c(1+|m_{i,t}|) \\
& + \left\| \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'} \right\| c(1+m_{i,t}^2) \\
& + \left\| \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'} \right\| c(1+m_{i,t}^2) \\
& + \left\| \frac{\partial r}{\partial\theta}\frac{\partial r}{\partial\theta'} \right\| c(1+m_{i,t}^4) \\
& + \left\| \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'} + \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'} \right\| c(1+m_{i,t}^2) \\
& + \left\| \frac{\partial m_{i,t}}{\partial\theta}\frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta}\frac{\partial m_{i,t}}{\partial\theta'} \right\| c(1+|m_{i,t}|^3) \\
& + \left\| \frac{\partial m_{i,t+j}}{\partial\theta}\frac{\partial r}{\partial\theta'} + \frac{\partial r}{\partial\theta}\frac{\partial m_{i,t+j}}{\partial\theta'} \right\| c(1+|m_{i,t}|^3)
\end{aligned}$$

Arranging the terms yields

$$
\begin{aligned}
&\left\| \frac{1}{\mathbb{P}_{00}} \frac{\partial^2 \mathbb{P}_{00}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{00}^2} \frac{\partial \mathbb{P}_{00}}{\partial\theta} \frac{\partial \mathbb{P}_{00}}{\partial\theta'} \right\| \\
&\leq c(1+|m_{i,t}|) \left(\left\| \frac{\partial^2 m_{i,t}}{\partial\theta\partial\theta'} \right\| + \left\| \frac{\partial^2 m_{i,t+j}}{\partial\theta\partial\theta'} \right\| + \left\| \frac{\partial^2 r}{\partial\theta\partial\theta'} \right\| \right) \\
&+ c(1+m_{i,t}^2) \left(\left\| \frac{\partial m_{i,t}}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'} \right\| + \left\| \frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'} \right\| + \left\| \frac{\partial m_{i,t}}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'} \right\| + \left\| \frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'} \right\| \right) \\
&+ c(1+|m_{i,t}|^3) \left(\left\| \frac{\partial m_{i,t}}{\partial\theta} \frac{\partial r}{\partial\theta'} \right\| + \left\| \frac{\partial r}{\partial\theta} \frac{\partial m_{i,t}}{\partial\theta'} \right\| + \left\| \frac{\partial m_{i,t+j}}{\partial\theta} \frac{\partial r}{\partial\theta'} \right\| + \left\| \frac{\partial r}{\partial\theta} \frac{\partial m_{i,t+j}}{\partial\theta'} \right\| \right) \\
&+ c(1+m_{i,t}^4) \left\| \frac{\partial r}{\partial\theta} \frac{\partial r}{\partial\theta'} \right\|
\end{aligned}
$$

A similar analysis not shown here indicates that the same upperbound is found for the other three norms $\left\| \frac{1}{\mathbb{P}_{kl}} \frac{\partial^2 \mathbb{P}_{kl}}{\partial\theta\partial\theta'} - \frac{1}{\mathbb{P}_{kl}^2} \frac{\partial \mathbb{P}_{kl}}{\partial\theta} \frac{\partial \mathbb{P}_{kl}}{\partial\theta'} \right\|$ for $\{k,l\} \in \{0,1\}$.

B.3 Bounds on functions of $m_t(\theta)$

In this subsection, I analyze the upperbounds for functions of $m_{i,t}(\theta)$, in particular the bounds for $m_{i,t}, m_{i,t}^2, m_{i,t}'$. First, remember that $m_{i,t}(\mathbf{x_i},\theta) = \sqrt{1-\rho^2}\sum_{k=0}^t \rho^k \beta' x_{i,t-k}$.

The bound for $m_{i,t}(\theta)$ is

$$
|m_{i,t}(\theta)| = \left| \sqrt{1-\rho^2} \sum_{k=0}^t \rho^k \beta' x_{i,t-k} \right| \leq \sqrt{1-\rho^2} \sum_{k=0}^t |\rho^k \beta' x_{i,t-k}| \leq \sum_{k=0}^t |\rho|^k \|\beta\| \|x_{i,t-k}\|
$$

The bound for $m_{i,t}^2(\theta)$ is

$$
\begin{aligned}
m_{i,t}^2(\theta) &= \left(\sqrt{1-\rho^2} \sum_{k=0}^t \rho^k \beta' x_{i,t-k} \right)^2 \\
&= (1-\rho^2) \sum_{k=0}^t \sum_{l=0}^t \rho^{k+l} \beta' x_{i,t-k} x_{i,t-l}' \beta \\
&\leq (1-\rho^2) \left| \sum_{k=0}^t \sum_{i=0}^t \rho^{k+l} \beta' x_{i,t-k} x_{t-i}' \beta \right| \\
&\leq (1-\rho^2) \sum_{k=0}^t \sum_{i=0}^t |\rho|^{k+l} \left| \beta' x_{i,t-k} x_{t-i}' \beta \right| \\
&\leq \sum_{k=0}^t \sum_{l=0}^t |\rho|^{k+l} \|\beta\|^2 \|x_{i,t-k}\| \|x_{i,t-l}\|
\end{aligned}
$$

The first derivative of $m_{i,t}(\theta)$ is

$$\begin{aligned}\frac{\partial m_{i,t}(\theta)}{\partial \theta'} &= -\sum_{k=0}^{t}\left[\frac{\partial\left(\sqrt{1-\rho^2}\rho^k\beta' x_{i,t-k}\right)}{\partial\rho}, \quad \frac{\partial\left(\sqrt{1-\rho^2}\rho^k\beta' x_{i,t-k}\right)}{\partial\beta'}\right] \\ &= -\sum_{k=0}^{t}\left[\left(\frac{-\rho}{\sqrt{1-\rho^2}}\rho^k + \sqrt{1-\rho^2}k\rho^{k-1}\right)\beta' x_{i,t-k}, \quad \sqrt{1-\rho^2}\rho^k x'_{i,t-k}\right] \\ &= -\sqrt{1-\rho^2}\sum_{k=0}^{t}\left[\left(\frac{-\rho^2}{1-\rho^2}+k\right)\rho^{k-1}\beta' x_{i,t-k}, \quad \rho^k x'_{i,t-k}\right]\end{aligned}$$

Next, let's write the norm of the derivative.

$$\begin{aligned}\left\|\frac{\partial m_{i,t}(\theta)}{\partial \theta'}\right\| &\le \sum_{k=0}^{t}\left\|\left[\left(\frac{-\rho^2}{1-\rho^2}+k\right)\rho^{k-1}\beta' x_{i,t-k}, \quad \rho^k x'_{i,t-k}\right]\right\| \\ &\le \sum_{k=0}^{t}\left|\left(\frac{-\rho^2}{1-\rho^2}+k\right)\rho^{k-1}\beta' x_{i,t-k}\right| + \sum_{k=0}^{t}\|\rho^k x'_{i,t-k}\| \\ &\le \sum_{k=0}^{t}\left(\frac{\rho^2}{1-\rho^2}+k\right)|\rho|^{k-1}\|\beta\|\|x_{i,t-k}\| + \sum_{k=0}^{t}|\rho|^k\|x_{i,t-k}\|\end{aligned}$$

Hence, $\sup_{\theta\in\Theta}\left\|\frac{\partial m_{i,t}(\theta)}{\partial\theta}\right\| \le M\sum_{k=0}^{t}(1+k)\bar{\rho}^k\|x_{i,t-k}\|$.

The second derivative of $m_{i,t}(\theta)$ is

$$\frac{\partial m_{i,t}(\theta)}{\partial\theta\partial\theta'} = -\sum_{k=0}^{t}\begin{bmatrix}\left(-\frac{2k+1}{\sqrt{1-\rho^2}}\rho^k - \frac{\rho^{k+2}}{\sqrt{1-\rho^2}^3} + \sqrt{1-\rho^2}k(k-1)\rho^{k-2}\right)\beta' x_{i,t-k} & \frac{\partial^2 m_{i,t}}{\partial\rho\partial\beta'} \\ \left(\frac{-\rho}{\sqrt{1-\rho^2}}\rho^k + \sqrt{1-\rho^2}k\rho^{k-1}\right)x_{i,t-k} & 0\end{bmatrix}$$

Hence, $\sup_{\theta\in\Theta}\left\|\frac{\partial m_{i,t}(\theta)}{\partial\theta\partial\theta'}\right\| \le M\sum_{k=0}^{t}(1+k^2)\bar{\rho}^k\|x_{i,t-k}\|$.