High-dimensional panel data with time heterogeneity: estimation and inference*

Yinchu Zhu

University of California, San Diego

yinchu.zhu@rady.ucsd.edu

January 15, 2017

Click here for the latest version

Job Market Paper

*This paper was previously circulated under the title "Uniform estimation and inference of time-heterogeneous dynamic panels with interactive fixed effects". I would like to thank Allan Timmermann, my advisor, for his support, guidance and patience throughout the development of this paper. I also thank Otilia Boldea, Jelena Bradic, Ivana Komunjer, Ulrich Mueller, Aureo de Paula, Hashem Pesaran, Andres Santos, Kaspar Wuthrich and the participants of the European Winter Meeting (2015), the North American Summer Meeting (2016) of the Econometric Society, UCSD econometrics seminars, California Econometrics Conference (2016) and Annual meeting of the Midwest Econometrics Group (2016) for helpful discussions and comments. I am particularly grateful to Jungbin Hwang and Roger Moon for discussions that inspired me to extend my earlier results to dynamic panel data models. All the errors are my own.

Abstract

We consider high-dimensional panel data models (large cross sections and long time horizons) with interactive fixed effects and allow the covariate/slope coefficients to vary over time without any restrictions. The parameter of interest is the vector that contains all the covariate effects across time. This vector has dimensionality tending to infinity, potentially much faster than the cross-sectional sample size. We develop methods for the estimation and inference of this high-dimensional vector, i.e., the entire trajectory of time variation in covariate effects. We show that both the consistency of our estimator and the asymptotic accuracy of the proposed inference procedure hold uniformly in time. Our methodology can be applied to several important issues in econometrics, such as constructing confidence bands for the entire path of covariate coefficients across time, testing the time-invariance of slope coefficients and estimation and inference of patterns of time variations, including structural breaks and regime switching. An important feature of our method is that it provides inference procedures for the time variation in pre-specified components of slope coefficients while allowing for arbitrary time variation in other components. Computationally, our procedures do not require any numerical optimization and are very simple to implement. Monte Carlo simulations demonstrate favorable properties of our methods in finite samples. We illustrate our methods through empirical applications in finance and economics.

Key words: inference in high-dimensional models, large dynamic panel data models, heterogeneous slope coefficients, interactive fixed effects, parameter instability

1 Introduction

How heterogeneity is modeled plays a key role in many empirical studies in economics and finance. Although linear panel data models have been extensively employed to account for cross-sectional and temporal heterogeneity, such heterogeneity is usually restricted to the error terms by various specifications of fixed effects and random effects. In contrast, slope coefficients are typically assumed to be homogeneous in cross sections and over time.

Allowing for heterogeneity in both error terms and slope coefficients can be very important in applied research. For example, consider the literature on predictability of stock returns. Most work applies linear regressions of stock returns against predictors such as the lagged dividend yield.[1] Suppose that we have a panel dataset containing observations of returns and dividend yields for a large number of stocks over a long time horizon. Heterogeneity in the error terms may arise as different stocks have different sensitivities to common shocks (e.g., macroeconomic activity and market-wide shocks) and firm-specific components (e.g., firm fixed effects). Meanwhile, it is also reasonable to expect time heterogeneity in the relationship between expected stock returns and the dividend yield since the instability of this relationship is well documented in the finance literature.[2] Researchers are often interested in whether return predictability from the dividend yield is stable across time, how such predictability evolves and whether the state of the macro economy affects such predictability. For example, an important question is how macroeconomic and/or financial turmoil, such as the Great Recession, affects the predictability of stock returns. Does the Great Recession only amount to shocks in the error terms or does it fundamentally change the relationship between stock returns and dividend yields?

The main contribution of this paper is to address these types of questions using a linear panel data model with general time-heterogeneous covariate effects. Suppose that for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, we observe dependent variables $y_{i,t} \in \mathbb{R}$ and covariates/regressors $x_{i,t} \in \mathbb{R}^k$ from the following model

$$y_{i,t} = x_{i,t}'\beta_t + \alpha_{i,t} + u_{i,t}, \tag{1.1}$$

where $\beta_t \in \mathbb{R}^k$ is the vector containing unobserved covariate effects at time t, $\alpha_{i,t}$ is unobserved fixed effects of individual i at time t and $u_{i,t}$ is an idiosyncratic error with $\mathbb{E}u_{i,t} = 0$ and $\mathbb{E}x_{i,t}u_{i,t} = 0$. We consider the interactive fixed effects for $\alpha_{i,t}$ and impose a factor structure on the regressors (similar to Pesaran (2006)); see Section 2 for details. We allow for dynamic structures since components of $x_{i,s}$ can be correlated with $u_{i,t}$ for $s \neq t$.

The most important feature of our model (1.1) is that the covariate effects $\{\beta_t\}_{t=1}^T$ are allowed to vary across time without any restrictions. The sequence $\{\beta_t\}_{t=1}^T$ can be viewed as either a

[1]See Campbell and Shiller (1988a), Keim and Stambaugh (1986), Campbell and Thompson (2008), Goyal and Welch (2003, 2008) and Rapach et al. (2010).

[2]See Paye and Timmermann (2006), Ang and Bekaert (2007), Pettenuzzo and Timmermann (2011) and Lettau and Van Nieuwerburgh (2008).

deterministic sequence or a stochastic process with arbitrary correlation with observed variables.[3] Throughout the paper, we assume that k is fixed and both n and T tend to infinity. Let β denote the high-dimensional vector representing the trajectory of β_t over time

$$\beta := (\beta_1', \ldots, \beta_T')' \in \mathbb{R}^{kT}.$$

We treat the high-dimensional vector β as the model parameter of interest and develop procedures for its estimation and inference. In particular, we propose a new methodology that can be used to construct confidence sets for β and simultaneously test multiple (or many) linear hypotheses of β.

Our general model eliminates the risk of misspecification in time-varying pattern of $\{\beta_t\}_{t=1}^T$. Time variation in model parameters has been recognized in many areas of applied research, such as macroeconomic forecasting (Stock and Watson, 1996, 2007; Giacomini and Rossi, 2006, 2009, 2010; Rossi, 2013). Most empirical work that addresses the issue of time-varying parameters uses a random coefficient approach and assumes that parameters evolve according to a particular stochastic process.[4] However, imposing parametric or non-parametric structures introduces the risk of misspecification, which might deliver misleading or even spurious results.[5] Our proposed methodology does not require any restriction on the time variation in the slope coefficients.

Moreover, the flexibility in our setup provides a natural framework of estimating parametric specifications for the time variation in β_t and testing the validity of these specifications. For example, one of the most popular models accounting for time-varying parameters is the structural break model in which β_t is assumed to have a piecewise constant pattern across t. If the true underlying trajectory of $\{\beta_t\}_{t=1}^T$ indeed follows such a pattern, then our method can be used to estimate the number and locations of structural breaks. Testing the validity of the structural break model is also straight-forward. Under the null hypothesis of correct specification, estimates of the break points are consistent and thus the stability of β_t between two breaks can be rewritten as multiple linear hypotheses of β, which can be tested using the proposed methodology. Similarly, our methodology can be used to estimate regime-switching models and test their specifications. If β_t is regime-dependent, then the proposed procedures can consistently recover the time series of regime membership and thus the hypothesis of homogeneity within each regime can be again formulated as multiple linear hypotheses on β. In addition, since our regime estimator does not assume any structure on the time variation of regimes, the time series of estimated regime membership can be used to test the validity of candidate specifications, e.g., whether the regimes evolve as a Markov chain (Hamilton, 1989) or a Bernoulli process.

[3] See Remark 3.3 for more discussion.

[4] Popular specifications either impose parametric models, such as piecewise constant parameters (structural breaks), Markov chains, Bernoulli distributions, random walks, autoregressive models, or assume non-parametric time variation with smooth paths.

[5] Even the flexible non-parametric specification that assumes only smoothness in time variation can fail to capture brief temporary changes, which may be due to momentary shocks in the economy and weather.

A distinct feature of the proposed method is that our results can be used for sub-vector (partial) inference allowing for flexible structures in the nuisance parameter. In practice, applied researchers are often interested in only a subset of the slope coefficients. For example, suppose $\beta_t = (\beta_{1,t}, \beta_{2,t})'$, where $\beta_{1,t}$ is of empirical interest and $\beta_{2,t}$ corresponds to a control variable. Our method can be used to test specifications of $\{\beta_{1,t}\}_{t=1}^T$ without imposing any restrictions on the time variation of the nuisance parameter $\{\beta_{2,t}\}_{t=1}^T$. Many existing specification tests, such as the popular test by Bai and Perron (1998) for structural breaks, can only handle the null hypothesis that specify the time variation in the entire $k \times 1$ vector β_t. In our example, a typical existing test for lack of structural breaks has the null hypothesis that both $\{\beta_{1,t}\}_{t=1}^T$ and $\{\beta_{2,t}\}_{t=1}^T$ are constant across time. Hence, our methodology provides specification tests that are robust to misspecifications of nuisance parameters.

Our results offer an intuitive setup to study and explain the time variation in the slope coefficients. For example, suppose that the researcher is interested in testing whether the slope coefficients vary with the business cycle. This question can be formulated in terms of the average value of β_t in economic recessions and expansions and thus can be phrased as inference of linear hypothesis of β. Alternatively, a regression-based approach can be applied. Since our methodology deliver consistent estimators for the entire path $\{\beta_t\}_{t=1}^T$, we can fit the estimated β_t in a time series regression against other explanatory variables.

Our paper contributes to econometric theories in several ways. First, we propose a new strategy for identification and estimation, overcoming difficulties due to flexibility in fixed effects and the general specification of β. Since the fixed effects in (1.1) is potentially correlated with the regressors, running the ordinary least square (OLS) estimation for each t does not guarantee consistent estimation of β_t. Even under strict exogeneity of the regressors, potential cross-sectional dependence in error terms could render traditional methods invalid; see e.g., Phillips and Sul (2003), Andrews (2005) and Pesaran and Tosetti (2011). This problem is illustrated in Appendix A.

Second, our methodology can be used for inference on the high-dimensional vector β. To the best of our knowledge, our work is the first in the literature on panel data models to address the inference problem of the entire path of unrestricted time variation in coefficients. Although model (2.1) with time or individual-specific covariate effects has been studied by authors such as Pesaran (2006)[6], inference results are only available for low-dimensional components of β (e.g., β_t for a fixed t). In contrast, our results deal with inference on the entire vector β by capitalizing on recent advances in high-dimensional statistics and probability. In existing work, inference on individual β_t's is based on the classical central limit theorem (CLT). Since T tends to infinity, $\beta \in \mathbb{R}^{kT}$ is a high-dimensional object and thus the classical CLT is not suitable for our purposes. One might attempt to construct a confidence set for the whole trajectory of β_t over time from confidence sets for each β_t. However, constructing confidence bands for the whole trajectory of β_t amounts to

[6]In fact, Pesaran (2006) considers panel data models with individual-specific covariate coefficients, but his method can be applied to models with time-heterogeneity by swapping the time and individual indices.

approximating the distribution of the maximal estimation error of β_t over all $t = 1, \ldots, T$. This is not straight-forward when T tends to infinity.[7] Building upon the recent results by Chernozhukov et al. (2013, 2014), we develop a multiplier bootstrap procedure, which is shown to be asymptotically exact in terms of size control.[8]

Finally, the estimation procedures proposed in this paper are computationally simple. For high-dimensional models, computational burden is often a key concern as naively extending algorithms designed for low-dimensional problems might not be computationally feasible. As will be introduced in Section 2, the fixed effect assumes a factor structure $\alpha_{i,t} = L'_{\alpha,i} F_{\alpha,t}$. Then the least squared estimator minimizes $\sum_{i=1}^{n} \sum_{t=1}^{T} (y_{i,t} - x'_{i,t}\beta_t - L'_{\alpha,i} F_{\alpha,t})^2$ over $\{\beta_t\}_{t=1}^{T}$, $\{L_{\alpha,i}\}_{i=1}^{n}$ and $\{F_{\alpha,t}\}_{t=1}^{T}$. This estimator has been applied intensively for low-dimensional problems (i.e., time-homogeneous β_t); see Bai (2009) and Moon and Weidner (2015). Since there is no closed-end solution to this optimization problem and the objective function is not jointly convex in $\{\beta_t\}_{t=1}^{T}$, $\{L_{\alpha,i}\}_{i=1}^{n}$ and $\{F_{\alpha,t}\}_{t=1}^{T}$, most numerical algorithms are not guaranteed to return the global maximizer. The usual remedy of trying many starting points is virtually infeasible since β is high-dimensional. We develop alternative identification and estimation strategies and derive procedures that only involve matrix multiplications and singular value decompositions, thereby considerably reducing the computational burden. Moreover, unlike most nonparametric methods, the methodology proposed in this paper does not require choosing any tuning parameters, except for the number of factors, which can also be consistently estimated in a manner free of tuning parameters. Our theoretical results still hold when the true number of factors are replaced by consistent estimators.

We demonstrate the advantage of the proposed methodology via three empirical studies in finance and economics. The first study is concerned with the predictability of stock returns using the lagged dividend yield and volatility as predictors. We find that the predictive power of both the dividend yield and volatility exhibits very different patterns of time variation; in particular, return predictability is linked to the macroeconomy but in different manners. We also find seasonality patterns in predictability, which is different from seasonality in the error term, often referred to as calendar effects. The second empirical study uses panel data on firms and focuses on the effects of several variables on firms' capital structure. We find that the patterns of time variation in the slope coefficients can be quite different from what is generated by simply applying time-homogeneous models to subsamples of the data. The third empirical application studies the effect of investment on economic growth. Using a multi-country panel dataset, we find strong evidence of time variation in this effect. Our methodology also finds group patterns in the fixed effects, suggesting that

[7] To illustrate these issues, suppose that $k = 1$ and that for each t, there is an estimator $\hat{\beta}_t$ such that $\sqrt{n}(\hat{\beta}_t - \beta_t) \to^d N(0, 1)$. Constructing a confidence band for all β_t's amounts to finding $c > 0$ such that $\mathbb{P}(\max_{1 \le t \le T} \sqrt{n}|\hat{\beta}_t - \beta_t| > c) \approx \eta$ for some pre-specified $\eta \in (0, 1)$. For large T, the difficulties of conducting inference based on existing methods arise as the validity of approximating $\sqrt{n}(\hat{\beta}_t - \beta_t)$ with Gaussian distributions might not be uniform in t and it is not straight-forward to account for the interdependence across t.

[8] One may address the issue of inter-temporal dependence from the perspective of multiple testing problems and use the Bonferroni method to control the family-wise error rate. Unfortunately, this approach usually results in a great loss of power and leads to conservative tests, especially in our case where the number of tests (in the multiple testing problem) can be much larger than the sample size.

developing and developed countries have different trends that are likely to be driven by the same factor but with different factor loadings. In all these studies, we find that time heterogeneity in the slope coefficients exists and displays complicated patterns that are difficult to capture by parametric models. Since no restrictions are imposed on the time heterogeneity in β_t, our findings are not subject to the misspecification risk in this regard.

Related literature

Our work builds upon the literature on large dynamic panel data models with fixed effects. The asymptotic framework in this literature allows both n and T to tend to infinity. The most common specification for the fixed effects is time-invariant individual-specific fixed effects (sometimes plus a time-specific component), e.g., see Phillips and Moon (1999), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003) and Hahn and Moon (2006). Bonhomme and Manresa (2015) propose a structure under which individuals are classified into several groups and the fixed effects are allowed to have unconstrained time variations but are homogeneous among individuals in the same group. Factor structures in fixed effects have also been considered, e.g., Andrews (2005), Bai (2009), Ahn et al. (2013), Su et al. (2015) and Moon and Weidner (2015).

Although most empirical work that uses panel data models assumes homogeneous covariate effects, numerous authors, such as Phillips and Sul (2003), Pesaran and Yamagata (2008) and Su and Chen (2013), have developed tests for assessing the reasonableness of this popular specification. In addition, the literature has seen work that directly considers models with heterogeneous slope coefficients. Just as the heterogeneity in the error terms can be treated as fixed or random effects, counterparts of these two approaches are also found in the study of heterogeneity in covariate effects. Under one approach, slope coefficients for different i and/or t are viewed as fixed parameters to be estimated, see e.g., Pesaran (2006), Zaffaroni (2009) and Lin and Ng (2012); under the other approach, the slope coefficients are assumed to be random variables generated from parametric models and the focus is the estimation and inference of these parametric models, see e.g., Swamy (1970), Rosenberg (1972) and Hsiao et al. (1993). Beyond the usual parametric/linear specification, several authors study nonparametric estimation and inference for heterogeneous covariate effects; see Qian and Wang (2012), Chen et al. (2013) and Boneva et al. (2015). An excellent survey for heterogeneous parameters in panel data models can be found in Chapter 6 of Hsiao (2014). Existing results on estimation and inference mainly focus on the average (across i or t) covariate effects and pointwise covariate effects (for given i or t).

An interesting paper by Freyberger (2012) considers heterogeneous nonparametric panel data models with interactive fixed effects. He treats the factor loadings as random variables and exploits their distributional properties to achieve nonparametric identification; the estimation strategy relies on the assumption that distribution of observables are identical in the cross section. In contrast, our result focuses on the linear models but can deal with non-random factor loadings and heterogeneous distributions across units.

Our specification of β falls into the category of high-dimensional models . Our theoretical results are based on the recent advances by Chernozhukov et al. (2013, 2014) on high-dimensional central limit theorems and bootstrap. To handle the high-dimensional nuisance parameter (fixed effects), we borrow tools from random matrix theory, see Vershynin (2010), and the literature on large factor models, see e.g., Forni et al. (2000), Stock and Watson (2002), Bai and Ng (2002) and Bai (2003).

Organization of the paper and notations

The rest of the paper is organized as follows. The formal setup of our model is introduced in Section 2. We provide the details of the main results in Section 3. In Section 4, we discuss several related econometric problems. Finite-sample properties of our procedures are demonstrated via Monte Carlo simulations in Section 5. We apply our methods to several empirical studies in Section 6. The appendix contains the proofs of theoretical results.

For any vector $x = (x_1, \ldots, x_{n_1})' \in \mathbb{R}^{n_1}$, $\|x\| = (\sum_{i=1}^{n} x_i^2)^{1/2} = \sqrt{x'x}$, $\|x\|_1 = \sum_{i=1}^{n} |x_i|$, $\|x\|_\infty = \max_{1 \le i \le n_1} |x_i|$ and $\|x\|_0$ denotes the number of nonzero entries in x. For any matrix $A \in \mathbb{R}^{n_1 \times n_2}$, $\|A\|$ denotes the spectral norm of A and we say that $A = U_A S_A V_A'$ is a singular value decomposition (SVD) if $U_A \in \mathbb{R}^{n_1 \times n_1}$ and $V_A \in \mathbb{R}^{n_2 \times n_2}$ are both orthogonal matrices and $S_A \in \mathbb{R}^{n_1 \times n_2}$ is a (rectangular) diagonal matrix with singular values of A on the diagonal in the non-increasing order. We also introduce the low rank approximation operator: for a non-negative integer r, define $\mathcal{T}_r(A) := U_A \bar{S}_r V_A'$, where $A = U_A S_A V_A'$ is an SVD and $\bar{S}_r$ is equal to S_A with all the diagonal entries of S_A set to zero except the first r diagonal entries. $s_j(A)$ denotes the jth largest singular value of A, counting multiplicity. For two positive sequences a_n and b_n, we use $a_n \asymp b_n$ to denote the condition that there exist constant $c_1, c_2 > 0$ such that $c_1 a_n \le b_n \le c_2 a_n$. We use $\sigma(\cdot)$ to denote the σ-algebra generated by random variables.

2 Model Setup and Assumptions

Suppose that for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, we observe dependent variables $y_{i,t} \in \mathbb{R}$ and covariates/regressors $x_{i,t} \in \mathbb{R}^k$ from the following model

$$y_{i,t} = x_{i,t}'\beta_t + \alpha_{i,t} + u_{i,t} \qquad \text{with} \qquad \alpha_{i,t} = F_{\alpha,t}' L_{\alpha,i}, \tag{2.1}$$

where $\beta_t \in \mathbb{R}^k$ is the vector containing unobserved covariate effects at time t, $\alpha_{i,t}$ is unobserved fixed effects of individual i at time t with $F_{\alpha,t} \in \mathbb{R}^{r_\alpha}$ and $L_{\alpha,i} \in \mathbb{R}^{r_\alpha}$ being the unobserved factor and its loading, and $u_{i,t}$ is an idiosyncratic error with $\mathbb{E}u_{i,t} = 0$ and $\mathbb{E}x_{i,t}u_{i,t} = 0$. We assume that k, r_α and r_Q are fixed and $T = T_n \to \infty$ as $n \to \infty$.

To achieve identification in this general model, we introduce assumptions on the regressors. Similar to Pesaran (2006), we assume a factor structure

$$x_{i,t} = Q_{i,t} + v_{i,t} \qquad \text{with} \qquad Q_{i,t} = F_{Q,t}' L_{Q,i}, \tag{2.2}$$

where $F_{Q,t} \in \mathbb{R}^{r_Q \times k}$ and $L_{Q,i} \in \mathbb{R}^{r_Q}$ are unobserved factors and their loadings, r_Q is fixed and $v_{i,t} \in \mathbb{R}^k$ is the idiosyncratic errors. Arbitrary correlations between $\{F_{Q,t}\}_{t=1}^T$ and $\{F_{\alpha,t}\}_{t=1}^T$ are permitted. The model (2.2) can be justified in many applications. Factor structures have been motivated on both theoretical and empirical grounds and have been widely used to model financial and macroeconomic data[9], to account for unobserved abilities (e.g., Lord et al. (1968), Hansen et al. (2004)) and to study consumer theory (e.g., Gorman (1981) and Lewbel (1991)).

The factor structure in $\alpha_{i,t}$, often referred to as interactive fixed effects, allows for a rich class of unobserved common effects and nests popular fixed effects models as special cases, see Bai (2009). The interactive fixed effects also allow for flexible cross-sectional and inter-temporal dependence among the regression residuals $\alpha_{i,t} + u_{i,t}$, see e.g., Andrews (2005) and Pesaran (2006).

The goal of this paper is to build a confidence set for $\beta \in \mathbb{R}^{kT}$ (a confidence band for β_t that is uniformly valid over t) and test hypotheses of the form

$$H_0: \ J\beta = a, \tag{2.3}$$

where $J \in \mathbb{R}^{m_J \times kT}$ and $a \in \mathbb{R}^{m_J}$ are nonrandom and m_J can be as large as $O(n^l)$ for some constant $0 \le l < \infty$.

We introduce the following definition, which is satisfied by a large class of random variables including polynomials of sub-Gaussian random variables as well as finite mixtures of random variables with thin-tailed distributions.

Definition 2.1. A random variable Z is said to have an exponential-type tail with parameter (b, γ) if $\forall z > 0$, $\mathbb{P}(|Z| > z) \le \exp\left[1 - (z/b)^\gamma\right]$.

We impose the following conditions for model (2.1) and (2.2).

Assumption 1. *Assume that the following hold:*

(i) There exist constants $b_*, \gamma_* > 0$ such that $\forall (i,t) \in \{1, \ldots, n\} \times \{1, \ldots, T\}$, each entry of $F_{\alpha,t}$, $L_{\alpha,i}$, $F_{Q,t}$, $L_{Q,i}$, $u_{i,t}$ and $v_{i,t}$ has an exponential-type tail with parameter (b_*, γ_*).

(ii) There exist constants $c_*, \gamma_{**} > 0$ such that $\alpha_{mixing}(t) \le c_* \exp(-t^{\gamma_{**}})$ $\forall t \ge 1$, where

$$\alpha_{mixing}(t) := \sup\Big\{\Big|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A \bigcap B)\Big| : \ A \in \sigma\left(\{(F_{Q,s}, F_{\alpha,s}, v_s, u_s) : s \le \tau\}\right), \\ B \in \sigma\left(\{(F_{Q,s}, F_{\alpha,s}, v_s, u_s) : s \ge \tau + t\}\right) \text{ and } \tau \in \mathbb{Z}\Big\}.$$

(iii) There exist constants $\kappa_1, \kappa_2 > 0$ and $\xi \in (6/7, 2)$ such that $\kappa_1 n^\xi \le T \le \kappa_2 n^\xi$.

(iv) There exist constants $C_1, C_2 > 0$ such that, with probability approaching one, all the eigenvalues of $n^{-1}L_Q'L_Q$, $T^{-1}F_Q'F_Q$, $n^{-1}L_\alpha'L_\alpha$ and $T^{-1}F_\alpha'F_\alpha$ lie in $[C_1, C_2]$.

[9] See e.g., Ross (1976), Campbell et al. (1997), Fama and French (1992, 2016), Ludvigson and Ng (2007), Forni and Lippi (1997), Stock and Watson (1998, 2002, 2006)

(v) $\{(\underline{v}_i, \underline{u}_i)\}_{i=1}^n$ *is independent across* i*, where* $\underline{v}_i = (v_{i,1}, \ldots, v_{i,T})' \in \mathbb{R}^{T\times k}$ *and* $\underline{u}_i = (u_{i,1}, \ldots, u_{i,T})' \in \mathbb{R}^T$*.*

(vi) $\{u, v\}$ *is independent of* $\{L_Q, F_Q, L_\alpha, F_\alpha\}$ *and* $\forall i, t$*,* $\mathbb{E}v_{i,t}u_{i,t} = 0$*.*

(vii) There exists a constant $C_5 > 0$ *such that* $\min_{1\le t\le T} s_k\left(n^{-1}\sum_{i=1}^n \mathbb{E}v_{i,t}v_{i,t}'\right) > C_5$*.*

Assumption 1(i) and (ii) enable us to apply large deviation theory, which is convenient in deriving bounds for the maximum of a large number of sums of random variables. Assumption 1(i) allows for thicker tails than the Gaussian and exponential distribution, although it rules out fat-tailed distributions such as student t distribution or the stationary distribution of GARCH processes. However, in Monte Carlo simulations, our procedure performs well with these fat-tailed distributions. With more careful arguments, it is possible that we can invoke the moderate deviation theory for self-normalized sums, such as Chen et al. (2016), and replace the exponential-type tails in Assumption 1(i) with bounded moment conditions. Assumption 1(ii) allows weak dependence across t and is satisfied in many situations.[10] Assumption 1(i) and (ii) are also imposed by Bonhomme and Manresa (2015) in their Assumption 2.

Assumption 1(iii) specifies the relative magnitude between n and T. Recent literature on dynamic panel data models considers three cases of sample size: $n/T \to 0$, $T/n \to 0$ and $n \asymp T$; see Hahn and Kuersteiner (2002), Moon and Phillips (2004), Arellano and Hahn (2007) and Bai (2009) among many others. We allow for all these three cases, which correspond to $\xi < 1$, $\xi > 1$ and $\xi = 0$, respectively. Assumption 1(iv) assumes strong factors in α and Q and is a standard condition in the large factor model literature; see Bai and Ng (2002), Bai (2003, 2009) and Moon and Weidner (2015).

Assumption 1(v) and (vi) say that the idiosyncratic terms are independent across i and are independent of the factors and their loadings. Similar conditions are routinely imposed in the literature on large factor models, e.g., see Bai (2003) and Bai and Ng (2006). Notice that Assumption 1(v) and (vi) still allow for arbitrary dependence across i for $L_{\alpha,i}$ and $L_{Q,i}$, as well as serial dependence within u, v, F_α and F_Q. Contemporaneous exogeneity of $v_{i,t}$ in Assumption 1(vi) is required for the identification of β_t. Heteroskedasticity is also allowed in $v_{i,t}$ and $u_{i,t}$ under Assumption 1. Finally, Assumption 1(vii) rules out asymptotically vanishing variances in the idiosyncratic terms of the regressors.

We now demonstrate Assumption 1 with a concrete example.

Example 2.1 (Time-heterogeneous dynamic panel data model)**.** Let $y_{i,t} = L_{\alpha,i}'(\sum_{j=0}^\infty \gamma_{t,j}F_{\alpha,t-j}) + \sum_{j=0}^\infty \gamma_{t,j}u_{i,t-j}$, where $\gamma_{t,j}$ is defined as the following: $\gamma_{t,0} = 1$, $\gamma_{t,j} = \Pi_{l=1}^j \beta_{t-l+1}$ for $j > 0$ and $\gamma_{t,j} = 0$ for $j < 0$. For simplicity, let $u_{i,t}, F_{\alpha,t}, L_{\alpha,i} \sim i.i.d\ N(0,1)$ and assume that $\sup_{t\ge 0}|\beta_t| \le c$

[10] For linear processes and GARCH processes, see Gorodetskii (1978) and Carrasco and Chen (2002). For Markov processes, one can actually show geometric decay of β-mixing coefficients using the so-called V-ergodicity property; see Meyn and Tweedie (2012).

for some constant $c \in (0,1)$. Then one can easily verify that $y_{i,t}$ defined above satisfies

$$y_{i,t} = L'_{\alpha,i} F_{\alpha,t} + \beta_t y_{i,t-1} + u_{i,t}.$$

Thus, in the notations of (2.1) and (2.2), $x_{i,t} = y_{i,t-1}$, $L_{Q,i} = L_{\alpha,i}$, $F_{Q,t} = \sum_{j=0}^{\infty} \gamma_{t-1,j} F_{\alpha,t-1-j}$ and $v_{i,t} = \sum_{j=0}^{\infty} \gamma_{t-1,j} u_{i,t-1-j}$. Assumption 1(i) holds by the Gaussianity and Assumptions 1(iii)-(vii) obviously hold. In Lemma B.16 of Appendix B.4, we show that Assumption 1(ii) also holds.

3 Main Results

In this section, we present the main results for estimation and inference of $\{\beta_t\}_{t=1}^T$. In Section 3.1, we discuss the key idea behind our identification strategy. Sections 3.2 and 3.3 develop the main methodology for estimation and inference and establish theoretical properties of the proposed procedures. Section 3.4 deals with the issue of determining the number of factors.

3.1 Identification strategy

Given the model (2.1) and (2.2), our estimation strategy is based on the following observation:

$$y_{i,t} = x'_{i,t}\beta_t + \alpha_{i,t} + u_{i,t} = v'_{i,t}\beta_t + \left(Q'_{i,t}\beta_t + \alpha_{i,t} + u_{i,t}\right).$$

We shall assume that $v_{i,t}$ is uncorrelated with $Q_{i,t}$, $\alpha_{i,t}$ and $u_{i,t}$. Therefore, at time t, we can view $Q'_{i,t}\beta_t + \alpha_{i,t} + u_{i,t}$ as the error term and simply use the cross-sectional variation to identify β_t:

$$\beta_t = \left(\sum_{i=1}^n \mathbb{E} v_{i,t} v'_{i,t}\right)^{-1} \left(\sum_{i=1}^n \mathbb{E} v_{i,t} y_{i,t}\right).$$

In other words, for each t, we run a cross-sectional regression of $y_{i,t}$ against $v_{i,t}$. Notice that $v_{i,t}$ is unobserved. To make this approach feasible, we exploit the factor structure (2.2) again and employ the technique of principal component analysis (PCA) to identify $v_{i,t}$.

Remark 3.1. Pesaran (2006) proposes the common correlated effect estimator (CCE), which can be adapted to our model. The strategy is the following. If $L_Q := (L_{Q,1}, \ldots, L_{Q,n})' \in \mathbb{R}^{n \times r_Q}$ were observed, then $v_{i,t}$ could be estimated as the residuals from projecting columns $X_t = (x_{1,t}, \ldots, x_{n,t})' \in \mathbb{R}^{n \times k}$ onto L_Q; since we do not observe L_Q, we need to replace it with an observed matrix $\tilde{L}$. Therefore, the plan is (1) to construct $\tilde{L}$ whose columns span a space that approximately contains columns of L_Q and (2) to take as estimates of $\{v_{i,t}\}_{i=1}^n$ the residuals of projecting columns of X_t onto $\tilde{L}$. To illustrate the idea of CCE, consider $\tilde{L} = (\bar{x}_{(1)}, \ldots, \bar{x}_{(n)})' \in \mathbb{R}^{n \times k}$ with $\bar{x}_{(i)} = T^{-1} \sum_{t=1}^T x_{i,t}$. Notice that under the specification (2.2), if we assume that the law of large numbers (LLN) applies across t, then $\bar{x}_{(i)} = A_T L_{Q,i} + T^{-1} \sum_{t=1}^T v_{i,t} \approx A_T L_{Q,i}$, where $A_T = T^{-1} \sum_{t=1}^T F'_{Q,t} \in \mathbb{R}^{k \times r_Q}$. Ignoring the approximation error due to LLN, we have $\tilde{L} = L_Q A'_T$. Then columns of $\tilde{L}$ span a space that contains columns of L_Q if and only if $\text{rank}A_T = r_Q$, which, in Pesaran (2006), is referred to as the rank condition. A necessary condition for the rank condition is $k \geq r_Q$, which may or may not hold in practice. In contrast, our method uses PCA and does not require this rank condition.

We now introduce some notations that will be used in the rest of the paper: $Y = [Y_1, \ldots, Y_T] \in \mathbb{R}^{n\times T}$, $X = [X_1, \ldots, X_T] \in \mathbb{R}^{n\times kT}$, $\alpha = [\alpha_1, \ldots, \alpha_T] \in \mathbb{R}^{n\times T}$, $u = [u_1, \ldots, u_T] \in \mathbb{R}^{n\times T}$, $v = [v_1, \ldots, v_T] \in \mathbb{R}^{n\times kT}$, $Q = [Q_1, \ldots, Q_T] \in \mathbb{R}^{n\times kT}$, $F_\alpha = [F_{\alpha,1}, \ldots, F_{\alpha,T}]' \in \mathbb{R}^{T\times r_\alpha}$, $F_Q = [F_{Q,1}, \ldots, F_{Q,T}]' \in \mathbb{R}^{kT\times r_Q}$ $L_Q = (L_{Q,1}, \ldots, L_{Q,n})' \in \mathbb{R}^{n\times r_Q}$ and $L_\alpha = (L_{\alpha,1}, \ldots, L_{\alpha,n})' \in \mathbb{R}^{n\times r_\alpha}$, where $y_t = (y_{1,t}, \ldots, y_{n,t})' \in \mathbb{R}^n$, $X_t = (x_{1,t}, \ldots, x_{n,t})' \in \mathbb{R}^{n\times k}$, $\alpha_t = (\alpha_{1,t}, \ldots, \alpha_{n,t})' \in \mathbb{R}^n$, $u_t = (u_{1,t}, \ldots, u_{n,t})' \in \mathbb{R}^n$, $v_t = (v_{1,t}, \ldots, v_{n,t})' \in \mathbb{R}^{n\times k}$ and $Q_t = (Q_{1,t}, \ldots, Q_{n,t})' \in \mathbb{R}^{n\times k}$. Notice that $Q = L_Q F_Q'$ and $\alpha = L_\alpha F_\alpha'$.

3.2 Estimation of β

For now, we assume that the values of r_Q and r_α are known and we will provide consistent estimators for r_Q and r_α later in Section 3.4. Since v_t is unknown, we first estimate it and use the estimated v_t to obtain an initial estimator for β_t. We define

$$\hat{\beta}_t = (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t' Y_t, \tag{3.1}$$

where $\hat{Q} = [\hat{Q}_1, \ldots, \hat{Q}_T] = \mathcal{T}_{r_Q}(X)$ and $\hat{v} = [\hat{v}_1, \ldots, \hat{v}_T] = X - \hat{Q}$. The following result establishes the theoretical properties of the above estimator.

Theorem 3.1 (Uniform estimation of β)**.** *Under Assumption 1, we have*

$$\|\hat{\beta} - \beta\|_\infty = O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right] \log^{c_0} n\right),$$

where $c_0 > 0$ is a constant and $\hat{\beta} := (\hat{\beta}_1', \ldots, \hat{\beta}_T')' \in \mathbb{R}^{kT}$ with $\hat{\beta}_t$ defined in (3.1).

This result says that $\hat{\beta}_t$ is a consistent estimator for β_t uniformly over t and the rate of convergence depends on the relative size of n and T. If $\xi \geq 1$ ($n/T = O(1)$), then the convergence rate is the parametric rate up to a logarithm factor, $n^{-1/2}\log^{c_0} n$. The logarithm factor is the price we pay for the high dimensionality of β and is common in the literature on high-dimensional statistics.[11] The exact value of c_0 is not important for our purposes. If $\xi < 1$ (n much larger than T), then the rate of convergence is strictly slower than $n^{-1/2}\log^{c_0} n$.

It turns out that the non-standard rate of convergence of $\hat{\beta}$ is due to the bias in the estimator; we now show that once the bias is removed, the rate of convergence in ℓ_∞-norm is $\sqrt{n^{-1}\log n}$. Notice that by the properties of SVD, $\hat{Q}_t'\hat{v}_t = 0$. Thus, it is not hard to see that

$$\sqrt{n}(\hat{\beta}_t - \beta_t) = (n^{-1}\hat{v}_t'\hat{v}_t)^{-1} n^{-1/2}\hat{v}_t'(\alpha_t + u_t). \tag{3.2}$$

[11] For example, see Bickel et al. (2009), Bühlmann and Van De Geer (2011) and Belloni and Chernozhukov (2011).

Our strategy is to remove the effect of $n^{-1/2}\hat{v}_t'\alpha_t$ by subtracting $(\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'\hat{\alpha}_t$ from $\hat{\beta}_t$, where $\hat{\alpha}_t$ is an estimator for α_t such that $n^{-1/2}\max_{1\leq t\leq T}\|\hat{v}_t'\alpha_t - \hat{v}_t'\hat{\alpha}_t\| = o_P(1)$. As we shall show, this is can be done in an intuitive manner. Since $\hat{\beta}_t$ is a consistent estimator for β_t, $y_t - X_t\hat{\beta}_t = \alpha_t + u_t + X_t(\beta_t - \hat{\beta}_t)$ is a consistent estimator for $\alpha_t + u_t$. Heuristically speaking, we have a consistent estimator for $\alpha + u$ and can simply apply PCA again to obtain an estimator for α.

Algorithm 1. *Implement the following steps:*

1. *Compute* $[\hat{\alpha}_1, \ldots, \hat{\alpha}_T] = \mathcal{T}_{r_\alpha}([y_1 - X_1\hat{\beta}_1, \ldots, y_T - X_T\hat{\beta}_T])$, *where* $\hat{\beta}_t$ *is defined in (3.1).*
2. *Compute* $\tilde{\beta}_t = \hat{\beta}_t - (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'\hat{\alpha}_t$.

The following result establishes the rate of convergence for the estimator in Algorithm 1.

Theorem 3.2. *Under Assumption 1, we have*

$$\|\tilde{\beta} - \beta\|_\infty = O_P\left(\sqrt{n^{-1}\log n}\right),$$

where $c_0 > 0$ *is a constant and* $\tilde{\beta} := (\tilde{\beta}_1', \ldots, \tilde{\beta}_T')' \in \mathbb{R}^{kT}$ *with* $\tilde{\beta}_t$ *defined in Algorithm 1.*

A comparison between Theorems 3.1 and 3.2 demonstrates the advantage of bias correction. When $\xi < 1$ (i.e., $n/T \to \infty$), $\tilde{\beta}$ is a strictly better estimator than $\hat{\beta}$ in terms of the rate of convergence in the ℓ_∞-norm; when $\xi \geq 1$ (i.e., $n = O(T)$), $\tilde{\beta}$ and $\hat{\beta}$ have the same rates of convergence up to logarithm factors.

3.3 Inference on β

Now we turn to the problem of testing high-dimensional linear combinations of β in the form (2.3). The idea is to approximate $\tilde{\beta}_t$ with an average of independent high-dimensional vectors. Let $G_i = (G_{i,1}', ..., G_{i,T}')' \in \mathbb{R}^{kT}$ with $G_{i,t} = \Sigma_t^{-1}v_{i,t}u_{i,t}$ and $\Sigma_t = n^{-1}\sum_{i=1}^n Ev_{i,t}v_{i,t}'$. We show, in the appendix, that

$$\left\|J\tilde{\beta} - J\beta - n^{-1}\sum_{i=1}^n JG_i\right\|_\infty = O_P(n^{-1/2-c})$$

for some constant $c > 0$, where $\tilde{\beta} = (\tilde{\beta}_1', \ldots, \tilde{\beta}_T')'$. The above display suggests the "obvious" strategy of approximating the distribution of $\sqrt{n}\|J\tilde{\beta} - J\beta\|_\infty$ by that of $\|N(0,\Omega)\|_\infty$, where $\Omega = n^{-1}\sum_{i=1}^n E(JG_iG_i'J')$. Since Ω is unknown, we replace it with a plug-in estimator. We will show that this intuitive approach can be justified even if the dimension of Ω is much larger than n and T.

To simplify the presentation, we introduce the following notation. For a random vector $Z \sim N(0,\Sigma)$, we define $\Phi(z,\Sigma) = \mathbb{P}(\|Z\|_\infty \leq z)$ and denote by $\Phi^{-1}(\cdot,\Sigma)$ the inverse of $\Phi(z,\Sigma)$ as a function of z. For a given Σ, the function $\Phi^{-1}(\cdot,\Sigma)$ can be easily computed by simulation. Our inference procedure for testing H_0 in (2.3) can be formally summarized as follows.

Algorithm 2. *For a test for H_0 (2.3) with nominal size $\eta \in (0,1)$, implement the following steps:*

1. *Compute $\hat{u}_t = y_t - X_t\hat{\beta}_t - \hat{\alpha}_t$, where $\hat{\beta}_t$ and $\hat{\alpha}_t$ are defined in (3.1) and Algorithm 1, respectively.*

2. *Comupute $\hat{G}_i = (\hat{G}'_{i,1}, \ldots, \hat{G}'_{i,T})' \in \mathbb{R}^{kT}$, where $\hat{G}_{i,t} = \hat{\bar{v}}_{i,t}\hat{u}_{i,t}$, $\hat{\bar{v}}_{i,t} = \hat{\Sigma}_t^{-1}\hat{v}_{i,t}$ and $\hat{\Sigma}_t = n^{-1}\hat{v}'_t\hat{v}_t$ with $\hat{v}_t$ defined in (3.1).*

3. *Generate $\{\zeta_i\}_{i=1}^n$ i.i.d $N(0,1)$ independent of the sample and compute $n^{-1/2}\sum_{i=1}^n J\hat{G}_i\zeta_i$.*

4. *Repeat the previous step as many times as computationally convenient to compute $\Phi^{-1}(1-\eta, \hat{\Omega})$, where $\hat{\Omega} = n^{-1}\sum_{i=1}^n J\hat{G}_i\hat{G}'_iJ'$.*

5. *Reject H_0 in (2.3) if and only if $\|J\tilde{\beta} - a\|_\infty > \Phi^{-1}(1-\eta, \hat{\Omega})$, where $\tilde{\beta} = (\tilde{\beta}'_1, \ldots, \tilde{\beta}'_T)'$ and $\tilde{\beta}_t$ is defined in Algorithm 1.*

Although the parameter of interest β is high-dimensional, we establish the validity of such procedures for our problem using recent tools developed by Chernozhukov et al. (2013). Even in light of their results, we still need to deal with the technical challenges arising due to the facts that $\tilde{\beta}$ is not exactly the mean of independent vectors and that the large-sample behavior of $\tilde{\beta}$ depends on the residuals, such as $u_{i,t}$, which are not observed and need to be replaced with estimates for the bootstrap procedure to be feasible. To justify Algorithm 2, we need some restrictions on J.

Assumption 2. *Assume that the following conditions hold for J in (2.3):*

(i) $m_J = O(n^l)$ *for some constant* $0 \leq l < \infty$.

(ii) *There exists a constant $A_1 > 0$ such that $\max_{1\leq j\leq m_J}\|J_j\|_1 \leq A_1$, where J_j is the transpose of the jth row of J.*

(iii) *There exists a constant $b_1 > 0$ such that $J'_j\left(n^{-1}\sum_{i=1}^n \mathbb{E}G_iG'_i\right)J_j \geq b_1$ $\forall j \in \{1, \ldots, m_J\}$, where $G_i = (G'_{i,1}, \ldots, G'_{i,T})' \in \mathbb{R}^{kT}$, $G_{i,t} = \bar{v}_{i,t}u_{i,t}$, $\bar{v}_{i,t} = \Sigma_t^{-1}v_{t,i}$ and $\Sigma_t = n^{-1}\sum_{i=1}^n \mathbb{E}v_{i,t}v'_{i,t}$.*

Assumption 2(i) allows us to test m_J linear transformations of β, where m_J can be fixed or grow polynomially fast in n. Notice that this allows for $m_J \gg \max\{n, T\}$. Building a confidence set for all the entries of β implies that $m_J = kT = O(n)$; inference on the individual β_t or on the average of β_t over t corresponds to a fixed m_J. Assumption 2(ii) can be viewed as a "near-sparsity" assumption on the rows of J, while it still allows $\|J_j\|_0 = kT$ $\forall 1 \leq j \leq m_J$. This is needed to control the bias of $J\tilde{\beta}$: although the maximal bias of all $\tilde{\beta}_t$'s can be shown to be small, the bias of each row of $J\tilde{\beta}$ is a linear combination of all the biases of $\tilde{\beta}_t$'s. Assumption 2(ii) allows us to control the bias of $J\tilde{\beta}$ via Holder's inequality. Assumption 2(iii) rules out "degenerate" linear combinations of G_i. This is needed for the theory of high-dimensional bootstrap.

The following theorem is our main theoretical result and establishes the validity of Algorithm 2.

Theorem 3.3 (High-dimensional inference)**.** *Under Assumptions 1 and 2, we have*

$$\limsup_{n\to\infty} \sup_{\eta\in(0,1)} \left| \mathbb{P}\left(\sqrt{n}\|J\tilde{\beta} - J\beta\|_\infty > \Phi^{-1}(1-\eta, \hat{\Omega})\right) - \eta \right| = 0,$$

where $\tilde{\beta}$ *and* $\Phi^{-1}(1-\eta,\hat{\Omega})$ *are defined in Algorithm 2.*

As our main result for inference, Theorem 3.3 says that Algorithm 2 can be used to test hypotheses that involve m_J linear combinations of a kT-dimensional vector, where both m_J and T can grow polynomial fast with n. We can easily invert the test to obtain confidence sets for $J\beta$.

Corollary 3.1. *Let Assumptions 1 and 2 hold. For any fixed* $\eta \in (0,1)$*, let* $\Phi^{-1}(1-\eta,\hat{\Omega})$ *and* $\tilde{\beta}$ *be defined as in Algorithm 2. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(J\beta \in \mathcal{C}_{1-\eta}(J)\right) = 1 - \eta,$$

where $\mathcal{C}_{1-\eta}(J) = \left\{ J\tilde{\beta} + v \,\middle|\, v \in \mathbb{R}^{m_J} \text{ and } \|v\|_\infty \le \Phi^{-1}(1-\eta,\hat{\Omega})\}/\sqrt{n} \right\}.$

In the following result, we show that the width of the above confidence set is $O_P\left(\sqrt{n^{-1}\log n}\right)$.

Theorem 3.4. *Suppose that Assumptions 1 and 2 hold. Then there exists a constant* $M > 0$ *such that,* $\forall \eta \in (0,1)$*,* $\mathbb{P}\left(\Phi^{-1}(1-\eta,\hat{\Omega}) \le M\sqrt{\log n}\right) \to 1$.

When the entries of $n^{-1/2}\sum_{i=1}^n JG_i$ have no correlation among each other, it can be shown that there exists a constant $M_0 > 0$ such that $\mathbb{P}(\Phi^{-1}(1-\eta,\hat{\Omega}) \ge M_0\sqrt{\log n}) \to 1$. This highlights the nonstandard nature of the problem as $\sqrt{n}\|J\tilde{\beta} - J\beta\|_\infty$ might not have a well-defined limiting distribution. Under certain conditions that guarantee weak dependence among entries of JG_i, one can employ tools from the extreme value theory and obtain a well-defined asymptotic distribution for a properly scaled version of $\|J\tilde{\beta} - J\beta\|_\infty$, such as $n\|J\tilde{\beta} - J\beta\|_\infty^2 + A_1 \log n + A_2 \log\log n$ with constants $A_1, A_2 \in \mathbb{R}$; see Cai and Jiang (2011). Although this alternative approach can provide a procedure with analytical critical values, it might require additional assumptions as well as very different theoretical techniques; we shall leave this possibility to future research.

Remark 3.2. Notice that the confidence set $\mathcal{C}_{1-\eta}(J)$ defined in Corollary 3.1 is a rectangle in $\mathbb{R}^{m_J}$, similar in spirit to a Kolmogorov-Smirnov-type test. One might wonder whether it is possible to build Cramer-von-Mises-type tests by considering $\|J\tilde{\beta} - J\beta\|_2$. Unfortunately, this is technically challenging since the tools from probability theories still appear inadequate in handling the ℓ_2-norm of the sum of independent high-dimensional vectors. To the best of our knowledge, existing tools can only handle the problems in which the dimensionality is much smaller than the sample size, e.g., $T = o(n^{1/4})$; see Peng and Schick (2012) and Pouzo (2015).

Remark 3.3. Our results are easy to understand when we treat $\{\beta_t\}_{t=1}^T$ as a deterministic sequence. However, it is worth pointing out that all the theoretical results so far still hold even if $\{\beta_t\}_{t=1}^T$ is stochastic process that is allowed to have arbitrary correlation with the observed variables Y and X. To see how this flexibility is possible, notice that deriving (3.2) the estimation error of the original estimator $\hat{\beta}$ is merely algebraic computations and does not require any knowledge of randomness in the data. Moreover, subsequent analysis used to derive Theorems 3.1, 3.2, 3.3 and 3.4 only involves properties of the factors, factor loadings and the idiosyncratic terms. Therefore, the estimation error of $\tilde{\beta}$ still decays at the rate $O_P(\sqrt{n^{-1}\log n})$ in ℓ_∞-norm and $\mathcal{C}_{1-\eta}(J)$ still contains the (random) vector $J\beta$ with probability approaching $1-\eta$. When β is random, the object of interest is typically parameters governing the randomness of β. In Sections 4.5 and 4.7, we illustrate how our results can be used for this purpose.

Sometimes, an applied researcher might be interested in the fixed effects. For example, when the fixed effects are assumed to be group-specific, see Bonhomme and Manresa (2015), consistent estimators for the fixed effects can be used to determine the group membership via k-means clustering (Forgy, 1965; Lloyd, 1982). The following result says that the fixed effects can be consistently estimated uniformly over i and t.

Theorem 3.5 (Uniform estimation of fixed effects)**.** *Under Assumption 1, for $\hat{\alpha}_{i,t}$ defined in Algorithm 1, we have that for some constant $c_0 > 0$,*

$$\max_{1\le i\le n,\ 1\le t\le T} |\hat{\alpha}_{i,t} - \alpha_{i,t}| = O_P\left(\left[n^{\xi/2-1} + n^{2-5\xi/2}\right]\log^{c_0} n\right).$$

Similar to Theorem 3.1, Theorem 3.5 says that the rate of convergence for the fixed effect depends on the relative size of n and T. If n and T have the same order of magnitude, then the convergence rate is $n^{-1/2}\log^{c_0} n$; if $\xi \neq 1$, then the rate would be strictly slower.

3.4 Determining the number of factors r_α and r_Q

So far, all the results are derived with the knowledge of the true values of r_Q and r_α for PCA. In practice, these values are often unknown. Now we derive two consistent estimators for these values. Although existing methods, such as Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013), can be used for estimating r_Q, these methods cannot be directly applied for the estimation of r_α due to the estimation errors in $\hat{\beta}_t$. We invoke results from random matrix theory and construct two simple estimators that are consistent under Assumption 1.

Theorem 3.6 (Information criterion)**.** *Let Assumption 1 hold. Define $\hat{r}_Q := \max\{r \mid s_r(X) \ge \mu_n\}$ and $\hat{r}_\alpha := \max\left\{r \mid s_r\left([y_1 - X_1\hat{\beta}_1, \dots, y_T - X_T\hat{\beta}_T]\right) \ge \tilde{\mu}_n\right\}$, where $\mu_n, \tilde{\mu}_n \to \infty$. Then*

(1) If $\mu_n/(\sqrt{n}\log^p n) \to \infty$ for any constant $p > 0$ and $\mu_n/\sqrt{nT} \to 0$, then $\mathbb{P}(\hat{r}_Q = r_Q) \to 1$.

(2) If $\tilde{\mu}_n/\left([\sqrt{T} + n/\sqrt{T}]\log^p n\right) \to \infty$ for any constant $p > 0$ and $\tilde{\mu}_n/\sqrt{nT} \to 0$, then $\mathbb{P}(\hat{r}_\alpha = r_\alpha) \to 1$.

The above estimator for r_Q and r_α is based on information criteria. Similar estimators are proposed by Bai and Ng (2002). One needs to choose a sequence of tuning parameters that satisfy certain rate conditions; however, it might not always be clear how to choose these tuning parameters in finite samples. For this reason, we also provide the following alternative estimators based on the ratio of singular values. These estimators are similar to the ones studied in Ahn and Horenstein (2013) and the only input is an upper bound on r_α and r_Q. In many situations, economic theories can shed some light on these upper bounds. For $r_{\max} \geq 1$, we define

$$\begin{aligned} \hat{r}_Q^{SV} &:= \underset{1 \leq r \leq r_{\max}}{\arg\max} \frac{s_r(X)}{s_{r+1}(X)} \\ \hat{r}_\alpha^{SV} &:= \underset{1 \leq r \leq r_{\max}}{\arg\max} \frac{s_r\left([y_1 - X_1\hat{\beta}_1, \ldots, y_T - X_T\hat{\beta}_T]\right)}{s_{r+1}\left([y_1 - X_1\hat{\beta}_1, \ldots, y_T - X_T\hat{\beta}_T]\right)} \end{aligned}$$

Theorem 3.7 (Singular value ratio estimator)**.** *Let Assumption 1 hold. Suppose that $1 \leq r_Q \leq r_{\max}$ and $1 \leq r_Q \leq r_{\max}$. Then $\mathbb{P}(\hat{r}_Q^{SV} = r_Q) \to 1$ and $\mathbb{P}(\hat{r}_\alpha^{SV} = r_\alpha) \to 1$.*

Remark 3.4. In practice, researchers might need to take additional care in applying the above results. For datasets that contain variables with very different scales, standardization is recommended, similar to the empirical applications in Stock and Watson (2002) and Boivin and Ng (2006).

4 Some Important Inference Problems

In this section, we discuss how several problems often encountered in applied research can be addressed using the methodology proposed in Section 3. It turns out that solving these problems reduces to finding the appropriate matrix J in Algorithm 2. Since empirical work typically focuses on single entries of β_t corresponding to variables of interest, we shall mainly discuss this case. Suppose that we are interested in inference on $\{\beta_{j_0,t}\}_{t=1}^T$, the trajectory of the j_0-th entry of $\beta_t \in \mathbb{R}^k$ across time. For $k \neq 1$, the inference problems only concern part of the parameter $\{\beta_t\}_{t=1}^T$ and shall be referred to as partial inference problems.

4.1 Uniform (over t) inference on $\beta_{j_0,t}$

In empirical research, the goal is often to find out whether some slope coefficient is zero (or some other pre-specified value of interest). When the slope coefficients are allowed to vary over time, the question often becomes whether the slope coefficient $\beta_{(j_0)} = (\beta_{j_0,1}, \ldots, \beta_{j_0,T})' \in \mathbb{R}^T$ is zero in all the time periods.

Notice that this is very different from the problem of testing simple hypotheses on β. Simple hypotheses completely specify the value for all the entries in β; as a result, one can plug-in the hypothesized value of β and test certain moment conditions, such as the orthogonality between $y_{i,t} - x_{i,t}'\beta_t$ and $x_{i,t}$. However, we are dealing with the more difficult problem of testing composite hypotheses on β. For example, consider the problem of testing $\beta_{(j_0)} = 0$. Since $\{\beta_{j,t}\}_{t=1}^T$ with $j \neq j_0$ are still allowed to take any values, the null hypothesis here does not determine the vector β and thus the aforementioned approach for testing simple hypotheses does not apply.

We now demonstrate how our method can be used to solve this inference problem. Let $J = I_T \otimes \tau'_{j_0,k}$ and $\tau_{j_0,k}$ denote the j_0-th column of I_k. Then we have $\beta_{(j_0)} = J\beta$. Notice that Assumption 2(i)-(ii) hold as $m_J = T$ and $\max_{1 \leq j \leq m_J} \|J_j\|_1 = 1$. Under the above notations, $\beta_{j_0,t} = 0 \; \forall 1 \leq t \leq T$ if and only if $J\beta = 0$. Hence, we only need to implement Algorithm 2 with $a = 0$. The problem of building confidence bands for $\{\beta_{j_0,t}\}_{t=1}^T$ reduces to constructing a rectangular confidence set for $\beta_{(j_0)}$ and can be easily solved using Corollary 3.1.

4.2 Inference on temporal difference in $\beta_{j_0,t}$

One of the simplest ways of studying time variation in parameters is to compare $\beta_{j_0,t}$ in different time periods. To formalize the idea, let $A, B \subset \{1, \dots, T\}$ be disjoint sets $A \bigcap B = \emptyset$. We construct confidence intervals for the difference in average parameter values between these two groups of time periods, i.e.,

$$d(A, B) = \frac{\sum_{t=1}^T \beta_{j_0,t} \mathbf{1}\{t \in A\}}{\sum_{t=1}^T \mathbf{1}\{t \in A\}} - \frac{\sum_{t=1}^T \beta_{j_0,t} \mathbf{1}\{t \in B\}}{\sum_{t=1}^T \mathbf{1}\{t \in B\}}. \tag{4.1}$$

As convention, we define $d(A, \emptyset) = \left[\sum_{t=1}^T \beta_t \mathbf{1}\{t \in A\}\right] / \left[\sum_{t=1}^T \mathbf{1}\{t \in A\}\right]$, which is the average parameter value for time periods in the set A. For example, A and B can denote the sets of time periods of economic recessions and expansions, respectively, and $d(A, B)$ is a measure of how the parameters differ across different stages of the business cycle.

We now phrase the problem as inference on a linear combination of β. Let M denote the $1 \times T$ row vector whose s-th entry is equal to $\mathbf{1}\{s \in A\}/|A| - \mathbf{1}\{s \in B\}/|B|$, where $|A|$ and $|B|$ denote the cardinality of the set A and B, respectively. Then it is not hard to see that $d(A, B) = J\beta$, where $J = M \otimes \tau'_{j_0,k}$. Therefore, a confidence interval can be used by implementing Algorithm 2 and computing $\mathcal{C}_{1-\eta}(J)$ defined in Corollary 3.1.

4.3 Estimation and inference of partial parameter instability

The estimate and the confidence set for $\beta_{(j_0)} \in \mathbb{R}^T$ give some indication on whether the slope coefficient is time-varying. We shall refer to changes in parameter values from one period to the next as parameter instability. Define the set of time periods of parameter instability as $\mathcal{B} = \{t \mid 2 \leq t \leq T \text{ and } \beta_{j_0,t} \neq \beta_{j_0,t-1}\}$. Our method can be used for the estimation and inference on $\mathcal{B}$.

This problem can be easily formulated into our framework. Notice that

$$\begin{pmatrix} \beta_{j_0,2}-\beta_{j_0,1} \\ \beta_{j_0,3}-\beta_{j_0,2} \\ \vdots \\ \beta_{j_0,T}-\beta_{j_0,T-1} \end{pmatrix} = J\beta \qquad \text{with} \qquad \underset{(T-1)\times kT}{J} = \underbrace{\begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}}_{(T-1)\times T} \otimes \tau'_{j_0,k}. \tag{4.2}$$

Clearly, the hypothesis of lack of parameter instability can be stated as $J\beta = 0$. Since Assumption 2(i)-(ii) are satisfied (due to $m_J = T-1$ and $\max_{1\le j\le m_J}\|J_j\|_1 = 2$), Theorem 3.3 says that the hypothesis of absence of parameter instability can be tested by applying Algorithm 2 with $a = 0$.

Remark 4.1. As explained in Section 1, a major advantage of our approach is that no assumptions are placed on the time variation in parameters not under testing, i.e., $\{\beta_{j,t}\}_{t=1}^T$ for $j \neq j_0$. Hence, the common approach of imposing the null hypothesis of $\beta_{j_0,1} = \ldots = \beta_{j_0,T}$, such as Su and Chen (2013), does not apply to the partial inference problem here.

Algorithm 2 also provides a natural estimate for the set $\mathcal{B}$. For $\eta \in (0,1)$, consider

$$\widehat{\mathcal{B}}(1-\eta) = \left\{ t \,\middle|\, 2 \le t \le T \text{ and } |\tilde\beta_{j_0,t} - \tilde\beta_{j_0,t-1}| > \Phi^{-1}(1-\eta,\hat\Omega)/\sqrt{n} \right\},$$

where $\tilde\beta_{j_0,t}$ denotes the j_0-th entry of $\tilde\beta_t$ (defined in Algorithm 2). By Theorem 3.3, it follows that

$$\liminf_{n\to\infty} \mathbb{P}\left(\widehat{\mathcal{B}}(1-\eta) \subseteq \mathcal{B}\right) \ge 1-\eta.$$

This means that, $\widehat{\mathcal{B}}(1-\eta)$, as an estimate for the set of instability periods, has asymptotic control on the false discovery rates: the probability of $\widehat{\mathcal{B}}(1-\eta)$ containing points that are not in $\mathcal{B}$ is asymptotically at most η.

Under additional assumptions, $\widehat{\mathcal{B}}$ can also serve as an estimator for $\mathcal{B}$. By Theorem 3.4, $\Phi^{-1}(1-\eta,\hat\Omega) = O_P(\sqrt{\log n})$. Therefore, if we assume that the structural breaks are not too small[12] (i.e., $\min_{t\in\mathcal{B}}|\beta_{j_0,t} - \beta_{j_0,t-1}|\sqrt{n/\log n} \to \infty$), then $\mathbb{P}(\mathcal{B} \subseteq \widehat{\mathcal{B}}(1-\eta)) \to 1$ and thus $\mathbb{P}(\mathcal{B} = \widehat{\mathcal{B}}(1-\eta))$ is asymptotically at least $1-\eta$.

Remark 4.2. To achieve consistent estimation of $\mathcal{B}$, we propose to replace $\widehat{\mathcal{B}}(1-\eta)$ with $\widehat{\mathcal{B}} = \{2 \le t \le T :\ |\tilde\beta_{j_0,t} - \tilde\beta_{j_0,t-1}| > z_n/\sqrt{n}\}$, where z_n is a sequence such that $\forall \eta \in (0,1)$, $\Phi^{-1}(1-\eta,\hat\Omega) \le z_n$ for large n and $z_n \asymp \sqrt{\log n}$. The above analysis implies that $\liminf \mathbb{P}(\mathcal{B} = \widehat{\mathcal{B}}) \ge 1-\eta$ for any $\eta \in (0,1)$ and therefore $\lim \mathbb{P}(\mathcal{B} = \widehat{\mathcal{B}}) = 1$. Now we propose a simple choice of z_n. By Lemma 2 in Chapter 7 of Feller (1968) and the union bound, we can easily show that $\forall x > 0$, $1 - \Phi(x,\hat\Omega) \le 2k(T-1)x^{-1}\|\hat\Omega\|_\infty^{1/2}\phi(x^{-1}\|\hat\Omega\|_\infty^{1/2})$, where $\phi(\cdot)$ is the p.d.f of $N(0,1)$. Hence, it is straight-forward

[12] In a sense, estimation of the set $\mathcal{B}$ is similar to the problem of model selection and thus requires similar regularity conditions, such as the so-called beta-min condition in high-dimensional models; see Bühlmann and Van De Geer (2011).

to show that $\forall \eta \in (0,1)$, we have that for large enough n, $\Phi^{-1}(1-\eta, \hat{\Omega}) \leq \sqrt{2\|\hat{\Omega}\|_\infty \log[k(T-1)]}$ almost surely. Therefore, a natural choice is $z_n = \sqrt{2\|\hat{\Omega}\|_\infty \log[k(T-1)]}$.

Remark 4.3. We note that power properties of tests for time-invariance in our panel setup might not follow existing results that deal with models for single time series. For example, consider the problem of testing $\beta_1 = \ldots = \beta_T$ versus $\beta_1 \neq 0$ and $\beta_2 = \ldots = \beta_T = 0$. When the sample consists of one time series, it is quite hard to detect the deviation that only occurs in one time period, regardless of how large T is. However, in our panel data setting, β_t is identified by the cross-sectional variations across n units and thus can be expected to estimated accurately for large n.

4.4 Partial inference on structural breaks

Sporadic changes in parameter values, often referred to as structural breaks, can be viewed as piecewise constant patterns of $\{\beta_{j_0,t}\}_{t=1}^T$. Unlike the setup considered in Section 4.3, structural breaks are viewed as infrequent jumps in parameter values, which remain stable for extended periods of time, say at least $2q$ periods. Inference on structural breaks for the full vector β_t has been widely studied, e.g., Andrews (1993), Bai and Perron (1998, 2003) and Qu and Perron (2007) among others. However, the partial inference problem of testing for structural breaks in $\{\beta_{j_0,t}\}_{t=1}^T$ without imposing any restrictions on $\{\beta_{j,t}\}_{t=1}^T$ for $j \neq j_0$ is rarely discussed.

Suppose that there are m structural breaks, which occur in periods $B_1 < \ldots < B_m$:

$$\begin{aligned} \beta_{j_0,1} = \ldots = \beta_{j_0,B_1} & \\ \beta_{j_0,B_1+1} = \ldots = \beta_{j_0,B_2} & \quad \text{and} \quad \beta_{j_0,B_1} \neq \beta_{j_0,B_1+1} \\ & \vdots \\ \beta_{j_0,B_m+1} = \ldots = \beta_{j_0,T} & \quad \text{and} \quad \beta_{j_0,B_m} \neq \beta_{j_0,B_m+1}. \end{aligned} \tag{4.3}$$

We follow the convention in the literature by setting $B_0 = 1$ and $B_{m+1} = T$. We consider the problem of testing the hypothesis that there are no structural breaks. Although we can simply use the test discussed in Section 4.3, we might obtain a more powerful test by taking into account the block structure of structural breaks. We consider a block average scheme under which, for any pair of two adjacent blocks of time periods, the average parameter values in these two blocks are compared.

Consider two adjacent blocks of time periods of length q and compute the difference in the block average of parameter values (DBA), i.e.,

$$\text{DBA}(t; \beta, q) := \frac{1}{q} \sum_{s=t-q+1}^{t} \beta_{j_0,s} - \frac{1}{q} \sum_{s=t+1}^{t+q} \beta_{j_0,s} = \left(B'_{(t,q)} \otimes \tau'_{j_0,k} \right) \beta,$$

where the s-th entry of the vector $B_{(t,q)} \in \mathbb{R}^T$ is defined as

$$B_{(t,q),s} = \frac{1}{q}\Big[\mathbf{1}\{t-q+1 \leq s \leq t\} - \mathbf{1}\{t+1 \leq s \leq t+q\}\Big].$$

The hypothesis of lack of structural breaks can be restated as $\mathrm{DBA}(t; \beta, q) = 0$, $\forall 2q \leq t \leq T-2q$, which corresponds to $J\beta = 0$, where the rows of J are $B'_{(t,q)} \otimes \tau'_{j_0,k}$ for all $t \in \{2q, \ldots, T-2q\}$. Notice that Assumption 2(i)-(ii) hold since $m_J = T - 4q < T$ and $\max_{1 \leq j \leq m_J} \|J_j\|_1 = 1$. By Theorem 3.3, we can test the hypothesis of lack of structural breaks by implementing Algorithm 2 with $a = 0$.

Under the alternative that $\{\beta_{j_0,t}\}_{t=1}^T$ has a piecewise constant structure as in (4.3), $\mathrm{DBA}(B_j; \beta, q) = \beta_{j_0,B_j+1} - \beta_{j_0,B_j}$; on the other hand, it is natural to expect $\mathrm{DBA}(B_j; \beta, q)$ to be better estimated than $\beta_{j_0,B_j+1} - \beta_{j_0,B_j}$ simply because the former is the difference between two averages, especially for large q. Hence, we expect that for large q, the test proposed in this subsection be more powerful in detecting structural breaks than the test discussed in Section 4.3.

Our block average setup also yields a natural estimator. The basic idea is the following. Suppose that we observe the true sequence $\{\beta_{j_0,t}\}_{t=1}^T$. Then $\mathrm{DBA}(t; \beta, q) \neq 0$ if and only if $B_j - q \leq t \leq B_j + q - 1$ for some $j \in \{1, \ldots, m\}$. Also notice that $t \mapsto |\mathrm{DBA}(t; \beta, q)|$ reaches the maximum (over $\{B_j - q, \ldots, B_j + q - 1\}$) at $t = B_j$. This suggests a recursive strategy. Suppose that we already found B_{j-1}. Let s_j denote the smallest number s such that $s \geq B_{j-1} + 2q - 1$ and $|\mathrm{DBA}(s; \beta, q)| > 0$. Then $B_j = \arg\max\{|\mathrm{DBA}(t; \beta, q)| \mid s_j \leq t \leq s_j + 2q - 1\}$.

Since $\|\tilde{\beta} - \beta\|_\infty = O_P(\sqrt{n^{-1}\log n})$, we consider a similar strategy with β replaced by $\tilde{\beta}$. Let $\delta_n = \Phi^{-1}(\hat{\Omega}, 1-\eta)/\sqrt{n}$, where $\Phi^{-1}(\hat{\Omega}, 1-\eta)$ is defined in Algorithm 2 using J described above. Starting with $\hat{B}_0 = 1$, we compute $\hat{B}_j$ recursively:

$$\hat{s}_j = \min\left\{s \mid s \geq \hat{B}_{j-1} + 2q - 1 \text{ and } |\mathrm{DBA}(s; \tilde{\beta}, q)| > \delta_n\right\} \text{ and } \hat{B}_j = \underset{\hat{s}_j \leq t \leq \hat{s}_j + 2q - 1}{\arg\max} \left|\mathrm{DBA}(t; \tilde{\beta}, q)\right|.$$

The iteration continues until $j = \hat{m}$, where $|\mathrm{DBA}(s; \tilde{\beta}, q)| \leq \delta_n$, $\forall s \geq \hat{m} + 2q - 1$. When the true $\{\beta_{j_0,t}\}_{t=1}^T$ follows the structural break pattern in (4.3) and the breaks are pronounced enough ($\sqrt{n/\log n}\min_{1\leq l \leq m} |\beta_{j_0,B_l} - \beta_{j_0,B_l+1}| \to \infty$), then Theorems 3.3 and 3.4 imply that both $\mathbb{P}(\hat{m} = m)$ and $\mathbb{P}(\{\hat{B}_1, \ldots, \hat{B}_{\hat{m}}\} = \{B_1, \ldots, B_m\})$ are asymptotically at least $1 - \eta$. For these probabilities to tend to one, we can simply choose $\delta_n = \sqrt{2n^{-1}\|\hat{\Omega}\|_\infty \log m_J}$, see Remark 4.2.

4.5 Estimating partial regime-dependence

Popular models for the time variation in parameter values often specify a pattern in which parameters take values in a small set, whose elements are often referred to as regimes. For example, models with structural breaks have parameters staying in one regime between breaks; Markov switching models, such as Hamilton (1989), often assume that the parameters follow Markov chain with a few states. Due to the flexibility of our setup, if the underlying parameters indeed follow such regime patterns and these regimes are different enough (from each other), then our results can be used to estimate the membership of these regimes, i.e., which regime contains which time periods. Notice that $\{\beta_{j_0,t}\}_{t=1}^T$ is allowed to be random here; see Remark 3.3.

Suppose that there m regimes for $\beta_{j_0,t}$, which can take value in $\{a_1, \ldots, a_m\}$. Since $\beta_{j_0,t}$ is a scalar, we assume, without loss of generality, that $a_1 < a_2 < \ldots < a_m$. For $1 \le r \le m$, define the set of time periods corresponding to the r-th regime: $\mathcal{Q}(r) = \{t \mid 1 \le t \le T \text{ and } \beta_{j_0,t} = a_r\}$. The goal is to estimate m as well as $\mathcal{Q}(r)$ for each $1 \le r \le m$.

Due to the monotonicity of $\{a_r\}_{r=1}^m$, we consider a simple sorting strategy. The basic idea is quite simple. If we could sort the true values $\{\beta_{j_0,t}\}_{t=1}^T$, then we would obtain a piecewise constant and non-decreasing path and different regimes are separated by jumps in the sorted sequence, leading to a structural break pattern in the sorted sequence. Hence, we could simply apply the techniques outlined in Section 4.4 to the sorted sequence. Since the true values $\{\beta_{j_0,t}\}_{t=1}^T$ are unknown, we simply implement this idea with $\{\tilde{\beta}_{j_0,t}\}_{t=1}^T$.

Formally, let $\pi : \{1, \ldots, T\} \mapsto \{1, \ldots, T\}$ be a permutation (bijective mapping) such that $\tilde{\beta}_{j_0,\pi(1)} \le \ldots \le \tilde{\beta}_{j_0,\pi(T)}$. Suppose that each regime contains at least $2q$ time periods. Discussions in Section 4.4 allow us to identify $\hat{\mu}$ breaks in the sequence $\{\tilde{\beta}_{j_0,\pi(s)}\}_{s=1}^T$, say $\varsigma_1, \ldots, \varsigma_{\hat{\mu}}$. Then our estimate for m is $\hat{\mu}+1$. We define

$$\widehat{\mathcal{Q}}(r) = \begin{cases} \{t \mid \pi(t) \ge \varsigma_{\hat{\mu}}\} & r = \hat{\mu}+1 \\ \{t \mid \varsigma_{r-1} \le \pi(t) < \varsigma_r\} & 2 \le r \le \hat{\mu} \\ \{t \mid \pi(t) < \varsigma_1\} & r = 1 \end{cases} \tag{4.4}$$

If $\min_{2\le j\le m}(a_j - a_{j-1})\sqrt{n}/\log n \to \infty$, then it follows, by Theorems 3.3 and 3.4, that $\mathbb{P}(\hat{\mu}+1 = m) \to 1$ and $\mathbb{P}(\bigcap_{r=1}^m \{\widehat{\mathcal{Q}}(r) = \mathcal{Q}(r)\}) \to 1$. In other words, when the regimes are different enough ($\min_{2\le j\le m}(a_j - a_{j-1})$ not too small), $\widehat{\mathcal{Q}}(r)$ can recover the regime pattern and thus be used to assess the specification of the time variation of parameters. For example, structural breaks should correspond to large blocks of time periods in which $\beta_{j_0,t}$ takes the same value, implying that, for each $1 \le r \le m$, $\mathcal{Q}(r)$ should contain consecutive time periods; regime switching patterns would imply the opposite for $\mathcal{Q}(r)$. We can conduct specifications tests. Consider, as an example, the problem of testing whether the switching pattern follows an i.i.d Bernoulli process or a first-order Markov chain. This problem reduces to testing the restrictions on transition probabilities using a sample of observed Markov chains.

4.6 Detecting general patterns of partial time-variation

In practice, big sudden shifts are not the only pattern of time variation. For example, changes in the parameters might be small at each point in time but accumulate to a large value over a long time horizon. In this case, the test discussed in previous subsections might not reveal evidence of structural breaks, but this does not mean that we should conclude that slope coefficients are time-invariant. Again, this is a partial inference problem in that the null hypothesis of time invariance only concerns $\beta_{(j_0)}$ and allows arbitrary time variation in $\{\beta_{j,t}\}_{t=1}^T$ for $j \neq j_0$.

To test for general patterns of time variation, we consider the maximal time variation. Notice that time invariance means that $\beta_{j_0,t_1} = \beta_{j_0,t_2}$, $\forall 1 \leq t_1 < t_2 \leq T$. Hence, we can consider $|\beta_{j_0,t_1} - \beta_{j_0,t_2}|$ all combinations of $t_1, t_2 \in \{1, \ldots, T\}$ with $t_1 < t_2$. We define $\iota_{t_1,t_2} \in \mathbb{R}^T$ as a vector of zeros, except that the t_1-th entry is 1 and the t_2-th entry is -1. Clearly, there are $T(T-1)/2$ vectors of this form and we form the matrix J as follows: the row $Tt_1 + t_2$ of J is $\iota'_{t_1,t_2} \otimes \tau'_{j_0,k}$. Under this notation, the null hypothesis of time invariance becomes $J\beta = 0$. Since Assumption 2(i)-(ii) hold with $m_J = T(T-1)/2$ and $\max_{1 \leq j \leq m_J} \|J_j\|_1 = 2$, Theorem 3.3 guarantees the validity of testing for time invariance using Algorithm 2 with $a = 0$.

Notice that the test statistics $\|J\tilde{\beta}\|_\infty$ is equal to $\max_{1 \leq t_1 < t_2 \leq T} |\tilde{\beta}_{j_0,t_1} - \tilde{\beta}_{j_0,t_2}|$, which is an estimate for $\max_{1 \leq t_1 < t_2 \leq T} |\beta_{j_0,t_1} - \beta_{j_0,t_2}|$, the distance between the peak and trough in $\tilde{\beta}_{j,t}$. Therefore, it follows, by Theorem 3.4, that this test can detect any time variation resulting in a parameter trajectory that cannot be contained in a band of constant width of $O(\sqrt{n^{-1}\log n})$. For this reason, this procedure can be used as a test for time invariance whenever the alternative does not specify a particular pattern of parameter changes.

4.7 Explaining time variations in the slope coefficients

One common method of explaining variations is to use regression analysis. Here, we treat $\{\beta_{j_0,t}\}_{t=1}^T$ as a stochastic process; see Remark 3.3. For example, applied researchers can analyze the randomness of $\{\beta_{j_0,t}\}_{t=1}^T$ using linear regressions:

$$\beta_{j_0,t} = z_t' \underset{m\times 1}{\theta} + \varepsilon_t, \tag{4.5}$$

where $z_t \in \mathbb{R}^m$ is the vector of observed explanatory variables with fixed m and ε_t is the error term. The goal is to conduct inference on θ. Notice that we still allow $\{\beta_{j,t}\}_{t=1}^T$ with $j \neq j_0$ to have completely different time variation patterns. Therefore, we cannot state the model (2.1) in terms of the low-dimensional parameter θ by imposing (4.5).

Were the process $\{\beta_{j_0,t}\}_{t=1}^T$ observed, one could simply estimate θ by the ordinary least squared estimator

$$\hat{\theta} = \left(\sum_{t=1}^T z_t z_t'\right)^{-1} \left(\sum_{t=1}^T z_t \beta_{j_0,t}\right).$$

However, in practice, the process $\{\beta_{j_0,t}\}_{t=1}^T$ is not observed and thus the estimator $\hat{\theta}$ is not feasible. Since Algorithm 2 delivers the estimated process $\{\tilde{\beta}_{j_0,t}\}_{t=1}^T$, we can consider the following estimator

$$\tilde{\theta} = \left(\sum_{t=1}^T z_t z_t'\right)^{-1} \left(\sum_{t=1}^T z_t \tilde{\beta}_{j_0,t}\right).$$

The following result says that, under certain conditions, $\tilde{\theta}$ and $\hat{\theta}$ are asymptotically equivalent.

Theorem 4.1. *Let Assumptions 1 and 2 hold with $J = I_{kT}$ and $\xi < 3/2$. Suppose that $\{z_t\}_{t=1}^T$ is independent of v and u. If $(T^{-1}\sum_{t=1}^T z_t z_t')^{-1} = O_P(1)$ and $\max_{1\leq t\leq T}\mathbb{E}\|z_t\|^2 = O(1)$. Then*

$$\sqrt{T}(\tilde{\theta} - \hat{\theta}) = o_P(1).$$

For inference of θ under the above linear regression framework, Theorem 4.1 says that it suffices to derive the limiting distribution of $\sqrt{T}(\hat{\theta} - \theta)$ since $\sqrt{T}(\tilde{\theta} - \theta)$ and $\sqrt{T}(\hat{\theta} - \theta)$ differ by $o_P(1)$. Therefore, the estimation error in $\tilde{\beta}$ does not contribute to the asymptotic distribution of the estimator $\tilde{\theta}$.

Remark 4.4. The key condition of Theorem 4.1 is the independence between $\{z_t\}_{t=1}^\infty$ and (v, u). Notice that we do not impose any assumption on the error term ε_t in (4.5). Our assumption here is similar in spirit to Assumption E in Bai and Ng (2006), who consider a related problem: factors are first estimated from a large panel dataset and then used as covariates in a separate regression. However, their conclusion is quite different from ours. Bai and Ng (2006) show that the estimation errors of the factors would in general influence the asymptotics in the latter regression, while Theorem 4.1 says that the estimation error in $\tilde{\beta}_t$ does not contribute to the limiting distribution of $\sqrt{T}(\tilde{\theta} - \theta)$. This is because the estimation errors of $\tilde{\beta}$ are noise in the dependent variable in regression (4.5), whereas in Bai and Ng (2006), the estimation errors of the factors affect the covariates in the regression of interest. In the regression setup, measurement errors in regressors cause a bigger problem than those in the response variable.

5 Monte Carlo Simulations

We consider both static (STA) and dynamic (DYN) models, which are specified as follows. The static model reads

$$y_{i,t} = F_{\alpha,t}' L_{\alpha,i} + \beta_{1,t} x_{1,i,t} + \beta_{2,t} x_{2,i,t} + u_{i,t}, \qquad \text{(STA)}$$

where $x_{i,t} = (x_{1,i,t}, x_{2,i,t})' = F_{Q,t}' L_{Q,i} + v_{i,t}$. Columns of F_α, F_Q and rows of v and u are generated as independent stochastic processes, which can take three specifications, denoted by GAUSS, STU-T and ARMA (specified later). We generate entries of $L_{\alpha,i}$ and $L_{Q,i}$ from i.i.d. $N(0, 1/2)$ and set the first column of $F_{Q,t}$ equal the first column of $F_{\alpha,t}$. $\beta_{1,t}$ and $\beta_{2,t}$ are drawn from i.i.d uniform distribution on $[-1, 1]$.

The dynamic model reads

$$y_{i,t} = F_{\alpha,t}' L_{\alpha,i} + \beta_{1,t} y_{i,t-1} + \beta_{2,t} \tilde{x}_{i,t} + u_{i,t}, \qquad \text{(DYN)}$$

where $\tilde{x}_{i,t} = F_{\tilde{Q},t}' L_{\tilde{Q},i} + \tilde{v}_{i,t}$ with $F_{\tilde{Q},t} \in \mathbb{R}^{r_Q - r_\alpha}$. Thus, the number of factors in the regressors are r_Q

(from both $F_{\alpha,t}$ and $F_{\tilde{Q},t}$)[13]. As before, columns of F_α, $F_{\tilde{Q}} = (F_{\tilde{Q},1}, \ldots, F_{\tilde{Q},T})'$, $\{\tilde{v}_{i,t}\}_{t=1}^T$ and rows of u are independent stochastic processes, which can take three specifications, denoted by GAUSS, STU-T and GARCH. We generate entries of $L_{\alpha,i}$ and $L_{\tilde{Q},i}$ from i.i.d. $N(0,1)$ and draw $\beta_{1,t}$ and $\beta_{2,t}$ from i.i.d uniform distribution on $[-0.7, 0.7]$.

Here, GAUSS denotes the process of i.i.d $N(0,1)$; STU-T denotes the process of i.i.d Student's t-distribution with 6 degrees of freedom normalized to have variance one. ARMA denotes the ARMA(1,1) zero-mean process with autoregressive and moving average coefficient being 0.924 and 0.592 (calibrated to quarterly data of real U.S. GDP growth), where the innovations are i.i.d zero-mean Gaussian with variance chosen such that the long-run variance of the ARMA process is one. GARCH denotes the GARCH(1,1) zero-mean process with ARCH and GARCH parameters being 0.12 and 0.85 (calibrated to monthly returns of the S&P500 index), where the standardized innovations are i.i.d zero-mean Gaussian with variance chosen such that the long-run mean variance of the GARCH process is one.

In all of our simulations, r_Q and r_α are estimated using $\hat{r}_Q^{SV}$ and $\hat{r}_\alpha^{SV}$ as discussed in Theorem 3.7. In all the tables and figures, the coverage probabilities of confidence bands and the rejection probabilities of tests are based on 2000 random samples.

For each simulated sample, we construct a 95% confidence band for the trajectory of the first entry of β_t; see Corollary 3.1 and Section 4.1. The results are reported in Table 1. As we can see, these results demonstrate decent finite-sample performance of the proposed confidence bands. The 95% confidence bands has empirical coverage probabilities around the nominal level, even for a sample size as small as $n = T = 60$. Strictly speaking, STU-T and GARCH processes do not satisfy the condition of exponential-type tails, but our procedures still perform quite well. For dynamic models, certain under coverage could occur for large n and relatively small T; this is only a finite sample problem since in our unreported results with larger sample sizes (e.g., $n = 900$ and $T = 200$), we find the coverage probability of the confidence bands close to their nominal levels.

We also consider the test for structural breaks. We keep the same specifications STA and DYN, except that $\beta_{1,t}$ is generated as

$$\beta_{1,1} = \ldots = \beta_{1,\lfloor \lambda T \rfloor} = w \qquad \text{and} \qquad \beta_{1,\lfloor \lambda T \rfloor + 1} = \ldots = \beta_{1,T} = w + \delta,$$

where $\lambda \in (0,1)$ is a fixed parameter and $\lfloor \lambda T \rfloor$ denotes the largest integer not exceeding λT. For STA and DYN specifications, w is from the uniform distribution on $[-1, 1]$ and on $[-0.7, 0.7]$, respectively. The null hypothesis of $\beta_1 = \ldots = \beta_T$ corresponds to $\delta = 0$. The deviation from the null hypothesis is measured in δ. Notice that this we only have structural breaks in $\{\beta_{1,t}\}_{t=1}^T$ since $\{\beta_{2,t}\}_{t=1}^T$ is still drawn from i.i.d uniform distributions as in STA and DYN.

We consider the test discussed in Section 4.3. The size properties of a 5% test are reported in

[13] Notice that when lagged $y_{i,t}$ is included as the regressors, we always have $r_Q \geq r_\alpha$. Since r_α factors drive $y_{i,t}$ and thus drive lagged $y_{i,t}$, which is only part of the regressors, the total number of factors driving the regressors is at least r_α.

Table 2. For static models, our test has decent size control in finite samples; for dynamic models, slight over-rejection could occur for large n and small T. In Figures 1 and 2, we plot the power curves of 5% tests under the STA-GAUSS and DYN-GAUSS specifications, respectively. As expected, the power increases with the sample size and the magnitude of δ. Interestingly, the power function is not sensitive to λ, the location of the structural break. Since we identify $\beta_{1,t}$ through cross-sectional units, rather than the time dimension, we do not need many time periods for each regime (i.e., before and after the break), a similar situation as discussed in Remark 4.3.

6 Empirical Applications

In this section, we illustrate the proposed methodology via three empirical problems: (1) stock return predictability, (2) firms' capital structure and (3) the effect of investment on economic growth.

6.1 Stock return predictability

A question of fundamental interest in finance is whether the equity risk premium is time-varying and, if so, can be predicted ahead of time as suggested by studies such as Campbell and Cochrane (1999) and Bansal and Yaron (2004). Two of the most popular predictors are the dividend yield[14] and volatility[15]. Here, we study the following regression using panel data

$$r_{i,t} = L'_{\alpha,i} F_{\alpha,t} + \theta_t d_{i,t-1} + \gamma_t \mathrm{VOL}_{i,t-1} + u_{i,t}, \tag{6.1}$$

where $r_{i,t}$ is the log excess return in period t on asset i, $d_{i,t-1}$ is the dividend yield in period $t-1$ for asset i and $\mathrm{VOL}_{i,t-1}$ denotes the variance of asset i in period t conditional on the information in period $t-1$.

We interpret θ_t (and γ_t) as capturing predictability in stock returns by means of time-variation in the dividend yield or conditional variance. In the specification in (6.1), we use a factor structure to model potential cross-sectional dependence among the error terms. These common factors include financial and macroeconomic shocks that drive the returns of all stocks, as well as time-specific and asset-specific fixed effects.[16] Due to the presence of these factors, methods based on OLS, such as the Fama-MacBeth regression, might provide inconsistent estimators even under strict exogeneity; see Appendix A for a simple example.

We use annual data on 100 equity portfolios sorted by size and book-to-market ratio and compute the dividend yield from the cum-dividend and ex-dividend return series.[17] The conditional volatility

[14]See e.g., Campbell and Shiller (1988a,b), Fama and French (1988), Hodrick (1992) and Koijen and Van Nieuwerburgh (2011)

[15]See e.g., Goyal and Santa-Clara (2003), Bakshi and Kapadia (2003), Ang et al. (2006) and Bollerslev et al. (2011).

[16]Well-known factors include the Fama-French factors (Fama and French (1992, 2016)) and macroeconomic factors in the large factor model literature, e.g., Stock and Watson (1998, 2002, 2006).

[17]The data is obtained from the website of Kenneth French.

is computed by fitting on AR(1) model with annual realized volatility. We specify the sampling period to run from 1960 to 2015 ($T = 56$ years) and retain the observations of $n = 89$ portfolios after removing 11 portfolios with missing data. We apply the methodology proposed in Sections 3 and 4. The estimate of $\beta = (\beta_1, \ldots, \beta_T)' \in \mathbb{R}^{2T}$ with $\beta_t = (\theta_t, \gamma_t)'$ and the 95% uniform confidence band are displayed in Figure 3.

The two plots in Figure 3 suggest quite different patterns of time variation for $\{\theta_t\}_{t=1}^T$ and $\{\gamma_t\}_{t=1}^T$. Since the red horizontal line in Panel A of Figure 3 representing the vector of zeros does not lie in the confidence band, we reject the hypothesis that $\theta_1 = \ldots = \theta_T = 0$. Time variation in θ_t is quite evident in Figure 3. Sporadic spikes in θ_t occur in the late 1960's and around 2000. This pattern indicates parameter instability, which is also documented in existing work[18]; the p-value of testing $\theta_1 = \ldots = \theta_T$ using the framework discussed in Section 4.3 is 0.001. Panel B of Figure 3 displays the path of time variation in the predictive power of conditional volatility. Such predictive power is mainly concentrated in the 1960's and 1970's and seems to have disappeared after 1980. We also cluster slope coefficients using structural break models; in particular, we assume that there are at least four years between breaks and apply the methodology outlined in Section 4.4. The results are reported in Table 3 and Figure 3. We find that the only structural break in θ_t occurred in the late 1990's and that there are three structural breaks in γ_t, which occurred in the late 1960's, late 1970's and early 1990's, respectively. However, we also reject the hypothesis that the parameter values are stable between the estimated structural breaks; this suggests that models with structural breaks in parameters might not be flexible enough to reveal all the features in return predictability.

Our method separates shocks in the error terms from those in the return predictability. Figure 4 plots the time series of the average noise level $n^{-1} \sum_{i=1}^n \hat{u}_{i,t}^2$, where $\hat{u}_{i,t}$ is defined in Algorithm 2. We compare Figures 3 and 4. During the recent Great Recession, the return predictability from the dividend yield and the conditional volatility was quite stable whereas large spikes are found in the average noise level. This indicates that the Great Recession only contributed to the noise in the error terms and did not change the relationship between stock returns and predictors, such as the dividend yield and conditional volatility. However, the collapse of the dot-com bubble appears to be a different kind of shock; we find large spikes in θ_t and the average noise level but not in γ_t. It is perhaps not surprising to see changes in the relationship between stock returns and the dividend yield as companies in the information technology sector, known for low dividends and realized profits, saw their stock prices soar and then plummet.

To study any seasonality in return predictability as well as its link to the macroeconomy, we also estimate model (6.1) using quarterly data over the same time periods ($T = 224$ quarters)[19]. Switching to quarterly data makes it more convenient to explore time variation related to macroeconomic variables, many of which are observed on a quarterly basis. We apply the framework outlined in Section 4.2. In Table 4, we construct confidence intervals for $d(A, B)$ (defined in (4.1)), where A and

[18] See Paye and Timmermann (2006); Lettau and Van Nieuwerburgh (2008); Viceira (1997).

[19] The conditional volatility is obtained by fitting the quarterly realized volatility to AR(4) model.

B are sets containing different time periods; average predictability corresponds to $A = \{1, \ldots, T\}$ and $B = \emptyset$. From Table 4, the average (across time) of return predictability from the dividend yield is estimated to be 0.49 and is not statistically significant from zero, while the average predictive power of volatility is negative and statistically significant, findings consistent with existing literature, see e.g., Glosten et al. (1993) and Goyal and Welch (2008).

Table 4 includes other intriguing findings. First, return predictability coefficients exhibit strong seasonality. A large literature has documented the presence of calendar effects in stock returns, i.e., different patterns of stock returns on certain days of the week, months of the year, etc.[20] Typically, these calendar effects are not conditional on other variables and thus should correspond to part of the fixed effects in (6.1). Our specification allows for both interactive fixed effects and time-heterogeneous slope coefficients and is thus flexible enough to distinguish seasonality in the error terms from seasonal changes in θ_t and γ_t. Table 4 and Figure 5 say that, on average, predictability using the dividend yield is particularly profound in the third quarter of the year and is not statistically different from zero in the other three quarters; on average, volatility has predictive power only in the second and third quarters. Our finding suggests that the calendar effects are present not only in the error terms but also in the slope coefficients.

Second, return predictability is related to the state of the macroeconomy. Numerous studies have found that stock returns are predictable only in certain stage of the business cycle, see e.g., Fama and French (1989), Rapach and Wohar (2006), Rapach et al. (2010) and Dangl and Halling (2012). Table 4 suggests that the dividend yield is informative only in economic recessions (defined by the NBER recession indicators); similar results hold if we treat as recessions periods in which the real GDP growth is smaller than its median. The predictive power of volatility is strong in NBER expansions, but not in recessions; on the other hand, this predictive power is only significant in periods with slow GDP growth. Unlike most work in the literature, we do not fit a two-regime parameter model to the data and thus our findings are not driven by specific model assumptions on the time variation in return predictability.

6.2 Firms' choice of capital structure

The study of firms' capital structure decisions is of fundamental interest in corporate finance. A large body of theoretical and empirical work has emerged to explain how corporations make decisions on the use of debt, see Titman and Wessels (1988), Harris and Raviv (1991), Rajan and Zingales (1995), Graham and Harvey (2001) and Welch (2004) among many others. In a survey paper, Frank and Goyal (2009) investigate numerous variables that can affect firms' capital structure. Following this literature, we consider the following regression:

$$LV_{i,t+1} = L'_{\alpha,i} F_{\alpha,t+1} + x'_{i,t} \beta_t + u_{i,t+1},$$

[20] See e.g., Jones et al. (1987), Keim and Stambaugh (1986), Haugen and Lakonishok (1988) and Kramer (1994).

where $LV_{i,t+1}$ is the leverage ratio of firm i at time $t+1$ and $x_{i,t}$ contains 11 covariates observed at time t for firm i.[21] We use the same data as Frank and Goyal (2009) and take the variables from Table II therein. We drop from $x_{i,t}$ variables that are either only time-specific (e.g. macroeconomic variables) or only firm-specific (e.g. whether the industry of the firm is regulated) since the effects of these variables are captured by the fixed effects. After removing missing data, we have a balanced panel with annual observations of $n = 167$ firms from 1963 to 2003 ($T = 41$).

We shall revisit the following conclusions of Frank and Goyal (2009):

(a) Firms with higher market-to-book ratios tend to have less leverage

(b) Firms with more tangible assets tend to have more leverage

(c) Firms with more profits tend to have less leverage

(d) Firms with more book assets tend to have more leverage

These conclusions are statements on the components of β_t corresponding to the following four regressors: profitability, book assets, market-to-book ratio and tangible assets. In this exercise, we focus on (1) estimates for $\beta = (\beta_1', \ldots, \beta_T)'$ and its 95% confidence sets, (2) testing for time-invariance of β_t and (3) inference on the average effect, i.e., $T^{-1}\sum_{t=1}^{T}\beta_t$. We consider two measures of the leverage ratio: the ratio of total debt to market assets (DM) and the ratio of total debt to book assets (DB). In Figures 6 and 7, we report the confidence bands for DM and DB, respectively. In Table 5, we report inference results for the average effects and time-invariance.

We find clear evidence of time variation in β_t. This is visually discernible in Figures 6 and 7. We also notice that the time variations are mostly slow changes in β_t rather than sudden abrupt changes. Applying the test for time invariance described in Section 4.6, we conclude, at the 5% significance level, that time variations are present in β_t for assets, profit and tangible assets; time invariance for the effects of market-to-book is also rejected at the 5% significance level when we use DM as the leverage ratio.

From Figures 6 and 7, we can reject the hypothesis that $\beta_{j,1} = \ldots = \beta_{j,T} = 0$ at the 5% significance level, for tangible assets, profits and book assets. From Table 5, we also reject, at the 5% level, that the average effect is zero for all the four variables of interest. Interestingly, the average effects of market-to-book ratio have different signs, depending on whether we use DM or DB as the leverage ratio, a finding consistent with Table V of Frank and Goyal (2009).

Overall, we confirm the findings in Frank and Goyal (2009), but our results also suggest quite different patterns of time variation. For example, Figures 6 and 7 show that the effects of the tangible assets change considerably and might have declined to zero or even switched signs at some point, whereas Table V of Frank and Goyal (2009) shows that the corresponding component of β_t

[21] These 11 variables are profitability, book assets, market-to-book ratio, change in assets, capital expenditure, median industry leverage, median industry growth, tangible assets, R&D expense, uniqueness and SGA (selling, general and administration) expense. See Appendix B of Frank and Goyal (2009) for detailed definitions.

has stayed away from zero in each decade and is relatively stable. Moreover, Frank and Goyal (2009) conclude that, for leverage measured by DM, the importance of profits has declined significantly since the 1950's and that of book assets has increased during that period, see Table V therein. From Figure 6, we see that the importance of profits has stayed stable if not increased. It is true that its importance might have temporarily dropped in the late 1980's, but quickly recovered in the early 1990's. Figure 6 also shows that the effect of book assets increased from zero to its peak in the late 1980's before it dropped to a level close to zero.

The above difference might suggest the benefit of our method, compared to the simple practice of dividing the sample into subsamples. In a sense, the approach adopted by Frank and Goyal (2009) amounts to specifying structural breaks that could occur only at the end of each decade for all the parameters. However, estimates from our model in Figures 6 and 7 indicate smooth and gradual changes for at least some of the parameters, such as book assets. We also see that certain trends in parameter values can reverse within one decade. These findings can serve as evidence supporting that a structural break model might not be a suitable specification for the parameters. Since different parameters can have completely different patterns of time variation, such as profits and book assets in Figure 6, it is advantageous to apply our flexible setup, which allows for any pattern of time variation in parameters.

6.3 Investment and economic growth

Our third application is related to the long-running debate on whether investment causes economic growth. Despite the obvious importance of this question, it appears that a consensus has yet to emerge. The literature contains studies that support such causality and perhaps equally many papers that conclude otherwise; see e.g., DeLong and Summers (1991), Mankiw et al. (1992), Islam (1995), Jones (1995), Blomström et al. (1996) and Bond et al. (2010).

To address this issue, we present a panel data analysis that allows for interactive fixed effects and unrestricted time-heterogeneous slope coefficients. Since the fixed effects can account for the endogeneity of investment, our setup could help shed light on any (time-varying) effect of investment on economic growth. We consider the following regression equation similar to the one studied in Blomström et al. (1996):

$$g_{i,t} = L'_{\alpha,i} F_{\alpha,t} + \theta_t INV_{i,t-1} + \gamma_t g_{i,t-1} + u_{i,t}, \tag{6.2}$$

where $g_{i,t}$ is the growth of real GDP per capita in country i in year t and $INV_{i,t-1}$ is the ratio of gross capital formation to GDP of country i in year $t-1$. The data is obtained from Penn World Table 9.0. After removing missing values, we have a balanced panel consisting of $n = 74$ countries over $T = 53$ years from 1962 to 2014.

In Table 6, we conduct inference regarding the average θ_t across time and test the time-invariance of θ_t. Figure 8 plots estimates for $\{\theta_t\}_{t=1}^{T}$ and its 95% confidence bands.

We find that the average value of θ_t across time is close to zero but that θ_t is not always zero. In Table 6, we see that the average θ_t across time is not statistically different from zero; however, time-invariance of θ_t is strongly rejected. From Figure 8, we see that the 95% confidence band for $\{\theta_t\}_{t=1}^T$ does not contain the red line representing the zero vector and does not contain any horizontal lines, implying time variation in $\{\theta_t\}_{t=1}^T$.

We also find that the average effect of investment on economic growth increased after the early 1990's. The methodology outlined in Section 4.4 is applied to the estimated $\{\theta_t\}_{t=1}^T$ in order to identify structural breaks. As shown in Figure 8, our method suggests that there is only one structural break, which occurred in the early 1990's. According to Table 6, the average effect of investment is not significantly different from zero in the pre-break periods and is significantly positive in the post-break periods. One explanation is related to advances in technology in the early 1990's. Several studies, such as Litan and Rivlin (2001) and Freund and Weinhold (2004), have found that the Internet has positive effects on productivity, management efficiency and international trade. Our findings are consistent with the possibility that the adoption of the Internet in the early 1990's increases the effect of investment on the economy. Moreover, in both pre-break and post-break periods, we reject time homogeneity in θ_t. This suggests that the usual structural break model might not be sufficient to describe the time-varying pattern in θ_t, highlighting the advantage of the proposed methodology.

We also consider the grouped fixed effects (GFE) discussed by Bonhomme and Manresa (2015). GFE assumes that the cross-sectional units can be categorized into a small number of groups and the time variation of the fixed effects is the same among nations in the same group. Since this specification can be viewed as a special case of the interactive fixed effects, Theorem 3.5 implies that the estimator $\hat{\alpha}_{i,t}$ from Algorithm 2 is a consistent estimator for GFE. Similar to Bonhomme and Manresa (2015), we estimate the group membership by applying the k-means clustering algorithm (Forgy, 1965; Lloyd, 1982) to $\hat{\alpha}_{i,t}$.

We comment on two findings under the GFE specification. First, we find a separation that roughly divides the countries in the sample into developed and developing nations. The result is reported in Figure 9. The red group in Figure 9 contains mostly developed countries, such as nations in North America, Western Europe, Australia, Japan and South Korea; the blue group in Figure 9 contains primarily developing countries, such as China, India, nations in Africa and South America.

Second, the estimated number of factors in the fixed effects is one and we find evidence supporting that the two groups are driven by this factor but with different sensitivities. In Figure 10, we plot the estimated trajectories of fixed effects in the two groups. The two paths of fixed effects display substantial co-movement but possess different volatilities: the red group has a more volatile path in the fixed effects. This suggests that the fixed effects in the two groups are driven by the same factor but the factor loading of the red group (mainly developed countries) is larger in magnitude than that of the blue group (mainly developing countries). One explanation is that developed nations, compared to developing nations, are more involved in international economic/political activities and are thus more sensitive to world-wide economic/political shocks. This also explains why the red group contains some countries that are usually classified as developing countries. For example, since the economy of Iran and Venezuela heavily relies on exporting petroleum-related products, which are closely connected to global economic trends, it is perhaps not surprising that these two countries are highly susceptible to international economic forces.

These results illustrate the benefit of our methodology. The patterns that admit economic interpretations, such as the group membership, are not results of a priori specifications that are explicitly imposed. In particular, we do not impose any restrictions on the group membership or on the co-movement of fixed effects between the two groups. Moreover, our results are robust to arbitrary time-heterogeneity in the slope coefficients. This is important since we find strong evidence of such time heterogeneity.

7 Conclusion

We consider panel data models with interactive fixed effects and time-heterogeneous slope coefficients. These models do not restrict the time-variation in the slope coefficients, while allowing for both cross-sectional and inter-temporal dependence in the error terms. As the data consists of a large number of cross-sectional observations over many time periods, the vector β containing all the slope coefficients across time has dimensionality tending to infinity.

We propose methods for estimating and conducting inference on β and establish their asymptotic properties. We treat the entire vector β as a high-dimensional parameter and provide tools for inference on the trajectory of the time-variation of slope coefficients. In particular, our results can be used to construct confidence bands for this trajectory of slope coefficients, to test for time-invariance and to conduct inference on specific patterns of time variations, including structural breaks and regime switching. Our methods are simple to implement and computationally convenient.

An interesting extension of our work is to allow covariate effects to be heterogeneous both across cross-sectional units and across time. Such a flexible framework could be quite natural in empirical applications. For example, certain treatments might have different effects on different individuals in different time periods; applied researchers might be interested in questions such as how the average (across individuals) treatment effects vary over time, whether certain (groups of) individuals are always more responsive to the treatment and whether time variation in the treatment effects is synchronized across individuals. Estimation and inference of these models would probably require certain structures on the heterogeneity in slope coefficients. To this end, one might borrow from popular specifications of fixed effects, although formal analysis is likely to encounter additional technical challenges and is left for future research.

Tables and Figures

Table 1: Coverage probability of 95% confidence bands for $\{\beta_{1,t}\}_{t=1}^{T}$

Panel A: static model (STA)

	GAUSS			STU-T			ARMA		
$r_Q = r_\alpha = 1$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.973	0.984	0.988	0.969	0.984	0.989	0.944	0.979	0.982
120	0.954	0.969	0.971	0.966	0.982	0.981	0.952	0.969	0.974
180	0.949	0.965	0.963	0.963	0.977	0.982	0.945	0.960	0.966
$r_Q = r_\alpha = 2$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.962	0.982	0.987	0.956	0.984	0.988	0.953	0.982	0.990
120	0.951	0.965	0.977	0.956	0.976	0.984	0.941	0.967	0.977
180	0.947	0.964	0.963	0.956	0.980	0.979	0.932	0.959	0.960
$r_Q = r_\alpha = 3$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.968	0.980	0.992	0.942	0.973	0.982	0.962	0.981	0.992
120	0.950	0.974	0.980	0.945	0.978	0.982	0.948	0.970	0.980
180	0.940	0.966	0.969	0.942	0.970	0.983	0.931	0.959	0.972

Panel B: dynamic model (DYN)

	GAUSS			STU-T			GARCH		
$r_\alpha = 1$ and $r_Q = 1$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.929	0.962	0.964	0.929	0.960	0.980	0.923	0.957	0.969
120	0.903	0.945	0.954	0.900	0.957	0.966	0.942	0.968	0.977
180	0.881	0.942	0.947	0.865	0.946	0.959	0.942	0.970	0.982
$r_\alpha = 1$ and $r_Q = 2$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.941	0.963	0.970	0.941	0.972	0.975	0.919	0.955	0.968
120	0.904	0.942	0.958	0.893	0.960	0.965	0.944	0.972	0.974
180	0.860	0.948	0.955	0.845	0.943	0.962	0.941	0.970	0.984
$r_\alpha = 1$ and $r_Q = 3$									
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.944	0.967	0.969	0.942	0.977	0.975	0.909	0.939	0.962
120	0.916	0.957	0.962	0.911	0.953	0.975	0.927	0.961	0.967
180	0.868	0.933	0.947	0.855	0.947	0.965	0.914	0.966	0.978

Table 2: Rejection probability under the null hypothesis that $\beta_{1,1} = \ldots = \beta_{1,T}$

Panel A: static model (STA)

				$r_Q = r_\alpha = 1$					
	GAUSS			STU-T			ARMA		
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.054	0.042	0.024	0.038	0.028	0.019	0.046	0.022	0.019
120	0.052	0.043	0.040	0.044	0.033	0.024	0.044	0.036	0.034
180	0.055	0.041	0.038	0.055	0.034	0.032	0.053	0.037	0.038
				$r_Q = r_\alpha = 2$					
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.044	0.023	0.019	0.059	0.026	0.019	0.044	0.014	0.012
120	0.057	0.039	0.031	0.056	0.030	0.020	0.048	0.030	0.033
180	0.055	0.048	0.037	0.056	0.046	0.035	0.061	0.033	0.032
				$r_Q = r_\alpha = 3$					
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.042	0.020	0.015	0.063	0.027	0.019	0.033	0.017	0.007
120	0.070	0.040	0.029	0.066	0.029	0.022	0.047	0.023	0.015
180	0.072	0.047	0.040	0.057	0.033	0.031	0.050	0.027	0.024

Panel B: dynamic model (DYN)

				$r_\alpha = 1$ and $r_Q = 1$					
	GAUSS			STU-T			GARCH		
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.078	0.049	0.041	0.055	0.032	0.031	0.073	0.056	0.047
120	0.072	0.054	0.048	0.070	0.042	0.038	0.046	0.039	0.033
180	0.084	0.062	0.056	0.086	0.051	0.040	0.047	0.028	0.027
				$r_\alpha = 1$ and $r_Q = 2$					
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.057	0.048	0.041	0.046	0.038	0.023	0.067	0.052	0.046
120	0.070	0.057	0.056	0.077	0.043	0.025	0.053	0.034	0.029
180	0.091	0.051	0.048	0.105	0.042	0.041	0.050	0.029	0.020
				$r_\alpha = 1$ and $r_Q = 3$					
$n \setminus T$	60	120	180	60	120	180	60	120	180
60	0.058	0.035	0.039	0.039	0.028	0.022	0.107	0.060	0.045
120	0.059	0.055	0.041	0.060	0.042	0.030	0.074	0.051	0.036
180	0.082	0.049	0.050	0.091	0.044	0.032	0.069	0.041	0.028

We report the rejection probabilities of tests for structural breaks with nominal size 5% under the null hypothesis that $\beta_{1,1} = \ldots = \beta_{1,T}$.

Table 3: Forecasting stock returns (annual data)

	Estimate	**t-stat**	**Conf interval**		**P-value**
					(Time variation)
Panel A: return predictability from the dividend yield $\{\theta_t\}_{t=1}^T$					
$T^{-1}\sum_{t=1}^{T}\theta_t$	-0.79	-1.44	-1.87	0.28	0.00
$T_{\theta,1}^{-1}\sum_{t=1}^{T_{\theta,1}}\theta_t$	1.05	5.09	0.64	1.45	0.00
$(T-T_{\theta,1})^{-1}\sum_{t=T_{\theta,1}+1}^{T}\theta_t$	-0.79	-1.44	-1.87	0.28	0.00
Panel B: return predictability from the conditional variance $\{\gamma_t\}_{t=1}^T$					
$T^{-1}\sum_{t=1}^{T}\gamma_t$	-0.27	-0.71	-1.01	0.47	0.00
$T_{\gamma,1}^{-1}\sum_{t=1}^{T_{\gamma,1}}\gamma_t$	5.28	3.33	2.17	8.38	0.00
$(T_{\gamma,2}-T_{\gamma,1})^{-1}\sum_{t=T_{\gamma,1}+1}^{T_{\gamma,2}}\gamma_t$	-6.02	-5.09	-8.33	-3.70	0.00
$(T_{\gamma,3}-T_{\gamma,2})^{-1}\sum_{t=T_{\gamma,2}+1}^{T_{\gamma,3}}\gamma_t$	0.60	0.71	-1.05	2.24	0.00
$(T-T_{\gamma,3})^{-1}\sum_{t=T_{\gamma,3}+1}^{T}\gamma_t$	-0.46	-0.96	-1.41	0.48	0.00

Structural break points are estimated using the methodology outlined in Section 4.4 and assuming that there are at least four years between structural breaks. For $\{\theta_t\}_{t=1}^T$, we find one break point denoted by $T_{\theta,1}$; for $\{\gamma_t\}_{t=1}^T$, we find three break points, denoted by $T_{\gamma,1}$, $T_{\gamma,2}$ and $T_{\gamma,3}$, respectively. See Figure 3 for plots for these breaks.

Table 4: Forecasting stock returns (quarterly data): difference in predictability

		θ_t (dividend yield)				γ_t (conditional volatility)			
Set A	Set B	**Est**	**t-stat**	**Conf Interval**		**Est**	**t-stat**	**Conf Interval**	
$\{1, \ldots, T\}$	$\emptyset$	0.49	1.45	-0.17	1.15	-1.66	-3.29	-2.65	-0.67
$\mathcal{Q}_1$	$\emptyset$	0.20	0.33	-0.99	1.38	0.90	0.64	-1.84	3.63
$\mathcal{Q}_2$	$\emptyset$	-0.10	-0.14	-1.51	1.30	-2.76	-3.55	-4.29	-1.23
$\mathcal{Q}_3$	$\emptyset$	2.18	4.19	1.16	3.19	-4.79	-6.91	-6.14	-3.43
$\mathcal{Q}_4$	$\emptyset$	-0.31	-0.59	-1.35	0.73	0.00	0.00	-1.79	1.78
$\mathcal{Q}_1$	$\mathcal{Q}_2$	0.30	0.38	-1.26	1.86	3.66	2.99	1.26	6.05
$\mathcal{Q}_2$	$\mathcal{Q}_3$	-2.28	-2.49	-4.07	-0.48	2.03	1.86	-0.10	4.16
$\mathcal{Q}_3$	$\mathcal{Q}_4$	2.49	3.54	1.11	3.86	-4.78	-5.61	-6.45	-3.11
$\mathcal{R}_{NBER}$	$\emptyset$	2.20	2.56	0.52	3.89	1.22	1.37	-0.53	2.96
$\mathcal{E}_{NBER}$	$\emptyset$	0.22	0.62	-0.48	0.93	-2.11	-3.70	-3.23	-0.99
$\mathcal{R}_{NBER}$	$\mathcal{E}_{NBER}$	1.98	2.16	0.18	3.77	3.33	3.12	1.23	5.42
$\mathcal{L}_{GDP}$	$\emptyset$	1.48	2.67	0.39	2.56	-4.10	-5.75	-5.50	-2.70
$\mathcal{H}_{GDP}$	$\emptyset$	-0.50	-1.52	-1.14	0.14	0.77	1.43	-0.29	1.83
$\mathcal{L}_{GDP}$	$\mathcal{H}_{GDP}$	1.98	3.24	0.78	3.17	-4.87	-6.40	-6.36	-3.38

For disjoint sets $A, B \subset \{1, \ldots, T\}$, we consider the difference in average predictability, i.e., $d(A, B)$ defined in (4.1). We report the estimates and 95% confidence intervals for $d(A, B)$, as well as the t-stat for testing $d(A, B) = 0$. The sets used in the table are defined as follows.

- For $j \in \{1, 2, 3, 4\}$, $\mathcal{Q}_j = \{t \mid 1 \leq t \leq T$ and time t is quarter j of some year$\}$. The results are also plotted in Figure 5.
- $\mathcal{R}_{NBER} = \{t \mid 1 \leq t \leq T$ and $NBER_t = 1\}$ and $\mathcal{E}_{NBER} = \{t \mid 1 \leq t \leq T$ and $NBER_t = 0\}$, where $NBER_t$ is the NBER indicator for economic recessions, which takes value one if the economy is in recession and takes value zero otherwise. Monthly data of the NBER indicators is obtained from the website of St. Louis Fed and the value of the indicator of the last month in a quarter is used as the value of that quarter.
- $\mathcal{L}_{GDP} = \{t \mid 1 \leq t \leq T$ and $GDP_t <$ median$(GDP)\}$ and $\mathcal{H}_{GDP} = \{t \mid 1 \leq t \leq T$ and $GDP_t >$ median$(GDP)\}$, where GDP_t denotes the real U.S. GDP growth in time period t and median(GDP) denotes the sample median of real GDP growth. We obtain the data from the website of St. Louis Fed.

Table 5: Determinants of firms' capital structures

	Estimate	**t-stat**	**Conf interval**		**P-value** (Time variation)
		Panel A: LV measured as DM			
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Profit},t}$	-0.72	-10.80	-0.84	-0.59	0.01
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Assets},t}$	0.06	6.65	0.04	0.08	0.01
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Mktbk},t}$	-0.03	-5.22	-0.05	-0.02	0.02
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Tang},t}$	0.19	6.96	0.13	0.24	0.02
		Panel B: LV measured as DB			
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Profit},t}$	-0.50	-6.87	-0.64	-0.35	0.00
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Assets},t}$	0.04	4.70	0.02	0.05	0.01
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Mktbk},t}$	0.02	2.79	0.00	0.03	0.25
$T^{-1}\sum_{t=1}^{T}\beta_{\text{Tang},t}$	0.18	7.67	0.13	0.23	0.00

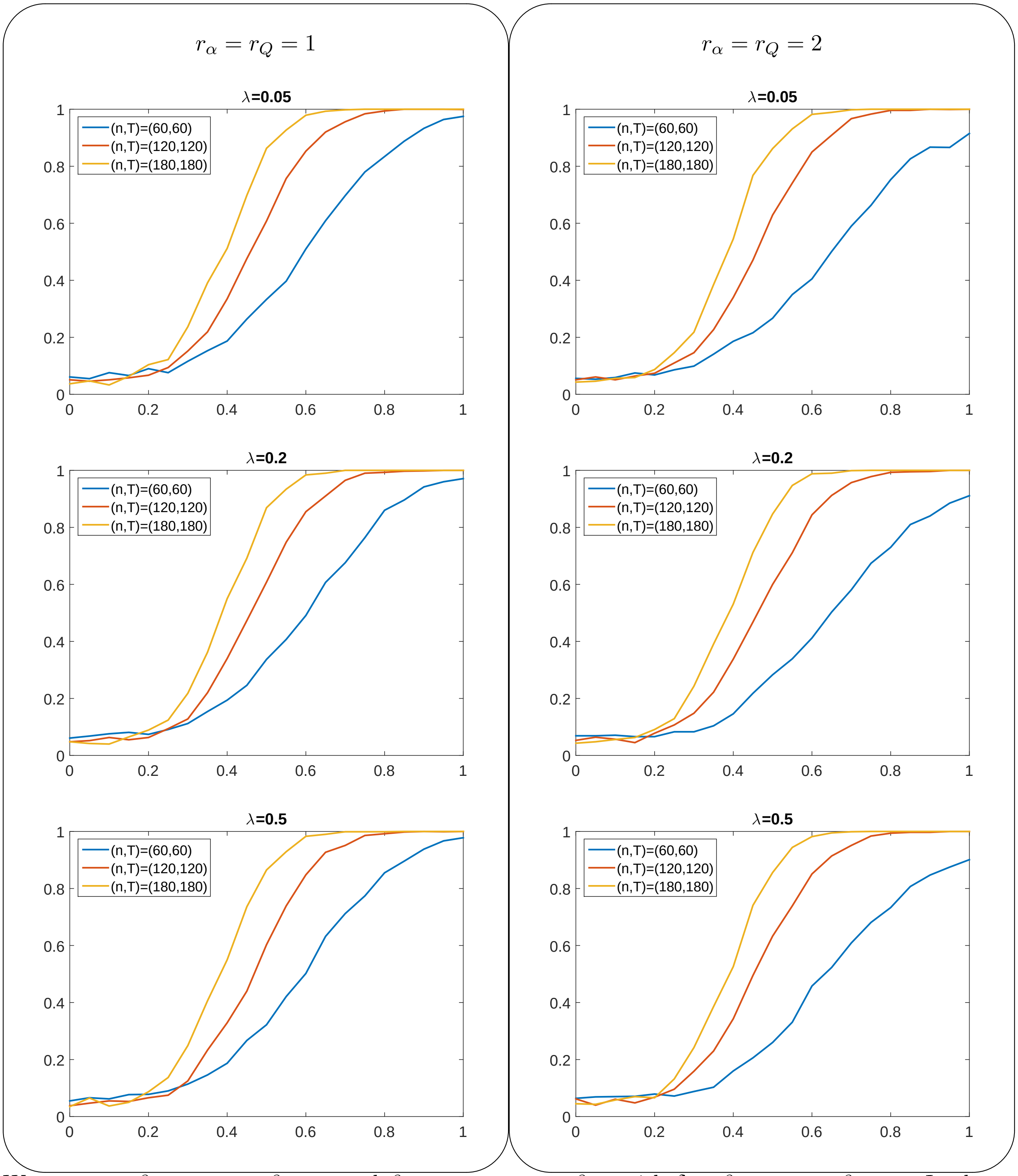
$\beta_{\text{Profit},t}$, $\beta_{\text{Assets},t}$, $\beta_{\text{Mktbk},t}$ and $\beta_{\text{Tang},t}$ represent the components of $\beta_t \in \mathbb{R}^{11}$ corresponding to profitability, assets, market-to-book ratio and tangibility, respectively. The above table reports the point estimate, t-statistic and confidence interval for the average β_t, as well as p-value of the test for lack of parameter instability of β_t described in Section 4.3.

Table 6: Fixed investment and economic growth

	Estimate	t-stat	Conf interval		P-value (Time variation)
$T^{-1}\sum_{t=1}^{T}\theta_t$	0.023	0.987	-0.022	0.067	0.000
$T_0^{-1}\sum_{t=1}^{T_0}\theta_t$	-0.055	-1.708	-0.117	0.008	0.000
$(T-T_0)^{-1}\sum_{t=T_0+1}^{T}\theta_t$	0.114	2.982	0.039	0.188	0.000

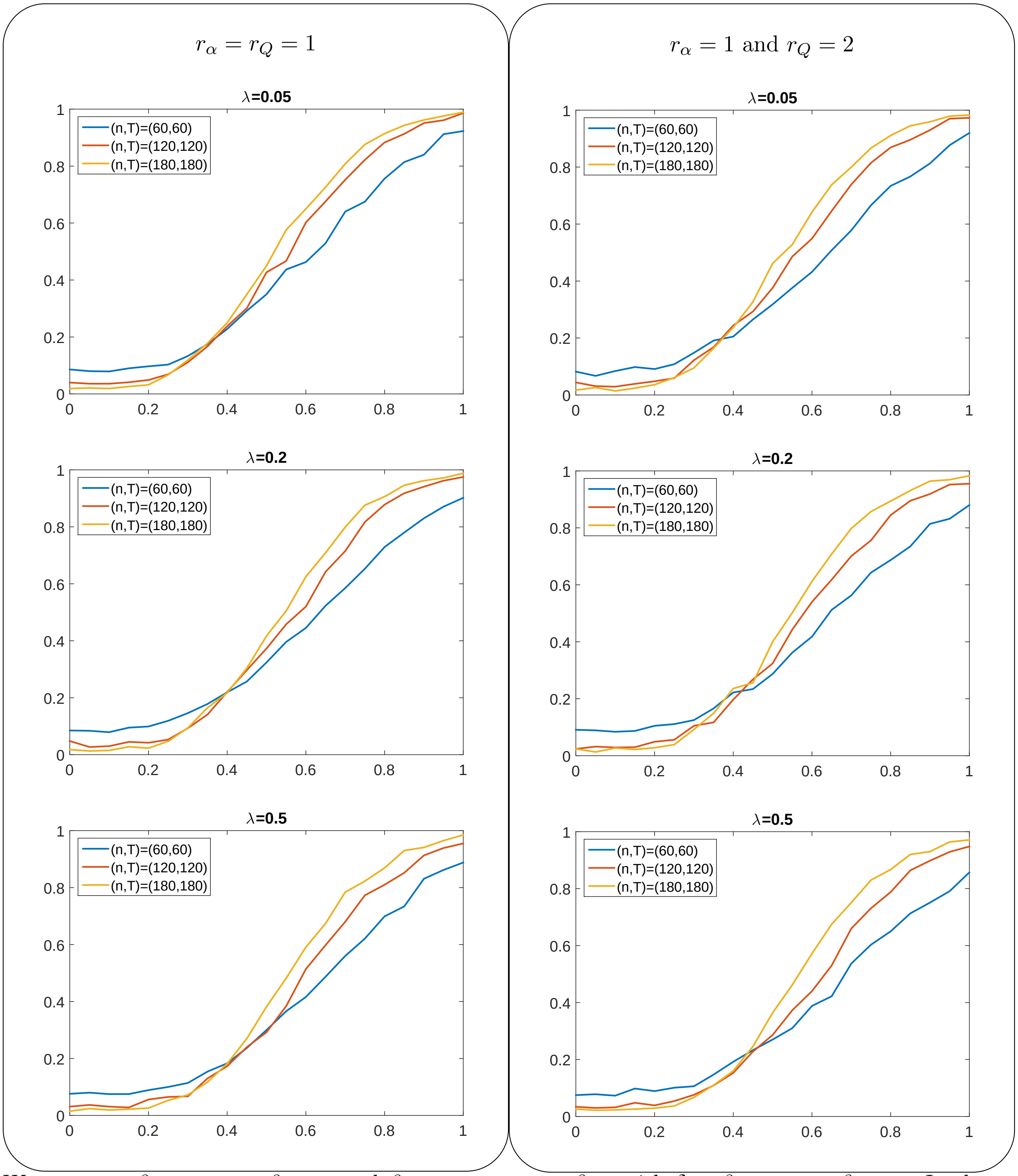
We consider the regression equation (6.2). In the above table, columns 1, 2 and 3 report the point estimate, t-statistic and confidence interval for $T^{-1}\sum_{t=1}^{T}\theta_t$. The last column reports the p-value of the test for lack of parameter instability of θ_t described in Section 4.3. T_0 is the structural break point estimated using the methodology outlined in Section 4.4; see Figure 8.

Figure 1: Power curves for testing structural breaks in $\{\beta_{1,t}\}_{t=1}^{T}$ (STA)



We generate $\beta_{1,1} = \ldots = \beta_{1,\lfloor \lambda T \rfloor}$ and $\beta_{1,\lfloor \lambda T \rfloor + 1} = \ldots = \beta_{1,T}$ with $\delta = \beta_{1,\lfloor \lambda T \rfloor + 1} - \beta_{1,\lfloor \lambda T \rfloor}$. In the above plots, we report the probability of rejecting $\beta_{1,1} = \ldots = \beta_{1,T}$ as a function of δ, for various values of (n, T) and λ.

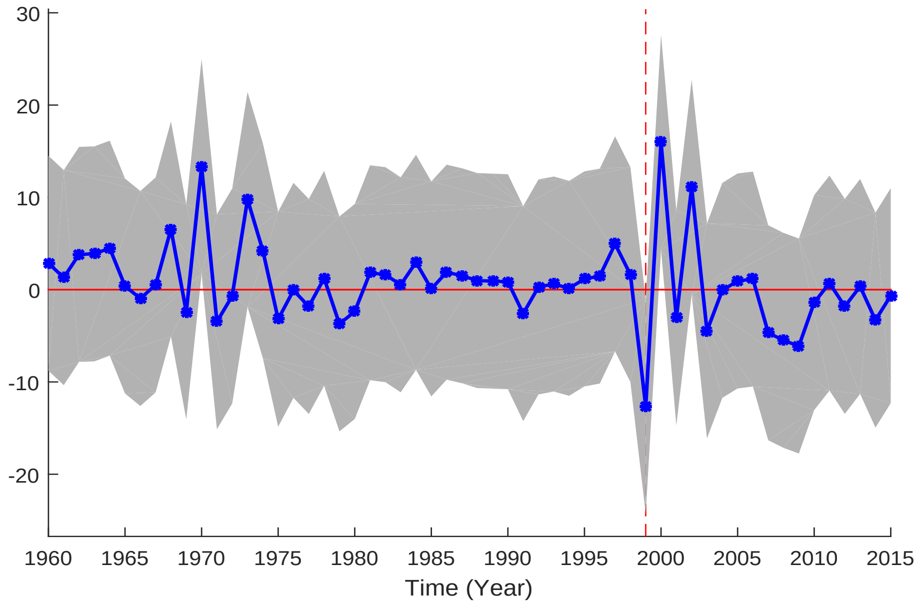
Figure 2: Power curves for testing structural breaks in $\{\beta_{1,t}\}_{t=1}^{T}$ (DYN)



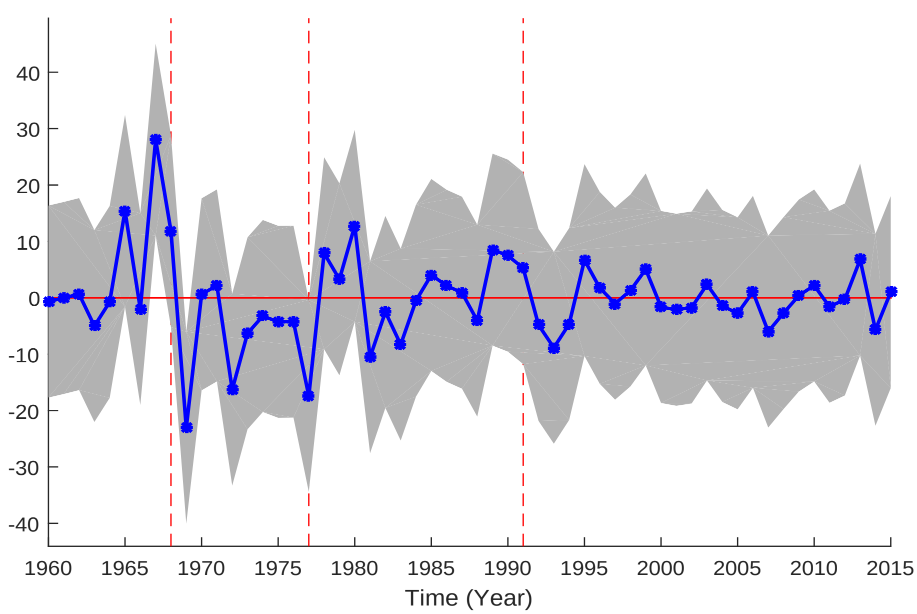
We generate $\beta_{1,1} = \ldots = \beta_{1,\lfloor \lambda T \rfloor}$ and $\beta_{1,\lfloor \lambda T \rfloor+1} = \ldots = \beta_{1,T}$ with $\delta = \beta_{1,\lfloor \lambda T \rfloor+1} - \beta_{1,\lfloor \lambda T \rfloor}$. In the above plots, we report the probability of rejecting $\beta_{1,1} = \ldots = \beta_{1,T}$ as a function of δ, for various values of (n, T) and λ.

Figure 3: Predictability of stock returns (annual data)

Panel A: Estimate and 95% confidence band for $\{\theta_t\}_{t=1}^T$ (predictive power of the dividend yield)

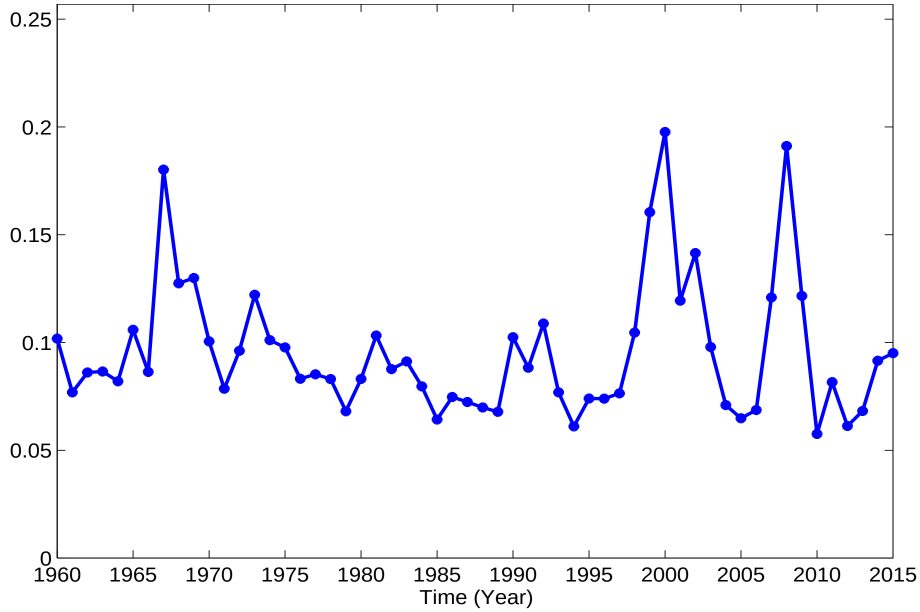


Panel B: Estimate and 95% confidence band for $\{\gamma_t\}_{t=1}^T$ (predictive power of volatility)



The blue line represents the estimate for $\{\theta_t\}_{t=1}^T$ (or $\{\gamma_t\}_{t=1}^T$) and the shaded area is the 95% confidence band. The red dashed vertical lines are the structural break points estimated using the methodology outlined in Section 4.4 and assuming that there are at least four years between structural breaks.

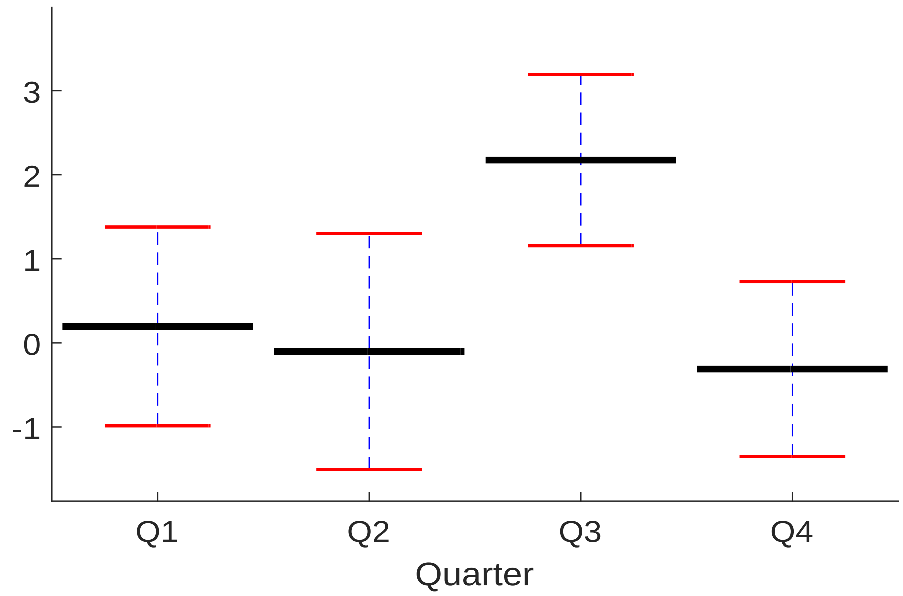
Figure 4: Predictability of stock returns (annual data): average noise level in error terms



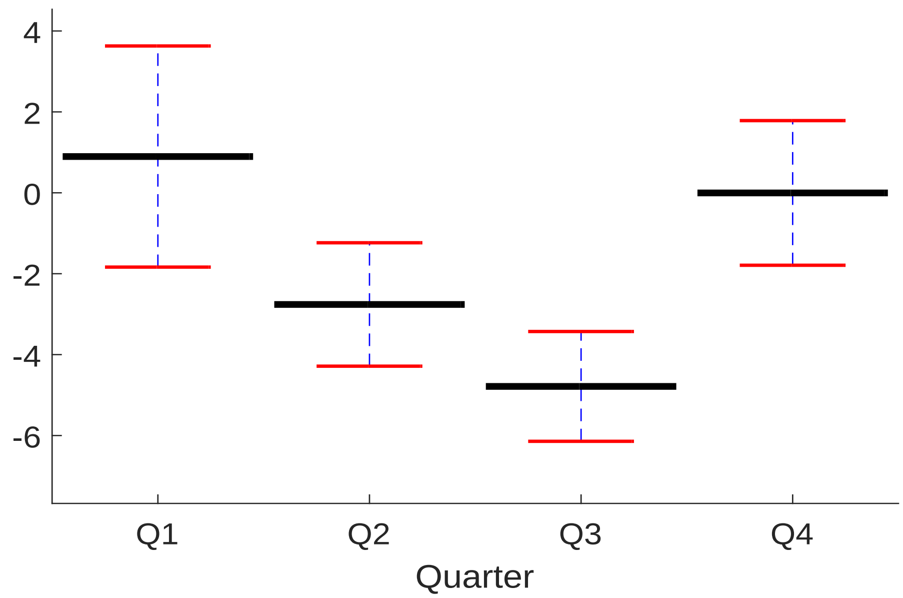
We plot the average noise level in error terms $\{\hat{\hat{\sigma}}_{u,t}\}_{t=1}^{T}$ defined by $\hat{\hat{\sigma}}_{u,t}^2 = n^{-1}\sum_{i=1}^{n}\hat{u}_{i,t}^2$, where $\hat{u}_{i,t}$ is defined in Algorithm 2.

Figure 5: Seasonality of return predictability (quarterly data)

Panel A: Average θ_t in each quarter of the year (predictive power of the dividend yield)

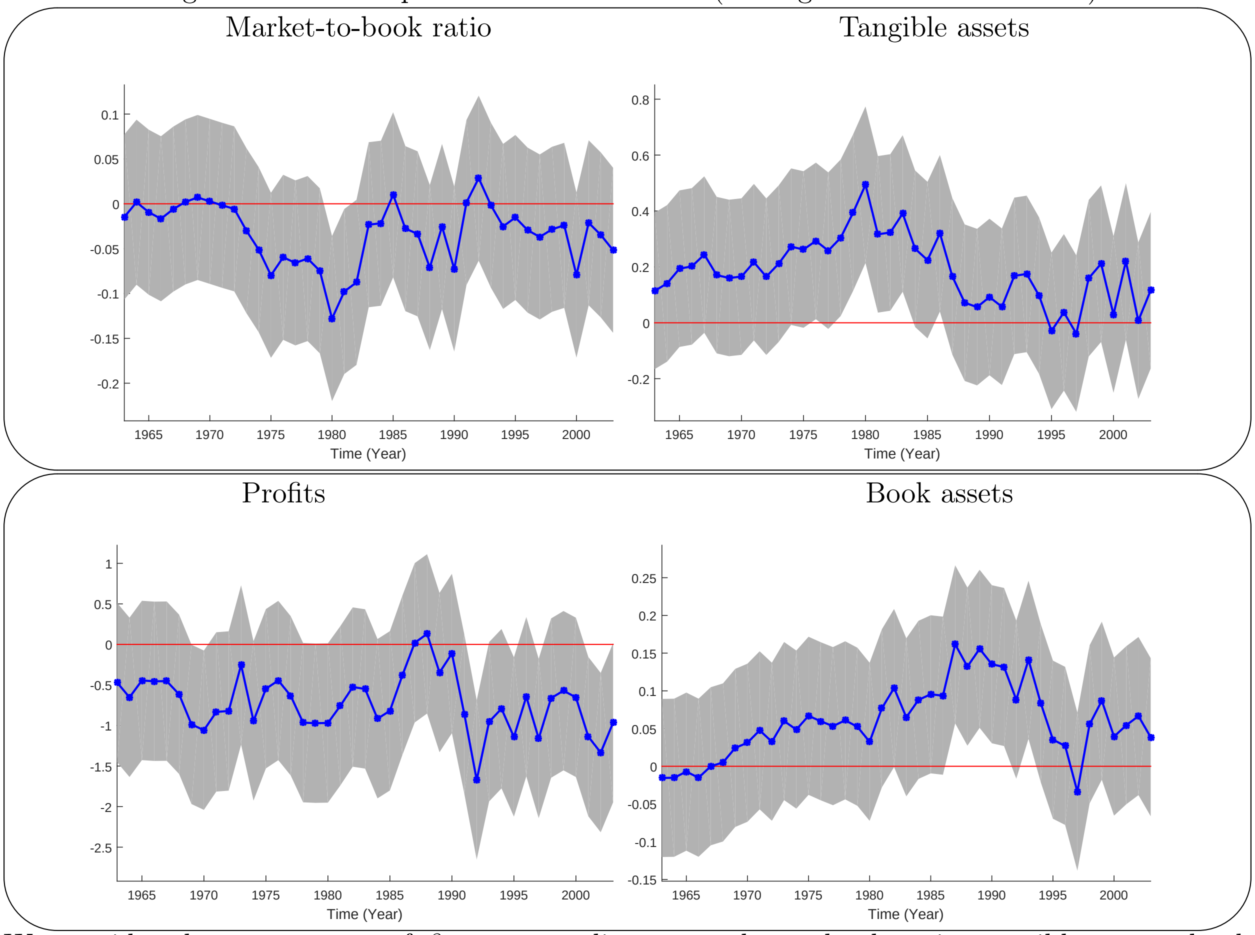


Panel B: Average γ_t in each quarter of the year (predictive power of volatility)



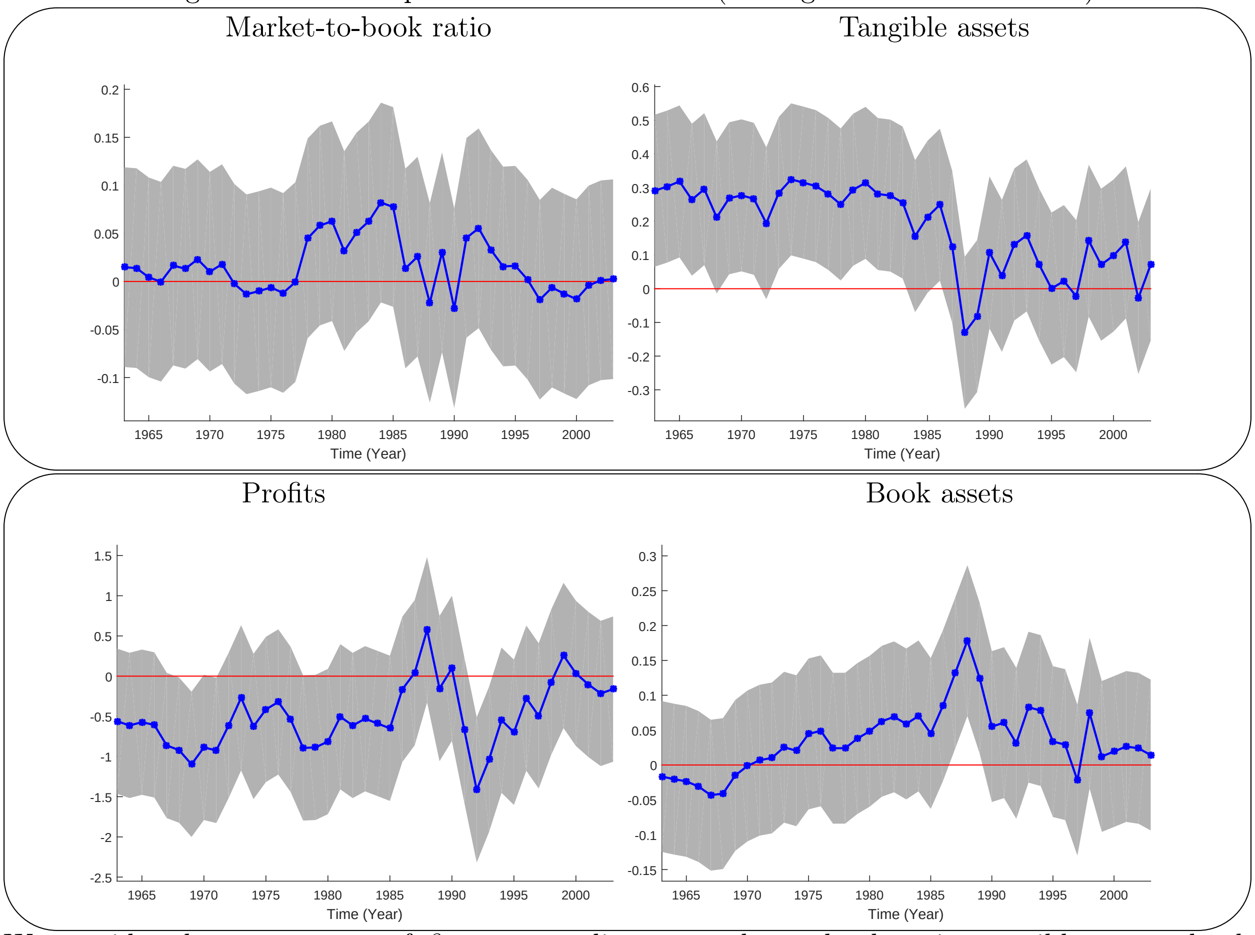
The black line represents the average θ_t (or γ_t) with one quarter of all the years and the red lines denote the 95% confidence interval of this average.

Figure 6: Firms' capital structure decisions (leverage ratio defined as DM)



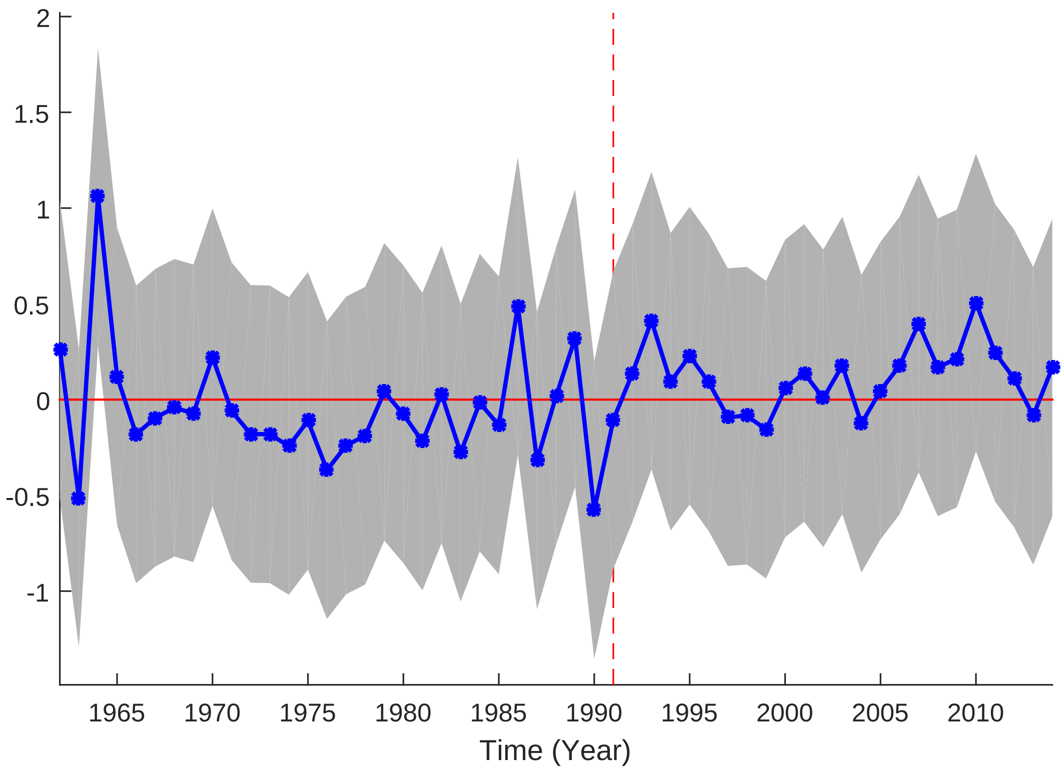
We consider the components of β_t corresponding to market-to-book ratio, tangible assets, book assets and profits. The blue lines are estimates for β. The shaded area is the 95% confidence set for β.

Figure 7: Firms' capital structure decisions (leverage ratio defined as DB)



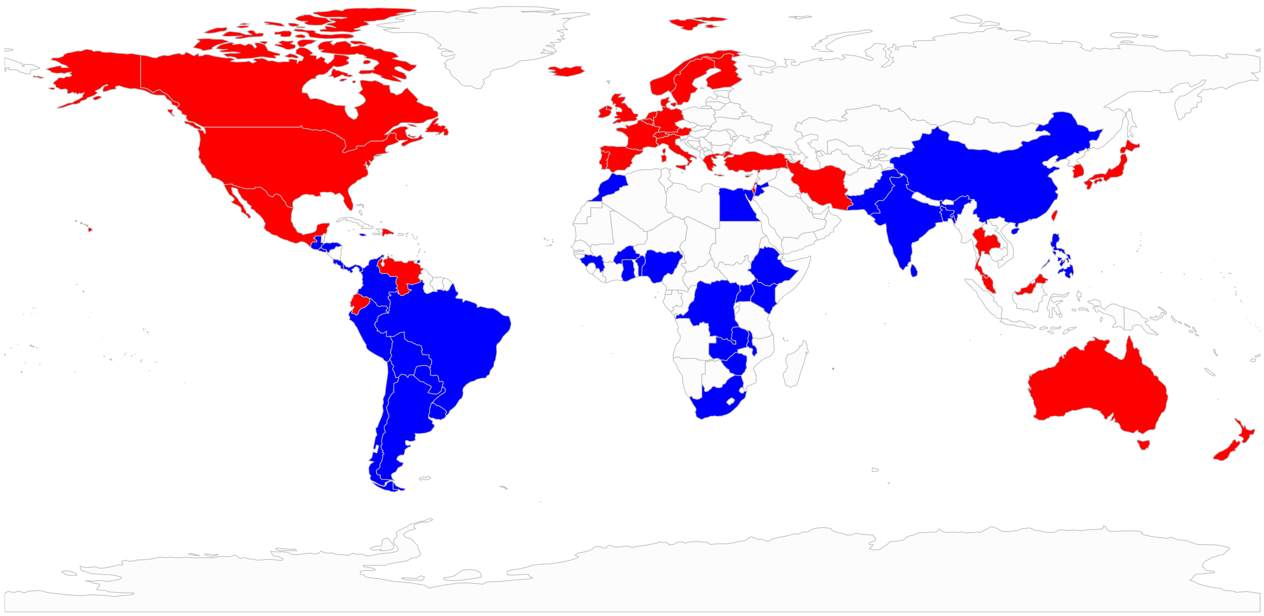
We consider the components of β_t corresponding to market-to-book ratio, tangible assets, book assets and profits. The blue lines are estimates for β. The shaded area is the 95% confidence set for β.

Figure 8: Fixed investment and economic growth



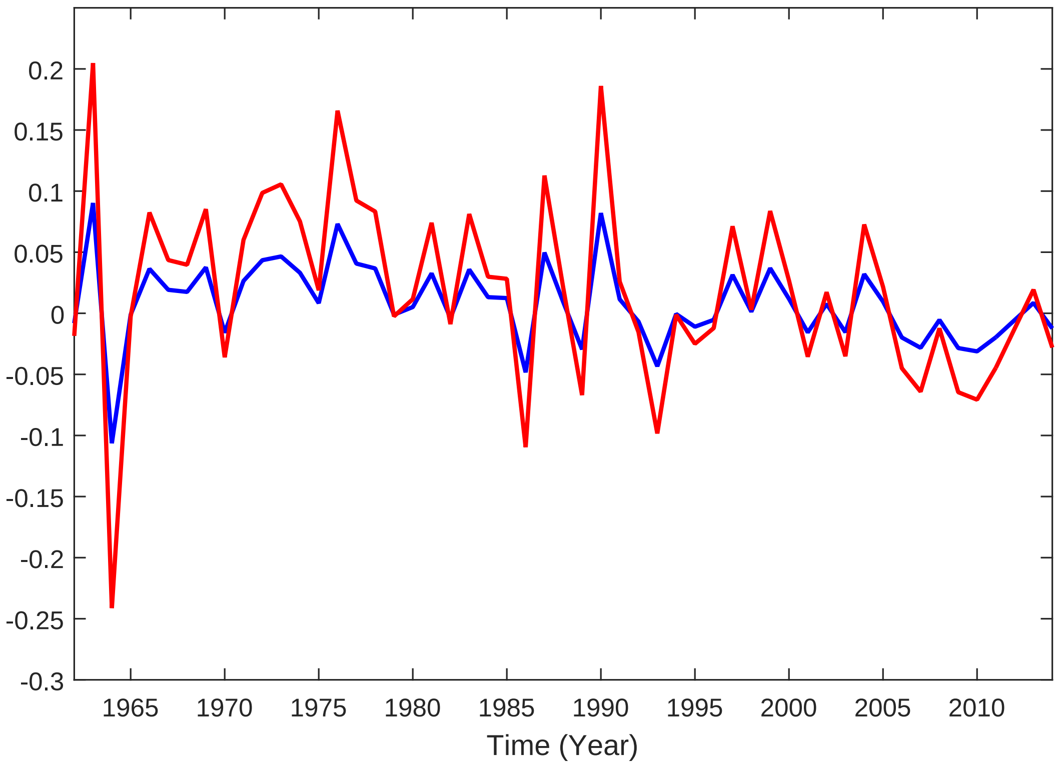
We consider the regression (6.2). The blue lines represent the estimate for $\{\theta_t\}_{t=1}^{T}$ and the shaded area is the 95% confidence band. The red dashed vertical line is the structural break point estimated using the methodology outlined in Section 4.4 and assuming that there are at least two years between structural breaks.

Figure 9: Fixed investment and economic growth: grouped pattern of fixed effects



We apply the the k-means clustering algorithm (Forgy, 1965; Lloyd, 1982) to the estimated fixed effects $\hat{\alpha}_{i,t}$ obtained in Algorithm 2. The estimated fixed effects are clustered into two groups, labeled by the red and blue colors.

Figure 10: Fixed investment and economic growth: trajectories of grouped pattern of fixed effects



We apply the the k-means clustering algorithm (Forgy, 1965; Lloyd, 1982) to the estimated fixed effects $\hat{\alpha}_{i,t}$ obtained in Algorithm 2. The estimated fixed effects are clustered into two groups, labeled by the red and blue colors. Here, we plot trajectories of the average fixed effects in the two groups under the same color label as in Figure 9. For example, the red color represents the same group in both this figure and Figure 9.

Appendix

In Appendix A, we provide a simple example illustrating the difficulties arising from the cross-sectional dependence in the error terms. The rest of the appendix contains proofs for the technical results in the main text. The theoretical results in the main text are proved in Appendix B. In Appendix C, we provide useful technical tools that are used in Appendix B.

We introduce some notations that will be used extensively for the rest of the paper. We denote $\max\{a, b\}$ and $\min\{a, b\}$ by $a \vee b$ and $a \wedge b$, respectively. For a matrix A, we define $\|A\|_\infty = \|\text{vec}A\|_\infty$. For a positive integer q, we define $[q] = \{1, \ldots, q\}$. For a set A, $|A|$ denotes the cardinality (number of elements) of A. We will repeatedly use the notation $O_P(\log^{O(1)} n)$ to denote a term of order $O_P((\log n)^r)$ for some constant $0 < r < \infty$. Finally, "wpa1" denotes the phrase "with probability approaching one".

A An example of difficulties due to cross-sectional dependence

Suppose that $y_{i,t} = x_{i,t}'\beta_t + \varepsilon_{i,t}$ with $\varepsilon_{i,t} = u_{i,t} + L_i'F_t$. Assume that, for each t, $\{(L_i, x_{i,t})\}_{i=1}^n$ is independent of $\{(F_t, u_{i,t})\}_{i=1}^n$, $\mathbb{E}F_t = 0$ and $\mathbb{E}u_{i,t} = 0$. Therefore, strict exogeneity holds: $\mathbb{E}(\varepsilon_{i,t} \mid \{x_{i,t}\}_{i=1}^n) = 0$.

However, the OLS estimator for each t might not be consistent for β_t. To see this, let $\hat{\beta}_{OLS,t} = (\sum_{i=1}^n x_{i,t}x_{i,t}')^{-1}(\sum_{i=1}^n x_{i,t}y_{i,t})$. The estimation error takes the form

$$\hat{\beta}_{OLS,t} - \beta_t = \left[n^{-1}\sum_{i=1}^n x_{i,t}x_{i,t}'\right]^{-1}\left[n^{-1}\sum_{i=1}^n x_{i,t}u_{i,t} + \left(n^{-1}\sum_{i=1}^n x_{i,t}L_i'\right)F_t\right]$$

The problem is that $n^{-1}\sum_{i=1}^n x_{i,t}L_i'$ need not be close to zero since the sequence $\{L_{\alpha,i}d_{i,t-1}\}_{i=1}^n$ might not have weak dependence across i and $\mathbb{E}x_{i,t}L_i'$ could be non-zero.

B Proofs of theoretical results in the main text

We provide the proof of Theorem 3.1 in Appendix B.1. Appendix B.2 contains proofs of Theorems 3.3, 3.4 and 3.5, as well as Theorem 3.2 and Corollary 3.1. Other results, including Theorems 3.6, 3.7 and 4.1, are proved in Appendix B.3. In Appendix B.4, we show Lemma B.16, which establishes strong mixing properties for the process described in Example 2.1. We recall some definitions used in the main text as well as introducing some new definitions that will be used in the rest of this section:

$$\begin{cases} \Sigma_t &= n^{-1} \sum_{i=1}^n \mathbb{E} v_{t,i} v_{t,i}' \\ \bar{v}_{i,t} &= \Sigma_t^{-1} v_{t,i} \\ G_{i,t} &= \bar{v}_{i,t} u_{i,t} \\ \hat{\Sigma}_t &= n^{-1} \hat{v}_t' \hat{v}_t = n^{-1} \sum_{i=1}^n \hat{v}_{i,t} \hat{v}_{i,t}' \\ \hat{\bar{v}}_{i,t} &= \hat{\Sigma}_t^{-1} \hat{v}_{i,t} \\ \hat{G}_{i,t} &= \hat{\bar{v}}_{i,t} \hat{u}_{i,t} \\ G_i &= (G_{i,1}', \ldots, G_{i,T}')' \\ \hat{G}_i &= (\hat{G}_{i,1}', \ldots, \hat{G}_{i,T}')' \\ \Omega &= n^{-1} \sum_{i=1}^n \mathbb{E} J G_i G_i' J' \\ \hat{\Omega} &= n^{-1} \sum_{i=1}^n J \hat{G}_i \hat{G}_i' J' \\ u_t &= (u_{1,t}, \ldots, u_{n,t})' \in \mathbb{R}^n \\ v_t &= (v_{1,t}, \ldots, v_{n,t})' \in \mathbb{R}^{n \times k} \\ \hat{v}_t &= (\hat{v}_{1,t}, \ldots, \hat{v}_{n,t})' \in \mathbb{R}^{n \times k} \\ \alpha_t &= (\alpha_{1,t}, \ldots, \alpha_{n,t})' \in \mathbb{R}^n \qquad \text{with } \alpha_{i,t} = L_{\alpha,i}' F_{\alpha,t} \\ \hat{\alpha}_t &= \hat{L}_\alpha \hat{F}_{\alpha,t} \\ \hat{u}_t &= y_t - X_t \hat{\beta}_t - \hat{\alpha}_t \\ D_{n,t} &= n^{-1/2} \hat{\Sigma}_t^{-1} \hat{v}_t' (\alpha_t - \hat{\alpha}_t) + n^{-1/2} \left(\hat{\Sigma}_t^{-1} \hat{v}_t' - \Sigma_t^{-1} v_t' \right) u_t \\ D_n &= (D_{n,1}', \ldots, D_{n,T}')' \end{cases} \tag{B.1}$$

B.1 Proof of Theorem 3.1

Lemma B.1. *Under Assumption 1, the following hold:*

(1) $\|L_Q\|_\infty$, $\|L_\alpha\|_\infty$, $\|F_Q\|_\infty$, $\|F_\alpha\|_\infty$, $\|u\|_\infty$, $\|v\|_\infty$, $\max_{(i,t)\in[n]\times[T]} \|\bar{v}_{i,t}\|$, $\max_{i,t} \|\bar{v}_{i,t} u_{i,t}\|$, *and* $\max_{(i,t)\in[n]\times[T]} \|x_{i,t}\|$ *are* $O_P(\log^{O(1)} n)$.

(2) both $\|u\|$ *and* $\|v\|$ *are* $O_P(\sqrt{n \log n})$.

Proof. **Proof of part (1).** The first six claims hold by Lemma C.7 and the exponential-type tails in Assumption 1.

To bound $\max_{i,t} \|\bar{v}_{i,t}\|$, notice that the $\|\cdot\|_1$-norm of rows of Σ_t^{-1} are bounded by some constants due to Assumption 1. Therefore, by Lemma C.3(1), entries of $\bar{v}_{i,t}$ have exponential-type tails with parameters that depend only on the constants in Assumption 1. Thus, Lemma C.7 implies $\max_{(i,t)\in[n]\times[T]} \|\bar{v}_{i,t}\| \leq \sqrt{k} \max_{(i,t)\in[n]\times[T]} \|\bar{v}_{i,t}\|_\infty = \sqrt{k} O_P(\log^{O(1)} |[n] \times [T]|) = O_P(\log^{O(1)} n)$.

To see the bound for $\max_{i,t} \|\bar{v}_{i,t} u_{i,t}\|$, notice that Lemma C.3(3) implies the exponential-type tail for entries of $\bar{v}_{i,t} u_{i,t}$. Then the bound follows by Lemma C.7.

To see the last claim of part (1), notice that $x_{i,t} = L'_{Q,i}F_{Q,t} + v_{i,t}$. Since entries of $L_{Q,i}$, $F_{Q,t}$ and $v_{i,t}$ have exponential-type tails, it follows, by Lemma C.3, that entries of $x_{i,t}$ also have exponential-type tails with parameters that only depend on the constants in Assumption 1. Thus, the bound for $\max_{(i,t)\in[n]\times[T]} \|x_{i,t}\|$ follows by Lemma C.7. We have proved part (1).

Proof of part (2). We apply the random matrix theory. By Theorem 5.48 and Remark 5.49 in Vershynin (2010),

$$\mathbb{E}\|u\| \leq C^{1/2}n^{1/2} + \bar{C}\sqrt{m\log(n \wedge T)}, \tag{B.2}$$

where $\bar{C}$ is an absolute constant and $m := \mathbb{E}\max_i \|\underline{u}_i\|^2$, where $\underline{u}_i = (u_{i,1}, \ldots, u_{i,T})' \in \mathbb{R}^T$. Let $s_i^2 = \mathbb{E}\|\underline{u}_i\|^2$.

By Lemma C.3(3)-(4), there exists a constant $b_* > 0$ such that $u_{i,t}^2 - \mathbb{E}u_{i,t}^2$ has an exponential-type tail with parameter (b_*, γ_1), where $\gamma_1 = \gamma_*/2$. Let $\gamma_2 = \min\{\gamma_{**}, 1/2\}$. Then $\alpha_n(t) \leq b_2\exp(-t^{\gamma_2})$ and $\gamma_2 < 1$. Hence, $\gamma = (\gamma_1^{-1} + \gamma_2^{-1})^{-1} < \gamma_2 < 1$. By Theorem 1 in Merlevède et al. (2011), there exist positive constants $C_1, \ldots, C_5 > 0$ depending only on b_*, b_2, γ and γ_2 such that $\forall x > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\left|\|\underline{u}_i\|^2 - s_i^2\right| > a_n x\right) = \mathbb{P}\left(\left|\sum_{t=1}^{T}\left(u_{i,t}^2 - \mathbb{E}u_{i,t}^2\right)\right| > a_n x\right) \\
&\leq T\exp\left(-C_1 a_n^\gamma x^\gamma\right) + \exp\left(-\frac{C_2 a_n^2 x^2}{1 + C_3 T}\right) + \exp\left[-\frac{C_4 a_n^2 x^2}{T}\exp\left(C_5 (a_n x)^{\gamma/(1-\gamma)}(\log a_n x)^{-\gamma}\right)\right],
\end{aligned}$$

where $a_n = d_*\sqrt{T\log n}$ and d_* is a constant to be determined. The union bound implies that $\forall x > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| > a_n x\right) \leq \sum_{i=1}^{n}\mathbb{P}\left(\left|\|\underline{u}_i\|^2 - s_i^2\right| > a_n x\right) \\
&\leq nT\exp\left(-C_1 a_n^\gamma x^\gamma\right) + n\exp\left(-\frac{C_2 a_n^2 x^2}{1 + C_3 T}\right) + n\exp\left[-\frac{C_4 a_n^2 x^2}{T}\exp\left(C_5 (a_n x)^{\gamma/(1-\gamma)}(\log a_n x)^{-\gamma}\right)\right].
\end{aligned}$$

Thus, by elementary computations, we can choose large constants $a_*, b_*, d_* > 0$ such that $\forall x \geq a_*$

$$\mathbb{P}\left(\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| / \left(d_*\sqrt{T\log n}\right) > x\right) = \mathbb{P}\left(\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| > a_n x\right) \leq b_*\exp\left(-x^\gamma\right). \tag{B.3}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| / \left(d_*\sqrt{T\log n}\right) &\overset{(i)}{=} \int_0^\infty \mathbb{P}\left(\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| / \left(d_*\sqrt{n\log n}\right) > x\right) dx \\
&\leq a_* + \int_{a_*}^\infty \mathbb{P}\left(\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| / \left(d_*\sqrt{n\log n}\right) > x\right) dx \\
&\overset{(ii)}{\leq} a_* + b_*\int_{a_*}^\infty \exp\left(-x^\gamma\right) dx
\end{aligned}$$

$$= \quad O(1),$$

where (i) follows by the identity $\mathbb{E}X = \int_0^\infty \mathbb{P}(X > x)dx$ for any non-negative random variable X and (ii) holds by (B.3). The above display implies that

$$m := \mathbb{E}\max_i \|\underline{u}_i\|^2 \leq \mathbb{E}\max_i \left|\|\underline{u}_i\|^2 - s_i^2\right| + \max_i s_i^2 = \sqrt{T\log n}O(1) + Cn$$
$$= O(n \vee \sqrt{T\log n}) \overset{(i)}{=} O(n),$$

where (i) holds by $T \asymp n^\xi$ with $\xi \in (6/7, 2)$. The above display and (B.2) implies that $\mathbb{E}\|u\| = O(\sqrt{n\log n})$ and thus $\|u\| = O_P(\sqrt{n\log n})$. This proves part (2) for $\|u\|$. The result for $\|v\|$ follows by an analogous argument. The proof is complete. □

Lemma B.2. *Let Assumption 1 hold. Then the following hold:*
(1) $\max_{(i,j,t)\in[n]\times[k]\times[T]} \sum_{s=1}^T |\mathbb{E}(v_{i,t,j}u_{i,s})| = O(1)$;
(2) $\max_{(i,t,j_1,j_2)\in[n]\times[T]\times[k]\times[k]} \sum_{s=1}^T |\mathbb{E}(v_{i,t,j_1}v_{i,s,j_2})| = O(1)$;
(3) $\max_{(i,s)} \sum_{t=1}^T |\mathbb{E}(G_{i,t}'G_{i,s})| = O(1)$.

Proof. We first show part (1). By the exponential-type tails, Lemma C.3(2) implies that there exists a constant $C > 0$ such that $\max_{i,t,j}\mathbb{E}|v_{i,t,j}|^4 \leq C$ and $\max_{i,s}\mathbb{E}|u_{i,s}|^4 \leq C$. By Corollary 16.2.4 of Athreya and Lahiri (2006), we have that, $\forall i,t,j,s$, $|\mathbb{E}(v_{i,t,j}u_{i,s})| \leq 4\left[2\alpha_{mixing}(|t-s|)\right]^{1/2} C^2$. Therefore,

$$\max_{i,j,t}\sum_{s=1}^T |\mathbb{E}(v_{i,t,j}u_{i,s})| \leq 4\sqrt{2}C^2 \max_t \sum_{s=1}^T \sqrt{\alpha_{mixing}(|t-s|)}$$
$$\overset{(i)}{\leq} 4\sqrt{2}C^2 c_* \max_t \sum_{s=1}^T \exp[-|t-s|^{\gamma_{**}}] \leq 8\sqrt{2}C^2 c_* \sum_{\tau=1}^\infty \exp[-\tau^{\gamma_{**}}],$$

where (i) holds by Assumption 1. Since $\sum_{\tau=1}^\infty \exp[-\tau^{\gamma_{**}}] < \infty$, the part (1) follows.

Notice that for i,t,s,j_1,j_2, Corollary 16.2.4 of Athreya and Lahiri (2006) still implies that $|\mathbb{E}(v_{i,t,j_1}v_{i,s,j_2})| \leq 4\left[2\alpha_{mixing}(|t-s|)\right]^{1/2} C^2$. Part (2) follows by the same argument.

To see part (3), let $\bar{v}_{i,t,j}$ denote the j-th component of $\bar{v}_{i,t}$ (defined in (B.1)). Since each row of Σ_t^{-1} is bounded in $\|\cdot\|_1$-norm and entries of $v_{i,t}$ have exponential-type tails, it follows, by Lemma C.3(1), that $\bar{v}_{i,t,j}$ has an exponential-type tail. Using Lemma C.3(3), we have that $\bar{v}_{i,t,j}u_{i,t}$ has an exponential-type tail. Then by the same argument as in the proof of part (1), we have that $\max_{(i,s,j)\in[n]\times[T]\times[k]} \sum_{t=1}^T |\mathbb{E}\bar{v}_{i,t,j}u_{i,t}\bar{v}_{i,s,j}u_{i,s}| = O(1)$. Since k is fixed, part (3) follows by $G_{i,t}'G_{i,s} = \sum_{j=1}^k \bar{v}_{i,t,j}u_{i,t}\bar{v}_{i,s,j}u_{i,s}$. The proof is complete. □

Lemma B.3. *Under Assumption 1, the following hold:*
(1) $\max_{i\in[n]} \|\sum_{t=1}^T v_{i,t}'F_{Q,t}'\| = O_P(T^{1/2}\log^{O(1)} n)$.
(2) $\max_{i\in[n]} \|\sum_{t=1}^T u_{i,t}F_{\alpha,t}'\| = O_P(T^{1/2}\log^{O(1)} n)$.

(3) $\max_{t\in[T]} \|L_Q' v_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(4) $\max_{t\in[T]} \|L_\alpha' v_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(5) $\max_t \|L_Q' u_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(6) $\max_t \|L_\alpha' u_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(7) $\max_{t\in[T]} \|v_t' u_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(8) $\max_{t\in[T]} \|n^{-1} v_t' v_t - \Sigma_t\| = O_P(n^{-1/2}\log^{O(1)} n)$.

(9) $\max_{s,t\in[T]} \|\sum_{i=1}^n (v_{i,t} x_{i,s}' - \mathbb{E} v_{i,t} x_{i,s}')\| = O_P(n^{1/2}\log^{O(1)} n)$.

(10) $\max_{t\in[T]} \|v_t' \alpha_t\| = O_P(n^{1/2}\log^{O(1)} n)$.

(11) $\max_{t\in[T]} \|v_t' u F_\alpha\| = O_P\left([\sqrt{nT}+n]\log^{O(1)} n\right)$.

(12) $\max_{t\in[T]} \|u_t' v F_Q\| = O_P\left([\sqrt{nT}+n]\log^{O(1)} n\right)$.

Proof. **Proof of part (1).** Let $j=(j_1,j_2)\in J := [n]\times[r_Q]$ and $\mathcal{F}_n$ the σ-algebra generated by F_Q. Since $\|F_Q\|_\infty = O_P(\log^{O(1)} n)$ (Lemma B.1), there exists a constant $c_0>0$ such that $\mathbb{P}(\mathcal{A}_n)\to 1$, where $\mathcal{A}_n = \{\|F_Q\|_\infty \le h_n\}$ with $h_n = \log^{c_0} n$.

Define $e_{t,j} = v_{j_1,t}' F_{Q,t}' \tau_{j_2} h_n^{-1} \mathbf{1}\{\|F_{Q,t}\|_\infty \le h_n\}$, where τ_{j_2} is the j_2th column of I_{r_Q}. Notice that $|e_{t,j}| \le |v_{j_1,t}|$. By Assumption 1, $\forall z>0$, $\mathbb{P}(|e_{t,j}|>z \mid \mathcal{F}_n) \le \mathbb{P}(|v_{j_1,t}|>z \mid \mathcal{F}_n) \le \exp[1-(z/b_*)^{\gamma_*}]$ a.s. Since $\{(e_{t,j})_{j\in J}\}_{t=1}^T$ is strong mixing with mixing coefficient $\alpha_{mixing}(\cdot)$ defined in Assumption 1 and v is independent of $\mathcal{F}_n$, $\{(e_{t,j})_{j\in J}\}_{t=1}^T$ is strong mixing in the sense of Lemma C.8. It follows, by Lemma C.8, that there exist constants $c,r>0$ depending only on the constants in Assumption 1 such that

$$\mathbb{P}\left(\left|T^{-1/2}\sum_{t=1}^T e_{t,j}\right| > z \middle| \mathcal{F}_n\right) \le \exp[1-(z/c)^r] \quad a.s. \qquad \forall j\in J,\ \forall z>0.$$

This exponential-type tail condition and Lemma C.7 imply that $\max_{j\in J} |T^{-1/2}\sum_{t=1}^T e_{t,j}| = O_P(\log^{O(1)} n)$. Therefore,

$$\max_{i\in[n]} \left\|\sum_{t=1}^n v_{i,t}' F_{Q,t}'\right\| \mathbf{1}_{\mathcal{A}_n} \le h_n r_Q^{1/2} T^{1/2} \max_{j\in J}\left|T^{-1/2}\sum_{t=1}^T e_{t,j}\right| \overset{(i)}{=} O_P(T^{1/2}\log^{O(1)} n),$$

where (i) holds by $h_n = O(\log^{O(1)} n)$. Since $\mathbb{P}(\mathcal{A}_n)\to 1$, part (1) follows.

Parts (2)-(7) follow by analogous arguments.

Proof of part (8). Notice that $n^{-1}v_t'v_t - \Sigma_t = n^{-1}\sum_{i=1}^n [v_{i,t}v_{i,t}' - \mathbb{E} v_{i,t} v_{i,t}']$. By Lemma C.3(3), there exist constants $c,r>0$ such that each entry of $v_{i,t}v_{i,t}'$ has an exponential-type tail with parameter (c,r) for all i,t. Then part (8) follows by Lemma C.6.

Part (9) follows by an analogous argument.

Part (10) follows by $\max_{t\in[T]} \|v_t'\alpha_t\| \le \max_{t\in[T]} \|v_t' L_\alpha\| \max_{t\in[T]} \|F_{\alpha,t}\|$, together with part (4) and Lemma B.1.

Proof of part (11). Let $j=(j_1,j_2)\in J := [k]\times[r_\alpha]$. Let $\mathcal{F}_n$ be σ-algebra generated by F_α. As before, since $\|F_\alpha\|_\infty = O_P(\log^{O(1)} n)$ (Lemma B.1), there exists a constant $c_0>0$ such that

$\mathbb{P}(\mathcal{A}_n) \to 1$, where $\mathcal{A}_n = \{\|F_\alpha\|_\infty \le h_n\}$ with $h_n = \log^{c_0} n$.

For $i \in [n]$ and $j = (j_1, j_2) \in J$, define $d_{i,j_2} = T^{-1/2} \sum_{s=1}^T u_{i,s} F_{\alpha,s,j_2} h_n^{-1} \mathbf{1}\{|F_{\alpha,s,j_2}| \le h_n\}$. Notice that $[v_t' u F_\alpha]_{j_1,j_2} \mathbf{1}_{\mathcal{A}_n} = T^{1/2} h_n \sum_{i=1}^n v_{i,t,j_1} d_{i,j_2}$. Notice that $\forall z > 0$,

$$\mathbb{P}(|u_{i,s} F_{\alpha,s,j_2} h_n^{-1} \mathbf{1}\{|F_{\alpha,s,j_2}| \le h_n\}| > z \mid \mathcal{F}_n) \le \mathbb{P}(|u_{i,s}| > z \mid \mathcal{F}_n) \le \exp[-(z/b_*)^{\gamma_*}].$$

Since u and F_α are independent, the sequence $\{u_{i,s} F_{\alpha,s,j_2} h_n^{-1} \mathbf{1}\{|F_{\alpha,s,j_2}|\}_{s=1}^T$ is strong mixing conditional on $\mathcal{F}_n$ in the sense of Lemma C.8. It follows, by Lemma C.8, that there exist constants $c_1, r_1 > 0$ such that $\mathbb{P}(|d_{i,j_2}| > z \mid \mathcal{F}_n) \le \exp[-(z/c_1)^{r_1}]$, $\forall z > 0$ for all i, j_2.

Since v_t is independent of $\mathcal{F}_n$, it follows, by the exponential-type tails of entries in v_t and Lemma C.3(3), that there exist constants $c_2, r_2 > 0$ such that $\mathbb{P}(|v_{i,t,j_1} d_{i,j_2}| > z \mid \mathcal{F}_n) \le \exp[-(z/c_2)^{r_2}]$ $\forall z > 0$. Thus, Lemma C.6 implies that

$$\max_{(t,j)\in[T]\times J} \left| \sum_{i=1}^n [v_{i,t,j_1} d_{i,j_2} - \mathbb{E}(v_{i,t,j_1} d_{i,j_2})] \right| = O_P(\sqrt{n \log |[T] \times J|}) = O_P(\sqrt{n \log n}). \tag{B.4}$$

Therefore,

$$\begin{aligned}
\max_t \|u_t' v F_Q\| \mathbf{1}_{\mathcal{A}_n} &\overset{(i)}{\le} T^{1/2} \sqrt{k r_\alpha} h_n \max_{j,t} \left| \sum_{i=1}^n v_{i,t,j_1} d_{i,j_2} \right| \\
&\le T^{1/2} \sqrt{k r_\alpha} h_n \left(\max_{j,t} \left| \sum_{i=1}^n [v_{i,t,j_1} d_{i,j_2} - \mathbb{E}(v_{i,t,j_1} d_{i,j_2})] \right| + \max_{j,t} \sum_{i=1}^n |\mathbb{E}(v_{i,t,j_1} d_{i,j_2})| \right) \\
&\overset{(ii)}{=} T^{1/2} O_P(\log^{O(1)} n) \left(O_P(\sqrt{n \log n}) + \max_{j,t} \sum_{i=1}^n |\mathbb{E}(v_{i,t,j_1} d_{i,j_2})| \right),
\end{aligned} \tag{B.5}$$

where (i) follows by $[v_t' u F_\alpha]_{j_1,j_2} \mathbf{1}_{\mathcal{A}_n} = T^{1/2} h_n \sum_{i=1}^n v_{i,t,j_1} d_{i,j_2}$, (ii) follows by (B.4) and $h_n = O_P(\log^{O(1)} n)$. Notice that

$$\begin{aligned}
\max_{j,t} \sum_{i=1}^n |\mathbb{E}(v_{i,t,j_1} d_{i,j_2})| &\le \max_{j,t} \sum_{i=1}^n \mathbb{E}|\mathbb{E}(v_{i,t,j_1} d_{i,j_2} \mid \mathcal{F}_n)| \\
&\overset{(i)}{\le} \max_{j,t} T^{-1/2} \sum_{i=1}^n \mathbb{E} \left\{ \sum_{s=1}^T |\mathbb{E}(v_{i,t,j_1} u_{i,s})| \cdot |F_{\alpha,s,j_2}| h_n^{-1} \mathbf{1}\{|F_{\alpha,s,j_2}| \le h_n\} \right\} \\
&\le n T^{-1/2} \max_{j,t,i} \sum_{s=1}^T |\mathbb{E}(u_{i,t,j_1} u_{i,s})| \\
&\overset{(ii)}{=} O(n T^{-1/2}),
\end{aligned}$$

where (i) holds by the definition of d_{i,j_2} and the independence between $\mathcal{F}_n$ and $v_{i,t,j_1} u_{i,s}$ and (ii) holds by Lemma B.2. The above display and (B.5) imply that $\max_t \|u_t' v F_Q\| \mathbf{1}_{\mathcal{A}_n} = O_P\left([\sqrt{nT} + n] \log^{O(1)} n \right)$. Since $\mathbb{P}(\mathcal{A}_n) \to 1$, part (11) follows.

Part (12) follows by an analogous argument as part (11). The proof is complete. □

Lemma B.4. *Under Assumption 1,*

$$\max_{1\leq i\leq n, 1\leq t\leq T} \|\hat{v}_{i,t} - v_{i,t}\| = O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right]\log^{O(1)} n\right).$$

Proof. First notice that $\|\hat{v}_{i,t} - v_{i,t}\| = \|\hat{L}'_{Q,i}\hat{F}_{Q,t} - L'_{Q,i}F_{Q,t}\| = \|\tau'_i(\hat{L}_Q\hat{F}_{Q,t} - L_QF_{Q,t})\|$. We apply Lemma C.11 with $L = L_Q$, $F = F_Q$, $e = v$ and $a = \tau_i$, where τ_i is the ith column of I_n. By Lemmas B.1 and B.3(4), we have $\max_t \|F_{Q,t}\| = O_P(\log^{O(1)} n)$, $\|v\| = O_P(\sqrt{n}\log^{O(1)} n)$ and $\max_t \|L'_Q v_t\| = O_P(\sqrt{n}\log^{O(1)} n)$. Therefore, Lemma C.11(4) and $T \asymp n^\xi$ imply that

$$\begin{aligned}
\max_{i,t} \|\tau'_i(\hat{L}_Q\hat{F}_{Q,t} - L_QF_{Q,t})\| \quad &\leq \quad O_P(n^{-\xi}\log^{O(1)} n)\max_t \left\|\sum_{t=1}^T v'_{i,t}F'_{Q,t}\right\| \\
&\quad +O_P\left([n^{-\xi/2} + n^{-1/2}]\log^{O(1)} n\right)\max_i \|L_{Q,i}\| \\
&\quad +O_P\left([n^{1/2-\xi} + n^{-\xi/2}]\log^{O(1)} n\right)\max_i \|\tau_i\|.
\end{aligned}$$

Notice that $\max_i \|\tau_i\| = 1$ and $\max_i \|L_{Q,i}\| \leq \sqrt{r_Q}\|L_Q\|_\infty = O_P(\log^{O(1)} n)$ (due to Lemma B.1). Thus, by the above display and Lemma B.3(1), we have

$$\max_{i,t} \|\tau'_i(\hat{L}_Q\hat{F}_{Q,t} - L_QF_{Q,t})\| = O_P\left(\left[n^{-1/2} + n^{1/2-\xi} + n^{-\xi/2}\right]\log^{O(1)} n\right).$$

Since $\max\left\{n^{-1/2},\ n^{1/2-\xi},\ n^{-\xi/2}\right\} \leq 2n^{-1/2} + 2n^{1/2-\xi}$, the desired result follows. □

Lemma B.5. *Under Assumption 1, both* $\max_t \|\hat{\Sigma}_t - \Sigma_t\|$ *and* $\max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\|$ *are* $O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right]\log^{O(1)} n\right)$.

Proof. First notice that

$$\begin{aligned}
\max_t \|\hat{\Sigma}_t - \Sigma_t\| &\leq \max_t \|n^{-1}(\hat{v}'_t\hat{v}_t - v'_tv_t)\| + \max_t \|n^{-1}v'_tv_t - \Sigma_t\| \\
&\overset{(i)}{\leq} \max_t \|n^{-1}(\hat{v}_t - v_t)'\hat{v}_t\| + \max_t \|n^{-1}v'_t(\hat{v}_t - v_t)\| + O_P(n^{-1/2}\log^{O(1)} n) \\
&\leq n^{-1}\max_t \|\hat{v}_t - v_t\| \max_t \|\hat{v}_t\| + n^{-1}\max_t \|v_t\| \max_t \|\hat{v}_t - v_t\| \\
&\quad + O_P(n^{-1/2}\log^{O(1)} n),
\end{aligned} \tag{B.6}$$

where (i) holds by Lemma B.3(8). By Lemma B.4, we have that

$$\max_t \|\hat{v}_t - v_t\| \leq n^{1/2}\max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\| = O_P\left(\left[1 + n^{1-\xi}\right]\log^{O(1)} n\right). \tag{B.7}$$

By Lemma B.1(2), $\max_t \|v_t\| \leq \|v\| = O_P(n^{1/2}\log^{O(1)} n)$. Since $\xi > 1/2$ (Assumption 1), it follows that

$$\max_t \|\hat{v}_t\| \le \max_t \|v_t\| + \max_t \|\hat{v}_t - v_t\| = O_P\left(n^{1/2} \log^{O(1)} n\right). \tag{B.8}$$

Now we combine (B.6) with (B.7) and (B.8) and obtain

$$\max_t \|\hat{\Sigma}_t - \Sigma_t\| = O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right] \log^{O(1)} n\right) = o_P(1).$$

Notice that $\|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\| = \|\hat{\Sigma}_t^{-1}(\Sigma_t - \hat{\Sigma}_t)\Sigma_t^{-1}\| \le \|\hat{\Sigma}_t^{-1}\|\|\Sigma_t - \hat{\Sigma}_t\|\|\Sigma_t^{-1}\| = \|\Sigma_t - \hat{\Sigma}_t\|/(s_{\min}(\hat{\Sigma}_t)s_{\min}(\Sigma_t))$. By Lemma C.10(1), $s_k(\hat{\Sigma}_t) + s_1(\Sigma_t - \hat{\Sigma}_t) \ge s_k(\Sigma_t)$. It follows that

$$\max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\| \le \frac{\max_t \|\hat{\Sigma}_t - \Sigma_t\|}{\min_t s_{\min}(\Sigma_t)\left(\min_t s_{\min}(\Sigma_t) - \max_t \|\hat{\Sigma}_t - \Sigma_t\|\right)} = O_P(1) \max_t \|\hat{\Sigma}_t - \Sigma_t\|.$$

The proof is complete. □

Proof of Theorem 3.1. Let $X = U_X S_X V_X'$ be an SVD, where $U_X \in \mathbb{R}^{n\times n}$ and $V_X \in \mathbb{R}^{kT\times kT}$ are orthogonal matrices and $S_X = \begin{bmatrix} S_{X,1} & 0 \\ 0 & S_{X,2} \end{bmatrix} \in \mathbb{R}^{n\times kT}$ with $S_{X,1} \in \mathbb{R}^{r_Q\times r_Q}$. By the definition of $\hat{Q}$ and $\hat{v}$, we have

$$\hat{Q} = U_X \begin{bmatrix} S_{X,1} & 0 \\ 0 & 0 \end{bmatrix} V_X' \qquad \text{and} \qquad \hat{v} = U_X \begin{bmatrix} 0 & 0 \\ 0 & S_{X,2} \end{bmatrix} V_X'.$$

Therefore, $\hat{v}'\hat{Q} = 0$. This means that $\hat{v}_t'\hat{Q}_t = 0 \ \forall 1 \le t \le T$. Since $Y_t = \alpha_t + X_t\beta_t + u_t$ and $X_t = \hat{Q}_t + \hat{v}_t$, it follows that

$$\begin{aligned}\hat{\beta}_t - \beta_t = (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'Y_t - \beta_t &= (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'(\alpha_t + u_t) + (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'\hat{Q}_t \\ &\overset{(i)}{=} (\hat{v}_t'\hat{v}_t)^{-1}\hat{v}_t'(\alpha_t + u_t) = n^{-1}\hat{\Sigma}_t^{-1}\hat{v}_t'(\alpha_t + u_t),\end{aligned} \tag{B.9}$$

where (i) holds by $\hat{v}_t'\hat{Q}_t = 0$. By Lemma B.3(7) and (10), $\max_t \|v_t'u_t\| = O_P(n^{1/2}\log^{O(1)} n)$ and $\max_t \|v_t'\alpha_t\| = O_P(n^{1/2}\log^{O(1)} n)$. Hence,

$$\max_t \|v_t'(\alpha_t + u_t)\| \le \max_t \|v_t'\alpha_t\| + \max_t \|v_t'u_t\| = O_P(n^{1/2}\log^{O(1)} n). \tag{B.10}$$

Notice that

$$\begin{aligned}\max_t \|(\hat{v}_t - v_t)'(\alpha_t + u_t)\| &\le \max_t \|\hat{v}_t - v_t\| \max_t \|\alpha_t + u_t\| \\ &\le n^{1/2} \max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\| \max_t \|\alpha_t + u_t\| \\ &\le n^{1/2} \max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\| \left[\max_t \|L_\alpha F_{\alpha,t}\| + \max_t \|u_t\|\right] \\ &\overset{(i)}{\le} n^{1/2} \max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\| O_P(\sqrt{n}\log^{O(1)} n)\end{aligned}$$

$$\begin{aligned} &\overset{(ii)}{=} n^{1/2} O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right] \log^{O(1)} n\right) O_P(\sqrt{n} \log^{O(1)} n) \\ &= O_P\left(\left[\sqrt{n} + n^{3/2-\xi}\right] \log^{O(1)} n\right), \end{aligned} \tag{B.11}$$

where (i) follows by $\max_t \|L_\alpha F_{\alpha,t}\| \leq \sqrt{n r_\alpha} \|L_\alpha\|_\infty \|F_\alpha\|_\infty = O_P(\sqrt{n} \log^{O(1)} n)$ (due to Lemma B.1) and $\max_t \|u_t\| \leq \|u\| = O_P(\sqrt{n} \log^{O(1)} n)$ (due to Lemma B.1(2)) and (ii) follows by Lemma B.4. Therefore, we obtain that

$$\begin{aligned} & \max_t \|\hat{\beta}_t - \beta_t\| \\ \overset{(i)}{=} \quad & n^{-1} \max_t \left\| \hat{\Sigma}_t^{-1} \left[v_t'(\alpha_t + u_t) + (\hat{v}_t - v_t)'(\alpha_t + u_t) \right] \right\| \\ \leq \quad & n^{-1} \max_t \|\hat{\Sigma}_t^{-1}\| \left(\max_t \|v_t'(\alpha_t + u_t)\| + \max_t \|(\hat{v}_t - v_t)'(\alpha_t + u_t)\| \right) \\ \leq \quad & n^{-1} \left(\max_t \|\Sigma_t^{-1}\| + \max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\| \right) \left(\max_t \|v_t'(\alpha_t + u_t)\| + \max_t \|(\hat{v}_t - v_t)'(\alpha_t + u_t)\| \right) \\ \overset{(ii)}{=} \quad & O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right] \log^{O(1)} n\right), \end{aligned}$$

where (i) holds by (B.9) and (ii) holds by (B.10), (B.11) and Lemma B.5. The desired result follows by $\max_t \|\hat{\beta}_t - \beta_t\|_\infty \leq \max_t \|\hat{\beta}_t - \beta_t\|$. □

B.2 Proofs for Theorems 3.3, 3.4 and 3.5 and Corollary 3.1

Lemma B.6. *Let $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_T) \in \mathbb{R}^{n \times T}$ with $\tilde{u}_t = (\tilde{u}_{1,t}, \ldots, \tilde{u}_{n,t})'$ and $\tilde{u}_{i,t} = u_{i,t} + x_{i.t}'(\beta_t - \hat{\beta}_t)$. Under Assumption 1, we have*

(1) $\max_t \|X_t(\hat{\beta}_t - \beta_t)\| = O_P\left(\left[1 + n^{1-\xi}\right] \log^{O(1)} n\right)$ *and* $\|\tilde{u}\| = O_P\left([n^{\xi/2} + n^{1-\xi/2}] \log^{O(1)} n\right)$.

(2) $\max_t \|v_t'(\tilde{u} - u) F_\alpha\| = O_P\left([n + n^{\xi}] \log^{O(1)} n\right)$.

(3) $\max_t \|u_t'(\hat{v}_t - v_t)\| = O_P\left([1 + n^{1-\xi}] \log^{O(1)} n\right)$.

Proof. **Proof for part (1).** Notice that

$$\begin{aligned} \max_t \|X_t(\hat{\beta}_t - \beta_t)\| &\leq \sqrt{n} \max_{i,t} |x_{i,t}'(\hat{\beta}_t - \beta_t)| \\ &\leq \sqrt{n} \max_{i,t} \|x_{i,t}\| \max_t \|\hat{\beta}_t - \beta_t\| \overset{(i)}{=} O_P\left(\left[1 + n^{1-\xi}\right] \log^{O(1)} n\right), \end{aligned} \tag{B.12}$$

where (i) holds by Lemma B.1 and Theorem 3.1. The definition of $\tilde{u}$ implies that

$$\|\tilde{u}\| \leq \|\tilde{u} - u\| + \|u\| \leq \sqrt{T} \max_t \|X_t(\hat{\beta}_t - \beta_t)\| + \|u\| \overset{(i)}{=} O_P\left([n^{\xi/2} + n^{1-\xi/2} + n^{1/2}] \log^{O(1)} n\right),$$

where (i) follows by (B.12), Lemma B.1(2) and $T \asymp n^{\xi}$.

Notice that $n^{\xi/2}+n^{1-\xi/2}\geq 2\sqrt{n^{\xi/2}n^{1-\xi/2}}=2n^{1/2}>n^{1/2}$. Thus, $\max\{n^{\xi/2},\ n^{1-\xi/2},\ n^{1/2}\}\leq n^{\xi/2}+n^{1-\xi/2}$. Thus, the stated bound for $\|\tilde{u}\|$ follows.

Proof for part (2). The definition of $\tilde{u}$ implies that

$$\begin{aligned}
& \max_t \|v_t'(\tilde{u}-u)F_\alpha\| && \text{(B.13)}\\
= & \max_t \left\|\sum_{s=1}^T(\sum_{i=1}^n v_{i,t}x_{i,s}')(\hat{\beta}_s-\beta_s)F_{\alpha,s}'\right\| \\
\leq & \max_t \sum_{s=1}^T\left\|(\sum_{i=1}^n v_{i,t}x_{i,s}')(\hat{\beta}_s-\beta_s)F_{\alpha,s}'\right\| \\
\leq & \left[\max_t \sum_{s=1}^T\left\|\sum_{i=1}^n v_{i,t}x_{i,s}'\right\|\right]\max_s\left\|(\hat{\beta}_s-\beta_s)F_{\alpha,s}'\right\| \\
\leq & \left[\max_t \sum_{s=1}^T\left\|\sum_{i=1}^n (v_{i,t}x_{i,s}'-\mathbb{E}v_{i,t}x_{i,s}')\right\|+\max_t \sum_{s=1}^T\left\|\sum_{i=1}^n \mathbb{E}v_{i,t}x_{i,s}'\right\|\right]\max_s\left\|(\hat{\beta}_s-\beta_s)F_{\alpha,s}'\right\|.
\end{aligned}$$

By Lemma C.3(3) and (4) (applied entry-wise), there exist constants $b,\gamma>0$ such that $\forall i,t,s$, each entry of $v_{i,t}x_{i,s}'-\mathbb{E}v_{i,t}x_{i,s}'$ has an exponential-type tail with parameter (b,γ). Hence,

$$\max_t \sum_{s=1}^T\left\|\sum_{i=1}^n (v_{i,t}x_{i,s}'-\mathbb{E}v_{i,t}x_{i,s}')\right\|\leq T\max_{s,t}\left\|\sum_{i=1}^n (v_{i,t}x_{i,s}'-\mathbb{E}v_{i,t}x_{i,s}')\right\|\overset{(i)}{=}O_P(n^{1/2+\xi}\log^{O(1)}n), \tag{B.14}$$

where (i) follows by Lemma B.3(9) and $T\asymp n^\xi$. Notice that

$$\begin{aligned}
\max_t \sum_{s=1}^T\left\|\sum_{i=1}^n \mathbb{E}v_{i,t}x_{i,s}'\right\| &\leq \max_t\sum_{i=1}^n\sum_{s=1}^T\|\mathbb{E}v_{i,t}x_{i,s}'\| \\
&\overset{(i)}{\leq} \max_t\sum_{i=1}^n\sum_{s=1}^T\|\mathbb{E}v_{i,t}v_{i,s}'\|+\max_t\sum_{i=1}^n\sum_{s=1}^T\|\mathbb{E}v_{i,t}Q_{i,s}'\| \\
&\leq \max_{i,t} n\sum_{s=1}^T\|\mathbb{E}v_{i,t}v_{i,s}'\|+\max_{i,t} n\sum_{s=1}^T\|\mathbb{E}v_{i,t}Q_{i,s}'\| \\
&\overset{(ii)}{=} O(n), && \text{(B.15)}
\end{aligned}$$

where (i) holds by $x_{i,s}=Q_{i,s}+v_{i,s}$ and (ii) follows by $\max_{i,t}\sum_{s=1}^T\|\mathbb{E}v_{i,t}v_{i,s}'\|\leq\sqrt{k}\max_{i,t}\sum_{s=1}^T\|\mathbb{E}v_{i,t}v_{i,s}'\|_\infty=O(1)$ (due to Lemma B.2) and $\mathbb{E}v_{i,t}Q_{i,s}'=0$ (by the independence between $v_{i,t}$ and $Q_{i,s}$). Also observe that

$$\max_s\left\|(\hat{\beta}_s-\beta_s)F_{\alpha,s}'\right\|\leq\sqrt{r_\alpha}\|F_\alpha\|_\infty\max_s\|\hat{\beta}_s-\beta_s\|\overset{(i)}{\leq}O_P\left([n^{-1/2}+n^{1/2-\xi}]\log^{O(1)}n\right), \tag{B.16}$$

where (i) holds by Theorem 3.1 and Lemma B.1.

Combining (B.13) with (B.14), (B.15) and (B.16), we obtain $\max_t \|v_t'(\tilde{u}-u)F_\alpha\| = O_P([n^\xi + n + n^{3/2-\xi}]\log^{O(1)} n)$. Since $3/2-\xi < 1$ (as $\xi > 6/7$), part (2) follows.

Proof of part (3). Notice that $\max_t \|u_t'(\hat{v}_t - v_t)\| = \max_t \|u_t'(\hat{L}_Q \hat{F}_{Q,t} - L_Q F_{Q,t})\|$. We apply Lemma C.11(4) with $L = L_Q$, $F = F_Q$ and $e = v$ (as well as $a = u_t$ in Lemma C.11(4)). By Lemmas B.1 and B.3(3), we have $\|v\| = O_P(\sqrt{n}\log^{O(1)} n)$, $\max_t \|F_{Q,t}\| = O_P(\log^{O(1)} n)$ and $\max_t \|L_Q' v_t\| = O_P(n^{1/2}\log^{O(1)} n)$. Therefore, Lemma C.11(4) and $T \asymp n^\xi$ imply that

$$\begin{aligned}\max_t \|u_t'(\hat{L}_Q\hat{F}_{Q,t} - L_Q F_{Q,t})\| \leq\ & O_P(n^{-\xi}\log^{O(1)} n)\max_t \|F_Q' v' u_t\| \\ & + O_P\left([n^{-\xi/2} + n^{-1/2}]\log^{O(1)} n\right)\max_t \|L_Q' u_t\| \\ & + O_P\left([n^{1/2-\xi} + n^{-\xi/2}]\log^{O(1)} n\right)\max_t \|u_t\|.\end{aligned}$$

By Lemmas B.3(5) and (9), $\max_t \|L_Q' u_t\| = O_P(n^{1/2}\log^{O(1)} n)$ and $\max_t \|u_t' v F_Q\| = O_P(n\log^{O(1)} n)$. Since $\max_t \|u_t\| \leq \|u\| = O_P(n^{1/2}\log^{O(1)} n)$ (due to Lemma B.1(2)), we have

$$\max_t \|u_t'(\hat{L}_Q\hat{F}_{Q,t} - L_Q F_{Q,t})\| = O_P\left([1 + n^{1-\xi} + n^{(1-\xi)/2}]\log^{O(1)} n\right).$$

Since $1 + n^{1-\xi} \geq 2\sqrt{1\cdot n^{1-\xi}} = 2n^{(1-\xi)/2} > n^{(1-\xi)/2}$, we have $\max\{1,\ n^{1-\xi},\ n^{(1-\xi)/2}\} \leq 1 + n^{1-\xi}$ and thus part (3) follows. ☐

Lemma B.7. *Under Assumption 1,*
(1) $\max_t \|v_t'(\hat{\alpha}_t - \alpha_t)\| = O_P\left([n^{5(1-\xi)/2} + n^{(\xi-1)/2}]\log^{O(1)} n\right)$.
(2) $\max_{i,t} |\hat{\alpha}_{i,t} - \alpha_{i,t}| = O_P\left([n^{\xi/2-1} + n^{2-5\xi/2}]\log^{O(1)} n\right)$.

Proof. **Proof for part (1).** Notice that $\max_t \|v_t'(\hat{\alpha}_t - \alpha_t)\| = \max_t \|v_t'(\hat{L}_\alpha \hat{F}_{\alpha,t} - L_\alpha F_{\alpha,t})\|$. Recall that $y_{i,t} - x_{i,t}'\hat{\beta}_t = \alpha_{i,t} + \tilde{u}_{i,t}$, where $\tilde{u}_{i,t}$ is defined in Lemma B.6. We apply Lemma C.11(4) with $L = L_\alpha$, $F = F_\alpha$ and $e = \tilde{u}$ (as well as $a = v_t$ for Lemma C.11(4)). By Lemmas B.6(1) and B.1, we have $\|\tilde{u}\| = O_P([n^{\xi/2} + n^{1-\xi/2}]\log^{O(1)} n)$ and $\max_t \|F_{\alpha,t}\| = O_P(\log^{O(1)} n)$. It follows, by Lemma C.11(4), $T \asymp n^\xi$ and straight-forward computations, that

$$\begin{aligned}& \max_t \|v_t'(\hat{L}_\alpha\hat{F}_{\alpha,t} - L_\alpha F_{\alpha,t})\| \\ \leq\ & O_P\left(n^{-\xi}\log^{O(1)} n\right)\max_t \|F_\alpha' \tilde{u}' v_t\| \\ & + O_P(\log^{O(1)} n)\left[n^{-1}M + n^{-1/2} + n^{1/2-\xi} + n^{\xi/2-1} + n^{1-3\xi/2}\right]\max_t \|L_\alpha' v_t\| \\ & + O_P(\log^{O(1)} n)\left[n^{1-3\xi/2} + n^{\xi/2-1} + n^{2-5\xi/2} + (n^{-1} + n^{-\xi})M\right]\max_t \|v_t\|, \qquad \text{(B.17)}\end{aligned}$$

where $M = \max_t \|L_\alpha' \tilde{u}_t\|$. Since $\tilde{u}_t - u_t = X_t(\beta_t - \hat{\beta}_t)$, we have that

$$M \leq \max_t \|L_\alpha' u_t\| + \|L_\alpha\| \max_t \|X_t(\hat{\beta}_t - \beta_t)\| \overset{(i)}{\leq} O_P\left([n^{1/2} + n^{3/2-\xi}]\log^{O(1)} n\right), \qquad \text{(B.18)}$$

where (i) holds by $\|L_\alpha\| \leq \sqrt{nr_\alpha}\|L_\alpha\|_\infty$ and Lemmas B.3(6), B.1(1) and B.6(1). Notice that

$$\max_t \|F_\alpha' \tilde{u}' v_t\| \leq \max_t \|F_\alpha' u' v_t\| + \max_t \|F_\alpha'(\tilde{u}-u)' v_t\| \overset{(i)}{=} O_P\left([n+n^{\xi}] \log^{O(1)} n\right), \tag{B.19}$$

where (i) holds by Lemmas B.3(11) and B.6(2), together with $T \asymp n^{\xi}$ and $n^{(1+\xi)/2} \leq n + n^{\xi}$.

Now we combine (B.17) with (B.18), (B.19), $\max_t \|L_\alpha' v_t\| = O_P(\sqrt{n}\log^{O(1)} n)$ (Lemma B.3(4)) and $\max_t \|v_t\| \leq \|v\| = O_P(\sqrt{n}\log^{O(1)} n)$ (Lemma B.1(2)). After some tedious computations, this yields

$$\max_t \|v_t'(\hat{L}_\alpha \hat{F}_{\alpha,t} - L_\alpha F_{\alpha,t})\| = O_P\left([1 + a_n^{-1/2} + a_n + a_n^{3/2} + a_n^2 + a_n^{5/2}] \log^{O(1)} n\right),$$

where $a_n = n^{1-\xi}$. By Lemma C.9, $1 + a_n^{-1/2} + a_n + a_n^{3/2} + a_n^2 + a_n^{5/2} \leq 6(a_n^{-1/2} + a_n^{5/2})$ for $a_n > 0$. Thus, part (1) follows.

Proof for part (2). The argument is similar to the one in part (1). We apply Lemma C.11(4) with $L = L_\alpha$, $F = F_\alpha$ and $e = \tilde{u}$ (as well as $a = \tau_i$ in Lemma C.11(4)), where τ_i is the ith column of I_n. Recall, from the proof of part (1), that $\|\tilde{u}\| = O_P([n^{\xi/2} + n^{1-\xi/2}]\log^{O(1)} n)$ and $\max_t \|F_{\alpha,t}\| = O_P(\log^{O(1)} n)$. Notice that $\max_i \|\tau_i\| = 1$ and $\max_i \|L_\alpha' \tau_i\| = O_P(\log^{O(1)} n)$ (due to B.1). Thus, Lemma C.11(4) and $T \asymp n^{\xi}$ imply that

$$\begin{aligned}
\max_{i,t} |\hat{\alpha}_{i,t} - \alpha_{i,t}| &= \max_{i,t} |\tau_i'(\hat{L}_\alpha \hat{F}_t - L_\alpha F_t)| \\
&\leq O_P(n^{-\xi}\log^{O(1)} n) \max_i \|F_\alpha' \tilde{u}' \tau_i\| \\
&\quad + O_P(\log^{O(1)} n)\Big[n^{-1}M + n^{-1/2} + n^{1/2-\xi} + n^{\xi/2-1} + n^{1-3\xi/2}\Big] \\
&\quad + O_P(\log^{O(1)} n)\Big[n^{1-3\xi/2} + n^{\xi/2-1} + n^{2-5\xi/2} + (n^{-1} + n^{-\xi})M\Big].
\end{aligned} \tag{B.20}$$

Notice that

$$\begin{aligned}
\max_i \|\tau_i' \tilde{u} F_\alpha\| &\leq \max_i \|\tau_i' u F_\alpha\| + \max_i \|\tau_i'(\tilde{u} - u) F_\alpha\| \\
&\overset{(i)}{=} O_P(T^{1/2}\log^{O(1)} n) + \max_i \left\| \sum_{t=1}^{T} x_{i,t}'(\hat{\beta}_t - \beta_t) F_{\alpha,t}' \right\| \\
&\leq O_P(T^{1/2}\log^{O(1)} n) + T \max_{i,t} \|x_{i,t}\| \cdot \max_t \|\hat{\beta}_t - \beta_t\| \cdot \sqrt{r_\alpha}\|F_\alpha\|_\infty \\
&\overset{(ii)}{=} O_P\left([n^{\xi/2} + n^{1/2} + n^{\xi-1/2}]\log^{O(1)} n\right),
\end{aligned} \tag{B.21}$$

where (i) follows by Lemma B.3(2) and (ii) follows by $T \asymp n^{\xi}$, Lemma B.1(1) and Theorem 3.1. Hence, We combine (B.20) with (B.21) and (B.18). After straight-forward (but tedious) computations, this yields

$$\max_{i,t} |\hat{\alpha}_{i,t} - \alpha_{i,t}| = O_P\left(n^{-1/2}[1 + a_n^{-1/2} + a_n^{1/2} + a_n + a_n^{3/2} + a_n^2 + a_n^{5/2}] \log^{O(1)} n\right),$$

where $a_n = n^{1-\xi}$. By Lemma C.9, $1 + a_n^{-1/2} + a_n^{1/2} + a_n + a_n^{3/2} + a_n^2 + a_n^{5/2} \leq 7(a_n^{-1/2} + a_n^{5/2})$ for $a_n > 0$. Thus, part (2) follows. The proof is complete. □

Lemma B.8. *Let $\hat{\bar{v}}_{i,t}$, $\bar{v}_{i,t}$ and $D_{n,t}$ be defined in (B.1). Suppose that Assumption 1 holds. Then (1) $\max_t \|\hat{\Sigma}_t^{-1}\| = O_P(1)$ and $\max_{i,t} \|\hat{\bar{v}}_{i,t} - \bar{v}_{i,t}\| = O_P\left([n^{-1/2} + n^{1/2-\xi}] \log^{O(1)} n\right)$,*
(2) $\|D_n\|_\infty = \max_t \|D_{n,t}\|_\infty = O_P\left([n^{\xi/2-1} + n^{3-7\xi/2}] \log^{O(1)} n\right)$.

Proof. **Proof for part (1).** By Lemma B.5,

$$\max_t \|\hat{\Sigma}_t^{-1}\| \leq \max_t \|\Sigma_t\| + \max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\| = O(1) + O_P\left(\left[n^{-1/2} + n^{1/2-\xi}\right] \log^{O(1)} n\right) \overset{(i)}{=} O_P(1), \tag{B.22}$$

where (i) holds by $\xi > 1/2$. Notice that

$$\begin{aligned}
\max_{i,t} \|\hat{\Sigma}_t^{-1}\hat{v}_{i,t} - \Sigma_t^{-1}v_{i,t}\| &\leq \max_{i,t} \|\hat{\Sigma}_t^{-1}(\hat{v}_{i,t} - v_{i,t})\| + \max_{i,t} \|(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1})v_{i,t}\| \\
&\leq \max_t \|\hat{\Sigma}_t^{-1}\| \max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\| + \max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\| \max_{i,t} \|v_{i,t}\| \\
&\overset{(i)}{=} O_P\left([n^{-1/2} + n^{1/2-\xi}] \log^{O(1)} n\right),
\end{aligned}$$

where (i) holds by the bounds for $\max_{i,t} \|\hat{v}_{i,t} - v_{i,t}\|$ and for $\max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\|$ (Lemmas B.4 and B.5), together with $\max_{i,t} \|v_{i,t}\| \leq \sqrt{k}\|v\|_\infty = O_P(\log^{O(1)} n)$ (Lemma B.1). This proves part (1).

Proof for part (2). By the definition of $D_{n,t}$ in (B.1), we have the following decomposition

$$D_{n,t} = n^{-1/2}\left(\hat{\Sigma}_t^{-1}\hat{v}_t' - \Sigma_t^{-1}v_t'\right) u_t + n^{-1/2}\hat{\Sigma}_t^{-1}v_t'(\alpha_t - \hat{\alpha}_t) + n^{-1/2}\hat{\Sigma}_t^{-1}(\hat{v}_t - v_t)'(\alpha_t - \hat{\alpha}_t). \tag{B.23}$$

Now we derive bounds for each of these three terms. Let $a_n = n^{1-\xi}$. For the first term, notice that

$$\begin{aligned}
&\max_t \left\|n^{-1/2}\left(\hat{\Sigma}_t^{-1}\hat{v}_t' - \Sigma_t^{-1}v_t'\right) u_t\right\| \\
= \quad &\max_t \left\|n^{-1/2}\hat{\Sigma}_t^{-1}(\hat{v}_t - v_t)'u_t\right\| + \max_t \left\|n^{-1/2}\left(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\right) v_t'u_t\right\| \\
\leq \quad &\max_t \left\|n^{-1/2}\hat{\Sigma}_t^{-1}\right\| \max_t \|(\hat{v}_t - v_t)'u_t\| + \max_t \left\|n^{-1/2}\left(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\right) v_t'u_t\right\| \\
\overset{(i)}{=} \quad &O_P\left(n^{-1/2}[1 + a_n] \log^{O(1)} n\right) + \max_t \left\|n^{-1/2}\left(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\right) v_t'u_t\right\|,
\end{aligned} \tag{B.24}$$

where (i) follows by (B.22) and Lemma B.6(3). Also notice that

$$\max_t \left\|n^{-1/2}\left(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\right) v_t'u_t\right\| \leq n^{-1/2} \max_t \left\|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\right\| \max_t \|v_t'u_t\|$$

$$\overset{(i)}{=} O_P\left(n^{-1/2}[1+a_n]\log^{O(1)} n\right), \tag{B.25}$$

where (i) follows by Lemmas B.3(7) and B.5. We combine (B.24) and (B.25) and obtain that

$$\max_t \left\| n^{-1/2}\left(\hat{\Sigma}_t^{-1}\hat{v}_t' - \Sigma_t^{-1}v_t'\right)u_t\right\| = O_P\left(n^{-1/2}[1+a_n]\log^{O(1)} n\right). \tag{B.26}$$

To bound the second term in (B.23), observe that

$$\begin{aligned}\max_t \|n^{-1/2}\hat{\Sigma}_t^{-1}v_t'(\alpha_t-\hat{\alpha}_t)\| &\le n^{-1/2}\max_t \|\hat{\Sigma}_t^{-1}\|\max_t \|v_t'(\hat{\alpha}_t-\alpha_t)\| \\ &\overset{(i)}{=} O_P\left([n^{2-5\xi/2}+n^{\xi/2-1}]\log^{O(1)} n\right) = O_P\left(n^{-1/2}[a_n^{5/2}+a_n^{-1/2}]\log^{O(1)} n\right),\end{aligned} \tag{B.27}$$

where (i) holds by (B.22) and Lemma B.7. To bound the third term in (B.23), we have that

$$\begin{aligned}&\max_t \|n^{-1/2}\hat{\Sigma}_t^{-1}(\hat{v}_t-v_t)'(\alpha_t-\hat{\alpha}_t)\| \\ &\le n^{-1/2}\max_t \|\hat{\Sigma}_t^{-1}\|\max_t \|\hat{v}_t-v_t\|\max_t \|\hat{\alpha}_t-\alpha_t\| \\ &\le n^{-1/2}\max_t \|\hat{\Sigma}_t^{-1}\| n^{1/2}\max_{i,t}\|\hat{v}_{i,t}-v_{i,t}\| n^{1/2}\max_{i,t}|\hat{\alpha}_{i,t}-\alpha_{i,t}| \\ &\overset{(i)}{=} O_P\left(n^{-1/2}[a_n^{-1/2}+a_n^{1/2}+a_n^{5/2}+a_n^{7/2}]\log^{O(1)} n\right),\end{aligned} \tag{B.28}$$

where (i) holds by (B.22) and Lemmas B.4 and B.7(2), together with $a_n = n^{1-\xi}$.

Now we combine (B.23) with (B.26), (B.27) and (B.28) and obtain

$$\max_t \|D_{n,t}\|_\infty = O_P\left(n^{-1/2}[1+a_n^{-1/2}+a_n+a_n^{1/2}+a_n^{5/2}+a_n^{7/2}]\log^{O(1)} n\right).$$

By Lemma C.9, $1+a_n^{-1/2}+a_n+a_n^{1/2}+a_n^{5/2}+a_n^{7/2} \le 6(a_n^{-1/2}+a_n^{7/2})$. Part (2) follows. □

Lemma B.9. *Recall $\hat{u}_{i,t}$ defined in Algorithm 2 and $\hat{\bar{v}}_{i,t}$ and $\bar{v}_{i,t}$ defined in (B.1). Under Assumption 1,* $\max_{1\le i\le n,1\le t\le T}\|\hat{\bar{v}}_{i,t}\hat{u}_{i,t}-\bar{v}_{i,t}u_{i,t}\| = O_P\left([n^{2-5\xi/2}+n^{\xi/2-1}]\log^{O(1)} n\right)$.

Proof. Let $a_n = n^{1-\xi}$. Notice that

$$\begin{aligned}\max_{i,t}|\hat{u}_{i,t}-u_{i,t}| = \max_{i,t}|y_{i,t}-x_{i,t}'\hat{\beta}_t-\hat{\alpha}_{i,t}-u_{i,t}| &\overset{(i)}{=} \max_{i,t}|\alpha_{i,t}+\tilde{u}_{i,t}-\hat{\alpha}_{i,t}-u_{i,t}| \\ &\le \max_{i,t}|\tilde{u}_{i,t}-u_{i,t}| + \max_{i,t}|\hat{\alpha}_{i,t}-\alpha_{i,t}| \\ &\le \max_{i,t}\|x_{i,t}\|\max\|\hat{\beta}_t-\beta_t\| + \max_{i,t}|\hat{\alpha}_{i,t}-\alpha_{i,t}| \\ &\overset{(ii)}{=} O_P\left(n^{-1/2}[1+a_n+a_n^{-1/2}+a_n^{5/2}]\log^{O(1)} n\right) \\ &\overset{(iii)}{=} O_P\left(n^{-1/2}[a_n^{-1/2}+a_n^{5/2}]\log^{O(1)} n\right),\end{aligned} \tag{B.29}$$

where (i) holds by $y_{i,t} = \alpha_{i,t}+x_{i,t}'\beta_t+u_{i,t}$ and the definition of $\tilde{u}_{i,t}$ in Lemma B.6, (ii) holds by

Lemma B.1, Theorem 3.1 and Lemma B.7(2) and (iii) holds by noticing that $1+a_n+a_n^{-1/2}+a_n^{5/2} \leq 4(a_n^{-1/2}+a_n^{5/2})$ (due to Lemma C.9). Notice that

$$\begin{aligned}
& \max_{i,t} \|\hat{\bar{v}}_{i,t}\hat{u}_{i,t} - \bar{v}_{i,t}u_{i,t}\| \\
\leq\ & \max_{i,t} \|\hat{\bar{v}}_{i,t} - \bar{v}_{i,t}\| \max_{i,t} |\hat{u}_{i,t}| + \max_{i,t} \|\bar{v}_{i,t}\| \max_{i,t} |\hat{u}_{i,t} - u_{i,t}| \\
\leq\ & \max_{i,t} \|\hat{\bar{v}}_{i,t} - \bar{v}_{i,t}\| \left(\|u\|_\infty + \max_{i,t} |\hat{u}_{i,t} - u_{i,t}| \right) + \max_{i,t} \|\bar{v}_{i,t}\| \max_{i,t} |\hat{u}_{i,t} - u_{i,t}| \\
\overset{(i)}{=}\ & O_P\left(\left\{ n^{-1/2}[1+a_n+a_n^{5/2}+a_n^{-1/2}] + n^{-1}[a_n^{-1/2}+a_n^{1/2}+a_n^{5/2}+a_n^{7/2}] \right\} \log^{O(1)} n \right) \\
\overset{(ii)}{=}\ & O_P\left(\left\{ n^{-1/2}[a_n^{5/2}+a_n^{-1/2}] + n^{-1}[a_n^{-1/2}+a_n^{7/2}] \right\} \log^{O(1)} n \right),
\end{aligned}$$

where (i) follows by (B.29) and Lemmas B.8(1) and B.1 and (ii) follows by $1+a_n+a_n^{5/2}+a_n^{-1/2} \leq 4(a_n^{-1/2}+a_n^{5/2})$ and $a_n^{-1/2}+a_n^{1/2}+a_n^{5/2}+a_n^{7/2} \leq 4(a_n^{-1/2}+a_n^{7/2})$ (due to Lemma C.9). Plugging in $a_n = n^{1-\xi}$, we obtain

$$\begin{aligned}
\max_{i,t} \|\hat{\bar{v}}_{i,t}\hat{u}_{i,t} - \bar{v}_{i,t}u_{i,t}\| &= O_P\left([n^{2-5\xi/2} + n^{\xi/2-1} + n^{\xi/2-3/2} + n^{5/2-7\xi/2}] \log^{O(1)} n \right) \\
&\overset{(i)}{=} O_P\left([n^{2-5\xi/2} + n^{\xi/2-1}] \log^{O(1)} n \right),
\end{aligned}$$

where (i) holds by $n^{\xi/2-3/2} = o(n^{\xi/2-1/2})$ and $n^{5/2-7\xi/2} = o(n^{2-5\xi/2})$ (since $\xi \in (6/7, 2)$). $\square$

Lemma B.10. *Recall Ω and $\hat{\Omega}$ defined in (B.1). Let Assumptions 1 and 2 hold. Then*

$$\sup_{x \in \mathbb{R}} \left| \Phi(x, \hat{\Omega}) - \Phi(x, \Omega) \right| = o_P(1).$$

Proof. **Step 1: derive the exponential-type tails for $(J_j'G_i)(J_k'G_i)$.** By Assumption 1, there is a constant $M > 0$ such that $\forall t \in [T]$, each row of Σ_t^{-1} is bounded (in $\|\cdot\|_1$) by M. It follows, by Lemma C.3(1) and Assumption 1, that there exist constants $b > 0$ depending only on M, k and γ_* such that $\forall (i,t) \in [n] \times [T]$, each entry of $\bar{v}_{i,t}$ has an exponential-type tail with parameter (b, γ_*). By Lemma C.3(3), there exists $b_G > 0$ such that $\forall (i,t) \in [n] \times [T]$, each entry of $G_{i,t} = \bar{v}_{i,t}u_{i,t}$ has an exponential-type tails with parameters $(b_G, \gamma_*/2)$.

By Assumption 2 and Lemma C.3(1), $J_j'G_i$ has an exponential-type tail with parameter $(c_n, \gamma_*/2)$, where $c_n = b_G A_1 \log^{2/\gamma_*}(m_J+2)$. Then Lemma C.3(3) implies that for $j, k \in [m_J]$, $(J_j'G_i)(J_k'G_i)$ has an exponential-type tail with parameter $(C_n, \gamma_*/4)$, where $C_n = 2^{4/\gamma_*}c_n^2$. Hence, $(J_j'G_i)(J_k'G_i)C_n^{-1}$ has an exponential-type tail with parameter $(1, \gamma_*/4)$.

Step 2: show the desired result by bounding $\|\hat{\Omega} - \Omega\|_\infty$. Since $\{(J_j'G_i)(J_k'G_i)\}_{i=1}^n$ is independent across i, it follows, by Lemma C.6, that

$$\max_{1 \leq j,k \leq m_J} \left| n^{-1} \sum_{i=1}^n \left[(J_j'G_i)(J_k'G_i) - \mathbb{E}(J_j'G_i)(J_k'G_i) \right] / C_n \right| = O_P\left(\sqrt{n^{-1}\log(m_J^2)} \right).$$

Let $\tilde{\Omega} = n^{-1}\sum_{i=1}^{n} G_i G_i'$. The above display implies that

$$\|\tilde{\Omega} - \Omega\|_\infty = \max_{1\le j,k\le m_J} \left| n^{-1}\sum_{i=1}^{n} \left[(J_j' G_i)(J_k' G_i) - \mathbb{E}(J_j' G_i)(J_k' G_i)\right]\right|$$
$$= C_n O_P\left(\sqrt{n^{-1}\log(m_J^2)}\right) \overset{(i)}{=} O_P\left(n^{-1/2}\log^{O(1)} n\right), \quad \text{(B.30)}$$

where (i) holds by the definition of C_n. Notice that

$$\begin{aligned}
& \left\| n^{-1}\sum_{i=1}^{n} (\hat{G}_i \hat{G}_i - G_i G_i') \right\|_\infty \\
\le\ & \max_i \|\hat{G}_i \hat{G}_i' - G_i G_i'\|_\infty \\
\le\ & \max_i \left(\|\hat{G}_i\|_\infty \|\hat{G}_i - G_i\|_\infty + \|G_i\|_\infty \|\hat{G}_i - G_i\|_\infty\right) \\
\overset{(i)}{\le}\ & \max_i \left(2\|G_i\|_\infty \|\hat{G}_i - G_i\|_\infty + \|\hat{G}_i - G_i\|_\infty^2\right) \\
\le\ & \left(2\max_{i,t}\|\bar{v}_{i,t}u_{i,t}\| \max_{i,t}\|\hat{\bar{v}}_{i,t}\hat{u}_{i,t} - \bar{v}_{i,t}u_{i,t}\|_\infty + \max_{i,t}\|\hat{\bar{v}}_{i,t}\hat{u}_{i,t} - \bar{v}_{i,t}u_{i,t}\|_\infty^2\right) \\
\overset{(ii)}{\le}\ & O_P\left([n^{2-5\xi/2} + n^{\xi/2-1}]\log^{O(1)} n\right), \qquad \text{(B.31)}
\end{aligned}$$

where (i) holds by $\|\hat{G}_i\|_\infty \le \|G_i\|_\infty + \|\hat{G}_i - G_i\|_\infty$ and (ii) follows by Lemmas B.1(1) and B.9. Therefore,

$$\|\hat{\Omega} - \tilde{\Omega}\|_\infty = \max_{1\le j,k\le m_J} \left| J_j' \left[n^{-1}\sum_{i=1}^{n} (\hat{G}_i \hat{G}_i - G_i G_i')\right] J_k \right|$$
$$\overset{(i)}{\le} \max_{1\le j\le m_J} \|J_j\|_1^2 \left\| n^{-1}\sum_{i=1}^{n} (\hat{G}_i \hat{G}_i - G_i G_i')\right\|_\infty \overset{(ii)}{=} O_P\left([n^{2-5\xi/2} + n^{\xi/2-1}]\log^{O(1)} n\right), \quad \text{(B.32)}$$

where (i) follows by Holder's inequality and (ii) follows by $\max_{1\le j\le m_J}\|J_j\|_1^2 \le A_1$ and (B.31). We combine (B.30) and (B.32) and obtain

$$\|\hat{\Omega} - \Omega\|_\infty = O_P\left([n^{2-5\xi/2} + n^{\xi/2-1} + n^{-1/2}]\log^{O(1)} n\right). \qquad \text{(B.33)}$$

By Assumption 2, the diagonal entries of Ω are bounded away from zero and infinity. Therefore, Lemma C.5 implies that

$$\sup_{x\in\mathbb{R}} \left|\Phi(x, \hat{\Omega}) - \Phi(x, \Omega)\right| \le M\Delta^{1/3}\left(1 \vee \log(2m_J/\Delta)\right)^{2/3}, \qquad \text{(B.34)}$$

where $M > 0$ is a constant and $\Delta = \|\hat{\Omega} - \Omega\|_\infty$. The desired result follows by (B.34), together with (B.33) and $T \asymp n^\xi$ with $\xi \in (6/7, 2)$. □

Lemma B.11. *Recall Ω defined in (B.1). Let Assumptions 1 and 2 hold. Then*

$$\limsup_{n\to\infty}\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|n^{-1/2}\sum_{i=1}^{n}JG_i\right\|_\infty \le x\right)-\Phi(x,\Omega)\right| = 0.$$

Proof. For $j\in[m_J]$ and $i\in[n]$, define $W_{i,j}=J_j'G_i$ and denote $W_i=JG_i=(W_{i,1},\ldots,W_{i,m_J})'\in\mathbb{R}^{m_J}$. By Assumption 2, $\min_{1\le j\le m_R} n^{-1}\sum_{i=1}^n \mathbb{E}W_{i,j}^2\ge c_1$.

As argued at the beginning of the proof of Lemma B.10, $W_{i,j}$ has an exponential-type tail with parameter (d_n,γ_1), where $d_n=c_0A_1\log^{1/\gamma_1}(m_J+2)$, $\gamma_1=\gamma_*/2$ and $c_0>0$ is a constant. Define $B_n=C_1n^{l/q}d_n$, where $q=4(l+1)$ and $C_1>0$ is a constant to be chosen later. Then by Lemma C.3(2), we have

$$\begin{cases} n^{-1}\sum_{i=1}^n\mathbb{E}|W_{i,j}|^3/B_n\le C_{\gamma_1,3}d_n^3B_n^{-1}=O\left(n^{-l/(4l+4)}\log^{O(1)}n\right)=o(1)\\ n^{-1}\sum_{i=1}^n\mathbb{E}W_{i,j}^4/B_n^2\le C_{\gamma_1,4}d_n^4B_n^{-2}=O\left(n^{-l/(2l+2)}\log^{O(1)}n\right)=o(1)\\ \max_{1\le i\le n}\mathbb{E}\max_{1\le j\le m_J}|W_{i,j}/B_n|^q\le C_{\gamma_1,q}m_Jd_n^qB_n^{-q}=O(1),\end{cases}$$

where $C_{\gamma_1,3}$, $C_{\gamma_1,4}$ and $C_{\gamma_1,q}$ are constants depending only on γ_1 and q. Therefore, we can choose a constant $C_1>0$ such that

$$\begin{cases} n^{-1}\sum_{i=1}^n\mathbb{E}|W_{i,j}|^3\le B_n & \forall 1\le j\le m_J\\ n^{-1}\sum_{i=1}^n\mathbb{E}W_{i,j}^4\le B_n^2 & \forall 1\le j\le m_J\\ \mathbb{E}\max_{1\le j\le m_J}|W_{i,j}/B_n|^q\le 2 & \forall 1\le i\le n.\end{cases}$$

Notice that $\{z\in\mathbb{R}^{m_J}\mid \|z\|_\infty\le x\}$ is a rectangle in $\mathbb{R}^{m_J}$. It follows, by Proposition 2.1 of Chernozhukov et al. (2014) (applied to $\{W_i\}_{i=1}^n$), that there exists a constant $C>0$ depending only on c_1 and q such that

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|n^{-1/2}\sum_{i=1}^{n}JG_i\right\|_\infty\le x\right)-\Phi(x,\Omega)\right|\\ \le\;& C\left\{\left(n^{-1}B_n^2\log^7(m_Jn)\right)^{1/6}+\left(n^{2/q-1}B_n^2\log^3(m_Jn)\right)^{1/3}\right\}\\ \overset{(i)}{=}\;& O\left(\left[n^{-\frac{l+2}{12(l+1)}}+n^{-\frac{1}{6(l+1)}}\right]\log^{O(1)}n\right)=o(1),\end{aligned}$$

where (i) holds by the definition of B_n. The proof is complete. □

Proof of Theorem 3.3. Since $\tilde\beta_t=\hat\beta_t-(\hat v_t'\hat v_t)^{-1}\hat v_t'(\alpha_t+u_t)$, we have that

$$\tilde\beta_t-\beta_t=\hat\beta_t-\beta_t-(\hat v_t'\hat v_t)^{-1}\hat v_t'\hat\alpha_t\overset{(i)}{=}(\hat v_t'\hat v_t)^{-1}\hat v_t'(\alpha_t-\hat\alpha_t)+(\hat v_t'\hat v_t)^{-1}\hat v_t'u_t\overset{(ii)}{=}n^{-1/2}D_{n,t}+n^{-1}\sum_{i=1}^{n}G_{i,t},$$

where (i) follows by (B.9) in the proof of Theorem 3.1 and (ii) follows by the definition of $D_{n,t}$ and $G_{i,t}$ in (B.1). For the rest of the proof, recall G_i, D_n, Ω and $\hat{\Omega}$ defined in (B.1). The above display means that

$$\sqrt{n}J(\tilde{\beta}-\beta) = JD_n + S_n^{JG}, \tag{B.35}$$

where $S_n^{JG} = n^{-1/2}\sum_{i=1}^n JG_i$. Define

$$\varepsilon = n^{-\kappa_*/2} \qquad \text{with} \qquad \kappa_* = \min\left\{1-\frac{\xi}{2},\ \frac{7\xi}{2}-3\right\}. \tag{B.36}$$

Notice that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde{\beta}-\beta)\right\|_\infty \le x\right) - \Phi(x,\Omega)\right| \\
\le &\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde{\beta}-\beta)\right\|_\infty \le x\right) - \mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x\right)\right| + \sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x\right) - \Phi(x,\Omega)\right| \\
\overset{(i)}{\le} &\mathbb{P}\left(\|JD_n\|_\infty > \varepsilon\right) + \sup_{x\in\mathbb{R}}\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \in (x-\varepsilon, x+\varepsilon]\right) + \sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x\right) - \Phi(x,\Omega)\right|,
\end{aligned} \tag{B.37}$$

where (i) follows by (B.35) and Lemma C.1. Also notice that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \in (x-\varepsilon, x+\varepsilon]\right) - \left[\Phi(x+\varepsilon,\Omega) - \Phi(x-\varepsilon,\Omega)\right]\right| \\
= &\left|\left[\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x+\varepsilon\right) - \Phi(x+\varepsilon,\Omega)\right] - \left[\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x-\varepsilon\right) - \Phi(x-\varepsilon,\Omega)\right]\right| \\
\le &2\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le t\right) - \Phi(t,\Omega)\right|.
\end{aligned} \tag{B.38}$$

Therefore, we have

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde{\beta}-\beta)\right\|_\infty \le x\right) - \Phi(x,\Omega)\right| \\
\overset{(i)}{\le} &\mathbb{P}\left(\|JD_n\|_\infty > \varepsilon\right) + \sup_{x\in\mathbb{R}}\left[\Phi(x+\varepsilon,\Omega) - \Phi(x-\varepsilon,\Omega)\right] + 3\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|S_n^{JG}\right\|_\infty \le x\right) - \Phi(x,\Omega)\right| \\
\overset{(ii)}{=} &\mathbb{P}\left(\|JD_n\|_\infty > \varepsilon\right) + \sup_{x\in\mathbb{R}}\left[\Phi(x+\varepsilon,\Omega) - \Phi(x-\varepsilon,\Omega)\right] + o(1) \\
\overset{(iii)}{\le} &\mathbb{P}\left(\|JD_n\|_\infty > \varepsilon\right) + C_0\varepsilon\sqrt{\log m_J} + o(1) \\
\overset{(iv)}{\le} &\mathbb{P}(A_1\|D_n\|_\infty > \varepsilon) + C_0\varepsilon\sqrt{\log m_J} + o(1),
\end{aligned} \tag{B.39}$$

where $C_0 > 0$ is a constant depending only on the constants in Assumption 2; in the above display, (i) follows by (B.37) and (B.38), (ii) follows by Lemma B.11, (iii) holds by Lemma C.4 and (iv) follows by Holder's inequality $\|JD_n\|_\infty \le \max_{1\le j\le m_J}\|J_j\|_1\|D_n\|_\infty$ and Assumption 2.

By Lemma B.8 and $T \asymp n^\xi$ with $\xi \in (6/7, 2)$ (Assumption 1), we have $\|D_n\|_\infty = O_P(n^{-\kappa_*})$ and

$\kappa_* > 0$. Therefore, (B.39) and (B.36) imply that

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde\beta-\beta)\right\|_\infty \le x\right) - \Phi(x,\Omega)\right| \\ \le \mathbb{P}\left(A_1 O_P(n^{-\kappa_*}) > n^{-\kappa_*/2}\right) + C_0 n^{-\kappa_*/2}\sqrt{\log m_J} + o(1) = o(1).$$

By Lemma B.10, $\sup_{x\in\mathbb{R}}|\Phi(x,\hat\Omega)-\Phi(x,\Omega)| = o_P(1)$. Hence, the above display implies that

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde\beta-\beta)\right\|_\infty \le x\right) - \Phi(x,\hat\Omega)\right| = o_P(1). \tag{B.40}$$

Fix an arbitrary constant $\delta > 0$. Then Lemma C.2 and (B.40) imply that

$$\limsup_{n\to\infty}\sup_{\eta\in(0,1)}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde\beta-\beta)\right\|_\infty > \Phi^{-1}(1-\eta,\hat\Omega)\right) - \eta\right| \\ \le \delta + \limsup_{n\to\infty}\mathbb{P}\left\{\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left\|\sqrt{n}J(\tilde\beta-\beta)\right\|_\infty \le x\right) - \Phi(x,\hat\Omega)\right| > \delta\right\} = \delta.$$

Since $\delta > 0$ is arbitrary, the desired result follows. □

Proof of Corollary 3.1. Notice that $J\beta \in \mathcal{C}_{1-\eta}(J)$ if and only if $\sqrt{n}\|J(\tilde\beta-\beta)\|_\infty \le \Phi^{-1}(1-\eta,\hat\Omega)$. It follows, by Theorem 3.3, that

$$P\left(\sqrt{n}\|J(\tilde\beta-\beta)\|_\infty \le \Phi^{-1}(1-\eta,\hat\Omega)\right) = 1 - P\left(\sqrt{n}\|J(\tilde\beta-\beta)\|_\infty > \Phi^{-1}(1-\eta,\hat\Omega)\right) \to 1-\eta.$$

The proof is complete. □

Proof of Theorem 3.4. If $l = 0$, the result clearly holds since $\Phi^{-1}(1-\eta,\hat\Omega) = O_P(1)$ for $m_J = O(1)$. Now we assume $l > 1$ and thus $m_J \to \infty$. Let $\zeta \sim N(0,\Omega)$ with $\zeta = (\zeta_1,\ldots,\zeta_{m_J})' \in \mathbb{R}^{m_J}$. At the beginning of the proof of Lemma B.10, we have showed that the entries of $G_{i,t}$ have exponential-type tails. Thus, $\|\Sigma_G\|_\infty \le K_1$ for some constant K_1, where $\Sigma_G = n^{-1}\sum_{i=1}^n \mathbb{E}G_iG_i'$. Since $\Omega_{j,j} = J_j'\Sigma_G J_j$ and $\|J_j\|_1 \le A_1$, Holder's inequality implies that $\Omega_{j,j} \le \|J_j\|_1^2\|\Sigma_G\|_\infty \le A_1^2 K_1$.

Since $\zeta_j \sim N(0,\Omega_{j,j})$, there exists a constant $K_2 > 0$ such that ζ_j has an exponential-type tail with parameter $(K_2, 2)$. Then,

$$\Phi\left(2K_2\sqrt{\log m_J},\Omega\right) = 1 - \mathbb{P}\left(\|\zeta\|_\infty > 2K_2\sqrt{\log m_J}\right) \ge 1 - \sum_{j=1}^{m_J}\mathbb{P}\left(|\zeta_j| > 2K_2\sqrt{\log m_J}\right) \\ \overset{(i)}{\ge} 1 - m_J\exp\left[1-(2K_2\sqrt{\log m_J}/K_2)^2\right] = 1 - e/m_J \to 1.$$

where (i) holds by the exponential-type tails of ζ_j. By Lemma B.10, $\Phi\left(2K_2\sqrt{\log m_J},\hat\Omega\right) = 1+o_P(1)$ and thus $2K_2\sqrt{\log m_J} = \Phi^{-1}(1+o_P(1),\hat\Omega)$. Since $\Phi^{-1}(1+o_P(1),\hat\Omega) \ge \Phi^{-1}(1-\eta,\hat\Omega)$ with probability approaching one, we have that

$$\mathbb{P}\left(\Phi^{-1}(1-\eta,\hat{\Omega}) \leq 2K_2\sqrt{\log m_J}\right) \to 1.$$

Since $m_J = O(n^l)$ (which means that $\log m_J = O(\log n)$, the desired result follows. ☐

Proof of Theorem 3.5. This is Lemma B.7(2). ☐

Proof of Theorem 3.2. The result follows by combining Theorems 3.3 and 3.4 using $J = I_{kT}$. ☐

B.3 Proof of Theorems 3.6, 3.7 and 4.1

The following result is useful for proving Theorem 3.6.

Lemma B.12. *Consider matrices $W, e \in \mathbb{R}^{n_1 \times n_2}$. Suppose that $s_{r+1}(W) = 0$ for some $r \geq 1$. Let $\mu > 0$ and define $\hat{r} = \max\{j \mid s_j(W+e) \geq \mu\}$. Then $\hat{r} \neq r$ implies that either $s_r(W) < 2\mu$ or $s_1(e) \geq \mu$.*

Proof. We proceed by contradiction. Suppose that $\hat{r} \neq r$, $s_r(W) \geq 2\mu$ and $s_1(e) < \mu$. We discuss two cases separately: case (A) with $\hat{r} > r$ and case (B) with $\hat{r} < r$.

We first consider case (A). By definition, $s_{\hat{r}}(W+e) \geq \mu$. Since $\hat{r} > r$, we have $\hat{r} \geq r+1$ and thus $s_{r+1}(W+e) \geq \mu$. By Lemma C.10(1), we have $s_{r+1}(W) + s_1(e) \geq s_{r+1}(W+e)$ and thus $s_{r+1}(W) + s_1(e) \geq \mu$. Therefore, the assumption of $s_{r+1}(W) = 0$ implies that $s_1(e) \geq \mu$, contradicting $s_1(e) < \mu$.

We now consider case (B). By definition $s_{\hat{r}+1}(W+e) < \mu$. Since $\hat{r} < r$, we have $\hat{r}+1 \leq r$ and thus $s_r(W+e) < \mu$. Hence, Lemma C.10(1) implies that $s_r(W) \leq s_r(W+e) + s_1(-e) < \mu + s_1(e)$. Since $s_1(e) < \mu$, we have that $s_r(W) < 2\mu$, contradicting $s_r(W) \geq 2\mu$.

Therefore, it is impossible that these three conditions hold simultaneously: $\hat{r} \neq r$, $s_r(W) \geq 2\mu$ and $s_1(e) < \mu$. Hence, $\hat{r} \neq r$ implies that at least one of the other two conditions does not hold. ☐

Proof of Theorem 3.6. **Proof of part (1).** Since $X = L_Q F_Q' + v$, Lemma B.12 implies that it suffices to verify

(1a) $\mathbb{P}[s_{r_Q}(L_Q F_Q') < 2\mu_n] \to 0$;

(1b) $\mathbb{P}[\|v\| \geq \mu_n] \to 0$.

Notice that (1b) follows by Lemma B.1 and $\sqrt{n}\log^{O(1)} n = o(\mu_n)$. By Assumption 1, both $s_1(F_Q/\sqrt{T})$ and $s_{r_Q}(L_Q/\sqrt{n})$ are bounded away from zero. It follows, by Lemma C.10(2), that there exists a constant $b_0 > 0$ such that $\mathbb{P}[s_{r_Q}(L_Q F_Q'/\sqrt{nT}) > b_0] \to 1$. Since $\sqrt{nT}/\mu_n \to \infty$, condition (1a) follows. We have proved part (1).

Proof of part (2). Recall $\tilde{u}_t = u_t + X_t(\beta_t - \hat{\beta}_t)$ (defined in Lemma B.6). Then $y_t - X_t\hat{\beta}_t = \alpha_t + \tilde{u}_t$. By Lemma B.12, it suffices to verify

(2a) $\mathbb{P}[s_{r_\alpha}(L_\alpha F_\alpha') < 2\tilde{\mu}_n] \to 0$;

(2b) $\mathbb{P}(\|\tilde{u}\| \geq \tilde{\mu}_n) \to 0$.

The same argument as in showing (1a) (with (L_Q, F_Q, r_Q) replaced by $(L_\alpha, F_\alpha, r_\alpha)$) proves (2a). Lemma B.6(1) and $T \asymp n^\xi$imply that $\|\tilde{u}\| = O_P([\sqrt{T} + n/\sqrt{T}]\log^{O(1)} n)$. Since $[\sqrt{T} + n/\sqrt{T}]\log^{O(1)} n = o(\tilde{\mu}_n)$, condition (2b) follows. We have proved part (2). □

The following results are useful for proving Theorem 3.7.

Lemma B.13. *Let Assumption 1 hold. Then for any fixed positive integer r, there exists a constant $c > 0$ such that $\mathbb{P}(s_r(u) > \sqrt{n}c) \to 1$. Moreover, $\mathbb{P}(s_r(\tilde{u}) > \sqrt{n}c/2) \to 1$, where $\tilde{u}$ is defined in Lemma B.6.*

Proof. Notice that there exist a constant $c_1 > 0$ and a sequence $\{t_j\}_{j=1}^r \subset [T]$ such that $t_1 < t_2 < \ldots < t_r$ and $d_n \geq c_1 T$, where $d_n = \min_{1\leq j\leq r-1}(t_{j+1} - t_j)$. Let $A \in \mathbb{R}^{n\times r}$ be defined as $A_{i,j} = u_{i,t_j}$ for $(i,j) \in [n]\times[r]$. Since A is a matrix consisting of r columns of u, Lemma C.10(4) implies

$$s_r(u) \geq s_r(A). \tag{B.41}$$

Let $\Sigma_A = n^{-1}\mathbb{E}A'A \in \mathbb{R}^{r\times r}$. We only need to show the lower bound for $s_r(A)$. We proceed in two steps. First, we show that singular values of Σ_A are bounded below; then we show the desired result. Finally, we shall show the result for $s_r(\tilde{u})$.

Step 1: derive lower bound for $s_r(\Sigma_A)$. Fix arbitrary $j_1, j_2 \in [r]$ with $j_1 \neq j_2$. Notice that $\Sigma_{A,j_1,j_2} = n^{-1}\sum_{i=1}^n \mathbb{E}u_{i,t_{j_1}}u_{i,t_{j_2}}$. By Lemma C.3(2) and the exponential-type tails of $u_{i,t}$'s, there exists a constant $c_1 > 0$ such that $\mathbb{E}|u_{i,t}|^4 \leq c_1$. It follows, by Corollary 16.2.4 of Athreya and Lahiri (2006), that

$$\begin{aligned}|\Sigma_{A,j_1,j_2}| \leq \max_i |\mathbb{E}(u_{i,t_{j_1}}u_{i,t_{j_2}})| &\leq 4c_1^2\sqrt{2\alpha_{mixing}(|t_{j_1} - t_{j_2}|)} \\ &\leq 4c_1^2\sqrt{c_*}\exp[-|t_{j_1} - t_{j_2}|^{\gamma_{**}}/2] \leq 4c_1^2\sqrt{c_*}\exp[-d_n^{\gamma_{**}}/2] = o(1).\end{aligned}$$

Let $\tilde{\Sigma}_A \in \mathbb{R}^{r\times r}$ be the diagonal matrix such that $\tilde{\Sigma}_{A,j,j} = \Sigma_{A,j,j}$ for $j \in [r]$. Then the above display implies that $\|\tilde{\Sigma}_A - \Sigma_A\| = o(1)$.

For $j \in [r]$, $\tilde{\Sigma}_{A,j,j} = \Sigma_{A,j,j} = n^{-1}\sum_{i=1}^n \mathbb{E}u_{i,t_j}^2$. By Assumption 1, there exists a constant $c_2 > 0$ with $\tilde{\Sigma}_{A,j,j} \geq c_2$ for $j \in [r]$. It follows, by Lemma C.10(1), that $s_r(\Sigma_A) + s_1(\tilde{\Sigma}_A - \Sigma_A) \geq s_r(\tilde{\Sigma}_A) \geq c_2$. Since $\|\tilde{\Sigma}_A - \Sigma_A\| = o(1)$, we have

$$s_r(\Sigma_A) \geq c_2/2. \tag{B.42}$$

Step 2: show the desired result for $s_r(u)$. By the law of large numbers, we have that for any $j_1, j_2 \in [r]$, $n^{-1}\sum_{i=1}^n \left(u_{i,t_{j_1}}u_{i,t_{j_2}} - \mathbb{E}u_{i,t_{j_1}}u_{i,t_{j_2}}\right) = o_P(1)$. Since r is fixed, this means that

$\|n^{-1}A'A - \Sigma_A\| = o_P(1)$. By Lemma C.10(1), we have

$$s_r(n^{-1}A'A) + s_1(\Sigma_A - n^{-1}A'A) \geq s_r(\Sigma_A) \overset{(i)}{\geq} c_2/2,$$

where (i) holds by (B.42). Since $\|\Sigma_A - n^{-1}A'A\| = o_P(1)$, we have that $s_r(n^{-1}A'A) \geq c_2/2 - o_P(1)$. By $s_r(A) = \sqrt{s_r(A'A)}$ and (B.41), the desired result for $s_r(u)$ holds with $c = c_2/3$.

Step 3: show the desired result for $s_r(\tilde{u})$. Let $B \in \mathbb{R}^{n\times r}$ with $B_{i,j} = \tilde{u}_{i,t_j} - u_{i,j}$. By the definition of $\tilde{u}_{i,t}$,

$$\max_{i,t} |\tilde{u}_{i,t} - u_{i,t}| \leq \max_{i,t} \|x_{i,t}\| \max_t \|\hat{\beta}_t - \beta_t\| \overset{(i)}{=} O_P\left([n^{-1/2} + n^{1/2-\xi}] \log^{O(1)} n\right) = o_P(1),$$

where (i) holds by Lemma B.1(1) and Theorem 3.1, together with $\xi > 6/7$. Since $\|B\| \leq \sqrt{nr} \max_{i,t} |\tilde{u}_{i,t} - u_{i,t}|$, we have that $\|B\| = o_P(\sqrt{n})$.

Notice that $A + B$ is a matrix consisting of r columns of $\tilde{u}$. Hence, Lemma C.10(4) implies $s_r(\tilde{u}) \geq s_r(A+B)$. By Lemma C.10(1), $s_r(A+B) + s_1(-B) \geq s_r(A)$. It follows, by $\|B\| = o_P(\sqrt{n})$, that $s_r(A+B)/\sqrt{n} \geq s_r(A)/\sqrt{n} - o_P(1)$. Since $s_r(A)/\sqrt{n} \geq \sqrt{c_2/3}$ with probability approaching one (due to Step 2), the desired result for $s_r(\tilde{u})$ follows. □

Lemma B.14. *Let Assumption 1 hold. Then for any fixed positive integer r, there exists a constant $c > 0$ such that $\mathbb{P}(s_r(v) > \sqrt{n}c) \to 1$.*

Proof. The proof is the same as that of Lemma B.13 with u replaced by v. □

Proof of Theorem 3.7. **Step 1: show consistency of $\hat{r}_Q^{SV}$.** It suffices to show the following:

(1a) $\max_{r_Q+1\leq r\leq r_{\max}} [s_r(X)/s_{r+1}(X)] = O_P(\sqrt{\log n})$;

(1b) $\max_{1\leq r\leq r_Q-1} [s_r(X)/s_{r+1}(X)] = O_P(1)$;

(1c) $\mathbb{P}(s_{r_Q}(X)/s_{r_Q+1}(X) > T^{1/3}) \to 1$.

We first show condition (1a). Lemma C.10(1) implies that, for $r > r_Q$,

$$s_r(X) = s_r(L_Q F_Q' + v) \leq s_r(L_Q F_Q') + s_1(v) \overset{(i)}{=} 0 + \|v\|, \tag{B.43}$$

where (i) holds by $\text{rank} L_Q = r_Q$. Lemma C.10(1) also implies that, for $r > r_Q$,

$$s_{r+1}(X) + s_1(-L_Q F_Q') \geq s_{r+1}(X - L_Q F_Q') = s_{r+1}(v). \tag{B.44}$$

By (B.43) and (B.44), we have

$$\max_{r_Q+1\leq r\leq r_{\max}} \frac{s_r(X)}{s_{r+1}(X)} \leq \frac{\|v\|}{s_{r_{\max}+1}(v)} \overset{(i)}{=} O_P(\sqrt{\log n}),$$

where (i) holds by Lemmas B.1(2) and B.14. This proves condition (1a).

Since $\|v\| = O_P(\sqrt{n \log n})$ (Lemma B.1(2)), $\|X - L_Q F_Q'\|/\sqrt{nT} = \|v\|/\sqrt{nT} = o_P(1)$. By Assumption 1, the largest r_Q singular values of $L_Q F_Q'/\sqrt{nT}$ are bounded away from zero and infinity. Therefore, there exist constants $c_1, c_2 > 0$ such that the largest r_Q singular values of $X/\sqrt{nT}$ lie in $[c_1, c_2]$ with probability approaching one. This proves condition (1b).

Since $\mathbb{P}[s_{r_Q}(X)/\sqrt{nT} \geq c_1] \to 1$ and $s_{r_Q+1}(X) \leq \|v\| = O_P(\sqrt{n \log n})$ (due to (B.43) and Lemma B.1(2)), we have that $\mathbb{P}[s_{r_Q}(X)/s_{r_Q+1}(X) \geq c_3\sqrt{T/\log n}] \to 1$ for some constant $c_3 > 0$. Condition (1c) follows. We have proved the consistency of $\hat{r}_Q^{SV}$.

Step 2: show consistency of $\hat{r}_\alpha^{SV}$. The argument is similar to Step 1. Notice that $y_t - X_t\hat{\beta}_t = \alpha_t + \tilde{u}_t$, where $\tilde{u}$ is defined in Lemma B.6. Recall that $\alpha = L_\alpha F_\alpha'$ with $\text{rank}\alpha = r_\alpha$. We shall verify the following conditions:

(2a) $\max_{r_\alpha+1 \leq r \leq r_{\max}} [s_r(\alpha + \tilde{u})/s_{r+1}(\alpha + \tilde{u})] = O_P\left([n^{(\xi-1)/2} + n^{(1-\xi)/2}] \log^{O(1)} n\right)$;

(2b) $\max_{1 \leq r \leq r_\alpha - 1} [s_r(\alpha + \tilde{u})/s_{r+1}(\alpha + \tilde{u})] = O_P(1)$;

(2c) $\mathbb{P}\left[s_{r_\alpha}(\alpha + \tilde{u})/s_{r_\alpha+1}(\alpha + \tilde{u}) > M_1 n^{(1+\xi)/2}[n^{\xi/2} + n^{1-\xi/2}]^{-1}/\log^{M_2} n\right] \to 1$ for constants $M_1, M_2 > 0$.

Notice that the above three conditions imply the desired result because for $\xi \in (6/7, 2)$,

$$\frac{n^{(1+\xi)/2}[n^{\xi/2} + n^{1-\xi/2}]^{-1}/\log^{M_2} n}{[n^{(\xi-1)/2} + n^{(1-\xi)/2}]\log^{O(1)} n} \to \infty \qquad \text{and} \qquad [n^{(\xi-1)/2} + n^{(1-\xi)/2}]\log^{O(1)} n \to \infty.$$

Similar to (B.43) and (B.44), we have that, for $r > r_\alpha$, $s_r(\alpha+\tilde{u}) \leq \|\tilde{u}\|$ and $s_{r+1}(\alpha+\tilde{u}) \geq s_{r+1}(\tilde{u})$. Therefore,

$$\max_{r_\alpha+1 \leq r \leq r_{\max}} \frac{s_r(\alpha + \tilde{u})}{s_{r+1}(\alpha + \tilde{u})} \leq \frac{\|\tilde{u}\|}{s_{r_{\max}+1}(\tilde{u})} \overset{(i)}{=} O_P\left([n^{(\xi-1)/2} + n^{(1-\xi)/2}]\log^{O(1)} n\right),$$

where (i) holds by Lemmas B.6(1) and B.13. This proves condition (2a).

Since $\|\tilde{u}\|/\sqrt{nT} = o_P(1)$ (due to Lemma B.6(1) and $T \asymp n^\xi$), condition (2b) follows from the same argument as the proof of condition (1b), except that (L_Q, F_Q, r_Q) is replaced with $(L_\alpha, F_\alpha, r_\alpha)$.

Similar to the proof of condition (1c), we notice that $\mathbb{P}[s_{r_\alpha}(\alpha + \tilde{u})/\sqrt{nT} \geq c_3] \to 1$ for some constant $c_3 > 0$ and $s_{r_\alpha+1}(\alpha + \tilde{u}) \leq \|\tilde{u}\| \overset{(i)}{=} O_P([n^{\xi/2} + n^{1-\xi/2}]\log^{O(1)} n)$ (with (i) due to Lemma B.6(1)). Hence, condition (2c) follows by $T \asymp n^\xi$. We have proved the consistency of $\hat{r}_\alpha^{SV}$. □

Proof of Theorem 4.1. To avoid complicated notations involving j_0, we prove the result for the full vector β_t (rather than $\beta_{j_0,t}$), i.e.,

$$\sqrt{T}(\tilde{\Theta} - \hat{\Theta}) = o_P(1), \tag{B.45}$$

where $\hat{\Theta} = (\sum_{t=1}^T \beta_t z_t')(\sum_{t=1}^T z_t z_t')^{-1}$ and $\tilde{\Theta} = (\sum_{t=1}^T \tilde{\beta}_t z_t')(\sum_{t=1}^T z_t z_t')^{-1}$. The result stated in Theorem 4.1 corresponds to the j_0-th row of the above display.

By (B.35) in the proof of Theorem 3.3 (with $J = I_{kT}$), we have $\tilde{\beta}_t - \beta_t = n^{-1/2} D_{n,t} + n^{-1/2}\delta_t$, where $\delta_t = n^{-1/2}\sum_{i=1}^n G_{i,t}$ and $D_{n,t}$ and $G_{i,t}$ are defined in (B.1). Thus, the definitions of $\tilde{\Theta}$ and $\hat{\Theta}$ imply that

$$\begin{aligned}\sqrt{T}(\tilde{\Theta} - \hat{\Theta}) &= \left(T^{-1/2}\sum_{t=1}^T (\tilde{\beta}_t - \beta_t) z_t'\right)\left(T^{-1}\sum_{t=1}^T z_t z_t'\right)^{-1} \\ &\overset{(i)}{=} \left(\frac{1}{\sqrt{nT}}\sum_{t=1}^T D_{n,t} z_t'\right) O_P(1) + \left(\frac{1}{\sqrt{nT}}\sum_{t=1}^T \delta_t z_t'\right) O_P(1), \qquad \text{(B.46)}\end{aligned}$$

where (i) holds by $(T^{-1}\sum_{t=1}^T z_t z_t')^{-1} = O_P(1)$. The rest of the proof proceeds in two steps in which we bound the two terms in (B.46).

Step 1: show $\frac{1}{\sqrt{nT}}\sum_{t=1}^T D_{n,t} z_t' = o_P(1)$. Notice that

$$\begin{aligned}\left\|\frac{1}{\sqrt{nT}}\sum_{t=1}^T D_{n,t} z_t'\right\| &\le \frac{1}{\sqrt{nT}}\sum_{t=1}^T \|D_{n,t}\|\|z_t\| \\ &\overset{(i)}{\le} \|D_n\|_\infty \sqrt{\frac{k}{nT}}\sum_{t=1}^T \|z_t\| \\ &\overset{(ii)}{=} \|D_n\|_\infty \sqrt{\frac{Tk}{n}} O_P(1) \\ &\overset{(iii)}{=} O_P\left([n^{\xi/2-1} + n^{3-7\xi/2}]\log^{O(1)} n\right)\sqrt{n^{\xi-1}} O_P(1) \overset{(iv)}{=} o_P(1), \qquad \text{(B.47)}\end{aligned}$$

where (i) follows by Holder's inequality and $\|D_{n,t}\| \le \sqrt{k}\|D_n\|_\infty$, (ii) follows by $T^{-1}\sum_{t=1}^T \|z_t\| = O_P(1)$ (due to $T^{-1}\sum_{t=1}^T \mathbb{E}\|z_t\| \le \max_t \mathbb{E}\|z_t\| = O(1)$), (iii) follows by $T \asymp n^\xi$ and Lemma B.8 and (iv) holds by $6/7 < \xi < 3/2$.

Step 2: show $\frac{1}{\sqrt{nT}}\sum_{t=1}^T \delta_t z_t' = o_P(1)$. Therefore, by Lemma C.6 (applied with $\mathcal{F}_n$ being the trivial σ-algebra), $\max_{1\le t\le T}\|\delta_t\|_\infty = O_P(\sqrt{\log(kT)})$. Let $r_n = \sum_{t=1}^T z_t \otimes \delta_t$. Then

$$\begin{aligned}\mathbb{E}\|r_n\|^2 = \sum_{s,t=1}^T \mathbb{E}\left[(z_t' \otimes \delta_t')(z_s \otimes \delta_s)\right] &= \sum_{s,t=1}^T \mathbb{E}\left[(z_t' z_s)(\delta_t'\delta_s)\right] \overset{(i)}{=} \sum_{s,t=1}^T \mathbb{E}(z_t' z_s)\mathbb{E}(\delta_t'\delta_s) \\ &\le T \max_{s,t}|\mathbb{E}(z_t' z_s)| \max_s \sum_{t=1}^T |\mathbb{E}(\delta_t'\delta_s)| \overset{(ii)}{=} O(T)\max_s \sum_{t=1}^T |\mathbb{E}(\delta_t'\delta_s)|, \qquad \text{(B.48)}\end{aligned}$$

where (i) holds by the independence between $\{z_t\}_{t=1}^T$ and (u, v) and (ii) holds by $\max_{s,t}|\mathbb{E}(z_t' z_s)| \le \max_t \mathbb{E}\|z_t\|^2 = O(1)$. Notice that

$$\max_s \sum_{t=1}^{T} |\mathbb{E}(\delta_t'\delta_s)| = \max_s \sum_{t=1}^{T} \left| n^{-1} \sum_{i,j=1}^{n} \mathbb{E}G_{i,t}'G_{j,s} \right| \overset{(i)}{=} \max_s \sum_{t=1}^{T} \left| n^{-1} \sum_{i=1}^{n} \mathbb{E}G_{i,t}'G_{i,s} \right|$$

$$\leq \max_s n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} |\mathbb{E}G_{i,t}'G_{i,s}| \leq \max_{(i,s)\in[n]\times[s]} \sum_{t=1}^{T} |\mathbb{E}G_{i,t}'G_{i,s}| \overset{(ii)}{=} O(1), \quad \text{(B.49)}$$

where (i) follows by the independence of $\{(G_{i,s}, G_{i,t})\}_{i=1}^{n}$ across i and (ii) holds by Lemma B.2(3). It follows, by (B.48) and (B.49), that $\mathbb{E}\|r_n\|^2 = O(T)$ and thus

$$\text{vec}\left(\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \delta_t z_t'\right) = \frac{1}{\sqrt{nT}} r_n = \frac{1}{\sqrt{nT}} O_P(\sqrt{T}) = o_P(1). \quad \text{(B.50)}$$

Hence, we obtain (B.45) by combining (B.46) with (B.47) and (B.50). The proof is complete. □

The following lemma is needed to prove Lemma B.16.

B.4 Strong mixing with geometric decay for Example 2.1

Lemma B.15. *Let $\varphi(x, y; r)$ be the p.d.f of $N\left(0, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}\right)$ for $|r| < 1$. Suppose that $\forall t, s \geq 0$, $\sigma_t, \sigma_s \geq 1$ and $|r_{t,s}| \leq C\exp(-|t-s|)$ for some constant $C > 0$. Then there exist constants $\tau, M > 0$ such that $\sup_{t,s,|t-s|>\tau} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_0^1 \varphi(i/\sigma_t, j/\sigma_s; hr_{t,s}/(\sigma_t\sigma_s))dh \leq M$.*

Proof. Using the formula for the p.d.f of bivariate Gaussian random vectors, we have that

$$\varphi(x, y; r) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left[-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right].$$

Choose $\tau > 0$ such that $|r_{t,s}|/(\sigma_t\sigma_s) \leq 1/2$ for $|t-s| \geq \tau$. Hence, for $0 \leq h \leq 1$,

$$\varphi(x, y; hr_{t,s}/(\sigma_t\sigma_s)) \leq \frac{1}{2\pi\sqrt{3/4}} \exp\left[-\frac{x^2 - xy + y^2}{3/2}\right]$$

$$\leq \frac{1}{\sqrt{3}\pi} \exp\left[-\frac{2}{3}\left((x^2+y^2)/2 + (x^2+y^2-xy)/2\right)\right] \leq \frac{1}{\sqrt{3}\pi} \exp\left[-\frac{1}{3}\left(x^2+y^2\right)\right].$$

It follows that for $|t-s| \geq \tau$,

$$\begin{aligned}
\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \int_0^1 \varphi(i/\sigma_t, j/\sigma_s; hr_{t,s}/(\sigma_t\sigma_s))dh &\leq \frac{1}{\sqrt{3}\pi} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \exp\left[-\frac{1}{3}\left(i^2+j^2\right)\right] \\
&< \frac{1}{\sqrt{3}\pi} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \exp\left[-\frac{1}{3}(i+j)\right] \\
&= \frac{1}{\sqrt{3}\pi} \left(\sum_{i=0}^{\infty} \exp(-i/3)\right)^2 < \infty.
\end{aligned}$$

Therefore, the desired result holds with $M = \left(\sum_{i=0}^{\infty} \exp\left(-i/3\right)\right)^2 / \left(\sqrt{3}\pi\right)$. ☐

Lemma B.16. *Under the setup of Example 2.1, there exist a constant $M > 0$ such that $\forall t \geq 1$, $\alpha_{mixing}(t) \leq Mc^t$.*

Proof. Recall the notations in Example 2.1. By Theorem 2.2 of Piterbarg (2012), it suffices to verify that

(i) For some constants $\tau_0, K_1 > 0$, we have that $\forall \tau \geq \tau_0$

$$\sup_{t,s,|t-s|>\tau} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_0^1 \varphi\left(i/\sigma_t, j/\sigma_s; hr(t,s)/(\sigma_t \sigma_s)\right) dh \leq K_1$$

(ii) For some constant $K_2 \in (0,1)$, we have that $\forall \tau \geq \tau_0$

$$\sup_{v} \sum_{t=v+\tau}^{\infty} \sum_{s=0}^{v} \frac{|r(t,s)|}{\sigma_t \sigma_s} \leq K_2^{|t-s|}.$$

Notice that $\sigma_t^2 = \sum_{j=0}^{\infty} \gamma_{t-1,j}^2 \geq \gamma_{t-1,0}^2 = 1$. For $t > s$,

$$\begin{aligned} |r(t,s)| = \left| \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \gamma_{t-1,j_1} \gamma_{s-1,j_2} \mathbb{E} u_{t-1-j_1} u_{s-1-j_2} \right| \overset{(i)}{=} \left| \sum_{j=0}^{\infty} \gamma_{t-1,j+t-s} \gamma_{s-1,j} \right| \\ \leq \sum_{j=0}^{\infty} c^{2j+t-s} = Dc^{t-s} \quad \text{for} \quad D = \frac{1}{1-c^2}, \end{aligned} \tag{B.51}$$

where (i) follows by u_t being i.i.d $N(0,1)$. Hence, claim (i) holds by Lemma B.15.

By (B.51), we have that for any $v \geq 0$,

$$\sum_{t=v+\tau}^{\infty} \sum_{s=0}^{v} \frac{|r(t,s)|}{\sigma_t \sigma_s} \leq D \sum_{t=v+\tau}^{\infty} \sum_{s=0}^{v} c^{t-s} = \frac{D}{1-c} \sum_{t=v+\tau}^{\infty} c^{t-v}(1 - c^{v+1}) \overset{(i)}{\leq} \frac{D}{1-c} \sum_{t=v+\tau}^{\infty} c^{t-v} = \frac{D}{(1-c)^2} c^{\tau},$$

where (i) holds by $c \in (0,1)$. Claim (ii) follows. The proof is complete. ☐

C Useful technical tools

C.1 Useful results on probability theory

Lemma C.1. *Let X and Y be two random vectors. Then $\forall t, \varepsilon > 0$, we have*

$$\left|\mathbb{P}\left(\|X\|_\infty \leq t\right) - \mathbb{P}\left(\|Y\|_\infty \leq t\right)\right| \leq \mathbb{P}\left(\|X - Y\|_\infty > \varepsilon\right) + \mathbb{P}\left(\|Y\|_\infty \in (t-\varepsilon, t+\varepsilon]\right).$$

Proof. By the triangular inequality,

$$\begin{aligned}\mathbb{P}\left(\|X\|_\infty > t\right) &\leq \mathbb{P}\left(\|X-Y\|_\infty > \varepsilon\right) + \mathbb{P}\left(\|Y\|_\infty > t-\varepsilon\right) \\ &= \mathbb{P}\left(\|X-Y\|_\infty > \varepsilon\right) + \mathbb{P}\left(\|Y\|_\infty > t\right) + \mathbb{P}\left(\|Y\|_\infty \in (t-\varepsilon, t]\right). \qquad \text{(C.1)}\end{aligned}$$

On the other hand, also by the triangular inequality,

$$\begin{aligned}\mathbb{P}\left(\|X\|_\infty > t\right) &\geq \mathbb{P}\left(\|Y\|_\infty > t+\varepsilon\right) - \mathbb{P}\left(\|X-Y\|_\infty > \varepsilon\right) \\ &= \mathbb{P}\left(\|Y\|_\infty > t\right) - \mathbb{P}\left(\|Y\|_\infty \in (t, t+\varepsilon]\right) - \mathbb{P}\left(\|X-Y\|_\infty > \varepsilon\right). \qquad \text{(C.2)}\end{aligned}$$

It follows, by (C.1) and (C.2), that

$$\left|\mathbb{P}\left(\|X\|_\infty > t\right) - \mathbb{P}\left(\|Y\|_\infty > t\right)\right| \leq \mathbb{P}\left(\|X-Y\|_\infty > \varepsilon\right) + \mathbb{P}\left(\|Y\|_\infty \in (t-\varepsilon, t+\varepsilon]\right).$$

The desired result follows by $|\mathbb{P}\left(\|X\|_\infty > t\right) - \mathbb{P}\left(\|Y\|_\infty > t\right)| = |\mathbb{P}\left(\|X\|_\infty \leq t\right) - \mathbb{P}\left(\|Y\|_\infty \leq t\right)|$. □

Lemma C.2. *Let X and Y be two random vectors. Define $F_X(x) = \mathbb{P}\left(\|X\|_\infty \leq x\right)$ and $F_Y(x) = \mathbb{P}\left(\|Y\|_\infty \leq x\right)$. Then $\forall \varepsilon > 0$,*

$$\sup_{\alpha\in(0,1)} \left|\mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha)\right) - \alpha\right| \leq \varepsilon + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon\right).$$

Proof. Fix $\alpha \in (0,1)$ and notice that

$$\begin{aligned}&\mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha)\right) \\ =\ & \mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha) \text{ and } \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| \leq \varepsilon\right) \\ &+ \mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha) \text{ and } \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon\right) \\ \overset{(i)}{\leq}\ & \mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha) \text{ and } \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| \leq \varepsilon\right) + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon\right) \\ \leq\ & \mathbb{P}\left(\|X\|_\infty > F_X^{-1}(1-\alpha-\varepsilon)\right) + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon\right) \\ \overset{(ii)}{=}\ & \alpha + \varepsilon + \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon\right), \qquad \text{(C.3)}\end{aligned}$$

where (i) follows from Lemma A.1(ii) in Romano and Shaikh (2012) (if $\sup_{x\in\mathbb{R}}[F_Y(x) - F_X(x)] \leq \varepsilon$ then $F_X^{-1}(1-\alpha-\varepsilon) \leq F_Y^{-1}(1-\alpha)$) and (ii) follows by the definition of $F_X(\cdot)$. Also notice that

$$\mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha)\right)$$

$$
\begin{aligned}
&\geq \quad \mathbb{P}\left(\|X\|_\infty > F_Y^{-1}(1-\alpha) \text{ and } \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| \leq \varepsilon \right) \\
&\overset{(i)}{\geq} \quad \mathbb{P}\left(\|X\|_\infty > F_X^{-1}(1-\alpha+\varepsilon) \text{ and } \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| \leq \varepsilon \right) \\
&\overset{(ii)}{\geq} \quad \mathbb{P}\left(\|X\|_\infty > F_X^{-1}(1-\alpha+\varepsilon) \right) - \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon \right) \\
&\overset{(iii)}{=} \quad \alpha - \varepsilon - \mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)| > \varepsilon \right) \qquad \text{(C.4)}
\end{aligned}
$$

where (i) follows from Lemma A.1(ii) in Romano and Shaikh (2012) (if $\sup_{x\in\mathbb{R}}[F_X(x) - F_Y(x)] \leq \varepsilon$ then $F_Y^{-1}(1-\alpha) \leq F_X^{-1}(1-\alpha+\varepsilon)$), (ii) follows by the elementary inequality that for any two events A and B, $\mathbb{P}(A \bigcap B) + \mathbb{P}(B^c) \geq \mathbb{P}(A \bigcap B) + \mathbb{P}(A \bigcap B^c) = \mathbb{P}(A)$ or equivalently $\mathbb{P}(A \bigcap B) \geq \mathbb{P}(A) - \mathbb{P}(B^c)$, and (iii) follows by the definition of $F_X(\cdot)$. The desired result follows by (C.3) and (C.4). $\square$

Lemma C.3. *The following hold.*

1. *Let $Z \in \mathbb{R}^{m_Z}$ be a random vector whose jth entry is denoted by Z_j. Suppose that there exist constants $b, \gamma > 0$ such that $\forall j \in [m_Z]$, Z_j has an exponential-type tail with parameter (b, γ). Then for any nonrandom vector $a \in \mathbb{R}^{m_Z}$, $a'Z$ has an exponential-type tail with parameter $\left(b\|a\|_1 \log^{1/\gamma}(\|a\|_0 + 2), \gamma\right)$.*

2. *Let $\{Z_j\}_{j=1}^{m_Z}$ be a sequence of random variables. Suppose that constants $b, \gamma > 0$ satisfy that $\forall j \in [m_Z]$, Z_j has an exponential-type tail with parameter (b, γ). Let $q > 0$ be any nonrandom number. Then there exists a constant $C_{\gamma,q} > 0$ depending only on γ and q such that $\mathbb{E} \max_{1 \leq j \leq m_Z} |Z_j|^q \leq C_{\gamma,q} m_Z b^q$ and $\mathbb{E}|Z_j|^q \leq C_{\gamma,q} b^q$ $\forall j \in [m_Z]$.*

3. *Let Z_1 and Z_2 be two random variables having exponential-type tails with parameters (b_1, γ_1) and (b_2, γ_2), respectively. Then $\forall \gamma \in \left(0, \gamma_1\gamma_2(\gamma_1+\gamma_2)^{-1}\right)$, $Z_1 Z_2$ has an exponential-type tail with parameter $\left(2^{1/\gamma_0} b_1 b_2, \gamma_0\right)$, where $\gamma_0 = \gamma_1\gamma_2(\gamma_1+\gamma_2)^{-1}$*

4. *Let X have an exponential-type tail with parameter (b_X, γ_X). Then $\forall a \in \mathbb{R}$, $X - a$ has an exponential-type tail with parameter $(b_X + |a|, \gamma_X)$.*

Proof. **Proof of part (1).** Let $A_0 := \{i \mid a_i \neq 0\}$. Then by Holder's inequality and the union bound, $\mathbb{P}(|a'Z| > x) \leq \mathbb{P}(\|a\|_1 \max_{i\in A_0} |Z_i| > x) \leq \sum_{i \in A_0} \mathbb{P}(\|a\|_1 |Z_i| > x) \leq \|a\|_0 \exp\left[1 - (xb^{-1}\|a\|_1^{-1})^\gamma\right]$. If $\|a\|_0 = 1$, then the result follows by $b\|a\|_1 < b\|a\|_1 \log^{1/\gamma}(3)$. For $\|a\|_0 > 1$, we let $c = b\|a\|_1 \log^{1/\gamma} \|a\|_0 < b\|a\|_1 \log^{1/\gamma}(\|a\|_0 + 2)$. For $x \leq c$, $\mathbb{P}(|a'Z| > x) \leq 1 \leq \exp(1 - (x/c)^\gamma)$. Since $\mathbb{P}(|a'Z| > x) \leq \|a\|_0 \exp\left[1 - (xb^{-1}\|a\|_1^{-1})^\gamma\right]$, it suffices to show that $\forall x > c$, $\log \|a\|_0 - (xb^{-1}\|a\|_1^{-1})^\gamma \leq 1 - (xc^{-1})^\gamma$. This is to say that $x^\gamma \geq (\log \|a\|_0 - 1)/((b\|a\|_1)^{-\gamma} - c^{-\gamma})$ $\forall x > c$. By simple computations, one can show that $c^\gamma = (\log \|a\|_0 - 1)/((b\|a\|_1)^{-\gamma} - c^{-\gamma})$. Part (1) follows.

Proof of part (2). Notice that, by the union bound, $\mathbb{P}(\max_{1\le j\le m_Z}|Z_j|>x)\le\sum_{j=1}^{m_Z}\mathbb{P}(|Z_j|>x)\le m_Z\exp\left[1-(x/b)^\gamma\right]$. Then

$$\begin{aligned}\mathbb{E}\max_{1\le j\le m_Z}|Z_j|^q \overset{(i)}{=} \int_0^\infty \mathbb{P}\left(\max_{1\le j\le m_Z}|Z_j|^q>x\right)dx &= \int_0^\infty \mathbb{P}\left(\max_{1\le j\le m_Z}|Z_j|>x^{1/q}\right)dx\\ &\overset{(ii)}{\le} m_Z\int_0^\infty \exp\left[1-\left(x^{1/q}/b\right)^\gamma\right]dx\\ &\overset{(iii)}{=} m_Z b^q\left(q\gamma^{-1}\int_0^\infty e^{1-z}z^{q/\gamma-1}dz\right),\end{aligned}$$

where (i) follows by the identity $\mathbb{E}X=\int_0^\infty\mathbb{P}(X>x)dx$ for any non-negative random variable X, (ii) follows by $\mathbb{P}(\max_{1\le j\le m_Z}|Z_j|>x)\le m_Z\exp\left[1-(x/b)^\gamma\right]$ and (iii) follows by a change of variable $z=\left(x^{1/q}/b\right)^\gamma$. The bound for $\mathbb{E}\max_{1\le j\le m_Z}|Z_j|^q$ follows with $C_{\gamma,q}=q\gamma^{-1}\int_0^\infty e^{1-z}z^{q/\gamma-1}dz$. The bound for $\mathbb{E}|Z_j|^q$ follows by the same reasoning with $\max_{1\le j\le m_Z}|Z_j|$ replaced by $|Z_j|$. This completes the proof for part (2).

Proof of part (3). The proof of Lemma A.2 of Fan et al. (2011) implies that $\forall\gamma\in(0,\gamma_0)$, Z_1Z_2 has an exponential-type tail with parameter (b_3,γ) $\forall b_3>b_0\max\{(\gamma/\gamma_0)^{1/\gamma_0},(1+\log 2)^{1/\gamma_0}\}$, where $b_0=b_1b_2$. Let $b_*=2^{1/\gamma_0}b_1b_2$. It is easy to check that $b_*>b_0\max\{(\gamma/\gamma_0)^{1/\gamma_0},(1+\log 2)^{1/\gamma_0}\}$. Thus, Z_1Z_2 has an exponential-type tail with parameter (b_*,γ) $\forall\gamma\in(0,\gamma_0)$. In other words, for any $x>0$, $\mathbb{P}(|Z_1Z_2|>x)\le\exp[-(x/b_*)^\gamma]$ $\forall\gamma\in(0,\gamma_0)$. We take the infimum of the right-hand side over γ and obtain for any $x>0$, $\mathbb{P}(|Z_1Z_2|>x)\le\exp[-(x/b_*)^{\gamma_0}]$. Part (3) follows.

Proof of part (4). Let $c=b_X+|a|$. Notice that $\mathbb{P}(|X-a|>t)\le\mathbb{P}(|X|+|a|>t)=\mathbb{P}(|X|>t-|a|)$. For $t\in(0,c]$, $\mathbb{P}(|X|>t-|a|)\le 1\le\exp\left[1-(t/c)^{\gamma_X}\right]$. For $t>c$, $t-|a|>0$ and $\mathbb{P}(|X|>t-|a|)\le\exp\left[1-((t-|a|)/b_x)^{\gamma_X}\right]$. It is easy to check that $(t-|a|)/b_x\ge t/c$ $\forall t>c$. Part (4) follows. □

Lemma C.4. *Let $\Sigma\in\mathbb{R}^{p\times p}$ be a positive-semi-definite matrix. Suppose that there exists a constant $b>0$ such that $\min_{1\le j\le p}\Sigma_{j,j}\ge b$. Then there exists a constant $C_b>0$ depending only on b such that $\forall\varepsilon>0$,*

$$\sup_{x\in\mathbb{R}}|\Phi(x-\varepsilon,\Sigma)-\Phi(x+\varepsilon,\Sigma)|\le C_b\varepsilon\sqrt{\log p}.$$

Proof. Let $Y\sim N(0,\Sigma)$ with its jth component denoted by Y_j. By Nazarov's anti-concentration inequality (Lemma A.1 in Chernozhukov et al. (2014)), there exists a constant $C_b'>0$ depending only on b such that

$$\begin{aligned}\sup_{x\in\mathbb{R}}\mathbb{P}\left(\max_{1\le j\le p}Y_j\in(x-\varepsilon,x+\varepsilon]\right) &\le 2C_b'\varepsilon\sqrt{\log p}\\ \sup_{x\in\mathbb{R}}\mathbb{P}\left(\max_{1\le j\le p}(-Y_j)\in(x-\varepsilon,x+\varepsilon]\right) &\le 2C_b'\varepsilon\sqrt{\log p}.\end{aligned}$$

Therefore,

$$\sup_{x\in\mathbb{R}} |\Phi(x-\varepsilon,\Sigma)-\Phi(x+\varepsilon,\Sigma)| = \sup_{x\in\mathbb{R}} \mathbb{P}\left(\|Y\|_\infty \in (x-\varepsilon, x+\varepsilon]\right)$$
$$\overset{(i)}{\leq} \sup_{x\in\mathbb{R}} \mathbb{P}\left(\max_{1\leq j\leq p} Y_j \in (x-\varepsilon,x+\varepsilon]\right) + \sup_{x\in\mathbb{R}} \mathbb{P}\left(\max_{1\leq j\leq p} (-Y_j) \in (x-\varepsilon,x+\varepsilon]\right) \leq 4C_b'\varepsilon\sqrt{\log p},$$

where (i) holds by $\|Y\|_\infty \in \{\max_{1\leq j\leq p} Y_j,\ \max_{1\leq j\leq p}(-Y_j)\}$. The proof is complete. □

Lemma C.5. *Let Σ_A and Σ_B be $p\times p$ positive semi-definite matrices. Define $\Delta = \max_{1\leq j,k\leq p} |\Sigma_{A,j,k} - \Sigma_{B,j,k}|$. Suppose that there exist constants $c, C > 0$ such that $c \leq \Sigma_{A,j,j} \leq C$ for $1\leq j\leq p$. Then there exists a constant $K>0$ depending only on c and C such that*

$$\sup_{x\in\mathbb{R}} |\Phi(x,\Sigma_A)-\Phi(x,\Sigma_B)| \leq C\Delta^{1/3}\left(1\vee \log(2p/\Delta)\right)^{2/3}.$$

Proof. Consider random vectors $X\sim N(0,\Sigma_A)$ and $Y\sim N(0,\Sigma_B)$. Define $\bar{X} = (X', -X')'$ and $\bar{Y} = (Y', -Y')'$. Notice that $\bar{X} \sim N(0,\bar{\Sigma}_A)$ and $\bar{Y}\sim N(0,\bar{\Sigma}_B)$, where $\bar{\Sigma}_A = D\otimes\Sigma_A$, $\bar{\Sigma}_B = D\otimes \Sigma_B$ and $D = \begin{pmatrix} 1 & -1 \\ -1 & 1\end{pmatrix}$.

The definition of $\bar{\Sigma}_A$ and $\bar{\Sigma}_B$ also implies that (1) $\max_{1\leq j,k\leq 2p} |\bar{\Sigma}_{A,j,k} - \bar{\Sigma}_{B,j,k}| = \max_{1\leq j,k\leq p} |\Sigma_{A,j,k} - \Sigma_{B,j,k}| = \Delta$ and (2) the diagonal entries of $\bar{\Sigma}_A$ lie in $[c, C]$. It follows, by Lemma 3.1 of Chernozhukov et al. (2013), that there exists a constant $M>0$ depending only on c and C such that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\max_{1\leq j\leq 2p} \bar{X}_j \leq x\right) - \mathbb{P}\left(\max_{1\leq j\leq 2p} \bar{Y}_j \leq x\right)\right| \leq M\Delta^{1/3}\left(1\vee\log(2p/\Delta)\right)^{2/3}.$$

We obtain the desired result by noticing that $\|X\|_\infty = \max_{1\leq j\leq 2p}\bar{X}_j$ and $\|Y\|_\infty = \max_{1\leq j\leq 2p}\bar{Y}_j$. □

Lemma C.6. *Let $\{u_{i,j}\}_{(i,j)\in[n]\times J}$ be an array of random variables and $\mathcal{F}_n$ be a σ-algebra. Suppose the following hold:*
(i) Condition on $\mathcal{F}_n$, u_i is independent across i, where $u_i = \{u_{i,j} \mid j\in J\}$.
(ii) There exist constants $b,\gamma > 0$ such that $\forall (i,j)\in[n]\times J$ and $\forall x>0$, $\mathbb{P}(|u_{i,j}| > x \mid \mathcal{F}_n) \leq \exp\left(1-(x/b)^\gamma\right)$ a.s.
(iii) $\forall 0 < c < \infty$, $n^{-c}\log|J| \to 0$, where $|J|$ denotes the cardinality of J.

Then
$$\max_{j\in J}\left|\sum_{i=1}^n \left[u_{i,j} - \mathbb{E}(u_{i,j}\mid\mathcal{F}_n)\right]\right| = O_P(\sqrt{n\log|J|}).$$

Proof. Let $\tilde{u}_{i,j} = u_{i,j} - \mathbb{E}(u_{i,j}\mid\mathcal{F}_n)$. By Lemma C.3(2) and (4) applied to the conditional probability measure $\mathbb{P}(\cdot\mid\mathcal{F}_n)$, we have that there exists a constant $b_1>0$ depending only on b and γ such that $\forall z>0$ $\forall (i,j)\in[n]\times J$, $\mathbb{P}(|\tilde{u}_{i,j}| > x\mid\mathcal{F}_n) \leq \exp\left(1-(x/b_1)^\gamma\right)$ a.s. Due to the conditional independence, the strong mixing coefficients are always zero.

Then by Theorem 1 in Merlevède et al. (2011) (applied to the conditional probability measure $\mathbb{P}(\cdot \mid \mathcal{F}_n)$), there exist positive constants C_1, C_2, C_3, C_4, C_5 and r depending only on bM_ε and γ such that $r < 1$ and $\forall z > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n} \tilde{u}_{i,j}\right| > z\sqrt{n\log|J|} \mid \mathcal{F}_n\right) \\
\leq\ & n\exp\left[-C_1\left(z\sqrt{n\log|J|}\right)^r\right] + \exp\left[-\frac{C_2 n z^2\log|J|}{1+nC_3}\right] \\
&+\exp\left\{-C_4 z^2\log|J|\exp\left[C_5\log^{-r}\left(z\sqrt{n\log|J|}\right)\left(z\sqrt{n\log|J|}\right)^{r/(1-r)}\right]\right\}\ a.s.
\end{aligned}$$

Then, by the union bound, we have that

$$\begin{aligned}
&\mathbb{P}\left(\max_{j\in J}\left|\sum_{i=1}^{n} \tilde{u}_{i,j}\right| > z\sqrt{n\log|J|} \mid \mathcal{F}_n\right) \\
\leq\ & \sum_{j\in J}\mathbb{P}\left(\max_{j\in J}\left|\sum_{i=1}^{n} \tilde{u}_{i,j}\right| > z\sqrt{n\log|J|} \mid \mathcal{F}_n\right) \\
\leq\ & |J|n\exp\left[-C_1\left(z\sqrt{n\log|J|}\right)^r\right] + |J|\exp\left[-\frac{C_2 n z^2\log|J|}{1+nC_3}\right] \\
&+|J|\exp\left\{-C_4 z^2\log|J|\exp\left[C_5\log^{-r}\left(z\sqrt{n\log|J|}\right)\left(z\sqrt{n\log|J|}\right)^{r/(1-r)}\right]\right\}\ a.s.
\end{aligned}$$

By assumption (iv), the first and third terms in the above display go to zero for any $z > 0$. Hence, $\forall\varepsilon > 0$, we can choose a large constant $z_* > 0$ such that

$$\mathbb{P}\left(\max_{j\in J}\left|\sum_{i=1}^{n} \tilde{u}_{i,j}\right| > z_*\sqrt{n\log|J|} \mid \mathcal{F}_n\right) \leq \varepsilon\ a.s. \tag{C.5}$$

Hence, we have proved the result since, for an arbitrary $\varepsilon > 0$, we can find $z_* > 0$ such that the above equation holds. The result follows by the law of iterated expectations. □

Lemma C.7. *Let $\{W_j\}_{j\in J}$ be random variables. If there exist constant $b, \gamma > 0$ such that $\forall j \in J$, W_j has an exponential-type tail with parameter (b,γ), then $\max_{j\in J}|W_j| = O_P(\log^{1/\gamma}|J|)$, where $|J|$ is the cardinality of J.*

Proof. By the union bound, we have

$$\begin{aligned}
\mathbb{P}\left(\max_{j\in J}|W_j| > (\log|J|)^{1/\gamma}x\right) &\leq \sum_{j\in J}\mathbb{P}\left(|W_j| > (\log|J|)^{1/\gamma}x\right) \\
&\leq |J|\exp\left[1-\left((\log|J|)^{1/\gamma}x/b\right)^\gamma\right] \\
&= \exp\left[1+(1-(x/b)^\gamma)\log|J|\right].
\end{aligned}$$

Hence, for any $\varepsilon > 0$, one can choose large enough x such that the right-hand side of the above display is smaller than ε. The result follows. □

Lemma C.8. *Let $\mathcal{F}_n$ be a σ-algebra and $\{W_t\}_{t=1}^T$ be random variables with $\mathbb{E}(W_t \mid \mathcal{F}_n) = 0$. Suppose that the following hold:*
(i) There exist constants $\gamma_1, b_1 > 0$ such that $\forall t \in [T]$ and $\forall z > 0$, $\mathbb{P}(|W_t| > z \mid \mathcal{F}_n) \leq \exp[1 - (z/b_1)^{\gamma_1}]$ a.s.
(ii) There exist constants $\gamma_2, b_2 > 0$ such that $\alpha_n(t \mid \mathcal{F}_n) \leq b_2 \exp(-t^{\gamma_2})$ a.s, where

$$\alpha_n(t \mid \mathcal{F}_n) := \sup\Big\{\Big|\mathbb{P}(A \mid \mathcal{F}_n)\mathbb{P}(B \mid \mathcal{F}_n) - \mathbb{P}(A \bigcap B \mid \mathcal{F}_n)\Big| : \ A \in \sigma\left(\{(W_s, \dots, W_s) \mid s \leq \tau\}\right), \\ B \in \sigma\left(\{(W_s, \dots, W_s) \mid s \geq \tau + t\}\right) \text{ and } \tau \in \mathbb{Z}\Big\}.$$

Then $\forall z > 0$, $\mathbb{P}(|T^{-1/2}\sum_{t=1}^T W_t| > z \mid \mathcal{F}_n) \leq \exp[1 - (z/b)^{\gamma}]$ a.s., where $b, \gamma > 0$ are constants depending only on γ_1, γ_2, b_1 and b_2.

Proof. Let $\gamma_3 = \min\{\gamma_2, 1/2\}$. Notice that $\alpha_n(t \mid \mathcal{F}_n) \leq b_2 \exp(-t^{\gamma_3})$ and $\gamma_3 < 1$. Thus, $\gamma := (\gamma_1^{-1} + \gamma_3^{-1})^{-1} < 1$. Hence, by Theorem 1 in Merlevède et al. (2011) (applied to the conditional probability measure $\mathbb{P}(\cdot \mid \mathcal{F}_n)$), there exist constants $C_1, C_2, C_3, C_4, C_5 > 0$ depending only on γ, γ_3, b_1 and b_2, such that $\forall z > 0$,

$$\mathbb{P}\left(\left|\sum_{t=1}^T W_t\right| > zT^{1/2} \mid \mathcal{F}_n\right) \leq \underbrace{T\exp\left(-C_1 T^{\gamma/2} z^{\gamma}\right)}_{J_{1,T}(z)} + \underbrace{\exp\left(-\frac{C_2 z^2 T}{1 + C_3 T}\right)}_{J_{2,T}(z)} \\ + \underbrace{\exp\left[-C_4 z^2 \exp\left(C_5 \frac{(T^{1/2}z)^{\gamma/(1-\gamma)}}{\left[\log(T^{1/2}z)\right]^{\gamma}}\right)\right]}_{J_{3,T}(z)} \ a.s.$$

It is not hard to see that one can choose a large enough constant $K > 0$ such that $\forall z \geq K$, $J_{1,T}(z) \leq \exp(-C_1 z^{\gamma})$, $J_{3,T}(z) \leq J_{1,T}(z)$ and $J_{2,T}(z) \leq \exp(-C_6 z^2)$, where $C_6 = C_2/(1 + C_3)$. Hence, $\forall z \geq K$, $J_{1,T}(z) + J_{2,T}(z) + J_{3,T}(z) \leq 2\exp(-C_1 z^{\gamma}) + \exp(-C_6 z^2)$. Since $\gamma < 1$, we have that $\forall z \geq K$,

$$\mathbb{P}\left(T^{-1/2}\left|\sum_{t=1}^T W_t\right| > z \mid \mathcal{F}_n\right) \leq 3\exp\left(-C_7 z^{\gamma}\right) \ a.s, \tag{C.6}$$

where $C_7 = \min\{C_1, C_6\}$. Let $b := \max\left\{K, (C_7^{-1}\log 3)^{1/\gamma}\right\}$.

For $z \in (0, b]$, $\exp[1 - (z/b)^{\gamma}] \geq 1 \geq \mathbb{P}(T^{-1/2}|\sum_{t=1}^T W_t| > z \mid \mathcal{F}_n)$. It is easy to verify that, $\forall z > b$, $3\exp(-C_7 z^{\gamma}) \leq \exp[1 - (z/b)^{\gamma}]$. It follows, by (C.6), that $\forall z > b$, $\mathbb{P}(T^{-1/2}|\sum_{t=1}^T W_t| > z \mid \mathcal{F}_n) \leq \exp[1 - (z/b)^{\gamma}]$. The proof is complete. □

Lemma C.9. *Let $x > 0$ and $\{b_j\}_{j=1}^q \subset \mathbb{R}$. Then $\sum_{j=1}^q x^{b_j} \leq q(x^{b_{\min}} + x^{b_{\max}})$, where $b_{\min} = \min_{1 \leq j \leq q} b_j$ and $b_{\max} = \max_{1 \leq j \leq q} b_j$.*

Proof. We discuss two cases: (A) $x \in (0,1]$ and (B) $x > 1$. In Case (A), $x^{b_{\min}} \geq x^{b_j}$ for $1 \leq j \leq q$ and thus $\sum_{j=1}^{q} x^{b_j} \leq q x^{b_{\min}} \leq q(x^{b_{\min}} + x^{b_{\max}})$. In Case (B), $x^{b_{\max}} \geq x^{b_j}$ for $1 \leq j \leq q$ and thus $\sum_{j=1}^{q} x^{b_j} \leq q x^{b_{\max}} \leq q(x^{b_{\min}} + x^{b_{\max}})$. The proof is complete. □

C.2 Useful results on PCA

Lemma C.10. *The following hold.*
(1) Let $A, B \in \mathbb{R}^{n_1 \times n_2}$ be two matrices. If $i + j - 1 \leq \min\{n_1, n_2\}$, then $s_{i+j-1}(A + B) \leq s_i(A) + s_j(B)$, where $s_j(\cdot)$ denotes the jth largest singular value.
(2) Let $A \in \mathbb{R}^{n_1 \times n_0}$ and $B \in \mathbb{R}^{n_0 \times n_2}$. If $1 \leq i \leq n_0$, then $s_i(AB) \geq s_i(A) s_{n_0 - i + 1}(B)$.
(3) Let $A, B \in \mathbb{R}^{n_1 \times n_2}$ be two matrices. If $\mathrm{rank} B \leq r$ and $1 \leq j \leq \min\{n_1, n_2\} - r$, then $s_j(A) \geq s_{j+r}(A + B) \geq s_{2r+j}(A)$.
(4) Let $A \in \mathbb{R}^{n_1 \times n_2}$. Let $B \in \mathbb{R}^{n_1 \times m}$ be the matrix consisting of the first m columns of A with $m \leq n_2$. Then for $j \in [m \wedge n_1]$, $s_j(B) \leq s_j(A)$.

Proof. Part (1) and (4) are Fact 6(b) and Fact 3, respectively, in Chapter 17.4 of Hogben (2006). Part (2) follows by Lemma 3 of Wang and Xi (1997). Part (3) follows by applying part (1): $s_j(A) = s_j(A) + s_{r+1}(B) \geq s_{j+r}(A+B)$ and $s_{j+r}(A+B) = s_{j+r}(A+B) + s_{r+1}(-B) \geq s_{2r+j}(A)$. □

Lemma C.11. *Let $W = LF' + e$ with $L \in \mathbb{R}^{n \times r}$ and $F \in \mathbb{R}^{T \times r}$. Let $W = \hat{U}\hat{\Sigma}\hat{V}'$ be an SVD and $\hat{U}_1 \in \mathbb{R}^{n \times r}$ the first r columns of $\hat{U}$. Define $\hat{L} = \sqrt{n}\hat{U}_1$ and $\hat{F}_t = n^{-1}\hat{L}'W_t$, where $W = (W_1, \ldots, W_T)$, $e = (e_1, \ldots, e_T)$ and $F = (F_1, \ldots, F_T)'$. Suppose that the following hold:*
(i) $\|e\| = o_P(\sqrt{nT})$ and $T \asymp n^{\kappa}$ for a constant $\kappa > 0$
(ii) There exist $0 < m_1 \leq m_2 < \infty$ such that all the eigenvalues of $\Sigma_F := T^{-1}F'F$ and $\Sigma_L := n^{-1}L'L$ belong to $[m_1, m_2]$ wpa1.
Then the following hold:

(1) $\hat{L}\hat{F}_t - LF_t = (n^{-1}\hat{L}\hat{L}' - I)LF_t + n^{-1}\hat{L}\hat{L}'e_t$.

(2) $\hat{L} - LH = \Delta_L$, *where* $H = F'FL'\hat{L}\hat{\Omega}_1^{-2}(nT)^{-1}$, $\Delta_L = (nT)^{-1}(LF'e' + eW')\hat{L}\hat{\Omega}_1^{-2}$ *and* $\hat{\Omega}_1 = \hat{\Sigma}_1(nT)^{-1/2}$ *and* $\hat{\Sigma}_1$ *is the upper-left* $r \times r$ *submatrix of* $\hat{\Sigma}$.

(3) $\|\hat{\Omega}_1^{-2}\| = O_P(1)$, $\|\Delta_L\| = O_P(\|e\|/\sqrt{T})$, $\|H\| = O_P(1)$ *and* $\|HH' - \Sigma_L^{-1}\| = O_P(\|e\|/\sqrt{nT})$.

(4) There exists a random variable $A_* = O_P(1)$ that does not depend on t such that, with probability one, $\forall a \in \mathbb{R}^n$,

$$\left| a'(\hat{L}\hat{F}_t - LF_t) \right| \leq n^{-\kappa} M_1 \|F'e'a\| A_* + \left[n^{-(1+\kappa)/2}\|e\|M_1 + n^{-1}M_2 + n^{-1-\kappa/2}\|e\|^2 \right] A_* \|L'a\| + \left[n^{-1/2-\kappa}\|e\|^2 M_1 + n^{-(1+\kappa)}\|e\|^3 + n^{-1-\kappa/2}\|e\|M_2 \right] A_* \|a\|,$$

where $M_1 = \|F\|_\infty$ and $M_2 = \max_t \|L'e_t\|$.

Proof. **Proof for part (1).** Since $\hat{F}_t = n^{-1}\hat{L}'W_t = n^{-1}\hat{L}'(LF_t + e_t)$, we have $\hat{L}\hat{F}_t = n^{-1}\hat{L}\hat{L}'(LF_t + e_t)$ and thus $\hat{L}\hat{F}_t - LF_t = (n^{-1}\hat{L}\hat{L}' - I)LF_t + n^{-1}\hat{L}\hat{L}'e_t$. Part (1) follows.

Proof for part (2). By the definition of $\hat{L}$, we have $WW'\hat{L} = \hat{L}\hat{\Sigma}_1^2$ and thus $\hat{L} = WW'\hat{L}\hat{\Sigma}_1^{-2}$. We obtain part (2) by noticing that

$$
\begin{aligned}
WW'\hat{L}\hat{\Sigma}_1^{-2} &= (LF' + e)W'\hat{L}\hat{\Sigma}_1^{-2} = LF'W'\hat{L}\hat{\Sigma}_1^{-2} + eW'\hat{L}\hat{\Sigma}_1^{-2} \\
&= LF'(FL' + e')\hat{L}\hat{\Sigma}_1^{-2} + eW'\hat{L}\hat{\Sigma}_1^{-2} = LH + (LF'e' + eW')\hat{L}\hat{\Sigma}_1^{-2}.
\end{aligned}
$$

Proof for part (3). Notice that by Lemma C.10(2), $s_r(LF') \geq s_1(L)s_r(F)$. Thus, by assumption (ii), it follows that there exists $b > 0$ such that $\mathbb{P}\left((nT)^{-1/2}s_r(LF') > b\right) \to 1$. By Lemma C.10(1), $s_r(W) + \|e\| = s_r(LF' + e) + s_1(-e) \geq s_r(LF')$. Thus,

$$
\begin{aligned}
\mathbb{P}\left((nT)^{-1/2}s_r(W) + (nT)^{-1/2}\|e\| \geq b\right) &= \mathbb{P}\left(s_r(W) + \|e\| > \sqrt{nT}b\right) \\
&\geq \mathbb{P}\left(s_r(LF') > \sqrt{nT}b\right) \to 1.
\end{aligned}
$$

Since $\|e\|/\sqrt{nT} = o_P(1)$, $\mathbb{P}\left(s_r(W)/\sqrt{nT} > b/2\right) \to 1$. Notice that $\|\hat{\Omega}_1^{-2}\| = nTs_r^{-2}(W)$. Therefore, $\|\hat{\Omega}_1^{-2}\|$ is bounded above by $4/b^2$ with probability approaching one. In other words, $\|\hat{\Omega}_1^{-2}\| = O_P(1)$.

The definition of Δ_L (in part (2)) implies that

$$
\begin{aligned}
\|\Delta_L\| &\leq (nT)^{-1}\Big[\|L\|\|F\|\|e\| + \|e\|\|LF' + e\|\Big]\|\hat{L}\|\|\hat{\Omega}_1^{-2}\| \\
&\leq (nT)^{-1}\Big[2\|L\|\|F\|\|e\| + \|e\|^2\Big]\|\hat{L}\|\|\hat{\Omega}_1^{-2}\| \overset{(i)}{=} O_P(T^{-1/2}\|e\|),
\end{aligned}
$$

where (i) follows by $\|L\| = O_P(n^{1/2})$, $\|F\| = O_P(T^{1/2})$, $\|\hat{L}\| = n^{1/2}$, $\|\hat{\Omega}_1^{-2}\| = O_P(1)$ and $\|e\|/\sqrt{nT} = o_P(1)$. Notice that

$$
\|H\| = \|F'F\|\|L\|\|\hat{L}\|\|\hat{\Omega}_1^{-2}\|/(nT) = O_P(T)O_P(n^{1/2})n^{1/2}O_P(1)/(nT) = O_P(1).
$$

Observe that

$$
I_r = n^{-1}\hat{L}'\hat{L} = n^{-1}(LH + \Delta_L)'(LH + \Delta_L) = H'\Sigma_L H + n^{-1}H'L'\Delta_L + n^{-1}\Delta_L' LH + n^{-1}\Delta_L'\Delta_L. \quad \text{(C.7)}
$$

Also observe that

$$
\begin{cases}
\|H'L'\Delta_L\| \leq \|H\| \cdot \|L\| \cdot \|\Delta_L\| = O_P(\|e\|\sqrt{n/T}) \\
\|\Delta_L'\Delta_L\| \leq \|\Delta_L\|^2 = O_P(\|e\|^2/T) \overset{(i)}{=} o_P(\|e\|\sqrt{n/T}),
\end{cases}
\quad \text{(C.8)}
$$

where (i) holds by $\|e\| = o_P(\sqrt{nT})$. Then (C.7) and (C.8) imply $H'\Sigma_L H + O_P(\|e\|/\sqrt{nT}) = I_r$.

By $O_P(\|e\|/\sqrt{nT}) = o_P(1)$, it follows that $I_r - (H'\Sigma_L H)^{-1} = O_P(\|e\|/\sqrt{nT})$ and thus

$$\|HH' - \Sigma_L^{-1}\| = \|H(I_r - (H'\Sigma_L H)^{-1})H'\| \leq \|H\| \cdot \|I_r - (H'\Sigma_L H)^{-1}\| \cdot \|H\| = O_P(\|e\|/\sqrt{nT}).$$

Proof for part (4). Let $A_{n,1} = \|n^{-1}L'L\|\|(HH' - \Sigma_L^{-1})\|$, $A_{n,2} = n^{-1}\|L\|\|\Delta_L\|^2$, $A_{n,3} = n^{-1}\|L\|\|\Delta_L\|\|H\|$, $A_{n,4} = \|n^{-1}L'L\|\|H\|$, $A_{n,5} = (nT)^{-1}\|\hat{\Omega}_1^{-2}\|\|\hat{L}\|$, $A_{n,6} = A_{n,1} + A_{n,3} + A_{n,4}A_{n,5}\|e\|\|F\|$, $A_{n,7} = A_{n,2} + A_{n,4}A_{n,5}\|e\|^2$ and $A_{n,8} = A_{n,4}A_{n,5}\|L\|$. Notice that

$$|a'(\hat{L}\hat{F}_t - LF_t)| \overset{(i)}{\leq} \|L'(n^{-1}\hat{L}\hat{L}' - I)a\|\|F_t\| + n^{-1}\|\hat{L}'a\|\|\hat{L}'e_t\|, \tag{C.9}$$

where (i) holds by part (1). Also notice that

$$\begin{aligned}
& \|L'(n^{-1}\hat{L}\hat{L}' - I)a\| \\
= \; & \|n^{-1}L'(\hat{L}\hat{L}' - L\Sigma_L^{-1}L')a\| \\
\overset{(i)}{=} \; & \|n^{-1}L'\left(L(HH' - \Sigma_L^{-1})L' + \Delta_L\Delta_L' + \Delta_L H'L' + LH\Delta_L'\right)a\| \\
\overset{(ii)}{\leq} \; & A_{n,1}\|L'a\| + A_{n,2}\|a\| + A_{n,3}\|L'a\| + A_{n,4}\|\Delta_L' a\| \\
\overset{(iii)}{\leq} \; & (A_{n,1} + A_{n,3})\|L'a\| + A_{n,2}\|a\| + A_{n,4}A_{n,5}\left(\|We'a\| + \|e\|\|F\|\|L'a\|\right) \\
\overset{(iv)}{\leq} \; & (A_{n,1} + A_{n,3})\|L'a\| + A_{n,2}\|a\| + A_{n,4}A_{n,5}\left(\|L\|\|F'e'a\| + \|e\|^2\|a\| + \|e\|\|F\|\|L'a\|\right) \\
= \; & A_{n,6}\|L'a\| + A_{n,7}\|a\| + A_{n,8}\|F'e'a\|,
\end{aligned} \tag{C.10}$$

where (i) follows by $\hat{L} = LH + \Delta_L$, (ii) follows by the triangular inequality and the submultiplicativity of $\|\cdot\|$, (iii) follows by the definition of Δ_L (part (2)) and (iv) follows by $W = LF' + e$. Then,

$$\begin{aligned}
|a'(\hat{L}\hat{F}_t - LF_t)| & \overset{(i)}{\leq} \sqrt{r}\Big[A_{n,6}\|L'a\| + A_{n,7}\|a\| + A_{n,8}\|F'e'a\|\Big]M_1 + n^{-1}\|\hat{L}'a\|\|\hat{L}'e_t\| \\
& \overset{(ii)}{\leq} \sqrt{r}\Big[A_{n,6}\|L'a\| + A_{n,7}\|a\| + A_{n,8}\|F'e'a\|\Big]M_1 \\
& \qquad + n^{-1}\left(\|H\|\|L'a\| + \|\Delta_L\|\|a\|\right)\left(\|H\|M_2 + \|\Delta_L\|\|e_t\|\right) \\
& \overset{(iii)}{\leq} \underbrace{\sqrt{r}A_{n,8}}_{J_1}\|F'e'a\|M_1 + \underbrace{\Big[\sqrt{r}A_{n,6}M_1 + n^{-1}\|H\|^2M_2 + n^{-1}\|H\|\|\Delta_L\|\|e\|\Big]}_{J_2}\|L'a\| \\
& \qquad + \underbrace{\Big[\sqrt{r}A_{n,7}M_1 + n^{-1}\|\Delta_L\|^2\|e\| + n^{-1}\|H\|\|\Delta_L\|M_2\Big]}_{J_3}\|a\|,
\end{aligned} \tag{C.11}$$

where (i) is due to (C.9), (C.10) and $\|F_t\| \leq \sqrt{r}M_1$, (ii) follows by $\hat{L} = LH + \Delta_L$ and (iii) follows by $\|e_t\| \leq \|e\|$.

By simple computations using part (3) and $T \asymp n^{\kappa}$, we have that

$$\begin{cases} J_1 = O_P(n^{-\kappa}) \\ J_2 = O_P\left(n^{-(1+\kappa)/2}\|e\|M_1 + n^{-1}M_2 + n^{-1-\kappa/2}\|e\|^2\right) \\ J_3 = O_P\left(n^{-1/2-\kappa}\|e\|^2 M_1 + n^{-(1+\kappa)}\|e\|^3 + n^{-1-\kappa/2}\|e\|M_2\right). \end{cases} \tag{C.12}$$

Notice that J_1, J_2 and J_3 do not depend on a. Therefore, part (4) follows by (C.11) and (C.12). The proof is complete. □

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14.

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4):1121–1159.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–56.

Andrews, D. W. (2005). Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial studies*, 20(3):651–707.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer Science & Business Media.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, pages 135–171.

Bai, J. (2009). 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.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Belloni, A. and Chernozhukov, V. (2011). L1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732.

Blomström, M., Lipsey, R. E., and Zejan, M. (1996). Is fixed investment the key to economic growth? *The Quarterly Journal of Economics*, 111(1):269–276.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bollerslev, T., Gibson, M., and Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of econometrics*, 160(1):235–245.

Bond, S., Leblebicioglu, A., and Schiantarelli, F. (2010). Capital accumulation and growth: a new look at the empirical evidence. *Journal of Applied Econometrics*, 25(7):1073–1099.

Boneva, L., Linton, O., and Vogt, M. (2015). A semiparametric model for heterogeneous panel data with fixed effects. *Journal of Econometrics*.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, 39(3):1496–1525.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2).

Campbell, J. Y., Lo, A. W.-C., MacKinlay, A. C., et al. (1997). *The econometrics of financial markets*, volume 2. princeton University press Princeton, NJ.

Campbell, J. Y. and Shiller, R. J. (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies*, 1(3):195–228.

Campbell, J. Y. and Shiller, R. J. (1988b). Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39.

Chen, J., Gao, J., and Li, D. (2013). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business & Economic Statistics*, 31(3):315–330.

Chen, X., Shao, Q.-M., and Wu, W. B. (2016). Self-normalized cramér-type moderate deviations under dependence. *The Annals of Statistics*, 44(4):1593–1617.

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.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Central limit theorems and bootstrap in high dimensions. *arXiv preprint arXiv:1412.3661*.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

DeLong, B. and Summers, L. H. (1991). Equipment investment and economic growth. *Quarterly Journal of Economics*, 106(2):445–502.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, 25(1):23–49.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies*, 29(1):69–103.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Feller, W. (1968). *An introduction to probability theory and its applications: volume I*, volume 3. John Wiley & Sons London-New York-Sydney-Toronto.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Forni, M. and Lippi, M. (1997). *Aggregation and the microfoundations of dynamic macroeconomics.* Oxford University Press.

Frank, M. Z. and Goyal, V. K. (2009). Capital structure decisions: which factors are reliably important? *Financial management*, 38(1):1–37.

Freund, C. L. and Weinhold, D. (2004). The effect of the internet on international trade. *Journal of international economics*, 62(1):171–189.

Freyberger, J. (2012). Nonparametric panel data models with interactive fixed effects.

Giacomini, R. and Rossi, B. (2006). How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics*, 68(s1):783–795.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, 76(2):669–705.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Glosten, L. R., Jaganathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Gorman, W. M. (1981). Some engel curves. *Essays in the theory and measurement of consumer behaviour in honor of sir Richard Stone.*

Gorodetskii, V. (1978). On the strong mixing property for linear sequences. *Theory of Probability & Its Applications*, 22(2):411–413.

Goyal, A. and Santa-Clara, P. (2003). Idiosyncratic risk matters! *The Journal of Finance*, 58(3):975–1007.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Graham, J. R. and Harvey, C. R. (2001). The theory and practice of corporate" nance: evidence from the" eld. *Journal of Financial Economics*, 60(187):243.

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 Moon, H. R. (2006). Reducing bias of mle in a dynamic panel model. *Econometric Theory*, 22(03):499–512.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84.

Hansen, K. T., Heckman, J. J., and Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1):39–98.

Harris, M. and Raviv, A. (1991). The theory of capital structure. *the Journal of Finance*, 46(1):297–355.

Haugen, R. A. and Lakonishok, J. (1988). *The incredible January effect: The stock market's unsolved mystery*. Business One Irwin.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies*, 5(3):357–386.

Hogben, L. (2006). *Handbook of linear algebra*. CRC Press.

Hsiao, C. (2014). *Analysis of panel data*, volume 54. Cambridge university press.

Hsiao, C., Appelbe, T. W., and Dineen, C. R. (1993). A general framework for panel data models with an application to canadian customer-dialed long distance telephone service. *Journal of Econometrics*, 59(1-2):63–86.

Islam, N. (1995). Growth empirics: a panel data approach. *The Quarterly Journal of Economics*, pages 1127–1170.

Jones, C. I. (1995). Time series tests of endogenous growth models. *The Quarterly Journal of Economics*, pages 495–525.

Jones, C. P., Pearce, D. K., and Wilson, J. W. (1987). Can tax-loss selling explain the january effect? a note. *The Journal of Finance*, 42(2):453–461.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17(2):357–390.

Koijen, R. and Van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3(1):467–491.

Kramer, C. (1994). Macroeconomic seasonality and the january effect. *Journal of Finance*, pages 1883–1891.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.

Lewbel, A. (1991). The rank of demand systems: theory and nonparametric estimation. *Econometrica: Journal of the Econometric Society*, pages 711–730.

Lin, C.-C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1):42–55.

Litan, R. E. and Rivlin, A. M. (2001). Projecting the economic impact of the internet. *The American Economic Review*, 91(2):313–317.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.

Lord, F., Novick, M., and Birnbaum, A. (1968). Statistical theories of mental test scores.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Moon, H. R. and Phillips, P. C. (2004). Gmm estimation of autoregressive roots near unity with panel data. *Econometrica*, 72(2):467–522.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Peng, H. and Schick, A. (2012). Asymptotic normality of quadratic forms with random vectors of increasing dimension. *Preprint*.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Pesaran, M. H. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1):50–93.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

Phillips, P. C. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Phillips, P. C. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal*, 6(1):217–259.

Piterbarg, V. I. (2012). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148. American Mathematical Society.

Pouzo, D. (2015). Bootstrap consistency for quadratic forms of sample averages with increasing dimension. *Electronic Journal of Statistics*, 9(2):3046–3097.

Qian, J. and Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics*, 167(2):483–493.

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.

Rajan, R. G. and Zingales, L. (1995). What do we know about capital structure? some evidence from international data. *Journal of Finance*, 50(5):1421–60.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2):821–862.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate us stock returns. *Journal of Financial Econometrics*, 4(2):238–274.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798–2822.

Rosenberg, B. (1972). The estimation of stationary stochastic regression parameters reexamined. *Journal of the American Statistical Association*, 67(339):650–654.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory*, 13(3):341–360.

Rossi, B. (2013). Advances in forecasting under instability. *Handbook of Economic Forecasting*, 2:1203–1324.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. Technical report, National bureau of economic research.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, 1:515–554.

Stock, J. H. and Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39(s1):3–33.

Su, L. and Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29(06):1079–1135.

Su, L., Jin, S., and Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*, 186(1):222–244.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–23.

Titman, S. and Wessels, R. (1988). The determinants of capital structure choice. *The Journal of finance*, 43(1):1–19.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. *Manuscript, Harvard University*, 4(3.5):3–0.

Wang, B.-Y. and Xi, B.-Y. (1997). Some inequalities for singular values of matrix products. *Linear algebra and its applications*, 264:109–115.

Welch, I. (2004). Capital structure and stock returns. *Journal of Political Economy*, 112(1):106–131.

Zaffaroni, P. (2009). Generalized least squares estimation of panel with common shocks. In *XXIX-th European Meeting of Statisticians, 2013, Budapest Contents*, page 41.