2015

Research related to high dimensional econometrics

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

RESEARCH RELATED TO HIGH DIMENSIONAL ECONOMETRICS

by

MINGLI CHEN

B.S., Shanghai University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2015

Approved by

First Reader

Pierre Perron, Ph.D.
Professor of Economics

Second Reader

Iván Fernández-Val, Ph.D.
Associate Professor of Economics

Third Reader

Marc Rysman, Ph.D.
Professor of Economics

Fourth Reader

Victor Chernozhukov, Ph.D.
MIT, Professor of Economics

Acknowledgments

I am deeply indebted to a long list of people for help, advice, and encouragement in writing this thesis and in completing my PhD.

First and foremost, I wish to express my great gratitude to my advisers, Victor Chernozhukov, Iván Fernández-Val, Pierre Perron, and Marc Rysman for their excellent advice, constant encouragement, and many helpful comments and discussions.

I thank Pierre for his wisdom, for many very helpful criticisms on both paper writing and oral presentations. I thank him for being strict with me while at the same time being very patient with me. I particularly thank him for spotting my mistakes and giving great suggestions, helping me form a practical view to econometrics, and for all the effort and time he spent nurturing me intellectually. His academic professionalism and everlasting strive for perfection inspire me greatly. I learned a lot through working with him, and there is still a lot to learn from him. Pierre has also pushed me to be meticulous and rigorous in my own work. It is with his help that I distinguish myself a little from a careless student and be a little closer to a scientific researcher. I also thank Pierre for helping me invite a seminar presenter from a statistics department.

I thank Ivan for being a good friend and great advisor. He is open to questions and discussions. I appreciate many helpful discussions with him, from which I benefited immensely. I also thank him for encouraging me to organize the econometrics reading group, and helping me with my first few presentations' slides. His insightful comments helped me greatly in the writing of my papers as well as in the job market. I admire his productivity and personality, as well as his great taste for food. I also thank him for buying me a birthday cake.

Special thanks goes to Marc, I thank him for hiring me as his research assistant. It was during this period that I generated the idea of working on nonlinear panel data models with factor structure. Chatting with Marc is stimulating, inspiriting, thought-provoking, and fun. Those chats also significantly improved this dissertation. I also thank him for stimulating me begin to be interested in empirical work. I highly recommend Marc as an adviser to anybody with an interest in empirical IO.

Running into Victor is a big structural change in my life. He is my first econometrics mentor. He taught me to use **R**. He broadens my view on econometrics. His constant help, guidance and support ever since I was an undergraduate has helped me go a long way. I like his short, concise and inspiring comments in his presentations, lectures and discussions. I thank his encouragement and support during my job market and some difficult times. I very much admire his works and interactions with people.

I thank Alexandre Belloni for being a great coauthor and many thoughtful discussions. I learn a lot through working with him, and sometimes I feel he is like one additional advisor. It has been great to work with Martin Weidner on a common project. I also thank him for useful job market advice. I also wish to thank all the teachers who taught me at various stages of my education. I especially thank Stephen Ryan for mentioning me in class, which gave me a lot of encouragement. His enthusiasm for IO is infectious. This also brings my thanks to my other IO professors Sergei Koulayev and Joshua Lustig. I thank Jiashun Jin for his teaching in statistics, in particular, his encouragement on in-class presentation.

I benefited a lot from the great research environment at BU. I thank Zhongjun Qu for many helpful discussions on econometrics and academic life in general. Our offices were neighboring for three years; I thank him for always having an open door and kindly sharing constructive criticisms. I thank Hiro Kaido for being my thesis defense chair, and offering useful econometrics comments. My thesis and academic life benefited from helpful conversations with numerous others, such as Manuel Arellano, Jushan Bai, William Greene, James Hamilton, Kengo Kato, Roger Koenker, Arthur Lewbel, Yuan Liao, Zhipeng Liao, Konrad Menzel, Whitney Newey, Morten Nielsen, Kevin Kyungchul Song, and many others at seminars and discussions during the job market. I especially thank Michael Wolf for giving us a box of Swiss chocolate. I am also grateful for my metrics friends at BU: Shuyang Bai, Julian Chan, Aparna Dutta, Francois Guay, Adam McCloskey, Andres Sagner, Yuan Tian, Denis Tkachenko, and Tomo Yabu. This thesis incorporates helpful suggestions from Matt

Johnson, Will Johnson, Ao Li, Caio Waisman, and Yiqing Xu. I'm particularly grateful to Brad Larsen for wonderful suggestions.

I would also like to express my warmest thanks to all the faculty and staff members at BU. I thank Randy Ellis for being such a great placement officer. I am in debt to him for many good suggestions and for answering my many questions. I owe him a beer. I appreciate Francesco Decarolis, Larry Epstein, Stefania Garetto, Keith Hylton, Jordi Jaumandreu, Kevin Lang, Bart Lipman, Alisdair McKay, Jianjun Miao, Dilip Mookherjee, Andrew Newman, Claudia Olivetti, Daniele Paserman, Johannes Schmieder for doing mock interviews with me, giving comments on my job market paper, or helping me with my job market (decision) process. Special thank to Sambuddha Ghosh for being a good friend and giving many very helpful suggestions. I thank Hsueh-Ling Huynh and Albert Ma for giving me candies, Andrew Campolieto and Gloria Murray for helping me with all administrative work, and Regina Cati for good meals.

I wish to thank my friends and classmates for accompanying and helping throughout my studies. I am very fortunate to have such a lovely group of classmates including everyone from my class and many from surrounding years. In no particular order, I thank Amrit Amirapu, Anindya Chakrabarti, Andrew Ellis, Buqu Gao, Jiannan Lu, Henry Mak, Elisabeth Perlman, William Shore, Ben Solow, Jordi Vidal-Robert, Christoph Walsh, Xufei Wang, Russell Weinstein, Mingzi Yi, Hongkai Zhang, Xiaojia Zhang and many others.

Most of all I wish to thank my big family. I thank my parents, for their kindness and examples. Their unconditional love gives me a lot of courage. I also want to express my thanks to my grandparents, especially my maternal grandfather and paternal grandmother, for loving me and taking care of me when I was a child. That also brings my thanks to my cousins for growing up together. I am also very fortunate to have many wonderful aunts and uncles. Last but not least, I thank my brother for supporting my decision for starting on a Ph.D. degree, which he likened to a journey. Indeed, it's a great journey, during which I met a lot of wonderful people.

RESEARCH RELATED TO HIGH DIMENSIONAL ECONOMETRICS

(Order No.)

MINGLI CHEN

Boston University, Graduate School of Arts and Sciences, 2015

Major Professor: Pierre Perron, Professor of Economics

ABSTRACT

This dissertation consists of three chapters related to high dimensional econometrics dealing with the estimation of nonlinear panel data models and networks models.

The first chapter proposes a fixed effects expectation-maximization estimator for a class of nonlinear panel data models with unobserved heterogeneity modeled as individual and/or time effects or an arbitrary interaction of the two. The estimator is obtained through a computationally simple iterative two-step procedure, both steps having a closed form solution. I show that the estimator is consistent in large panels, derive the asymptotic distribution for a probit model with interactive effects, and develop analytical bias corrections to deal with the incidental parameter problem.

The second chapter considers estimation and inference for semiparametric nonlinear panel single index models with interactive effects. These include static and dynamic probit, logit, and Poisson models. An iterative two-step procedure to maximize the likelihood is proposed. The estimator is consistent, but has bias due to the incidental parameter problem. Analytical and jackknife bias corrections are developed to remove the bias without increasing variance.

The third chapter proposes Quantile Graphical Models (QGMs) to characterize predictive and conditional dependence relationships within a set of random variables in non-Gaussian settings. These characterize the best linear predictor under asymmetric losses and the conditional dependence at each quantile. Estimators based on high-dimensional techniques are proposed. Each QGM represents the tail interdependence and the associated

tail risk network and can be used to measure systemic risk contributions for the study of financial contagion and hedging under a market downturn.

Contents

List of Tables

List of Figures

Chapter 1

Estimation of Nonlinear Panel Models with Multiple Unobserved Effects

1.1 Introduction

Panel data allow the possibility of controlling for unobserved heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it can result in misleading inference. For example, in demand estimation, unobserved individual heterogeneity is an important source of variation.

In this paper, I model unobserved heterogeneity as individual-specific effects to control for individual heterogeneity, and/or time specific effects to control for common shocks that occur to each individual. The way I control for those individual and time effects in nonlinear models is to treat each effect as a separate parameter to be estimated, and I propose a fixed effects expectation-maximization (EM) estimator that can be applied to a class of nonlinear panel data models with those individual and/or time effects. Of particular interest is the case of interactive effects, i.e., when the unobserved heterogeneity is modeled as a factor analytical structure. To the best of the author's knowledge, the current paper presents the first fixed effects EM-type estimator for nonlinear panel data models.

Interactive effects relax the invariant heterogeneity assumption and allow a more general model of time-varying heterogeneity. These interactive effects can be arbitrarily correlated with the observable covariates, which accommodates endogeneity and, at the same time, allows correlations between individual effects. As an example of why these interactive effects are important, (Moon et al., 2014), in a demand estimation setting, demonstrate that interactive fixed effects can capture strong persistence in market shares across products and markets, and find evidence that the factors are indeed capturing much of the unobservable product and time effects leading to price endogeneity.

The nonlinear panel data models with unobserved fixed effects that I consider in this paper have the following latent representation:

$$Y_{it}^{*} = X_{it}^{'}\beta + g(\alpha_i, \gamma_t) + \varepsilon_{it}, \tag{1.1}$$

$$Y_{it} = r(Y_{it}^{*}), \tag{1.2}$$

for $t = 1, ..., T$ and $i = 1, ..., N$. The econometrian observes Y_{it}, the dependent variable for individual i at time t (or t can be a group), and X_{it}, the time-variant $K \times 1$ regressor matrix. The econometrician does not observe Y_{it}^{*} (the latent dependent variable), α_i (the unobserved time-invariant individual effect), γ_t (the unobserved time effect), or ε_{it} (the unobserved error term). The vector β contains the main structural parameters of interest. The function $r(\cdot)$ is a known transformation of the unobserved latent variable. The individual effects α_i and time effects γ_t are allowed to be correlated with the regressor matrix. I do not make parametric assumptions on the distribution of either individual effects or time effects, hence the model is semiparametric.[1] The method proposed here can be applied to many functional forms between α_i and γ_t. The leading case I consider is when $g(\alpha_i, \gamma_t) = \alpha_i^{'}\gamma_t$ where both α_i and γ_t are $R \times 1$ vectors; note that this includes the special case settings with only individual effects or settings with additive individual and time effects.

Substantial theoretical and computational challenges are present in nonlinear panel models involving a large number of individual and time effects. In particular, in these models it is in general not possible to remove the unobserved effects by differencing as is commonly done in linear models. The incidental parameter problem, first pointed out by (Neyman and

[1] Relaxing parametric assumptions on the distribution of unobserved heterogeneity in nonlinear models is important, as often such restrictions cannot be justified by economic theory.

Scott, 1948), may also be present due to the fact that an estimator of β will be a function of the estimators of α_i and γ_t, which converges to their limits at slower convergence rates than that of β.

To deal with these problems, I propose a fixed effects expectation-maximization (EM) type estimator, which I denote IF-EM when applied to the interactive effects case. The estimator is obtained through an iterative two-step procedure, where the two steps have closed-form solutions. The first step (the "E"-step) involves obtaining the expectation of the mean utility function (the latent index) conditional on the observed dependent data.[2] The second step (the "M"-step) involves maximizing the resulting "linear" model. In practice, the estimator is simple and straightforward to compute. Monte Carlo simulations demonstrate it has good small-sample properties.

The incidental parameters problem might be present because estimates of fixed effects are partially consistent, and structural parameters of interest are functions of these estimates.[3] For example, I discuss a panel probit model with interactive fixed effects (which I denote PPIF) and demonstrate that its estimator PPIF is biased. I develop analytical bias corrections to deal with the incidental parameter problem. The correction is based on adapting to my setting the general asymptotic expansion of fixed effects estimators with incidental parameters in multiple dimensions under asymptotic sequences where both dimensions of the panel grow with the sample size (as in (Fernández-Val and Weidner, 2013)). In addition to model parameters, I provide bias corrections for average partial effects, which are functions of the data, parameters, and individual and time effects in nonlinear models.

The proposed model and estimates can have wide applications in economics. For example, factor structures have been used in a probit setting to represent market structure (as in (Elrod and Keane, 1995)) or, in a linear setting, to explain labor and behavioral out-

[2]As shown later, this is essentially an inverse distribution approach. For the exponential class of distributions, under Bregman loss, the conditional expectation is optimal in terms of MSE.

[3]The incidental parameters problem has different effects in different contexts and might not be present in some nonlinear models, e.g., Poisson models or slope coefficients in Tobit models. Additionally, marginal effects in probit models with individual fixed effects might not have bias or might have small bias, as shown in (Fernández-Val, 2009).

comes ((Heckman et al., 2006)) or estimate the evolution of cognitive and noncognitive skills ((Cunha and Heckman, 2008; Cunha et al., 2010)). Another example where the fixed effects approaches are used is the international trade partner choice (as in (Helpman et al., 2008)). The estimator is also particularly useful in empirical finance and in the setting with long time series, such as empirical work using PSID data. Furthermore, the estimation procedure can easily be extended to multinomial choice models.

This paper is related to multiple strands of the literature. First, it is related to the literature on linear panel data models with factor structures. (Bai, 2009a) estimates factors using the method of principal components. (Moon et al., 2014) extend the standard BLP random coefficients discrete choice demand model and propose a two-step procedure to calculate the estimator. Other related papers include (Holtz-Eakin et al., 1988; Ahn et al., 2001; Bai and Ng, 2002; Bai, 2003; Ahn et al., 2013; Andrews, 2005; Pesaran, 2006; Bai, 2009b; Moon and Weidner, 2010a), and (Moon and Weidner, 2010b). Some of these papers (e.g. (Bai, 2009b)) let $N \to \infty$ and $T \to \infty$ while others (e.g. (Ahn et al., 2013)) have T fixed and $N \to \infty$.

This paper is also related to the literature on nonlinear panel data models and bias correction, such as (Arellano and Hahn, 2007; Hahn and Newey, 2004; Hahn and Kuersteiner, 2002; Fernández-Val, 2009; Bester and Hansen, 2009; Carro, 2007; Fernández-Val and Vella, 2011; Bonhomme, 2012; Chamberlain, 1980), and (Dhaene and Jochmans, 2010). (Charbonneau, 2012) extends the conditional fixed effects estimators to logit and Poisson models with exogenous regressors and additive individual and time effects. (Fernández-Val and Weidner, 2013) develop analytical and jackknife bias corrections for nonlinear panel data models with additive individual and time effects. (Freyberger, 2012) studies nonparametric panel data models with multidimensional, unobserved individual effects when T is fixed. (Chen et al., 2013) develop analytical and jackknife estimators for a class of nonlinear panel data models with individual and time effects which enter the model interactively.

A final contribution of this paper is on the computation front, relating to the EM algorithm and latent backfiting procedure. Related work includes (Orchard and Woodbury, 1972; Dempster et al., 1977; Pan, 2002; Meng and Rubin, 1993; Laird, 1985), and (Pastorello et al., 2003).

The remainder of the paper is structured as follows. Section 1.2 introduces the model, the leading examples and their estimators. I also discuss the convergence of the estimation procedure. Section 1.3 presents consistency and asymptotic results for probit with interactive fixed effects. Section 1.4 presents some extensions and discussions. Section 1.5 contains Monte Carlo simulation results and Section 1.6 presents the empirical examples. Section 1.7 concludes. All proofs are contained in the Appendix.

1.2 Models and Estimators

In this section, I start with the panel probit with interactive individual and time effects case. I first specify the model and present the parameters and functional of interest and then show how the model can be estimated using the proposed EM procedure.

1.2.1 Panel probit with interactive fixed effects (PPIF)

1.2.1.1 Model

I consider the following interactive fixed effects probit model

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \alpha_i'\gamma_t + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{1.3}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, and β is a finite dimensional parameter vector. The variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively. The model is semiparameteric in that I do not specify the distribution of these effects nor their relationship with the explanatory variables, but, given that I consider probit in this section, I do specify ε to be normally distributed with unit variance.

Denoting the cumulative distribution function of ε_{it} as $\Phi(\cdot)$, the standard normal distribution, the conditional distribution of Y_{it} can then be written using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i, \gamma_t) = \Phi(X_{it}\beta + \alpha_i^{'}\gamma_t).$$

For estimation, I adopt a fixed effects approach, treating the unobserved individual and time effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, ..., \alpha_N, \gamma_1, ..., \gamma_N)'$. The model parameter β usually includes regression coefficients of interest, while the unobserved effects ϕ_{NT} are treated as nuisance parameters. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0, \gamma_1^0, ..., \gamma_T^0)'$. Other quantities of interest involve averages over the data and unobserved effects, such as average partial effects, which are often the ultimate quantities of interest in nonlinear models. These can be denoted

$$\delta_{NT}^0 = \mathbb{E}_\phi[\Delta_{NT}(\beta^0, \phi_{NT}^0)], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha_i^{'}\gamma_t), \tag{1.4}$$

where $\Delta(X_{it}, \beta, \alpha_i^{'}\gamma_t)$ represents some partial effect of interest and $\mathbb{E}_\phi$ denotes the expectation with respect to the distribution of the data, conditional on ϕ_{NT}^0 and β^0.

Some examples of partial effects are the following:

Example 1.2.1. (Average partial effects) If $X_{it,k}$, the k-th element of X_{it}, is binary, its partial effect for model specified by (1.3) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha_i^{'}\gamma_t) = \Phi(\beta_k + X_{it,-k}^{'}\beta_{-k} + \alpha_i^{'}\gamma_t) - \Phi(X_{it,-k}^{'}\beta_{-k} + \alpha_i^{'}\gamma_t), \tag{1.5}$$

where β_k is the k-th element of β, and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the k-th element. If $X_{it,k}$ is continuous, the partial effects of $X_{it,k}$ for model (1.3) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i, \gamma_t) = \beta_k \phi_f(X_{it}'\beta + \alpha_i'\gamma_t), \tag{1.6}$$

here $\phi_f(\cdot)$ is the derivative of Φ.

A particular application of this model is the study of international trade partner choice. For example, (Helpman et al., 2008) consider panel of unilateral trade flows between 158 countries for the year 1986. They use a probit model for the extensive margin of a gravity equation with exporter and importer country effects to allow for asymmetric trade.

Example 1.2.2. (International Trade)

$$Pr(Trade_{ij} = 1 | X_{ij}, \alpha_i \gamma_j) \quad = \Phi(X_{ij}'\beta + \alpha_i'\gamma_j), \quad \forall i, j \in V, \quad i \neq j,$$

here V contains the identities of all the countries considered.

Here $Trade_{ij}$ is an indicator for positive trade from country j to country i, X_{ij} includes log of bilateral distance, and nine indicators for geography, institution and culture differences.[4] In this setting, $N \approx T$. The estimated fixed effects can be used for forecasting network linkages or calculating average partial effects as well.

1.2.1.2 Estimator for panel probit with interactive fixed effects

In this section, I describe how the model with interactive fixed effects can be estimated using the proposed EM procedure. I discuss the case where the model has a known number of factors R.[5] I will start with $R = 1$; the case for $R > 1$ will be discussed in Section 1.4. For full identification, I assume $\gamma_1 = 1$, though different normalization restrictions can be imposed and will require different maximization steps, but this does not affect the estimation of β as the factor structure enters into the model jointly as $\alpha_i \gamma_t$.

[4]See (Helpman et al., 2008) for additional details.

[5]Choosing the number of factors is beyond the scope of this paper.

Definition 1.2.1. (PPIF) The EM procedure for estimating the panel probit model with interactive fixed effects is as follows:

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)} + \alpha_i^{(k)}\gamma_t^{(k)}$,

(2) E-step: Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} : &= E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)})\}, \end{aligned}$$

(3) M-step: This contains three conditional maximization (CM) steps

CM-step 1: Given α_i and γ_t, the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X_{it}' \right)^{-1} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} \left(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)} \gamma_t^{(k)} \right) \right\},$$

CM-step 2: Given β and γ_t, the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \left\{ \sum_{t=1}^{T} (\hat{Y}_{it}^{(k)} - X_{it}'\beta^{(k+1)}) \gamma_t^{(k)} \right\} \Big/ \sum_{t=1}^{T} \left\{ \gamma_t^{(k)} \right\}^2,$$

CM-step 3: Given β and α_i, the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \left\{ \sum_{i=1}^{N} (\hat{Y}_{it}^{(k)} - X_{it}'\beta^{(k+1)}) \alpha_i^{(k+1)} \right\} \Big/ \sum_{i=1}^{N} \left\{ \alpha_i^{(k+1)} \right\}^2,$$

(4) Iterate the above steps until convergence.

Convergence and consistency, along with the asymptotic distribution of β will be discussed in the next sections.

Note that the estimation procedure can be adapted to linear panel data models with interactive fixed effects, e.g. (Bai, 2009b). In a linear panel data model, Y^* is observed, and hence the E-step described here will not be needed. However, the conditional maximization procedure can still be applied to estimate a linear model.

The EM procedure proposed here is simple, easy to implement and has closed-form solutions in each step. The conditional maximization steps involves replacing the functional of the current estimates of the other parameters.[6]

Remark 1.2.1. Different normalizations for the individual and time effects can lead to different estimation procedures, even for linear models. For example, with the normalization $\gamma_1 = 1$, the linear panel data model with interactive fixed effects

$$Y_{it} = X_{it}'\beta + \alpha_i\gamma_t + \varepsilon_{it},$$

can be estimated as follows

CM-step 1: Given α_i and γ_t, the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'\right)^{-1}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\left(Y_{it} - \alpha_i^{(k)}\gamma_t^{(k)}\right)\right\},$$

CM-step 2: Given β and γ_t, the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \left\{\sum_{t=1}^{T}(Y_{it} - X_{it}'\beta^{(k+1)})\gamma_t^{(k)}\right\} \Big/ \sum_{t=1}^{T}\left\{\gamma_t^{(k)}\right\}^2,$$

CM-step 3: Given β and α_i, the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \left\{\sum_{i=1}^{N}(Y_{it} - X_{it}'\beta^{(k+1)})\alpha_i^{(k+1)}\right\} \Big/ \sum_{i=1}^{N}\left\{\alpha_i^{(k+1)}\right\}^2,$$

Iterate until convergence.

Since individual effects and additive individual and time effects are special cases of interactive effects, I will present results for the individual effects case only.[7] For the case with additive individual and time effects, see Appendix A.1.1.

[6]This is an expectation and conditional maximization (ECM) procedure, see (Meng and Rubin, 1993) for more details about ECM.

[7]More precisely, when the unobserved individual and time effects are multidimensional, the additive individual and time effects case is a special case of the interactive effects case.

1.2.2 Panel probit with only individual fixed effects

In this setting, I consider the following model

$$
\begin{aligned}
Y_{it}^* &= X_{it}'\beta + \alpha_i + \varepsilon_{it}, \\
Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\},
\end{aligned} \tag{1.7}
$$

for $i = 1, ..., N$ and $t = 1, ..., T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, β is a finite-dimensional parameter vector, α_i are unobserved individual effects.

Similarly to Section (1.2.1), I model the conditional distribution of Y_{it} using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i) = \Phi(X_{it}\beta + \alpha_i),$$

and for estimation I adopt a fixed effects approach treating the unobserved individual effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, ..., \alpha_N)'$. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0)'$. Other quantities of interest involve averages over the data and unobserved effects

$$\delta_{NT}^0 = \mathbb{E}[\Delta_{NT}(\beta^0, \phi_{NT}^0)], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha_i), \tag{1.8}$$

and examples of partial effects (Δ) are the following:

Example 1.2.3. (Average partial effects) If $X_{it,k}$, the k-th element of X_{it}, is binary, its partial effect for model (1.7) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha_i) = \Phi(\beta_k + X_{it,-k}'\beta_{-k} + \alpha_i) - \Phi(X_{it,-k}'\beta_{-k} + \alpha_i), \tag{1.9}$$

where β_k is the k-th element of β, and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β

except for the k-th element. If $X_{it,k}$ is continuous, for model (1.7) the partial effects of $X_{it,k}$ on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i) = \beta_k \phi_f(X_{it}^{'}\beta + \alpha_i), \tag{1.10}$$

where $\phi_f(\cdot)$ is the derivative of Φ.

Definition 1.2.2. The fixed effects EM estimator for panel probit with individual fixed effects is defined by

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)})$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)} + \alpha_i^{(k)}$,

(2) E-step: Calculate

$$\begin{aligned}\hat{Y}_{it}^{(k)} : &= E[Y_{it}^{*}|Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)})/\{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)})\},\end{aligned}$$

(3) **M-step**: This contains two conditional maximization steps

CM-step 1: Given α_i, the parameter β can be updated by

$$\beta^{(k+1)} = (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{'})^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)})\},$$

CM-step 2: Given β, the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_{it}^{(k)} - X_{it}^{'}\beta^{(k+1)}),$$

(4) Iterate until converge.

This is essentially the case $\gamma_t = 1, \forall t = 1, .., T$. Note that the CM-step 2 here is just the average over time using $\hat{Y}_{it}^{(k)}$ as surrogate for Y_{it}^{*}. This estimation procedure does not involve computing the inverse of the Hessian.

1.2.3 Nonlinear panel models with multiple unobserved effects

In this section, I describe how a general nonlinear panel data model with individual and time effects can be estimated using the proposed EM procedure.

Definition 1.2.3. The fixed effect EM estimator for a class of nonlinear panel data model with individual and time effects is defined by

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$;

(2) E-step: calculate $\hat{Y}_{it}^{(k)} := E[Y_{it}^*|Y_{it}, X_{it}, \beta^{(k)}, g(\alpha_i^{(k)}, \gamma_t^{(k)})]$,

(3) M-step:

$$(\beta^{(k+1)}, \alpha^{(k+1)}, \gamma^{(k+1)}) \in \underset{\beta,\alpha,\gamma}{\arg\min} S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)}) = (\hat{Y}_{it}^{(k)} - X_{it}'\beta - g(\alpha_i, \gamma_t))^2), \quad (1.11)$$

(4) Iterate until convergence.

Convergence and consistency of $\hat{\beta}$, defined as the output from the iteration, will be discussed in the following sections. Note that this procedure is different from the traditional EM algorithm (discussed in (Dempster et al., 1977)), which is used to maximize the expected log-likelihood function when there are latent variables, and its E-step is to augment the incomplete likelihood with conditional likelihood for $Y_{it}^*|Y_{it}$; while here, the E-step is to calculate a surrogate, $\hat{Y}_{it}$, for the unobserved Y_{it}^* when there are unobserved individual and time effects. This difference leads to a different strategy of proof. Specifically, I adopt the approach of using the conditional expectation of Y_{it}^* because under Bregman loss the conditional expectation is optimal in terms of mean squared error. Under certain conditions, e.g., the density of the error term is in the exponential class of distributions, as shown in Section 1.3, as well as for probit, those two have the same score functions. This is due to the quadratic loss function of the probit model.

Remark 1.2.2. Depending on the functional form of the individual and/or time effects, the M-step can be as follows:

CM-step 1: Given α_i and γ_t, the parameter β is updated via

$$\beta^{(k+1)} = (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{'})^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}(\hat{Y}_{it}^{(k)} - g(\alpha_i^{(k)}, \gamma_t^{(k)}))\},$$

CM-step 2: Given β, the parameters α_i and γ_t are updated by maximizing

$$-\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{Y}_{it}^{(k)} - X_{it}^{'}\beta - g(\alpha_i^{(k)}, \gamma_t^{(k)}))^2,$$

and this step can be implemented by using the method of least squares (or principal components).

1.2.3.1 Convergence

In this section, I show the resulting estimate from the estimation procedure converges to a point that maximizes the observed log-likelihood function. I focus on the interactive fixed effects case, which is more complex due to the high degree of nonlinearity of the unobserved effects term (all the other cases are concave in the fixed effects, though the convergence rates are different). Consistency results are discussed in Section 1.4. The IF-EM for probit suffers from asymptotic bias because the fixed effects converge slowly, which I address in Section 1.3.

For a binary model, denote the negative log-likelihood function

$$-\mathcal{L}_{NT} = -\sum_{i,t} logF(q_{it}(X_{it}^{\prime}\beta + \alpha_i^{\prime}\gamma_t)),$$

where $q_{it} := 2Y_{it} - 1$ and F is the cdf of Y_{it} conditional on X_{it},α_i and γ_t. For brevity, assume F is symmetric. Define the hazard function $h(\theta_1) := -\partial logF(\theta_1)/\partial\theta_1$ for a particular argument θ_1.

Recall the quadratic loss function $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)}) = (\hat{Y}_{it}^{(k)} - X_{it}^{'}\beta - g(\alpha_i, \gamma_t))^2$ of the M-step that the proposed fixed effects EM-type estimator depends on. The strategy of the proof is to show the negative log likelihood function of the model under consideration is majorized by this quadratic function (up to some constant), which is satisfied by the following propositions

Proposition 1.2.1. *Suppose X is a three-dimensional matrix with p sheets $(N \times T \times p)$, β and $\tilde{\beta}$ are $p \times 1$ vectors, α and $\tilde{\alpha}$ are $N \times R$ matrices, and γ and $\tilde{\gamma}$ are $T \times R$ matrices. Define $\tilde{h}_{it} := h(q_{it}(X_{it}'\tilde{\beta} + \tilde{\alpha}_i'\tilde{\gamma}_t))$, then*

$$-\mathcal{L}_{NT}(\beta, \alpha, \gamma) \leq -\mathcal{L}_{NT}(\tilde{\beta}, \tilde{\alpha}, \tilde{\gamma}) - \frac{1}{2}\sum_{i,t}\tilde{h}_{it}^2 + \frac{1}{2}\sum_{i,t}(\tilde{z}_{it} - X_{it}'\beta - \alpha_i'\gamma_t)^2.$$

Proof: See Appendix A.1.2.

Proposition 1.2.2. *(i) Up to a constant that depends on $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ but not on (β, α, γ), the function $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ majorizes $-\mathcal{L}_{NT}(\beta, \alpha, \gamma)$ at $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$.*

(ii) Let $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$, $k = 1, 2, ...,$ be a sequence obtained by the IF-EM procedure. Then $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ decreases as k increases and converges to a local minimum of $-\mathcal{L}_{NT}(\beta, \alpha, \gamma)$ as k goes to infinity.

The proof of part (i) follows by applying the result from Proposition 1.2.1. The proof of part (ii) follows from the property of the quadratic majorization.

This proves the convergence of the general EM procedure. Note that although I show it for an interactive fixed effects model, the same proof procedure can be adapted to other single index models with individual and time fixed effects. I discuss consistency in Section 1.4. Since the asymptotic distribution differs for different models, in the next section I will show the asymptotic distribution for the probit model, in which case the incidental parameter problem occurs, and provide an analytical bias correction solution.

The EM procedure proposed here is simple, easy to implement, and has a closed form solution in each step. The method can be extended in a straightforward way to handle composite data which consists of both binary and continuous variables. While the binary variables are modeled with Bernoulli distributions, the continuous variables can be modeled with Gaussian distributions. Including some continuous variables corresponds to adding some Gaussian log-likelihood terms to the existing log-likelihood expression. Since the Gaussian log-likelihood is quadratic, the ultimate function would still be majorized by a quadratic function.[8]

1.3 Asymptotic theory for panel probit with interactive fixed effects

In this section, I discuss consistency and asympototic bias of the proposed estimator. I do so in the context of PPIF but my method of proof can be extended to a wider class of models.

1.3.1 Consistency

I show PPIF is consistent but suffers from incidental parameters bias. I will also discuss bias corrections to the parameter and average partial effects in the next section.

I consider a panel probit model with scalar individual and time effects that enter the likelihood function interactively through $\pi_{it} = \alpha_i \gamma_t$. In this model, the dimension of the incidental parameters is $\dim \phi_{NT} = N + T$. I prove the consistency of PPIF under assumptions on the indexes. Since the proposed fixed effects EM estimator has the same score as that of MLE, I derive its properties directly through the expansion of the score of its profile likelihood function.

In this section, the parametric part of the model takes the form

$$log\Phi(q_{it}(X_{it}'\beta + \pi_{it})) = \ell_{it}(\beta, \pi_{it}).$$

Hence, the log-likelihood function is

$$\mathcal{L}_{NT}(\beta, \phi_{NT}) = \mathcal{L}_{NT}(\beta, \pi) = \frac{1}{NT}\sum_{i,t}\ell_{it}(\beta, \pi) = \frac{1}{NT}\sum_{i,t} log\Phi(q_{it}(X_{it}'\beta + \pi_{it})).$$

[8] When there are no fixed effects, convergence is proved by the contraction mapping theorem argument. See (Gourieroux et al., 1987)

I make the following assumptions:

Assumption 1. *Let* $v > 0$ *and* $\mu > 4(8+v)/v$. *Let* $\varepsilon > 0$ *and let* $\mathcal{B}^0_\varepsilon$ *be a subset of* $\mathbb{R}^{\dim \beta + 1}$ *that contains an* ε*-neighborhood of* (β^0, π^0_{it}) *for all* i, t, N, T.

(i) Asymptotics: Consider limits of sequences where $N/T \to \kappa^2$, $0 < \kappa < \infty$, *as* $N, T \to \infty$.

(ii) Sampling: Conditional on ϕ, $\{(Y_i^T, X_i^T) : 1 \leq i \leq N\}$ *is independent across* i, *and for each* i, $\{Y_{it}, X_{it} : 1 < t \leq T\}$ *is* α*-mixing with mixing coefficients satisfying* $\sup_i a_i(m) = O(m^{-\mu})$ *as* $m \to \infty$, *where*

$$a_i(m) := \underset{t}{sup} \underset{A \in \mathcal{A}^i_t, B \in \mathcal{B}^i_{t+m}}{sup} |P(A \cap B) - P(A)P(B)|$$

and for $Z_{it} = (Y_{it}, X_{it})$, $\mathcal{A}^i_t$ *is the sigma field generated by* $(Z_{it}, Z_{i,t-1}, ...)$, *and* $\mathcal{B}^i_t$ *is the sigma field generated by* $(Z_{it}, Z_{i,t+1}, ...)$.

(iii) Moments: The partial derivatives of $\ell_{it}(\beta, \pi)$ *w.r.t. the elements of* (β, π) *up to fourth order are bounded in absolute value uniformly over* $(\beta, \pi) \in \mathcal{B}^0_\varepsilon$ *by a function* $M(Z_{it}) > 0$ *a.s., and* $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+v}]$ *is a.s. uniformly bounded over* N, T. *There exist constants* $b_{\min}$ *and* $b_{\max}$ *such that for all* $(\beta, \pi) \in \mathcal{B}^0_\varepsilon$, $0 < b_{\min} \leq -\mathbb{E}_\phi[\partial_{\pi^2} \ell_{it}(\beta, \pi)] \leq b_{\max}$ *a.s. uniformly over* i,t, N, T.

(iv) Non-colinearity condition: $\exists c > 0$, *such that w.p.a.1,*

$$\underset{\{\alpha \in \mathbb{R}, \|\alpha\| = 1\}}{min} \underset{\Lambda \in \mathbb{R}^{N \times 2}}{min} \frac{1}{NT} Tr[(\alpha \cdot X)' M_\alpha (\alpha \cdot X)] > c$$

Assumption (i) defines the large-T asymptotic framework as in (Hahn and Kuersteiner, 2002; Fernández-Val and Weidner, 2013; Chen et al., 2013). Assumption (ii) defines the data sampling conditions. Assumption (iii) defines the finite moment condition. Assumption (iv) states that no linear combination of the regressors converges to zero, even after projecting any two-dimensional factor loading α. Note that this rules out time-invariant and cross-sectional invariant regressors.

Define the fixed effects EM estimator for PPIF as $\hat{\beta}_{PPIF}$.

Lemma 1.3.1. *Under Assumption 1,* $\hat{\beta}_{PPIF} = \beta^0 + o_P(1)$.

The proof is found in Appendix A.2.1 and contains two steps. I first show the consistency of the index with the generalized residuals from the E-step. Then, in step two I show that the residuals satisfy the conditions imposed on the linear panel data models with interactive fixed effects as in (Bai, 2009b). The consistency of $\hat{\beta}_{PPIF}$ follows.

1.3.2 Asymptotic results

Define the nonlinear differencing operator

$$D_{\beta\pi^q}\ell_{it} := \partial_{\pi^{q+1}}\ell_{it}(X_{it} - \Xi_{it}), \quad for\ q = 0, 1, 2$$

where Ξ_{it} is a $\dim\beta$-vector including the least squares projections of X_{it} on the space of incidental parameters spanned by $\alpha_i^0\gamma_t^0(\alpha_i + \gamma_t)$ weighted by $\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})$, i.e.,

$$\Xi_{it,k} = \alpha_i^0\gamma_t^0(\alpha_{i,k}^* + \gamma_{t,k}^*), \tag{1.12}$$

$$(\alpha_k^*, \gamma_k^*) \in \underset{\alpha_{i,k},\gamma_{t,k}}{\arg\min}\sum_{i,t}\mathbb{E}_\phi[-\partial_{z^2}\ell_{it}(X_{it} - \alpha_i^0\gamma_t^0(\alpha_{i,k} + \gamma_{t,k}))^2].$$

Let $\overline{\mathcal{H}}$ be the $(N+T)\times(N+T)$ expected value of the Hessian matrix of the log-likelihood with respect to the nuisance parameters evaluated at the true parameters, i.e.,

$$\overline{\mathcal{H}} = \mathbb{E}_\phi[-\partial_{\phi\phi'}\mathcal{L}] = \begin{bmatrix} \overline{\mathcal{H}}_{(\alpha\alpha)} & \overline{\mathcal{H}}_{(\alpha\gamma)} \\ \overline{\mathcal{H}}'_{(\alpha\gamma)} & \overline{\mathcal{H}}_{(\gamma\gamma)} \end{bmatrix},$$

where $\overline{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(\sum_t(\gamma_t^0)^2\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$, $\overline{\mathcal{H}}_{(\alpha\gamma)it} = (\alpha_i^0\gamma_t^0\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$, and $\overline{\mathcal{H}}_{(\gamma\gamma)} = \text{diag}(\sum_i(\alpha_i^0)^2\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$. Furthermore, let $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\overline{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)}$, and $\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}$

denote the $N \times N$, $N \times T$, $T \times N$ and $T \times T$ blocks of the inverse $\overline{\mathcal{H}}^{-1}$ of $\overline{\mathcal{H}}$. Then

$$\begin{aligned} \Xi_{it} &= -\frac{1}{NT}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}\gamma^0_\tau\gamma^0_t + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}\alpha^0_j\gamma^0_t \\ &+ \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\alpha^0_i\gamma^0_\tau + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\alpha^0_i\alpha^0_j)\mathbb{E}_\phi(\partial_{\beta\pi}\ell_{j\tau}). \end{aligned} \tag{1.13}$$

This nonlinear differencing operator generalizes to nonlinear models the partialing-out of individual and time effects in linear models. For example, if the model is linear, $\partial_{\pi^2}\ell_{it} = -1$, $\partial_{\beta\pi}\ell_{it} = -X_{it}$, and

$$\Xi_{it} = T^{-1}\sum_{t=1}^{T}\mathbb{E}_\phi[X_{it}] + N^{-1}\sum_{i=1}^{N}\mathbb{E}_\phi[X_{it}] - (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi[X_{it}],$$

so that $D_\beta\ell_{it} = -(X_{it} - \Xi_{it})\partial_\pi\ell_{it}$, $D_{\beta\pi}\ell_{it} = -(X_{it} - \Xi_{it})$, and $D_{\beta\pi^2}\ell_{it} = 0$.

Let $\overline{\mathbb{E}} := \text{plim}_{N,T\to\infty}$. The following theorem establishes the asymptotic distribution of the fixed effects EM estimator for PPIF, $\hat{\beta}_{PPIF}$.

Theorem 1.3.1. *(Asymptotic distribution of $\hat{\beta}_{PPIF}$). Suppose that Assumption 1 holds, that the following limits exist*

$$\begin{aligned} \overline{B}_\infty &= -\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\sum_{\tau=t}^{T}\gamma^0_t\gamma^0_\tau\mathbb{E}_\phi[\partial_\pi\ell_{it}D_{\beta\pi}\ell_{i\tau}] + \frac{1}{2}\sum_{t=1}^{T}(\gamma^0_t)^2\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum_{t=1}^{T}(\gamma^0_t)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}\right], \\ \overline{D}_\infty &= -\overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha^0_i)^2\mathbb{E}_\phi(\partial_\pi\ell_{it}D_{\beta\pi}\ell_{it} + \frac{1}{2}D_{\beta\pi^2}\ell_{it})}{\sum_{i=1}^{N}(\alpha^0_i)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}\right], \\ \overline{W}_\infty &= -\overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi(\partial_{\beta\beta'}\ell_{it} - \partial_{\pi^2}\ell_{it}\Xi_{it}\Xi_{it}')\right], \end{aligned}$$

and that $\overline{W}_\infty > 0$. Then,

$$\sqrt{NT}(\hat{\beta}_{PPIF} - \beta^0) \xrightarrow{d} \overline{W}_\infty^{-1}N(\kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty, \overline{W}_\infty).$$

The detailed proof is in Appendix A.2.2.

Let $\tilde{X}_{it} = X_{it} - \Xi_{it}$ be the residual of the least squares projection of X_{it} on the space spanned by the incidental parameters weighted by $\mathbb{E}_\phi(\omega_{it})$, for $\omega_{it} = (\phi_f(X_{it}'\beta + \alpha_i^0\gamma_t^0))^2/[\Phi(X_{it}'\beta^0 + \alpha_i^0\gamma_t^0)(1 - \Phi(X_{it}'\beta + \alpha_i^0\gamma_t^0))]$.

Remark 1.3.1. For the probit model with X_{it} strictly exogenous, observe that

$$\begin{aligned}
\overline{B}_\infty &= \overline{\mathbb{E}}[\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi[\omega_{it}\tilde{X}_{it}\tilde{X}_{it}']}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi[\omega_{it}]}]\beta^0, \\
\overline{D}_\infty &= \overline{\mathbb{E}}[\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi[\omega_{it}\tilde{X}_{it}\tilde{X}_{it}']}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi[\omega_{it}]}]\beta^0, \\
\overline{W}_\infty &= \overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi[\omega_{it}\tilde{X}_{it}\tilde{X}_{it}']\right].
\end{aligned}$$

The asymptotic bias is therefore a positive-definite-matrix of the weighted average of the true parameters as in the case of the probit model with additive effects (see (Fernández-Val and Weidner, 2013)).

1.3.3 Asymptotic distribution of the average partial effects

In nonlinear models, the researcher is often interested in average partial effects in addition to the model structural parameters. These effects are averages of the data, parameters and unobserved effects as in equation (1.4). I impose the following sampling and moment conditions on the function Δ that defines the partial effects:

Assumption 2. *(Partial effects). Let $v > 0$, $\epsilon > 0$, and $\mathcal{B}_\epsilon^0$ all be as in Assumption 1*

(i) Sampling: for all N, T,$\{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are deterministic;

(ii) Model: for all i, t, N, T, the partial effects depend on α_i and γ_t through $\alpha_i\gamma_t$:

$$\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i\gamma_t).$$

The realizations of the partial effects are denoted by $\Delta_{it} := \Delta_{it}(\beta^0, \alpha_i^0\gamma_t^0)$.

(iii) Moments: The partial derivatives of $\Delta_{it}(\beta, \pi)$ with respect to the elements of (β, π) up to fourth order are bounded in absolute value uniformly over $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$ by a function $M(Z_{it}) > 0$ a.s., and $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+v}]$ is a.s. uniformly bounded over N, T.

(iv) Non-degeneracy and moments: $\min_{i,t} Var(\Delta_{it}) > 0$ and $\max_{i,t} Var(\Delta_{it}) < \infty$, uniformly over N, T.

Analogous to Ξ_{it} in equation (1.13), define

$$\Psi_{it} = -\frac{1}{NT}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}\gamma_\tau^0\gamma_t^0 + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}\alpha_j^0\gamma_t^0 + \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\alpha_i^0\gamma_\tau^0 + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\alpha_i^0\alpha_j^0)\partial_\pi\Delta_{j\tau},$$

which is the population projection of $\partial_\pi\Delta_{it}/\mathbb{E}_\phi[\partial_{\pi^2}\ell_{it}]$ on the space spanned by the incidental parameters under the metric given by $\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]$. I use a analogous notation to the previous section for the derivatives with respect to β and higher order derivatives with respect to π.

Let δ^0_{NT} be the APE as defined in equation (1.4), and $\hat{\delta}$ be its estimator $\Delta_{NT}(\hat{\beta}, \hat{\phi}_{NT}) = (NT)^{-1}\sum_{i,t}\Delta(X_{it}, \hat{\beta}, \hat{\alpha}_i\hat{\gamma}_t)$. The following theorem establishes the asymptotic distribution of $\hat{\delta}$.

Theorem 1.3.2. *(Asymptotic distribution of $\hat{\delta}$). Suppose that the assumptions of Theorem 1.3.1 and Assumption 2 hold, and that the following limits exist:*

$$\begin{aligned}
\overline{(D_\beta\Delta)}_\infty &= \overline{\mathbb{E}}[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi(\partial_\beta\Delta_{it} - \Xi_{it}\partial_\pi\Delta_{it})],\\
\overline{B}^\delta_\infty &= \overline{(D_\beta\Delta)}'_\infty\overline{W}^{-1}_\infty\overline{B}_\infty + \overline{\mathbb{E}}[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\sum_{\tau=t}^{T}\gamma_t^0\gamma_\tau^0\mathbb{E}_\phi(\partial_\pi\ell_{it}\partial_{\pi^2}\ell_{i\tau}\Psi_{i\tau})}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}]\\
&\quad -\overline{\mathbb{E}}[\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\gamma_t^0)^2[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})]}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}],\\
\overline{D}^\delta_\infty &= \overline{(D_\beta\Delta)}'_\infty\overline{W}^{-1}_\infty\overline{D}_\infty + \overline{\mathbb{E}}[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi(\partial_\pi\ell_{it}\partial_{\pi^2}\ell_{it}\Psi_{it})}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}]\\
&\quad -\overline{\mathbb{E}}[\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})]}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}],
\end{aligned}$$

$$\overline{V}^{\delta}_{\infty} = \overline{\mathbb{E}}\{\frac{1}{NT}\sum_{i=1}^{N}[\sum_{t,\tau=1}^{T}\mathbb{E}_{\phi}(\widetilde{\Delta}_{it}\widetilde{\Delta}^{'}_{i\tau}) + \sum_{t=1}^{T}\mathbb{E}_{\phi}(\Gamma_{it}\Gamma^{'}_{it})]\},$$

for some $\overline{V}^{\delta}_{\infty} > 0$*, where* $\tilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$ *and* $\Gamma_{it} = \overline{(D_{\beta}\Delta)}^{'}_{\infty}\overline{W}^{-1}_{\infty}D_{\beta}\ell_{it} - \mathbb{E}_{\phi}(\Psi_{it})\partial_{\pi}\ell_{it}$*. Then,*

$$\sqrt{NT}(\hat{\delta} - \delta^{0}_{NT} - T^{-1}\overline{B}^{\delta}_{\infty} - N^{-1}\overline{D}^{\delta}_{\infty}) \xrightarrow{d} N(0, \overline{V}^{\delta}_{\infty}).$$

The bias and variance are of the same order asymptotically under the asymptotic sequence of Assumption 1(i).

Remark 1.3.2. (Average effects from bias-corrected estimators). As in the case of the probit with additive effects ((Fernández-Val and Weidner, 2013)), the first term in the expressions of the biases $\overline{B}^{\delta}_{\infty}$ and $\overline{D}^{\delta}_{\infty}$ comes from the bias of the estimator of β. It drops out when the APEs are constructed from asymptotically unbiased or bias-corrected estimators of the parameter β, i.e.,

$$\tilde{\delta} = \Delta(\tilde{\beta}, \hat{\phi}(\tilde{\beta})),$$

where $\tilde{\beta}$ is such that $\sqrt{NT}(\tilde{\beta} - \beta^{0}) \xrightarrow{d} N(0, \overline{W}^{-1}_{\infty})$. The asymptotic variance of $\tilde{\delta}$ is the same as in Theorem 1.3.2.

In the following examples I assume that the APEs are constructed from asymptocially unbiased estimators of the model parameters.

Example 1.3.1. Consider the partial effects defined in (1.5) and (1.6) with

$$\Delta_{it}(\beta, \alpha_i\gamma_t) = \Phi(\beta_k + X^{'}_{it,-k}\beta_{-k} + \alpha_i\gamma_t) - \Phi(X^{'}_{it,-k}\beta_{-k} + \alpha_i\gamma_t)$$

and

$$\Delta_{it}(\beta, \alpha_i\gamma_t) = \beta_k\phi_f(X^{'}_{it}\beta + \alpha_i\gamma_t).$$

Denote $H_{it} = \phi_f(X^{'}_{it}\beta + \alpha^{0}_{i}\gamma^{0}_{t})/[\Phi(X^{'}_{it}\beta^{0} + \alpha^{0}_{i}\gamma^{0}_{t})(1 - \Phi(X^{'}_{it}\beta + \alpha^{0}_{i}\gamma^{0}_{t}))]$ and use notations previously introduced, the components of the asymptotic bias of $\tilde{\delta}$ are

$$\overline{B}^{\delta}_{\infty} = \overline{\mathbb{E}}[\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}[2\sum_{\tau=t+1}^{T}\mathbb{E}_{\phi}(H_{it}(Y_{it}-\Phi_{it})\omega_{i\tau}\tilde{\Psi}_{i\tau})-\mathbb{E}_{\phi}(\Psi_{it})\mathbb{E}_{\phi}(H_{it}\partial^2\Phi_{it})}{\sum_{t=1}^{T}\mathbb{E}_{\phi}(\omega_{it})}]$$
$$+ \overline{\mathbb{E}}[\frac{1}{2N}\sum_{i=1}^{N}\frac{\mathbb{E}_{\phi}(\partial_{\pi^2}\Delta_{it})]}{\sum_{t=1}^{T}\mathbb{E}_{\phi}(\omega_{it})}]$$
$$\overline{D}^{\delta}_{\infty} = \overline{\mathbb{E}}[\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}[-\mathbb{E}_{\phi}(\Psi_{it})]\mathbb{E}_{\phi}(H_{it}\partial^2\Phi_{it})+\mathbb{E}_{\phi}(\partial_{\pi^2}\Delta_{it})}{\sum_{i=1}^{N}\mathbb{E}_{\phi}(\omega_{it})}]$$

where $\tilde{\Psi}_{it}$ is the residual of the population regression of $-\partial_{\pi}\Delta_{it}/\mathbb{E}_{\phi}[\omega_{it}]$ on the space spanned by the incidental parameters under the metric given by $\mathbb{E}_{\phi}[\omega_{it}]$. If all the components of X_{it} are strictly exogenous, the first term in the numerator of $\overline{B}^{\delta}_{\infty}$ is zero.

1.3.4 Bias-corrected estimators

The results of the previous sections show that the asymptotic distributions of the interactive fixed effects estimators of the model parameters and APEs can have asymptotic bias under sequences where T grows at the same rate as N, as also discussed in (Chen et al., 2013). This large-T version of the incidental parameters problem can invalidate any inference based on the asymptotic distribution. In this section I discuss how to construct analytical bias corrections for PPIF and give conditions for the asymptotic validity of the analytical bias corrections. The proof strategy here is similar to (Fernández-Val and Weidner, 2013) which is under the additive individual and time effects setting.

The analytical corrections are constructed using sample analogs of the expressions in Theorems 1.3.1 and 1.3.2, replacing the true values of β and ϕ by the estimated ones. To describe these corrections, I introduce some additional notation. For any function of the data, unobserved effects and parameters $\varphi_{itj}(\beta, \alpha_i\gamma_t, \alpha_i\gamma_{t-j})$ with $0 \leq j < t$, let $\hat{\varphi}_{itj} = \varphi_{it}(\hat{\beta}, \hat{\alpha}_i\hat{\gamma}_t, \hat{\alpha}_i\hat{\gamma}_{t-j})$ be its estimator, e.g., $\mathbb{E}_{\phi}[\widehat{\partial_{\pi^2}\ell_{it}}]$ denotes the estimator of $\mathbb{E}_{\phi}[\partial_{\pi^2}\ell_{it}]$. Let

$\hat{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\hat{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\hat{\mathcal{H}}^{-1}_{(\gamma\alpha)}$ and $\hat{\mathcal{H}}^{-1}_{(\gamma\gamma)}$ denote the blocks of the matrix $\hat{\mathcal{H}}^{-1}$, where

$$\hat{\mathcal{H}} = \begin{pmatrix} \hat{\mathcal{H}}_{(\alpha\alpha)} & \hat{\mathcal{H}}_{(\alpha\gamma)} \\ \hat{\mathcal{H}}'_{(\alpha\gamma)} & \hat{\mathcal{H}}_{(\gamma\gamma)} \end{pmatrix},$$

$\hat{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(-\sum_t (\hat{\gamma}_t)^2 \mathbb{E}_\phi[\widehat{\partial_{\pi^2}\ell_{it}}])/(NT)$, $\hat{\mathcal{H}}_{(\alpha\gamma)it} = -\hat{\alpha}_i \hat{\gamma}_t \mathbb{E}_\phi[\widehat{\partial_{\pi^2}\ell_{it}}]/(NT)$, and $\hat{\mathcal{H}}_{(\gamma\gamma)} = \text{diag}(-\sum_i (\hat{\alpha}_i)^2 \mathbb{E}_\phi[\widehat{\partial_{\pi^2}\ell_{it}}])/(NT)$. Let

$$\hat{\Xi}_{it} = -\frac{1}{NT}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\hat{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}\hat{\gamma}_\tau\hat{\gamma}_t + \hat{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}\hat{\alpha}_j\hat{\gamma}_t + \hat{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\hat{\alpha}_i\hat{\gamma}_\tau + \hat{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\hat{\alpha}_i\hat{\alpha}_j)\mathbb{E}_\phi(\widehat{\partial_{\beta\pi}\ell_{j\tau}}),$$

the kth component of $\hat{\Xi}_{it}$ corresponds to a least square regression of X_{it} on the space spanned by the incidental parameters weighted by $-\mathbb{E}_\phi(\widehat{\partial_{\beta\pi}\ell_{it}})$. The analytical bias-corrected estimator of β^0 is

$$\tilde{\beta}^A = \hat{\beta} - \hat{B}/T - \hat{D}/N,$$

where

$$\hat{B} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=0}^{L}(T/(T-j))\sum_{t=j+1}^{T}\hat{\gamma}_t\hat{\gamma}_\tau\mathbb{E}(\partial_\pi\widehat{\ell_{it}D_{\beta\pi}}\ell_{i\tau}) + \frac{1}{2}\sum_{t=1}^{T}(\hat{\gamma}_t)^2\mathbb{E}(\widehat{D_{\beta\pi^2}\ell_{it}})}{\sum_{t=1}^{T}(\hat{\gamma}_t)^2\mathbb{E}(\widehat{\partial_{\pi^2}\ell_{it}})},$$

$$\hat{D} = -\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\hat{\alpha}_i)^2\mathbb{E}(\partial_\pi\widehat{\ell_{it}D_{\beta\pi}}\ell_{it} + \frac{1}{2}\widehat{D_{\beta\pi^2}\ell_{it}})}{\sum_{i=1}^{N}(\hat{\alpha}_i)^2\mathbb{E}(\widehat{\partial_{\pi^2}\ell_{it}})},$$

and L is a trimming parameter for estimation of spectral expectations such that $L \to \infty$ and $L/T \to 0$, see (Hahn and Kuersteiner, 2011).

Asymptotic $(1-p)$- confidence intervals for the components of β^0 can be formed as

$$\tilde{\beta}^A_k \pm z_{1-p}\sqrt{\widehat{W}^{-1}_{kk}/(NT)}, \quad k = \{1, ..., \dim \beta^0\}.$$

where z_{1-p} is the $(1-p)$ quantile of the standard normal distribution, and $\widehat{W}^{-1}_{kk}$ is the

(k,k)-element of the matrix $\widehat{W}^{-1}$ with

$$\widehat{W} = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi(\widehat{\partial_{\beta\beta'}\ell_{it}}) - \mathbb{E}_\phi(\partial_{\pi^2}\widehat{\ell_{it}\Xi_{it}}\Xi'_{it}).$$

The analytical bias-corrected estimator of δ^0_{NT} is

$$\tilde{\delta}^A = \tilde{\delta} - \hat{B}^\delta/T - \hat{D}^\delta/N,$$

where I use $\tilde{\tilde{\delta}}$, i.e., the APE constructed from a bias corrected estimator of β. Let

$$\hat{\Psi}_{it} \quad = \quad -\frac{1}{NT}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\hat{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}\hat{\gamma}_\tau\hat{\gamma}_t + \hat{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}\hat{\alpha}_j\hat{\gamma}_t + \hat{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\hat{\alpha}_i\hat{\gamma}_\tau + \hat{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\hat{\alpha}_i\hat{\alpha}_j)\widehat{\partial_\pi\Delta_{j\tau}},$$

then the estimated asymptotic biases are

$$\begin{aligned}
\hat{B}^\delta &= \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=0}^{L}[T/(T-j)]\sum_{t=j+1}^{T}\hat{\gamma}_t\hat{\gamma}_\tau\mathbb{E}_\phi(\widehat{\partial_\pi\ell_{i,t-j}}\widehat{\partial_{\pi^2}\ell_{it}}\hat{\Psi}_{it})}{\sum_{t=1}^{T}(\hat{\gamma}_t)^2\mathbb{E}_\phi(\widehat{\partial_{\pi^2}}\ell_{it})} \\
&\quad -\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\hat{\gamma}_t)^2[\mathbb{E}_\phi(\widehat{\partial_{\pi^2}\Delta_{it}}) - \mathbb{E}_\phi(\widehat{\partial_{\pi^3}\ell_{it}})\mathbb{E}_\phi(\widehat{\Psi}_{it})]}{\sum_{t=1}^{T}(\hat{\gamma}_t)^2\mathbb{E}_\phi(\widehat{\partial_{\pi^2}}\ell_{it})} \\
\hat{D}^\delta &= \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\hat{\alpha}_i)^2[\mathbb{E}_\phi(\partial_\pi\widehat{\ell_{it}\partial_{\pi^2}\ell}_{it}\Psi_{it}) - \frac{1}{2}\mathbb{E}_\phi(\widehat{\partial_{\pi^2}\Delta_{it}}) + \frac{1}{2}\mathbb{E}_\phi(\widehat{\partial_{\pi^3}}\ell_{it})\mathbb{E}_\phi(\widehat{\Psi}_{it})]}{\sum_{i=1}^{N}(\hat{\alpha}_i)^2\mathbb{E}_\phi(\widehat{\partial_{\pi^2}}\ell_{it})}].
\end{aligned}$$

The estimator of the asymptotic variance depends on the assumptions about the distribution of the unobserved effects and explanatory variables. Assumption 2(i) requires imposing a homogeneity assumption on the distribution of the explanatory variables to estimate the first term of the asymptotic variance. For example, if $\{X_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ is identically distributed over i, this term is given by

$$\hat{V}^\delta = \frac{1}{NT}\sum_{i=1}^{N}[\sum_{t,\tau=1}^{T}\hat{\tilde{\Delta}}_{it}\hat{\tilde{\Delta}}'_{i\tau} + \sum_{t=1}^{T}\mathbb{E}_\phi(\widehat{\Gamma_{it}\Gamma'_{it}})],$$

for $\hat{\tilde{\Delta}}_{it} = \hat{\Delta}_{it} - N^{-1}\sum_{i=1}^{N}\hat{\Delta}_{it}$. Bias corrected estimators and confidence intervals can be constructed in the same fashion as for the model parameter.

The following theorems show that the analytical bias corrections eliminate the bias from the asymptotic distribution of the fixed effects estimators of the model parameters and APEs without increasing the variance, and that the estimators of the asymptotic variances are consistent. Those are the main results of this section.

Theorem 1.3.3. *(Bias correction for $\hat{\beta}$) Under the conditions of Theorem 1.3.1,*

$$\widehat{W} \xrightarrow{p} \overline{W}_{\infty},$$

and, if $L \to \infty$ and $L/T \to 0$,

$$\sqrt{NT}(\tilde{\beta}^A - \beta^0) \xrightarrow{d} N(0, \overline{W}_{\infty}{}^{-1}).$$

Theorem 1.3.4. *(Bias correction for $\hat{\delta}$) Under the conditions of Theorems 1.3.1 and 1.3.2,*

$$\hat{V}^{\delta} \xrightarrow{p} \overline{V}^{\delta}_{\infty},$$

and, if $L \to \infty$ and $L/T \to 0$,

$$\sqrt{NT}(\tilde{\delta}^A - \delta^0_{NT}) \xrightarrow{d} N(0, \overline{V}^{\delta}_{\infty}).$$

Remark 1.3.3. Split-panel jackknife as described in (Chen et al., 2013; Fernández-Val and Weidner, 2013) can also be applied.

1.4 Discussions and Extensions

1.4.1 Comparison with the existing estimators: No fixed effects or only individual effects

When there are no fixed effects, the model becomes

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{1.14}$$

where all objects are as defined previously. The conditional distribution of Y_{it} is given by

$$P(Y_{it} = 1|X_{it}, \beta) = \Phi(X_{it}\beta),$$

and for estimation the following EM procedure can be used:

Definition 1.4.1. (1) Given initial $\beta^{(k)}$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)}$;

(2) E-step: Calculate $\hat{Y}_{it}^{(k)} := E[Y_{it}^*|Y_{it}, X_{it}, \beta^{(k)}]$;

(3) M-step: The parameter β is updated via

$$\beta^{(k+1)} = (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}')^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\hat{Y}_{it}^{(k)}\}.$$

(4) Iterate until convergence.

I start by comparing this estimation with existing methods.

Proposition 1.4.1. *For panel probit models, the proposed EM-type estimator is asymptotically equivalent to the MLE.*

Proof: See Appendix A.3.1.1. When applying the proposed fixed effects EM-type estimator to probit (or for the general exponential family), its E-step involves calculating the conditional expectation of the error, which is exactly the score of expected, complete data, log-likelihood function or the score of the observed log-likelihood (it also corresponds to the notion of generalized residuals proposed in (Gourieroux et al., 1987) for cross-sectional data). Hence, the fixed effects EM-type estimator directly works with the observed score. For the case when there are no unobserved effects, the EM method is asymptotically equivalent to MLE and there is no asymptotic bias. For the cases when there are unobserved effects, and when there are incidental parameter problems, an iterated bias correction to the score can be easily implemented through the E-step.

Proposition 1.4.2. *For the panel probit model with individual effects, the difference between the proposed fixed effects EM-type estimator and Newton's method lies in whether inverting the Hessian of the observed data log-likelihood function.*

Proof: See Appendix A.3.1.2. I explicitly compare the two iterative steps of the fixed effects EM-type estimator and the Netwon's method. Each iteration of the proposed fixed effects EM-type estimator is a least squares calculation (with the generalized residual); it does not use the inverse of the Hessian of the observed data log-likelihood function like Newton's method.[9]

1.4.2 PPIF with multiple factors

In this setting, the model, written in matrix notation, is

$$Y = \mathbf{1}(X\beta + \alpha\gamma' + \varepsilon \geq 0),$$

where $Y = (Y_1, ..., Y_N)'$ (with $Y_i = (Y_{i1}, ..., Y_{iT})'$, a $T \times 1$ vector) is an $N \times T$ matrix and X (with $X_i = [X_{i1}, ..., X_{iT}]'$ is a $T \times p$ matrix) is a three-dimensional matrix with p sheets $(N \times T \times p)$, the ℓ-th sheet of which is associated with the ℓ-th element of $\beta (\ell = 1, ..., p)$. $\alpha = (\alpha_1, ..., \alpha_N)'$ is an $N \times R$ matrix, while $\gamma = (\gamma_1, ..., \gamma_T)'$ is a $T \times R$ matrix. The product $X\beta$ is an $N \times T$ matrix and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)$ is an $N \times T$ matrix.

[9] See (Greene, 2004) for more about estimation of nonlinear panel data models with individual fixed effects.

Since $\alpha\gamma' = \alpha A^{-1} A\gamma'$ for any $R \times R$ invertible A, identification is not possible without restrictions.

Condition 1. (Normalization) (i) $\gamma'\gamma/T = I_R$; (ii) $\alpha'\alpha =$ diagonal.

Under different normalization conditions, the estimation procedure (the conditional maximization steps) for the factor is different.

Definition 1.4.2. The EM procedure for estimating a panel probit model with multi-dimensional interactive fixed effects under Condition 1 is defined by the following:

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)} + (\alpha_i^{(k)})'\gamma_t^{(k)}$,

(2) E-step: Calculate

$$\begin{aligned}\hat{Y}_{it}^{(k)} : &= E[Y_{it}^*|Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)})/\{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)})\},\end{aligned}$$

(3) M-step: This contains three conditional maximization (CM) steps

CM-step 1: Given α_i and γ_t, the parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^{N} X_i' X_i\right)^{-1} \left\{\sum_{i=1}^{N} X_i'(\hat{Y}_i^{(k)} - \alpha_i^{(k)}\gamma^{(k)})\right\},$$

CM-step 2: Given β and α_i, the parameter γ is updated via

$$\gamma^{(k+1)} = \mathrm{eig}[\frac{1}{NT}\sum_{i=1}^{N}(\hat{Y}_i^{(k)} - X_i\beta^{(k+1)})(\hat{Y}_i^{(k)} - X_i\beta^{(k+1)})'],$$

CM-step 3: Given β and γ_t, the parameter α is updated via

$$\alpha^{(k+1)} = T^{-1}(\hat{Y}^{(k)} - X\beta^{(k+1)})\gamma^{(k+1)},$$

(4) Iterate until convergence.

The CM-step 2 calculates the R largest eigenvector of the matrix in brackets, arranged in decreasing order. It imposes the normalizations of Condition 1 by using eigenvectors. An alternative estimation procedure based on a QR decomposition that does not impose Condition 1(ii) is also proposed below.

Definition 1.4.3. The QR-based decomposition EM procedure for estimating a panel probit model with multi-dimensional interactive fixed effects is defined by the following:

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)} + (\alpha_i^{(k)})'\gamma_t^{(k)}$,

(2) E-step: Calculate

$$\begin{aligned}\hat{Y}_{it}^{(k)} :&= E[Y_{it}^*|Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)})/\{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)})\},\end{aligned}$$

(3) M-step: This contains three conditional maximization (CM) steps

CM-step 1: Given α_i and γ_t, the parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1} \left\{\sum_{i=1}^{N} X_i'(\hat{Y}_i^{(k)} - \alpha_i^{(k)}\gamma^{(k)})\right\},$$

CM-step 2: Given β and α_i, the parameter γ is updated via

$$\gamma^{(k+1)} = (\hat{Y}^{(k)} - X\beta^{(k+1)})'\alpha^{(k)}((\alpha^{(k)})'\alpha^{(k)})^{-1}.$$

Compute the QR decomposition $\gamma^{(k+1)} = \tilde{\gamma}^{(k+1)}R_M$ and replace $\gamma^{(k+1)}$ by $\tilde{\gamma}^{(k+1)}$,

CM-step 3: Given β and $\tilde{\gamma}$, the parameter α is updated via

$$\alpha^{(k+1)} = (\hat{Y}^{(k)} - X\beta^{(k+1)})\tilde{\gamma}^{(k+1)},$$

(4) Iterate until convergence.

Through the iterations, the columns of the updated values of γ are made orthonormal via the QR decomposition (imposing normalization, but other decomposition methods can also be used), i.e., $(\tilde{\gamma}^{(k+1)})'\tilde{\gamma}^{(k+1)}$ is orthonormal (I_R). The QR decomposition is often used to solve the linear least squares problem, and is the basis for a particular eigenvalue algorithm. With additional restrictions, such as a full rank condition on γ and a sign restriction on R_M, the QR decomposition method can achieve unique values of α and γ.

Note that the orthogonalization does not alter the convergence property. Let $\gamma^{(k+1)}$ be the optimizer before orthogonalization. Then $S(\beta, \gamma^{(k+1)}, \alpha^{(k)}) \leq S(\beta, \gamma^{(k)}, \alpha^{(k)})$. Let $\gamma^{(k+1)} = \tilde{\gamma}^{(k+1)} R_M$ be the QR decomposition of $\gamma^{(k+1)}$, and let $\tilde{\alpha}^{(k)} = \alpha^{(k)} R_M'$. Then $\tilde{\alpha}^{(k)}(\tilde{\gamma}^{(k+1)})' = \alpha^{(k)}(\gamma^{(k+1)})'$, so $S(\beta, \tilde{\gamma}^{(k+1)}, \tilde{\alpha}^{(k)}) = S(\beta, \gamma^{(k+1)}, \alpha^{(k)})$, and, consequently, $S(\beta, \tilde{\gamma}^{(k+1)}, \tilde{\alpha}^{(k)}) \leq S(\beta, \gamma^{(k)}, \alpha^{(k)})$.

1.4.2.1 Consistency

In general, the consistency proof contains two steps as shown in the proof for PPIF. The first step involves the consistency of the conditional expectation, and the second checks the assumptions needed for the consistency of the "linearized" model.

Assumption 3. *(Bounded second-order derivative)* $\partial_{\pi^2} \mathcal{L}_{NT}(\beta, \pi) \geq b_{\min}$.

Lemma 1.4.1. *Under Assumption 3 and Assumption 1(i), (ii), and (iv),* $\hat{\beta}_{IF-EM} = \beta^0 + o_p(1)$.

Proof: See Appendix A.3.2.

1.5 Simulations

This section reports evidence on the finite sample behavior of fixed effects estimators in static models with strictly exogenous regressors. This includes several cases: no unobserved effects, individual effects, additive individual and time effects, and interactive individual and time effects. I analyze the performance of the generalized least square (GLS) method using the **R**-package **glm**, which is available on CRAN, and the fixed effects EM-type estimators in terms of bias and inference accuracy based on their asymptotic distribution. I also analyze the performance of the uncorrected and bias-corrected interactive fixed effects EM-type estimators in terms of bias and inference accuracy. In particular, I compute the biases, standard deviations, and root mean squared errors (RMSE) of the estimators, the ratio of averaged standard errors to the simulation standard deviations (SE/SD); and the empirical coverages of confidence intervals with 95% nominal value (p; .95). All results are based on 500 replications.

The data generating processes are:

- DGP-1: $Y_{it} = \mathbf{1}\{X_{it}\beta + \varepsilon_{it} > 0\}, \quad (i = 1, ..., N;\ t = 1, ..., T),$
- DGP-2: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i + \varepsilon_{it} > 0\}, \quad (i = 1, ..., N;\ t = 1, ..., T),$
- DGP-3: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i + \gamma_t + \varepsilon_{it} > 0\}, \quad (i = 1, ..., N;\ t = 1, ..., T),$
- DGP-4: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i\gamma_t + \varepsilon_{it} > 0\}, \quad i = 1, ..., N;\ t = 1, ..., T,$

where $\beta = 1$, $\alpha_i \sim N(0,1)$, $\gamma_t \sim N(0,1)$, and $X_{it} \sim N(0,1)$ are strictly exogenous with respect to ε_{it} with $\varepsilon_{it} \sim N(0,1)$.

Throughout, “No FE” refers to the probit without fixed effects; “FE i” refers to the probit with individual fixed effects; “FE 2” refers to the probit with additive individual and time fixed effects; “IF” refers to the probit with interactive fixed effects; “glm” refers to the GLS estimator in **R**, while “EM” refers to the fixed effects EM-type estimators proposed. For interactive fixed effects, I also implement the bias correction procedure proposed here; “BC-IF” refers to the bias-corrected estimator. All the results are reported in percentages of the true parameter value.

The simulation results are summarized in Table 1.1 for $N = 100$ and $T = 8, 12, 20$, and in Table 1.2 for N=52 and $T = 14, 26, 52$. They show that in all the cases analyzed EM has smaller biases and variances and compares favorably to **glm**. For example, for the case with additive individual and time effects, when $N = 100$ and $T = 12$, the bias for **glm** is 21%, whereas the EM estimator is only 11%. Even for the case without unobserved effects, when $N = 100$ and $T = 20$, the bias for **glm** is 0.52%, whereas the EM estimator is only 0.11%. In terms of RMSE, for the case of individual effects, when $N = 52$ and $T = 14$, the RMSE for **glm** is 16%, whereas for the EM estimator it is 15%. When there is a bias, the results also show that it is of the same order of magnitude as the standard deviation for the uncorrected EM and **glm** estimator, and this causes severe undercoverage of the confidence intervals. The analytical bias correction removes the bias without increasing dispersion and produces substantial improvements in terms of RMSE and coverage probabilities. For example, the analytical bias correction reduces the RMSE by more than 4% and increases coverage by around 20% in the $N = 100$ and $T = 12$ case.

1.6 Empirical example

1.6.1 A gravity equation and the extensive margins of trade

Understanding how different trade barriers influence trade flows is key when one wants to study the impact of distance, trade agreements, and other trade frictions. See (Helpman et al., 2008; Bernard et al., 2007; Charbonneau, 2012). For my application, I use the same data set as in (Helpman et al., 2008), which consists of information on who trades with whom for a large set of countries.

I illustrate the estimation and difference when including differing degrees of fixed effects, namely the cases with no fixed effects, only individual fixed effects, additive individual and time fixed effects, and interactive fixed effects. The fixed effects are importer and exporter fixed effects for a single year, the year 1986. I obtain a balanced panel of 158 countries that account for the majority of world trade. The probability of country j exporting to country i is

$$Prob[Trade_{ij} = 1 | X_{ij}, g(\alpha_i, \gamma_j)] = \Phi(X_{ij}'\beta + g(\alpha_i, \gamma_j)).$$

Here X_{ij} contains D_{ij}, representing the distance between country i's and country j's most populated cities; $Border_{ij}$, a dummy that takes the value 1 if i and j share a border; $Legal_{ij}$, a dummy that takes the value 1 if the two countries have the same legal system; $Language_{ij}$, a dummy that takes the value 1 if i and j have the same official language; $Colony_{ij}$, a dummy that takes the value 1 if i and j were ever in a colonial relationship; $Currency_{ij}$, a dummy that takes the value of 1 if the two countries use the same currency; RTA_{ij}, a dummy that takes the value 1 if i and j are in a regional trade agreement; and, finally, α_i and γ_j, respectively representing importer and exporter fixed effects.

The results of the effects of trade barriers are summarized in Table 1.3. After accounting for exporter fixed effects the effect of a common currency decreases in magnitude from about -0.45 to -0.16. This suggests that excluding exporter effects may overstate the decrease in the likelihood of trade when trading partners share a common currency. The changes of magnitude on language and region suggest that excluding exporter effects may understate the importance of having the same language and the same religion. Similarly, the magnitude changes of distance, from about -0.19 to -0.29, suggesting that excluding exporter effects may understate the importance of distance. Importantly, the magnitude of the coefficient for border changes from 0.16 to -0.03 suggests overstating the importance of sharing a border. Note also that the effect of free trade agreements is rather robust to the inclusion or complete omission of fixed effects. This suggests that perhaps the effect of a free trade agreement on the likelihood of trade between a pair of countries does not depend on the exact trade network of those countries; FTAs appear to increase the likelihood of trade regardless of which fixed effects are included.

1.7 Conclusion

This paper presents an EM type method of estimating nonlinear panel data models with multiple unobserved effects, allowing for interactions between the unobserved individual and time specific effects. The method can be applied to models with individual effects, additive individual and time effects, interactive effects and other general functional form of unobserved effects. In finite-sample simulations, the method outperform the existing generalized least square methods for the models with individual effects and additive individual and time effects in terms of both bias and variance. Furthermore, I derive the asymptotic distribution of the proposed EM estimator for the panel probit model with interactive fixed effects. Analytical bias corrections are developed to deal with the incidental parameter problem for both the estimates of the coefficients and its associated average partial effects. Simulations demonstrate the correction works well in reducing the bias and root mean squared error and improves coverage rates. Finally for purpose of illustration, I use the example of international trade networks demonstrating that misspecifying the fixed effects model can over or understate the importance of certain factors on the likelihood of trade. A wide range of future theoretical and empirical work can build upon the results of this paper. For example, sample selection models with interactive effects or models with strategic interactions, such as binary game models, could benefit from and build on the approach proposed here.

Table 1.1: Finite Sample Properties of Static Probit Estimators, N=100

Model	Estimator	Bias	Std.Dev.	RMSE	SE/SD	P;.95
			T=8			
No FE	EM	0.26	7.48	7.49	1.03	0.97
	glm	0.69	7.59	7.61	1.02	0.96
FE i	EM	20.74	10.37	23.18	0.73	0.29
	glm	22.38	11.73	25.26	0.85	0.39
Add-FE	EM	20.73	9.24	22.69	0.86	0.28
	glm	29.21	13.95	32.36	0.83	0.32
IF		8.95	10.08	13.47	0.72	0.69
	BC-IF	-4.69	8.91	10.06	0.81	0.84
			T=12			
No FE	EM	-0.10	6.01	6.02	1.04	0.96
	glm	0.31	6.09	6.09	1.03	0.96
FE i	EM	12.53	7.61	14.65	0.79	0.45
	glm	13.43	8.11	15.68	0.89	0.53
Add-FE	EM	10.88	6.62	12.73	0.99	0.64
	glm	20.81	10.20	23.17	0.89	0.38
IF		7.64	6.94	10.32	0.83	0.73
	BC-IF	-0.45	6.42	6.43	0.9	0.92
			T=20			
No FE	EM	0.11	4.93	4.94	0.98	0.94
	glm	0.52	5.00	5.02	0.97	0.95
FE i	EM	6.44	5.22	8.28	0.85	0.67
	glm	7.20	5.50	9.06	0.95	0.70
Add-FE	EM	3.56	4.60	5.82	1.02	0.89
	glm	10.88	6.57	12.71	0.93	0.60
IF		4.03	4.86	6.31	0.90	0.83
	BC-IF	-0.99	4.62	4.72	0.95	0.94

Notes: All the entries are in percentage of the true parameter value. 500 replications.

Table 1.2: Finite Sample Properties of Static Probit Estimators, N=52

Model	Estimator	Bias	Std.Dev.	RMSE	SE/SD	P;.95
			T=14			
No FE	EM	-0.02	7.83	7.84	1.03	0.94
	glm	0.43	7.97	7.98	1.01	0.95
FE i	EM	11.3	9.55	14.79	0.81	0.68
	glm	12.47	10.53	16.31	0.9	0.77
Add-FE	EM	2.92	7.74	8.27	1.02	0.94
	glm	24.05	15.28	28.48	0.8	0.53
IF		4.8	9.28	10.44	0.79	0.83
	BC-IF	-3.56	8.52	9.22	0.86	0.87
			T=26			
No FE	EM	-0.13	5.92	5.92	0.99	0.94
	glm	0.27	5.99	5.99	0.99	0.94
FE i	EM	4.88	6	7.73	0.88	0.85
	glm	5.33	6.21	8.17	0.98	0.89
Add-FE	EM	0.53	5.63	5.65	1	0.95
	glm	10.94	8.08	13.59	0.93	0.7
IF		3.43	6.28	7.16	0.85	0.87
	BC-IF	-1.3	5.96	6.09	0.9	0.92
			T=52			
No FE	EM	-0.18	4.22	4.22	0.98	0.95
	glm	0.22	4.27	4.27	0.98	0.95
FE i	EM	2.2	4.07	4.62	0.91	0.89
	glm	2.48	4.2	4.88	1	0.92
Add-FE	EM	1.21	3.97	4.15	1	0.94
	glm	6.99	5.17	8.69	0.96	0.71
IF		1.5	3.91	4.18	0.96	0.91
	BC-IF	-1.48	3.78	4.05	0.99	0.94

Notes: All the entries are in percentage of the true parameter value. 500 replications.

Table 1.3: Coefficients of Static Probit Model for Trade

	(1)	(2)	(3)	(4)
Distance	-0.185	-0.177	-0.294	-0.297
Border	0.161	0.152	-0.027	-0.041
Island	-0.175	-0.178	-0.153	-0.16
Landlock	-0.357	-0.358	-0.471	-0.474
Legal	-0.308	-0.309	-0.208	-0.212
Language	0.08	0.079	0.166	0.173
Colony	2.222	2.245	2.06	1.962
Currency	-0.446	-0.449	-0.158	-0.19
FTA	1.685	1.629	1.645	1.648
Religion	0.2	0.191	0.367	0.36
Importer effects		Yes	Yes	Yes
Exporter effects			Yes	Yes
Interactive				Yes

Chapter 2

Nonlinear Panel Models with Interactive Effects[1]

2.1 Introduction

Panel data models are useful to identify causal effects because they allow the researcher to control for multiple sources of unobserved heterogeneity modeled as individual and time effects. The general idea is to use variation across time to control for unobserved time invariant individual effects and to use contemporaneous variation across individuals to control for aggregate time effects. We consider estimation and inference on semiparametric nonlinear panel models with predetermined explanatory variables and interactive individual and time effects. We focus on single index models, which cover static and dynamic probit, logit, and Poisson models. We adopt a fixed effects approach that treats the realizations of the unobserved individual and time effects as parameters to be estimated, and therefore does not impose any restriction on the relationship between these effects and the observable explanatory variables. Fixed effects estimation in nonlinear models with interactive effects, however, is computationally challenging and suffers from the incidental parameter problem.

Maximum likelihood estimation of standard single index models with cross section data is computationally tractable because the likelihood function is concave in all the model parameters. This computational tractability is preserved in panel models with additive individual and time effects, but it breaks down in panel models with interactive effects because the index is no longer linear in the individual and time effects. Moreover, the principal components algorithm proposed by (Bai, 2009b) for linear models with interactive effects

[1]This chapter is based on a joint work with Iván Fernández-Val and Martin Weidner

cannot be applied to nonlinear models. We deal with this challenge by proposing an iterative two-step algorithm to compute the fixed effects conditional maximum likelihood estimator (FE-CMLE), where each step solves a concave optimization program. The algorithm is based on the observation that the likelihood program is concave on the individual effects after fixing the time effects and vice versa. We show that the algorithm converges to a local maximum, as the likelihood function decreases at each step of the algorithm. In a simple model where the FE-CMLE can be obtained by principal components methods, the iterative algorithm finds the same estimates as principal components up to numerical tolerance error.

We characterize the asymptotic properties of the FE-CMLE under sequences where the cross section (N) and time series (T) dimensions of the panel pass to infinity at the same rate. We give conditions for consistency of the estimators of the index coefficients. Consistency is hard to establish in this setting because the dimension of the parameter space grows with the sample size and we cannot resort to concavity, unlike in models with additive individual and time effects. While consistent, the FE-CMLE has a bias in the asymptotic distribution of the same order as the variance. This is the large-T version of the well-known incidental parameter problem (Neyman and Scott, 1948), where the bias arises from the large number of estimated parameters and the nonlinearity of the model. We characterize the first order bias, and propose analytical and jackknife corrections that remove the bias from the asymptotic distribution. Asymptotically the correction does not increase variance and the confidence intervals constructed around the corrected estimator have correct coverage. We also derive asymptotic theory for fixed effects estimators of average partial effects (APEs). These APEs are often the quantities of interest in nonlinear models and are functions of the data, index coefficients and unobserved individual and time effects. As (Fernández-Val and Weidner, 2013), we find that in general the incidental parameter bias is asymptotically of second order because the estimators of the APEs have slower rate of convergence than the estimators of the index coefficients. In numerical simulations, we show that the asymptotic results provide a good approximation to the behavior of the FE-CMLE and the bias corrections perform well in finite samples for multiple values of N and T.

Literature review: (Neyman and Scott, 1948), (Heckman, 1981), (Lancaster, 2000), and (Greene, 2004) discussed the incidental parameter problem in panel data models. (Phillips and Moon, 1999), (Hahn and Kuersteiner, 2002), (Lancaster, 2002), (Woutersen, 2001), (Hahn and Newey, 2004), (Carro, 2007), (Arellano and Bonhomme, 2009), (Fernández-Val, 2009), (Hahn and Kuersteiner, 2011), (Fernández-Val and Vella, 2011), and (Kato et al., 2012) proposed large-T bias corrections for fixed effects estimators in linear and nonlinear panel models with additive individual effects; see also (Arellano and Hahn, 2007) for a recent survey on this literature. (Bai, 2009b) and (Moon and Weidner, 2010a; Moon and Weidner, 2010b) considered large-T bias corrections for FE-CMLE estimators of linear models with interactive individual and time effects. (Charbonneau, 2012) and (Fernández-Val and Weidner, 2013)considered fixed effects estimation of nonlinear panel models with additive individual and time effects.

In Section 2.2, we introduce the model and fixed effects estimators. Section 2.3 describes the bias corrections to deal with the incidental parameters problem and illustrates how the bias corrections work through an example. Section 2.4 provides the asymptotic theory. Section 2.5 gives numerical examples. We collect the proofs of all the results and additional technical details in the Appendix.

2.2 Model and Estimators

2.2.1 Model

The data consist of $N \times T$ observations $\{(Y_{it}, X_{it}')' : 1 \leq i \leq N, 1 \leq t \leq T\}$, for a scalar outcome variable of interest Y_{it} and a vector of explanatory variables X_{it}. We assume that the outcome for individual i at time t is generated by the sequential conditionally independent process:

$$Y_{it} \mid X_i^t, \alpha, \gamma, \beta \sim f_Y(\cdot \mid X_{it}'\beta + \alpha_i \gamma_t), \quad (i = 1, ..., N; t = 1, ..., T),$$

where $X_i^t = (X_{i1}, \dots, X_{it})$, $\alpha = (\alpha_1, \dots, \alpha_N)$, $\gamma = (\gamma_1, \dots, \gamma_T)$, f_Y is a known probability function, and β is a finite dimensional parameter vector.

The variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively. The model is semiparametric because we do not specify the distribution of these effects nor their relationship with the explanatory variables. The conditional distribution f_Y represents the parametric part of the model. The vector X_{it} contains predetermined variables with respect to Y_{it}. Note that X_{it} can include lags of Y_{it} to accommodate dynamic models. The model is a single index model because the explanatory variables and unobserved effects enter f_Y through the index $z_{it} := X_{it}'\beta + \alpha_i \gamma_t$ and is interactive because the individual and time effects enter the index z_{it} multiplicatively as $\alpha_i \gamma_t = \alpha_i \times \gamma_t$.

We consider three running examples throughout the analysis:

Example 2.2.1. [Linear model] Let Y_{it} be a continuous outcome. We can model the conditional distribution of Y_{it} using the Gaussian linear model

$$f_Y(y \mid X_{it}'\beta + \alpha_i \gamma_t) = \varphi((X_{it}'\beta + \alpha_i \gamma_t)/\sigma)/\sigma, \quad y \in \mathbb{R},$$

where φ is the density function of the standard normal and σ is a positive scale parameter.

Example 2.2.2. [Binary response model] Let Y_{it} be a binary outcome and F be a cumulative distribution function of the standard normal or logistic distribution. We can model the conditional distribution of Y_{it} using the probit or logit model

$$f_Y(y \mid X_{it}'\beta + \alpha_i \gamma_t) = F(X_{it}'\beta + \alpha_i \gamma_t)^y [1 - F(X_{it}'\beta + \alpha_i \gamma_t)]^{1-y}, \quad y \in \{0, 1\}.$$

Example 2.2.3. [Count response model] Let Y_{it} be a non-negative interger-valued outcome, and $f(\cdot;\lambda)$ be the probability mass function of a Poisson random variable with mean $\lambda > 0$. We can model the conditional distribution of Y_{it} using the Poisson model

$$f_Y(y \mid X_{it}'\beta + \alpha_i\gamma_t) = f(y; \exp[X_{it}'\beta + \alpha_i\gamma_t]), \quad y \in \{0, 1, 2, ...\}.$$

For estimation, we adopt a fixed effects approach treating the realization of the unobserved individual and time effects as parameters to be estimated. We collect all these effects in the vector $\phi_{NT} = (\alpha_1, ..., \alpha_N, \gamma_1, ..., \gamma_T)'$. The model parameter β includes the index coefficients of interest, while the unobserved effects ϕ_{NT} are treated as a nuisance parameter. The true values of the parameters, denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0, \gamma_1^0, ..., \gamma_T^0)'$, are the solution to the population fixed effects conditional maximum likelihood program

$$\max_{(\beta,\phi_{NT})\in\mathbb{R}^{\dim\beta+\dim\phi_{NT}}} \mathbb{E}_\phi[\mathcal{L}_{NT}(\beta,\phi_{NT})],$$
$$\mathcal{L}_{NT}(\beta,\phi_{NT}) := (NT)^{-1/2}\sum_{i,t}\log f_Y(Y_{it} \mid X_{it}'\beta + \alpha_i\gamma_t), \tag{2.1}$$

for every N, T, where $\mathbb{E}_\phi$ denotes the expectation with respect to the distribution of the data conditional on the unobserved effects and initial conditions including strictly exogenous variables. We need to impose a scale normalization on ϕ_{NT}^0 because multiplying by a constant to all α_i, while dividing by same constant to all γ_t, does not change $\alpha_i\gamma_t$. We normalize ϕ_{NT}^0 to satisfy $\sum_i[\alpha_i^0]^2 = \sum_t[\gamma_t^0]^2$. Existence and uniqueness of the solution to the population problem up to the scale normalization will be guaranteed by our assumptions in Section 2.4 below, including concavity of the objective function in the index $X_{it}'\beta + \alpha_i\gamma_t$.

The pre-factor $(NT)^{-1/2}$ in $\mathcal{L}_{NT}(\beta,\phi_{NT})$ is just a convenient rescaling for the asymptotic analysis.

Other quantities of interest involve averages over the data and unobserved effects

$$\delta^0_{NT} = \mathbb{E}[\Delta_{NT}(\beta^0, \phi^0_{NT})], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(Y_{it}, X_{it}, \beta, \alpha_i \gamma_t), \tag{2.2}$$

where $\mathbb{E}$ denotes the expectation with respect to the joint distribution of the data and the unobserved effects, provided that the expectation exists. They are indexed by N and T because the marginal distribution of $\{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$ can be heterogeneous across i and/or t; see Section 2.4.2. These averages include scale parameters and other average partial effects (APEs), which are often the ultimate quantities of interest in nonlinear models. Some examples of partial effects are the following:

Example 2.2.1 (Linear model). *The variance σ^2 in the linear model can be expressed as an APE with*

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i \gamma_t) = (Y_{it} - X_{it}'\beta - \alpha_i \gamma_t)^2. \tag{2.3}$$

Example 2.2.2 (Binary response model). *If $X_{it,k}$, the kth element of X_{it}, is binary, its partial effect on the conditional probability of Y_{it} is*

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i \gamma_t) = F(\beta_k + X_{it,-k}'\beta_{-k} + \alpha_i \gamma_t) - F(X_{it,-k}'\beta_{-k} + \alpha_i \gamma_t), \tag{2.4}$$

where β_k is the kth element of β, and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the kth element. If $X_{it,k}$ is continuous and F is differentiable, the partial effect of $X_{it,k}$ on the conditional probability of Y_{it} is

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i \gamma_t) = \beta_k \partial F(X_{it}'\beta + \alpha_i \gamma_t), \tag{2.5}$$

where ∂F is the derivative of F.

Example 2.2.3 (Count response model). *If $X_{it,k}$, the kth element of X_{it}, is binary, its* *partial effect on the conditional probability of* Y_{it} *is*

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i\gamma_t) = \exp(\beta_k + X'_{it,-k}\beta_{-k} + \alpha_i\gamma_t) - \exp(X'_{it,-k}\beta_{-k} + \alpha_i\gamma_t), \tag{2.6}$$

where β_k *is the kth element of* β*, and* $X_{it,-k}$ *and* β_{-k} *include all elements of* X_{it} *and* β *except for the kth element. If* $X_{it,k}$ *is continuous, the partial effect of* $X_{it,k}$ *on the conditional expectation of* Y_{it} *is*

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i\gamma_t) = \beta_k \exp(X'_{it}\beta + \alpha_i\gamma_t). \tag{2.7}$$

2.2.2 Fixed effects estimators

The sample analog of the program (2.1) is

$$\max_{(\beta,\phi_{NT})\in\mathbb{R}^{\dim\beta+\dim\phi_{NT}}} \mathcal{L}_{NT}(\beta, \phi_{NT}). \tag{2.8}$$

As in the population case, we shall impose conditions guaranteeing that the solutions to the previous programs exist and are unique with probability approaching one as N and T become large, including the scale normalization on ϕ_{NT}. The program (2.8) cannot be solved using standard optimization algorithms because it is not concave in ϕ_{NT} due to the multiplicative structure. We propose an iterative two-step algorithm for the case where the log-likelihood is concave in the index z_{it}, where each step solves a concave maximization program. The algorithm is based on the observation that the log-likelihood program is concave on the individual effects after fixing the time effects and vice versa. To describe the algorithm it is convenient to separate $\phi_{NT} = (\alpha, \gamma)$, so that $\mathcal{L}_{NT}(\beta, \phi_{NT}) = \mathcal{L}_{NT}(\beta, \alpha, \gamma)$.

Algorithm 2.2.1 (IFE-CMLE). *1. Iteration 0: find initial values* $(\widehat{\beta}^{(0)}, \widehat{\alpha}^{(0)}, \widehat{\gamma}^{(0)})$ *solv-*

ing

$$\widehat{\gamma}^{(0)} \in \arg \max_{(\beta,\gamma)\in\mathbb{R}^{\dim\beta+T}} \mathcal{L}_{NT}(\beta, 1_N, \gamma), \quad (\widehat{\beta}^{(0)}, \widehat{\alpha}^{(0)}) \in \arg \max_{(\beta,\alpha)\in\mathbb{R}^{\dim\beta+N}} \mathcal{L}_{NT}(\beta, \alpha, \widehat{\gamma}^{(0)}),$$

where 1_N is a N-vector of ones.

2. *Iteration k: update $(\widehat{\beta}^{(k-1)}, \widehat{\alpha}^{(k-1)}, \widehat{\gamma}^{(k-1)})$ in two steps solving*

   (a) *Step 1:* $\widehat{\gamma}^{(k)} \in \arg\max_{\gamma\in\mathbb{R}^T} \mathcal{L}_{NT}(\widehat{\beta}^{(k-1)}, \widehat{\alpha}^{(k-1)}, \gamma)$,

   (b) *Step 2:* $(\widehat{\beta}^{(k)}, \widehat{\alpha}^{(k)}) \in \arg\max_{(\beta,\alpha)\in\mathbb{R}^{\dim\beta+N}} \mathcal{L}_{NT}(\beta, \alpha, \widehat{\gamma}^{(k)})$.

3. *Repeat 2 until convergence in m iterations, e.g. when*

$$\mathcal{L}_{NT}(\widehat{\beta}^{(m)}, \widehat{\alpha}^{(m)}, \widehat{\gamma}^{(m)}) - \mathcal{L}_{NT}(\widehat{\beta}^{(m-1)}, \widehat{\alpha}^{(m-1)}, \widehat{\gamma}^{(m-1)}) < \epsilon_{tol},$$

where ϵ_{tol} is a tolerance level (e.g., 10^{-4}).

4. *Final iteration: define the IFE-CMLE as*

$$\widehat{\beta}_{NT} = \widehat{\beta}^{(m)}, \quad \widehat{\phi}_{NT} = (c\widehat{\alpha}^{(m)}, \widehat{\gamma}^{(m)}/c),$$

where $c^4 = \widehat{\gamma}^{(m)\prime}\widehat{\gamma}^{(m)}/\widehat{\alpha}^{(m)\prime}\widehat{\alpha}^{(m)}$. The rescaling by c imposes the scale normalization $\sum_i \widehat{\alpha}_i^2 = \sum_t \widehat{\gamma}_t^2$ in $\widehat{\phi}_{NT}$.

Remark 2.2.1. [Convergence of IFE-MLE] If $z_{it} \mapsto \log f_Y(Y_{it} \mid z_{it})$ is concave, then the objective functions in each step $\gamma \mapsto \mathcal{L}_{NT}(\beta, \alpha, \gamma)$ and $(\beta, \alpha) \mapsto \mathcal{L}_{NT}(\beta, \alpha, \gamma)$ are also concave. Moreover, in view of the fact

$$\mathcal{L}_{NT}(\widehat{\beta}^{(k-1)}, \widehat{\alpha}^{(k-1)}, \widehat{\gamma}^{(k-1)}) \leq \mathcal{L}_{NT}(\widehat{\beta}^{(k-1)}, \widehat{\alpha}^{(k-1)}, \widehat{\gamma}^{(k)}) \leq \mathcal{L}_{NT}(\widehat{\beta}^{(k)}, \widehat{\alpha}^{(k)}, \widehat{\gamma}^{(k)}),$$

the convergence of the algorithm to a local maximum of the program (2.8) is guaranteed. We find that the speed of convergence is fast in simulations.

To analyze the statistical properties of the estimator of β it is conceptually convenient to solve the program (2.8) in two steps. First, we concentrate out the nuisance parameter ϕ_{NT}. For given β, we define the optimal $\widehat{\phi}_{NT}(\beta)$ as

$$\widehat{\phi}_{NT}(\beta) = \arg \max_{\phi_{NT} \in \mathbb{R}^{\dim \phi_{NT}}} \mathcal{L}_{NT}(\beta, \phi_{NT}) \,. \tag{2.9}$$

The fixed effects estimators of β^0 and ϕ^0_{NT} are then

$$\widehat{\beta}_{NT} \;=\; \arg \max_{\beta \in \mathbb{R}^{\dim \beta}} \mathcal{L}_{NT}(\beta, \widehat{\phi}_{NT}(\beta)) \,, \qquad \widehat{\phi}_{NT} \;=\; \widehat{\phi}_{NT}(\widehat{\beta}). \tag{2.10}$$

Estimators of APEs can be formed by plugging-in the estimators of the model parameters in the sample version of (2.2), i.e.

$$\widehat{\delta}_{NT} = \Delta_{NT}(\widehat{\beta}, \widehat{\phi}_{NT}). \tag{2.11}$$

2.3 Incidental parameter problem and bias corrections

In this section we give a heuristic discussion of the main results, leaving the technical details to Section 2.4.

2.3.1 Incidental parameter problem

Fixed effects estimators in nonlinear or dynamic models suffer from the incidental parameter problem ((Neyman and Scott, 1948)). The individual and time effects are incidental parameters that cause the estimators of the model parameters to be inconsistent under asymptotic sequences where either N or T are fixed. To describe the problem let

$$\overline{\beta}_{NT} := \arg \max_{\beta \in \mathbb{R}^{\dim \beta}} \mathbb{E}_{\phi} \left[\mathcal{L}_{NT}(\beta, \widehat{\phi}_{NT}(\beta)) \right]. \tag{2.12}$$

In general, $\text{plim}_{N\to\infty}\overline{\beta}_{NT} \neq \beta^0$ and $\text{plim}_{T\to\infty}\overline{\beta}_{NT} \neq \beta^0$ because of the estimation error in $\widehat{\phi}_{NT}(\beta)$ when one of the dimensions is fixed. If $\widehat{\phi}_{NT}(\beta)$ is replaced by $\phi_{NT}(\beta) = \arg\max_{\phi_{NT}\in\mathbb{R}^{\dim\phi_{NT}}} \mathbb{E}_\phi[\mathcal{L}_{NT}(\beta, \phi_{NT})]$, then the resulting $\overline{\beta}_{NT} = \beta^0$. We consider analytical and jackknife corrections for the bias $\overline{\beta}_{NT} - \beta^0$.

2.3.2 Bias Corrections

Some expansions can be used to explain our corrections. Under suitable sampling conditions, the bias is small for large enough N and T, i.e., $\text{plim}_{N,T\to\infty}\overline{\beta}_{NT} = \beta^0$. For smooth likelihoods and under appropriate regularity conditions, as $N, T \to \infty$,

$$\overline{\beta}_{NT} = \beta^0 + \overline{B}^{\beta}_{\infty}/T + \overline{D}^{\beta}_{\infty}/N + o_P(T^{-1} \vee N^{-1}), \tag{2.13}$$

for some $\overline{B}^{\beta}_{\infty}$ and $\overline{D}^{\beta}_{\infty}$ that we characterize in Theorem 2.4.1, where $a \vee b := \max(a, b)$. Unlike in nonlinear models without incidental parameters, the order of the bias is higher than the inverse of the sample size $(NT)^{-1}$ due to the slow rate of convergence of $\widehat{\phi}_{NT}$. Note also that by the properties of the maximum likelihood estimator

$$\sqrt{NT}(\widehat{\beta}_{NT} - \overline{\beta}_{NT}) \to_d \mathcal{N}(0, \overline{V}_{\infty}).$$

Under asymptotic sequences where $N/T \to \kappa^2$ as $N, T \to \infty$, the fixed effects estimator is asymptotically biased because

$$\begin{aligned}\sqrt{NT}(\widehat{\beta}_{NT} - \beta^0) &= \sqrt{NT}(\widehat{\beta}_{NT} - \overline{\beta}_{NT}) + \sqrt{NT}(\overline{B}^{\beta}_{\infty}/T + \overline{D}^{\beta}_{\infty}/N + o_P(T^{-1} \vee N^{-1})) \\ &\to_d \mathcal{N}(\kappa\overline{B}^{\beta}_{\infty} + \kappa^{-1}\overline{D}^{\beta}_{\infty}, \overline{V}_{\infty}).\end{aligned} \tag{2.14}$$

This is the large-N large-T version of the incidental parameters problem that invalidates any inference based on the asymptotic distribution. Relative to fixed effects estimators with only individual effects, the presence of time effects introduces additional asymptotic bias through $\overline{D}^{\beta}_{\infty}$.

The analytical bias correction consists of removing estimates of the leading terms of the bias from the fixed effect estimator of β^0. Let $\widehat{B}^{\beta}_{NT}$ and $\widehat{D}^{\beta}_{NT}$ be estimators of $\overline{B}^{\beta}_{\infty}$ and $\overline{D}^{\beta}_{\infty}$, respectively. The bias corrected estimator can be formed as

$$\widetilde{\beta}^{A}_{NT} = \widehat{\beta}_{NT} - \widehat{B}^{\beta}_{NT}/T - \widehat{D}^{\beta}_{NT}/N.$$

If $N/T \to \kappa^2$, $\widehat{B}^{\beta}_{NT} \to_P \overline{B}^{\beta}_{\infty}$, and $\widehat{D}^{\beta}_{NT} \to_P \overline{D}^{\beta}_{\infty}$, then

$$\sqrt{NT}(\widetilde{\beta}^{A}_{NT} - \beta^0) \to_d \mathcal{N}(0, \overline{V}_{\infty}).$$

The analytical correction therefore centers the asymptotic distribution at the true value of the parameter, without increasing asymptotic variance.

We consider a jackknife bias correction method that does not require explicit estimation of the bias, but is computationally more intensive. This method is the double split panel jackknife (SPJ) correction of (Fernández-Val and Weidner, 2013), which extended the jackknife correction of (Dhaene and Jochmans, 2010) to models with additive individual and time effects. Alternative jackknife corrections based on the leave-one-observation-out panel jackknife (PJ) of (Hahn and Newey, 2004) and combinations of PJ and SPJ are also possible. We do not consider corrections based on PJ because they are theoretically justified by second-order expansions of $\overline{\beta}_{NT}$ that are beyond the scope of this paper.

To describe the double SPJ correction, let $\widetilde{\beta}_{N,T/2}$ be the average of the 2 split jackknife estimators that leave out the first and second halves of the time periods, and let $\widetilde{\beta}_{N/2,T}$ be the average of the 2 split jackknife estimators that leave out half of the individuals.[2] In choosing the cross sectional division of the panel, one might want to take into account individual clustering structures to preserve and account for cross sectional dependencies. If

[2]When T is odd we define $\widetilde{\beta}_{N,T/2}$ as the average of the 2 split jackknife estimators that use overlapping subpanels with $t \leq (T+1)/2$ and $t \geq (T+1)/2$. We define $\widetilde{\beta}_{N/2,T}$ similarly when N is odd.

there are no cross sectional dependencies, $\widetilde{\beta}_{N/2,T}$ can be constructed as the average of the estimators obtained from all possible partitions of $N/2$ individuals to avoid ambiguity and arbitrariness in the choice of the division.[3] The bias corrected estimator is

$$\widetilde{\beta}^{J}_{NT} = 3\widehat{\beta}_{NT} - \widetilde{\beta}_{N,T/2} - \widetilde{\beta}_{N/2,T}. \tag{2.15}$$

To give some intuition about how the corrections works, note that

$$\widetilde{\beta}^{J}_{NT} - \beta_0 = (\widehat{\beta}_{NT} - \beta_0) - (\widetilde{\beta}_{N,T/2} - \widehat{\beta}_{NT}) - (\widetilde{\beta}_{N/2,T} - \widehat{\beta}_{NT}),$$

where $\widetilde{\beta}_{N,T/2} - \widehat{\beta}_{NT} = \overline{B}^{\beta}_{\infty}/T + o_P(T^{-1} \vee N^{-1})$ and $\widetilde{\beta}_{N/2,T} - \widehat{\beta}_{NT} = \overline{D}^{\beta}_{\infty}/N + o_P(T^{-1} \vee N^{-1})$. The time series split removes the bias term $\overline{B}^{\beta}_{\infty}$ and the cross sectional split removes the bias term $\overline{D}^{\beta}_{\infty}$.

2.4 Asymptotic Theory for Bias Corrections

In nonlinear panel data models the population problem (2.12) generally does not have closed form solution, so we need to rely on asymptotic arguments to characterize the terms in the expansion of the bias (2.13) and to justify the validity of the corrections.

2.4.1 Asymptotic distribution of model parameters

We consider single index panel models with predetermined explanatory variables and scalar interactive individual and time effects that enter the likelihood function through $z_{it} = X'_{it}\beta + \alpha_i \gamma_t$. In these models the dimension of the incidental parameters is $\dim \phi_{NT} = N + T$. These models cover the linear, probit and Poisson specifications of Examples 2.2.1–2.2.3.

[3] There are $P = \binom{N}{2}$ different cross sectional partitions with $N/2$ individuals. When N is large, we can approximate the average over all possible partitions by the average over $S \ll P$ randomly chosen partitions to speed up computation.

The parametric part of our panel models takes the form

$$\log f_Y(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta) = \ell_{it}(z_{it}). \tag{2.16}$$

We denote the derivatives of the log-likelihood function ℓ_{it} by $\partial_{z^q}\ell_{it}(z) := \partial^q \ell_{it}(z)/\partial z^q$, $q = 1, 2, \ldots$ We drop the argument z_{it} when the derivatives are evaluated at the true value of the index $z_{it}^0 := X_{it}'\beta^0 + \alpha_i^0\gamma_t^0$, i.e., $\partial_{z^q}\ell_{it} := \partial_{z^q}\ell_{it}(z_{it}^0)$. We also drop the dependence on NT from all the sequences of functions and parameters, e.g. we use $\mathcal{L}$ for $\mathcal{L}_{NT}$ and ϕ for ϕ_{NT}.

We make the following assumptions:

Assumption 4. *[Panel models] Let $\nu > 0$ and $\mu > 4(8+\nu)/\nu$. Let $\varepsilon > 0$ and let $\mathcal{B}_\varepsilon^0$ be a bounded subset of $\mathbb{R}$ that contains an ε-neighbourhood of z_{it}^0 for all i, t, N, T.*

- *(i) Asymptotics: we consider limits of sequences where $N/T \to \kappa^2$, $0 < \kappa < \infty$, as $N, T \to \infty$.*
- *(ii) Sampling: conditional on ϕ, $\{(Y_i^T, X_i^T) : 1 \leq i \leq N\}$ is independent across i and, for each i, $\{(Y_{it}, X_{it}) : 1 \leq t \leq T\}$ is α-mixing with mixing coefficients satisfying $\sup_i a_i(m) = \mathcal{O}(m^{-\mu})$ as $m \to \infty$, where*

$$a_i(m) := \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+m}^i} |P(A \cap B) - P(A)P(B)|,$$

  and for $Z_{it} = (Y_{it}, X_{it})$, $\mathcal{A}_t^i$ is the sigma field generated by $(Z_{it}, Z_{i,t-1}, \ldots)$, and $\mathcal{B}_t^i$ is the sigma field generated by $(Z_{it}, Z_{i,t+1}, \ldots)$.

- *(iii) Model: for $X_i^t = \{X_{is} : s = 1, ..., t\}$, we assume that for all i, t, N, T,*

$$Y_{it} \mid X_i^t, \phi, \beta \sim \exp[\ell_{it}(X_{it}'\beta + \alpha_i\gamma_t)].$$

  The realizations of the parameters and unobserved effects that generate the observed *data are denoted by* β^0 *and* ϕ^0*.*

- *(iv) Smoothness and moments: We assume that* $z \mapsto \ell_{it}(z)$ *is four times continuously differentiable over* $\mathcal{B}_\varepsilon^0$ *a.s. The partial derivatives of* $\ell_{it}(z)$ *with respect to* z *up to fourth order are bounded in absolute value uniformly over* $z \in \mathcal{B}_\varepsilon^0$ *by a function* $M(Z_{it}) > 0$ *a.s., and* $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+\nu}]$ *is a.s. uniformly bounded over* N, T*. In addition, we assume that* X_{it} *is bounded uniformly over* i, t, N, T*.*

- *(v) Concavity: For all* N, T, $z \mapsto \ell_{it}(z)$ *is strictly concave over* $z \in \mathbb{R}$ *a.s. Furthermore, there exist positive constants* $b_{\min}$ *and* $b_{\max}$ *such that for all* $z \in \mathcal{B}_\varepsilon^0$, $b_{\min} \leq -\partial_{z^2}\ell_{it}(z) \leq b_{\max}$ *a.s. uniformly over* i, t, N, T*.*

- *(vi) Strong factors:* $\frac{1}{N}\sum_i(\alpha_i^0)^2 \to_P \sigma_\alpha^2 > 0$ *and* $\frac{1}{T}\sum_t(\gamma_t^0)^2 \to_P \sigma_\gamma^2 > 0$.

- *(vii) Generalized noncolinearity: For any* d_v*-vector* v*, define the coprojection matrix as* $\mathcal{M}_v = I_{d_v} - v(v'v)v'$*, where* I_{d_v} *denotes the identity matrix of order* d_v*. The* $\dim\beta \times \dim\beta$ *matrix with elements*

$$D_{k_1k_2}(\gamma) = (NT)^{-1}\mathrm{Tr}(\mathcal{M}_{\alpha^0}X_{k_1}\mathcal{M}_\gamma X_{k_2}'), \quad k_1, k_2 \in \{1, ..., \dim\beta\},$$

  satisfies $D(\gamma) > c > 0$ *for all* $\gamma \in \mathbb{R}^T$*, wpa1.*

We assume that the index z_{it}^0 is bounded. This condition holds if X_{it}, α_i and γ_t are bounded. The relative rate of growth of N and T is chosen to produce a non-degenerate asymptotic distribution. Assumption $4(i) - (v)$ are similar to (Fernández-Val and Weidner, 2013), so we do not discuss them further here. The strong factor and generalized noncolinearity assumptions were previously imposed in (Bai, 2009b) and (Moon and Weidner, 2010a; Moon and Weidner, 2010b) for linear models with interactive effects. Generalized noncolinearity rules out time and cross section invariant explanatory variables.

To describe the asymptotic distribution of the fixed effects estimator $\widehat{\beta}$, it is convenient to introduce some additional notation. Let $\overline{\mathcal{H}}$ be the $(N+T) \times (N+T)$ expected Hessian matrix of the log-likelihood with respect to the nuisance parameters evaluated at the true parameters, i.e.

$$\overline{\mathcal{H}} = \mathbb{E}_\phi[-\partial_{\phi\phi'}\mathcal{L}] = \begin{pmatrix} \overline{\mathcal{H}}_{(\alpha\alpha)} & \overline{\mathcal{H}}_{(\alpha\gamma)} \\ [\overline{\mathcal{H}}_{(\alpha\gamma)}]' & \overline{\mathcal{H}}_{(\gamma\gamma)} \end{pmatrix}, \tag{2.17}$$

where $\overline{\mathcal{H}}_{(\alpha\alpha)} = \operatorname{diag}(\sum_t \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}])/\sqrt{NT}$, $\overline{\mathcal{H}}_{(\alpha\gamma)it} = \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}]/\sqrt{NT}$, and $\overline{\mathcal{H}}_{(\gamma\gamma)} = \operatorname{diag}(\sum_i \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}])/\sqrt{NT}$. Furthermore, let $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\overline{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)}$, and $\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}$ denote the $N \times N$, $N \times T$, $T \times N$ and $T \times T$ blocks of the Moore-Penrose pseudoinverse $\overline{\mathcal{H}}^{-1}$ of $\overline{\mathcal{H}}$. It is convenient to define the projection vector Ξ_{it} and the residual $\tilde{X}_{it}$ by

$$\begin{aligned} \Xi_{it} &:= -\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{\tau=1}^{T} (\gamma_t^0 \gamma_\tau^0 \, \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \alpha_i^0 \gamma_\tau^0 \, \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \\ &+ \gamma_t^0 \alpha_j^0 \, \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \alpha_i^0 \alpha_j^0 \, \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}) \, \mathbb{E}_\phi \left(\partial_{z^2}\ell_{j\tau} X_{j\tau}\right), \\ \tilde{X}_{it} &:= X_{it} - \Xi_{it}. \end{aligned} \tag{2.18}$$

The k-th component of Ξ_{it} corresponds to the following population least squares projection

$$\begin{aligned} \Xi_{it,k} &= \alpha^*_{i,k}\gamma_t^0 + \alpha_i^0 \gamma^*_{t,k}, \\ (\alpha^*_k, \, \gamma^*_k) &= \arg \min_{\alpha_{i,k}, \gamma_{t,k}} \sum_{i,t} \mathbb{E}_\phi(-\partial_{z^2}\ell_{it}) \left(\frac{\mathbb{E}_\phi(\partial_{z^2}\ell_{it} X_{it})}{\mathbb{E}_\phi(\partial_{z^2}\ell_{it})} - \alpha^*_{i,k}\gamma_t^0 - \alpha_i^0 \gamma^*_{t,k} \right)^2. \end{aligned}$$

Let $\overline{\mathbb{E}} := \operatorname{plim}_{N,T\to\infty}$. The following theorem establishes the asymptotic distribution of the fixed effects estimator $\widehat{\beta}$.

Theorem 2.4.1 (Asymptotic distribution of $\widehat{\beta}$)**.** *Suppose that Assumption 4 holds, that the* *following limits exist*

$$\overline{B}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\sum_{\tau=t}^{T}\gamma_t^0\gamma_\tau^0\mathbb{E}_\phi\left(\partial_z\ell_{it}\partial_{z^2}\ell_{i\tau}\tilde{X}_{i\tau}\right)+\frac{1}{2}\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(\partial_{z^3}\ell_{it}\tilde{X}_{it})}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right],$$

$$\overline{D}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_z\ell_{it}\partial_{z^2}\ell_{it}\tilde{X}_{it}+\frac{1}{2}\partial_{z^3}\ell_{it}\tilde{X}_{it}\right)}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right],$$

$$\overline{W}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\tilde{X}_{it}\tilde{X}_{it}'\right)\right],$$

and that $\overline{W}_\infty > 0$. *Then,*

$$\sqrt{NT}\left(\widehat{\beta}-\beta^0\right) \to_d \overline{W}_\infty^{-1}\mathcal{N}(\kappa\overline{B}_\infty+\kappa^{-1}\overline{D}_\infty,\ \overline{W}_\infty),$$

so that $\overline{B}_\infty^\beta = \overline{W}_\infty^{-1}\overline{B}_\infty$ *and* $\overline{D}_\infty^\beta = \overline{W}_\infty^{-1}\overline{D}_\infty$ *in equation* (2.13).

It is instructive to evaluate the expressions of the bias also in our running examples.

Example 2.2.1 (Linear model). *In the linear model with strictly exogenous explanatory variables,* $Y_{it} \mid X_i^T, \alpha, \gamma \sim \mathcal{N}(X_{it}'\beta + \alpha_i\gamma_t, \sigma^2)$ *independently over* i *and* t, *the expressions of the bias of Theorem 2.4.1 yield*

$$\overline{B}_\infty = \overline{D}_\infty = 0,$$

which agree with the no asymptotic bias result in (Bai, 2009b) for homoskedastic linear models with interactive effects.

Example 2.2.2 (Binary response model). *In this case*

$$\ell_{it}(z) = Y_{it}\log F(z) + (1-Y_{it})\log[1-F(z)],$$

so that $\partial_z\ell_{it} = H_{it}(Y_{it}-F_{it})$, $\partial_{z^2}\ell_{it} = -H_{it}\partial F_{it} + \partial H_{it}(Y_{it}-F_{it})$, *and* $\partial_{z^3}\ell_{it} = -H_{it}\partial^2 F_{it} - 2\partial H_{it}\partial F_{it} + \partial^2 H_{it}(Y_{it}-F_{it})$, *where* $H_{it} = \partial F_{it}/[F_{it}(1-F_{it})]$, *and* $\partial^j G_{it} := \partial^j G(Z)|_{Z=z_{it}^0}$ *for any function* G *and* $j = 0,1,2$. *Substituting these values in the expressions of the bias of*

Theorem 2.4.1 for the probit model with all the components of X_{it} strictly exogenous yields

$$\begin{aligned}\overline{B}_\infty &= \overline{\mathbb{E}}\left[\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi[\partial_{z^2}\ell_{it}\tilde{X}_{it}\tilde{X}_{it}']}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right]\beta^0,\\ \overline{D}_\infty &= \overline{\mathbb{E}}\left[\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi[\partial_{z^2}\ell_{it}\tilde{X}_{it}\tilde{X}_{it}']}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right]\beta^0.\end{aligned}$$

The asymptotic bias is therefore a positive definite matrix weighted average of the true parameter value as in the case of the probit model with additive individual and time effects (Fernández-Val and Weidner, 2013).

Example 2.2.3 (Count response model). *In this case*

$$\ell_{it}(z) = zY_{it} - \exp(z) - \log Y_{it}!,$$

so that $\partial_z\ell_{it} = Y_{it} - \omega_{it}$ *and* $\partial_{z^2}\ell_{it} = \partial_{z^3}\ell_{it} = -\omega_{it}$*, where* $\omega_{it} = \exp(z_{it}^0)$*. Substituting these values in the expressions of the bias of Theorem 2.4.1 yields*

$$\overline{B}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\sum_{\tau=t+1}^{T}\gamma_t^0\gamma_\tau^0\mathbb{E}_\phi\left[(Y_{it}-\omega_{it})\omega_{i\tau}\tilde{X}_{i\tau}\right]}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi\left(\omega_{it}\right)}\right],$$

and $\overline{D}_\infty = 0$*. If in addition all the components of* X_{it} *are strictly exogenous, then we get the no asymptotic bias result* $\overline{B}_\infty = \overline{D}_\infty = 0$*.*

2.4.2 Asymptotic distribution of APEs

In nonlinear models we are often interested in APEs, in addition to the model parameters. These effects are averages of the data, parameters and unobserved effects; see expression (2.2). For the panel models of Assumption 4 we specify the partial effects as $\Delta(Y_{it}, X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i\gamma_t)$. The restriction that the partial effects depend on α_i and

γ_t through $\pi_{it} = \alpha_i \gamma_t$ is natural in our panel models since

$$\mathbb{E}[Y_{it} \mid X_i^t, \alpha_i, \gamma_t, \beta] = \int y \exp[\ell_{it}(X_{it}' \beta + \pi_{it})] dy,$$

and the partial effects are usually defined as differences or derivatives of this conditional expectation with respect to the components of X_{it}. For example, the partial effects for the probit and Poisson models and the scale parameter in the linear model described in Section 2.2 satisfy this restriction.

The distribution of the unobserved individual and time effects in general is not ancillary for the APEs, unlike for model parameters. We therefore need to make assumptions on this distribution to define and interpret the APEs, and to derive the asymptotic distribution of their estimators. Here, we control the heterogeneity of the partial effects assuming that the individual effects and explanatory variables are identically distributed cross sectionally and stationary over time so that the APE δ^0_{NT} does not change with N and T, i.e. $\delta^0_{NT} = \delta^0$. We also impose smoothness and moment conditions on the function Δ that defines the partial effects. We use these conditions to derive higher-order stochastic expansions for the fixed effect estimator of the APEs and to bound the remainder terms in these expansions. Let $\pi^0_{it} = \alpha^0_i \gamma^0_t$, $\{\alpha_i\}_N := \{\alpha_i : 1 \leq i \leq N\}$, $\{\gamma_t\}_T := \{\gamma_t : 1 \leq t \leq T\}$, and $\{X_{it}, \alpha_i, \gamma_t\}_{NT} := \{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$.

Assumption 5. *[Partial effects] Let $\nu > 0$, $\epsilon > 0$, and let $\mathcal{B}^0_\varepsilon$ be a subset of $\mathbb{R}^{\dim \beta + 1}$ that contains an ε-neighbourhood of (β^0, π^0_{it}) for all i, t, N, T..*

- *(i) Sampling: for all N, T, $\{X_{it}, \alpha_i, \gamma_t\}_{NT}$ is identically distributed across i and/or stationary across t.*
- *(ii) Model: for all i, t, N, T, the partial effects depend on α_i and γ_t through $\alpha_i \gamma_t$:*

$$\Delta(Y_{it}, X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i \gamma_t).$$

The realizations of the partial effects are denoted by $\Delta_{it} := \Delta_{it}(\beta^0, \alpha_i^0 \gamma_t^0)$.

- *(iii) Smoothness and moments: The function* $(\beta, \pi) \mapsto \Delta_{it}(\beta, \pi)$ *is four times continuously differentiable over* $\mathcal{B}_\varepsilon^0$ *a.s. The partial derivatives of* $\Delta_{it}(\beta, \pi)$ *with respect to the elements of* (β, π) *up to fourth order are bounded in absolute value uniformly over* $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$ *by a function* $M(Z_{it}) > 0$ *a.s., and* $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+\nu}]$ *is a.s. uniformly bounded over* N, T.

- *(iv) Non-degeneracy and moments:* $0 < \min_{i,t}[\mathbb{E}(\Delta_{it}^2) - \mathbb{E}(\Delta_{it})^2] \leq \max_{i,t}[\mathbb{E}(\Delta_{it}^2) - \mathbb{E}(\Delta_{it})^2] < \infty$, *uniformly over* N, T.

Analogous to Ξ_{it} in equation (2.18) we define

$$\Psi_{it} = -\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{\tau=1}^{T} \left(\gamma_t^0 \gamma_\tau^0 \, \overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} + \alpha_i^0 \gamma_\tau^0 \, \overline{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} + \gamma_t^0 \alpha_j^0 \, \overline{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} + \alpha_i^0 \alpha_j^0 \, \overline{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \right) \partial_\pi \Delta_{j\tau}, \tag{2.19}$$

which also corresponds to a weighted least squares population projection. We denote the derivatives of the partial effects $\Delta_{it}(\beta, \pi)$ by $\partial_\beta \Delta_{it}(\beta, \pi) := \partial \Delta_{it}(\beta, \pi)/\partial \beta$, $\partial_{\beta\beta'} \Delta_{it}(\beta, \pi) := \partial^2 \Delta_{it}(\beta, \pi)/(\partial \beta \partial \beta')$, $\partial_{\pi^q} \Delta_{it}(\beta, \pi) := \partial^q \Delta_{it}(\beta, \pi)/\partial \pi^q$, $q = 1, 2, 3$, *etc.* We drop the arguments β and π when the derivatives are evaluated at the true parameters β^0 and $\pi_{it}^0 := \alpha_i^0 \gamma_t^0$, e.g. $\partial_{\pi^q} \Delta_{it} := \partial_{\pi^q} \Delta_{it}(\beta^0, \pi_{it}^0)$.

Let δ^0 and $\widehat{\delta}$ be the APE and its fixed effects estimator, defined as in equations (2.2) and (2.11), where $\widehat{\delta}$ is constructed from a bias corrected estimators of the parameter β, i.e. $\widehat{\delta} = \Delta(\widetilde{\beta}, \widehat{\phi}(\widetilde{\beta}))$, where $\widetilde{\beta}$ is such that $\sqrt{NT}(\widetilde{\beta} - \beta^0) \to_d N(0, \overline{W}_\infty^{-1})$. The following theorem establishes the asymptotic distribution of $\widehat{\delta}$.

Theorem 2.4.2 (Asymptotic distribution of $\widehat{\delta}$)**.** *Suppose that the assumptions of Theorem 2.4.1 and Assumption 5 hold, and that the following limits exist:*

$$\overline{B}_\infty^\delta = \overline{\mathbb{E}} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \sum_{\tau=t}^{T} \gamma_t^0 \gamma_\tau^0 \mathbb{E}_\phi \left(\partial_z \ell_{it} \partial_{z^2} \ell_{i\tau} \Psi_{i\tau} \right)}{\sum_{t=1}^{T} (\gamma_t^0)^2 \mathbb{E}_\phi \left(\partial_{z^2} \ell_{it} \right)} \right]$$

$$
- \overline{\mathbb{E}}\left[\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\gamma_t^0)^2\left[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it})-\mathbb{E}_\phi(\partial_{z^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})\right]}{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right],
$$

$$
\overline{D}_\infty^\delta = \overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_z\ell_{it}\partial_{z^2}\ell_{it}\Psi_{it}\right)}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right]
- \overline{\mathbb{E}}\left[\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\left[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it})-\mathbb{E}_\phi(\partial_{z^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})\right]}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi\left(\partial_{z^2}\ell_{it}\right)}\right],
$$

$$
\overline{V}_\infty^\delta = \overline{\mathbb{E}}\left\{\frac{r_{NT}^2}{N^2T^2}\mathbb{E}\left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\Delta}_{it}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\Delta}_{it}\right)' + \sum_{i=1}^{N}\sum_{t=1}^{T}\Gamma_{it}\Gamma_{it}'\right]\right\},
$$

for some deterministic sequence $r_{NT}\to\infty$ *such that* $r_{NT}=\mathcal{O}(\sqrt{NT})$ *and* $\overline{V}_\infty^\delta>0$, *where* $\widetilde{\Delta}_{it}=\Delta_{it}-\delta^0$ *and* $\Gamma_{it}=\overline{\mathbb{E}}\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\partial_\beta\Delta_{it}\right]'\overline{W}_\infty^{-1}\partial_z\ell_{it}X_{it}-\mathbb{E}_\phi(\Psi_{it})\partial_z\ell_{it}$. *Then,*

$$
r_{NT}(\widehat{\delta}-\delta^0-T^{-1}\overline{B}_\infty^\delta-N^{-1}\overline{D}_\infty^\delta)\to_d\mathcal{N}(0,\ \overline{V}_\infty^\delta).
$$

Remark 2.4.1. [Convergence rate, bias and variance] The rate of convergence r_{NT} is determined by the inverse of the first term of $\overline{V}_\infty^\delta$, which corresponds to the asymptotic variance of $\delta := (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta_{it}$,

$$
r_{NT}^2=\mathcal{O}\left(\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbb{E}[\widetilde{\Delta}_{it}\widetilde{\Delta}_{js}']\right)^{-1}.
$$

Assumption 5(iv) and the condition $r_{NT}\to\infty$ ensure that we can apply a central limit theorem to δ. The exact rate of convergence in general depends on the sampling properties of the unobserved effects. For example, if $\{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are independent sequences, and α_i and γ_t are independent for all i,t, then in general $r_{NT}=\sqrt{NT/(N+T-1)}$,

$$
\overline{V}_\infty^\delta = \overline{\mathbb{E}}\left\{\frac{r_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\left[\sum_{t,\tau=1}^{T}\mathbb{E}(\widetilde{\Delta}_{it}\widetilde{\Delta}_{i\tau}')+\sum_{j\neq i}\sum_{t=1}^{T}\mathbb{E}(\widetilde{\Delta}_{it}\widetilde{\Delta}_{jt}')+\sum_{t=1}^{T}\mathbb{E}(\Gamma_{it}\Gamma_{it}')\right]\right\},
$$

and the asymptotic bias is of order $T^{-1/2}+N^{-1/2}$. The bias and the last term of $\overline{V}_\infty^\delta$ are asymptotically negligible in this case under the asymptotic sequences of Assumption 4(i).

Example 2.2.1 (Linear model). *For $\delta = \sigma^2$, the convergence rate is $r_{NT} = \sqrt{NT}$ regardless of the sampling properties of the unobserved individual and time effects because $\Delta_{it} = (Y_{it} - X_{it}'\beta^0 - \pi_{it}^0)^2$ is independent over i and α-mixing over t. The distribution of the unobserved effects is ancillary for the APE because the information matrix of the log-likelihood $\ell_{it} = -.5 \log 2\pi - .5 \log \delta - .5(Y_{it} - X_{it}'\beta - \pi_{it})^2/\delta$ is orthogonal in π_{it} and δ at $\pi_{it} = \pi_{it}^0$ and $\delta = \delta^0$.*

2.4.3 Bias corrected estimators

The results of the previous sections show that the asymptotic distributions of the fixed effects estimators of the model parameters and APEs can have biases of the same order as the variances under sequences where T grows at the same rate as N. This is the large-T version of the incidental parameters problem that invalidates any inference based on the asymptotic distribution. In this section we describe how to construct analytical bias corrections for panel models and give conditions for the asymptotic validity of analytical and jackknife bias corrections.

The jackknife correction for the model parameter β in equation (2.15) is generic and applies to the panel model. For the APEs, the jackknife correction is formed similarly as

$$\widetilde{\delta}_{NT}^{J} = 3\widehat{\delta}_{NT} - \widetilde{\delta}_{N,T/2} - \widetilde{\delta}_{N/2,T},$$

where $\widetilde{\delta}_{N,T/2}$ is the average of the 2 split jackknife estimators of the APE that leave out the first and second halves of the time periods, and $\widetilde{\delta}_{N/2,T}$ is the average of the 2 split jackknife estimators of the APE that leave out half of the individuals.

The analytical corrections are constructed using sample analogs of the expressions in Theorems 2.4.1 and 2.4.2, replacing the true values of β and ϕ by the fixed effects estimators. To describe these corrections, we introduce some additional notation. For any function of the data, unobserved effects and parameters $g_{itj}(\beta, \alpha_i\gamma_t, \alpha_i\gamma_{t-j})$ with $0 \leq j < t$, let $\widehat{g}_{itj} = g_{it}(\widehat{\beta}, \widehat{\alpha}_i\widehat{\gamma}_t, \widehat{\alpha}_i\widehat{\gamma}_{t-j})$ denote the fixed effects estimator, e.g., $\widehat{\mathbb{E}_\phi[\partial_{z^2}\ell_{it}]}$ denotes the fixed effects estimator of $\mathbb{E}_\phi[\partial_{z^2}\ell_{it}]$. Let $\widehat{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\widehat{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\widehat{\mathcal{H}}^{-1}_{(\gamma\alpha)}$, and $\widehat{\mathcal{H}}^{-1}_{(\gamma\gamma)}$ denote the blocks of the Moore-Penrose pseudo inverse matrix $\widehat{\mathcal{H}}^{-1}$, where

$$\widehat{\mathcal{H}} = \begin{pmatrix} \widehat{\mathcal{H}}_{(\alpha\alpha)} & \widehat{\mathcal{H}}_{(\alpha\gamma)} \\ [\widehat{\mathcal{H}}_{(\alpha\gamma)}]' & \widehat{\mathcal{H}}_{(\gamma\gamma)} \end{pmatrix},$$

$\widehat{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(-\sum_t \widehat{\mathbb{E}_\phi[\partial_{z^2}\ell_{it}]})/\sqrt{NT}$, $\widehat{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(-\sum_i \widehat{\mathbb{E}_\phi[\partial_{z^2}\ell_{it}]})/\sqrt{NT}$, and $\widehat{\mathcal{H}}_{(\alpha\gamma)it} = -\widehat{\mathbb{E}_\phi[\partial_{z^2}\ell_{it}]}/\sqrt{NT}$. Let

$$\begin{aligned}
\widehat{\Xi}_{it} &:= -\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\widehat{\gamma}_t\widehat{\gamma}_\tau\,\widehat{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \widehat{\alpha}_i\widehat{\gamma}_\tau\,\widehat{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \\
&+ \widehat{\gamma}_t\widehat{\alpha}_j\,\widehat{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \widehat{\alpha}_i\widehat{\alpha}_j\,\widehat{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau})\,\mathbb{E}_\phi\,(\widehat{\partial_{z^2}\ell_{j\tau}}X_{j\tau}), \\
\widehat{\widetilde{X}}_{it} &:= X_{it} - \widehat{\Xi}_{it}.
\end{aligned}$$

The k-th component of $\widehat{\Xi}_{it}$ corresponds to the following least squares projection

$$\begin{aligned}
\widehat{\Xi}_{it,k} &= \widehat{\alpha}^*_{i,k}\widehat{\gamma}_t + \widehat{\alpha}_i\widehat{\gamma}^*_{t,k}, \\
(\widehat{\alpha}^*_k,\, \widehat{\gamma}^*_k) &= \arg\min_{\alpha_{i,k},\gamma_{t,k}}\sum_{i,t}\mathbb{E}_\phi(\widehat{-\partial_{z^2}}\ell_{it})\left(\frac{\mathbb{E}_\phi(\widehat{\partial_{z^2}\ell_{it}}X_{it})}{\mathbb{E}_\phi(\widehat{\partial_{z^2}\ell_{it}})} - \alpha^*_{i,k}\widehat{\gamma}_t - \widehat{\alpha}_i\gamma^*_{t,k}\right)^2.
\end{aligned}$$

The analytical bias corrected estimator of β^0 is

$$\widetilde{\beta}^A = \widehat{\beta} - \widehat{W}^{-1}\widehat{B}/T - \widehat{W}^{-1}\widehat{D}/N,$$

where

$$\widehat{B} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=0}^{L}[T/(T-j)]\sum_{t=j+1}^{T}\widehat{\gamma}_t\widehat{\gamma}_{t-j}\mathbb{E}_\phi\left(\widehat{\partial_z\ell_{i,t-j}\partial_{z^2}\ell_{it}\tilde{X}_{it}}\right) + \frac{1}{2}\sum_{t=1}^{T}\widehat{\gamma}_t^2\mathbb{E}_\phi(\widehat{\partial_{z^3}\ell_{it}}\tilde{X}_{it})}{\sum_{t=1}^{T}\widehat{\gamma}_t^2\mathbb{E}_\phi\,\widehat{(\partial_{z^2}\ell_{it})}},$$

$$\widehat{D} = -\frac{1}{T}\sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \widehat{\alpha}_i^2 \left[\mathbb{E}_\phi\left(\widehat{\partial_z \ell_{it} \partial_{z^2} \ell_{it} \tilde{X}_{it}}\right) + \frac{1}{2}\mathbb{E}_\phi\left(\widehat{\partial_{z^3} \ell_{it} \tilde{X}_{it}}\right)\right]}{\sum_{i=1}^{N} \widehat{\alpha}_i^2 \mathbb{E}_\phi\left(\widehat{\partial_{z^2} \ell_{it}}\right)},$$

$$\widehat{W} = -(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}_\phi\left(\widehat{\partial_{z^2} \ell_{it} \tilde{X}_{it} \tilde{X}'_{it}}\right),$$

and L is a trimming parameter for estimation of spectral expectations such that $L \to \infty$ and $L/T \to 0$ (Hahn and Kuersteiner, 2011). The factor $T/(T-j)$ is a degrees of freedom adjustment that rescales the time series averages $T^{-1}\sum_{t=j+1}^{T}$ by the number of observations instead of by T. Unlike for variance estimation, we do not need to use a kernel function because the bias estimator does not need to be positive. Asymptotic $(1-p)$–confidence intervals for the components of β^0 can be formed as

$$\widetilde{\beta}_k^A \pm z_{1-p}\sqrt{\widehat{W}_{kk}^{-1}/(NT)}, \quad k = \{1, ..., \dim \beta^0\},$$

where z_{1-p} is the $(1-p)$–quantile of the standard normal distribution, and $\widehat{W}_{kk}^{-1}$ is the (k,k)-element of the matrix $\widehat{W}^{-1}$.

The analytical bias corrected estimator of δ^0 is

$$\widetilde{\delta}^A = \widehat{\delta} - \widehat{B}^\delta/T - \widehat{D}^\delta/N,$$

where $\widehat{\delta}$ is the APE constructed from a bias corrected estimator of β. Let

$$\widehat{\Psi}_{it} = -\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{\tau=1}^{T}\left(\widehat{\gamma}_t\widehat{\gamma}_\tau\, \widehat{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \widehat{\alpha}_i\widehat{\gamma}_\tau\, \widehat{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} + \widehat{\gamma}_t\widehat{\alpha}_j\, \widehat{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \widehat{\alpha}_i\widehat{\alpha}_j\, \widehat{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\right)\widehat{\partial_\pi \Delta_{j\tau}}.$$

The fixed effects estimators of the components of the asymptotic bias are

$$\widehat{B}^\delta = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j=0}^{L}[T/(T-j)]\sum_{t=j+1}^{T} \widehat{\gamma}_t\widehat{\gamma}_{t-j}\mathbb{E}_\phi\left(\widehat{\partial_z \ell_{i,t-j} \partial_{z^2} \ell_{it} \Psi_{it}}\right)}{\sum_{t=1}^{T} \widehat{\gamma}_t^2 \mathbb{E}_\phi\left(\widehat{\partial_{z^2} \ell_{it}}\right)}$$

$$- \frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \widehat{\gamma}_t^2 \left[\widehat{\mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it})} - \widehat{\mathbb{E}_\phi(\partial_{z^3} \ell_{it})} \widehat{\mathbb{E}_\phi(\Psi_{it})} \right]}{\sum_{t=1}^{T} \widehat{\gamma}_t^2 \widehat{\mathbb{E}_\phi \left(\partial_{z^2} \ell_{it} \right)}},$$

$$\widehat{D}^\delta = \frac{1}{T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \widehat{\alpha}_i^2 \left[\widehat{\mathbb{E}_\phi \left(\partial_z \ell_{it} \partial_{z^2} \ell_{it} \Psi_{it} \right)} - \frac{1}{2} \widehat{\mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it})} + \frac{1}{2} \widehat{\mathbb{E}_\phi(\partial_{z^3} \ell_{it})} \widehat{\mathbb{E}_\phi(\Psi_{it})} \right]}{\sum_{i=1}^{N} \widehat{\alpha}_i^2 \widehat{\mathbb{E}_\phi \left(\partial_{z^2} \ell_{it} \right)}}.$$

The estimator of the asymptotic variance in general depends on the sampling properties of the unobserved effects. Under the independence assumption of Remark 2.4.1,

$$\widehat{V}^\delta = \frac{r_{NT}^2}{N^2 T^2} \sum_{i=1}^{N} \left[\sum_{t,\tau=1}^{T} \widehat{\widetilde{\Delta}}_{it} \widehat{\widetilde{\Delta}}_{i\tau}' + \sum_{t=1}^{T} \sum_{j \neq i} \widehat{\widetilde{\Delta}}_{it} \widehat{\widetilde{\Delta}}_{jt}' + \sum_{t=1}^{T} \widehat{\mathbb{E}_\phi(\Gamma_{it} \Gamma_{it}')} \right], \tag{2.20}$$

where $\widehat{\widetilde{\Delta}}_{it} = \widetilde{\Delta}_{it} - \widehat{\delta}$. Note that we do not need to specify the convergence rate to make inference because the standard errors $\sqrt{\widehat{V}^\delta}/r_{NT}$ do not depend on r_{NT}. Bias corrected estimators and confidence intervals can be constructed in the same fashion as for the model parameter.

We use the following homogeneity assumption to show the validity of the jackknife corrections for the model parameters and APEs. It ensures that the asymptotic bias is the same in all the partitions of the panel. The analytical corrections *do not* require this assumption.

Assumption 6. *[Unconditional homogeneity] The sequence $\{(Y_{it}, X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$ is identically distributed across i and strictly stationary across t, for each N, T.*

Remark 2.4.2. [Test of homogeneity] Assumption 6 is a sufficient condition for the validity of the jackknife corrections. The weaker condition that the asymptotic biases are the same in all the partitions of the panel can be tested using the Chow-type test recently proposed in (Dhaene and Jochmans, 2014).

The following theorems are the main result of this section. They show that the analytical and jackknife bias corrections eliminate the bias from the asymptotic distribution of the fixed effects estimators of the model parameters and APEs without increasing variance, and that the estimators of the asymptotic variances are consistent.

Theorem 2.4.3 (Bias corrections for $\widehat{\beta}$). *Under the conditions of Theorems 2.4.1,*

$$\widehat{W} \to_P \overline{W}_\infty,$$

and, if $L \to \infty$ *and* $L/T \to 0$,

$$\sqrt{NT}(\widetilde{\beta}^A - \beta^0) \to_d \mathcal{N}(0, \overline{W}_\infty^{-1}).$$

Under the conditions of Theorems 2.4.1 and Assumption 6,

$$\sqrt{NT}(\widetilde{\beta}^J - \beta^0) \to_d \mathcal{N}(0, \overline{W}_\infty^{-1}).$$

Theorem 2.4.4 (Bias corrections for $\widehat{\delta}$). *Under the conditions of Theorems 2.4.1 and 2.4.2,*

$$\widehat{V}^\delta \to_P \overline{V}_\infty^\delta,$$

and, if $L \to \infty$ *and* $L/T \to 0$,

$$r_{NT}(\widetilde{\delta}^A - \delta_{NT}^0) \to_d \mathcal{N}(0, \overline{V}_\infty^\delta).$$

Under the conditions of Theorems 2.4.1 and 2.4.2, and Assumption 6,

$$r_{NT}(\widetilde{\delta}^J - \delta^0) \to_d \mathcal{N}(0, \overline{V}_\infty^\delta).$$

Remark 2.4.3. [Rate of convergence] The rate of convergence r_{NT} depends on the properties of the sampling process for the explanatory variables and unobserved effects (see remark 2.4.1).

2.5 Numerical Examples

To illustrate how the bias corrections work in finite samples, we consider the non-regression version of Example 2.2.1, $Y_{it} \mid \alpha, \gamma \sim \mathcal{N}(\alpha_i \gamma_t, \sigma^2)$ independently over i and t. In this linear model the fixed effects estimator of ϕ_{NT} can be obtained by the principal component method of (Bai, 2009b) or by Algorithm 2.2.1 with $\mathcal{L}_{NT}(\delta, \phi_{NT}) = -\sum_{i,t}(Y_{it} - \alpha_i \gamma_t)^2/2$. Then, the fixed effects estimator of the APE $\delta = \sigma^2$ is

$$\widehat{\delta}_{NT} = (NT)^{-1} \sum_{i,t} (Y_{it} - \widehat{\alpha}_i \widehat{\gamma}_t)^2 .$$

Applying the results of Theorem 2.4.2 to $\Delta_{it} = (Y_{it} - \alpha_i \gamma_t)^2$, the probability limit of $\widehat{\delta}_{NT}$ admits the expansion

$$\widehat{\delta}_{NT} = \delta^0 \left(1 - \frac{1}{T} - \frac{1}{N}\right) + o_P\left(\frac{1}{T} \vee \frac{1}{N}\right),$$

as $N, T \to \infty$, so that $\overline{B}^{\delta}_{\infty} = -\delta^0$ and $\overline{D}^{\delta}_{\infty} = -\delta^0$.

To form the analytical bias correction we can set $\widehat{B}^{\delta}_{NT} = -\widehat{\delta}_{NT}$ and $\widehat{D}^{\delta}_{NT} = -\widehat{\delta}_{NT}$. This yields $\widetilde{\delta}^{A}_{NT} = \widehat{\delta}_{NT}(1 + 1/T + 1/N)$ with

$$\widetilde{\delta}^{A}_{NT} = \delta^0 + o_P(T^{-1} \vee N^{-1}).$$

This correction reduces the order of the bias, but it increases finite-sample variance because the factor $(1 + 1/T + 1/N) > 1$. We compare the biases and standard deviations of the fixed effects estimator and the corrected estimator in a numerical example below. For the Jackknife correction, straightforward calculations give

$$\widetilde{\delta}^{J}_{NT} = 3\widehat{\delta}_{NT} - \widehat{\delta}_{N,T/2} - \widehat{\delta}_{N/2,T} = \delta^0 + o_P(T^{-1} \vee N^{-1}).$$

Table 2.1 presents numerical results for the bias and standard deviations of the fixed effects and bias corrected estimators in finite samples obtained from 50,000 simulations. We consider panels with $N, T \in \{10, 25, 50\}$, and only report the results for $T \leq N$ since all the expressions are symmetric in N and T. All the numbers in the table are in percentage of the true parameter value, so we do not need to specify the value of δ^0. We only report results based on the fixed effects estimator that uses Algorithm 2.2.1, because the results based on the estimator that uses principal components are identical up to the tolerance level.[4] By comparing the first two rows of the table, we find that the first order approximation captures most of the bias of the fixed effects estimator. The analytical and jackknife corrections offer substantial improvements in terms of bias reduction. The second and sixth row of the table show that the bias of the fixed effects estimator is of the same order of magnitude as the standard deviation, where $\overline{V}_{NT} = \text{Var}[\widehat{\delta}_{NT}] = 2(N-1)(T-1)(\delta^0)^2/(NT)^2$ under independence of Y_{it} over i and t conditional on the unobserved effects. The seventh row shows the increase in standard deviation due to analytical bias correction is small compared to the bias reduction, where $\overline{V}^A_{NT} = \text{Var}[\widetilde{\delta}^A_{NT}] = (1+1/N+1/T)^2\overline{V}_{NT}$. The last row shows that the jackknife yields less precise estimates than the analytical correction in small panels. The asymptotic variance $\overline{V}_\infty = 2(\delta^0)^2/(NT)$ in the fifth row provides a good approximation to the finite sample variance of all the estimators.

Table 2.2 illustrates the effect of the bias on the inference based on the asymptotic distribution. It shows the coverage probabilities of 95% asymptotic confidence intervals for δ^0 constructed in the usual way as

$$\text{CI}_{.95}(\widehat{\delta}) = \widehat{\delta} \pm 1.96\widehat{V}^{1/2}_{NT} = \widehat{\delta}(1 \pm 1.96\sqrt{2/(NT)})$$

where $\widehat{\delta} = \{\widehat{\delta}_{NT}, \widetilde{\delta}^A_{NT}, \widetilde{\delta}^J_{NT}\}$ and $\widehat{V}_{NT} = 2\widehat{\delta}^2/(NT)$ is an estimator of the asymptotic variance $\overline{V}_\infty$. Here we find that the confidence intervals based on the fixed effect estimator display severe undercoverage for all the sample sizes. The confidence intervals based on the corrected estimators have high coverage probabilities, which approach the nominal level as the sample

[4]We set the tolerance criterium to $|\widehat{\delta}^{(m)} - \widehat{\delta}^{(m-1)}| < \epsilon_{tol} = 10^{-4}$.

size grows, as expected from the asymptotic results.

Table 2.1: Biases and Standard Deviations for $Y_{it} \mid \alpha, \gamma, \delta \sim \mathcal{N}(\alpha_i \gamma_t, \delta)$

	N = 10	N=25		N=50		
	T = 10	T=10	T=25	T=10	T=25	T=50
$(\overline{B}_\infty/T + \overline{D}_\infty/N)/\delta^0$	-.20	-.14	-.08	-.12	-.06	-.04
$(\widehat{\delta}_{NT} - \delta^0)/\delta^0$	-.20	-.14	-.08	-.12	-.06	-.04
$(\widetilde{\delta}^A_{NT} - \delta^0)/\delta^0$	-.04	-.02	-.01	-.01	.00	.00
$(\widetilde{\delta}^J_{NT} - \delta^0)/\delta^0$	.01	.00	-.01	.00	.00	.00
$\sqrt{\overline{V}_\infty}/\delta^0$	.14	.09	.06	.06	.04	.03
$\sqrt{\overline{V}_{NT}}/\delta^0$	.13	.08	.05	.06	.04	.03
$\sqrt{\overline{V}^A_{NT}}/\delta^0$	.15	.09	.06	.07	.04	.03
$\sqrt{\overline{V}^J_{NT}}/\delta^0$	.18	.10	.06	.07	.04	.03

Notes: Results obtained by 50,000 simulations

Table 2.2: Coverage probabilities for $Y_{it} \mid \alpha, \gamma, \delta \sim \mathcal{N}(\alpha_i \gamma_t, \delta)$

	N = 10	N=25		N=50		
	T = 10	T=10	T=25	T=10	T=25	T=50
$\mathrm{CI}_{.95}(\widehat{\delta}_{NT})$	.52	.53	.63	.43	.62	.67
$\mathrm{CI}_{.95}(\widetilde{\delta}^A_{NT})$	.88	.91	.93	.92	.94	.94
$\mathrm{CI}_{.95}(\widetilde{\delta}^J_{NT})$	.89	.90	.92	.92	.93	.94

Results obtained by 50,000 simulations. Nominal coverage probability is .95.

Chapter 3

Quantile Graphical Models: Prediction and Conditional Independence with Applications to Financial Risk Management[1]

3.1 Introduction

We propose Quantile Graphical Models (QGMs) to characterize and visualize the dependence structure of a set of random variables. The proposed framework allows us to understand prediction and conditional independence between these variables. Moreover, it also enable us to focus on specific parts of the distributions of these variables such as tail events. Such understanding plays an important role in applications like financial contagion and systemic risk measuring where extreme events are the main interest for regulators. Our techniques are intended to be applied in high-dimensional settings where the total number of variables (or additional conditioning variables) is large – possibly larger than the sample size.

Our work is complementary to a large body of work that focused on the case of jointly Gaussian random variables (Lauritzen, 1996). In such setting, it is well known that conditional independence structure is completely characterized by the covariance matrix of the random variables of interest. Indeed, a zero entry in the precision matrix (inverse of the covariance matrix) identifies a pair of conditionally independent variables. Thus the precision matrix can be directly translated into a Gaussian graphical model (GGM) which can be used to study the interdependence. Further this approach characterize the conditional

[1] This chapter is based on a joint work with Alexandre Belloni and Victor Chernozhukov

mean predictability of one set of the variables by linear combinations of the other variables.

In this work we provide an alternative route to estimate conditional independence and predictability under asymmetric loss functions that is appealing to Gaussian and non-Gaussian settings. It hinges on the equivalence between conditional probabilities and conditional quantiles to characterize a random variable. We build upon the quantile regression literature (Koenker, 2005) to represent dependence. Furthermore, we exploit recent works on penalized quantile regression methods that allow the estimation of the conditional quantile function in high dimensional settings which enables us to handle many controls and transformations of the original variables to achieve a flexible specification.

Our interest lies on understanding the dependence between the components of a d-dimensional random vector X_V, where the set V contains the labels of the components. Quantile graph models (QGMs) allow us to visualize dependence for each specific quantile index τ through a graph where the set of nodes V represents the components of X_V and edges represent a relation between the corresponding components. Therefore we have a graph process indexed by $\tau \in (0,1)$. The structure represented by the τ-quantile graph represents a local relation and can be valuable in applications where the tail interdependence (high or low quantile index) is the main interest.

The network produced by QGMs has several important features. First, it enables different strength of the links in different directions. This is important because for undirected networks, the distinction between exposure and contribution is unclear. Second, compared with Gaussian Graphical Models (which is characterized by the covariance matrix), QGMs are able to capture the tail interdependence through estimating at up or low quantiles. Third, QGMs can capture the asymmetric dependence structure at different quantiles, which can be particularly useful in applications (e.g., stock market returns, exchange rate dependence). In addition, by considering all the quantiles we can characterize the conditional independence structure between the variables. This is useful specially when the variables are not jointly Gaussian distributed, in which case the covariance matrix cannot completely characterize establish conditional independence.

We also provide estimation methods to learn QGMs from data. The estimators are geared to cover high-dimensional settings where the size of the model is large relative to the sample size. These estimators are based on ℓ_1-penalized quantile regression and low biased equations. Under mild regularities conditions, we discuss rates of convergence and properties of the selected graph structure that hold uniformly over a large class of data generating process. Furthermore, based on proper thresholding, recovery of the QGMs pattern is possible when coefficients are well separated from zero which parallel the results for graph recovery in the Gaussian case based on the estimation of the precision matrix. (Similar to the graph recovery in the Gaussian case the exact recover is subject to the lack of uniformity validity critiques of Leeb and Pötscher (Leeb and Pötscher, 2008).) Of independent interest, the analysis of the ℓ_1-penalized quantile regression derived here considers a set of index $\mathcal{T}$ that grows to (asymptotically) cover $(0,1)$. Under additional weak conditions, the same rate of convergence established in (Belloni et al., 2011) can be achieved when $\mathcal{T}$ grows (provided it does not grow too fast relative to the sample size).

QGMs can play an important role in applications where tail events are relevant. With certain rescaling of the edge weights, we are able to capture the importance of each node or measuring its systemic risk contribution. In parallel with (Andersen et al., 2013), many approaches to systemic risk measurement fit naturally into the QGMs. For example, one can view the $\Delta CoVaR_\tau^{b|a}$, $a, b \in V$ (suitably scaled), as a measure of the impact of firm a on firm b, as the weight in the edge of a QGM at quantile τ. Then, the systemic risk of firm a, $\Delta CoVaR_\tau^{sys|a}$ which measures contributions of individual firms to systemic network event, equals to the sum of coefficients over $b \in V$, $\sum_{b \in V} \Delta CoVaR_\tau^{b|a}$. Similarly, the sum over $a \in V$ measures exposures of individual firms to systemic shocks from the network.

QGMs can also be used to study contagion and network spillover effects since it is useful for studying tail risk spillovers. We consider the study of international financial contagion in volatilities, specializing in estimating the risk transmission channels, see (Claessens and

Forbes, 2001) for an overview on international financial contagions. After estimating the risk transmission channels, we can use our $\Delta CoVaR$ measure to calculate the contribution and exposure of each country to the whole market. Our method applies to the case where many countries involved, overcome the problem of curse of dimensionality that traditional methods might have.

Understanding the dependence between stock returns plays a key role in hedging strategies. However, these strategies are critical precisely during downside movement of the market. The union of QGMs can be more informative in representing conditional independence than Gaussian graphical models in this setting. Indeed, recent empirical evidence (Ang et al., 2006; Ang and Chen, 2002; Patton, 2004) points to non-Gaussianity of the distribution of stock returns, especially during market downturns. Further, hedging decisions are typically interested on extreme outcomes rather than average outcomes. Finally, it is also instructive to understand how the dependence (and policies) would change as the downside movement of the market becomes more extreme. This application motivated us to consider conditional QGMs that extend the previous models to be conditional on additional events (e.g. downside movement of the market).

Regarding the conditional independence structure for high dimensional models, this paper relates to the large statistic literature on estimating high dimensional Gaussian Graphical Models. It is well known that recovering the structure of an undirected Gaussian graph is equivalent to recovering the support of the precision matrix, i.e. covariance matrix estimation, (Dempster, 1972) and (Lauritzen, 1996; Cox and Wermuth, 1996; Edwards, 2000). Several methods for covariance matrix estimation involves hypothesis testing, (Edwards, 2000; Drton and Perlman, 2004; Drton and Perlman, 2007; Drton and Perlman, 2008). In the high-dimensional setting, (Meinshausen and Bühlmann, 2006) propose neighborhood selection with the Lasso for each node in the graph and combine the results column-by-column to get the final Gaussian graphs. (Yuan and Lin, 2007; Banerjee et al., 2008; Friedman et al., 2008) directly estimate the precision matrix through penalizing the log-likelihood function directly. Other refinement estimators including (Yuan, 2010; Cai et al., 2011; Liu and Luo, 2012; Sun and Zhang, 2012; Liu and Wang, 2012). (Liu et al., 2009) extended the result to a more general class of models called nonparanormal models or semiparametric Gaussian copula models, i.e., the variables follow a joint normal distribution after a set of unknown monotone transformations. See also (Liu et al., 2012; Xue and Zou, 2012; Xue et al., 2012). However, all those methods assume the (transformed) random matrix follows a joint normal distribution. The proposed work provides a complementary method for additional settings by giving up efficiency in the Gaussian case.

The rest of the paper is organized as follows. Section 3.2 lays out the foundation of QGMs. Section 3.3 contains estimators for QGMs. Section 3.4 contains some simulation evidence. Section 3.5 provides empirical applications of QGMs to measure systemic risk contribution and to hedging conditional on the downside movements of the US stock market.

Notation. For an integer k, we let $[k] := \{1, \dots, k\}$ denote the set of integers from 1 to k. For a random variable X we note by $\mathcal{X}$ its support. We use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use $\|\cdot\|_p$ to denote the p-norm of a vector as well as the induced p-norm of a matrix. We denote the ℓ_0-"norm" by $\|\cdot\|_0$ (i.e., the number of nonzero components), the max norm by $\|\mathbf{A}\|_{max} = \max\{|A_{ij}|\}$, the Frobenius norm by $\|\mathbf{A}\|_F = \{\sum_{i \in V} \sum_{j \in V} A_{ij}^2\}^{1/2}$. We denote by $\|\beta\|_{1,n} = \sum_{j=1}^{d} \hat{\sigma}_j |\beta_j|$ the ℓ_1-norm weighted by $\hat{\sigma}_j$'s. Finally, given a vector $\delta \in \mathbb{R}^p$, and a set of indices $T \subset \{1, ..., d\}$, we denote by δ_T the vector in which $\delta_{Tj} = \delta_j$ if $j \in T$, $\delta_{Tj} = 0$ if $j \notin T$.

3.2 Quantile Graph Models

In this section we describe quantile graph models associated with a d-dimensional random vector $X = X_V$ where the set $V = [d] = \{1, \dots, d\}$ denotes the labels of the components. These models aim to provide a description of the interdependence between the random variables in X_V. In particular, they induce graphs that allow for visualization of dependence structures. However, different models arise from different objectives as we discuss below.

3.2.1 Conditional Independence Quantile Graphs

Conditional independence graphs have been used to provide a visualization and insight on the dependence structure between random variables. Each node of the graph is associated with a component of X_V. We denote the conditional independence graph as $G^I = (V, E^I)$ where G^I is an undirected graph with vertex set V and edge set E which is represented by an adjacency matrix ($E^I_{a,b} = 1$ if the edge $(a,b) \in G^I$, and $E^I_{a,b} = 0$ otherwise). An edge (a,b) is not contained in the graph if and only if

$$X_a \perp X_b \ \mid \ X_{V\setminus\{a,b\}}, \tag{3.1}$$

namely X_b and X_a are independent conditionally on all remaining variables $X_{V\setminus\{a,b\}} = \{X_k; k \in V\setminus\{a,b\}\}$.

Remark 3.2.1 (Conditional Independence Under Gaussianity). In the case that X is jointly normally distributed, $X_V \sim N_d(0,\Sigma)$ with Σ as the covariance matrix of X_V, the conditional independence structure between two components is determined by the inverse of covariance matrix, i.e. the precision matrix $\Theta = \Sigma^{-1}$. It follows that the nonzero elements in the precision matrix corresponds to the nonzero coefficients of the associated (high dimensional) mean regression. The family of Gaussian distributions with this property is known as a Gauss-Markov random field with respect to the graph G. This observation has motivated a large literature (Lauritzen, 1996) and some extension that allow transformations of Gaussian variables.

Our main interest is to allow for non-Gaussian distributions. In order to achieve a tractable concept in such generality, we use that (3.1) occurs if and only if

$$F_{X_a}(\cdot \mid X_{V\setminus\{a\}}) = F_{X_a}(\cdot \mid X_{V\setminus\{a,b\}}) \ \text{ for all } \ X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{3.2}$$

In turn, by the equivalence between conditional probabilities and conditional quantiles to characterize a random variable, we have that (3.1) occurs if and only if

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \quad \text{for all } \tau \in (0,1), \quad \text{and} \quad X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{3.3}$$

For a quantile index $\tau \in (0,1)$, we define the τ-quantile conditional independence graph as the directed graph $G(\tau) = (V, E^I(\tau))$ with vertex set V and edge set $E^I(\tau)$. An edge (a,b) is not contained in the edge set $E^I(\tau)$ if and only if

$$Q_{X_a}(\tau \mid X_{V\setminus\{a\}}) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}) \;\; \text{for all} \;\; X_{V\setminus\{a\}} \in \mathcal{X}_{V\setminus\{a\}}. \tag{3.4}$$

By the equivalence between (3.2) and (3.3), the union of τ-quantile graphs over $\tau \in (0,1)$ represents the conditional independence structure of X, namely $E^I = \cup_{\tau\in(0,1)} E^I(\tau)$. This motivates us to consider a relaxation of (3.1). For a set of quantile indices $\mathcal{T} \subset (0,1)$, we say that

$$X_a \perp_{\mathcal{T}} X_b \;\mid\; X_{V\setminus\{a,b\}}, \tag{3.5}$$

X_a and X_b are $\mathcal{T}$-conditionally independent given $X_{V\setminus\{a,b\}}$, if (3.4) holds for all $\tau \in \mathcal{T}$. Thus, we have that (3.1) implies (3.5).[2] We define the quantile graph associated with $\mathcal{T}$ as

$$E^I(\mathcal{T}) = \cup_{\tau\in\mathcal{T}} E^I(\tau).$$

Although the conditional independence concept relates to all quantile indices, the quantile characterization described above also lends itself to quantile specific impacts which can be of independent interest.[3]

[2] In our analysis we will allow $\mathcal{T}$ to change with n so that it approaches $(0,1)$ asymptotically.

[3] For example, we might be interested in some extreme events which typically correspond to crises in financial systems.

3.2.2 Prediction Quantile Graphs under Asymmetric Check Function Loss

Prediction Quantile Graph Models (PCGMs) are concerned with prediction accuracy (instead of conditional independence as in Section 3.2.1). More precisely, for each $a \in V$, we are interested on the predicting X_a based on linear combinations of the remaining variables, $X_{V\setminus\{a\}}$, where accuracy is measured with respect to an asymmetric loss function. Formally, PQGMs measure accuracy as

$$\mathcal{LQ}_\tau(a \mid V \setminus \{a\}) = \min_\beta E[\rho_\tau(X_a - X'_{V\setminus\{a\}}\beta)] \tag{3.6}$$

where the asymmetric loss function $\rho_\tau(t) = \tau - 1\{t \leq 0\}t$ is the check function used in Koenker and Basset (1978).

Importantly, PQGMs are concerned with the best linear predictor under the asymmetric loss function ρ_τ. This is a fundamental distinction with respect to CIQGMs discussed in Section 3.2.1 where the specification of the conditional quantile was approximately a linear function of transformations Z_a. Indeed, we note that under suitable conditions the linear predictor that solves the minimization problem in (3.6) approximates the conditional quantile regression as shown in (Belloni et al., 2011). (In fact, the conditional quantile function would be linear if the vector X_V was jointly Gaussian.) However, PQGMs do not assume that the conditional quantile function of X_a is well approximated by a linear function and instead it focuses on the best linear predictor.

In principle each component of X_V can have predictive power for other components. However, we say that X_b is predictively uninformative for X_a given $X_{V\setminus\{a,b\}}$ if

$$\mathcal{LQ}_\tau(a \mid V \setminus \{a\}) = \mathcal{LQ}_\tau(a \mid V \setminus \{a, b\}) \quad \text{for all} \quad \tau \in (0, 1).$$

Therefore, considering a linear function of X_b does not improve our performance of predicting X_a with respect to the asymmetric loss function ρ_τ.

Again we can visualize the prediction relations using a graph process indexed by $\tau \in$

$(0,1)$. PQGMs allow us to visualize which variables are predictively informative to another variable by using a directed graph $G^P(\tau) = (V, E^P(\tau))$ where edge (a,b) is in the graph only if X_b is predictively informative for X_a given $X_{V\setminus\{a,b\}}$ at the quantile τ. Finally we define the PQGM associated with $\mathcal{T} \subset (0,1)$ as

$$E^P(\mathcal{T}) = \cup_{\tau\in\mathcal{T}} E^P(\tau).$$

3.2.3 Conditional Quantile Graphical Models

In this section we discuss an useful extension of the QGMs discussed in Sections 3.2.1 and 3.2.2. It allows for conditioning on additional events Z where the index set $\mathcal{Z}$ is possibly infinite. This is motivated by several applications where the interdependence between the random variables in X_V maybe substantially impacted by additional observable events.

This general framework allows to accommodate different forms of conditioning: (i) Z might denote additional variables; or (ii) Z can be an event. The main implication of this extension is that the QGMs are now graph processes indexed not only by $\tau \in (0,1)$ but also by $Z \in \mathcal{Z}$.

In order to generalize CIQGMs, we say that $(\mathcal{T}, Z)$-conditionally independent,

$$X_a \perp_{\mathcal{T}} X_b \ \mid \ X_{V\setminus\{a,b\}}, Z \tag{3.7}$$

provided that for all $\tau \in \mathcal{T}$ we have

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}, Z) = Q_{X_a}(\tau|X_{V\setminus\{a,b\}}, Z). \tag{3.8}$$

The conditional independence edge set associated with (τ, Z) is defined analogously as before. We denote them by $E^I(\tau, Z)$ and $E^I(\mathcal{T}, Z) = \cup_{\tau\in\mathcal{T}} E^I(\tau, Z)$ for $Z \in \mathcal{Z}$.

The extension of PQGMs proceeds by defining the accuracy under the asymmetric loss function conditionally on Z. More precisely, we define

$$\mathcal{LQ}_\tau(a \mid V \setminus \{a\}, Z) = \min_\beta E[\rho_\tau(X_a - X'_{V\setminus\{a\}}\beta) \mid Z]. \tag{3.9}$$

The predictive edge set associated with (τ, Z) is also defined analogously as before. We denote as $E^P(\tau, Z)$ and $E^P(\mathcal{T}, Z) = \cup_{\tau\in\mathcal{T}} E^P(\tau, Z)$.

Example 3.2.1 (Predictive QGMs of Stock Returns Under Downside Market Movement)**.** Hedging decisions rely on the dependence of the returns of various stocks. However, hedging's performance is more relevant during downside movements of the market. In such setting it is of interest to understand interdependence conditionally on downside movements. We can parameterize the downside movements by using a random variable M, which denotes a market index, and condition the on the event $Z = \{M \leq m\}$. This allows us to define conditional quantile graphical models $G^P(\tau, Z) = (V, E^P(\tau, Z))$, for $Z \in \mathcal{Z}$. □

3.3 Estimators for High-Dimensional Quantile Graphical Models

In this section we propose and discuss estimators for QGMs introduced in Section 3.2. Throughout this section it is assumed that we observe i.i.d. observations of the d-dimensional random vector X_V, namely $\{X_{iV} : i = 1, \ldots, n\}$. Given the finite data, unless additional assumptions are imposed we cannot estimate the quantities of interest for all $\tau(0,1)$. We will consider a (compact) set of quantile index $\mathcal{T} \subset (0,1)$. Nonetheless, the estimators are intended to handle high dimensional models. In particular we consider a sequence of models where d and $\mathcal{T}$ are indexed by the sample size n and allowed to grow as n grows.

3.3.1 Estimators for Conditional Independence Quantile Graphs

Next we discuss the specification and propose an estimator for CIQGMs. In general it is potentially hard to correctly specify coherent models. The following examples provide us with a starting point.

Example 3.3.1 (Gaussian Case)**.** Consider the Gaussian case, $X_V \sim N(\mu, \Sigma)$ and $V = [d]$. It follows that for $a \in V$, the conditional distribution $X_a \mid X_{V\setminus\{a\}}$ satisfies

$$X_a \mid X_{V\setminus\{a\}} \sim N\left(\mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j - \mu_j), \frac{1}{(\Sigma^{-1})_{aa}}\right).$$

Therefore the conditional quantile function of X_a is linear in $X_{V\setminus\{a\}}$ and is given by

$$Q_{X_a}(\tau|X_{V\setminus\{a\}}) = \frac{\Phi^{-1}(\tau)}{(\Sigma^{-1})_{aa}^{1/2}} + \mu_a - \sum_{j\in V\setminus\{a\}} \frac{(\Sigma^{-1})_{aj}}{(\Sigma^{-1})_{aa}}(X_j - \mu_j).$$

Example 3.3.2 (Multiplicative Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Assume that X_2 and ε are independent positive random variables. Assume further that they relate to X_1 as

$$X_1 = \alpha + \varepsilon X_2.$$

In this case we have that the conditional quantile functions are linear and given by

$$Q_{X_1}(\tau|X_2) = \alpha + F_\varepsilon^{-1}(\tau)X_2 \quad \text{and} \quad Q_{X_2}(\tau|X_1) = (X_1 - \alpha)/F_\varepsilon^{-1}(1-\tau).$$

□

Example 3.3.3 (Additive Error Model)**.** Consider $d = 2$ so that $V = \{1, 2\}$. Let $X_2 \sim U(0, 1)$ and $\epsilon \sim U(0, 1)$ be independent random variables. Also define the random variable X_1 is defined as

$$X_1 = \alpha + \beta X_2 + \varepsilon.$$

It follows that $Q_{X_1}(\tau|X_2) = \alpha + \beta X_2 + \tau$. However, if $\beta = 0$, we have $Q_{X_2}(\tau|X_1) = \tau$, and for $\beta > 0$, direct calculations yield that

$$Q_{X_2}(\tau|X_1) = \begin{cases} \frac{\tau}{\beta}(X_1 - \alpha), & \text{if} \quad X_1 \le \alpha + \beta \\ \tau + (1-\tau)(X_1 - \alpha - \beta), & \text{if} \quad X_1 \ge \alpha + \beta \end{cases}$$

where we note that $X_1 \in [\alpha, 1+\alpha+\beta]$. □

Although a linear specification is correct for Examples 3.3.1 and 3.3.2, Example 3.3.3 above illustrates that we need to consider more general transformation of the basic covariates X_V in the specification for each conditional quantile function. Nonetheless, specifications with additional non-linear terms can approximate non-drastic departures from normality.

We will consider a conditional quantile representation for each $a \in V$. It is based on transformations of the original covariates $X_{V\setminus\{a\}}$ that creates a p-dimensional vector $Z_a = P(X_{V\setminus\{a\}}) \in \mathbb{R}^p$ so that

$$Q_{X_a}(\tau | X_{V\setminus\{a\}}) = Z_a' \beta^a(\tau) + r_a(\tau), \quad \beta^a(\tau) \in \mathbb{R}^p, \quad \text{for all } \tau \in \mathcal{T} \tag{3.10}$$

where $r_a(\tau)$ denotes a small approximation error. For $b \in V \setminus \{a\}$ we let $I_a(b) := \{j : Z_{aj} \text{ depends on } X_b\}$. That is $I_a(b)$ has the components of Z_a that are functions of X_b. Under correct specification, if X_a and X_b are conditionally independent, we have $\beta^a_j(\tau) = 0$ for all $j \in I_a(b)$.

This allows us to connect the conditional independence quantile graph estimation problem with a model selection within quantile regression. Indeed, the representation (3.10) has been used in several quantile regression models, see (Koenker, 2005). Under mild conditions this model allows us to identify the process $(\beta^a(\tau))_{\tau\in\mathcal{T}}$ as the solution of the following optimization problem

$$\beta^a(\tau) \in \arg\min_{\beta} \; E[\rho_\tau(X_a - Z_a'\beta)]. \tag{3.11}$$

In order to allow a flexible specification, it is attractive to consider a high-dimensional vector of Z_a where its dimension p is possibly larger than the sample size. In turn, having a large number of technical controls creates an estimation challenge if the number of coefficients p is not negligible with respect to the sample size n. A useful condition that makes estimation possible in such high dimensional setting that applies to several applications is approximate sparsity (Fan et al., 2011; Belloni et al., 2012; Belloni et al., 2014). Formally we require

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\|\beta^a(\tau)\|_0 \le s \quad \text{and} \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}}\{E[r_a^2(\tau)]\}^{1/2} \lesssim \sqrt{s/n}, \tag{3.12}$$

where the sparsity parameter s of the model is allowed to grow (at a slower rate) as n grows.

Algorithm 1 below contains our proposal to estimate $\beta^a(\tau), a \in V, \tau \in \mathcal{T}$. It is based on three procedures in order to overcome the high-dimensionality. In the first step we apply a ℓ_1-penalized quantile regression. The second step applies Lasso where the data is weighted by

$$f_{a\tau} = f_{X_a|X_{V\setminus\{a\}}}(Q_{X_a}(\tau \mid X_{V\setminus\{a\}}) \mid X_{V\setminus\{a\}}),$$

the conditional density at the true quantile. (We note that an estimate for $f_{a\tau}$ is directly available based on the estimator for ℓ_1-penalized quantile regression for $\tau+h$ and $\tau-h$ where h is a bandwidth parameter, see (Koenker, 2005; Belloni et al., 2013b).) Finally the third step aims to remove the bias from penalization and applies instrumental quantile regression.

There are several parameters that need to be specified for Algorithm 1. First, it assumes that the columns have been normalized such that $\mathbb{E}_n[Z_{iaj}^2] = 1$, $a \in V$, $j \in [p]$. The penalty parameter λ_I is chosen as the $(1-\xi)$-quantile of the ℓ_∞-norm of the score at the true quantile function where $1-\xi$ is the desired confidence level. It was shown in (Belloni and Chernozhukov, 2011) that the score has a pivotal distribution

$$W = \max_{a\in V}\sup_{\tau\in\mathcal{T}} \frac{\|\mathbb{E}_n[(1\{U_i \le \tau\} - \tau)Z_a]\|_\infty}{\sqrt{\tau(1-\tau)}} \tag{3.13}$$

where $\{U_i : i = 1, \ldots, n\}$ are i.i.d. uniform $(0,1)$ random variables. Regarding the parameters for the weighted Lasso in Step 2, the choice of penalty level $\lambda := 1.1n^{-1/2}2\Phi^{-1}(1-\xi)$ and penalty loading $\widehat{\Gamma}_\tau = \text{diag}[\hat{\Gamma}_{\tau kk}, k \in [p] \setminus \{j\}]$ is a diagonal matrix defined by the the following procedure: (1) Compute the Post Lasso estimator $\widetilde{\theta}_\tau^0$ based on λ and initial values $\hat{\Gamma}_{\tau jj} = \max_{i\le n} f_{ia\tau}\{\mathbb{E}_n[Z_{iaj}^2 Z_{iak}^2]\}^{1/2}$. (2) Compute the residuals $\widehat{v}_i = f_{ia\tau}(Z_{iaj} - Z'_{ia\setminus\{j\}}\widetilde{\theta}_\tau^0)$ and update

$$\hat{\Gamma}_{\tau kk} = \sqrt{\mathbb{E}_n[f^2_{ia\tau} Z^2_{iak} \widehat{v}^2_i]},\ k \in [p] \setminus \{j\}. \tag{3.14}$$

Finally, Step 3 uses $\mathcal{A}_\tau = \{\alpha \in \mathbb{R} : |\alpha - \widetilde{\alpha}_\tau| \leq 10\{\mathbb{E}_n[Z^2_{iaj}]\}^{-1/2}/\log n\}$.

Algorithm 1 (Conditional Independence Quantile Graphical Model)
For each $a \in V$, and $j \in [p]$, and $\tau \in \mathcal{T}$, perform the following:

1. Run Post-ℓ_1-quantile regression of X_a on Z_{aj} and $Z_{a\setminus\{j\}}$;
   keep fitted value $Z'_{a\setminus\{j\}}\widetilde{\beta}_\tau$,

$$\begin{aligned} (\hat{\alpha}_\tau, \hat{\beta}_\tau) &\in \arg\min_{\alpha,\beta} \mathbb{E}_n[\rho_\tau(X_{ia} - Z_{iaj}\alpha - Z'_{ia\setminus\{j\}}\beta)] + \lambda_I \sqrt{\tau(1-\tau)}\|\beta\|_1 \\ (\widetilde{\alpha}_\tau, \widetilde{\beta}_\tau) &\in \arg\min_{\alpha,\beta} \mathbb{E}_n[\rho_\tau(X_{ia} - Z_{iaj}\alpha - Z'_{ia\setminus\{j\}}\beta)] \ : \\ &\qquad support(\beta) \subseteq support(\hat{\beta}^{(2s)}_\tau). \end{aligned}$$

2. Run Post-Lasso of $f_{ia\tau}Z_{iaj}$ on $f_{ia\tau}Z_{ia\setminus\{j\}}$;
   keep the residual $\widetilde{v}_i := f_{ia\tau}(Z_{iaj} - Z'_{ia\setminus\{j\}}\widetilde{\theta}_\tau)$,

$$\begin{aligned} \hat{\theta}_\tau &\in \arg\min_\theta \mathbb{E}_n[f^2_{ia\tau}(Z_{iaj} - Z'_{ia\setminus\{j\}}\theta)^2] + \lambda\|\widehat{\Gamma}_\tau\theta\|_1 \\ \widetilde{\theta}_\tau &\in \arg\min_\theta \mathbb{E}_n[f^2_{ia\tau}(Z_{iaj} - Z'_{ia\setminus\{j\}}\theta)^2] \ : \ support(\theta) \subseteq support(\hat{\theta}_\tau). \end{aligned}$$

3. Run Instrumental Quantile Regression of $X_{ia} - Z'_{ia\setminus\{j\}}\widetilde{\beta}_\tau$ on Z_{iaj}
   using $\widetilde{v}_i$ as the instrument for Z_{iaj},

$$\begin{aligned} &\check{\beta}^a_j(\tau) \in \arg\min_{\alpha\in\mathcal{A}_\tau} L_n(\alpha), \\ &\text{where } L_n(\alpha) := \frac{\{\mathbb{E}_n[(1\{X_{ia} \leq Z_{iaj}\alpha + Z'_{ia\setminus\{j\}}\widetilde{\beta}_\tau\} - \tau)\widetilde{v}_i]\}^2}{\mathbb{E}_n[(1\{X_{ia} \leq Z_{iaj}\alpha + Z'_{ia\setminus\{j\}}\widetilde{\beta}_\tau\} - \tau)^2\widetilde{v}^2_i]}. \end{aligned}$$

Algorithm 1 above has been studied in (Belloni et al., 2013b) when it is applied to a single triple (a, j, τ). Under similar conditions, results that hold uniformly over $(a, j, \tau) \in V \times [p] \times \mathcal{T}$ are achieved based on the tools developed in (Belloni and Chernozhukov, 2011) and (Chernozhukov et al., 2012). Algorithm 1 is tailored to achieve good rates of convergence in the ℓ_∞-norm. In particular, under regularity conditions with probability going to 1 we have

$$\sup_{\tau\in\mathcal{T}} \|\beta^a(\tau) - \check{\beta}^a(\tau)\|_\infty \lesssim \sqrt{\frac{\log(p \vee n)}{n}}.$$

In order to create an estimate of $E^I(\tau) = \{(a,b) \in V \times V : \max_{j\in I_a(b)} |\beta^a_j(\tau)| > 0\}$, we define

$$\hat{E}^I(\tau) = \left\{ (a,b) \in V \times V : \max_{j\in I_a(b)} \frac{|\check{\beta}^a_j(\tau)|}{\text{se}(\check{\beta}^a_j(\tau))} > \bar{\lambda} \right\}$$

where $\text{se}(\check{\beta}^a_j(\tau)) = \{\tau(1-\tau)\mathbb{E}_n[\widetilde{v}_i^2]\}^{1/2}$ is an estimate of the standard deviation of the estimator, and $\bar{\lambda}$ is set to be of the order of $\sqrt{\log(p \vee n)/n}$ to account for the uniformity over $a \in V$, $j \in [p]$, and $\tau \in \mathcal{T}$.

Remark 3.3.1 (Stepdown procedure for $\bar{\lambda}$). Setting a critical value $\bar{\lambda}$ that accounts for the multiple hypothesis that are being tested plays an important role to select the graph $\hat{E}^I(\tau)$. Further improvements can be obtained by considering the stepdown procedure of (Romano and Wolf, 2005) for multiple hypothesis testing that was studied for the high-dimensional case in (Chernozhukov et al., 2013). The procedure iteratively creates a suitable sequence of decreasing critical values. In each step only null hypotheses that were not rejected are considered to determine the critical value. Thus, as long as any hypothesis is rejected at a step, the critical value decreases and we continue to the next iteration. The procedure stops when no hypothesis in the current active set is rejected.

3.3.2 Estimators for Prediction Quantile Graphs

Next we discuss the specification and propose an estimator for PQGMs. In this case we are interested on studying prediction of X_a, $a \in V$, using a linear combination of $X_{V\setminus\{a\}}$ under the asymmetric loss discussed in (3.6). Given the loss function ρ_τ, the target d-dimensional vector of parameters $\beta^a(\tau)$ is defined as (part of) the solution of the following optimization problem

$$(\alpha^a(\tau), \beta^a(\tau)) \in \arg\min_{\alpha,\beta} \; E[\rho_\tau(X_a - \alpha - X'_{V\setminus\{a\}}\beta)]. \tag{3.15}$$

By considering the case that d is large, the use of high-dimensional tools to achieve good estimators is of interest. The estimation procedure we propose is based on ℓ_1-penalized quantile regression. Again we consider models that satisfy an approximately sparse condition. Formally, we require the existence of sparse coefficients $\{\bar{\beta}^a(\tau) : a \in V, \tau \in \mathcal{T}\}$ such that

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\|\bar{\beta}^a(\tau)\|_0 \leq s \quad \text{and} \quad \max_{a\in V}\sup_{\tau\in\mathcal{T}}\{E[\{X'_{V\setminus\{a\}}(\beta^a(\tau)-\bar{\beta}^a(\tau))\}^2]\}^{1/2} \lesssim \sqrt{s/n}, \quad (3.16)$$

where the sparsity parameter s of the model is allowed to grow as n grows. A key issue is to set the penalty parameter properly so that it upper bounds

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\max_{j\in V\setminus\{a\}} \frac{|\mathbb{E}_n[(1\{X_{ia} \leq \alpha^a(\tau) + X'_{iV\setminus\{a\}}\beta^a(\tau)\} - \tau)X_{ij}]|}{\hat{\sigma}_j\sqrt{\tau(1-\tau)}} \quad (3.17)$$

where $\hat{\sigma}_j = \{\mathbb{E}_n[X_{ij}^2]\}^{1/2}$. However, it is important to note that we do not assume that the conditional quantile of X_a is a linear function of $X_{V\setminus\{a\}}$. Thus the penalty parameter in the penalized quantile regression needs to account for such misspecification and is no longer pivotal as in (Belloni and Chernozhukov, 2011).

In order to handle this issue we make a two step estimation. In the first step the penalty parameter λ_0 is conservative and set via bounds constructed based on symmetrization arguments, see (van de Geer, 2008; Belloni et al., 2013a). The second steps uses the preliminary estimator to bootstrap (3.17) based on the tools in (Chernozhukov et al., 2013). The following algorithm states the procedure.

Under correct linear specification, ℓ_1-QR has been studied in (Belloni and Chernozhukov, 2011). The work (Belloni et al., 2013b) allows for a vanishing approximation error. It was shown to achieve good rates of convergence in the ℓ_2-norm. In particular, under regularity conditions with probability going to 1 we have

$$\max_{a\in V}\sup_{\tau\in\mathcal{T}}\|\beta^a(\tau) - \check{\beta}^a(\tau)\| \lesssim \sqrt{\frac{s\log(d\vee n)}{n}}.$$

Algorithm 2 (Predictive Quantile Graph Model)
For each $a \in V$, and $\tau \in \mathcal{T}$, perform the following:

1. Run ℓ_1-quantile regression of X_a on $X_{V\setminus\{a\}}$ with penalty λ_0

$$(\hat{\alpha}^a_\tau, \hat{\beta}^a_\tau) \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_{ia} - \alpha - X'_{iV\setminus\{a\}}\beta)] + \lambda_0 \textstyle\sum_{j\in V\setminus\{a\}} \hat{\sigma}_j|\beta_j|$$

   where $\hat{\sigma}_j = \{\mathbb{E}_n[X^2_{ij}]\}^{1/2}$.

2. Set $\hat{\varepsilon}_{ia\tau} = 1\{X_{ia} \leq \hat{\alpha}^a_\tau + X'_{iV\setminus\{a\}}\hat{\beta}^a_\tau\} - \tau$ for $i \in [n]$, $a \in V$ and $\tau \in \mathcal{T}$. Compute the penalty level $\lambda_B = (1-\xi)$-quantile of W where

$$W := \max_{a\in V} \sup_{\tau\in\mathcal{T}} \max_{j\in V\setminus\{a\}} \frac{|\mathbb{E}_n[\hat{\varepsilon}_{ia\tau} e_i X_{ij}]|_\infty}{\hat{\sigma}_j\sqrt{\tau(1-\tau)}}$$

   where $\{e_i : i = 1, \ldots, n\}$ is a sequence of i.i.d. standard Gaussian random variables. For each $a \in V$, and $\tau \in \mathcal{T}$, perform the following:

3. Run ℓ_1-quantile regression of X_a on $X_{V\setminus\{a\}}$ with penalty λ_B

$$\begin{aligned}&(\check{\alpha}^a(\tau), \check{\beta}^a(\tau)) \in \arg\min_\beta \mathbb{E}_n[\rho_\tau(X_{ia} - \alpha - X'_{iV\setminus\{a\}}\beta)]\\ &+\lambda_B\sqrt{\tau(1-\tau)}\textstyle\sum_{j\in V\setminus\{a\}} \hat{\sigma}_j|\beta_j|\end{aligned}$$

The estimate of the prediction quantile graph is given by

$$\hat{E}^P(\tau) = \left\{(a,b) \in V \times V :\ |\check{\beta}^a_b(\tau)| > 0\right\},$$

that is, it is induced by the covariates selected by the ℓ_1-penalized estimator.

3.3.3 Conditional Quantile Graph Models

In order to handle the additional conditional event $Z \in \mathcal{Z}$ we propose to modify the Algorithms 1 and 2 based on kernel smoothing. To that extent, we assume that the observed data is of the form $\{(X_{iV}, Z_i) : i = 1, \ldots, n\}$, where Z_i might be defined through additional variables. Furthermore, we assume that for each $Z \in \mathcal{Z}$ we have access to a kernel function K_Z.

Example 3.3.4 (Predictive QGMs of Stock Returns Under Downside Market Movements,

Algorithm 1' (Z-Conditional Independence Quantile Graphical Model)
For each $a \in V$, and $j \in [p]$, $\tau \in \mathcal{T}$, and $Z \in \mathcal{Z}$, perform the following:

1. Run (local) Post-ℓ_1-quantile regression of X_a on Z_{aj} and $Z_{a\backslash\{j\}}$; keep fitted value $Z'_{a\backslash\{j\}}\widetilde{\beta}_\tau$,

$$\begin{aligned}&(\hat{\alpha}_\tau, \hat{\beta}_\tau) \in \arg\min_{\alpha,\beta} \mathbb{E}_n[K_Z(Z_i)\rho_\tau(X_{ia} - Z_{iaj}\alpha - Z'_{ia\backslash\{j\}}\beta)] + \lambda_I\sqrt{\tau(1-\tau)}\|\beta\|_1\\ &(\widetilde{\alpha}_\tau, \widetilde{\beta}_\tau) \in \arg\min_{\alpha,\beta} \mathbb{E}_n[K_Z(Z_i)\rho_\tau(X_{ia} - Z_{iaj}\alpha - Z'_{ia\backslash\{j\}}\beta)]\\ &\quad \textit{with support}(\beta) \subseteq \textit{support}(\hat{\beta}_\tau^{(2s)}).\end{aligned}$$

2. Run (local) Post-Lasso of $f_{ia\tau}Z_{iaj}$ on $f_{ia\tau}Z_{ia\backslash\{j\}}$; keep the residual $\widetilde{v}_i := f_{ia\tau}(Z_{iaj} - Z'_{ia\backslash\{j\}}\widetilde{\theta}_\tau)$,

$$\begin{aligned}&\hat{\theta}_\tau \in \arg\min_\theta \mathbb{E}_n[K_Z(Z_i)f_{ia\tau}^2(Z_{iaj} - Z'_{ia\backslash\{j\}}\theta)^2] + \lambda\|\widehat{\Gamma}_\tau\theta\|_1\\ &\widetilde{\theta}_\tau \in \arg\min_\theta \mathbb{E}_n[f_{ia\tau}^2(Z_{iaj} - Z'_{ia\backslash\{j\}}\theta)^2]\\ &\quad \textit{with support}(\theta) \subseteq \textit{support}(\hat{\theta}_\tau).\end{aligned}$$

3. Run (local) Instrumental Quantile Regression of $X_{ia} - Z'_{ia\backslash\{j\}}\widetilde{\beta}_\tau$ on Z_{iaj} using $\widetilde{v}_i$ as the instrument for Z_{iaj},

$$\begin{aligned}&\check{\beta}_j^a(\tau) \in \arg\min_{\alpha\in\mathcal{A}_\tau} L_n(\alpha),\\ &\text{where } L_n(\alpha) := \frac{\{\mathbb{E}_n[K_Z(Z_i)(1\{X_{ia} \le Z_{iaj}\alpha + Z'_{ia\backslash\{j\}}\widetilde{\beta}_\tau\} - \tau)\widetilde{v}_i]\}^2}{\mathbb{E}_n[K_Z^2(Z_i)(1\{X_{ia} \le Z_{iaj}\alpha + Z'_{ia\backslash\{j\}}\widetilde{\beta}_\tau\} - \tau)^2\widetilde{v}_i^2]}.\end{aligned}$$

continued). In Example 3.2.1, we have $Z_i = M_i$ denote the market return and the conditioning event to be $Z = 1\{M \le m\}$. We might be interest on a fixed m or on a family of values $m \in (-\bar{m}, 0]$. The latter induces $\mathcal{Z} = \{\ \{M \le m\} : m \in (-\bar{m}, 0]\}$. The kernel function is simply $K_Z(t) = 1\{t \le m\}/\sum_{i=1}^n 1\{Z_i \le m\}$.

3.4 Simulations of Predictive Quantile Graph Models

In this section we perform numerical simulations to illustrate the performance of the estimators for PQGMs. We will consider several different designs. In order to compare with other proposals we will consider Gaussian and non-Gaussian examples.

Algorithm 2' (Z-Conditional Predictive Quantile Graph Model)
For each $a \in V$, $\tau \in \mathcal{T}$, and $Z \in \mathcal{Z}$ perform the following:

1. Run (local) ℓ_1-quantile regression of X_a on $X_{V\setminus\{a\}}$ with penalty λ_0

$$
\begin{aligned}
&(\hat{\alpha}^a_\tau, \hat{\beta}^a_\tau) \in \arg\min\nolimits_\beta \mathbb{E}_n[K_Z(Z_i)\rho_\tau(X_{ia} - \alpha - X'_{iV\setminus\{a\}}\beta)] \\
&+\lambda_0 \textstyle\sum_{j\in V\setminus\{a\}} \hat{\sigma}_j|\beta_j|
\end{aligned}
$$

   where $\hat{\sigma}_j = \{\mathbb{E}_n[K_Z(Z_i)X_{ij}^2]\}^{1/2}$.

2. Set $\hat{\varepsilon}_{ia\tau} = 1\{X_{ia} \leq \hat{\alpha}^a_\tau + X'_{iV\setminus\{a\}}\hat{\beta}^a_\tau\} - \tau$ for $i \in [n]$, $a \in V$ and $\tau \in \mathcal{T}$. Compute the penalty level $\lambda_B = (1-\xi)$-quantile of W where

$$
W := \max_{a\in V} \sup_{\tau\in\mathcal{T}} \max_{j\in V\setminus\{a\}} \frac{|\mathbb{E}_n[K_Z(Z_i)\hat{\varepsilon}_{ia\tau}e_iX_{ij}]|_\infty}{\hat{\sigma}_j\sqrt{\tau(1-\tau)}}
$$

   where $\{e_i : i = 1, \ldots, n\}$ is a sequence of i.i.d. standard Gaussian random variables. For each $a \in V$, and $\tau \in \mathcal{T}$, perform the following:

3. Run (local) ℓ_1-quantile regression of X_a on $X_{V\setminus\{a\}}$ with penalty λ_B

$$
\begin{aligned}
&(\check{\alpha}^a(\tau), \check{\beta}^a(\tau)) \in \arg\min\nolimits_\beta \mathbb{E}_n[K_Z(Z_i)\rho_\tau(X_{ia} - \alpha - X'_{iV\setminus\{a\}}\beta)] \\
&+\lambda_B\sqrt{\tau(1-\tau)} \textstyle\sum_{j\in V\setminus\{a\}} \hat{\sigma}_j|\beta_j|
\end{aligned}
$$

3.4.1 Isotropic Non-Gaussian Example

The equivalence between a zero in the inverse covariance matrix and a pair of conditional independent variables break down for non-gaussian distribution. The nonparanormal extends Gaussian graphical models to semiparametric Gaussian copula models by transforming the variables by smooth functions. We illustrate the applicability of QGM in representing the independence structure of a set of variables when the random variables are not jointly (non-para)normal.

Consider i.i.d. copies of an d-dimensional random vector $W = (W_1, \ldots, W_d)$ from the following multivariate normal distribution, $W \sim N(0, I_{d\times d})$, where $I_{d\times d}$ is the identity matrix. Further, we generate

$$
Y = -\sqrt{\tfrac{2}{3\pi-2}} + \sqrt{\tfrac{\pi}{3\pi-2}}W_{d-1}^2|W_d|. \tag{3.18}
$$

It follows that $E[Y] = \sqrt{\frac{\pi}{3\pi-2}}(E[|W_d|] - \sqrt{2/\pi}) = 0$ and $Var(Y) = \frac{\pi}{3\pi-2}(E[W_d^2 \cdot W_{d-1}^4] - \frac{2}{\pi}) = 1$. In addition, equation (3.18) is a location-scale-shift model in which the conditional median of the response is zero while quantile functions other than the median are nonzero. We define the vector X_V as

$$X_V = (W_1, ..., W_{d-1}, Y)'.$$

In this new set of variables, only X_{d-1} and X_d (i.e. W_{d-1} and Y) are not (conditionally) independent. Nonetheless, the new covariance matrix of X_V is still $I_{d\times d}$.[4]

Next we consider an i.i.d. sample with a sample size of $n = 300$ and $d = 15$. We show graphs of independence structure estimated by using both the GGM and QGM(s) in this the non-Gaussian setting,

Gaussian is estimated by using graphical lasso without any transformation of X_V, and the final graph is chosen by Extended Bayesian information criterion (ebic), see (Foygel and Drton, 2010). Nonparanormal is estimated by using graphical lasso with nonparanormal transformation of X_V, see (Liu et al., 2009), and the final graph is chosen by ebic. Both graphs are estimated by using R-package **huge**.

We also compare our estimation results using QGM with neighborhood selection methods, e.g. TIGER of (Liu and Wang, 2012) in R-package **flare**, the left graph is when choosing the turning parameter to be $\sqrt{\frac{\log d}{n}}$ while the right graph is when choosing the tuning parameter to be $2\sqrt{\frac{\log d}{n}}$. Throughout, we use Tiger2 (or TIGER2) represent TIGER with penalty level $2\sqrt{\frac{\log d}{n}}$.

As expected, GGM cannot detect the correct dependence structure when the joint distribution is non-Gaussian while QGM can still represent the right independence structure.

[4]Indeed, for any $k \leq d-1$ we have

$$\begin{aligned} E[X_d \cdot X_k] &= E[Y \cdot W_k] = E[W_k \cdot (-\sqrt{\tfrac{2}{3\pi-2}} + \tfrac{1}{\sqrt{3-2/\pi}} W_k^2 |W_d|)] \\ &= \sqrt{\tfrac{\pi}{3\pi-2}} E[|W_d| W_k^3] - \sqrt{\tfrac{2}{3\pi-2}} E[W_k] = 0. \end{aligned}$$

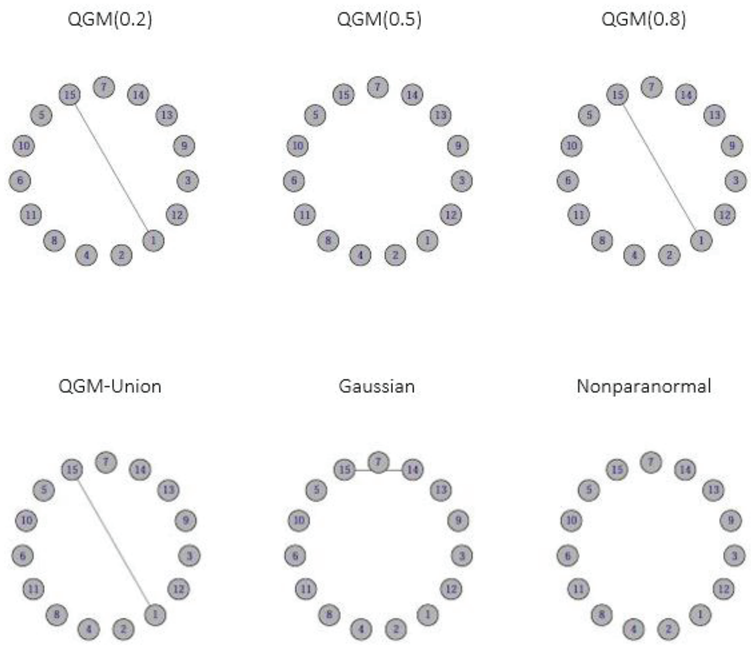

Figure 3.1: QGM(s) and GGM

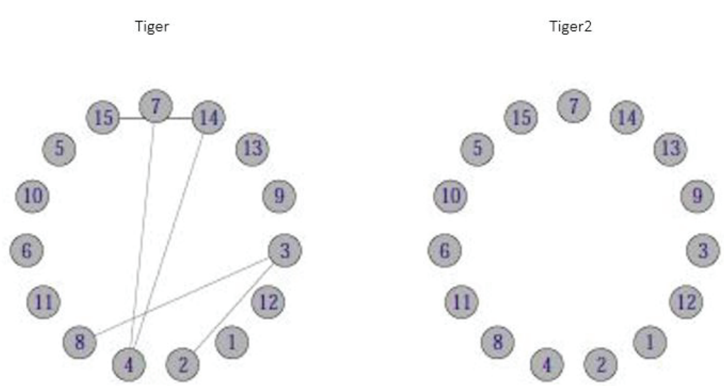

Figure 3.2: Tiger1 and Tiger2

3.5 Empirical Applications of QGM

3.5.1 Financial Contagion

In this section we apply QGM for the study of international financial contagion. We focus on examining financial contagion through the volatility spillover perspective. (Engle and Susmel, 1993) reported that international stock markets are related through their volatilities instead of returns. (Diebold and Yilmaz, 2009) studied the return and volatility spillovers of 19 countries and found differences in return and volatility spillovers. For a survey of financial contagion see (Claessens and Forbes, 2001). We also illustrate how QGM can highlight asymmetric dependence between the random variables.

We use daily equity index returns, September 2009 to September 2013 (1044 observations), from Morgan Stanley Capital International (MSCI). The returns are all translated into dollar-equivalents as of September 6th 2013. We use absolute returns as a proxy for volatility. We have a total of 45 countries in our sample, there are 21 developed markets (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom, the United States), 21 emerging markets (Brazil, Chile, Mexico, Greece, Israel, China, Colombia, Czech Republic, Egypt, Hungary, India, Indonesia, Korea, Malaysia, Peru, Philippines, Poland, Russia, Taiwan, Thailand, Turkey), and 3 frontier markets (Argentina, Morocco, Jordan).

Below we provide a full-sample analysis of global volatility spillovers at different tails. We denote 20% quantile as Low Tail, 50% quantile as Median, 80% quantile as Up Tail. Both QGMs and GGM are estimated. Our purpose is to show the usefulness of QGM in representing nonlinear tail interdependence allowing for heteroskedasticity and to show that QGM measures correlation asymmetry by looking at behavior in the tails of the distribution (not specific to any model).

There are significant differences in the network structure in terms of volatility spillovers

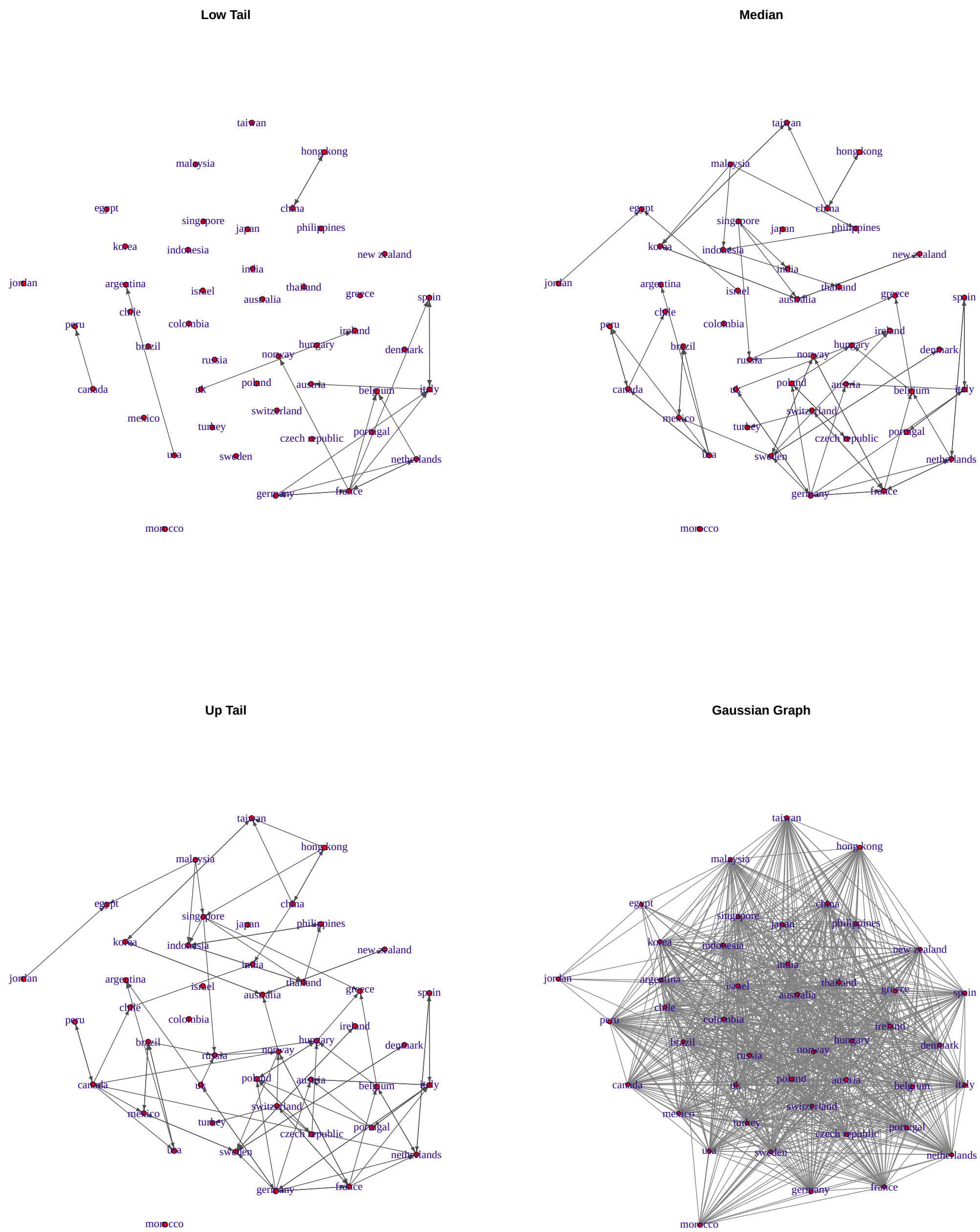

Figure 3.3: International Financial Contagion

when using QGM and Gaussian graph. QGM permits conditional asymmetries in correlation dynamics, suited to investigate the presence of asymmetric responses. We find significant increase at the up tail interdependence between the volatility series, i.e. we find downside correlation (high volatility) are much larger than upside correlation (low volatility). This confirms findings in finance literature that financial markets become more interdependent during high volatility periods.

We also find if two countries are located in the same geographic region, with many similarities in terms of market structure and history, they tend to be closely connected (the homophily effect as in network terminology); while two economies located in separate geographic regions are less likely directly connected. We find among European Union member countries, Germany appears to play a major role in the transmission of shocks to others. While in Asia, Hong Kong, Thailand, and Singapore appears to play a major role. Among all the north and south American countries, Canada and US play a major role in risk transmission.

We also report *net*-$\Delta CoVaR$ to measure spillover accounting for the network (see Appendix C.1) for the volatility series through QGM at up tail in Figure 3.4.

Figure 3.4 shows that, globally, total volatility spillovers from Germany, France, US and Hong Kong to the others are much larger than total volatility spillovers from the others to them; while the opposite happens to Greece and Spain. Both Greece and Spain receive larger volatility spillovers from others than contribute to the others. The estimated network structure is important here as it demonstrates that shocks originating in some stock markets, e.g. Germany and Hong Kong, may be amplified in their transmission throughout the system, posing greater risks to the whole market than other shock's origination.

3.5.2 Stock Returns Conditional on Market Downside Movement

Stock markets are in general non-Gaussian. (Ang and Chen, 2002) find correlation asymmetries in the data and reject the null hypothesis of multivariate normal distributions at

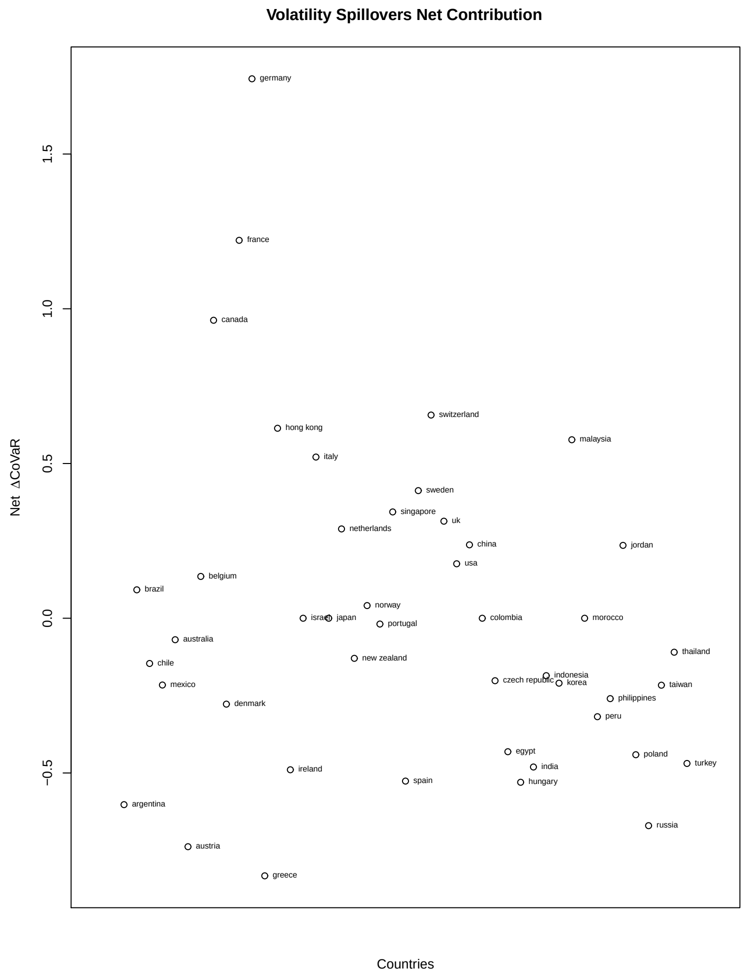

Figure 3.4: Net-$\Delta CoVaR$ of Each Country

daily, weekly, and monthly frequencies, conditional on market "downside" movements. See also (Longin and Solnik, 2001; Patton, 2006) among other studies in the empirical finance literature for the non-Gaussian feature of financial markets. Hence, generally in the financial market context, conditional correlation only conveys partial and often misleading information on the actual underlying conditional dependencies.

We contribute to the literature by showing the union of a set of QGMs can be used to obtain a conditional independence graph when the main interest lies in estimating the conditional independence structure of stocks under a market downturn. While the joint distribution of stocks considered is generally non-Gaussian, since QGM does not impose any parametric assumption on the joint distribution of stocks, the union of QGMs allows for both Gaussian and non-Gaussian joint distributions in estimating the conditional independence structure.

This will be modelled with a conditional quantile graph models. We consider the conditioning events to be $Z = \{\text{Market return} \leq m_u\}$ for we set $m_u = u$-th quantile of the market index return to capture downside movement of the market (note that $u = 1$ corresponds to regular market). We obtain daily stock returns from CRSP. The full sample consists of 2769 observations of daily stock returns for 86 stocks in the S&P 500 from Jan 2, 2003 to December 31, 2013. The total number of stocks is 86 due to data availability at CRSP. We define market downside as when the market index returns are below a pre-specified level and we use S&P 500 as market index. In this case, the conditioning on a particular Z corresponds simply to consider the subsample based on whether the corresponding date's market return is less equal to the u-th quantile of the market index returns. We reported the number of edges, there is no linkage between two stocks if there are conditional independent, at different subsamples in Table 3.1 below.

For estimators based on QGM and GGM, the number of edges increases with the quantile index. However, potentially due to asymmetry in relations, there are significant differences between the results of QGMs and GGM. There are significantly higher interdependence in GGM. Nonetheless, increase in conditional correlation could be a result of assuming conditional normality for the return distribution – estimation bias in correlation conditional on market upside or downside moves will cause false correlation. These empirical findings support evidence from the empirical finance literature.

Table 3.1: Edges Produced by Different Graph Estimators

Quantile of market index (u)	PQGM	Glasso(eBIC)	TGalasso	TIGER
0.15	406	1752	1804	3372
0.5	744	2152	2278	5734
0.75	842	2380	2478	6180
0.9	978	2461	2564	6344
1	1062	2518	2660	6290

Appendix A

Proofs for Chapter 1

A.1 Results of Section 1.2

A.1.1 Panel probit with additive individual and time effects

In this setting, I consider the following model

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \alpha_i + \gamma_t + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{A.1}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, β is a finite-dimensional parameter vector, the variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks respectively.

Similarly to Section (1.2.1), I model the conditional distribution of Y_{it} using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i, \gamma_t) = \Phi(X_{it}\beta + \alpha_i + \gamma_t),$$

and for estimation I adopt a fixed effects approach treating the unobserved individual and time effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, ..., \alpha_N, \gamma_1, ..., \gamma_N)'$. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0, \gamma_1^0, ..., \gamma_T^0)'$. Other quantities of interest involve averages over the data and unobserved effects

$$\delta^0_{NT} = \mathbb{E}_\phi[\Delta_{NT}(\beta^0, \phi^0_{NT})], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1}\sum_{i,t}\Delta(X_{it}, \beta, \alpha_i, \gamma_t), \tag{A.2}$$

and examples of partial effects (Δ) are the following:

Example A.1.1. (Average partial effects) If $X_{it,k}$, the k-th element of X_{it}, is binary, its partial effect for model (A.1) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha_i + \gamma_t) = \Phi(\beta_k + X'_{it,-k}\beta_{-k} + \alpha_i + \gamma_t) - \Phi(X'_{it,-k}\beta_{-k} + \alpha_i + \gamma_t), \tag{A.3}$$

where β_k is the k-th element of β, and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the k-th element. If $X_{it,k}$ is continuous, for model (A.1) the partial effects of $X_{it,k}$ on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i, \gamma_t) = \beta_k \phi_f(X'_{it}\beta + \alpha_i + \gamma_t), \tag{A.4}$$

where $\phi_f(\cdot)$ is the derivative of Φ.

Definition A.1.1. The fixed effect EM estimator for panel probit with additive fixed effects is defined by

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)} + \alpha_i^{(k)} + \gamma_t^{(k)}$,

(2) E-step: Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} : &= E[Y_{it}^*|Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)})/\{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)})\}, \end{aligned}$$

(3) **M-step**: This contains three conditional maximization steps

CM-step 1: Given α_i and γ_t, the parameter β can be updated by

$$\beta^{(k+1)} = (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}')^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)} - \gamma_t^{(k)})\},$$

CM-step 2: Given β and γ_t, the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_{it}^{(k)} - X_{it}'\beta^{(k+1)} - \gamma_t^{(k)}),$$

CM-step 3: Given β and α_i, the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_{it}^{(k)} - X_{it}'\beta^{(k+1)} - \alpha_i^{(k+1)})$$

(4) Iterate until convergence.

Note that the CM-step 2 and CM-step 3 here are just the average over time and individual using $\hat{Y}_{it}^{(k)}$ as surrogate for Y_{it}^*.

A.1.2 Proof of Proposition 1.2.1

By second-order Taylor expansion, for any two arguments θ_1 and θ_2,

$$-logF(\theta_1) = -logF(\theta_2) - \frac{\partial logF(\theta_2)}{\partial\theta_2}(\theta_1 - \theta_2) - \frac{1}{2}\frac{\partial^2 logF(\theta)}{\partial^2\theta}|_{\theta^*}(\theta_1 - \theta_2)^2.$$

Denote $h(\theta) = -\frac{\partial logF(\theta)}{\partial\theta}$. Using the fact that $-logF(q_{it}z_{it})$ is strictly convex on $(0,1)$ for logit and probit, and simple calculation shows $0 < -\frac{\partial^2 logF(\theta)}{\partial^2\theta}|_{\theta^*} < 1$, one has

$$-logF(\theta_1) \le -logF(\theta_2) + h(\theta_2)(\theta_1 - \theta_2) + \frac{1}{2}(\theta_1 - \theta_2)^2,$$

by completing the square, this can be written as

$$-logF(\theta_1) \le -logF(\theta_2) + \frac{1}{2}(\theta_1 - \theta_2 + h(\theta_2))^2 - \frac{1}{2}h^2(\theta_2).$$

Now substitute $q_{it}(X_{it}'\beta+\alpha_i'\gamma_t)$ for θ_1 and $q_{it}(X_{it}'\tilde{\beta}+\tilde{\alpha}_i'\tilde{\gamma}_t)$ for θ_2, one has

$$\begin{aligned}-logF(q_{it}(X_{it}'\beta+\alpha_i'\gamma_t)) \quad &\leq \quad -logF(q_{it}(X_{it}'\tilde{\beta}+\tilde{\alpha}_i'\tilde{\gamma}_t)-\frac{1}{2}h^2(q_{it}(X_{it}'\tilde{\beta}+\tilde{\alpha}_i'\tilde{\gamma}_t)) \\ &+\frac{1}{2}((X_{it}'\beta+\alpha_i'\gamma_t)-(X_{it}'\tilde{\beta}+\tilde{\alpha}_i'\tilde{\gamma}_t)+q_{it}h(q_{it}(X_{it}'\tilde{\beta}+\tilde{\alpha}_i'\tilde{\gamma}_t)))^2\end{aligned}$$

sum over i and t to obtain the required results.

A.2 Proofs of Section 1.3

A.2.1 Proof of Consistency for $\hat{\beta}_{PPIF}$

The proof contains two steps. In Step 1, I show the estimated index $\tilde{z}_{it}$ is a good approximation to z_{it} with some structural error (the generalized residuals). In Step 2, I show the structural error satisfies the assumption in (Bai, 2009b) for linear panel data models with interactive fixed effects. With a little abuse of notation, in this section I use $\hat{\beta}$ to denote $\hat{\beta}_{PPIF}$ which is the estimate of the EM procedure for panel probit models.

Step 1. Denote $q_{it}=2Y_{it}-1$. I prove the consistence directly from the likelihood function

$$\ell_{it}(\beta,\alpha_i,\gamma_t) \quad = \quad log\Phi(q_{it}(X_{it}^{'}\beta+\alpha_i\gamma_t)), \quad \mathcal{L}_{NT}=\frac{1}{NT}\sum_{i,t}\ell_{it}=\sum_{i,t}log\Phi(q_{it}z_{it}),$$

for any θ_1 and θ_2, the following is an upper bound for the negative log-likelihood:

$$\begin{aligned}-log\Phi(\theta_1) \quad &\leq \quad -log\Phi(\theta_2)-\frac{\phi_f(\theta_2)}{\Phi(\theta_2)}(\theta_1-\theta_2)+\frac{1}{2}(\theta_1-\theta_2)^2 \\ &= \quad -log\Phi(\theta_2)+\frac{1}{2}(\theta_1-\theta_2-\frac{\phi_f(\theta_2)}{\Phi(\theta_2)})^2-\frac{1}{2}(\frac{\phi_f(\theta_2)}{\Phi(\theta_2)})^2,\end{aligned}$$

where $\phi_f(\cdot)$ is the Gaussian density. Substitute $q_{it}z_{it}$ for θ_1 and $q_{it}\tilde{z}_{it}$ for θ_2, then

$$-log\Phi(q_{it}z_{it}) \leq -log\Phi(q_{it}\tilde{z}_{it}) + \frac{1}{2}(z_{it} - \tilde{z}_{it} + q_{it}\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})})^2 - \frac{1}{2}(\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})})^2. \quad (A.5)$$

Note, from the proof here, one can also infer using $\tilde{z}_{it} = z_{it} + q_{it}\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})} = z_{it} + \frac{Y_{it}-\Phi(z_{it})}{\Phi(z_{it})(1-\Phi(z_{it}))}\phi_f(q_{it}z_{it})$ is a good next step approximation, as the quadratic loss is a surrogate for the Bernoulli log-likelihood function.

Step 2. Denote the structural error (generalized residual) as $e_{it} = \frac{Y_{it}-\phi_f(z_{it})}{\Phi(z_{it})\Phi(z_{it})}\phi_f(q_{it}z_{it})$. One has $\mathbb{E}_\phi[e_{it}] = 0$. Since the estimated parameters minimize the objective function, with equation (A.5) one has

$$0 \geq \mathcal{L}_{NT}(\beta^0, \phi^0) - \mathcal{L}_{NT}(\hat{\beta}, \hat{\phi}) \geq \frac{1}{2NT}\sum_{i,t}[(z_{it}^0 - \hat{z}_{it} + e_{it})^2 - e_{it}^2]$$

The consistency proof for $\hat{\beta}$ is equivalent to that for the linear regression model with interactive fixed effects. In matrix notation, as in Section 1.4, the above inequality would be

$$\begin{aligned}\frac{1}{NT}Tr(e'e) &\geq \frac{1}{NT}Tr[(X'(\hat{\beta} - \beta^0) + \hat{\alpha}\hat{\gamma} - \alpha^0\gamma^0 - e)'(X'(\hat{\beta} - \beta^0) + \hat{\alpha}\hat{\gamma}' - \alpha^0\gamma^0 - e)] \\ &\geq \frac{1}{NT}Tr[(X'(\hat{\beta} - \beta^0) - e)'M_{(\hat{\alpha},\alpha^0)}(X'(\hat{\beta} - \beta^0) - e)]\end{aligned}$$

where $M_{(\hat{\alpha},\alpha^0)} = \mathbf{1}_T - (\hat{\alpha}, \alpha^0)[(\hat{\alpha}, \alpha^0)'(\hat{\alpha}, \alpha^0)]^{-1}(\hat{\alpha}, \alpha^0)'$ is the projector that projects orthogonal to $(\hat{\alpha}, \alpha^0)$.

With Assumption 1 (iv), which says that no linear combination of the regressors converges to zero, even after projecting any factor loading α, one has $\frac{1}{NT}Tr(Xe') = o_p(1)$, and $\mathbb{E}[e_{it}] = 0$. One can also check that $\|e\| = o_p(\sqrt{NT})$. The assumption $\frac{1}{NT}Tr(XX') = O_p(1)$ is satisfied from the distributional assumption on the regressors above. One then has

$$|\frac{1}{NT}Tr(e'M_{(\hat{\alpha},\alpha^0)}X_k)| \leq \frac{1}{NT}|Tr(e'X_k)| + \frac{1}{NT}|Tr(e'P_{(\hat{\alpha},\alpha_0)}X_k)|$$

$$\leq \quad o_p(1) + \frac{2}{NT}\|e\|\|X_k\| = o_p(1).$$

Under these, one has

$$0 \geq c\|\hat{\beta} - \beta\| + o_p\|\hat{\beta} - \beta^0\| + o_p(1),$$

from which it is concluded that $\hat{\beta} = \beta^0 + o_p(1)$.

A.2.2 Proofs of Theorems 1.3.1 and 1.3.2

In the section, I suppress the dependence on NT of all the sequences of functions and parameters to lighten the notation, e.g. I write $\mathcal{L}$ for $\mathcal{L}_{NT}$ and ϕ for ϕ_{NT}. It is also convenient to introduce some notation that will be extensively used in the analysis. Let

$$\mathcal{S}(\beta, \phi) = \partial_\phi \mathcal{L}(\beta, \phi) \qquad\qquad \mathcal{H}(\beta, \phi) = -\partial_{\phi\phi'} \mathcal{L}(\beta, \phi),$$

where $\partial_x f$ denotes the partial derivative of f with respect to x, and additional subscripts denote higher-order partial derivatives. I refer to the dim ϕ -vector $\mathcal{S}(\beta, \phi)$ as the incidental parameter score, and to the dim $\phi\times$ dim ϕ matrix $\mathcal{H}(\beta, \phi)$ as the incidental parameter Hessian. I omit the argument of the functions when they are evaluated at the true parameter values (β^0, ϕ^0), e.g. $\mathcal{H} = \mathcal{H}(\beta^0, \phi^0)$. I use a bar to indicate expectations, e.g. $\partial_\beta \bar{\mathcal{L}} = \mathbb{E}[\partial_\beta \mathcal{L}]$, and a tilde to denote that the variables are in deviation with respect to their expectations, e.g. $\partial_\beta \bar{\mathcal{L}} = \partial_\beta \mathcal{L} - \partial_\beta \bar{\mathcal{L}}$. For $c \geq 0$, I define the sets $\mathcal{B}(c, \beta^0) = \{\beta : \|\beta - \beta^0\|_\infty \leq c\}$, and $\mathcal{B}_q(c, \beta^0, \phi^0) = \{(\beta, \phi) : \|\beta - \beta^0\| < c, \|\phi - \phi^0\|_q < c\}$, which are closed balls of radius c around the true parameters β^0 and (β^0, ϕ^0), respectively, under the L_2 norm and L_q-norm.

Analogous to Ξ_{it} defined in Eq (1.13), I define

$$\Lambda_{it} \quad = \quad -\frac{1}{NT}\sum_{j=1}^{N}\sum_{\tau=1}^{T}(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}\gamma^0_\tau\gamma^0_t + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}\alpha^0_j\gamma^0_t + \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\gamma^0_\tau\alpha^0_i + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\alpha^0_j\alpha^0_i)\partial_\pi \ell_{j\tau}$$

and analogous to $D_\beta \ell_{it}$ defined in the main text I also define $D_\beta \Delta_{it} = \partial_\beta \Delta_{it} - \partial_\pi \Delta_{it} \Xi_{it}$.

With a little abuse of notation, in this section I use $\hat{\beta}$ to denote $\hat{\beta}_{PPIF}$ which is the estimate of the EM procedure for panel probit models.

A close look at the iterative EM procedure yields

$$\begin{aligned} \hat{\beta}^{(k+1)} &= (\sum_{i,t} X_{it} X_{it}')^{-1} \sum_{i,t} X_{it} (\hat{Y}_{it}^{(k)} - \hat{\alpha}_i^{(k)} \gamma_t^{(k)}) \\ &= \beta^{(k)} + (X'X)^{-1} \partial_\beta \mathcal{L}(\beta^{(k)}, \hat{\phi}(\beta^{(k)})), \end{aligned} \tag{A.6}$$

which depends on the score of the profile likelihood function.

For $r \geq 0$, define the sets $\mathcal{B}(r, \beta^0) = \{\beta : \|\beta - \beta^0\| \leq r\}$, and $\mathcal{B}_q(r, \phi^0) = \{\phi : \|\phi - \phi^0\|_q \leq r\}$, which are closed balls of radius r around the true parameter values β^0 and ϕ^0, respectively.

Before going to the proof of Theorems 1.3.1 and 1.3.2, I first introduce two lemmas that will be used.

Lemma A.2.1. *(Asymptotic expansions of $\hat{\beta}$). Let Assumption 1 hold. Then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \overline{W}_\infty^{-1} U + o_p(1),$$

where $U = U^{(0)} + U^{(1)}$, $\overline{W}_\infty := \lim_{N,T\to\infty} \overline{W}$ exists with $\overline{W}_\infty > 0$, and

$$\begin{aligned} \overline{W} &= -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\mathbb{E}_\phi(\partial_{\beta\beta'} \ell_{it}) + \mathbb{E}_\phi(-\partial_{\pi^2} \ell_{it}) \Xi_{it} \Xi_{it}'], \\ U^{(0)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T D_\beta \ell_{it}, \\ U^{(1)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \{-\Lambda_{it} [D_{\beta\pi} \ell_{it} - \mathbb{E}(D_{\beta\pi} \ell_{it})] + \frac{1}{2} \Lambda_{it}^2 \mathbb{E}(D_{\beta\pi^2} \ell_{it})\}. \end{aligned}$$

Proof. The proof follows from using Theorem B.1 of (Fernández-Val and Weidner, 2013) and applying Lemma A.4.1. From Theorem B.1 of (Fernández-Val and Weidner, 2013),

$$\sqrt{NT}\partial_\beta \mathcal{L}(\beta, \hat{\phi}(\beta)) = U - \overline{W}\sqrt{NT}(\beta - \beta^0) + R(\beta),$$

with

$$\overline{W} = -(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}[\partial_{\phi\beta'}\overline{\mathcal{L}}]),$$

hence applying Lemma A.4.1 (ii) yields

$$\overline{W} \quad = \quad -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbb{E}_\phi(\partial_{\beta\beta'}\ell_{it}) + \mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})\Xi_{it}\Xi_{it}']. \qquad \text{(A.7)}$$

Similarly, applying Theorem B.1 of (Fernández-Val and Weidner, 2013) yields

$$U^{(0)} = \sqrt{NT}(\partial_\beta \mathcal{L} + [\partial_{\beta\phi'}\bar{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S}),$$

$$\begin{aligned} U^{(1)} \quad &= \quad \sqrt{NT}([\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S}) \\ &\quad + \sqrt{NT}\sum_{g=1}^{\dim\phi}(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}])[\overline{\mathcal{H}}^{-1}\mathcal{S}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g/2. \end{aligned}$$

By using Lemma A.4.1 (i),

$$U^{(0)} \quad = \quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\partial_\beta \ell_{it} - \Xi_{it}\partial_\pi \ell_{it}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}D_\beta \ell_{it}. \qquad \text{(A.8)}$$

Decompose $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$U^{(1a)} = \sqrt{NT}([\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S}),$$

and

$$U^{(1b)} = \sqrt{NT}\sum_{g=1}^{\dim\phi}(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}])[\overline{\mathcal{H}}^{-1}\mathcal{S}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g/2.$$

By using Lemma A.4.1 (i) and (iii),

$$U^{(1a)} = -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}(\partial_{\beta\pi}\tilde{\ell}_{it} + \Xi_{it}\partial_{\pi^2}\tilde{\ell}_{it}) = -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}[D_{\beta\pi}\ell_{it} - \mathbb{E}_{\phi}(D_{\beta\pi}\ell_{it})],$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}^2[\mathbb{E}_{\phi}(\partial_{\beta\pi^2}\ell_{it}) + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathbb{E}_{\phi}(\partial_{\phi}\partial_{\pi^2}\ell_{it})],$$

where for each i, t it is the case that $\partial_{\phi}\partial_{\pi^2}\ell_{it}$ is a dim ϕ-vector, which can be written as $\partial_{\phi}\partial_{\pi^2}\ell_{it} = \begin{pmatrix} A1_T \\ A'1_N \end{pmatrix}$ for an $N \times T$ matrix A with elements $A_{j\tau} = \partial_{\pi^3}\ell_{j\tau}$ if $j = i$ and $\tau = t$, and $A_{j\tau} = 0$ otherwise. Thus, again applying Lemma A.4.1(i) yields $[\partial_{\beta\phi'}\bar{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\partial_{\phi}\partial_{\pi^2}\ell_{it} = -\sum_{j,\tau}\Xi_{j\tau}\delta_{(i=j)}\delta_{(t=\tau)}\partial_{\pi^3}\ell_{it} = -\Xi_{it}\partial_{\pi^3}\ell_{it}$. Therefore

$$U^{(1b)} = \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}^2\mathbb{E}(\partial_{\beta\pi^2}\ell_{it} - \Xi_{it}\partial_{\pi^3}\ell_{it}) = \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}^2\mathbb{E}_{\phi}(D_{\beta\pi^2}\ell_{it}),$$

hence

$$U^{(1)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{-\Lambda_{it}[D_{\beta\pi}\ell_{it} - \mathbb{E}(D_{\beta\pi}\ell_{it})] + \frac{1}{2}\Lambda_{it}^2\mathbb{E}(D_{\beta\pi^2}\ell_{it})\}. \tag{A.9}$$

□

Lemma A.2.2. *(Asymptotic expansion of $\hat{\delta}$). Let Assumptions 1 and 2 hold and let $\|\hat{\beta} - \beta^0\| = O_p((NT)^{-1/2}) = o_p(r_\beta)$. Then*

$$\sqrt{NT}(\hat{\delta} - \delta) = V_{\Delta}^{(0)} + V_{\Delta}^{(1)} + o_p(1),$$

where

$$V_{\Delta}^{(0)} = [\frac{1}{NT}\sum_{i,t}\mathbb{E}_{\phi}(D_{\beta}\Delta_{it})]'\overline{W}_{\infty}^{-1}U^{(0)} - \frac{1}{\sqrt{NT}}\sum_{i,t}\mathbb{E}_{\phi}(\Psi_{it})\partial_{\pi}\ell_{it},$$

$$V_{\Delta}^{(1)} = [\frac{1}{NT}\sum_{i,t}\mathbb{E}_{\phi}(D_{\beta}\Delta_{it})]'\overline{W}_{\infty}^{-1}U^{(1)} + \frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}[\Psi_{it}\partial_{\pi^2}\ell_{it} - \mathbb{E}_{\phi}(\Psi_{it})\mathbb{E}_{\phi}(\partial_{\pi^2}\ell_{it})] + \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2[\mathbb{E}_{\phi}(\partial_{\pi^2}\ell_{it}) - \mathbb{E}_{\phi}(\partial_{\pi^3}\ell_{it})\mathbb{E}_{\phi}(\Psi_{it})].$$

Proof. The proof follows from using Theorem B.4 of (Fernández-Val and Weidner, 2013) and applying Lemma A.4.1. Theorem B.4 of (Fernández-Val and Weidner, 2013) implies

$$\hat{\delta} - \delta = [\partial_{\beta'}\overline{\Delta} + (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}(\partial_{\phi\beta'}\overline{\mathcal{L}})](\hat{\beta} - \beta^0) + U_{\Delta}^{(0)} + U_{\Delta}^{(1)} + o_p(1/\sqrt{NT}), \qquad \text{(A.10)}$$

with

$$U_{\Delta}^{(0)} = (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}\mathcal{S},$$

$$U_{\Delta}^{(1)} = (\partial_{\phi'}\tilde{\Delta})\overline{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\phi}\overline{\Delta})\overline{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} + \frac{1}{2}\mathcal{S}'\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'}\overline{\Delta} + \sum_{g=1}^{\dim\phi}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}(\partial_{\phi}\overline{\Delta})]_g]\overline{\mathcal{H}}^{-1}\mathcal{S}.$$

By using Lemma A.4.1,

$$\sqrt{NT}U_{\Delta}^{(0)} = -\frac{1}{\sqrt{NT}}\sum_{i,t}\mathbb{E}_{\phi}(\Psi_{it})\partial_{\pi}\ell_{it}, \qquad \text{(A.11)}$$

$$\sqrt{NT}U_{\Delta}^{(1)} = \frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}[\Psi_{it}\partial_{\pi^2}\ell_{it} - \mathbb{E}_{\phi}(\Psi_{it})\mathbb{E}_{\phi}(\partial_{\pi^2}\ell_{it})] + \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2[\mathbb{E}_{\phi}(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_{\phi}(\partial_{\pi^3}\ell_{it})\mathbb{E}_{\phi}(\Psi_{it})]. \qquad \text{(A.12)}$$

From the proof of Lemma A.2.1 and the following proof of Theorem 1.3.1, it follows that $\sqrt{NT}(\hat{\beta} - \beta^0) = \overline{W}_{\infty}^{-1}U + o_p(1) = O_p(1)$, by Lemma A.4.1,

$$\sqrt{NT}[\partial_{\beta'}\overline{\Delta} + (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}(\partial_{\phi\beta'}\overline{\mathcal{L}})](\hat{\beta} - \beta^0) = [\frac{1}{NT}\sum_{i,t}\mathbb{E}_{\phi}(D_{\beta}\Delta_{it})]'\overline{W}_{\infty}^{-1}(U^{(0)} + U^{(1)}) + o_p(1). \qquad \text{(A.13)}$$

Combining equations A.10, A.11, A.12 and A.13 gives the result. □

A.2.2.1 Proof of Asymptotics for $\hat{\beta}_{PPIF}$

I characterize the asymptotic distribution of $\hat{\beta}$ from the limit average Hessian $\overline{W}_\infty$ and the limiting distribution of the approximated score U. Next two steps are to get the eventual result.

Step 1 shows $U^{(0)} \xrightarrow{d} N(0, \overline{W}_\infty)$. In the likelihood setting $\mathbb{E}\partial_\beta \mathcal{L} = 0$, $\mathbb{E}\mathcal{S} = 0$, and, by the Bartlett identities $\mathbb{E}(\partial_\beta \mathcal{L} \partial_{\beta'} \mathcal{L}) = -\frac{1}{NT}\partial_{\beta\beta'}\overline{\mathcal{L}}$, $\mathbb{E}(\partial_\beta \mathcal{L}\mathcal{S}') = -\frac{1}{NT}\partial_{\beta\phi'}\bar{\mathcal{L}}$, and $\mathbb{E}(\mathcal{S}\mathcal{S}') = \frac{1}{NT}\overline{\mathcal{H}}$. Denote $v = ((\alpha^0)', -(\gamma^0)')'$, $\mathcal{S}'v = 0$ and $\partial_{\beta\phi'}\bar{\mathcal{L}}v = 0$.

From the definitions $\overline{W} = -(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}[\partial_{\phi\beta'}\overline{\mathcal{L}}])$ and $U^{(0)} = \sqrt{NT}(\partial_\beta \mathcal{L} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S})$,

$$\mathbb{E}(U^{(0)}) = 0, \qquad Var(U^{(0)}) = \overline{W} \tag{A.14}$$

which implies $lim_{N,T\to\infty} Var(U^{(0)}) = \overline{W}_\infty$.

According to Lemma A.2.1

$$U^{(0)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} D_\beta \ell_{it}, \tag{A.15}$$

where $D_\beta \ell_{it} := \partial_\beta \ell_{it} - \partial_\pi \ell_{it} \Xi_{it}$ is a martingale difference sequence for each i and independent across i, conditional on ϕ. Applying Theorem 2.3 in (McLeish, 1974) yields

$$U^{(0)} \xrightarrow{d} N[0, \lim_{N,T\to\infty} Var(U^{(0)})] \sim N(0, \overline{W}_\infty) \tag{A.16}$$

Step 2 shows that $U^{(1)} \to_P \kappa\bar{B}_\infty + \kappa^{-1}\bar{D}_\infty$. Since $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$U^{(1a)} = -\frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}[D_{\beta\pi}\ell_{it} - E_\phi(D_{\beta\pi}\ell_{it})]$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}^{2}\mathbb{E}_{\phi}(D_{\beta\pi^2}\ell_{it})$$

Plugging-in the definition of Λ_{it}, I decompose $U^{(1a)} = U^{(1a,1)} + U^{(1a,2)} + U^{(1a,3)} + U^{(1a,4)}$, where

$$U^{(1a,1)} = \frac{1}{(NT)^{3/2}}\sum_{i,j}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij}(\sum_{\tau}\partial_{\pi}\ell_{j\tau}\gamma_{\tau}^{0})\sum_{t}(D_{\beta\pi}\ell_{it} - \mathbb{E}_{\phi}D_{\beta\pi}\ell_{it})\gamma_{t}^{0},$$

$$U^{(1a,2)} = \frac{1}{(NT)^{3/2}}\sum_{t,j}\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}(\sum_{\tau}\partial_{\pi}\ell_{j\tau}\gamma_{\tau}^{0})\sum_{i}(D_{\beta\pi}\ell_{it} - \mathbb{E}_{\phi}D_{\beta\pi}\ell_{it})\alpha_{i}^{0},$$

$$U^{(1a,3)} = \frac{1}{(NT)^{3/2}}\sum_{i,\tau}\overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau}(\sum_{j}\partial_{\pi}\ell_{j\tau}\alpha_{j}^{0})\sum_{t}(D_{\beta\pi}\ell_{it} - \mathbb{E}_{\phi}D_{\beta\pi}\ell_{it})\gamma_{t}^{0},$$

$$U^{(1a,4)} = \frac{1}{(NT)^{3/2}}\sum_{t,\tau}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}(\sum_{j}\partial_{\pi}\ell_{j\tau}\alpha_{j}^{0})\sum_{i}(D_{\beta\pi}\ell_{it} - E_{\phi}D_{\beta\pi}\ell_{it})\alpha_{i}^{0}.$$

By the Cauchy-Schwarz inequality applied to the sum over t in $U^{(1a,2)}$,

$$\begin{aligned}(U^{(1a,2)})^2 &\leq \frac{1}{(NT)^3}[\sum_{t}(\sum_{j,\tau}\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}\partial_{\pi}\ell_{j\tau}\gamma_{\tau}^{0})^2][\sum_{t}(\sum_{i}(D_{\beta\pi}\ell_{it} - \mathbb{E}D_{\beta\pi}\ell_{it})\alpha_{i}^{0})^2] \\ &= \frac{1}{(NT)^3}[\sum_{t}O_p(NT)][\sum_{t}O_p(N)] = O_p(1/N) = o_p(1)\end{aligned}$$

Using that both $\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)}\partial_{\pi}\ell_{j\tau}\gamma_{\tau}^{0}$ and $(D_{\beta\tau}\ell_{it} - \mathbb{E}D_{\beta\pi}\ell_{it})\alpha_{i}^{0}$ are mean zero, independent across i.

Therefore, $U^{(1a,2)} = o_p(1)$. Analogously $U^{(1a,3)} = o_p(1)$.

According to Lemma A.2.5, it is the case that $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} = -\text{diag}[(\frac{1}{NT}\sum_{t=1}^{T}\mathbb{E}_{\phi}(\partial_{\pi^2}\ell_{it}(\gamma_{t}^{0})^2))^{-1}] + O_p(1)$. Analogously to the proof of $U^{(1a,2)}$, the $O_p(1)$ part of $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}$ has an asymptotically negligible contribution to $U^{(1a,1)}$. Thus,

$$\begin{aligned} U^{(1a,1)} &= \frac{1}{(NT)^{3/2}} \sum_{i,j} \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} \Big(\sum_{\tau} \partial_\pi \ell_{j\tau} \gamma^0_\tau\Big) \sum_t (D_{\beta\pi}\ell_{it} - \mathbb{E}_\phi D_{\beta\pi}\ell_{it}) \gamma^0_t \\ &= -\frac{1}{(NT)^{1/2}} \sum_i \frac{\big(\sum_\tau \partial_\pi \ell_{i\tau}\gamma^0_\tau\big) \sum_t (D_{\beta\pi}\ell_{it} - \mathbb{E}_\phi D_{\beta\pi}\ell_{it})\gamma^0_t}{\sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2}\ell_{it}(\gamma^0_t)^2)} + o_p(1) \end{aligned}$$

previous assumptions guarantee that $\mathbb{E}_\phi[(U_i^{(1a,1)})^2] = \mathcal{O}_p(1)$, uniformly over i. Note that both the denominator and the numerator of $U_i^{(1a,1)}$ are of order T. For the denominator this is obvious because of the sum over T. For the numerator there are two sums over T, but both $\partial_\pi \ell_{i\tau}\gamma^0_\tau$ and $(D_{\beta\pi}\ell_{it} - \mathbb{E}_\phi(D_{\beta\pi}\ell_{it}))\gamma^0_t$ are mean zero weakly correlated processes, the sum over which is of order $\sqrt{T}$ each. By applying the WLLN over i, $\frac{1}{N}\sum_i U_i^{(1a,1)} = \frac{1}{N}\mathbb{E}_\phi U_i^{(1a)} + o_P(1)$, and therefore

$$U^{(1a,1)} = \underbrace{-\sqrt{\frac{N}{T}} \frac{1}{N} \sum_{i=1}^N \frac{\sum_{t=1}^T \sum_{\tau=t}^T \mathbb{E}_\phi(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{i\tau}\gamma^0_t\gamma^0_\tau)}{\sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2}\ell_{it}(\gamma^0_t)^2)}}_{\equiv \sqrt{\frac{N}{T}}\overline{B}^{(1)}} + o_p(1).$$

Here, I use that $\mathbb{E}_\phi(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{i\tau}) = 0$ for $t > \tau$. Analogously,

$$U^{(1a,4)} = \underbrace{-\sqrt{\frac{T}{N}} \frac{1}{T} \sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}_\phi(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{it}(\alpha^0_i)^2)}{\sum_{i=1}^N \mathbb{E}_\phi(\partial_{\pi^2}\ell_{it}(\alpha^0_i)^2)}}_{\equiv \sqrt{\frac{T}{N}}\overline{D}^{(1)}} + o_p(1).$$

hence $U^{(1a)} = \kappa \overline{B}^{(1)} + \kappa^{-1}\overline{D}^{(1)} + o_p(1)$.

Next, I analyze $U^{(1b)}$. I decompose $\Lambda_{it} = \Lambda^{(1)}_{it} + \Lambda^{(2)}_{it} + \Lambda^{(3)}_{it} + \Lambda^{(4)}_{it}$, where

$$\Lambda_{it}^{(1)} = -\frac{1}{NT}\sum_{j=1}^{N}\overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1}\gamma_t^0\sum_{\tau=1}^{T}\partial_\pi \ell_{j\tau}\gamma_\tau^0, \qquad \Lambda_{it}^{(2)} = -\frac{1}{NT}\sum_{j=1}^{N}\overline{\mathcal{H}}_{(\gamma\alpha)tj}^{-1}\alpha_i^0\sum_{\tau=1}^{T}\partial_\pi \ell_{j\tau}\gamma_\tau^0,$$

$$\Lambda_{it}^{(3)} = -\frac{1}{NT}\sum_{\tau=1}^{T}\overline{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1}\gamma_t^0\sum_{j=1}^{N}\partial_\pi \ell_{j\tau}\alpha_j^0, \qquad \Lambda_{it}^{(4)} = -\frac{1}{NT}\sum_{\tau=1}^{T}\overline{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1}\alpha_i^0\sum_{j=1}^{N}\partial_\pi \ell_{j\tau}\alpha_j^0.$$

This decomposition of Λ_{it} includes the following decomposition of $U^{(1b)}$

$$U^{(1b)} = \sum_{p,q=1}^{4} U^{(1b,p,q)}, \quad U^{(1b,p,q)} = \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^{(p)}\Lambda_{it}^{(q)}\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it}).$$

Due to symmetry $U^{(1b,p,q)} = U^{(1b,q,p)}$ this is a decomposition into 10 distinct terms. Consider $U^{(1b,1,2)}$,

$U^{(1b,1,2)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}U_i^{(1b,1,2)}$, with

$U_i^{(1b,1,2)} = \frac{1}{2T}\sum_{t=1}^{T}\gamma_t^0\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})\frac{1}{N^2}\sum_{j_1,j_2=1}^{N}\overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)tj_2}^{-1}\alpha_i^0(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\partial_\pi\ell_{j_1\tau}\gamma_\tau^0)(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\partial_\pi\ell_{j_2\tau}\gamma_\tau^0)$.

Using $\mathbb{E}_\phi(\sum_t \partial_\pi \ell_{it}\gamma_t^0) = 0$, $\mathbb{E}_\phi(\sum_t \partial_\pi \ell_{it}\gamma_t^0 \sum_j \partial_{j\tau}\gamma_\tau^0)$ for $i \neq j$, and the properties of the inverse expected Hessian from Theorem A.2.5 one finds $\mathbb{E}_\phi[U_i^{(1b,1,2)}] = O_p(1/N)$, uniformly over i, and $\mathbb{E}_\phi[(U_i^{(1b,1,2)})^2] = O_p(1)$, uniformly over i, and $\mathbb{E}_\phi[U_i^{(1b,1,2)}U_j^{(1b,1,2)}] = O_p(1/N)$, uniformly over $i \neq j$. This implies that $\mathbb{E}_\phi U^{(1b,1,2)} = O_p(1/N)$, and $\mathbb{E}_\phi[(U^{(1b,1,2)} - \mathbb{E}_\phi U^{(1b,1,2)})^2] = O_p(1/\sqrt{N})$, and therefore $U^{(1b,1,2)} = o_p(1)$. By similar arguments one obtains $U^{(1b,p,q)} = o_p(1)$ for all combinations of $p,q = 1,2,3,4$, except for $p = q = 1$ and $p = q = 4$.

For $p = q = 1$, $U^{(1b,1,1)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}U_i^{(1b,1,1)}$, and

$U_i^{(1b,1,1)} = \frac{1}{2T}\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})\frac{1}{N^2}\sum_{j_1,j_2=1}^{N}\overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1}\overline{\mathcal{H}}_{(\alpha\alpha)ij_2}^{-1}(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\partial_\pi\ell_{j_1\tau}\gamma_\tau^0)(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\partial_\pi\ell_{j_2\tau}\gamma_\tau^0)$.

Analogous to the result for $U^{(1b,1,2)}$ one finds $\mathbb{E}_\phi[(U^{(1b,1,1)} - \mathbb{E}_\phi U^{(1b,1,1)})^2] = O_p(1/\sqrt{N})$, and therefore $U^{(1b,1,1)} = \mathbb{E}_\phi U^{(1b,1,1)} + o_p(1)$.

Furthermore,

$$\begin{aligned}\mathbb{E}_\phi U^{(1b,1,1)} &= \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})\mathbb{E}_\phi[(\partial_\pi\ell_{it}\gamma_t^0)^2]}{[\sum_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})]^2} + o(1)\\ &= \underbrace{-\sqrt{\frac{N}{T}}\frac{1}{2N}\sum_{i=1}^{N}\frac{\sum\limits_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum\limits_{t=1}^{T}(\gamma_t^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}}_{\equiv\sqrt{\frac{N}{T}}\overline{B}^{(2)}} + o(1),\end{aligned}$$

analogously,

$$U^{(1b,4,4)} = \mathbb{E}_\phi U^{(1b,4,4)} + o_p(1) = \underbrace{-\sqrt{\frac{T}{N}}\frac{1}{2T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum_{i=1}^{N}(\alpha_i^0)^2\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})}}_{\equiv\sqrt{\frac{T}{N}}\overline{D}^{(2)}} + o_p(1),$$

thus $U^{(1b)} = \kappa\overline{B}^{(2)} + \kappa^{-1}\overline{D}^{(2)} + o_p(1)$.

Since $\overline{B}_\infty = \lim_{N,T\to\infty}[\overline{B}^{(1)} + \overline{B}^{(2)}]$ and $\overline{D}_\infty = \lim_{N,T\to\infty}[\overline{D}^{(1)} + \overline{D}^{(2)}]$, then $U^{(1)} = \kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty + o_p(1)$.

I have shown $U^{(0)} \xrightarrow{d} N(0, \overline{W}_\infty)$, and $U^{(1)} \xrightarrow{p} \kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty$. Using this and Lemma A.2.1 I obtain

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} \overline{W}_\infty^{-1} N(\kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty, \overline{W}_\infty).$$

A.2.2.2 Proof of asymptotic distribution of APE

I consider the case of scalar Δ_{it} to simplify the notation. Decompose

$$\sqrt{NT}(\hat{\delta} - \delta_{NT}^0 - \overline{B}_\infty^\delta/T - \overline{D}_\infty^\delta/N) = \sqrt{NT}(\delta - \delta_{NT}^0) + \sqrt{NT}(\hat{\delta} - \delta - \overline{B}_\infty^\delta/T - \overline{D}_\infty^\delta/N).$$

Part (1): Limit of $\sqrt{NT}(\hat{\delta} - \delta - \overline{B}_\infty^\delta/T - \overline{D}_\infty^\delta/N)$. An argument analogous to the

proof of 1.3.1 using Lemma A.2.2 yields

$$\sqrt{NT}(\hat{\delta} - \delta) \xrightarrow{d} N(\kappa \overline{B}_{\infty}^{\delta} + \kappa^{-1}\overline{D}_{\infty}^{\delta}, \overline{V}_{\infty}^{\delta(1)}),$$

where $\overline{V}_{\infty}^{\delta(1)} = \overline{\mathbb{E}}\{(NT)^{-1}\sum_{i,t}\mathbb{E}_{\phi}[\Gamma_{it}^2]\}$, for the expressions of $\overline{B}_{\infty}^{\delta}$, $\overline{D}_{\infty}^{\delta}$, and Γ_{it} given in the statement of the theorem. Then, by Mann-Wald theorem

$$\sqrt{NT}(\hat{\delta} - \delta - \overline{B}_{\infty}^{\delta}/T - \overline{D}_{\infty}^{\delta}/N) \xrightarrow{d} N(0, \overline{V}_{\infty}^{\delta(1)}).$$

\# Part (2): Limit of $\sqrt{NT}(\delta - \delta_{NT}^0)$. Here I show that $\sqrt{NT}(\delta - \delta_{NT}^0) \xrightarrow{d} N(0, \overline{V}_{\infty}^{\delta(2)})$ and characterize the asymptotic variance $\overline{V}_{\infty}^{\delta(2)}$. I characterize $\overline{V}_{\infty}^{\delta(2)}$ as $\overline{V}_{\infty}^{\delta(2)} = \overline{\mathbb{E}}\{NT\mathbb{E}[(\delta - \delta_{NT}^0)^2]\}$, because $\mathbb{E}[\delta - \delta_{NT}^0] = 0$. Note, the rate $\sqrt{NT}$ is determined through $\mathbb{E}[(\delta - \delta_{NT}^0)^2]$, where

$$\mathbb{E}[(\delta - \delta_{NT}^0)^2] = \mathbb{E}[(\frac{1}{NT}\sum_{i,t}\tilde{\Delta}_{it})^2] = \frac{1}{N^2T^2}\sum_{i,j,t,s}\mathbb{E}[\tilde{\Delta}_{it}\tilde{\Delta}_{js}], \tag{A.17}$$

for $\tilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$. The order of $\mathbb{E}[(\delta - \delta_{NT}^0)^2]$ is equal to the number of terms of the sums in equation (A.17) that are nonzero, which is determined by the sample properties of $\{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$. Under Assumption 2(i)

$$\mathbb{E}[(\delta - \delta_{NT}^0)^2] = \frac{1}{N^2T^2}\sum_{i,t,s}\mathbb{E}[\tilde{\Delta}_{it}\tilde{\Delta}_{is}] = O(N^{-1}),$$

because $\{\tilde{\Delta}_{it} : 1 \leq i \leq N; 1 \leq t \leq T\}$ is independent across i and α-mixing across t.

\#Part(3): Limit of $\sqrt{NT}(\hat{\delta} - \delta_{NT}^0 - T^{-1}\overline{B}_{\infty}^{\delta} - N^{-1}\overline{D}_{\infty}^{\delta})$.The conclusion of the Theorems follows because $(\delta - \delta_{NT}^0)$ and $(\hat{\delta} - \delta - T^{-1}\overline{B}_{\infty}^{\delta} - N^{-1}\overline{D}_{\infty}^{\delta})$ are asymptotically independent and $\overline{V}_{\infty}^{\delta} = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)}$.

A.2.3 Proofs of Theorems 1.3.3 and 1.3.4

I start by stating a lemma that is going to be used for this section. It corresponds to Lemma C.2 of (Fernández-Val and Weidner, 2013) and the proof is omitted for brevity.

Lemma A.2.3. *Let* $G(\beta,\phi) := \frac{1}{N(T-j)}\sum_{i,t\geq j+1} g(X_{it}, X_{i,t-j}, \beta, \alpha_i\gamma_t, \alpha_i\gamma_{t-j})$ *for* $0 \leq j < T$, *and* $\mathcal{B}^0_\varepsilon$ *be a subset of* $\mathbb{R}^{\dim\beta+2}$ *that contains an* ε*-neighborhood of* $(\beta, \pi^0_{it}, \pi^0_{i,t-j})$ *for all* i, t, j, N, T, *and for some* $\varepsilon > 0$.
Assume that $(\beta, \pi_1, \pi_2) \to g_{itj}(\beta, \pi_1, \pi_2) := g(X_{it}, X_{i,t-j}, \beta, \pi_1, \pi_2)$ *is Lipschitz continuous over* $\mathcal{B}^0_\varepsilon$ *a.s., i.e.* $|g_{itj}(\beta_1, \pi_{11}, \pi_{21}) - g_{itj}(\beta_0, \pi_{10}, \pi_{20})| \leq M_{itj}\|(\beta_1, \pi_{11}, \pi_{21}) - (\beta_0, \pi_{10}, \pi_{20})\|$ *for all* $(\beta_1, \pi_{11}, \pi_{21}) \in \mathcal{B}^0_\varepsilon$, $(\beta_0, \pi_{10}, \pi_{20}) \in \mathcal{B}^0_\varepsilon$, *and some* $M_{itj} = O_p(1)$ *for all* i, t, j, N, T. *Let* $(\hat\beta, \hat\phi)$ *be an estimator of* (β, ϕ) *such that* $\|\hat\beta - \beta^0\| \xrightarrow{p} 0$ *and* $\|\hat\phi - \phi^0\|_\infty \xrightarrow{p} 0$. *Then,*

$$G(\hat\beta, \hat\phi) \xrightarrow{p} \overline{\mathbb{E}}[G(\beta^0, \phi^0)],$$

provided that the limit exists.

This lemma shows the consistency of the estimators of averages of the data and parameters. I will use this result to show the validity of the analytical bias corrections and the consistency of the variance estimators.

A.2.3.1 Proof of Theorem 1.3.3

I separate the proof in two parts corresponding to the two statements of the theorem.

Part I: Proof of $\hat W \xrightarrow{p} \overline{W}_\infty$. The asymptotic variance and its estimators can be expressed as $\overline{W}_\infty = \overline{\mathbb{E}}[W(\beta^0, \phi^0)]$ and $\widehat{W} = W(\hat\beta, \hat\phi)$, where $W(\beta, \phi)$ has a first order representation as a continuously differentiable transformation of terms that have the form of $G(\beta, \phi)$ in Lemma A.2.3.The result then follows by the continuous mapping theorem noting that $\|\hat\beta - \beta^0\| \xrightarrow{p} 0$ and $\|\hat\phi - \phi^0\|_\infty \xrightarrow{p} 0$.

Part II: Proof of $\sqrt{NT}(\tilde\beta^A - \beta^0) \xrightarrow{d} N(0, \overline{W}^{-1}_\infty)$. I show that $\hat B \xrightarrow{p} \overline{B}_\infty$ and $\hat D \xrightarrow{p} \overline{D}_\infty$. These asymptotic biases and their fixed effects estimators are either time-series averages of fractions of cross-sectional averages, or vice versa. The nesting of the averages makes the analysis a bit more cumbersome than the analysis of $\widehat{W}$, but the results follows by similar standard arguments, also using that $L \to \infty$ and $L/T \to 0$ guarantee that the trimmed estimator in $\hat{B}$ is also consistent for the spectral expectations; see Lemma 6 in (Hahn and Kuersteiner, 2011).

A.2.3.2 Proof of Theorem 1.3.4

I separate the proof into two parts corresponding to the two statements of the theorem.

Part I: $\hat{V}^{\delta} \xrightarrow{p} \overline{V}^{\delta}_{\infty}$. $\overline{V}^{\delta}_{\infty}$ and $\hat{V}^{\delta}$ have a similar structure to $\overline{W}_{\infty}$ and $\widehat{W}$ in part I of the proof of Theorem 1.3.3, so that the consistency follows by an analogous argument.

Part II: $\sqrt{NT}(\tilde{\delta}^{A} - \delta^{0}_{NT}) \xrightarrow{d} N(0, \overline{V}^{\delta}_{\infty})$. As in the proof of Theorem 1.3.2, I decompose

$$\sqrt{NT}(\tilde{\delta}^{A} - \delta^{0}_{NT}) = \sqrt{NT}(\delta - \delta^{0}_{NT}) + \sqrt{NT}(\tilde{\delta}^{A} - \delta).$$

Then, by Mann-Wald theorem,

$$\sqrt{NT}(\tilde{\delta}^{A} - \delta) = \sqrt{NT}(\hat{\delta} - \hat{B}^{\delta}/T - \hat{D}^{\delta}/N - \delta) \xrightarrow{d} N(0, \overline{V}^{\delta(1)}_{\infty}),$$

provided that $\hat{B}^{\delta} \xrightarrow{p} \overline{B}^{\delta}_{\infty}$ and $\hat{D}^{\delta} \xrightarrow{p} \overline{D}^{\delta}_{\infty}$, and $\sqrt{NT}\delta - \delta^{0}_{NT}) \xrightarrow{d} N(0, \overline{V}^{\delta(2)}_{\infty})$, where $\overline{V}^{\delta(1)}_{\infty}$ and $\overline{V}^{\delta(2)}_{\infty}$ are defined as in the proof of Theorem 1.3.2. The statement thus follows by using a similar argument to part II of the proof of Theorem 1.3.3 to show the consistency of $\hat{B}^{\delta}$and $\hat{D}^{\delta}$, and because $(\delta - \delta^{0}_{NT})$ and $(\tilde{\delta}^{A} - \delta)$ are asymptotically independent, and $\overline{V}^{\delta}_{\infty} = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)}$.

A.2.4 Properties of the Inversed Expected Incidental Parameter Hessian

The following two lemmas would be used in the proof of asymptotic distributions of β and δ.

Lemma A.2.4. *Let Assumption 1 hold, then* $\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty < 1 - \frac{b_{\min}}{b_{\max}}$, *and* $\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty < 1 - \frac{b_{\min}}{b_{\max}}$.

Proof. Let $h_{it} = \mathbb{E}(-\partial_{\pi^2}\ell_{it})$, Assumption 1 guarantees that $b_{\min} \le h_{it} \le b_{\max}$, therefore

$$
\begin{aligned}
\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty &= \max_i \frac{\sum_t |\alpha_i^0\gamma_t^0 h_{it}|}{\sum_t (\gamma_t^0)^2 h_{it}} = 1 - \max_i \frac{\sum_t((\gamma_t^0)^2 - |\alpha_i^0\gamma_t^0|)h_{it}}{\sum_t(\gamma_t^0)^2 h_{it}} \\
&\le 1 - \frac{\|\gamma^0\|^2 - \min_i|\alpha_i^0|\|\gamma^0\|_1}{\|\gamma^0\|^2}\frac{b_{\min}}{b_{\max}}
\end{aligned}
$$

similar,

$$
\begin{aligned}
\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty &= \max_t \frac{\sum_i |\alpha_i^0\gamma_t^0 h_{it}|}{\sum_i (\alpha_i^0)^2 h_{it}} = 1 - \max_t \frac{\sum_i((\alpha_i^0)^2 - |\alpha_i^0\gamma_t^0|)h_{it}}{\sum_i(\alpha_i^0)^2 h_{it}} \\
&\le 1 - \frac{\|\alpha^0\|^2 - \min_t|\gamma_t^0|\|\alpha^0\|_1}{\|\alpha^0\|^2}\frac{b_{\min}}{b_{\max}}
\end{aligned}
$$

Since $\|\alpha^0\|^2 \ge \frac{1}{N}\|\alpha^0\|_1^2$, as long as $\frac{1}{N}\|\alpha^0\|_1 \ge \min_t|\gamma_t^0|$, $\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty \le 1 - \frac{b_{\min}}{b_{\max}}$; similarly since$\|\gamma^0\|^2 \ge \frac{1}{T}\|\gamma^0\|_1^2$, as long as $\frac{1}{T}\|\gamma^0\|_1 \ge \min_i|\alpha_i^0|$, $\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \le 1 - \frac{b_{\min}}{b_{\max}}$. □

Lemma A.2.5. *Under Assumption 1,*

$$
\|\overline{\mathcal{H}}^{-1} - \operatorname{diag}(\overline{\mathcal{H}}_{(\alpha\alpha)}, \overline{\mathcal{H}}_{(\gamma\gamma)})^{-1}\|_{\max} = O_p(1).
$$

Proof. By the inversion formula for partitioned matrices

$$
\overline{\mathcal{H}}^{-1} = \begin{pmatrix} A & -A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \\ -\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}A & \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \end{pmatrix},
$$

with

$$
A \equiv (\overline{\mathcal{H}}_{(\alpha\alpha)} - \overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1} = \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}(\mathbf{I} - \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1}
$$

$$= \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\sum_{n=0}^{\infty}(\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^{n}.$$

Define

$$B \equiv \sum_{n=1}^{\infty}(\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^{n},$$

then $A = \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}+\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}B$. By using the matrix norm property that$\|AB\|_{\max} \leq \|A\|_{\infty}\|B\|_{\max}$ and Lemma A.2.4

$$\begin{aligned}
\|B\|_{\max} &\leq \sum_{n=1}^{\infty}(\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^{n}\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty}\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}\|\overline{\mathcal{H}}_{(\gamma\alpha)}^{-1}\|_{\max} \\
&\leq [\sum_{n=1}^{\infty}(1-\frac{b_{\min}}{b_{\max}})^{2n}]T\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty}\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}^{2} = O(N^{-1}).
\end{aligned}$$

From this I obtain

$$\|A\|_{\infty} \leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} + N\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty}\|B\|_{max} = O(N).$$

From the different blocks of

$$\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1} = \begin{pmatrix} A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} & -A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A & \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \end{pmatrix}$$

it can be seen that

$$\begin{aligned}
\|A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\max} &= \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}B\|_{\max} \leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty}\|B\|_{\max} = O_p(1), \\
\|-A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\max} &\leq \|A\|_{\infty}\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty} = O_p(1) \\
\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\max} &\leq \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}^{2}\|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\infty}\|A\|_{\infty}\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \\
&\leq N\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}^{2}\|A\|_{\infty}\|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max}^{2} = O_p(1)
\end{aligned}$$

Having the bound $O_p(1)$ for the max-norm of each block of the matrix yields also the same bound for the max-norm of the matrix itself, as desired. □

This result establishes that $\overline{\mathcal{H}}^{-1}$ can be uniformly approximated by a diagonal matrix, which is given by the inverse of the diagonal terms of $\overline{\mathcal{H}}$. The diagonal elements of $\text{diag}(\overline{\mathcal{H}}_{(\alpha\alpha)}, \overline{\mathcal{H}}_{(\gamma\gamma)})^{-1}$ are of order N and T respectively, hence the order of difference established by the lemma is relatively small.

With this result, $\|\overline{\mathcal{H}}^{-1}\|_\infty \leq \|\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1}\|_\infty + \|\overline{\mathcal{D}}^{-1}\|_\infty \leq (N+T)\|\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1}\|_{\max} + \|\overline{\mathcal{D}}^{-1}\|_\infty = O_p(N)$ which can be used to verify the assumption in the proof of Theorem B.1 of (Fernández-Val and Weidner, 2013).

A.3 Proof of Section 1.4

A.3.1 Compare with existing methods

A.3.1.1 Proof of Proposition 1.4.1

The proof is mainly for the case without unobserved effects, but similarly argument can be used to the proof of other cases.

The model looks $Y_{it} = \mathbf{1}\{X_{it}'\beta + \varepsilon_{it} \geq 0\}$, and ε_{it} is normally distributed with variance 1. When estimating the structural parameter of probit using MLE,

$$\beta \in \arg\max_{\beta\in\Theta} \mathcal{L}_{NT} = \sum_{i,t} \ell_{it} = \sum_{i,t} Y_{it} log\Phi(X_{it}'\beta) + (1-Y_{it})log(1-\Phi(X_{it}'\beta)),$$

and then the score of β is

$$\sum_{i,t} X_{it}\{\underbrace{Y_{it}\frac{\phi_f(X_{it}'\beta)}{\Phi(X_{it}'\beta)} - (1-Y_{it})\frac{\phi_f(X_{it}'\beta)}{1-\Phi(X_{it}'\beta)}}_{\tilde{g}_{it}(\beta)}\} = 0$$

$$\Leftrightarrow \sum_{i,t} X_{it}\{\frac{Y_{it} - \Phi(X_{it}'\beta)}{\Phi(X_{it}'\beta)(1-\Phi(X_{it}'\beta))}\phi_f(X_{it}'\beta)\} = 0,$$

which relates to the generalized residuals part of EM,

$$\begin{aligned}\hat{Y}_{it} &= X_{it}\beta + \underbrace{Y_{it}\cdot\phi_f(X_{it}\beta)/\Phi(X_{it}\beta) - (1-Y_{it})\cdot\phi_f(X_{it}\beta)/\{1-\Phi(X_{it}\beta)\}}_{g_{it}(\beta)}, \\ &= X_{it}\beta + (Y_{it}-\Phi(X_{it}\beta))\cdot\phi(X_{it}\beta)/\{\Phi(X_{it}\beta)(1-\Phi(X_{it}\beta))\},\end{aligned}$$

and

$$\beta = (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{'})^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{'}\hat{Y}_{it}\}.$$

Denote $\mu_{it}^{(k)} = X_{it}^{'}\beta^{(k)}$, the score function is of β is zero, i.e. the unique fixed-point property, means that,

$$\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{'}((Y_{it}-\Phi(X_{it}^{'}\beta))\cdot\phi_f(X_{it}^{'}\beta)/\{\Phi(X_{it}^{'}\beta)(1-\Phi(X_{it}^{'}\beta))\}) = 0 \Rightarrow \beta^{(k)} = \beta^{0},$$

this is due to the identification condition that

$$E^0[g_{it}(\beta^0)|X_{it}] = E^0[E[\varepsilon_{it}|Y_{it}, X_{it}, \beta^0]|X_{it}] = E^0[\varepsilon_{it}|X_{it}] = 0.$$

By central limit theory for the score

$$\sqrt{NT}E[\nabla_\beta l_{it}] = \sqrt{NT}E[\sum_{i,t} X_{it}g_{it}(\beta)] \xrightarrow{d} N(0, E\frac{\phi_{it}^2}{\Phi_{it}(1-\Phi_{it})}X_{it}X_{it}'),$$

with $Var(\sum_{i,t} X_{it}\tilde{g}_{it}(\beta)) = Var(\sum_{i,t} X_{it}\frac{Y_{it}-\Phi(X_{it}\beta)}{\Phi(X_{it}\beta)(1-\Phi(X_{it}\beta))}\phi_f(X_{it}^{'}\beta))$.
Since $Var(Y_{it}-\Phi(X_{it}'\beta)|X_{it}) = \Phi(X_{it}'\beta)(1-\Phi(X_{it}'\beta))$,

$$\sqrt{NT}(\hat{\beta}-\beta) \xrightarrow{d} N(0, [E\frac{\phi_{it}^2}{\Phi_{it}(1-\Phi_{it})}X_{it}X_{it}]^{-1})$$

for both EM and MLE.

A.3.1.2 Proof of Proposition 1.4.2

This is to show the difference between the proposed fixed effects EM-type estimator and the Newton's method as described in (Greene, 2003).

From the **E-**step, one has $\hat{Y}_{it}^{(k)} = X_{it}'\beta^{(k)} + \alpha_i^{(k)} + \underbrace{\frac{Y_{it} - \Phi(\mu_{it}^{(k)})}{\Phi(\mu_{it}^{(k)})(1-\Phi(\mu_{it}^{(k)}))}\phi_{it}(\mu_{it}^{(k)})}_{g_{it}^{(k)}}.$

For fixed effects EM-type estimator, given α_i, parameter β can be updated by

$$\begin{aligned}\beta^{(k+1)} &= (\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}')^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)})\} \\ &= \beta^{(k)} + \underbrace{(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}')^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}g_{it}^{(k)}\}}_{\Delta_{\beta_{EM}}^{(k)}},\end{aligned}$$

hence α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_{it}^{(k)} - X_{it}'\beta^{(k+1)}) = \alpha_i^{(k)} + g_{ii}^{(k)} - \frac{1}{T}\sum_{t=1}^{T} X_{it}'\Delta_{\beta_{EM}}^{(k)}.$$

For Newton's method as described in (Greene, 2003) Chapter 21

$$\begin{aligned}\beta^{(k+1)} &= \beta^{(k)} - \{\sum_{i=1}^{N}\sum_{t=1}^{T} h_{it}(X_{it} - \overline{X}_i)(X_{it} - \overline{X}_i)'\}^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}^{(k)}(X_{it} - \overline{X}_i)\} \\ &= \beta^{(k)} + \Delta_{\beta_{NR}}^{(k)},\end{aligned}$$

and

$$\alpha_i^{(k+1)} = \alpha_i^{(k)} - g_{ii}^{(k)}/h_{ii}^{(k)} - \overline{X}_i'\Delta_{\beta_{NR}}^{(k)},$$

here $h_{it} = g_{it}' = \frac{\phi_f(z_{it}q_{it})}{\Phi(z_{it}q_{it})} - (\frac{\phi_f(z_{it}q_{it})}{\Phi(q_{it}z_{it})})^2$, $z_{it} = X_{it}'\beta + \alpha_i$, $q_{it} = 1 - 2Y_{it}$, $h_{ii} = \sum_{t=1}^{T} h_{it}$, and $g_{ii} = \sum_{t=1}^{T} g_{it}$. The sign difference is due to that h_{it} is negative for all values of $z_{it}q_{it}$.

A.3.2 Proof of Consistency for general $\hat{\beta}$

In general, the consistency proof will contain two steps as shown in the proof of PPIF.

Denote $z_{it} = X_{it}'\beta + \alpha_i\gamma_t$, under the bounded from below of the second order derivatives assumption

$$\forall y \in \mathcal{Y},\ z \in \mathcal{Z}:\ b_{min} < \partial_{z^2}\mathcal{L}(y,z),$$

also assume that $\mathcal{Z}$ is convex, i.e. since $\mathcal{Z} \subset \mathbb{R}$ it is an interval (either open or closed). From this it follows that for all $z_1,\ z_2 \in \mathcal{Z}$ one has

$$\begin{aligned}
\mathcal{L}(y,z_1) - \mathcal{L}(y,z_2) &= [\partial_z\mathcal{L}(y,z_1)](z_1 - z_2) + \frac{1}{2}[\partial_{z^2}\mathcal{L}(y,\tilde{z})](z_1 - z_2)^2 \\
&\geq [\partial_z\mathcal{L}(y,z_1)](z_1 - z_2) + \frac{b_{min}}{2}(z_1 - z_2)^2 \\
&= \frac{b_{min}}{2}(z_1 - z_2 + \frac{1}{b_{min}}[\partial_z\mathcal{L}(y,z_1)])^2 - \frac{1}{2b_{min}}[\partial_z\mathcal{L}(y,z_1)]^2,
\end{aligned}$$

where $z_1 \leq \tilde{z} \leq z_2$. Define $\hat{z}_{it} = z_{it}(\hat{\beta}, \hat{\alpha}_i, \hat{\gamma}_t)$, and $e_{it} = \frac{1}{b_{min}}[\partial_z\mathcal{L}_{it}]$. Note that $\mathbb{E}(e_{it}) = 0$. Since the estimated parameters minimize the objective function, observe that

$$\begin{aligned}
0 &\geq \mathcal{L}_{NT}(\beta^0,\phi^0) - \mathcal{L}_{NT}(\hat{\beta},\hat{\phi}) = \frac{1}{NT}\sum_{i,t}[\mathcal{L}_{it}(z_{it}^0) - \mathcal{L}_{it}(\hat{z}_{it})] \\
&\geq \frac{b_{min}}{2NT}\sum_{i,t}[(z_{it}^0 - \hat{z}_{it} + e_{it})^2 - e_{it}^2] = \frac{b_{min}}{2NT}\sum_{i,t}\{[X_{it}'(\hat{\beta} - \beta^0) + \hat{\alpha}_i\hat{\gamma}_t - \alpha_i^0\gamma_t^0 - e_{it}]^2 - e_{it}^2\}.
\end{aligned}$$

Once the last inequality is obtained, the consistency proof for $\hat{\beta}$ is equivalent to that for the linear regression model with interactive fixed effects. In matrix notation, the above inequality reads

$$\begin{aligned}\frac{1}{NT}Tr(e'e) &\geq \frac{1}{NT}Tr[(X'(\hat{\beta}-\beta^0)+\hat{\alpha}\hat{\gamma}-\alpha^0\gamma^0-e)'(X'(\hat{\beta}-\beta^0)+\hat{\alpha}\hat{\gamma}'-\alpha^0\gamma^0-e)] \\ &\geq \frac{1}{NT}Tr[(X'(\hat{\beta}-\beta^0)-e)'M_{(\hat{\alpha},\alpha^0)}(X'(\hat{\beta}-\beta^0)-e)]\end{aligned}$$

where $M_{(\hat{\alpha},\alpha^0)} = \mathbf{1}_T - (\hat{\alpha},\alpha^0)[(\hat{\alpha},\alpha^0)'(\hat{\alpha},\alpha^0)]^{-1}(\hat{\alpha},\alpha^0)'$ is the projector that projects orthogonal to $(\hat{\alpha},\alpha^0)$.

The assumptions on the panel model already guarantee that $\frac{1}{NT}Tr(Xe') = o_P(1)$. One can furthermore show that $\|e\| = o_P(\sqrt{NT})$, also the assumption $\frac{1}{NT}Tr(XX') = O_p(1)$ is satisfied from the distribution assumption on the regressors above. Then,

$$\begin{aligned}|\frac{1}{NT}Tr(e'M_{(\hat{\alpha},\alpha^0)}X_k)| &\leq \frac{1}{NT}|Tr(e'X_k)| + \frac{1}{NT}|Tr(e'P_{(\hat{\alpha},\alpha_0)}X_k)| \\ &\leq o_p(1) + \frac{2}{NT}\|e\|\|X_k\| = o_p(1).\end{aligned}$$

Under these, one has

$$0 \geq c\|\hat{\beta}-\beta\| + o_p\|\hat{\beta}-\beta^0\| + o_p(1)$$

from which $\hat{\beta} = \beta^0 + o_p(1)$.

A.4 Some useful algebraic results

For any $N \times T$ matrix A, define the $N \times T$ matrix $\mathbb{P}A$ as follows

$$(\mathbb{P}A)_{it} = \alpha_i^0\gamma_t^0(\alpha_i^* + \gamma_t^*), \quad (\alpha^*,\gamma^*) \in arg\min_{\alpha_i,\gamma_t}\sum_{i,t}\mathbb{E}(-\partial_{\pi^2}\ell_{it})(A_{it} - \alpha_i^0\gamma_t^0(\alpha_i+\gamma_t))^2.$$

Here, the minimization is over $\alpha \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}^T$, and $\mathbb{P}$ is the projection operator. It is a linear projection, i.e. $\mathbb{P}\mathbb{P} = \mathbb{P}$. It is also convenient to define

$$\widetilde{\mathbb{P}}A = \mathbb{P}\tilde{A}, \quad where \quad \tilde{A}_{it} = \frac{A_{it}}{\mathbb{E}(-\partial_{\pi^2}\ell_{it})}. \tag{A.18}$$

$\widetilde{\mathbb{P}}$ is a linear operator, but not a projection. Note that Ξ and Λ defined before can be written as $\Xi_k = \widetilde{\mathbb{P}}B_k$ and $\Lambda = \widetilde{\mathbb{P}}C$, where $C_{it} = -\partial_\pi \ell_{it}$ and $B_{k,it} = -\mathbb{E}_\phi(\partial_{\beta_k \pi}\ell_{it})$, for $k = 1, ..., \dim\beta$. [1]

The linear operator $\widetilde{\mathbb{P}}$ is closely related to the projection operator $\mathbb{P}$. The following lemma shows how in the context of panel probit model some expressions that regularly appear in the general expansions can conveniently be expressed by using the operator $\widetilde{\mathbb{P}}$.

Lemma A.4.1. *Let A, B and C be $N \times T$ matrices, and let the expected incidental parameter Hessian $\overline{\mathcal{H}}$ be invertible. Define the $N+T$ vectors $\mathcal{A}$ and $\mathcal{B}$ and the $(N+T) \times (N+T)$ matrix $\mathcal{C}$ as follows*

$$\mathcal{A} = \frac{1}{NT}\begin{pmatrix} A\gamma^0 \\ A'\alpha^0 \end{pmatrix}, \quad \mathcal{B} = \frac{1}{NT}\begin{pmatrix} B\gamma^0 \\ B'\alpha^0 \end{pmatrix},$$

and

$$\mathcal{C} = \frac{1}{NT}\begin{pmatrix} diag(C(\gamma^0 \circ \gamma^0)) & C \circ (\alpha^0(\gamma^0)') \\ (C \circ (\alpha^0(\gamma^0)'))' & diag(C'(\alpha^0 \circ \alpha^0)) \end{pmatrix}$$

where $\circ$ denotes the Hadamard product, i.e., element-by-element product. Then

(i) $\mathcal{A}'\overline{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT}\sum_{i,t}(\widetilde{\mathbb{P}}A_{it})B_{it} = \frac{1}{NT}\sum_{i,t}(\widetilde{P}B)_{it}A_{it}$,

(ii) $\mathcal{A}'\overline{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT}\sum_{i,t}\mathbb{E}(-\partial_{\pi^2}\ell_{it})(\widetilde{\mathbb{P}}A)_{it}(\widetilde{\mathbb{P}}B)_{it}$,

(iii) $\mathcal{A}'\overline{\mathcal{H}}^{-1}\mathcal{C}\overline{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT}\sum_{i,t}(\widetilde{\mathbb{P}}A)_{it}C_{it}(\widetilde{\mathbb{P}}B)_{it}$.

Proof. Let $\alpha_i^0\gamma_t^0(\tilde{\alpha}_i^* + \tilde{\gamma}_t^*) = (\mathbb{P}\tilde{A})_{it} = (\widetilde{\mathbb{P}}A)_{it}$, with $\tilde{A}$ as defined in eq (A.18). The FOC of the minimization problem in the definition of $(\mathbb{P}\tilde{A})_{it}$ can be written as $\overline{\mathcal{H}}\begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix} = \mathcal{A}$.

---

[1] B_k and Ξ_k are $N \times T$ matrices with entries $B_{k,it}$ and $\Xi_{k,it}$ respectively, while B_{it} and Ξ_{it} are $\dim\beta$-vectors with entries $B_{k,it}$and $\Xi_{k,it}$.

One solution to this is $\begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix} = \overline{\mathcal{H}}^{-1}\mathcal{A}$. Therefore,

$$\mathcal{A}'\overline{\mathcal{H}}^{-1}\mathcal{B} = \begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix}' \mathcal{B} = \frac{1}{NT}\sum_{i,t} \alpha_i^0 \gamma_t^0 (\tilde{\alpha}_i^* + \tilde{\gamma}_t^*) B_{it} = \frac{1}{NT}\sum_{i,t} (\widetilde{\mathbb{P}}A)_{it} B_{it}.$$

This is the first equality of the Statement (i) in the lemma. The second equality of Statement (i) follows by symmetry. Statement (ii) is a special case of Statement (iii) with $\mathcal{C} = \overline{\mathcal{H}}$, so Statement (iii) needs to be proved.

Let $\alpha_i^0 \gamma_t^0 (\alpha_i^* + \gamma_t^*) = (\mathbb{P}\tilde{B})_{it} = (\widetilde{\mathbb{P}}B)_{it}$, where $\tilde{B}_{it} = \frac{B_{it}}{\mathbb{E}(-\partial_{\pi^2}\ell_{it})}$. Analogous to the above, choose $\begin{pmatrix} \alpha^0 \circ \alpha^* \\ \gamma^0 \circ \gamma^* \end{pmatrix} = \overline{\mathcal{H}}^{-1}\mathcal{B}$ as one solution to the minimization problem. Then

$$\begin{aligned} & \mathcal{A}'\overline{\mathcal{H}}^{-1}\mathcal{C}\overline{\mathcal{H}}^{-1}\mathcal{B} \\ = \; & \frac{1}{NT}\sum_{i,t} (\alpha_i^0 \gamma_t^0)^2 [\tilde{\alpha}_i^* C_{it} \alpha_i^* + \tilde{\gamma}_t^* C_{it} \alpha_i^* + \tilde{\alpha}_i^* C_{it} \gamma_t^* + \tilde{\gamma}_t^* C_{it} \gamma_t^*] \\ = \; & \sum_{i,t} (\widetilde{\mathbb{P}}A)_{it} C_{it} (\widetilde{\mathbb{P}}B)_{it} \end{aligned}$$

□

Appendix B

Proofs for Chapter 2

B.1 Results

We drop the subscript NT on ϕ_{NT} and $\mathcal{L}_{NT}(\beta, \phi)$, and we denote the unpenalized objective function (denoted by $\mathcal{L}_{NT}(\beta, \phi)$ in the main text) as

$$\mathcal{L}^*(\beta, \phi) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \ell_{it}(z_{it}),$$

where $\phi = (\alpha', \gamma')'$ and $z_{it} = X_{it}'\beta + \alpha_i \gamma_t$. To fix the rescaling freedom in α_i and γ_t we introduce the penalized objective function

$$\mathcal{L}(\beta, \phi) = \mathcal{L}^*(\beta, \phi) - \frac{b}{8\sqrt{NT}} \left(\sum_{i=1}^{N} \alpha_i^2 - \sum_{t=1}^{T} \gamma_t^2 \right)^2,$$

where $b > 0$ is a constant. Let $\widehat{\beta}$ and $\widehat{\phi} = (\widehat{\alpha}', \widehat{\gamma}')'$ be the maximizers of $\mathcal{L}(\beta, \phi)$. The penalty term guarantees that the estimator satisfies the normalization $\sum_{i=1}^{N} \widehat{\alpha}_i^2 = \sum_{t=1}^{T} \widehat{\gamma}_t^2$. Note that we also normalize the true parameters such that the same normalization holds, i.e. $\sum_{i=1}^{N} (\alpha_i^0)^2 = \sum_{t=1}^{T} (\gamma_t^0)^2$. In addition, let $\widehat{\phi}(\beta) = (\widehat{\alpha}(\beta)', \widehat{\gamma}(\beta)')'$ be the maximizer of $\mathcal{L}(\beta, \phi)$ for given β.

B.1.1 Consistency

Lemma B.1.1. *Let Assumption 4 be satisfied. Then we have* $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_P(N^{-3/8})$ *and*

$$\frac{1}{\sqrt{NT}} \left\| \widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} \right\|_F = \mathcal{O}_P(N^{-3/8} + \|\beta - \beta^0\|),$$

uniformly over β *in a* ϵ*-neighborhood around* β^0 *for some* $\epsilon > 0$*. This implies*[1]

$$\frac{1}{\sqrt{N}} \|\widehat{\phi}(\beta) - \phi^0\| = \mathcal{O}_P(N^{-3/8} + \|\beta - \beta^0\|),$$

uniformly over β *in a neighborhood around* β^0.

Proof. Let $\partial_z \ell_{it} = \partial_z \ell_{it}(z^0_{it})$, etc. For all $z_1, z_2 \in \mathcal{Z}$ a second order Taylor expansion of $\ell_{it}(z_1)$ around z_2 gives

$$\begin{aligned}
\ell_{it}(z_1) - \ell_{it}(z_2) &= [\partial_z \ell_{it}(z_1)](z_1 - z_2) - \frac{1}{2}[\partial_{z^2} \ell_{it}(\tilde{z})]\,(z_1 - z_2)^2 \\
&\geq [\partial_z \ell_{it}(z_1)](z_1 - z_2) + \frac{b_{\min}}{2}\,(z_1 - z_2)^2 \\
&= \frac{b_{\min}}{2}\left(z_1 - z_2 + \frac{1}{b_{\min}}[\partial_z \ell_{it}(z_1)]\right)^2 - \frac{1}{2b_{\min}}[\partial_z \ell_{it}(z_1)]^2. \qquad \text{(B.1)}
\end{aligned}$$

where $\tilde{z} \in [\min(z_1, z_2), \max(z_1, z_2)]$. Let $e_{it} := \partial_z \ell_{it}/b_{\min}$. We have

$$\begin{aligned}
0 &\geq \sqrt{NT}\left[\mathcal{L}(\beta^0, \phi^0) - \mathcal{L}(\widehat{\beta}, \widehat{\phi})\right] \\
&= \sum_{i,t} \left[\ell_{it}(z^0_{it}) - \ell_{it}(\widehat{z}_{it})\right] \\
&\geq \frac{b_{\min}}{2} \sum_{i,t} \left[(z^0_{it} - \widehat{z}_{it} + e_{it})^2 - e^2_{it}\right] \\
&= \frac{b_{\min}}{2} \sum_{i,t} \left\{ \left[X'_{it}(\widehat{\beta} - \beta^0) + \widehat{\alpha}_i\widehat{\gamma}_t - \alpha^0_i\gamma^0_t - e_{it}\right]^2 - e^2_{it} \right\}.
\end{aligned}$$

Note that the penalty term of the objective function does not enter here, because it is zero

[1] For this we need the strong-factor assumption (not required before in this theorem) and the normalization $\sum_{i=1}^N \widehat{\alpha}_i^2 = \sum_{t=1}^T \widehat{\gamma}_t^2$ and $\sum_{i=1}^N (\alpha^0_i)^2 = \sum_{t=1}^T (\gamma^0_t)^2$.

when evaluated both at the estimates or at the true values of the parameters.

Let e be the $N \times T$ matrix with entries e_{it}. Let X_k be the $N \times T$ matrix with entries $X_{k,it}$, $k = 1, \ldots, \dim \beta$. Let $\beta \cdot X = \sum_k \beta_k X_k$. In matrix notation, the above inequality reads

$$\mathrm{Tr}(e'e) \geq \mathrm{Tr}\left[\left((\widehat{\beta} - \beta^0) \cdot X + \widehat{\alpha}\widehat{\gamma}' - \alpha^0 \gamma^{0\prime} - e\right)\left((\widehat{\beta} - \beta^0) \cdot X + \widehat{\alpha}\widehat{\gamma}' - \alpha^0 \gamma^{0\prime} - e\right)'\right].$$

Analogous to the consistency proof for linear regression models with interactive fixed effects in (Bai, 2009b) and (Moon and Weidner, 2010a) we can conclude that

$$\frac{1}{NT}\mathrm{Tr}(e'e) \geq \frac{1}{NT}\mathrm{Tr}\left[\mathcal{M}_{\alpha^0}\left((\widehat{\beta} - \beta^0) \cdot X - e\right)\mathcal{M}_{\widehat{\gamma}}\left((\widehat{\beta} - \beta^0) \cdot X - e\right)'\right] \tag{B.2}$$

$$= \frac{1}{NT}\left[\mathrm{Tr}(e'e) + \mathrm{Tr}\left[\mathcal{M}_{\alpha^0}\left((\widehat{\beta} - \beta^0) \cdot X\right)\mathcal{M}_{\widehat{\gamma}}\left((\widehat{\beta} - \beta^0) \cdot X\right)'\right]\right. \tag{B.3}$$

$$\left. + 2\mathrm{Tr}\left[\left((\widehat{\beta} - \beta^0) \cdot X\right)' e\right] + \mathcal{O}_P(\|e\|^2) + \mathcal{O}_P(\|\widehat{\beta} - \beta^0\| \|e\| \max_k \|X_k\|)\right], \tag{B.4}$$

where we used that e.g.

$$\left|\mathrm{Tr}\left(X_k' \mathcal{P}_{\alpha^0} e\right)\right| \leq \mathrm{rank}\left(X_k' \mathcal{P}_{\alpha^0} e\right) \left\|X_k' \mathcal{P}_{\alpha^0} e\right\| \leq \|X_k\| \|e\|,$$
$$\left|\mathrm{Tr}\left(e' \mathcal{P}_{\alpha^0} e\right)\right| \leq \mathrm{rank}\left(e' \mathcal{P}_{\alpha^0} e\right) \left\|e' \mathcal{P}_{\alpha^0} e\right\| \leq \|e\|^2.$$

Lemma D.6 in (Fernández-Val and Weidner, 2013) shows that under our assumptions we have $\|\partial_z \ell\| = \mathcal{O}_P(N^{5/8})$, where $\partial_z \ell$ is the $N \times T$ matrix with entries $\partial_z \ell_{it}$. We thus also have $\|e\| = \mathcal{O}_P(N^{5/8})$. We furthermore have $\|X_k\|^2 \leq \|X_k\|_F^2 = \sum_{it} X_{k,it}^2 = \mathcal{O}_P(NT)$, and therefore $\|X_k\| = \mathcal{O}_P(\sqrt{NT})$. We thus have $\|X_k\| \|e\| = \mathcal{O}_P(N^{13/8})$ and $\|e\|^2 = \mathcal{O}_P(N^{5/4})$. Furthermore

$$\mathrm{Tr}\left(X_k' e\right) = \frac{1}{b_{\min}} \sum_{it} X_{it} \partial_z \ell_{it} = \mathcal{O}_P(\sqrt{NT}).$$

Applying those results and the generalized collinearity assumption to (B.4) gives

$$0 \geq c\|\widehat{\beta} - \beta^0\| + \mathcal{O}_P(N^{-3/8}\|\widehat{\beta} - \beta^0\|) + \mathcal{O}_P(N^{-3/4}).$$

This implies that $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_P(N^{-3/8})$.

Define $e_{it}(\beta) = \partial_z \ell_{it}(X_{it}'\beta + \alpha_i^0\gamma_t^0)/b_{\min}$. Analogous to the above argument we find from $\mathcal{L}(\beta, \widehat{\phi}(\beta)) \geq \mathcal{L}(\beta, \phi^0)$ that

$$\begin{aligned} 0 &\geq \sqrt{NT}\left[\mathcal{L}(\beta, \phi^0) - \mathcal{L}(\beta, \widehat{\phi}(\beta))\right] \\ &= \sum_{i,t}\left[\ell_{it}(X_{it}'\beta + \alpha_i^0\gamma_t^0) - \ell_{it}(X_{it}'\beta + \widehat{\alpha}_i(\beta)\widehat{\gamma}_t(\beta))\right] \\ &= \frac{b_{\min}}{2}\sum_{i,t}\left\{\left[\widehat{\alpha}_i(\beta)\widehat{\gamma}_t(\beta) - \alpha_i^0\gamma_t^0 - e_{it}(\beta)\right]^2 - [e_{it}(\beta)]^2\right\}. \end{aligned}$$

This implies that

$$\begin{aligned} &\mathrm{Tr}(e(\beta)'e(\beta)) \\ &\geq \mathrm{Tr}\left[\left(\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} - e(\beta)\right)\left(\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} - e(\beta)\right)'\right] \\ &= \mathrm{Tr}(e(\beta)'e(\beta)) + \underbrace{\mathrm{Tr}\left[\left(\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right)\left(\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right)'\right]}_{=\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\|_F^2} \\ &+ \mathcal{O}_P\left(\left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\|_F \|e(\beta)\|\right) \end{aligned}$$

Note that since $\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}$ is at most rank 2 we have that $\frac{1}{\sqrt{2}}\left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\|_F \leq \left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\| \leq \left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\|_F$, i.e. the Frobenius and the spectral norm are equivalent.

We have $e_{it}(\beta) = e_{it} + [X_{it}'(\beta - \beta^0)]\partial_{z^2}\ell_{it}(X_{it}'\tilde{\beta} + \alpha_i^0\gamma_t^0)/b_{\min}$, where $\tilde{\beta}$ lies between β and β^0. Therefore $\|e(\beta)\| \leq \|e\| + \mathcal{O}_P(\sqrt{NT}\|\beta - \beta^0\|)$. We thus find

$$0 \geq \frac{1}{NT}\left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\|_F^2 + \mathcal{O}_P\left[\left(N^{-3/8} + \|\beta - \beta^0\|\right)\left\|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime}\right\|_F/\sqrt{NT}\right].$$

From this we conclude that

$$\frac{1}{\sqrt{NT}} \left\| \widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} \right\|_F = \mathcal{O}_P(N^{-3/8} + \|\beta - \beta^0\|).$$

Next, let $d := \left\| \widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} \right\|_F$. By the triangular inequality, $\left\|\alpha^0\gamma^{0\prime}\right\|_F - d \leq \|\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)'\|_F \leq \left\|\alpha^0\gamma^{0\prime}\right\|_F + d$, or equivalently $\|\alpha^0\|\|\gamma^0\| - d \leq \|\widehat{\alpha}(\beta)\|\|\widehat{\gamma}(\beta)\| \leq \|\alpha^0\|\|\gamma^0\| + d$. Using our normalization this gives $\|\alpha^0\|^2 - d \leq \|\widehat{\alpha}(\beta)\|^2 \leq \|\alpha^0\|^2 + d$. This implies that $\|\widehat{\alpha}(\beta)\| = \|\alpha^0\| + \mathcal{O}(d/\|\alpha^0\|) = \|\alpha^0\| + \mathcal{O}(d/\sqrt{N})$, or equivalently $\|\widehat{\gamma}(\beta)\| = \|\gamma^0\| + \mathcal{O}(d/\sqrt{N})$.

Let θ_α be the angle between the vectors α^0 and $\widehat{\alpha}$. We have

$$\begin{aligned} d = \left\| \widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} \right\|_F &\geq \left\| \mathcal{M}_{\widehat{\alpha}(\beta)} \left(\widehat{\alpha}(\beta)\widehat{\gamma}(\beta)' - \alpha^0\gamma^{0\prime} \right) \right\|_F \\ &= \left\| \mathcal{M}_{\widehat{\alpha}(\beta)} \alpha^0\gamma^{0\prime} \right\|_F = \left\| \mathcal{M}_{\widehat{\alpha}(\beta)} \alpha^0 \right\| \left\| \gamma^0 \right\| = \cos(\theta_\alpha)\|\alpha^0\|\|\gamma^0\|. \end{aligned}$$

Therefore $\cos(\theta_\alpha) \leq d/(\|\alpha^0\|\|\gamma^0\|) = \mathcal{O}(d/N)$. Together with $\|\widehat{\alpha}(\beta)\| = \|\alpha^0\| + \mathcal{O}(d/\sqrt{N})$ this implies that $\|\widehat{\alpha}(\beta) - \alpha^0\| = \mathcal{O}(d/\sqrt{N})$. Analogously we conclude that $\|\widehat{\gamma}(\beta) - \gamma^0\| = \mathcal{O}(d/\sqrt{N})$. □

B.1.2 Inverse Expected Incidental Parameter Hessian

The expected incidental parameter Hessian evaluated at the true parameter values is

$$\overline{\mathcal{H}} = \mathbb{E}_\phi[-\partial_{\phi\phi'}\mathcal{L}] = \begin{pmatrix} \overline{\mathcal{H}}^*_{(\alpha\alpha)} & \overline{\mathcal{H}}^*_{(\alpha\gamma)} \\ [\overline{\mathcal{H}}^*_{(\alpha\gamma)}]' & \overline{\mathcal{H}}^*_{(\gamma\gamma)} \end{pmatrix} + \frac{b}{\sqrt{NT}} vv',$$

where $v = v_{NT} = (\alpha^{0\prime}, -\gamma^{0\prime})'$, $\overline{\mathcal{H}}^*_{(\alpha\alpha)} = \operatorname{diag}(\frac{1}{\sqrt{NT}} \sum_t (\gamma_t^0)^2 \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}])$,
$\overline{\mathcal{H}}^*_{(\alpha\gamma)it} = \frac{1}{\sqrt{NT}} \alpha_i^0 \gamma_t^0 \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}]$, and $\overline{\mathcal{H}}^*_{(\gamma\gamma)} = \operatorname{diag}(\frac{1}{\sqrt{NT}} \sum_i (\alpha_i^0)^2 \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}])$.

Lemma B.1.2. *Under Assumptions 4 we have*

$$\left\| \overline{\mathcal{H}}^{-1} - \operatorname{diag}\left(\overline{\mathcal{H}}^*_{(\alpha\alpha)}, \overline{\mathcal{H}}^*_{(\gamma\gamma)} \right)^{-1} \right\|_{\max} = \mathcal{O}_P\left((NT)^{-1/2} \right).$$

The goal of this appendix subsection is to prove Lemma B.1.2, but before doing so it is useful to present two more intermediate lemmas.

In the following we assume that $\alpha_i^0 \neq 0$ and $\gamma_t^0 \neq 0$ holds for all i, t. However, this is only assumed for notational simplicity of the proof. Concretely, $|\alpha_i^0|^{-1}$ and $|\gamma_t^0|^{-1}$ will occur below, but actually only in expressions where $|\alpha_i^0|^{-1}$ is eventually multiplied with α_i^0, and $|\gamma_t^0|^{-1}$ is eventually multiplied with γ_t^0. Therefore, all results also hold without this assumption. More importantly, the proof does never require that α_i^0 and γ_t^0 are bounded away from zero.

Lemma B.1.3. *If the statement of Lemma B.1.2 holds for some constant $b > 0$, then it holds for any constant $b > 0$.*

Proof. Write $\overline{\mathcal{H}} = \overline{\mathcal{H}}^* + \frac{b}{\sqrt{NT}} vv'$, where $\overline{\mathcal{H}}^* = \mathbb{E}_\phi \left[-\frac{\partial^2}{\partial\phi\partial\phi'} \mathcal{L}^* \right]$. Since $\overline{\mathcal{H}}^* v = 0$,

$$\overline{\mathcal{H}}^{-1} = \left(\overline{\mathcal{H}}^*\right)^\dagger + \left(\frac{b}{\sqrt{NT}} vv'\right)^\dagger = \left(\overline{\mathcal{H}}^*\right)^\dagger + \frac{\sqrt{NT}}{b\|vv'\|^2} vv' = \left(\overline{\mathcal{H}}^*\right)^\dagger + \frac{\sqrt{NT}}{b\left[\sum_i (\alpha^0)^2 + \sum_t (\gamma^0)^2\right)]^2} vv',$$

where $\dagger$ refers to the Moore-Penrose pseudo-inverse. Thus, if $\overline{\mathcal{H}}_1$ is the expected Hessian for $b = b_1 > 0$ and $\overline{\mathcal{H}}_2$ is the expected Hessian for $b = b_2 > 0$, $\left\| \overline{\mathcal{H}}_1^{-1} - \overline{\mathcal{H}}_2^{-1} \right\|_{\max} = \left\| \left(\frac{1}{b_1} - \frac{1}{b_2}\right) \frac{\sqrt{NT}}{\left[\sum_i (\alpha^0)^2 + \sum_t (\gamma^0)^2\right)]^2} vv' \right\|_{\max} = \mathcal{O}_P\left((NT)^{-1/2}\right)$. Here we used that $\max_i |\alpha_i^0|$ and $\max_t |\gamma_t^0|$ are bounded and that $\frac{1}{N}\sum_i (\alpha^0)^2$ and $\frac{1}{T}\sum_t (\gamma^0)^2$ converge to positive constants. $\square$

In the following, let $|\alpha^0|$ be the N-vector with entries $|\alpha_i^0|$, and let $|\gamma^0|$ be the T-vector with entries $|\gamma_t^0|$.

Lemma B.1.4. *Let Assumptions 4 hold and let $0 < b \leq b_{\min} \left(1 + \frac{b_{\max}}{b_{\min}}\right)^{-1}$. Then,*

$$\left\| \operatorname{diag}(|\alpha^0|)^{-1}\, \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \operatorname{diag}(|\gamma^0|) \right\|_\infty < 1 - \frac{b}{b_{\max}},$$

and

$$\left\| \operatorname{diag}(|\gamma^0|)^{-1} \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\, \overline{\mathcal{H}}_{(\gamma\alpha)} \operatorname{diag}(|\alpha^0|) \right\|_\infty < 1 - \frac{b}{b_{\max}}.$$

Proof. Let $h_{it} = \mathbb{E}_\phi(-\partial_{z^2}\ell_{it})$, and define

$$\tilde{h}_{it} = (h_{it} - b) - \frac{1}{b^{-1} + \sum_j (\alpha_j^0)^2 \left(\sum_\tau (\gamma_\tau^0)^2 h_{j\tau}\right)^{-1}} \sum_j \frac{(\alpha_j^0)^2 (h_{jt} - b)}{\sum_\tau (\gamma_\tau^0)^2 h_{j\tau}}.$$

By definition, $\overline{\mathcal{H}}_{(\alpha\alpha)} = \overline{\mathcal{H}}^*_{(\alpha\alpha)} + b\alpha^0\alpha^{0\prime}/\sqrt{NT}$ and $\overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}^*_{(\alpha\gamma)} - b\alpha^0\gamma^{0\prime}/\sqrt{NT}$. The matrix $\overline{\mathcal{H}}^*_{(\alpha\alpha)}$ is diagonal with elements $\sum_t (\gamma_t^0)^2 h_{it}/\sqrt{NT}$. The matrix $\overline{\mathcal{H}}^*_{(\alpha\gamma)}$ has elements $\alpha_i^0 \gamma_t^0 h_{it}/\sqrt{NT}$. The Woodbury identity states that

$$\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\alpha^0 \left(\sqrt{NT}\, b^{-1} + \alpha^{0\prime}\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\alpha^0\right)^{-1} \alpha^{0\prime}\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}.$$

Then, $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\,\overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\tilde{H}/\sqrt{NT}$, where $\tilde{H}$ is the $N \times T$ matrix with elements $\alpha_i^0\gamma_t^0\tilde{h}_{it}$. Therefore

$$\left\|\mathrm{diag}(|\alpha^0|)^{-1}\,\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\mathrm{diag}(|\gamma^0|)\right\|_\infty = \max_i \frac{\sum_t (\gamma_t^0)^2 \tilde{h}_{it}}{\sum_t (\gamma_t^0)^2 h_{it}}.$$

The assumption guarantees that $b_{\max} \geq h_{it} \geq b_{\min}$, which implies $h_{jt} - b \geq b_{\min} - b > 0$, and

$$\begin{aligned}\tilde{h}_{it} &> h_{it} - b - \frac{1}{b^{-1}} \sum_j \frac{(\alpha_j^0)^2 (h_{jt} - b)}{\sum_\tau (\gamma_\tau^0)^2 h_{j\tau}} \geq b_{\min} - b \left(1 + \frac{\sum_j (\alpha_j^0)^2}{\sum_\tau (\gamma_\tau^0)^2}\,\frac{b_{\max}}{b_{\min}}\right) \\ &= b_{\min} - b\left(1 + \frac{b_{\max}}{b_{\min}}\right) \geq 0,\end{aligned}$$

where we used the normalization $\sum_j (\alpha_j^0)^2 = \sum_\tau (\gamma_\tau^0)^2$ and the upper bound we impose on b. We conclude that

$$\begin{aligned}&\left\|\mathrm{diag}(|\alpha^0|)^{-1}\,\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\mathrm{diag}(|\gamma^0|)\right\|_\infty \\ &= \max_i \frac{\sum_t (\gamma_t^0)^2 \tilde{h}_{it}}{\sum_t (\gamma_t^0)^2 h_{it}} \\ &= 1 - \min_i \frac{1}{\sum_t (\gamma_t^0)^2 h_{it}} \sum_t (\gamma_t^0)^2 \left(b + \frac{1}{b^{-1} + \sum_j (\alpha_j^0)^2 \left(\sum_\tau (\gamma_\tau^0)^2 h_{j\tau}\right)^{-1}} \sum_j \frac{(\alpha_j^0)^2 (h_{jt} - b)}{\sum_\tau (\gamma_\tau^0)^2 h_{j\tau}}\right)\end{aligned}$$

$$< 1 - \frac{\sum_t (\gamma_t)^2 b}{\sum_t (\gamma_t)^2 b_{\max}}$$
$$= 1 - \frac{b}{b_{\max}}.$$

Analogously one finds that $\left\| \text{diag}(|\gamma^0|)^{-1} \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \, \overline{\mathcal{H}}_{(\gamma\alpha)} \text{diag}(|\alpha^0|) \right\|_\infty < 1 - \frac{b}{b_{\max}}$. □

Proof. [Proof of Lemma B.1.2] We choose $b \le b_{\min} \left(1 + \frac{b_{\max}}{b_{\min}}\right)^{-1}$, so that Lemma B.1.4 becomes applicable. According to Lemma B.1.3 the choice of b has no effect on the general validity of the lemma for all $b > 0$.

By the inversion formula for partitioned matrices,

$$\overline{\mathcal{H}}^{-1} = \begin{pmatrix} A & -A \, \overline{\mathcal{H}}_{(\alpha\gamma)} \, \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \, \overline{\mathcal{H}}_{(\gamma\alpha)} \, A & \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} + \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \, \overline{\mathcal{H}}_{(\gamma\alpha)} \, A \, \overline{\mathcal{H}}_{(\alpha\gamma)} \, \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \end{pmatrix},$$

with $A := (\overline{\mathcal{H}}_{(\alpha\alpha)} - \overline{\mathcal{H}}_{(\alpha\gamma)} \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)})^{-1}$. The Woodbury identity states that

$$\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} - \underbrace{\overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} \alpha^0 \left(\sqrt{NT}/b + \alpha^{0\prime} \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} \alpha^0\right)^{-1} \alpha^{0\prime} \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}}_{=: C_{(\alpha\alpha)}},$$
$$\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} = \overline{\mathcal{H}}_{(\gamma\gamma)}^{*-1} - \underbrace{\overline{\mathcal{H}}_{(\gamma\gamma)}^{*-1} \gamma^0 \left(\sqrt{NT}/b + \gamma^{0\prime} \overline{\mathcal{H}}_{(\gamma\gamma)}^{*-1} \gamma^0\right)^{-1} \gamma^{0\prime} \overline{\mathcal{H}}_{(\gamma\gamma)}^{*-1}}_{=: C_{(\gamma\gamma)}}.$$

By our assumptions we have $\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{*-1}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}_{(\alpha\gamma)}^{*}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$. Therefore[2]

$$\|C_{(\alpha\alpha)}\|_{\max} \le \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}\|_\infty^2 \left\|\alpha^0 \alpha^{0\prime}\right\|_{\max} \left(\sqrt{NT}/b + \alpha^{0\prime} \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} \alpha^0\right)^{-1} = \mathcal{O}_P(1/\sqrt{NT}),$$
$$\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_\infty \le \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}\|_\infty + N \|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1).$$

Analogously, $\|C_{(\gamma\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$ and $\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_\infty = \mathcal{O}_P(1)$. Furthermore, $\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \le \|\overline{\mathcal{H}}_{(\alpha\gamma)}^{*}\|_{\max} + b\|\alpha^0 \gamma^{0\prime}\|/\sqrt{NT} = \mathcal{O}_P(1/\sqrt{NT})$.

[2]Here and in the following me make use of the inequalities $\|AB\|_{\max} < \|A\|_\infty \|B\|_{\max}$, $\|AB\|_{\max} < \|A\|_{\max} \|B'\|_\infty$, $\|A\|_\infty \le n\|A\|_{\max}$, which hold for any $m \times n$ matrix A and $n \times p$ matrix B.

We also have $\left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$ and $\left\|\text{diag}(|\alpha^0|)^{-1}C_{\alpha\alpha}\right\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$. Those last two results do not require α_i^0 to be bounded away from zero, because in those expressions the $|\alpha_i^0|^{-1}$ gets multiplied with α_i^0 and we have $|\alpha_i^0|^{-1}\alpha_i^0 = \mathcal{O}(1)$. We thus have

$$
\begin{aligned}
&\left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} \\
&= \left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)} - \text{diag}(|\alpha^0|)^{-1}C_{\alpha\alpha}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} \\
&= \left\|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)} - \text{diag}(|\alpha^0|)^{-1}C_{\alpha\alpha}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} \\
&\leq \left\|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\right\|_{\infty}\left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} + \left\|\text{diag}(|\alpha^0|)^{-1}C_{\alpha\alpha}\right\|_{\infty}\left\|\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} \\
&\leq N\left\|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\right\|_{\infty}\left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\max} + N\left\|\text{diag}(|\alpha^0|)^{-1}C_{\alpha\alpha}\right\|_{\max}\left\|\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_{\infty} \\
&= \mathcal{O}_P(1).
\end{aligned}
$$

Define $D := \text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\text{diag}(|\alpha^0|)$ and

$$
\begin{aligned}
B &:= \left(\mathbf{1}_N - \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\right)^{-1} - \mathbf{1}_N \\
&= \text{diag}(|\alpha^0|)\left[(\mathbf{1}_N - D)^{-1} - \mathbf{1}_N\right]\text{diag}(|\alpha^0|)^{-1} \\
&= \text{diag}(|\alpha^0|)\left(\sum_{n=1}^{\infty} D^n\right)\text{diag}(|\alpha^0|)^{-1} \\
&= \text{diag}(|\alpha^0|)\left(\sum_{n=0}^{\infty} D^n\right)\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}.
\end{aligned}
$$

Note that $A = \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - C_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B$. By Lemma B.1.4, we have

$$
\begin{aligned}
\|D\|_{\infty} &= \left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\text{diag}(|\gamma^0|)\text{diag}(|\gamma^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\text{diag}(|\alpha^0|)\right\|_{\infty} \\
&\leq \left\|\text{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\text{diag}(|\gamma^0|)\right\|_{\infty}\left\|\text{diag}(|\gamma^0|)^{-1}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\text{diag}(|\alpha^0|)\right\|_{\infty} \\
&< \left(1 - \frac{b}{b_{\max}}\right)^2 < 1.
\end{aligned}
$$

We thus have

$$\begin{aligned}\|B\|_{\max} &\leq \left\|\operatorname{diag}(|\alpha^0|)\right\|_\infty \left(\sum_{n=0}^{\infty} \|D\|_\infty^n\right) \left\|\operatorname{diag}(|\alpha^0|)^{-1}\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\right\|_\infty \left\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\right\|_\infty \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max} \\ &\leq \max_i |\alpha_i^0| \left(\sum_{n=0}^{\infty}\left(1-\frac{b}{b_{\max}}\right)^{2n}\right) \mathcal{O}_P(1)\,\mathcal{O}_P(1)\,\mathcal{O}_P(1/\sqrt{NT}) = \mathcal{O}_P(1/\sqrt{NT}).\end{aligned}$$

By the triangle inequality,

$$\|A\|_\infty \leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_\infty + N\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_\infty \|B\|_{\max} = \mathcal{O}_P(1).$$

Thus, for the different blocks of

$$\overline{\mathcal{H}}^{-1} - \begin{pmatrix} \overline{\mathcal{H}}_{(\alpha\alpha)}^{*} & 0 \\ 0 & \overline{\mathcal{H}}_{(\gamma\gamma)}^{*} \end{pmatrix}^{-1} = \begin{pmatrix} A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} & -A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A & \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} - C_{(\gamma\gamma)} \end{pmatrix},$$

we find

$$\begin{aligned}\left\|A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}\right\|_{\max} &= \left\|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}B - C_{(\alpha\alpha)}\right\|_{\max} \\ &\leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_\infty \|B\|_{\max} - \|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT}), \\ \left\|-A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\right\|_{\max} &\leq \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_\infty = \mathcal{O}_P(1/\sqrt{NT}), \\ \left\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} - C_{(\gamma\gamma)}\right\|_{\max} &\leq \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_\infty^2 \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} + \|C_{(\gamma\gamma)}\|_{\max} \\ &\leq N\|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_\infty^2 \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}^2 + \|C_{(\gamma\gamma)}\|_{\max} \\ &= \mathcal{O}_P(1/\sqrt{NT}).\end{aligned}$$

The bound $\mathcal{O}_P(1/\sqrt{NT})$ for the max-norm of each block of the matrix yields the same bound for the max-norm of the matrix itself. $\square$

B.1.3 Local Concavity of the Objective Function

The consistency result for $\widehat{\phi}(\beta)$ in Lemma B.1.1 is not sufficient to apply the general expansion results in Ferndez-Val and Weidner (2013).[3] The goal of this section is to close this gap by using local concavity of $\mathcal{L}(\beta,\phi)$ in ϕ around ϕ^0.

In the following we only consider parameter values that satisfy the constraint $\sum_i \alpha_i^2 = \sum_t \gamma_t^2$ (otherwise there are additional terms in the Hessian from the penalty terms, which we do not want to consider). Let $\ell_{it}(\beta,\pi_{it}) = \ell_{it}(z_{it})$, where $\pi_{it} = \alpha_i\gamma_t$ and $z_{it} = X_{it}'\beta + \alpha_i\gamma_t$. Let $h_{it}(\beta,\pi_{it}) = -\partial_\pi \ell_{it}(\beta,\pi_{it})$. The incidental parameter Hessian reads

$$\mathcal{H}(\beta,\phi) = -\partial_{\phi\phi'}\mathcal{L}(\beta,\phi) = \begin{pmatrix} \mathcal{H}^*_{(\alpha\alpha)}(\beta,\phi) & \mathcal{H}^*_{(\alpha\gamma)}(\beta,\phi) \\ [\mathcal{H}^*_{(\alpha\gamma)}(\beta,\phi)]' & \mathcal{H}^*_{(\gamma\gamma)}(\beta,\phi) \end{pmatrix} + \frac{b}{\sqrt{NT}} v(\phi)[v(\phi)]',$$

where $v(\phi) = (\alpha', -\gamma')'$, $\mathcal{H}^*_{(\alpha\alpha)}(\beta,\phi) = \operatorname{diag}[\frac{1}{\sqrt{NT}}\sum_t \gamma_t^2 h_{it}(\beta,\alpha_i\gamma_t)]$,
$\mathcal{H}^*_{(\alpha\gamma)it}(\beta,\phi) = \frac{1}{\sqrt{NT}}\alpha_i\gamma_t h_{it}(\beta,\alpha_i\gamma_t) - \frac{1}{\sqrt{NT}}\partial_z \ell_{it}(z_{it})$,
and $\mathcal{H}^*_{(\gamma\gamma)}(\beta,\phi) = \operatorname{diag}[\frac{1}{\sqrt{NT}}\sum_i \alpha_i^2 h_{it}(\beta,\alpha_i\gamma_t)]$. We decompose the Hessian as $\mathcal{H}(\beta,\phi) = H(\beta,\phi) + F(\beta,\phi)$, where

$$\begin{aligned} H(\beta,\phi) &= \begin{pmatrix} H_{(\alpha\alpha)}(\beta,\phi) & H_{(\alpha\gamma)}(\beta,\phi) \\ [H_{(\alpha\gamma)}(\beta,\phi)]' & H_{(\gamma\gamma)}(\beta,\phi) \end{pmatrix} \\ &= \begin{pmatrix} H^*_{(\alpha\alpha)}(\beta,\phi) & H^*_{(\alpha\gamma)}(\beta,\phi) \\ [H^*_{(\alpha\gamma)}(\beta,\phi)]' & H^*_{(\gamma\gamma)}(\beta,\phi) \end{pmatrix} + \frac{b}{\sqrt{NT}} v(\phi)[v(\phi)]', \\ F(\beta,\phi) &= \begin{pmatrix} 0_{N\times N} & F_{(\alpha\gamma)}(\beta,\phi) \\ [F_{(\alpha\gamma)}(\beta,\phi)]' & 0_{T\times T} \end{pmatrix}, \end{aligned}$$

where $H^*_{(\alpha\alpha)}(\beta,\phi) = \mathcal{H}^*_{(\alpha\alpha)}(\beta,\phi)$, $H^*_{(\alpha\gamma)it}(\beta,\phi) = \frac{1}{\sqrt{NT}}\alpha_i\gamma_t h_{it}(\beta,\alpha_i\gamma_t)$,
$H^*_{(\gamma\gamma)}(\beta,\phi) = \mathcal{H}^*_{(\gamma\gamma)}(\beta,\phi)$, and $F_{(\alpha\gamma)it}(\beta,\phi) = -\frac{1}{\sqrt{NT}}\partial_z \ell_{it}(z_{it})$.

Lemma B.1.5. *For $\lambda_{\min}[H(\beta,\phi)]$, the smallest eigenvalue of $H(\beta,\phi)$, we have*

$$\lambda_{\min}[H(\beta,\phi)] \geq \min\left\{ \min_{i\in\{1,\ldots,N\}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \gamma_t^2 \left[h_{it}(\beta,\alpha_i\gamma_t) - |h_{it}(\beta,\alpha_i\gamma_t) - b|\right] \right.,$$

[3] Assumption B.1(iii) of the general expansion requires $\|\widehat{\phi}(\beta) - \phi^0\|_q = o_P\left((NT)^{-\epsilon}\right)$ for some $q > 4$ and some $\epsilon \geq 0$.

$$\min_{t\in\{1,\ldots,T\}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \alpha_i^2 \left[h_{it}(\beta,\alpha_i\gamma_t) - |h_{it}(\beta,\alpha_i\gamma_t) - b|\right] \Bigg\}.$$

Thus, if $h_{it}(\beta,\alpha_i\gamma_t) \geq b$ for all i,t, then we have

$$\lambda_{\min}[H(\beta,\phi)] \geq \min\left\{ \frac{b}{\sqrt{NT}} \sum_{t=1}^{T} \gamma_t^2,\ \frac{b}{\sqrt{NT}} \sum_{i=1}^{N} \alpha_i^2 \right\}.$$

We will only use the second bound for $\lambda_{\min}[H(\beta,\phi)]$ provided in the lemma, but the first bound for $\lambda_{\min}[H(\beta,\phi)]$ provided in the lemma shows that the condition $h_{it}(\beta,\alpha_i\gamma_t) \geq b$ is not necessary to appropriately bound $\lambda_{\min}[H(\beta,\phi)]$, but it is convenient.

Proof. In the following proof we drop all parameter arguments from the functions. Define $g_i^{(1)} := \frac{b}{\sqrt{NT}} \sum_{t=1}^{T} \gamma_t^2 - \frac{2}{\sqrt{NT}} \sum_{t=1}^{T} 1(b > h_{it})\gamma_t^2(b - h_{it})$ and $g_t^{(2)} := \frac{b}{\sqrt{NT}} \sum_{i=1}^{N} \alpha_i^2 - \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} 1(b > h_{it})\alpha_i^2(b - h_{it})$. Equivalently we can write $g_i^{(1)} = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \gamma_t^2[h_{it}(\beta,\alpha_i\gamma_t) - |h_{it}(\beta,\alpha_i\gamma_t) - b|]$ and $g_t^{(2)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \alpha_i^2 \left[h_{it}(\beta,\alpha_i\gamma_t) - |h_{it}(\beta,\alpha_i\gamma_t) - b|\right]$.

Let G be the diagonal $(N+T)\times(N+T)$ matrix with diagonal elements given by $g_i^{(1)}$, $i = 1,\ldots,N$ and $g_t^{(2)}$, $t = 1,\ldots,T$, in that order. It is easy to verify that $H = H(\beta,\phi)$ satisfies

$$\begin{aligned} H = G &+ \frac{b}{\sqrt{NT}}(\alpha', 0_{1\times T})'(\alpha', 0_{1\times T}) + \frac{b}{\sqrt{NT}}(0_{1\times N}, \gamma')'(0_{1\times N}, \gamma') \\ &+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} 1(h_{it} \geq b)(h_{it} - b)(\gamma_t e'_{N,i}, \alpha_i e'_{T,t})'(\gamma_t e'_{N,i}, \alpha_i e'_{T,t}) \\ &+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} 1(b > h_{it})(b - h_{it})(\gamma_t e'_{N,i}, -\alpha_i e'_{T,t})'(\gamma_t e'_{N,i}, -\alpha_i e'_{T,t}). \end{aligned}$$

This shows that $H - G$ is positive definite, i.e. $H \geq G$, which implies that $\lambda_{\min}(H) \geq \lambda_{\min}(G)$. Since G is diagonal we have $\lambda_{\min}(G) = \min\{\min_i g_i^{(1)}, \min_t g_t^{(2)}\}$. $\square$

Lemma B.1.6. *Let Assumption 4 be satisfied, and let $r_\beta = r_{\beta,NT} = o_P(1)$ and $r_\phi = r_{\phi,NT} = o_P(\sqrt{N})$. Then, $\mathcal{H}(\beta,\phi)$ is positive definite for all $\beta \in \mathcal{B}(r_\beta,\beta^0)$ and $\phi \in \mathcal{B}(r_\phi,\phi^0)$,* *wpa1, where $\mathcal{B}(r_\beta, \beta^0)$ and $\mathcal{B}(r_\phi, \phi^0)$ are balls under the Euclidian norm. This implies that $\mathcal{L}(\beta, \phi)$ is strictly concave in $\phi \in \mathcal{B}(r_\phi, \phi^0)$, for all $\beta \in \mathcal{B}(r_\beta, \beta^0)$.*

Proof. Let $\beta \in \mathcal{B}(r_\beta, \beta^0)$ and $\phi \in \mathcal{B}(r_\phi, \phi^0)$. We have $\mathcal{H}(\beta, \phi) = H(\beta, \phi) + F(\beta, \phi)$. Weyl's inequality guarantees that $\lambda_{\min}[\mathcal{H}(\beta, \phi)] \geq \lambda_{\min}[H(\beta, \phi)] - \|F(\beta, \phi)\|$, where $\|F(\beta, \phi)\|$ is the spectral norm of $F(\beta, \phi)$.

By choosing $b = b_{\min}$ in Lemma B.1.5,
we find $\lambda_{\min}[H(\beta, \phi)] \geq b_{\min} \min\left\{\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \gamma_t^2,\ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \alpha_i^2\right\}$. Thus, the desired result follows if we can show that $\|F(\beta, \phi)\| = o_P(1)$, or equivalently $\|F_{(\alpha\gamma)}(\beta, \phi)\| = o_P(1)$.

Remember that $F_{(\alpha\gamma)it}(\beta, \phi) = -\frac{1}{\sqrt{NT}} \partial_\pi \ell_{it}(\beta, \alpha_i \gamma_t)$. A Taylor expansion gives

$$\partial_\pi \ell_{it}(\beta, \alpha_i\gamma_t) = \partial_\pi \ell_{it}(\beta^0, \alpha_i^0 \gamma_t^0) + (\beta - \beta^0)' \partial_{\beta\pi} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it}) + (\alpha_i \gamma_t - \alpha_i^0 \gamma_t^0) \partial_{\pi^2} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it}).$$

The spectral norm of the $N \times T$ matrix with entries $\partial_{\beta_k \pi} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it})$ is bounded by the Frobenius norm of this matrix, which is of order $\sqrt{NT}$, since we assume uniformly bounded moments for $\partial_{\beta_k \pi} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it})$. The spectral norm of the $N \times T$ matrix with entries $(\alpha_i \gamma_t - \alpha_i^0 \gamma_t^0) \partial_{\pi^2} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it})$ is also bounded by the Frobenius norm of this matrix, which is equal to $\sqrt{\sum_{it} (\alpha_i \gamma_t - \alpha_i^0 \gamma_t^0)^2 [\partial_{\pi^2} \ell_{it}(\tilde{\beta}_{it}, \tilde{\pi}_{it})]^2}$ and thus bounded by $b_{\max} \sqrt{\sum_{it} (\alpha_i \gamma_t - \alpha_i^0 \gamma_t^0)^2} = b_{\max} \|\alpha\gamma' - \alpha^0 \gamma^{0\prime}\|_F$. We thus find

$$\begin{aligned}
\left\|F_{(\alpha\gamma)it}(\beta, \phi)\right\| &\leq \frac{1}{\sqrt{NT}} \left(\|\partial_\pi \ell_{it}\| + \mathcal{O}_P(\sqrt{NT}) \|\beta - \beta^0\| + b_{\max} \|\alpha\gamma' - \alpha^0 \gamma^{0\prime}\|_F\right) \\
&= \mathcal{O}_P\left(\frac{1}{\sqrt{NT}} N^{5/8}\right) + \mathcal{O}_P(r_\beta) + \mathcal{O}_P(r_\phi / \sqrt{N}) \\
&= o_P(1),
\end{aligned}$$

where we also used that $\|\alpha\gamma' - \alpha^0 \gamma^{0\prime}\|_F = \mathcal{O}_P(\sqrt{N}) \|\phi - \phi^0\|$. We thus have $\|F_{(\alpha\gamma)}(\beta, \phi)\| = o_P(1)$, which was left to show. □

B.1.4 Proof of Theorem 2.4.1

Proof. The above results show that all regularity conditions are satisfied to apply the expansion results in Theorem B.1 and Corollary B.2 of (Fernández-Val and Weidner, 2013). Note that the objective function is not globally concave, but is locally concave according to Lemma B.1.6, and due to the consistency result in Lemma B.1.1 the local concavity is sufficient here. From (Fernández-Val and Weidner, 2013) we thus know that

$$\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}_\infty^{-1} U + o_P(1),$$

where $\overline{W}_\infty = \text{plim}_{N,T\to\infty} \overline{W}$, $U = U^{(0)} + U^{(1)}$, and

$$\begin{aligned}
\overline{W} &= -\frac{1}{\sqrt{NT}} \left(\partial_{\beta\beta'} \overline{\mathcal{L}} + [\partial_{\beta\phi'} \overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\, [\partial_{\phi\beta'} \overline{\mathcal{L}}] \right), \\
U^{(0)} &= \partial_\beta \mathcal{L} + [\partial_{\beta\phi'} \overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1} \mathcal{S}, \\
U^{(1)} &= [\partial_{\beta\phi'} \widetilde{\mathcal{L}}] \overline{\mathcal{H}}^{-1} \mathcal{S} - [\partial_{\beta\phi'} \overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\, \widetilde{\mathcal{H}}\, \overline{\mathcal{H}}^{-1}\, \mathcal{S} \\
&\quad + \frac{1}{2} \sum_{g=1}^{\dim \phi} \left(\partial_{\beta\phi'\phi_g} \overline{\mathcal{L}} + [\partial_{\beta\phi'} \overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1} [\partial_{\phi\phi'\phi_g} \overline{\mathcal{L}}] \right) [\overline{\mathcal{H}}^{-1} \mathcal{S}]_g \overline{\mathcal{H}}^{-1} \mathcal{S}.
\end{aligned} \tag{B.5}$$

We could use these formulas as a starting point to derive the result of the theorem. It is, however, convenient to note that the first order asymptotic results for the interactive model $\ell_{it}(\beta,\, \alpha_i \gamma_t) = \ell_{it}(z_{it})$ are closely related to those obtained from the infeasible model $\ell_{it}^{\dagger}(\beta, \alpha_i, \gamma_t) := \ell_{it}(\beta,\, \alpha_i \gamma_t^0 + \alpha_i^0 \gamma_t - \alpha_i^0 \gamma_t^0)$. This infeasible model can also be written in terms of a "standard" additive model by defining $\alpha_i^{(\dagger)} := \alpha_i / \alpha_i^0$, $\gamma_t^{(\dagger)} = \gamma_t / \gamma_t^0$, and $\ell_{it}^{(\dagger)}(\beta, \alpha_i^{(\dagger)} + \gamma_t^{(\dagger)}) \equiv \ell_{it}\left(\beta, \alpha_i^0 \gamma_t^0 (\alpha_i^{(\dagger)} + \gamma_t^{(\dagger)} - 1)\right)$, where we have to assume $\alpha_i^0 \neq 0$ and $\gamma_t^0 \neq 0$, however (ignore this for the moment). The estimator for β in model $\ell_{it}^{\dagger}$ and $\ell_{it}^{(\dagger)}$ are identical, i.e. $\widehat{\beta}^{\dagger} = \widehat{\beta}^{(\dagger)}$. The asymptotic results for the model $\ell_{it}^{(\dagger)}(\beta, \alpha_i^{(\dagger)} + \gamma_t^{(\dagger)})$ are known from (Fernández-Val and Weidner, 2013), namely $\sqrt{NT}\left(\widehat{\beta}^{(\dagger)} - \beta^0\right) \to_d \left[\overline{W}_\infty^{(\dagger)}\right]^{-1} \mathcal{N}(\kappa \overline{B}_\infty^{(\dagger)} + \kappa^{-1} \overline{D}_\infty^{(\dagger)},\, \overline{W}_\infty^{(\dagger)})$, with $\overline{B}_\infty^{(\dagger)}$, $\overline{D}_\infty^{(\dagger)}$ and $\overline{W}_\infty^{(\dagger)}$ defined there.

The relation between certain derived quantities of model $\ell_{it}^{(\dagger)}$ and ℓ_{it} is given by:

$$\begin{aligned}
\left[\overline{\mathcal{H}}^{-1}\right]^{(\dagger)} &= \operatorname{diag}(\alpha^{0\prime}, \gamma^{0\prime})^{-1}\, \overline{\mathcal{H}}^{-1}\, \operatorname{diag}(\alpha^{0\prime}, \gamma^{0\prime})^{-1}, \\
\partial_{z^q} \ell_{it}^{(\dagger)} &= (\alpha_i^0 \gamma_t^0)^q \partial_{\pi} \ell_{it}, \\
\partial_{\beta \pi^q} \ell_{it}^{(\dagger)} &= (\alpha_i^0 \gamma_t^0)^q \partial_{\beta\pi} \ell_{it}, \\
\Xi_{it}^{(\dagger)} &= (\alpha_i^0 \gamma_t^0)^{-1}\, \Xi_{it}.
\end{aligned}$$

Using this we find that $\overline{B}_{\infty}^{(\dagger)}$, $\overline{D}_{\infty}^{(\dagger)}$ and $\overline{W}_{\infty}^{(\dagger)}$ can be written in terms of model ℓ_{it} quantities as

$$\begin{aligned}
\overline{B}_{\infty}^{(\dagger)} &= -\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{t=1}^{T}\sum_{\tau=t}^{T} \gamma_t^0 \gamma_\tau^0 \mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi} \ell_{i\tau}\right) + \frac{1}{2}\sum_{t=1}^{T} (\gamma_t^0)^2 \mathbb{E}_\phi (D_{\beta\pi^2} \ell_{it})}{\sum_{t=1}^{T} (\gamma_t^0)^2 \mathbb{E}_\phi\left(\partial_{\pi^2} \ell_{it}\right)}\right], \\
\overline{D}_{\infty}^{(\dagger)} &= -\overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\sum_{i=1}^{N} (\alpha_i^0)^2 \mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi} \ell_{it} + \frac{1}{2} D_{\beta\pi^2} \ell_{it}\right)}{\sum_{i=1}^{N} (\alpha_i^0)^2 \mathbb{E}_\phi\left(\partial_{\pi^2} \ell_{it}\right)}\right], \\
\overline{W}_{\infty}^{(\dagger)} &= -\overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}_\phi\left(\partial_{\beta\beta'} \ell_{it} - \partial_{\pi^2} \ell_{it} \Xi_{it} \Xi_{it}'\right)\right].
\end{aligned}$$

What is left to do is to adjust these known results for $\widehat{\beta}^{\dagger} = \widehat{\beta}^{(\dagger)}$ for the discrepancy between $\widehat{\beta}$ and $\widehat{\beta}^{\dagger}$, i.e. accounting the difference between model ℓ_{it} and $\ell_{it}^{\dagger}$, using the expansion results in (B.5) above.

We only consider correctly specified models here, which implies that $\operatorname{Var}(\mathcal{S}) = \mathbb{E}[\mathcal{S}\mathcal{S}'] = \frac{1}{\sqrt{NT}} \overline{\mathcal{H}}^{*}$ (Bartlett identity). Using this we find that

$$\begin{aligned}
&\mathbb{E}_\phi\left\{\frac{1}{2}\sum_{g=1}^{\dim\phi}\left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \overline{\mathcal{H}}^{-1}\mathcal{S}\right\} \\
&\qquad = \frac{1}{2\sqrt{NT}}\sum_{g,h=1}^{\dim\phi}\left(\partial_{\beta\phi_g\phi_h}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi_g\phi_h}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}]_{gh}, \qquad \text{(B.6)}
\end{aligned}$$

where the difference between $\overline{\mathcal{H}}^{*}$ and $\overline{\mathcal{H}}$ does not matter. Since $U^{(1)}$ only contributes bias and no variance to $\widehat{\beta}$ it is thus sufficient to evaluate the second line in (B.6), instead of the more complicated first line.

Comparing model ℓ_{it} and $\ell_{it}^{\dagger}$ we find that

$$
\begin{aligned}
\mathcal{S} &= \mathcal{S}^{\dagger}, \\
\partial_{\beta}\mathcal{L} &= \partial_{\beta}\mathcal{L}^{\dagger}, \\
\overline{\mathcal{H}} &= \overline{\mathcal{H}}^{\dagger}, \\
\widetilde{\mathcal{H}} &= \widetilde{\mathcal{H}}^{\dagger} + \frac{1}{\sqrt{NT}} \begin{pmatrix} 0_{N\times N} & [-\partial_{\pi}\ell_{it}]_{N\times T} \\ [-\partial_{\pi}\ell_{it}]_{T\times N} & 0_{T\times T} \end{pmatrix}, \\
\partial_{\beta\phi'}\overline{\mathcal{L}} &= \partial_{\beta\phi'}\overline{\mathcal{L}}^{\dagger}, \\
\partial_{\beta\phi'}\widetilde{\mathcal{L}} &= \partial_{\beta\phi'}\widetilde{\mathcal{L}}^{\dagger}, \\
\partial_{\beta\beta'}\overline{\mathcal{L}} &= \partial_{\beta\beta'}\overline{\mathcal{L}}^{\dagger}, \\
\partial_{\beta_k\phi\phi'}\overline{\mathcal{L}} &= \partial_{\beta_k\phi\phi'}\overline{\mathcal{L}}^{\dagger} + \frac{1}{\sqrt{NT}} \begin{pmatrix} 0_{N\times N} & [\partial_{\beta_k\pi}\overline{\ell}_{it}]_{N\times T} \\ [\partial_{\beta_k\pi}\overline{\ell}_{it}]_{T\times N} & 0_{T\times T} \end{pmatrix}, \\
\partial_{\alpha_i\alpha_j\alpha_k}\overline{\mathcal{L}} &= \partial_{\alpha_i\alpha_j\alpha_k}\overline{\mathcal{L}}^{\dagger}, \\
\partial_{\alpha_i\alpha_j\gamma_t}\overline{\mathcal{L}} &= \partial_{\alpha_i\alpha_j\gamma_t}\overline{\mathcal{L}}^{\dagger} + 1(i=j)\ \frac{2}{\sqrt{NT}}\ \gamma_t^0\ \partial_{\pi^2}\overline{\ell}_{it}, \\
\partial_{\alpha_i\gamma_t\gamma_s}\overline{\mathcal{L}} &= \partial_{\alpha_i\gamma_t\gamma_s}\overline{\mathcal{L}}^{\dagger} + 1(t=s)\ \frac{2}{\sqrt{NT}}\ \alpha_i^0\ \partial_{\pi^2}\overline{\ell}_{it}, \\
\partial_{\gamma_t\gamma_s\gamma_u}\overline{\mathcal{L}} &= \partial_{\gamma_t\gamma_s\gamma_u}\overline{\mathcal{L}}^{\dagger}.
\end{aligned}
$$

Thus, we have $U^{(0)} = U^{(0)\dagger}$ (this term contributes variance, but no bias) and for the terms in $U^{(1)}$ (which contribute bias, but no variance)

$$
[\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\beta\phi'}\widetilde{\mathcal{L}}^{\dagger}]\left[\overline{\mathcal{H}}^{-1}\right]^{\dagger}\mathcal{S}^{\dagger} = 0,
$$

i.e. no additional bias contribution from this term.

$$
\begin{aligned}
&- [\partial_{\beta_k\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,\widetilde{\mathcal{H}}\,\overline{\mathcal{H}}^{-1}\,\mathcal{S} - \left\{-[\partial_{\beta_k\phi'}\overline{\mathcal{L}}]^{\dagger}\,[\overline{\mathcal{H}}^{-1}]^{\dagger}[\widetilde{\mathcal{H}}]^{\dagger}[\overline{\mathcal{H}}^{-1}]^{\dagger}[\mathcal{S}]^{\dagger}\right\} \\
&= -\frac{1}{\sqrt{NT}}[\partial_{\beta_k\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1} \begin{pmatrix} 0_{N\times N} & [-\partial_{\pi}\ell_{it}]_{N\times T} \\ [-\partial_{\pi}\ell_{it}]_{T\times N} & 0_{T\times T} \end{pmatrix} \overline{\mathcal{H}}^{-1}\mathcal{S}
\end{aligned}
$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \underbrace{\left\{ [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_i \; \partial_\pi \ell_{it} \; [\overline{\mathcal{H}}^{-1}]_{tt} \sum_{j=1}^{N} \alpha_j^0 \partial_\pi \ell_{jt} \} \right\}}_{=:T_{\text{new}_1}}$$

$$+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \underbrace{\left\{ [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_t \; \partial_\pi \ell_{it} \; [\overline{\mathcal{H}}^{-1}]_{ii} \sum_{s=1}^{T} \gamma_s^0 \partial_\pi \ell_{is} \right\}}_{=:T_{\text{new}_2}}$$

$$+ o_P(1),$$

where $T_{\text{new}} = T_{\text{new}_1} + T_{\text{new}_2}$,and the off-diagonal elements of the second $\overline{\mathcal{H}}^{-1}$ only give vanishing contributions. Taking expectations and using that $\mathbb{E}_\phi \left[\partial_\pi \ell_{it} \partial_\pi \ell_{js} \right] = -1(i = j)1(t = s)\partial_{\pi^2} \overline{\ell}_{it}$ we obtain the following non-vanishing bias contribution:

$$\begin{aligned}
\mathbb{E}_\phi T_{\text{new}} &= -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_i \; \partial_{\pi^2} \overline{\ell}_{it} \; \alpha_i^0 \; [\overline{\mathcal{H}}^{-1}]_{tt} + [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_t \; \partial_{\pi^2} \overline{\ell}_{it} \; \gamma_t^0 \; [\overline{\mathcal{H}}^{-1}]_{ii} \right\} \\
&= \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_i \alpha_i^0 \; \partial_{\pi^2} \overline{\ell}_{it}}{\sum_{i=1}^{N} (\alpha_i^0)^2 \partial_{\pi^2} \overline{\ell}_{it}} \\
&+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} [\partial_{\beta_k \phi'} \overline{\mathcal{L}}\, \overline{\mathcal{H}}^{-1}]_t \gamma_t^0 \; \partial_{\pi^2} \overline{\ell}_{it}}{\sum_{t=1}^{T} (\gamma_t^0)^2 \partial_{\pi^2} \overline{\ell}_{it}} + \mathcal{O}_P(1/\sqrt{NT}),
\end{aligned}$$

where we used our result on the structure of $\overline{\mathcal{H}}^{-1}$.

$$\begin{aligned}
&\frac{1}{2\sqrt{NT}} \sum_{g,h=1}^{\dim \phi} \partial_{\beta_k \phi_g \phi_h} \overline{\mathcal{L}} \; [\overline{\mathcal{H}}^{-1}]_{gh} - \frac{1}{2\sqrt{NT}} \sum_{g,h=1}^{\dim \phi} \partial_{\beta_k \phi_g \phi_h} \overline{\mathcal{L}}^\dagger \; [\overline{\mathcal{H}}^{-1}]_{gh}^\dagger \\
&\quad = \frac{1}{2NT} \text{Tr} \left[\begin{pmatrix} 0_{N \times N} & [\partial_{\beta_k \pi} \, \ell_{it}]_{N \times T} \\ [\partial_{\beta_k \pi} \, \ell_{it}]_{T \times N} & 0_{T \times T} \end{pmatrix} \overline{\mathcal{H}}^{-1} \right] = \mathcal{O}_P(1/\sqrt{NT}),
\end{aligned}$$

because the diagonal elements of $\overline{\mathcal{H}}^{-1}$ do not contribute here, while the off-diagonal terms elements contribute as $\frac{1}{NT} \sum_{it} \mathcal{O}_P(1/\sqrt{NT}) = \mathcal{O}_P(1/\sqrt{NT})$, according to the lemma on $\overline{\mathcal{H}}^{-1}$.

$$
\begin{aligned}
&\frac{1}{2\sqrt{NT}} \sum_{g,h=1}^{\dim\phi} [\partial_{\beta_k\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi_g\phi_h}\overline{\mathcal{L}}]\ [\overline{\mathcal{H}}^{-1}]_{gh} \\
&- \frac{1}{2\sqrt{NT}} \sum_{g,h=1}^{\dim\phi} \partial_{\beta_k\phi'}\overline{\mathcal{L}}^{\dagger}[\overline{\mathcal{H}}^{-1}]^{\dagger}\partial_{\phi\phi_g\phi_h}\overline{\mathcal{L}}^{\dagger}\ [\overline{\mathcal{H}}^{-1}]^{\dagger}_{gh} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{[\partial_{\beta_k\phi'}\overline{\mathcal{L}}\,\overline{\mathcal{H}}^{-1}]_i\ \alpha_i^0\ \partial_{\pi^2}\overline{\ell}_{it}\ [\overline{\mathcal{H}}^{-1}]_{tt} \\
&+ [\partial_{\beta_k\phi'}\overline{\mathcal{L}}\,\overline{\mathcal{H}}^{-1}]_t\ \gamma_t^0\ \partial_{\pi^2}\overline{\ell}_{it}\ [\overline{\mathcal{H}}^{-1}]_{ii}\} + \mathcal{O}_P(1/\sqrt{NT}) \\
&= -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}[\partial_{\beta_k\phi'}\overline{\mathcal{L}}\,\overline{\mathcal{H}}^{-1}]_i\alpha_i^0\ \partial_{\pi^2}\overline{\ell}_{it}}{\sum_{i=1}^{N}(\alpha_i^0)^2\partial_{\pi^2}\overline{\ell}_{it}} \\
&- \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}[\partial_{\beta_k\phi'}\overline{\mathcal{L}}\,\overline{\mathcal{H}}^{-1}]_t\gamma_t^0\ \partial_{\pi^2}\overline{\ell}_{it}}{\sum_{t=1}^{T}(\gamma_t^0)^2\partial_{\pi^2}\overline{\ell}_{it}} + \mathcal{O}_P(1/\sqrt{NT}),
\end{aligned}
$$

where the off-diagonal elements of the second $[\overline{\mathcal{H}}^{-1}]$ only contribute terms of order $1/\sqrt{NT}$. Thus, we find that for the correctly specified case the two additional bias contributions (that occur for the model ℓ_{it} but are not present in model $\ell_{it}^{\dagger}$) from the terms $-[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,\widetilde{\mathcal{H}}\,\overline{\mathcal{H}}^{-1}\,\mathcal{S}$ and $\frac{1}{2}\sum_{g=1}^{\dim\phi}[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}$ exactly cancel. We have thus shown that the asymptotic distribution of $\widehat{\beta}$ and $\widehat{\beta}^{\dagger}$ are identical.

$\square$

Appendix C

Supplemental Materials for Chapter 3

C.1 Incorporating network structure: CoVaR, network spillover effects, and systemic risk

Traditional risk measures, such as Value of Risk (VaR), focus on the loss of an individual institution only. CoVaR proposed by (Adrian and Brunnermeier, 2011) measures the VaR of the whole financial system or a particular financial institution by conditioning on another institution being in distress. Thus, it relates systemic risk to tail spillover effects from individual institutions to the whole system. (Adrian and Brunnermeier, 2011) define firm b's CoVaR at level τ conditional on a particular outcome from firm a, as the value of $CoVaR_\tau^{b|a}$ that solves

$$Pr(X_b \leq CoVaR_\tau^{b|a} | \mathbb{C}(X_a)) \quad = \quad \tau,$$

A particular case is $\mathbb{C}(X_a) = \{X_a = VaR_\tau^a\}$ for a low quantile index τ, which is interpreted as with probability τ institution b is in trouble given that institution a is in trouble. They also define institution a's contribution to b as

$$\Delta CoVaR_\tau^{b|a} = CoVaR_\tau^{b|X_a = VaR_\tau^a} - CoVaR_\tau^{b|X_a = Median_a}.$$

They mainly use quantile regression to estimate the $CoVaR$ measure. More precisely, the predicted value from the quantile regression of X_b on X_a gives the value at risk of institution b conditional on institution a since VaR_τ^b given X_a is just the conditional quantile, i.e. conditional VaR

$$VaR_\tau^b|X_a = \alpha^b(\tau) + \beta^b(\tau)X_a,$$

Replacing variable X_a by its unconditional quantile, i.e. VaR_τ^a, yields

$$CoVaR_\tau^{b|X_a} = \alpha^b(\tau) + \beta^b(\tau)VaR_\tau^a \;\text{ and }\; \Delta CoVaR_\tau^{b|a} = \beta^b(\tau)(VaR_\tau^a - VaR_{50\%}^a)$$

We incorporate network spillover effects into risk measuring. We show that with QGM, individual institution's contribution to systemic risk can incorporate tail risk interconnections between institutions in the whole financial system (in the network, each node represents a financial institution now). The identified risk connections between all financial institutions constitute a systemic risk network. Note, institution a's overall systemic risk contribution, $\Delta CoVaR^{sys|a}$ measures the contribution of institution a to overall systemic risk $\sum_a \Delta CoVaR^{sys|a}$.

We define

$$Pr(X_b \leq CoVaR_\tau^{b|a,V\setminus\{a,b\}}|\mathbb{C}(X_a, X_{V\setminus\{a,b\}})) = \tau$$

then

$$CoVaR_\tau^{b|X_a=VaR_\tau^a, X_{V\setminus\{a,b\}}=VaR_q^{V\setminus\{a,b\}}} = \beta_0^b(\tau) + \beta_a^b(\tau)VaR_\tau^a + \beta_{V\setminus\{a,b\}}^b(\tau)VaR_\tau^{V\setminus\{a,b\}}$$

$$\Delta CoVaR_\tau^{b|a,V\setminus\{a,b\}} = \beta_a^b(\tau)(VaR_\tau^a - VaR_{50\%}^a)$$

where $\beta^b(\tau) = \{\beta_0^b(\tau), \beta_{V\setminus\{b\}}^b(\tau)\}$ is estimated via ℓ_1-penalized quantile regression.

We stack $\Delta CoVaR_\tau^{b|a,V\setminus\{a,b\}}$ as the (a,b)-th element of an $d \times d$ matrix $E^\beta(\tau)$ representing a weighted directed network of institutions. Here d is the number of total financial institutions considered. Following (Andersen et al., 2013), the systemic risk contribution of firm a, $\Delta CoVaR^{sys|a}$, is the network to-degree of institution a which is defined as $\delta_a^{to} = \Delta CoVaR^{sys|a} = \sum_k \Delta CoVaR^{k|a,V\setminus\{a,k\}}$. To-degrees measure contributions of individual institutions to the overall risk of systemic network events.

Similarly, from-degree of node a is defined as $\delta_a^{from} = \Delta CoVaR^{a|sys} = \sum_b \Delta CoVaR_\tau^{a|b,V\setminus\{a,b\}}$. From-degrees measure exposure of individual institutions to systemic shocks from the network. The total degree δ, i.e. $\sum_a \Delta CoVaR^{sys|a}$, aggregates institution-specific systemic risk across institutions hence provides a measure of total systemic risk in the whole financial system.

Finally , we define the net contribution as $net\text{-}\Delta CoVaR^a = \delta_a^{to} - \delta_a^{from}$. For more about network theory, see (Kolaczyk, 2009).

C.2 ℓ_1-Penalized Quantile Regression for Near Extreme Quantiles Indices

In this section we revisit the rate of convergence of ℓ_1-penalized quantile regression estimators. We are concerned to the case that the compact set $\mathcal{T} \subset (0,1)$ grows so that it asymptotically covers $(0,1)$. Namely, the measure of the estimated set of indices goes to on, $|\mathcal{T}| \to 1$. Our results build upon the prior work (Belloni and Chernozhukov, 2011) which focused on the case that $\mathcal{T}$ is bounded away from the extreme quantiles. In what follows we let $\underline{\tau} := \min_{\tau\in\mathcal{T}} \tau(1-\tau)$ to characterise how fast $\mathcal{T}$ approaches the extremes. We use the notation and assumptions (D.1-D.4) in (Belloni and Chernozhukov, 2011)

In what follows we let K_n such that $\max_{i\leq n} \|x_i\|_\infty \leq K_n$ with probability $1-\varepsilon \to 1$.

Lemma C.2.1 (Rate of Convergence of ℓ_1-QR). *Suppose that Assumptions D1-D4 hold and*

$K_n^2 \log(n \vee p) = o(n \min_{\tau\in\mathcal{T}} \tau(1-\tau))$. *Then, we have with probability* $1-\alpha-4\gamma-\varepsilon_n$

$$\sup_{\tau\in\mathcal{T}} \|J_\tau^{1/2}(\hat\beta(\tau)-\beta(\tau))\| \lesssim_P \frac{1}{\underline{f}^{1/2}\kappa_{c_0}}\sqrt{\frac{s\log(n\vee p)}{n}}$$

where $J_\tau^{1/2} = E[f_{y|x}(x'\beta(\tau) \mid x)xx']$

Lemma C.2.1 complements the rates of convergence derived in Theorem 2 of (Belloni and Chernozhukov, 2011). The latter does not assume the additional requirement $K_n^2\log(n\vee p) = o(n\underline{\tau})$ but it was established for a fixed set $\mathcal{T}$, that is, $\underline{\tau}$ bounded away from zero. Indeed Theorem 2 of (Belloni and Chernozhukov, 2011) yields the rate

$$\frac{1}{\underline{\tau}^{1/2}\underline{f}^{1/2}\kappa}\sqrt{\frac{s\log(n\vee p)}{n}}$$

which is potentially slower than the rate established in Lemma C.2.1 as $\underline{\tau}$ can go to zero with n (provided the additional requirement $K_n^2\log(n\vee p) = o(n\underline{\tau})$ holds).

The proof of Lemma C.2.1 follows the proof of Theorem 2 of (Belloni and Chernozhukov, 2011) and the improvement is achieved by controlling the penalty choice under the additional requirement $K_n^2\log(n\vee p) = o(n\underline{\tau})$. This is done in the following technical lemma.

Lemma C.2.2 (Penalty Parameter Bound). *Let* $\underline{\tau} = \min_{\tau\in\mathcal{T}}\tau(1-\tau)$ *and* $K_n = \max_{i\le n}\|x_i\|_\infty$. *Under* $K_n^2\log(d/\underline{\tau}) = o(n\underline{\tau})$, *for* n *large enough we have that for some constant* $\bar C$

$$\Lambda(1-\alpha \mid X) \le \sqrt{1+\frac{\log(16/\alpha)}{\log(d/\underline{\tau})}}\,\bar C\sqrt{n\log(d/\underline{\tau})}.$$

Proof. Conditionally on $x_1,\ldots,x_n$, letting $\hat\sigma_j^2 = \mathbb{E}_n[x_{ij}^2]$, we have that

$$\Lambda = \sup_{\tau\in\mathcal{T}}\left|\frac{nE_n[x_{ij}(\tau-1\{u_i\le\tau\})]}{\hat\sigma_j\sqrt{\tau(1-\tau)}}\right|.$$

Step 1. (Entropy Calculation) Let $\mathcal{F} = \{x_{ij}(\tau-1\{u_i\le\tau\}) : \tau\in\mathcal{T}\}$, $h_\tau = \sqrt{\tau(1-\tau)}$, and

$\mathcal{G} = \{f_\tau/h_\tau : \tau \in \mathcal{T}\}$. We have that

$$\begin{array}{rl} d(f_\tau/h_\tau, f_{\bar\tau}/h_{\bar\tau}) & \leq d(f_\tau, f_{\bar\tau})/h_\tau + d(f_{\bar\tau}/h_\tau, f_{\bar\tau}/h_{\bar\tau}) \\ & \leq d(f_\tau, f_{\bar\tau})/h_\tau + d(0, f_{\bar\tau}/h_{\bar\tau})|h_\tau - h_{\bar\tau}|/h_\tau \end{array}$$

Therefore, since $\|F\|_Q \leq \|G\|_Q$ by $h_\tau \leq 1$, and $d(0, f_{\bar\tau}/h_{\bar\tau}) \leq 1/h_{\bar\tau}$ we have

$$N(\varepsilon\|G\|_Q, \mathcal{G}, Q) \leq N(\varepsilon\|F\|_Q/\{2\min_{\tau\in\mathcal{T}} h_\tau\}, \mathcal{F}, Q)N(\varepsilon/\{2\min_{\tau\in\mathcal{T}} h_\tau^2\}, \mathcal{T}, |\cdot|).$$

Thus we have for some constants K and v that

$$N(\varepsilon\|G\|_Q, \mathcal{G}, Q) \leq d(K/\{\epsilon\min_{\tau\in\mathcal{T}} h_\tau^2\})^v.$$

Step 2.(Symmetrization) Since we have $E[g^2] = 1$ for all $g \in \mathcal{G}$, by Lemma 2.3.7 in (van der Vaart and Wellner, 1996) we have

$$P(\Lambda \geq t\sqrt{n}) \quad \leq 4P(\max_{j\leq d}\sup_{\tau\in\mathcal{T}} |\mathbb{G}_n^o(g)| \geq t/4)$$

where $\mathbb{G}_n^o : \mathcal{G} \to \mathbb{R}$ is the symmetrized process generated by Rademacher variables. Conditional on $(x_1, u_1), \ldots, (x_n, u_n)$, we have that $\{\mathbb{G}_n^o(g) : g \in \mathcal{G}\}$ is sub-Gaussian with respect to the $L_2(\mathbb{P}_n)$-norm by the Hoeffding inequality. Thus, by Lemma 16 in (Belloni and Chernozhukov, 2011), for $\delta_n^2 = \sup_{g\in\mathcal{G}} \mathbb{E}_n[g_i^2]$ and $\bar\delta_n = \delta_n/\|G\|_{\mathbb{P}_n}$, we have

$$P(\sup_{g\in\mathcal{G}} |\mathbb{G}_n^o(g)| > C\bar{K}\delta_n\sqrt{\log(dK/\underline{\tau})} \mid \{X_i, U_i\}_{i=1}^n) \leq \int_0^{\bar\delta_n/2} \epsilon^{-1}\{d(K/\{\epsilon\min_{\tau\in\mathcal{T}} h_\tau^2\})^v\}^{-C^2+1}d\varepsilon$$

for some universal constant $\bar{K}$.

In order to control δ_n, note that $\delta_n^2 = \sup_{g\in\mathcal{G}} \frac{1}{\sqrt{n}}\mathbb{G}_n(g^2) + E[g^2]$. In turn, since $\sup_{g\in\mathcal{G}} \mathbb{E}_n[g^4] \leq \delta_n^2 \max_{i\leq n} G_i^2$, we have

$$P(\sup_{g\in\mathcal{G}} |\mathbb{G}_n^o(g^2)| > C\bar{K}\delta_n \max_{i\leq n} G_i\sqrt{\log(dK/\underline{\tau})} \mid \{X_i, U_i\}_{i=1}^n) \leq \int_0^{\bar\delta_n/2} \epsilon^{-1}\{d(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\varepsilon$$

Thus with probability $1-\int_0^{1/2}\epsilon^{-1}\{d(K/\epsilon\underline{\tau})^v\}^{-C^2+1}d\varepsilon$, since $E[g^2]=1$ and $\max_{i\leq n} G_i \leq K_n/\sqrt{\underline{\tau}}$, we have

$$\delta_n \leq 1 + \frac{C' K_n \sqrt{\log(dK/\underline{\tau})}}{\sqrt{n}\sqrt{\underline{\tau}}}.$$

Therefore, under $K_n\sqrt{\log(dK/\underline{\tau})} = o(\sqrt{n}\sqrt{\underline{\tau}})$, conditionally on $\{X_i\}_{i=1}^n$ and n sufficiently large, with probability $1-2\int_0^{1/2}\epsilon^{-1}\{d(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\varepsilon$ we have that

$$\sup_{g\in\mathcal{G}}|\mathbb{G}_n^o(g)| \leq 2C\bar{K}\sqrt{\log(dK/\underline{\tau})}$$

The stated bound follows since for $C>2$

$$2\int_0^{1/2}\epsilon^{-1}\{d(K/\{\epsilon\underline{\tau}\})^v\}^{-C^2+1}d\varepsilon \leq \{d/\underline{\tau}\}^{-C^2+1}2\int_0^{1/2}\epsilon^{-2+C^2}d\varepsilon \leq \{d/\underline{\tau}\}^{-C^2+1}.$$

□

Bibliography

Adrian, T. and Brunnermeier, M. (2011). CoVaR. Technical report, National Bureau of Economic Research.

Ahn, S., Lee, Y., and Schmidt, P. (2001). Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

Ahn, S., Lee, Y., and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14.

Andersen, T., Bollerslev, T., Christoffersen, P., and Diebold, F. (2013). Financial risk measurement for financial risk management. *Handbook of the Economics of Finance*, 2:1127–1220.

Andrews, D. (2005). Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Ang, A., Chen, J., and Xing, Y. (2006). Downside risk. *Review of Financial Studies*, 19(4):1191–1239.

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2009a). Likelihood approach to small T dynamic panel models with interactive effects. *Unpublished manuscript.*

Bai, J. (2009b). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression for high dimensional sparse models. *Annals of Statistics*, 39(1):82–130.

Belloni, A., Chernozhukov, V., and Fernández-Val, I. (2011). Conditional quantile processes based on series or many regressors. *arXiv:1105.6154*.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013a). Inference methods for high-dimensional sparse econometric models. *Advances in Economics and Econometrics, 10th World Congress of Econometric Society*, III:245–295.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Belloni, A., Chernozhukov, V., and Kato, K. (2013b). Robust inference in high-dimensional approximately sparse quantile regression models. *arXiv:1312.7186*.

Bernard, A., Jensen, B., Redding, S., and Schott, P. (2007). Firms in international trade. Technical report, National Bureau of Economic Research.

Bester, A. and Hansen, C. (2009). Identification of marginal effects in a nonparametric correlated random effects model. *Journal of Business & Economic Statistics*, 27(2):235–250.

Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80(4):1337–1385.

Cai, T., Liu, W., and Luo, X. (2011). A constrained $\ell 1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Carro, J. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140(2):503–528.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):225–238.

Charbonneau, K. (2012). Multiple fixed effects in nonlinear panel data models. *Unpublished manuscript.*

Chen, M., Fernández-Val, I., and Weidner, M. (2013). Interactive fixed effects in nonlinear panel data models with large N, T. *Unpublished manuscript.*

Chernozhukov, V., Chetverikov, D., and Kato, K. (2012). Gaussian approximation of suprema of empirical processes. *arXiv:1212.6885*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Claessens, S. and Forbes, K. (2001). *International Financial Contagion*. Springer.

Cox, D. and Wermuth, N. (1996). *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press.

Cunha, F. and Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Dempster, A. (1972). Covariance selection. *Biometrics*, pages 157–175.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Dhaene, G. and Jochmans, K. (2010). Split-panel jackknife estimation of fixed-effect models. *Unpublished manuscript*.

Dhaene, G. and Jochmans, K. (2014). Split-panel jackknife estimation of fixed-effect models. *Unpublished manuscript*.

Diebold, F. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

Drton, M. and Perlman, M. (2004). Model selection for gaussian concentration graphs. *Biometrika*, 91(3):591–602.

Drton, M. and Perlman, M. (2007). Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449.

Drton, M. and Perlman, M. (2008). A sinful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200.

Edwards, D. (2000). *Introduction to graphical modelling*. Springer.

Elrod, T. and Keane, M. (1995). A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, 32(1):1–16.

Engle, R. and Susmel, R. (1993). Common volatility in international equity markets. *Journal of Business & Economic Statistics*, 11(2):167–176.

Fan, J., Lv, J., and Qi, L. (2011). Sparse high dimensional models in economics. *Annual review of economics*, 3(1):291–317.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.

Fernández-Val, I. and Vella, F. (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics*, 163(2):144–162.

Fernández-Val, I. and Weidner, M. (2013). Individual and time effects in nonlinear panel data models with large N, T. *Unpublished manuscript*.

Foygel, R. and Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. In *Advances in Neural Information Processing Systems*.

Freyberger, J. (2012). Nonparametric panel data models with interactive fixed effects. *Unpublished manuscript*.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987). Generalised residuals. *Journal of Econometrics*, 34(1):5–32.

Greene, W. (2003). *Econometric analysis*. Pearson Education India.

Greene, W. (2004). The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal*, 7(1):98–119.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. *Econometrica*, 70(4):1639–1657.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–195.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487.

Holtz-Eakin, D., Newey, W., and Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–1395.

Kato, K., Galvao, A., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, New York.

Kolaczyk, E. (2009). *Statistical analysis of network data*. Springer.

Laird, N. (1985). Missing information principle. *Encyclopedia of Statistical Sciences*.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2):391–413.

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies*, 69(3):647–666.

Lauritzen, S. (1996). *Graphical models*. Oxford University Press.

Leeb, H. and Pötscher, B. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(02):338–376.

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326.

Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10:2295–2328.

Liu, H. and Wang, L. (2012). Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *arXiv:1209.2437*.

Liu, W. and Luo, X. (2012). High-dimensional sparse precision matrix estimation via sparse column inverse operator. *arXiv:1203.3896*.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

McLeish, D. (1974). Dependent central limit theorems and invariance principles. *The Annals of Probability*, 2(4):620–628.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

Meng, X.-L. and Rubin, D. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278.

Moon, H. R., Shum, M., and Weidner, M. (2014). Estimation of random coefficients logit demand models with interactive fixed effects. *Unpublished manuscript.*

Moon, H. R. and Weidner, M. (2010a). Dynamic linear panel regression models with interactive fixed effects. *Unpublished manuscript.*

Moon, H. R. and Weidner, M. (2010b). Linear regression for panel with unknown number of factors as interactive fixed effects. *Unpublished manuscript.*

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Orchard, T. and Woodbury, M. (1972). A missing information principle: theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 697–715. University of California Press Berkeley, CA.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of financial economics*, 63(1):3–50.

Pastorello, S., Patilea, V., and Renault, E. (2003). Iterative and recursive estimation in structural nonadaptive models. *Journal of Business & Economic Statistics*, 21(4):449–509.

Patton, A. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Pesaran, H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Phillips, P. and Moon, H. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Romano, J. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Sun, T. and Zhang, C.-H. (2012). Sparse matrix inversion with scaled lasso. *arXiv:1202.2723.*

van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *Annals of Statistics*, 36(2):614–645.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics.

Woutersen, T. (2001). Robustness against incidental parameters and mixing distributions. Technical report, Department of Economics, University of Western Ontario.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional non-paranormal graphical models. *The Annals of Statistics*, 40(5):2541–2571.

Xue, L., Zou, H., and Cai, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse ising models. *The Annals of Statistics*, 40(3):1403–1429.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research*, 99:2261–2286.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Curriculum Vitae

Mingli Chen

Education

Ph.D. in Economics, Boston University, Boston MA, May 2015 (expected)
Dissertation Title: *Research Related to High Dimensional Econometrics*
Dissertation Committee: Pierre Perron, Iván Fernández-Val, Marc Rysman, Victor Chernozhukov (MIT)

B.A., Information and Computing Science, Shanghai University, Shanghai, China, July 2009

Fields of Interest

Econometrics, Time Series Econometrics, Financial Econometrics, Empirical IO

Teaching Experience

Teaching Assistant, Statistics for Business and Economics, Department of Economics, Boston University, Spring 2011
Teaching Assistant, Introduction to Econometrics, Department of Economics, Boston University, Spring 2011

Academic Work Experience

Visiting Graduate Fellow, Federal Reserve Bank of Boston, August 2013-present
Research Assistant, Professor Pierre Perron, Department of Economics, Boston University, Fall 2010, Spring 2014, Fall 2014, Spring 2015
Research Assistant, Professor Iván Fernández-Val, Department of Economics, Boston University, Summer 2011-Spring 2012, Spring 2013- Fall 2013
Research Assistant, Professor Marc Rysman, Department of Economics, Boston University, Fall 2011, Spring 2012, Summer 2012, Fall 2012
Research Assistant, Professor Joshua Lustig, Department of Economics, Boston University, Summer 2011

Fellowships, Honors and Awards

Travel Grant, Department of Economics, Boston University, June 2013

Dean's Fellowship, Boston University, Fall 2009-Spring 2011
Teaching Fellowship, Department of Economics, Boston University, Fall 2009, Spring 2010
Research Grant, Shanghai Government, China, Fall 2008 - Spring 2009
Academic Scholarship, Shanghai University, 2005-2009
Outstanding Graduate, Shanghai University, 2009
2nd prize, China Undergraduate Mathematical Contest in Modeling, October 2008

Working Papers

"Estimation of Nonlinear Panel Models with Multiple Unobserved Effects", October 2014
"Nonlinear Panel Models with Interactive Effects" (with Iván Fernández-Val and Martin Weidner), May 2013
"Quantile Graphical Models: Prediction and Conditional Independence with Applications to Financial Risk Management" (with Alexandre Belloni and Victor Chernozhukov), July 2014

R-Packages

"Rearrangement" (with Wesley Graybill, Victor Chernozhukov, Iván Fernández-Val and Alfred Galichon), published on the Comprehensive R Archive Network
"Counterfactual" (with Victor Chernozhukov, Iván Fernández-Val and Blaise Melly)

Work in Progress

"Consumer Payment Choice: Heterogeneity in Transaction-Level Data" (with Marc Rysman)
"Dating Structural Breaks using Lasso-Type Estimators" (with Pierre Perron)
"Bayesian Inference for Nonlinear Panel Data Models with Interactive Fixed Effects"

Referee Experience

Journal of Econometrics, *Journal of Business & Economic Statistics*

Conferences and Presentations

International Symposium on the Analysis of Panel Data, Xiamen University, June 2013
Boston University/Boston College Green Line Econometrics Meeting, Boston, MA, November 2013

Languages: English (fluent), Chinese (native)

Computer Skills: R, Matlab, STATA, GAUSS, EViews, SPSS, SAS, Mathematica, Maple, AMPL, Lindo, Lingo, C, C++, AutoCAD, LaTeX, Microsoft Office

Citizenship/Visa: P.R. China/F1