Uniform Inference for Characteristic Effects of Large Continuous-Time Linear Models[*]

Yuan Liao   Xiye Yang[†]

November 26, 2018

Abstract

We consider continuous-time models with a large panel of moment conditions, where the structural parameter depends on a set of characteristics, whose effects are of interest. The leading example is the linear factor model in financial economics where factor betas depend on observed characteristics such as firm specific instruments and macroeconomic variables, and their effects pick up long-run time-varying beta fluctuations. We specify the factor betas as the sum of characteristic effects and an orthogonal idiosyncratic parameter that captures high-frequency movements. It is often the case that researchers do not know whether or not the latter exists, or its strengths, and thus the inference about the characteristic effects should be valid uniformly over a broad class of data generating processes for idiosyncratic parameters. We construct our estimation and inference in a two-step continuous-time GMM framework. It is found that the limiting distribution of the estimated characteristic effects has a discontinuity when the variance of the idiosyncratic parameter is near the boundary (zero), which makes the usual "plug-in" method using the estimated asymptotic variance only valid pointwise and may produce either over- or under- coveraging probabilities. We show that the uniformity can be achieved by cross-sectional bootstrap. Our procedure allows both known and estimated factors, and also features a bias correction for the effect of estimating unknown factors.

Key words: Large dimensions, high-frequency data, cross-sectional bootstrap, GMM

[†]Address: 75 Hamilton St., New Brunswick, NJ 08901, USA. Email: `yuan.liao@rutgers.edu`, `xiyeyang@econ.rutgers.edu`

1   Introduction

Conditional factor models have been playing an important role in capturing the time-varying sensitivities of individual outcomes to the factors, in which the factor betas are varying over time. In this paper, we study a continuous-time conditional factor model:

$$\mathbf{Y}_t = \mathbf{Y}_0 + \int_0^t \boldsymbol{\alpha}_s ds + \int_0^t \boldsymbol{\beta}_s d\mathbf{F}_s + \mathbf{U}_t \tag{1.1}$$

where $\mathbf{Y}_0$ is the starting value of the outcome process at time 0; the model contains a factor, idiosyncratic, and a drift process $\{(\mathbf{F}_t, \mathbf{U}_t, \boldsymbol{\alpha}_t) : t \geq 0\}$. Here $(\mathbf{Y}_t, \boldsymbol{\alpha}_t, \mathbf{U}_t)$ are all high-dimensional (whose dimension $p \to \infty$), and the dimension of the factor process $\mathbf{F}_t$ is fixed. In addition, the process $\{\boldsymbol{\beta}_t : t \geq 0\}$ represents the factor loadings (or "betas"), which is assumed to be stochastic in this paper. Model (1.1) covers many uses of factor models in asset pricing. In financial economics, extensive empirical studies have shown that assets' individual betas can be largely explained by their "characteristics" (or called "instruments"). These include lagged characteristics that are common to all stocks, characteristics specific to individual stocks, as well as observations of other firm instruments (Gagliardini et al., 2016). Estimated betas as functions of the conditioning characteristics represent the effects of characteristics on firm specific sensitivities to the risk factors. They pick up long-run patterns and fluctuations in the betas. Therefore, estimating the characteristic effects on the individual betas is one of the central econometric tasks in financial economics. Given the importance of the topic, the econometric problem, however, is challenging for the reasons we shall elucidate below.

Let $\boldsymbol{\beta}_{lt}$ denote the K-dimensional factor betas of the l-th individual at time t. Let $\mathbf{X}_{lt}$ be a vector of observed characteristics that may be varying across individuals and times, and $\mathbf{X}_t = \{\mathbf{X}_{lt} : l \leq p\}$. We model: for each $l = 1, ..., p$,

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad \mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0, \tag{1.2}$$

where $\mathbf{g}_t(\cdot)$ is an unknown time-varying nonparametric function that is assumed to be well approximable by a sieve representation. Here, each component $\gamma_{lt,r}$ of $\boldsymbol{\gamma}_{lt}$ ($r \leq K$) satisfies, for some constant $C > 0$,

$$\text{Var}(\gamma_{lt,r}|\mathbf{X}_t) \in [0, C] \text{ almost surely.} \tag{1.3}$$

The main message from (1.2)-(1.3) is that in the identification condition $\mathbf{g}_t(\mathbf{X}_{lt}) = \mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{X}_{lt})$, the variance of the "error components" $\boldsymbol{\gamma}_{lt} := \boldsymbol{\beta}_{lt} - \mathbf{g}_t(\mathbf{X}_{lt})$ is defined on a compact set that includes zero, and can be either exactly on, arbitrarily close to, or bounded away from its "boundary". Thus we allow *arbitrarily unknown signal strengths* of $\boldsymbol{\gamma}_{lt}$. So the economic meaning of model (1.2) is that the factor beta is decomposed as the sum of two components: (i) a nonparametric function of the observed instruments, $\mathbf{g}_{lt} := \mathbf{g}_t(\mathbf{X}_{lt})$, which we call "characteristic betas", and (ii) a time-varying and individual specific component, $\boldsymbol{\gamma}_{lt}$, which we call "idiosyncratic betas". The characteristic beta $\mathbf{g}_{lt}$ picks up long-run beta patterns and fluctuations, and possesses less volatile, while $\boldsymbol{\gamma}_{lt}$ captures high frequency movements in beta, and represents the remaining time-varying individual factor sensitivities after conditioning on the observed characteristics. In addition, the strength of the idiosyncratic betas is allowed to be arbitrary, reflecting the nature that the explainability from the characteristics is unknown.

The goal of this paper is to provide a *uniformly valid* inference of $\mathbf{g}_{lt}$, the characteristic effects on factor betas. By "uniformly valid", we mean the coverage probability is asymptotically correct uniformly over a broad class of data generating processes (DGPs) that allows various possible signal strengths of $\boldsymbol{\gamma}_{lt}$, measured by a weighted cross-sectional variance:

$$\mathbf{V}_{\gamma,t} := \mathrm{Var}\left(\frac{1}{\sqrt{p}} \sum_{m=1}^{p} h_{t,ml} \boldsymbol{\gamma}_{mt} \bigg| \mathbf{X}_t \right).$$

We assume $\boldsymbol{\gamma}_{lt}$ to be conditionally cross-sectionally independent given $\mathbf{X}_t$, and find that the strength of $\mathbf{V}_{\gamma,t}$ plays a crucial role in the asymptotic behavior of estimated characteristic effect, and affects both the rate of convergence and limiting distributions. Here $h_{t,ml}$ is a function of $\mathbf{X}_t$ whose definition will be clear in the paper. In particular, the asymptotic distribution of $\mathbf{g}_{lt}$ has a "discontinuity" when the strength of $\boldsymbol{\gamma}_{lt}$, the eigenvalues of $\mathbf{V}_{\gamma,t}$, are near zero. As a consequence, the usual *pointwise* inference procedures under a fixed DGP only produce confidence intervals that are valid for specific DGPs, therefore potentially produce misleading inferences. Specifically, benchmark methods in the literature, which ignore the high-frequency beta dynamics in $\boldsymbol{\gamma}_{lt}$, would produce *under-coveraging* confidence intervals of the characteristic effects. On the other hand, we show that even if $\boldsymbol{\gamma}_{lt}$ is allowed, standard "plug-in" procedures using the estimated asymptotic variances do not produce uniformly correct coverage probabilities, because they require very strong signal strengths of $\mathbf{V}_{\gamma,t}$, leading to *over-coveraging* confidence intervals when the signal strength is weak. The discontinuity issue here is similar to the problem of estimating parameters on a boundary. As is shown by, e.g., (Andrews, 1999) and (Ketz, 2017), when a test statistic has a discontinuity

in its limiting distribution, as occurs in estimating parameters on a boundary and in random coefficients models, pointwise asymptotics can be very misleading.

We reply on a cross-sectional bootstrap to achieve the uniform inference. It is important to note that the employed bootstrap is cross-sectional, which resamples the cross-sectional individuals while keeping all the serial observations for each resampled individual. This procedure is essential because the discontinuity arises when the cross-sectional variance $\mathbf{V}_{\gamma,t}$ is near the boundary, and the cross-sectional bootstrap avoids estimating $\mathbf{V}_{\gamma,t}$ in the current context. We show that the bootstrap procedure leads to a correct asymptotic coverage probability and is uniformly valid over a large class of DGPs, and explain the reasons in detail.

Allowing $\mathbf{V}_{\gamma,t}$ to be unrestricted on a compact set that includes zero as the boundary point is the main motivation of this paper, which arises from the following practical consideration. While characteristics may fully explain the factor betas at times when they are just updated and made publicly available ($\boldsymbol{\gamma}_{lt} \approx 0$), there are also times when factor betas contain either unmeasurable or high-frequency components that are more volatile and cannot be captured by the characteristics ($\boldsymbol{\gamma}_{lt}$ very different from zero). In those occasions, modeling the beta as fully specified functions of observed characteristics can be very restrictive. This is particularly true for high-dimensional and high-frequency factor models in empirical asset pricing, where individual factor betas demonstrate large heterogeneity when the number of assets is large, and assets' returns are available at a very high frequency. On the contrary, characteristics such as the firm sizes and book-market values, often vary more smoothly and are measured at a much lower frequency, often (but not always) leaving large portions of stock betas' dynamics unexplained. As we show in this paper, without taking into account the high-frequency movements of betas after conditioning on the characteristics, the inference procedures of characteristics' effects are not asymptotically valid. Unfortunately, this is often the case in the financial economic literature, which has been dominated by modeling betas as fully specified functions of the observed characteristics, including both parametric (e.g., Shanken (1990); Cochrane (1996); Ferson and Harvey (1999); Avramov and Chordia (2006); Gagliardini et al. (2016)) and nonparametric models, e.g., Connor and Linton (2007) and Connor et al. (2012). As is shown by Ghysels (1998), misspecifying beta risk may result in serious pricing errors that might even be larger than those produced by an unconditional asset pricing model.

4

1.1 Many moment conditions in continuous-time

The model can be generalized to the high-dimensional continuous-time linear moment conditions framework. For each $l \leq p$, the parameter $\boldsymbol{\beta}_{lt}$ is identified by the following moment condition:

$$\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) = 0, \quad \forall t \in [0, T] \tag{1.4}$$

where $\mathbf{c}_{z,lt} = d[\mathbf{Z}_l, \mathbf{Z}_l]_t/dt$ is the instantaneous quadratic variation process of $\mathbf{Z}_l = \{\mathbf{Z}_{lt}\}_{t \geq 0}$, and $\Psi_l(\boldsymbol{\beta}, \mathbf{c})$ is a known function linear with respect to $\boldsymbol{\beta}$. In addition, we observe a set of time-varying characteristics $\mathbf{X}_{lt}$, so that we have the following decomposition:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l = 1, \cdots, p, \tag{1.5}$$

where $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$. The model admits the linear factor model as a special case by setting $\mathbf{Z}_{lt} = (Y_{lt}, \mathbf{F}_t)$ and

$$\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) = \mathbf{c}_{FF,t}\boldsymbol{\beta}_{lt} - \mathbf{c}_{YF,lt}, \quad \mathbf{c}_{z,mt} = (\mathbf{c}_{FF,t}, \mathbf{c}_{YF,mt}).$$

Here $\mathbf{c}_{FF,t}$ and $\mathbf{c}_{YF,lt}$ are the quadratic covariances for the processes $\mathbf{F}_t$ and Y_{lt}. Moreover, the above model also admits a continuous-time linear regression model with individual-specific regressors: e.g., Barndorff-Nielsen and Shephard (2004); Mykland et al. (2006); Kalnina (2012), as well as the idiosyncratic volatility model applied by Ang et al. (2009); Herskovic et al. (2016); Li et al. (2016).

We provide a general two-step estimation procedure to make uniform inferences about the characteristic effect $\mathbf{g}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt})$ for each fixed $l \leq p$. In step (i), we estimate $\boldsymbol{\beta}_{lt}$ by either directly solving (1.4) or using generalized method of moments (GMM), with the sample quadratic variation in place of $\mathbf{c}_{z,lt}$, and in step (ii), we estimate $\mathbf{g}_{lt}$ by a standard nonparametric sieve regression on (1.5) with the estimated $\boldsymbol{\beta}_{lt}$. We aim to construct a confidence interval CI_τ for $\mathbf{g}_{lt}$ for any specific $l \leq p$ and at a specific time t, so that at the nominal level $1 - \tau$,

$$\lim_{p,T \to \infty} \sup_{\mathbb{P} \in \mathcal{P}} |\mathbb{P}(\mathbf{g}_{lt} \in CI_\tau) - (1 - \tau)| = 0$$

Here the probability measure $\mathbb{P}$ is taken uniformly over a broad DGP class $\mathcal{P}$, which admits various cross-sectional variations in $(\boldsymbol{\gamma}_{lt}, \mathbf{g}_{lt})$ and dynamics if they are also time-varying. While our framework belongs to a more general class of two-step GMM estimators, we encounter a new feature as in the linear factor model we described earlier: the estimated $\mathbf{g}_{lt}$

possesses a discontinuity on its limiting distribution because the strengths of $\mathbf{V}_{\gamma,t}$ is unknown and can be near the boundary. This brings new challenges to achieving the uniformity in the inference. Uniformity in the above sense is essential in this context, because it makes the inference valid and robust to the unknown degrees of dynamics in $\boldsymbol{\beta}_{lt}$, especially, the strengths of the cross-sectional variations in $\boldsymbol{\gamma}_{lt}$.

In the presence of "boundary parameters", the asymptotic inference becomes nonstandard, and modified inference procedures have been proposed in both time series and cross-sectional models, e.g., Ketz (2017, 2018); Pedersen (2017). The main difference between our model and those considered in the literature is that we do not model $\mathbf{V}_{\gamma,t}$ explicitly as an unknown "parameter", so it does not appear in the GMM objective function. In fact, the mapping from $\boldsymbol{\gamma}_{lt}$ to the limiting distribution of the estimated $\mathbf{g}_{lt}$ is still continuous, which makes the cross-sectional bootstrap first-order valid. We provide more detailed explanations in Section 2.

1.2   The literature

The linear factor model covers many useful models in the arbitrage pricing theory and models of linkages between international stock markets, as well as macroeconomics. Earlier literature include, e.g., Chamberlain and Rothschild (1983); Connor and Korajczyk (1993); King et al. (1994); Stock and Watson (2002); Bai and Ng (2002). With the massive use of the newly available datasets of intraday asset prices and large number of cross-sectional data, the continuous-time factor model of high dimensions have received extensive attentions in the recent high-frequency literature, such as Pelger (2016); Aït-Sahalia and Xiu (2017); Fan et al. (2016a); Li et al. (2018).

The study of the effects of characteristics on betas is an essential subject in financial economics. For instance, it is commonly known that firm sensitivities to risk factors depend on the firm specific raw size and value characteristics. As is noted by Daniel and Titman (1997), "It is the firms' characteristics (size and ratios) rather than the covariance structure of returns that appear to explain the cross sectional variation in stock returns." Ang and Kristensen (2012) also found that the market risk premium is less correlated with value stocks' beta (stocks with high book-to-market ratio) than with growth stocks' beta. Firms' momentum is also one of the commonly used characteristics, whose effect on the factor sensitivities has been found to be linearly growing with the momentum, indicating a constant effect. In addition, Ferson and Harvey (1999) found that the lagged characteristics track

variations in expected returns that is not captured by the Fama-French (Fama and French, 1992) three-factor model, and that these characteristics have explanatory power on the factor loadings because they pick up betas' time-variation. In addition, the effect of common characteristics such as the term spread and default spread demonstrate significantly different volatiles among betas of individual stocks and portfolios, explaining the larger heterogeneity of the factor loadings for the former. Other empirical evidence that systematic risk is related to firm characteristics and business cycle variables is provided by Jagannathan and Wang (1996); Lettau and Ludvigson (2001), among many others.

While most of the aforementioned works assume that the betas are fully explained by observed characteristics, a similar decomposition to (1.2) was given by Kelly et al. (2017), where betas are decomposed into a linear function of lagged characteristics as well as an unobservable loading component. They specifically require $\boldsymbol{\gamma}_{lt}$ to be "strong", with cross-sectional variances that are bounded away from zero. Fan et al. (2016b) and Kim et al. (2018) respectively studied a model whose betas have a similar decomposition. They did not study the inference problem. Our paper is also related to the continuous-time GMM framework of Li and Xiu (2016), but is different on the inference aspect, where we study the continuous-GMM estimation in the presence of high-dimensional linear moment conditions, and when combined with the nonparametric regression, there is a discontinuity issue on the limiting distribution for the characteristic effect. Other literature on continuous-time regression models can be found from Barndorff-Nielsen and Shephard (2004); Mykland et al. (2006); Li et al. (2017), among others. Also note that we focus on the continuous components of the factor models, and factor model for the jump components was studied by Li et al. (2018).

The rest of this paper is organized as follows. Section 2 informally discusses the issue of uniformity and the intuitive solutions using the cross-sectional bootstrap. Section 3 describes the continuous-time conditional factor model driven by stochastic processes. We separately study the known and the unknown factor cases, and present the asymptotic results of the estimators. Section 4 extends the model to the more general continuous-time GMM framework, with many moment conditions. Section 5 presents real data applications on the high-frequency stock return data of firms from S&P500. Finally, in the supplement, we give simulated example to examine the uniformity of the proposed inference in finite sample, additional empirical findings, as well as all the technical proofs.

Notation: We observe data every Δ_n unit of time and let Δ_n go to zero in the limit. For any process Z, let $\Delta_i^n Z = Z_{i\Delta_n} - Z_{(i-1)\Delta_n} = \int_{(i-1)\Delta_n}^{i\Delta_n} dZ_t$. For simplicity, we will

denote $Z_{i\Delta_n}$ by Z_i. We use the symbol $\xrightarrow{\mathcal{L}\text{-}s}$ to denote stable convergence in law. We say a constant a *absolute constant* if it does not depend on any pointwise DGP. Let $\mathbf{I}_d$ be the $d \times d$ dimensional identity matrix. For a matrix $\mathbf{A}$, we use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to respectively denote its smallest and largest eigenvalues. In addition, let $\|\mathbf{A}\| := \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, and $\|\mathbf{A}\|_\infty = \max_{ij}|(\mathbf{A})_{ij}|$. In addition, we shall achieve inferences uniformly valid over a large class of data generating process $\mathcal{P}$. For a random sequence X_n, we write $X_n \asymp O_P(a_n)$ if $X_n = O_P(a_n)$ and $a_n/X_n = O_P(1)$.

2 Heuristic Discussions on the Uniformity Issue

2.1 A discrete-time unconditional model

While this paper studies continuous-time models, in this section we heuristically discuss the issue we encounter on the uniform inference using a discrete-time, unconditional factor model with observed factors. Consider the following discrete-time one-factor model:

$$y_{mt} = [\mathbf{x}_m'\boldsymbol{\theta} + \boldsymbol{\gamma}_m]\mathbf{f}_t + u_{mt}, \quad m = 1, ..., p; \quad t = 1, ..., T. \tag{2.1}$$

Here $(y_{mt}, \mathbf{f}_t, \mathbf{x}_m)$ are observable and of finite dimensions, in particular for ease of presentation, $\dim(\mathbf{f}_t) = \dim(\boldsymbol{\gamma}_m) = 1$. The goal is to make inference about $\boldsymbol{\theta}$, which is a $d \times 1$ vector. We assume that $\mathbb{E}(\boldsymbol{\gamma}_m|\mathbf{X}) = 0$, $\mathbb{E}(u_{mt}|\mathbf{X}, \mathbf{F}, \boldsymbol{\Gamma}) = 0$, where $(\mathbf{X}, \boldsymbol{\Gamma}) = \{(\mathbf{x}_m, \boldsymbol{\gamma}_m) : m \leq p\}$ and $\mathbf{F} = \{\mathbf{f}_t : t \leq T\}$, and that $\{u_{mt}, \boldsymbol{\gamma}_m\}$ are cross-sectionally independent across $m \leq p$.

A natural estimator for $\boldsymbol{\theta}$ is based on a combination of cross-sectional and time-series regression:

$$\widehat{\boldsymbol{\theta}} = \mathbf{s}_f^{-1}\mathbf{s}_x^{-1}\frac{1}{pT}\sum_{t=1}^{T}\sum_{m=1}^{p}\mathbf{x}_m\mathbf{f}_t y_{mt},$$

where $\mathbf{s}_f = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t^2$, and $\mathbf{s}_x = \frac{1}{p}\sum_{m=1}^{p}\mathbf{x}_m\mathbf{x}_m'$. Then $\widehat{\boldsymbol{\theta}}$ has the following expansion

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \underbrace{\mathbf{s}_x^{-1}\frac{1}{p}\sum_{m=1}^{p}\mathbf{x}_m\boldsymbol{\gamma}_m}_{\text{(b)}} + \underbrace{\mathbf{s}_f^{-1}\mathbf{s}_x^{-1}\frac{1}{pT}\sum_{t=1}^{T}\sum_{m=1}^{p}\mathbf{x}_m\mathbf{f}_t u_{mt}}_{\text{(a)}}. \tag{2.2}$$

Thus the asymptotic distribution depends on the interplay of two leading terms. Term **(b)**

8

arises from the cross-sectional estimation, which has a rate $O_P(p^{-1/2}\|\mathbf{V}_\gamma\|^{1/2})$, with

$$\mathbf{V}_\gamma = \mathbf{s}_x^{-1}\frac{1}{p}\sum_{m=1}^{p}\mathbf{x}_m\mathbf{x}_m'\operatorname{Var}(\boldsymbol{\gamma}_m|\mathbf{X})\mathbf{s}_x^{-1}, \tag{2.3}$$

where $\operatorname{Var}(\cdot|\mathbf{X})$ denotes the conditional variance given $\mathbf{X}$. Because $\boldsymbol{\gamma}_m$'s are cross-sectionally independent so this term admits a cross-sectional central limit theorem (CLT). In addition, term $(\mathbf{a})$ has a rate $O_P((Tp)^{-1/2})$ because u_{mt}'s are conditionally independent across (i,t).

The first question to address is, what are the final rate of convergence and the limiting distribution? The key to this question is that the strengths of the eigenvalues of $\mathbf{V}_\gamma$ are unknown, and are arbitrarily supported on a compact set $[0,C]$ for $C > 0$. If $\mathbf{V}_\gamma$ is *weak and near boundary* (zero), whose eigenvalues, treated as sequences, decay at rate faster than $O_P(T^{-1})$, then $(\mathbf{a})$ is the dominating term, leading to, for some covariance $\mathbf{v}_1$,

$$\sqrt{Tp}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to^d \mathcal{N}(0, \mathbf{v}_1),$$

whose asymptotic distribution and $\mathbf{v}_1$ are determined by $(\mathbf{a})$. Intuitively, this occurs when the observed characteristics capture almost all the beta fluctuations, leading to a fast rate of convergence. On the other hand, if $\mathbf{V}_\gamma$ is *strong* with all eigenvalues bounded away from zero, $(\mathbf{b})$ becomes the dominating term, and we simply have, for $\mathbf{v}_2 := \operatorname{plim}_{p\to\infty}\mathbf{V}_\gamma$,

$$\sqrt{p}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to^d \mathcal{N}(0, \mathbf{v}_2).$$

In this case, the limiting distribution is determined by the cross-sectional CLT of $(\mathbf{b})$. Intuitively, this means when the idiosyncratic betas have strong cross-sectional variations, time series regression is not helpful to remove their effects on estimating $\boldsymbol{\theta}$, and the cross-sectional projection dominates. This leads to a slower rate of convergence.

Consequently, there is a discontinuity on the limiting distribution of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ when $\mathbf{V}_\gamma$ is near the boundary. In practice, anything in between the above two extreme cases might also happen, leading to an unknown rate of convergence $O_P(a_{pT})$, where $a_{pT} \in [(pT)^{-1/2}, p^{-1/2}]$. This issue is similar to the problems in estimating parameters that are possibly on the boundary of the parameter space (Andrews, 1999; Andrews and Soares, 2010). The problem arises as we do not pretest or know how strong $\boldsymbol{\gamma}$'s cross-sectional variation is, which can vary in a large class of data generating process. Most of the financial econometric studies take the "weak" case as the default assumption (e.g., Ferson and Harvey (1999); Gagliardini

et al. (2016); Connor et al. (2012)), while more recent studies (e.g., Fan et al. (2016b); Kelly et al. (2017); Kim et al. (2018)) provide evidence of the presence of the "strong" case in some sampling periods. Above all, to our best knowledge, all the existing inferences are pointwise, and is not robust to the strength of $\boldsymbol{\gamma}$'s variations. Pointwise inferences, therefore, can be misleading.

2.2 Drawbacks of the usual "plug-in" method

The second question to address is, what is the impact of the unknown rate for $\mathbf{V}_\gamma$ on the inference about $\boldsymbol{\theta}$? The "standard" inference procedure is to plug-in the estimated asymptotic covariances for $\mathbf{V}_\gamma$ and $\mathbf{v}_1$, using their sample analogues. This procedure, however, works only pointwise, and does not provide a uniformly valid confidence interval. To understand the issue, consider the estimation of $\mathbf{V}_\gamma$. If $\boldsymbol{\gamma}_m$ were known, White (1980)'s heteroskedastic covariance estimator can be applied:

$$\widetilde{\mathbf{V}}_\gamma = \mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^p \mathbf{x}_m \mathbf{x}_m' \boldsymbol{\gamma}_m^2 \mathbf{s}_x^{-1}. \tag{2.4}$$

Replacing $\boldsymbol{\gamma}_m$ with its consistent estimator $\widehat{\boldsymbol{\gamma}}_m$, we obtain $\widehat{\mathbf{V}}_\gamma = \mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^p \mathbf{x}_m \mathbf{x}_m' \widehat{\boldsymbol{\gamma}}_m^2 \mathbf{s}_x^{-1}$. Then $\widehat{\mathbf{V}}_\gamma - \mathbf{V}_\gamma$ has a decomposition

$$\underbrace{\mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^p \mathbf{x}_m \mathbf{x}_m' [\widehat{\boldsymbol{\gamma}}_m^2 - \boldsymbol{\gamma}_m^2] \mathbf{s}_x^{-1}}_{\gamma\text{-estimation error}} + \underbrace{\mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^p \mathbf{x}_m \mathbf{x}_m' \left[\boldsymbol{\gamma}_m^2 - \operatorname{Var}(\boldsymbol{\gamma}_m | \mathbf{X}) \right] \mathbf{s}_x^{-1}}_{\text{LLN error}} \tag{2.5}$$

where "LLN error" refers to the error associated with the law of large number. The main issue is that the γ-estimation error cannot be uniformly controlled. Note that in the ideal case where $\boldsymbol{\theta}$ were known, one would estimate $\boldsymbol{\gamma}_m$ from (2.1) by running time series regression of $y_{mt} - \mathbf{x}_m' \boldsymbol{\theta} \mathbf{f}_t$ on $\mathbf{f}_t$ for each fixed $m \le p$. Then $\widehat{\boldsymbol{\gamma}}_m - \boldsymbol{\gamma}_m = \mathbf{r}_m$, where $\mathbf{r}_m := \mathbf{s}_f^{-1} \frac{1}{T} \sum_{t \le T} \mathbf{f}_t u_{mt}$. This leads to

$$\gamma\text{-estimation error} \ge \mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^p \mathbf{x}_m \mathbf{x}_m' \mathbf{r}_m^2 \mathbf{s}_x^{-1} \asymp O_P(T^{-1}).$$

This results in an estimation error $\|\widehat{\mathbf{V}}_\gamma - \mathbf{V}_\gamma\|$ being lower bounded by an order $O_P(T^{-1})$, which is not negligible whenever $\lambda_{\min}(\mathbf{V}_\gamma) = O_P(T^{-1})$ (corresponding to the case of weak γ-signal).

Consequently, the usual plug-in covariance estimator using $\widehat{\mathbf{V}}_\gamma$ would lead to an asymptotically incorrect distribution, and over-coveraging probabilities. On the other hand, ignoring $\mathbf{V}_\gamma$ would result in under-coveraging probabilities when it is present. Hence it is not uniformly valid.[1]

2.3   The cross-sectional bootstrap

To resolve the uniformity issue, we propose to use the cross-sectional bootstrap, which is intuitive and very easy to implement. We simply take random samples with replacement across cross-sectional individuals $\{1, ..., p\}$. Once an individual $l^* \in \{1, ..., p\}$ is sampled, its associated entire time series $\{y_{l^*, t}\}_{t \leq T}$ is sampled. Then we obtain the estimator $\widehat{\boldsymbol{\theta}}^*$ using the bootstrap data. Finally, we calculate the critical value of $\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}$ from a set of bootstrap estimators. This procedure is very simple, but perhaps surprisingly, leads to the desired uniform coverage for $\boldsymbol{\theta}$.

To prove the bootstrap validity, it is essential to show that this procedure directly mimics the cross-sectional variations in $\{\boldsymbol{\gamma}_m\}$. To see this intuitively, we note that we can expand the bootstrap estimator as: for $\mathbf{R}$ as an asymptotically negligible term,

$$\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}} = \mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^{p} (\mathbf{x}_m^* \boldsymbol{\gamma}_m^* - \mathbf{x}_m \boldsymbol{\gamma}_m) + \mathbf{s}_f^{-1} \mathbf{s}_x^{-1} \frac{1}{pT} \sum_{t=1}^{T} \mathbf{f}_t \sum_{m=1}^{p} (\mathbf{x}_m^* u_{it}^* - \mathbf{x}_m u_{mt}) + \mathbf{R},$$

where $\{(\mathbf{x}_m^*, \boldsymbol{\gamma}_m^*) : i \leq p\}$ is a simple random sample from $\{(\mathbf{x}_m, \boldsymbol{\gamma}_m) : i \leq p\}$ with replacement. Then the bootstrap asymptotic variance of $\widehat{\boldsymbol{\theta}}^*$ is analogously $\frac{1}{Tp} \mathbf{v}_1 + \frac{1}{p} \widetilde{\mathbf{V}}_\gamma$, where $\frac{1}{Tp} \mathbf{v}_1$ is the asymptotic variance of term (**a**) in (2.2), and $\widetilde{\mathbf{V}}_\gamma$ is defined in (2.4). The only approximation error for $\mathbf{V}_\gamma$ is therefore:

$$\widetilde{\mathbf{V}}_\gamma - \mathbf{V}_\gamma = \underbrace{\mathbf{s}_x^{-1} \frac{1}{p} \sum_{m=1}^{p} \mathbf{x}_m \mathbf{x}_m' \left[\boldsymbol{\gamma}_m^2 - \mathrm{Var}(\boldsymbol{\gamma}_m | \mathbf{X}) \right] \mathbf{s}_x^{-1}}_{\text{LLN error}}.$$

Consequently, the γ-estimation error component in (2.5) is avoided. The LLN error is of a higher order than $\mathbf{V}_\gamma$, regardless of the signal strength of $\mathbf{V}_\gamma$. For instance, suppose $\boldsymbol{\gamma}_m$ is generated from a rescaled sequence, that is, $\boldsymbol{\gamma}_m = b_T \bar{\boldsymbol{\gamma}}_m$, where $b_T \geq 0$ is a non-random

---

[1]More precisely, when $\mathbf{V}_\gamma = O_P(T^{-1})$, plugging in its consistent estimator over-estimates the asymptotic variance, leading to valid but severely conservative inferences.

arbitrary sequence, and $\bar{\boldsymbol{\gamma}}_m$ satisfies, for $C, c_0 > 0$, almost surely,

$$\lambda_{\min}\left[\frac{1}{p}\sum_{m=1}^{p}\mathbf{x}_m\mathbf{x}_m' \,\mathrm{Var}(\bar{\boldsymbol{\gamma}}_m|\mathbf{X})\right] > c_0, \quad \mathbb{E}(\|\bar{\boldsymbol{\gamma}}_m\|^4|\mathbf{X}) < C.$$

Then the LLN-error $= o_P(1)\mathbf{V}_\gamma$. Hence the approximation error for the asymptotic variance of $\widehat{\boldsymbol{\theta}}$ is negligible regardless of the strength b_T. The bootstrap validity can be achieved.

Andrews (2000) gave a generic counter-example showing that the usual bootstrap is inconsistent when the parameter is near the boundary of its space. So modified inference procedures have been proposed, e.g., see more recently, Ketz (2017, 2018); Pedersen (2017). We note several important differences between our problem and that of the cited literature. The main difference is that the mapping from the underlying DGP to the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ is continuous in our setting. Such a mapping is essentially $\boldsymbol{\Gamma} :\rightarrow \mathcal{M}(\boldsymbol{\Gamma})$,

$$\mathcal{M}(\boldsymbol{\Gamma}) := \mathbf{s}_x^{-1}\underbrace{\frac{1}{p}\sum_{m=1}^{p}\mathbf{x}_m\boldsymbol{\gamma}_m}_{(\mathbf{b})},$$

where $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma}_m\}$. The main reason of achieving a continuous mapping is that we *do not* specify $\mathrm{Var}(\boldsymbol{\gamma}_m|\mathbf{X})$ or $\mathbf{V}_\gamma$ as unknown parameters. All the parameters in our model, $\{(\boldsymbol{\theta}, \boldsymbol{\gamma}_m) : i \leq p\}$, are inside the interior of their parameter space $\Theta_\theta \times \otimes_{m=1}^{p}\Theta_\gamma$, where Θ_θ and Θ_γ are compact subsets respectively in $\mathbb{R}^K$ and $\mathbb{R}$. Therefore when estimating $\boldsymbol{\theta}$, the loss function (e.g., least squares) does not explicitly depend on the unknown "boundary parameter" $\mathrm{Var}(\boldsymbol{\gamma}_m|\mathbf{X})$. In the absence of $\mathrm{Var}(\boldsymbol{\gamma}_m|\mathbf{X})$ in the loss function and with the continuous mapping $\mathcal{M}(\boldsymbol{\Gamma})$, the cross-sectional bootstrap is asymptotically valid. In sharp contrast, in the model of Andrews (2000) and Ketz (2017), $\mathrm{Var}(\boldsymbol{\gamma}_m|\mathbf{X})$ is explicitly modeled as an unknown parameter and appears in the loss function. Therefore, the discontinuity in the asymptotic distribution explicitly arises from the presence of the boundary parameter in their loss functions, but is not due to the issue of interplay among multiple terms in the asymptotic expansion. Another reason why the cross-sectional bootstrap is valid in our context is that, as we illustrated in the above, the effect of estimating $\boldsymbol{\gamma}_m$ can be avoided by resampling the cross-sectional units, and the bootstrap variance directly estimates the asymptotic variance of term (**b**). In contrast, the usual "plug-in" method for the estimated asymptotic variance is not uniformly valid because the estimation error for $\mathrm{Var}(\boldsymbol{\gamma}_m|\mathbf{X})$ dominates the estimand.[2]

---

[2] A possible alternative approach is to employ the thresholding: estimate $\mathbf{V}_\gamma$ using $\widehat{\mathbf{V}}_\gamma 1_{\{\|\widehat{\mathbf{V}}_\gamma\| < c_T \log T\}}$ for

The discussions in this section are based on a very simple setting, assuming that: (1) the model is discrete-time; (2) the betas are time-invariant; (3) factors are directly observable; (4) the parameter of interest is finite dimensional; (5) there are no drifts. In Section 3 we shall formally explore this idea in a continuous-time conditional factor model with drifts using high-frequency data, and separately consider observed and estimated factors. We also extend the regression model to more general high-dimensional moment condition based models in Section 4.

3   The Continuous-Time Factor Model with Characteristics

3.1   The model

Consider a large panel of p time series $\mathbf{Y}_t = (Y_{1t}, \cdots, Y_{pt})$, where p is large. In financial asset pricing applications, $\mathbf{Y}_t$ can be the vector of log-prices of p stocks at time t. We assume $\mathbf{Y} = \{\mathbf{Y}_t\}_{t \geq 0}$ is a multivariate Itô semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. For simplicity, we begin with a model without jumps, as we are interested in the continuous components of log-prices and factors. The jump-robust estimators are given in Section 3.3.3, where we employ a standard procedure to truncate jumps out.[3]

We assume the following (continuous) factor structure:

$$\mathbf{Y}_t = \mathbf{Y}_0 + \int_0^t \boldsymbol{\alpha}_s ds + \int_0^t \boldsymbol{\beta}_s d\mathbf{F}_s + \mathbf{U}_t \tag{3.1}$$

where $\mathbf{Y}_0$ is the starting value of the process $\mathbf{Y}$ at time 0, the drift process $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_s\}_{s \geq 0}$ is an optional $\mathbb{R}^p$-valued process, the factor loading process $\boldsymbol{\beta} = \{\boldsymbol{\beta}_t\}_{t \geq 0}$ is an optional $p \times K$ matrix process. The K dimensional continuous factor process $\mathbf{F}_t$ and the idiosyncratic

---

some sequence $c_T \asymp \min\{T, \sqrt{p}\}^{-1}$, so that $c_T \log T$ "just dominates" $\|\mathbf{V}_\gamma - \widehat{\mathbf{V}}_\gamma\|$. The similar approach has been employed to deal with the distribution discontinuity in the context of random coefficient models, and moment inequalities (e.g., Andrews and Soares (2010)). But in the current context, it has a few drawbacks. One is that it is hard to cover the entire space of all possible sequences for the eigenvalues of $\mathbf{V}_\gamma$. It also leaves a question of choosing the constant in c_T. So we do not pursue it in this paper.

[3]In addition, we assume there is no micro-structure noises. In empirical studies we use data of five-min frequency. In the presence of micro-structure noises, other solutions include sub-sampling (Zhang et al. (2005)), realized kernel (Barndorff-Nielsen et al. (2008)) and pre-averaging (Jacod et al. (2009)). Our main results remain valid when using those more complicated noise-robust estimators.

continuous risk $\mathbf{U}_t$ can be represented as

$$\mathbf{F}_t = \int_0^t \boldsymbol{\alpha}_s^F ds + \int_0^t \boldsymbol{\sigma}_s^F d\mathbf{W}_s^F,$$

$$\mathbf{U}_t = \int_0^t \boldsymbol{\sigma}_s^U d\mathbf{W}_s^U, \tag{3.2}$$

where $\mathbf{W}^U$ and $\mathbf{W}^F$ are two multi-dimensional Brownian motions and are orthogonal to each other (that is, their quadratic covariation is zero), and $\boldsymbol{\alpha}^F = \{\boldsymbol{\alpha}_s^F\}_{s \geq 0}$ is the drift process of the factors. At any time point t, we write $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_{1t}, ..., \boldsymbol{\beta}_{pt})'$ and in general, each $\boldsymbol{\beta}_{lt}$ $(l = 1, \cdots, p)$ is a $K \times 1$ vector of adapted stochastic processes. In the literature, this beta is referred to as the continuous beta (Bollerslev et al. (2016)).

In addition, for each firm $l \leq p$, we observe a set of (possibly) time-varying characteristics:

$$\mathbf{X}_{lt} = (\mathbf{x}_{l,t}', \mathbf{x}_l', \mathbf{x}_t')', \quad l = 1, \cdots, p.$$

We allow the characteristics $\mathbf{X}_{lt}$ to consist of (1) common time-varying characteristics $\mathbf{x}_t$ (such as term and default spread and macroeconomic variables); (2) individual specific characteristics $\mathbf{x}_l$ that are time-invariant over the sampling period $[0, T]$ (such as size and value which change annually); and (3) characteristics $\mathbf{x}_{l,t}$ that are both time-varying and individual specific. Here we present $(\mathbf{x}_{l,t}', \mathbf{x}_l', \mathbf{x}_t')'$ with a bit abuse of notation.

In this paper, we consider the following decomposition of the continuous betas:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l = 1, \cdots, p. \tag{3.3}$$

The overall effect of characteristics on the factor loadings is represented by $\mathbf{g}_{lt} := \mathbf{g}_t(\mathbf{X}_{lt})$, and is called "characteristic beta". Here $\mathbf{g}_t(\cdot)$ is a nonparametric function of macroeconomic and firm variables, possessing less volatile and picks up long run beta fluctuations. On the other hand, $\boldsymbol{\gamma}_{lt}$ represents the remaining time-varying individual factor risks after conditioning on the observed characteristics, and captures high frequency movements in betas. The two components capture different aspects of beta dynamics. For the identification purpose, we assume $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$ for $\mathbf{X}_t = \{\mathbf{X}_{lt} : l \leq p\}$, which well separates the characteristic effects from the remaining effects. The goal is to make uniform inference about $\mathbf{g}_{lt}$ over a broad DGP class, which admits various cross-sectional variations in $(\boldsymbol{\gamma}_{lt}, \mathbf{g}_{lt})$ and dynamics.

We separately study two cases: known and unknown factor cases. In the "known factor case", the high-frequency return data of the common factors are observable, as in the case of

Ait-Sahalia et al. (2014) who constructed Fama-French factors using high-frequency returns. On the other hand, the "unknown factor case" refers to situations in which we do not observe the high-frequency factors, but can estimate them from a large continuous-time panel (up to a locally time-invariant rotation matrix). We shall show that the effect of estimating $\mathbf{F}_t$ leads to an asymptotic bias that needs to be corrected.

3.2 Discussion of the Condition $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$

The condition $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$ serves as a central condition to achieve the identification of the characteristic effects, under which both components in the beta decomposition are well separated. We now discuss the plausibility of this condition and possible approaches to relaxing it. In the presence of omitted characteristics, say $\widetilde{\boldsymbol{\gamma}}_{lt}$, its "explanable component" $\mathbb{E}(\widetilde{\boldsymbol{\gamma}}_{lt}|\mathbf{X}_t)$ is "absorbed" in $\mathbf{g}_{lt}$, and $\boldsymbol{\gamma}_{lt}$ only contains the orthogonal component $\widetilde{\boldsymbol{\gamma}}_{lt} - \mathbb{E}(\widetilde{\boldsymbol{\gamma}}_{lt}|\mathbf{X}_t)$.

In the absence of this condition, identification is lost, and we need further exogenous variables to identify the effect of characteristics. Consider the ideal case that $\boldsymbol{\beta}_{lt}$ were known. Then in (3.3), $\mathbf{X}_{lt}$ is endogenous. To identify $\mathbf{g}_t(\cdot)$, consider an instrumental variable approach: we need to find an exogenous instrumental variable $\mathbf{w}_{lt}$ so that $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{w}_{lt}) = 0$. Define the operator:

$$\mathcal{T} : \mathbf{g} \to \mathbb{E}(\mathbf{g}(\mathbf{X}_{lt})|\mathbf{w}_{lt}).$$

We then have $\mathcal{T}(\mathbf{g}_t) = \mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{w}_{lt})$. The identification of $\mathbf{g}_t$ depends on the invertibility of $\mathcal{T}$, and holds if and only if the conditional distribution of $\mathbf{X}_{lt}|\mathbf{w}_{lt}$ is complete, which is an untestable condition (see, e.g., Newey and Powell (2003)). Suppose $\mathcal{T}$ is indeed invertible, it is well known that estimating $\mathbf{g}_t$ becomes an ill-posed inverse problem, and regularizations are needed, with possibly a very slow rate of convergence. We refer to the literature for related estimation and identification issues: Hall and Horowitz (2005); Darolles et al. (2011); Chen and Pouzo (2012), etc. Therefore, while relaxing the condition $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$ is possible using the nonparametric instrumental variable approach, it requires a very different argument for the identification and estimation. We do not pursue it in this paper.

3.3 Estimation

Suppose there are in total n sample intervals on the interval $[0, T]$, with equal interval length, being Δ_n. For any $t \in [0, T)$, define $I_t^n = \{\lfloor t/\Delta_n \rfloor + 1, \cdots, \lfloor t/\Delta_n \rfloor + k_n\}$, where $\lfloor \cdot \rfloor$ is the floor (greatest integer) function, and k_n is the number of high frequency observations within the

15

window I_t^n. Ignoring the jumps, fix any t, and for all $i \in I_t^n$, by the Burkholder-Davis-Grundy inequality (cf. Chapter 2 of Jacod and Protter (2011)), we have the following approximation for $\Delta_i^n \mathbf{Y} := \mathbf{Y}_{i\Delta_n} - \mathbf{Y}_{(i-1)\Delta_n}$ (which is a $p \times 1$ vector):

$$
\begin{aligned}
\Delta_i^n \mathbf{Y} &= \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\boldsymbol{\alpha}_s ds + \boldsymbol{\beta}_s d\mathbf{F}_s + d\mathbf{U}_s \right) \\
&= \boldsymbol{\alpha}_t \Delta_n + \boldsymbol{\beta}_t \Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U} + o_P(\Delta_n \sqrt{k_n}).
\end{aligned}
\tag{3.4}
$$

where $o_P(\Delta_n \sqrt{k_n})$ holds for each fixed element of $\Delta_i^n \mathbf{Y}$, and is uniform in $i \in I_t^n$. Let $\mathbf{G}_t$ be the $p \times K$ matrix of $\{\mathbf{g}_{lt}\}_{l=1}^p$ and $\boldsymbol{\Gamma}_t$ be the $p \times K$ matrix of $\{\boldsymbol{\gamma}_{lt}\}_{l=1}^p$. We shall then use all observations on I_t^n to estimate $\mathbf{G}_t$. Let $\dim(\boldsymbol{\beta}_{lt}) = K$, and $\dim(\mathbf{X}_{lt}) = K_x$. Throughout the paper, we shall assume $p, k_n \to \infty$, $\Delta_n \to 0$, while K, K_x are fixed constants.

To define the estimators of $\boldsymbol{\Gamma}_t$ and $\mathbf{G}_t$, we introduce the following notation. Let $\boldsymbol{\phi}_{lt} = (\phi_1(\mathbf{X}_{lt}), ..., \phi_J(\mathbf{X}_{lt}))'$ be a $J \times 1$ vector of sieve basis functions of $\mathbf{X}_{lt}$, which can be taken as, e.g., Fourier basis, B-splines, and wavelets. Let $\boldsymbol{\Phi}_t = (\boldsymbol{\phi}_{1t}, ..., \boldsymbol{\phi}_{pt})'$ be the $p \times J$ basis matrix, and define the projection matrix:

$$
\mathbf{P}_t = \boldsymbol{\Phi}_t (\boldsymbol{\Phi}_t' \boldsymbol{\Phi}_t)^{-1} \boldsymbol{\Phi}_t', \quad p \times p.
$$

We subsequently discuss the estimation procedures for the known and unknown factor cases.

3.3.1 The Known Factor Case

In the known factor case, we also observe $\{\Delta_i^n \mathbf{F}\}_{i \in I_t^n}$ in each interval. We use the following two-step estimation:

Step 1. Run time-series regression:

$$
\widehat{\boldsymbol{\beta}}_t := \sum_{i \in I_t^n} \Delta_i^n \mathbf{Y} \, \Delta_i^n \mathbf{F}' \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \, \Delta_i^n \mathbf{F}' \right)^{-1}
\tag{3.5}
$$

Write $\widehat{\boldsymbol{\beta}}_t = (\widehat{\boldsymbol{\beta}}_{1t}, ..., \widehat{\boldsymbol{\beta}}_{pt})'$ to be the $p \times K$ matrix.

Step 2. Run cross-sectional regression:

$$
\begin{aligned}
\widehat{\mathbf{G}}_t &= (\widehat{\mathbf{g}}_{1t}, ..., \widehat{\mathbf{g}}_{pt})' := \mathbf{P}_t \widehat{\boldsymbol{\beta}}_t, \\
\widehat{\boldsymbol{\Gamma}}_t &= (\widehat{\boldsymbol{\gamma}}_{1t}, ..., \widehat{\boldsymbol{\gamma}}_{pt})' := (\mathbf{I}_p - \mathbf{P}_t) \widehat{\boldsymbol{\beta}}_t.
\end{aligned}
$$

Putting together, the estimators can be expressed as:

$$\widehat{\mathbf{G}}_t = \sum_{i \in I_t^n} \mathbf{P}_t \,\Delta_i^n \mathbf{Y}\, \Delta_i^n \mathbf{F}' \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F}\, \Delta_i^n \mathbf{F}' \right)^{-1},$$

$$\widehat{\mathbf{\Gamma}}_t = \sum_{i \in I_t^n} (\mathbf{I}_p - \mathbf{P}_t)\Delta_i^n \mathbf{Y}\, \Delta_i^n \mathbf{F}' \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F}\, \Delta_i^n \mathbf{F}' \right)^{-1}. \tag{3.6}$$

3.3.2 The Unknown Factor Case

When factors are unknown, we first estimate the latent factors and use them in place of $\Delta_i^n \mathbf{F}$ in (3.6). In the continuous-time factor model literature such as Aït-Sahalia and Xiu (2017) and Pelger (2016), these factors are estimated using the regular principal component (PCA) method, extended from Connor and Korajczyk (1986). But different from these works, we employ the PCA on the "projected returns", a method that was proposed by Fan et al. (2016b) in the discrete unconditional factor model. Here we extend the method of this procedure to the continuous-time conditional factor model, and study its asymptotic effect for inference about the characteristic betas.

We use the simplified notation $\mathbf{P}_{i-1} := \mathbf{P}_{(i-1)\Delta_n}$, $\mathbf{G}_{i-1} := \mathbf{G}_{(i-1)\Delta_n}$, and $\mathbf{\Gamma}_{i-1} := \mathbf{\Gamma}_{(i-1)\Delta_n}$. On each local window I_t^n, we can define the following $p \times k_n$ matrix:

$$(\mathbf{P}\Delta^n \mathbf{Y})_t = (\mathbf{P}_{i-1}\Delta_i^n \mathbf{Y} : i \in I_t^n)$$

Define the estimated factors, a $k_n \times K$ matrix

$$\widehat{\Delta^n \mathbf{F}} = (\widehat{\Delta_i^n \mathbf{F}} : i \in I_t^n)', \quad k_n \times K,$$

whose columns equal $\sqrt{k_n \Delta_n}$ times the eigenvectors of the $k_n \times k_n$ matrix $\frac{1}{p k_n \Delta_n}(\mathbf{P}\Delta^n \mathbf{Y})_t'(\mathbf{P}\Delta^n \mathbf{Y})_t$ corresponding to the first K eigenvalues. We then use estimated factors in place of $\{\Delta_i^n \mathbf{F}\}_{i \in I_t^n}$ in (3.6):

$$\widehat{\mathbf{G}}_t^{\text{latent}} = \mathbf{P}_t \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} (\Delta_i^n \mathbf{Y})\widehat{\Delta_i^n \mathbf{F}}',$$

$$\widehat{\mathbf{\Gamma}}_t^{\text{latent}} = (\mathbf{I}_p - \mathbf{P}_t) \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} (\Delta_i^n \mathbf{Y})\widehat{\Delta_i^n \mathbf{F}}' \tag{3.7}$$

17

and note that $\frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \widehat{\Delta_i^n \mathbf{F}} \widehat{\Delta_i^n \mathbf{F}}' = \mathbf{I}_K$. The l-th components $\widehat{\mathbf{g}}_{lt}^{\text{latent}}$ and $\widehat{\boldsymbol{\gamma}}_{lt}^{\text{latent}}$, respectively estimate the characteristic and idiosyncratic betas for the l-th individual. The superscript "latent" indicates that the estimators are defined for the case of latent factors.

We now give an intuitive explanation on the estimated factors. Note that $\mathbf{P}_{i-1}$ is the cross-sectional projection matrix onto the space expanded by the sieve transformations of the characteristics at time $(i-1)\Delta_n$. Apply the projection to the discretized model:

$$\mathbf{P}_{i-1}\Delta_i^n \mathbf{Y} = \mathbf{G}_{i-1}\Delta_i^n \mathbf{F} + \underbrace{\mathbf{P}_{i-1}\boldsymbol{\psi}_{i-1}\Delta_n}_{\text{higher-order term}} + \underbrace{\mathbf{P}_{i-1}\boldsymbol{\Gamma}_{i-1}\Delta_i^n \mathbf{F} + \mathbf{P}_{i-1}\Delta_i^n \mathbf{U}}_{\text{projection errors}}$$
$$+ \underbrace{(\mathbf{P}_{i-1}\mathbf{G}_{i-1} - \mathbf{G}_{i-1})\Delta_i^n \mathbf{F}}_{\text{sieve approximation errors}},$$

where $\boldsymbol{\psi}_{i-1}$ denotes the higher-order drift and diffusion terms. By the identification conditions $\mathbb{E}(\boldsymbol{\Gamma}_t|\mathbf{X}_t) = 0$ and that $\mathbb{E}(\mathbf{U}_{t+s} - \mathbf{U}_t|\mathbf{X}_t) = 0$, the two components of the "projection errors" are projected off, whose rate of decay (after standardized by $\Delta_n^{-1/2}$) is of $O_P(p^{-1/2})$. Ignoring the higher-order term, we have

$$\mathbf{P}_{i-1}\Delta_i^n \mathbf{Y} \approx \mathbf{G}_{i-1}\Delta_i^n \mathbf{F}, \tag{3.8}$$

which is nearly "idiosyncratic-free", and leads to

$$(\mathbf{P}\Delta^n \mathbf{Y})_t'(\mathbf{P}\Delta^n \mathbf{Y})_t \approx \Delta^n \mathbf{F}\mathbf{G}_t'\mathbf{G}_t\Delta^n \mathbf{F}'.$$

Therefore the columns of $\Delta^n \mathbf{F}$ are approximately the eigenvectors of the "idiosyncratic-free" matrix $(\mathbf{P}\Delta^n \mathbf{Y})_t'(\mathbf{P}\Delta^n \mathbf{Y})_t$, up to a rotation. Hence we can estimate them by applying PCA on $(\mathbf{P}\Delta^n \mathbf{Y})_t'(\mathbf{P}\Delta^n \mathbf{Y})_t$.

Furthermore, (3.8) also yields:

$$(\mathbf{I}_p - \mathbf{P}_{i-1})\Delta_i^n \mathbf{Y} \approx \boldsymbol{\Gamma}_{i-1}\Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U}. \tag{3.9}$$

It shows that $\boldsymbol{\Gamma}_{i-1}$ represents the loading on the risk factors of the remaining components of returns, after the characteristic effect is conditioned.

18

3.3.3 Jump-robust estimators

In the general case with jumps, we employ the truncation method to remove those jumps. For notation simplicity, we omit the details and simply assume the jumps are of finite variation. In the known factor case, we replace each $\Delta_i^n \mathbf{Y}$ and $\Delta_i^n \mathbf{F}$ (previously assumed to be continuous) with their truncated versions:

$$\widehat{\mathbf{G}}_t = \sum_{i \in I_t^n} \mathbf{P}_{i-1} \Delta_i^n \mathbf{Y}_{\psi_n^Y} \Delta_i^n \mathbf{F}'_{\psi_n^F} \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F}_{\psi_n^F} \Delta_i^n \mathbf{F}'_{\psi_n^F} \right)^{-1},$$

$$\widehat{\mathbf{\Gamma}}_t = \sum_{i \in I_t^n} (\mathbf{I}_N - \mathbf{P}_{i-1}) \Delta_i^n \mathbf{Y}_{\psi_n^Y} \Delta_i^n \mathbf{F}'_{\psi_n^F} \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F}_{\psi_n^F} \Delta_i^n \mathbf{F}'_{\psi_n^F} \right)^{-1},$$

where $\Delta_i^n Z_{\psi_n^Z} := \Delta_i^n Z_l \, 1_{\{\|\Delta_i^n Z_l\| \leq \psi_n^{Z_l}\}}$ denotes the usual truncated process for the process $\Delta_i^n Z$, with some random sequence ψ_n^Z that depends on certain property of Z and converges in probability to zero as $\Delta_n \to 0$ (e.g., Mancini (2001)).[4] In the unknown factor case, we only need to replace each $\Delta_i^n \mathbf{Y}$ with its corresponding truncated versions.

3.4 Assumptions

We now present the technical assumptions for the asymptotic properties of the estimated characteristic effect $\widehat{\mathbf{g}}_{lt}$ for a fixed $l \leq p$. This subsection presents the required conditions to prove the limiting distribution. In particular, we allow the idiosyncratic components $\{\mathbf{U}_t, \mathbf{\Gamma}_t\}$ to be cross-sectionally weakly dependent. In Section 3.6, we shall present the required conditions for the cross-sectional bootstrap, where we assume them to be cross-sectionally independent.

We assume that the following conditions hold uniformly over a class of DPG's: $\mathbb{P} \in \mathcal{P}$. We apply the standard assumptions to define the stochastic processes as follows (e.g., Protter (2005)).

Assumption 3.1 (Data Generating Process). *(i) The process* $\mathbf{Y}$ *is an Itô semimartingale, whose continuous component is given by* (1.1), *and the continuous component of* $\mathbf{F}$ *and* $\mathbf{U}$ *are given by* (3.2). *The jump components of* $\mathbf{Y}$ *and* $\mathbf{F}$ *are of finite variation.*

---

[4]The common practice is the set $\psi_n^{Z_l} = \alpha_l \Delta_n^\varpi$, where $\varpi \in (0, 1/2)$, $\alpha_l = C(\frac{1}{t}\mathrm{IV}(Z_l)_t)^{1/2}$ with $C = 3, 4$ or 5 and $\mathrm{IV}(Z_l)_t$ is the integrated volatility of Z_l over $[0, t]$.

(ii) Write $\boldsymbol{\theta}_{lt} := (\mathbf{X}_{lt}, \boldsymbol{\gamma}_{lt}, \alpha_{lt}, \sigma_{lt}^U, \boldsymbol{\sigma}_t^F)$. For each $l \leq p$, $\boldsymbol{\theta}_{lt}$ is a multivariate continuous and locally bounded (uniformly in $l \leq p$) Itô semimartingale with the form:

$$\boldsymbol{\theta}_{lt} = \boldsymbol{\theta}_{l0} + \int_0^t \tilde{\boldsymbol{\alpha}}_{ls} ds + \int_0^t \tilde{\boldsymbol{\sigma}}_{ls} d\mathbf{W}_s^l.$$

Here $\{\tilde{\boldsymbol{\alpha}}_{ls}\}_{s \geq 0}$ and $\{\tilde{\boldsymbol{\sigma}}_{ls}\}_{s \geq 0}$ are optional processes and locally bounded uniformly in $l \leq p$. We allow $\mathbf{W}_t^l$ to be correlated with $(\mathbf{W}_t^F, \mathbf{W}_t^U)$ defined in (3.2).

(iii) $\mathbb{E}(\boldsymbol{\gamma}_{lt} | \mathbf{X}_t) = 0$ for all $t \in [0, T), l \leq p$.

(iv) The quadratic covariation $[\mathbf{F}, \mathbf{U}]_t = 0$ for all t.

Assumption 3.2 (Smoothness with respect to time). *There are absolute constants $C, \eta > 0$,*

(i) *Write* $\tilde{\mathbf{g}}(t, \mathbf{x}) := \mathbf{g}_t(\mathbf{x})$. *Then $\tilde{\mathbf{g}}$ and ϕ_j are differentiable, satisfying*

$$\sup_{t \in [0,T], \mathbf{x} \in \mathcal{X}} \left| \frac{\partial \tilde{\mathbf{g}}(t, \mathbf{x})}{\partial t} \right| + \sup_{t \in [0,T], \mathbf{x} \in \mathcal{X}} \left| \frac{\partial \tilde{\mathbf{g}}(t, \mathbf{x})}{\partial \mathbf{x}} \right| + \max_{l \leq N, j \leq J, \mathbf{x} \in \mathcal{X}} \left| \frac{\partial \phi_j(\mathbf{x})}{\partial \mathbf{x}} \right| < C,$$

where $\mathcal{X}$ is the domain of $\{\mathbf{X}_{lt}\}_{l \leq p, t \in [0,T]}$.

(ii) $\max_{t \in [0,T]} \|\mathbf{G}_t - \mathbf{P}_t \mathbf{G}_t\|_\infty \leq CJ^{-\eta}$ *almost surely for $\eta > 0$ so that $J^{-\eta} \sqrt{pk_n} = o(1)$.*

The above assumption ensures that $\mathbf{g}_t(\mathbf{X}_{lt})$ and $\boldsymbol{\phi}_{lt}$ are smooth transformations of $\mathbf{X}_{lt}$. In particular, condition (i) is regarding the smoothness with respect to time, so $\mathbf{g}_t(\mathbf{X}_{lt})$ and $\boldsymbol{\phi}_{lt}$ are also semimartingales; Condition (ii) is regarding the smoothness with respect to cross-sections. It holds if $\max_{t,\mathbf{x}} \|\mathbf{g}_t(\mathbf{x}) - \sum_{j=1}^J \mathbf{b}_j \phi_j(\mathbf{x})\| \leq CJ^{-\eta}$ for some sieve coefficients $\{\mathbf{b}_j : j \leq J\}$.

We now describe the asymptotic variance of $\hat{\mathbf{g}}_{lt}$, and introduce further notation. Let $\mathbf{c}_{FF,t}$ $(K \times K)$ and $\mathbf{c}_{uu,t}$ $(p \times p)$ be the instantaneous quadratic variation processes of $\mathbf{F} = \{\mathbf{F}_t\}_{t \geq 0}$ and $\mathbf{U} = \{\mathbf{U}_t\}_{t \geq 0}$. Let $U_{l,t}$ denote the l-th component of $\mathbf{U}_t$, where $l \leq p$. Also, let

$$h_{t,lm} = \boldsymbol{\phi}_{lt}' (\frac{1}{p} \boldsymbol{\Phi}_t' \boldsymbol{\Phi}_t)^{-1} \boldsymbol{\phi}_{mt}, \quad l, m \leq p.$$

The asymptotic variance depends on the following matrices.

$$
\begin{aligned}
\mathbf{V}_{u,t} &= \lim_{s \to 0} \mathbf{c}_{FF,t}^{-1} \mathrm{Var} \left(\frac{1}{s\sqrt{p}} \sum_{m=1}^p h_{t,ml} (\mathbf{F}_{t+s} - \mathbf{F}_t)(U_{m,t+s} - U_{m,t}) \Big| \mathcal{F}_t \right) \mathbf{c}_{FF,t}^{-1} \\
\mathbf{V}_{\gamma,t} &= \mathrm{Var} \left(\frac{1}{\sqrt{p}} \sum_{m=1}^p h_{t,ml} \boldsymbol{\gamma}_{mt} \Big| \mathbf{X}_t \right).
\end{aligned}
$$

20

Assumption 3.3 (Moment Bounds). *There are absolute constants $c, C > 0$, so that*

(i) $\max_{l \leq p, t \in [0,T]} \mathbb{E} \left\| \mathbf{g}_t(\mathbf{X}_{lt}) \right\|^4 < C$, $\mathbb{E} \left\| \boldsymbol{\gamma}_{lt} \right\|^4 < C$.

(ii) $\max_{t \in [0,T], l, m \leq p} \mathbb{E}\, h_{t,lm}^4 \leq C$.

(iii) $c < \min_{t \leq T} \lambda_{\min}(\frac{1}{p} \boldsymbol{\Phi}_t' \boldsymbol{\Phi}_t) \leq \max_{t \leq T} \lambda_{\max}(\frac{1}{p} \boldsymbol{\Phi}_t' \boldsymbol{\Phi}_t) < C$ *almost surely.*

Assumption 3.4 (Cross-sectional Weak Dependence). *There are absolute constants $c, C > 0$ so that almost surely,*

(i) $\max_{m \leq p} \sup_{t \in [0,T]} \mathrm{Var}(\boldsymbol{\gamma}_{mt} | \mathbf{X}_t) < C$, *and*

$$c < \inf_{t \in [0,T]} \lambda_{\min}(\mathbf{V}_{u,t}) \leq \sup_{t \in [0,T]} \lambda_{\max}(\mathbf{V}_{u,t}) < C, \qquad \lambda_{\max}(\mathbf{V}_{\gamma,t}) \leq C \lambda_{\min}(\mathbf{V}_{\gamma,t}).$$

(ii) If $\lambda_{\min}(\mathbf{V}_{\gamma,t}) \neq 0$, then

$$\mathbf{V}_{\gamma,t}^{-1/2} \frac{1}{\sqrt{p}} \sum_{m=1}^{p} \boldsymbol{\gamma}_{mt} h_{t,ml} \xrightarrow{\mathcal{L}\text{-}s} N(0, \mathbf{I}_K).$$

(iii) Uniformly in $t \in [0,T]$, almost surely, $\|\mathbf{c}_{uu,t}\| < C$, $\lambda_{\min}(\mathbf{c}_{FF,t}) > c$ and

$$\max_{s \in [0, \Delta_n]} \mathbb{E} \left(\left\| \frac{1}{s\sqrt{p}} \sum_{m=1}^{p} h_{t,ml} (\mathbf{F}_{t+s} - \mathbf{F}_t)(U_{m,t+s} - U_{m,t}) \right\|^4 \Big| \mathcal{F}_t \right) < C.$$

The asymptotic distribution of $\widehat{\mathbf{g}}_{lt}$ is jointly determined by two uncorrelated components:

$$\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} = \mathbf{c}_{FF,t}^{-1} \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{U}' \mathbf{P}_{t,l} + \boldsymbol{\Gamma}_t' \mathbf{P}_{t,l} + o_P((k_n p)^{-1/2}), \qquad (3.10)$$

where $\mathbf{P}_{t,l}$ denotes the l-th column of $\mathbf{P}_t$. Assumptions 3.4 condition (ii) requires that $\boldsymbol{\gamma}_{lt}$ be cross-sectionally weakly dependent, and $\boldsymbol{\Gamma}_t' \mathbf{P}_{t,l}$ admits a cross-sectional CLT. In addition, condition (iii) requires $\{U_{mt} : m \leq p\}$ be cross-sectionally weakly dependent. These conditions ensure the asymptotic normality of (3.10).

Assumption 3.5 (For estimated factors). *Define $\boldsymbol{\Sigma}_{G,t} = \frac{1}{p} \mathbf{G}_t' \mathbf{G}_t$. Almost surely:*

(i) $c < \inf_{t \leq T} \lambda_{\min}(\boldsymbol{\Sigma}_{G,t}) \leq \sup_{t \leq T} \lambda_{\max}(\boldsymbol{\Sigma}_{G,t}) < C$ *for absolute constants $c, C > 0$.*

(ii) The eigenvalues of $\boldsymbol{\Sigma}_{G,t}^{1/2} \mathbf{c}_{FF,t} \boldsymbol{\Sigma}_{G,t}^{1/2}$ are distinct: there is an absolute constant $c > 0$, so that the ordered eigenvalues $\lambda_1 < \lambda_2 ... < \lambda_K$ of $\boldsymbol{\Sigma}_{G,t}^{1/2} \mathbf{c}_{FF,t} \boldsymbol{\Sigma}_{G,t}^{1/2}$ satisfy:

$$\lambda_1 > c, \quad \lambda_{j+1} - \lambda_j > c, \quad j = 1, ..., K-1.$$

Assumption 3.5 is similar to the *pervasive condition* in the approximate factor model's literature, which identifies the latent factors (up to a rotation). In particular, condition (i) requires that the betas should nontrivially depend on $\mathbf{X}_t$.

3.5   Asymptotic Distributions and Uniform Bias Correction

We first present the estimated $\mathbf{g}_{lt}$ for a fixed (l, t) when factors are observable.

Theorem 3.1 (known factor case). *Suppose $J, p, k_n \to \infty$ and $\Delta_n \to 0$ satisfy: $J^2 = O(p)$, $k_n p J^{-2\eta} + p k_n^2 \Delta_n = o(1)$. Under Assumptions 3.1-3.4, for a fixed (l, t),*

$$\left(\frac{1}{k_n p} \mathbf{V}_{u,t} + \frac{1}{p} \mathbf{V}_{\gamma,t} \right)^{-1/2} \left(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} \right) \xrightarrow{\mathcal{L}\text{-s}} N(0, \mathbf{I}_K).$$

When the factors are latent and estimated, $\widehat{\mathbf{g}}_{lt}$ consistently estimates a rotated $\mathbf{g}_{lt}$. Up to the rotation, the asymptotic variance is identical to that of the known factor case. However, the effect of estimating the factors gives rise to a bias term. Let $\widehat{\mathbf{V}}_t$ be a $k_n \times k_n$ diagonal matrix consisting of the first K eigenvalues of $\frac{1}{p k_n \Delta_n} (\mathbf{P} \Delta^n \mathbf{Y})'_t (\mathbf{P} \Delta^n \mathbf{Y})_t$. Let

$$\mathbf{M}_t = \frac{1}{k_n \Delta_n \sqrt{p}} \sum_{i \in I_t} \widehat{\mathbf{V}}_t^{-1} \widehat{\Delta_i^n \mathbf{F}} \, \Delta_i^n \mathbf{F}' \, \boldsymbol{\beta}'_{(i-1)\Delta_n} \mathbf{P}_{i-1}$$

$$\text{BIAS}_g = \mathbf{M}_t \frac{1}{k_n \sqrt{p}} \sum_{i \in I_t} \mathbf{P}_{i-1} \, \mathbf{c}_{uu,(i-1)\Delta_n} \, \mathbf{P}_{i-1,l}.$$

We have the following theorem.

Theorem 3.2 (unknown factor case). *In addition to the assumptions for Theorem 3.1, assume Assumption 3.5 and $k_n J = o(p^2)$. Then there is a $K \times K$ rotation matrix $\boldsymbol{\Upsilon}_{nt}$,*

$$\boldsymbol{\Upsilon}_{nt}^{'-1/2} \left(\frac{1}{k_n p} \mathbf{V}_{u,t} + \frac{1}{p} \mathbf{V}_{\gamma,t} \right)^{-1/2} \boldsymbol{\Upsilon}_{nt}^{-1/2} \left(\widehat{\mathbf{g}}_{lt}^{\text{latent}} - \boldsymbol{\Upsilon}_{nt} \mathbf{g}_{lt} - \text{BIAS}_g \right) \xrightarrow{\mathcal{L}\text{-s}} N(0, \mathbf{I}_K).$$

We make several remarks.

Remark 3.1. If $\|\mathbf{V}_{\gamma,t}\| = o_P(k_n^{-1})$, then the rate of convergence is $O_P((k_n p)^{-1/2})$, and

$$\mathbf{V}_{u,t}^{-1/2} \sqrt{k_n p} \left(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} \right) \xrightarrow{\mathcal{L}\text{-s}} N(0, \mathbf{I}_K).$$

Intuitively, this occurs when the cross-sectional variation of $\boldsymbol{\gamma}_{lt}$ is weak. As a result, the observed characteristics capture most of the beta fluctuations in $\boldsymbol{\beta}_t$, leading to a fast rate of convergence. On the other hand, if $\lambda_{\min}(\mathbf{V}_{\gamma,t}) \gg k_n^{-1}$,

$$\mathbf{V}_{\gamma,t}^{-1/2} \sqrt{p} \, (\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}) \quad \xrightarrow{\mathcal{L}\text{-}s} \quad N(0, \mathbf{I}_K).$$

In particular, the rate of convergence is $O_P(p^{-1/2})$ if the eigenvalues of $\mathbf{V}_{\gamma,t}$ are bounded away from zero, corresponding to the case of strong cross-sectional variations in $\boldsymbol{\Gamma}_t$. Intuitively, this means when idiosyncratic betas have strong cross-sectional variations, time-series regression in step 1 is not informative to estimating $\mathbf{g}_{lt}$, and the main statistical error arises from the step 2 cross-sectional regression. This leads to a slower rate of convergence.

Remark 3.2. As in the known factor case, the two sources of the randomness $\frac{1}{k_n p}\mathbf{V}_{u,t}$ and $\frac{1}{p}\mathbf{V}_{\gamma,t}$ jointly determine the rate of convergence and the limiting distribution of $\widehat{\mathbf{g}}_{lt}^{\text{latent}}$. In addition, estimating the latent factors leads to a bias term, whose order is $O_P(p^{-3/2})$. If $\|\mathbf{V}_{\gamma,t}\| = o_P(k_n^{-1})$,

$$(\boldsymbol{\Upsilon}_{nt}\mathbf{V}_{u,t}\boldsymbol{\Upsilon}_{nt}')^{-1/2}\sqrt{k_n p} \, \left(\widehat{\mathbf{g}}_{lt}^{\text{latent}} - \boldsymbol{\Upsilon}_{nt}\mathbf{g}_{lt} - \text{BIAS}_d\right) \xrightarrow{\mathcal{L}\text{-}s} N(0, \mathbf{I}_K).$$

But if $\lambda_{\min}(\mathbf{V}_{\gamma,t}) \gg \max\{k_n^{-1}, p^{-2}\}$, then $\widehat{\mathbf{g}}_{lt}^{\text{latent}}$ is asymptotically unbiased:

$$(\boldsymbol{\Upsilon}_{nt}\mathbf{V}_{\gamma,t}\boldsymbol{\Upsilon}_{nt}')^{-1/2}\sqrt{p} \, \left(\widehat{\mathbf{g}}_{lt}^{\text{latent}} - \boldsymbol{\Upsilon}_{nt}\mathbf{g}_{lt}\right) \quad \xrightarrow{\mathcal{L}\text{-}s} \quad N(0, \mathbf{I}_K).$$

Therefore when the signal from $\boldsymbol{\gamma}_{lt}$ is sufficiently strong, the rate of convergence slows down, and dominates the bias arising from the effect of estimating factors. However, as the magnitude of the eigenvalues of $\mathbf{V}_{\gamma,t}$ is unknown and may change over time, we nevertheless need a bias-correction procedure.

Generally, while $\|\mathbf{V}_{\gamma,t}\| = o_P(k_n^{-1})$ and $\lambda_{\min}(\mathbf{V}_{\gamma,t}) \gg k_n^{-1}$ are two special cases, we do not know the actual strength of the eigenvalues of $\mathbf{V}_{\gamma,t}$ in practice. In fact, its eigenvalues can be any sequences in a large range, resulting in an unknown rate of convergence for $(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt})$ between $O_P((k_n p)^{-1/2})$ and $O_P(p^{-1/2})$. This calls for a need of uniform inference. We shall rely on the cross-sectional bootstrap, as we formally present in the next subsection.

We now derive a bias-corrected spot estimated $\mathbf{g}_{lt}$ in the case of estimated factors. The bias correction is valid uniformly over various signal strengths. Note that in BIAS_g, $\mathbf{M}_t$ can be naturally estimated by $\widehat{\mathbf{M}}_t = \frac{1}{\sqrt{p}}\widehat{\mathbf{V}}_t^{-1}\widehat{\mathbf{G}}_t'$. The major challenge arises in estimating the

$p \times p$ quadratic variation $\mathbf{c}_{uu,t}$, which is high-dimensional when p is large. We consider two cases for the bias correction.

CASE I: cross-sectionally uncorrelated

When $\{\Delta_i^n U_1, \cdots, \Delta_i^n U_p\}$ are cross-sectionally uncorrelated, the $\mathcal{F}_{i-1}$ conditional quadratic variation $\mathbf{c}_{uu,(i-1)\Delta_n}$ is a diagonal matrix. Let $\widehat{\Delta_i^n \mathbf{U}} = \Delta_i^n \mathbf{Y} - (\widehat{\mathbf{G}}_{i-1} + \widehat{\mathbf{\Gamma}}_{i-1})\Delta_i^n \mathbf{F}$. Apply White (1980)'s covariance estimator using the residuals:

$$\widehat{\mathrm{BIAS}}_g = \widehat{\mathbf{M}}_t \frac{1}{k_n \Delta_n \sqrt{p}} \sum_{i \in I_t^n} \mathbf{P}_{i-1} \operatorname{diag}\{\widehat{\Delta_i^n \mathbf{U}} \widehat{\Delta_i^n \mathbf{U}}'\} \mathbf{P}_{i-1,l}$$

CASE II: cross-sectionally weakly correlated (sparse)

In this case $\mathbf{c}_{uu,(i-1)\Delta_n}$ is no longer diagonal. We shall assume it is a sparse covariance matrix, in the sense that many of its off-diagonal entries are zero or nearly so. Then the thresholding estimator of Fan et al. (2013) can be applied, yielding a nearly $\min\{k_n, p\}^{1/2}$-consistent sparse covariance estimator $\widehat{\mathbf{c}}_{uu,t}$. More specifically, let s_{dl} be the (d, l)-th element of $\frac{1}{\Delta_n k_n} \sum_{i \in I_t^n} \widehat{\Delta_i^n \mathbf{U}} \widehat{\Delta_i^n \mathbf{U}}'$. Let the (d, l)-th entry of the estimated covariance be:

$$(\widehat{\mathbf{c}}_{uu,t})_{dl} = \begin{cases} s_{dd}, & \text{if } d = l, \\ \operatorname{th}(s_{dl}) 1_{\{|s_{dl}| > \varrho_{dl}\}} & \text{if } d \neq l, \end{cases}$$

where $\operatorname{th}(\cdot)$ is a thresholding function, whose typical choices are the hard-thresholding and soft-thresholding. Here the threshold value $\varrho_{dl} = \bar{C}(s_{dd} s_{ll})^{1/2} \omega_{np}$ for some constant $\bar{C} > 0$, with $\omega_{np} = \sqrt{\frac{\log p}{k_n}} + \frac{J}{p} \max_{j,d} \frac{1}{J} \|\phi(\mathbf{x}_{jd})\|^2 \sqrt{\log J}$.[5] Estimate the bias by:

$$\widehat{\mathrm{BIAS}}_g = \widehat{\mathbf{M}}_t \frac{1}{k_n \sqrt{p}} \sum_{i \in I_t^n} \mathbf{P}_{i-1} \widehat{\mathbf{c}}_{uu,t} \mathbf{P}_{i-1,l}.$$

Formally, we have the following theorem for the bias-corrected estimator.

Theorem 3.3 (Bias correction). *Suppose $Jk_n = o(p^2)$. Consider CASE I and CASE II for estimated factors. In particular, assume that for $\max_{i \in I_t^n} \|\widehat{\mathbf{c}}_{u,t} - \mathbf{c}_{u,i}\| = o_P(\sqrt{\frac{p}{Jk_n}})$[6] in CASE II. Define the bias-corrected characteristic beta estimator $\widetilde{\mathbf{g}}_{lt}^{\mathrm{latent}} = \widehat{\mathbf{g}}_{lt}^{\mathrm{latent}} - \widehat{\mathrm{BIAS}}_g$. Under*

---

[5]Hard-threolding takes $\operatorname{th}(s_{dl}) = s_{dl}$, while soft-thresholding takes $\operatorname{th}(s_{dl}) = \operatorname{sgn}(s_{dl})(|s_{dl}| - \varrho_{dl})$. We shall justify the choice of ω_{np} in Section E.2.5. In addition, the choice of the constant $\bar{C}$ can be either guided using cross-validation, or simply a constant near one. For returns of S&P 500, the rule of thumb choice $\bar{C} = 0.5$ empirically works very well. We refer to Fan et al. (2013) for more discussions on thresholding.

[6]We shall verify this condition in Section E.2.5 for sparse covariance estimators.

Assumptions 3.1-3.5, we have

$$\mathbf{\Upsilon}_{nt}^{'-1/2} \left(\frac{1}{k_n p} \mathbf{V}_{u,t} + \frac{1}{p} \mathbf{V}_{\gamma,t} \right)^{-1/2} \mathbf{\Upsilon}_{nt}^{-1/2} \left(\widetilde{\mathbf{g}}_{lt}^{\text{latent}} - \mathbf{\Upsilon}_{nt} \mathbf{g}_{lt} \right) \xrightarrow{\mathcal{L}\text{-}s} N(0, \mathbf{I}_K).$$

3.6 Uniform Confidence Intervals Using Cross-Sectional Bootstrap

3.6.1 Independent cross-sectional bootstrap

As we explained in Section 2, the major difficulty is to handle $\mathbf{\Gamma}_t' \mathbf{P}_{t,l} = \frac{1}{p} \sum_{m=1}^{p} \boldsymbol{\gamma}_{mt} h_{t,ml}$ in the asymptotic expansion of $\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}$, which contributes to the asymptotic variance by $\frac{1}{p} \mathbf{V}_{\gamma,t}$. If we use the standard "plug-in" method to estimate $\mathbf{V}_{\gamma,t}$, we would have to introduce a γ-estimation error

$$\frac{1}{p} \sum_{m=1}^{p} h_{t,ml}^2 \left[\widehat{\boldsymbol{\gamma}}_{mt} \widehat{\boldsymbol{\gamma}}_{mt}' - \boldsymbol{\gamma}_{mt} \boldsymbol{\gamma}_{mt}' \right],$$

which is not negligible when $\mathbf{V}_{\gamma,t}$ is near zero. Instead, we propose to use cross-sectional bootstrap to achieve uniform inference, and let the bootstrap distribution "mimic" cross-sectional variations.

Let $\mathcal{M} := \{m_1, ..., m_p\}$ be a simple random sample with replacement from $\{1, ..., p\}$, and we always fix $m_l = l$ when we are interested $\mathbf{g}_{lt}$ for the l-th specific individual. As we do not need to mimic the time series variations, so for each sampled index $m \in \mathcal{M}$, the entire time series $\{\Delta_i^n Y_m, i \in I_t^n\}$ and $\{\mathbf{X}_{m,i}, i \in I_t^n\}$ are kept. Therefore, we independently resample the time series and obtain the bootstrap data: $\{\Delta_i^n Y_m, i \in I_t^n\}_{m \in \mathcal{M}}$ and $\{\mathbf{X}_{mi} : i \in I_t^n\}_{m \in \mathcal{M}}$. In addition, we also keep the entire time series $\{\Delta_i^n \mathbf{F}\}$ in the case of known factors, and $\{\widehat{\Delta_i^n \mathbf{F}}\}$ in the case of unknown factors. The effect of estimating $\Delta_i^n \mathbf{F}$ does not play a role in the cross-sectional variations. Hence we do not re-estimate the factors in each bootstrapped sample.

Let $\Delta_i^n \mathbf{Y}^* = (\Delta_i^n Y_{m_1}, ..., \Delta_i^n Y_{m_p})'$, $\mathbf{\Phi}_t^* = (\boldsymbol{\phi}_{m_1,t}, ..., \boldsymbol{\phi}_{m_p,t})'$ and $\mathbf{P}_t^* = \mathbf{\Phi}_t^* (\mathbf{\Phi}_t^{*'} \mathbf{\Phi}_t^*)^{-1} \mathbf{\Phi}_t^{*'}$. Let $\mathbf{P}_{t,l}^*$ be the l-th column of $\mathbf{P}_t^*$. Define

$$\begin{aligned}
\widehat{\mathbf{g}}_{lt}^* &= \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{F}' \right)^{-1} \sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{t,l}^*, && \text{when factors are known} \\
\widehat{\mathbf{g}}_{lt}^{* \, \text{latent}} &= \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \widehat{\Delta_i^n \mathbf{F}} \, \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{t,l}^*, && \text{when factors are estimated} && (3.11)
\end{aligned}$$

When $\mathbf{g}_{lt}$ is multidimensional, it is easier to present the confidence interval for a lin-

ear transformation $\mathbf{v}'\mathbf{g}_{lt}$. The following algorithm summarizes the steps for computing the confidence intervals.

Algorithm 3.1. Compute the confidence interval for $\mathbf{v}'\mathbf{g}_{lt}$ (or $\mathbf{v}'\mathbf{\Upsilon}_{nt}\mathbf{g}_{lt}$ in the estimated factor case) as follows.

Step 1. Take a simple random sample $\{m_1, ..., m_p\}$ with replacement from $\{1, ..., p\}$. Fix $m_l = l$. Obtain $\Delta_i^n \mathbf{Y}^* = (\Delta_i^n Y_{m_1}, ..., \Delta_i^n Y_{m_p} : i \in I_t^n)'$ and $\mathbf{P}_t^*$.

Step 2. Compute $\widehat{\mathbf{g}}_{lt}^*$, or $\widehat{\mathbf{g}}_{lt}^{*\,\text{latent}}$ in the case of estimated factors, as in (3.11).

Step 3. Repeat Step 1-2 for B times and obtain either $\{\widehat{\mathbf{g}}_{lt}^{*b}\}_{b \leq B}$ or $\{\widehat{\mathbf{g}}_{lt}^{*\,\text{latent},b}\}_{b \leq B}$, depending on whether factors are observable. For the confidence level $1 - \tau$, let q_τ be the $1 - \tau$ bootstrap quantile of $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}|\}_{b \leq B}$; or q_τ^{latent} as the quantile of $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*\,\text{latent},b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}^{\text{latent}}|\}_{b \leq B}$.

Step 4. Compute the confidence interval as:

$$
\begin{aligned}
CI_{nt,\tau} &= [\mathbf{v}'\widehat{\mathbf{g}}_{lt} - q_\tau, \mathbf{v}'\widehat{\mathbf{g}}_{lt} + q_\tau], \\
(\text{or } CI_{nt,\tau}^{\text{latent}} &= [\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{\text{latent}} - \mathbf{v}'\widehat{\text{BIAS}}_g - q_\tau^{\text{latent}}, \mathbf{v}'\widehat{\mathbf{g}}_{lt}^{\text{latent}} - \mathbf{v}'\widehat{\text{BIAS}}_g + q_\tau^{\text{latent}}] \quad).
\end{aligned}
$$

We need the following conditions for the bootstrap validity.

Assumption 3.6. *Conditionally on $\mathbf{X}_t$, $\{(\boldsymbol{\gamma}_{mt}, U_{m,t+s} - U_{mt}) : m \leq p\}$ are cross-sectionally independent, for each $t \in [0, T]$ and $s \in [0, \Delta_n]$.*

Assumption 3.7. *There are absolute constants $C, \eta > 0$,*
(i) Almost surely,

$$
\frac{\frac{1}{p}\sum_{m=1}^p h_{t,mm}^2 h_{t,ll}^2 \mathbb{E}(\|\boldsymbol{\gamma}_{mt}\|^4 | \mathbf{X}_t)}{\lambda_{\min}^2(\frac{1}{p}\sum_{m=1}^p h_{t,ml}^2 \text{Var}(\boldsymbol{\gamma}_{mt}|\mathbf{X}_t))} < C.
$$

If the denominator equals zero, then the above ratio is defined as zero.
(ii) Almost surely in the bootstrap sampling space, $\sup_{t \in [0,T]} \|\mathbf{G}_t^* - \mathbf{P}_t^*\mathbf{G}_t^*\|_\infty \leq CJ^{-\eta}$, where $\mathbf{G}_t^* = (\mathbf{g}_{m_1,t}, ..., \mathbf{g}_{m_p,t})'$.

Note that the bootstrap asymptotic variance for $\widehat{\mathbf{g}}_{lt}^*$ is analogously $\frac{1}{k_n p}\mathbf{V}_{u,t} + \frac{1}{p}\widetilde{\mathbf{V}}_{\gamma,t}$, where $\widetilde{\mathbf{V}}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^p h_{i,ml}^2 \boldsymbol{\gamma}_{mt}\boldsymbol{\gamma}_{mt}'$. The only approximation error for the $\mathbf{V}_{\gamma,t}$ part is the "law of large numbers error":

$$
\widetilde{\mathbf{V}}_{\gamma,t} - \mathbf{V}_{\gamma,t} = \frac{1}{p}\underbrace{\sum_{m=1}^p h_{t,ml}^2 \left[\boldsymbol{\gamma}_{mt}\boldsymbol{\gamma}_{mt}' - \text{Var}(\boldsymbol{\gamma}_{mt}|\mathbf{X}_t)\right]}_{\text{LLN error}}.
$$

26

Consequently, the γ-estimation error component is avoided. This forms the foundation of the bootstrap asymptotic validity. The moment bound in Assumption 3.7 (i) on $\mathbf{\Gamma}_t$ is used to ensure that the LLN-error is negligible regardless of the strength of $\mathbf{V}_{\gamma,t}$.

Theorem 3.4 (Uniformly valid confidence intervals). *Let $\mathcal{P}$ be the collection of all data generating processes $\mathbb{P}$ for which Assumptions 3.1 - 3.7 hold. Then for any fixed vector $\mathbf{v} \in \mathbb{R}^K \backslash \{0\}$ such that $\|\mathbf{v}\| > c > 0$, for each fixed $l \leq p, t \in [0, T]$,*

$$
\begin{aligned}
\text{known factor case:} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \left| \mathbb{P}(\mathbf{v}' \mathbf{g}_{lt} \in CI_{nt,\tau}) - (1 - \tau) \right| \to 0 \\
\text{unknown factor case:} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \left| \mathbb{P}(\mathbf{v}' \mathbf{\Upsilon}_{nt} \mathbf{g}_{lt} \in CI_{nt,\tau}^{\text{latent}}) - (1 - \tau) \right| \to 0.
\end{aligned}
$$

3.6.2 Block cross-sectional bootstrap

Theorem 3.4 requires the idiosyncratic components $\{\{\Delta_i^n U_m\}_{t \in [0,T]}, \{\boldsymbol{\gamma}_{mt}\}_{t \in [0,T]}\}_{m \leq p}$ be cross-sectionally independent. We can relax this assumption to allow for cross-sectionally block-dependent idiosyncratic components when these blocks are known, and rely on block cross-sectional bootstraps. More specifically, suppose the cross-sectional index set has a non-overlapping partition $\{1, ..., p\} = B_1 \cup ... \cup B_H$, with $H \to \infty$, and the cardinality of each "block" B_h is finite: $\max_{h \leq H} |B_h|_0 = O(1)$. For a fixed $k \leq K$. We assume:

(i) Individuals' block mememberships are known; (that is, $\{B_h\}_{h \leq H}$ are known)[7],

(ii) $\text{Cov}(\Delta_t^n U_{l_1}, \Delta_t^n U_{l_2}) = 0$ and $\text{Cov}(\gamma_{mt,k_1}, \gamma_{mt,k_2}) = 0$ for any $t \in [0, T]$ if the two individuals (l_1, l_2) belong to different blocks, for any $k_1, k_2 \leq K$; $\gamma_{mt,k}$ denotes the k-th element of $\boldsymbol{\gamma}_{mt}$.

Therefore, conditionally on the factors, individuals are possibly correlated only within the same block. Empirically, the assumption of known blocks of finite size can be supported by setting blocks as industry sectors, which is motivated from the economic intuition that firms within similar industries are expected to have higher correlations conditioning on the factors, e.g., Aït-Sahalia and Xiu (2017). These blocks form a natural basis for the application of non-overlapping block bootstraps. Suppose we are interested in the inference for $\mathbf{g}_{lt}$ for a specific firm $l \leq p$, and it is known that $l \in B_{h_0}$ for a particular $h_0 \leq H$. Set l as the first element of B_{h_0}. We employ the block-bootstrap on the cross-sectional units, and can proceed with the following algorithm.

Algorithm 3.2. Compute the confidence interval for $\mathbf{v}' \mathbf{g}_{lt}$ as follows.

---

[7]In the more complicated case where blocks are unknown, one could first apply a block-thresholding method to estimate the cross-sectional covariance matrix of $\Delta_t^n \mathbf{U}$, to consistently recover the block structures first. See, e.g., Cai et al. (2012).

Step 1: Fix $B_1^* = B_{h_0}$. Take a simple random sample $\{B_2^*, ..., B_H^*\}$ with replacement from $\{B_1, ..., B_H\}$.

Step 2: For each sampled block B_h^*, all the individuals in the block are sampled associated with their entire time series. We obtain

$$\begin{aligned}
\Delta_i^n \mathbf{Y}^* &= (\Delta_i^n Y_m : m \in B_h^*, h \le H), \quad i \in I_t^n \\
\mathbf{\Phi}_t^* &= (\phi_{mt} : m \in B_h^*, h \le H).
\end{aligned}$$

and $\mathbf{P}_t^* = \mathbf{\Phi}_t^* (\mathbf{\Phi}_t^{*'} \mathbf{\Phi}_t^*)^{-1} \mathbf{\Phi}_t^{*'}$.

Step 3: Define

$$\begin{aligned}
\widehat{\mathbf{g}}_{lt}^* &= \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{F}' \right)^{-1} \sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{t,l}^* \quad \text{when factors are known} \\
\widehat{\mathbf{g}}_{lt}^{* \, \text{latent}} &= \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \widehat{\Delta_i^n \mathbf{F}} \, \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{t,l}^* \quad \text{when factors are estimated.}
\end{aligned}$$

Step 4: Repeat Steps 1-3 for B times, and obtain either $\{\widehat{\mathbf{g}}_{lt}^{*b}\}_{b \le B}$ or $\{\widehat{\mathbf{g}}_{lt}^{* \, \text{latent}, b}\}_{b \le B}$, depending on whether factors are observable. Let q_τ (or q_τ^{latent}) be the $1 - \tau$ bootstrap quantile of $\{|\mathbf{v}' \widehat{\mathbf{g}}_{lt}^{*b} - \mathbf{v}' \widehat{\mathbf{g}}_{lt}|\}_{b \le B}$ (or $\{|\mathbf{v}' \widehat{\mathbf{g}}_{lt}^{* \, \text{latent}, b} - \mathbf{v}' \widehat{\mathbf{g}}_{lt}^{\text{latent}}|\}_{b \le B}$). Compute the confidence interval for $\mathbf{v}' \mathbf{g}_{lt}$ (or $\mathbf{v}' \mathbf{\Upsilon}_{nt} \mathbf{g}_{lt}$ in the estimated factor case) as:

$$\begin{aligned}
CI_{nt,\tau} &= [\mathbf{v}' \widehat{\mathbf{g}}_{lt} - q_\tau, \mathbf{v}' \widehat{\mathbf{g}}_{lt} + q_\tau], \\
(\text{or } CI_{nt,\tau}^{\text{latent}} &= [\mathbf{v}' \widehat{\mathbf{g}}_{lt}^{\text{latent}} - \mathbf{v}' \widehat{\text{BIAS}}_g - q_\tau^{\text{latent}}, \mathbf{v}' \widehat{\mathbf{g}}_{lt}^{\text{latent}} - \mathbf{v}' \widehat{\text{BIAS}}_g + q_\tau^{\text{latent}}] \quad).
\end{aligned}$$

The proof of the first-order bootstrap validity is very similar to that of Theorem 3.4, building on the results of the validity of block-bootstrap (Andrews, 2004; Lahiri, 1999). We omit the formal proof for technical simplicity.

3.7   Inference about the integrated characteristic-beta

We can also estimate the long-run characteristic effect: $\int_0^T \mathbf{g}_{lt} dt$ in the case of known factors.[8] Using the standard overlapping spot estimates (see, e.g. Jacod and Rosenbaum (2013)), we

---

[8]Due to the rotation discrepancy, estimating the long-run effect with unknown factors subjects to the issue of time-varying rotations, and is a very challenging problem in the presence of time-varying betas, and we shall leave it for the future research.

estimate it by

$$\widehat{\int_0^T \mathbf{g}_{lt}dt} := \sum_{t=1}^{[T/\Delta_n]-k_n} \widehat{\mathbf{g}}_{lt}\Delta_n.$$

Asymptotically, it has the following decomposition:

$$\widehat{\int_0^T \mathbf{g}_{lt}dt} - \int_0^T \mathbf{g}_{lt}dt = \sum_{t=1}^{[T/\Delta_n]-k_n} \Delta_n\mathbf{\Gamma}_t'\mathbf{P}_{t,l} + \sum_{t=1}^{[T/\Delta_n]-k_n} \Delta_n\mathbf{c}_{FF,t}^{-1}\frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\Delta_i^n\mathbf{U}'\mathbf{P}_{t,l}$$

$$+o_P\left(\sqrt{\frac{\Delta_n}{p}}\right).$$

As in the spot estimation case, the asymptotic expansion also admits two components: the effect of estimating integrated betas and the effect of cross-sectional estimation. The final limiting distribution is determined by the interplay of both terms. Due to the unknown signal strength of the idiosyncratic beta $\sum_{t=1}^{[T/\Delta_n]-k_n}\Delta_n\mathbf{\Gamma}_t'\mathbf{P}_{t,l}$, we still rely on the bootstrap to make uniform inference about $\mathbf{v}'\int_0^T \mathbf{g}_{lt}dt$ for any specific vector $\mathbf{v}$ of interest.

Denote by $\widehat{\int_0^T \mathbf{g}_{lt}dt}^{*b} = \sum_{t=1}^{[T/\Delta_n]-k_n}\widehat{\mathbf{g}}_{lt}^{*b}\Delta_n$ as the bootstrap estimator in the b-th generated sample. Let $\widetilde{q}_\tau$ be the $1-\tau$ bootstrap quantile of $\{|\mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt}^{*b} - \mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt}|\}_{b\leq B}$. The confidence interval for $\mathbf{v}'\int_0^T \mathbf{g}_{lt}dt$ is given by

$$\widetilde{CI_{n,\tau}} = \left[\mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt} - \widetilde{q}_\tau, \mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt} + \widetilde{q}_\tau\right].$$

The following condition plays a similar role as that of Assumption 3.7, but for estimating the integrated characteristic effects.

Assumption 3.8. *There is an absolute constant $C > 0$, almost surely,*

$$\frac{1}{p}\sum_{m=1}^p \frac{1}{[T/\Delta_n]}\sum_{t=1}^{[T/\Delta_n]-k_n}\mathbb{E}(\|\boldsymbol{\gamma}_{mt}\|^4 h_{t,mm}^2 h_{t,ll}^2|\mathbf{X}_t) \leq C\lambda_{\min}^2\left[\frac{1}{p}\sum_{m=1}^p \text{Var}(\frac{1}{[T/\Delta_n]}\sum_{t=1}^{[T/\Delta_n]-k_n}\boldsymbol{\gamma}_{mt}h_{t,ml}|\mathbf{X}_t)\right].$$

Theorem 3.5 (long-run g). *Consider the known factor case Let $\mathcal{P}$ be the collection of all data generating processes $\mathbb{P}$ for which Assumptions 3.1-3.8 hold. Then for any fixed vector*

$\mathbf{v} \in \mathbb{R}^K \backslash \{0\}$ *such that* $\|\mathbf{v}\| > c > 0$*, for each fixed* $l \leq p$,

$$\sup_{\mathbb{P} \in \mathcal{P}} \left| \mathbb{P} \left(\mathbf{v}' \int_0^T \mathbf{g}_{lt} dt \in \widetilde{CI_{n,\tau}} \right) - (1 - \tau) \right| \to 0.$$

4 Models of Many Continuous-Time Moment Conditions

4.1 The model

We consider a more general continuous-time model with linear moment conditions. Suppose we observe data that are discretized realizations from a continuous-time stochastic process $\{\mathbf{Z}_t = (\mathbf{Z}_{1t}, ..., \mathbf{Z}_{pt}) : t \in [0, T]\}$. We assume $\mathbf{Z} = \{\mathbf{Z}_t\}_{t \geq 0}$ be a multivariate Itô semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. Igonoring jumps, we have

$$\mathbf{Z}_{lt} = \int_0^t \boldsymbol{\alpha}_{ls}^Z ds + \int_0^t \boldsymbol{\sigma}_{ls}^Z d\mathbf{W}_{ls}^Z, \quad l = 1, ..., p, \quad \forall t \in [0, T]$$

where $\{\mathbf{W}_{lt}^Z\}_{t \leq 0}$ is a Brownian motion, and $\{\boldsymbol{\alpha}_{lt}^Z\}_{t \geq 0}$ is the drift process. For each $t \in [0, T]$, let the parameter $\boldsymbol{\beta}_{lt}$ be identified by the following moment condition:

$$\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) = 0, \quad l = 1, ..., p, \tag{4.1}$$

where $\mathbf{c}_{z,lt} = d[\mathbf{Z}_l, \mathbf{Z}_l]_t / dt$ is the instantaneous quadratic variation process of $\mathbf{Z}_l = \{\mathbf{Z}_{lt}\}_{t \geq 0}$, and $\Psi_l(\boldsymbol{\beta}, \mathbf{c})$ is a known function linear with respect to $\boldsymbol{\beta}$ and continuously differentiable with respect to $\mathbf{c}$. In addition, we assume that $\boldsymbol{\beta}_{lt}$ depends on a set of characteristics $\mathbf{X}_{lt}$, so has the following decomposition:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l = 1, \cdots, p, \tag{4.2}$$

where $\mathbb{E}(\boldsymbol{\gamma}_{lt} | \mathbf{X}_t) = 0$. Here $\mathbf{X}_t$ a set of (possibly) time-varying characteristics. The effect of characteristics on $\boldsymbol{\beta}_{lt}$ is represented by $\mathbf{g}_t(\mathbf{X}_{lt})$. The goal is to make uniform inference about $\mathbf{g}_t(\mathbf{X}_{lt})$ that allows the cross-sectional variations of $\boldsymbol{\gamma}_{lt}$ to be possibly arbitrarily close to the "boundary".

Many applications in economics and finance give rise to the continuous-time *many linear moment conditions* as in (4.1) and (4.2), as we now illustrate with a few examples.

Example 4.1 (Multivariate regression models). Consider a system of multivariate continuous-time linear regression models with individual-specific regressors, e.g., Barndorff-Nielsen and Shephard (2004); Mykland et al. (2006); Kalnina (2012); Li et al. (2017):

$$Y_{lt} = Y_{l0} + \int_0^t \alpha_{ls} ds + \int_0^t \boldsymbol{\beta}_{ls}' d\mathbf{F}_{ls} + U_{lt} : \quad l \leq p \tag{4.3}$$

where we observe realizations of $(Y_{lt}, \mathbf{F}_{lt})$ for $l = 1, ..., p$. Assume that the quadratic covariation of U_{lt} and $\mathbf{F}_{lt}$ be zero, then we have

$$\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) = \mathbf{c}_{FF,lt} \boldsymbol{\beta}_{lt} - \mathbf{c}_{YF,lt}$$

with $(\mathbf{c}_{FF,lt}, \mathbf{c}_{YF,lt})$ respectively being the quadratic variation of the individual-specific regressor $\mathbf{F}_{lt}$, and the quadratic covariation between Y_{lt} and $\mathbf{F}_{lt}$. Here we consider a high-dimensional system where $p \to \infty$ fast. In particular, this model admits the linear factor model described in Section 3 as a special case, by setting $\mathbf{F}_{lt} = \mathbf{F}_t$ as a common factor, for all $l \leq p$.

Example 4.2 (Idioscynractic variance models). Consider a continuous-time factor model

$$Y_{lt} = Y_{l0} + \int_0^t \alpha_{ls} ds + \int_0^t \boldsymbol{\theta}_{ls}' d\mathbf{F}_s + U_{lt} : \quad l \leq p$$

but we use $\boldsymbol{\theta}_{lt}$ to denote the factor "betas". Under the condition that U_{lt} and $\mathbf{F}_t$ are two orthogonal processes, the factor betas can be identified as $\boldsymbol{\theta}_{lt} = \mathbf{c}_{FF,t}^{-1} \mathbf{c}_{YF,lt}$, and yield the following covariance decomposition:

$$c_{YY,lt} = \boldsymbol{\theta}_{lt}' \mathbf{c}_{FF,t} \boldsymbol{\theta}_{lt} + c_{uu,lt} = \mathbf{c}_{YF,lt}' \mathbf{c}_{FF,t}^{-1} \mathbf{c}_{YF,lt} + c_{uu,lt} \tag{4.4}$$

where $c_{YY,lt}$ and $c_{uu,lt}$ respectively denote the quadratic variations of Y_{lt} and U_{lt}. Consider a model where firms' idiosyncratic variances possess a high degree of comovement: the idiosyncratic variance $c_{uu,lt}$ is proportional to the market factor, with a firm-specific scalar coefficient $\boldsymbol{\beta}_{lt} \in \mathbb{R}$. That is,

$$c_{uu,lt} = \boldsymbol{\beta}_{lt} c_{FF1,t} + \bar{c}_l, \tag{4.5}$$

where $c_{FF1,t}$ denotes the quadratic variation of the market factor, and to identify $\boldsymbol{\beta}_{lt}$ we assume that $\bar{c}_l$ is a known intercept in (4.5). The model is in spirit similar to the moment

31

restriction on the quadratic variations studied by Li et al. (2016)[9]. Here $\boldsymbol{\beta}_{lt}$ is referred to as the "idiosyncratic variance beta", and can be used to measure the relation between the idiosyncratic volatility for stock l and the market factor (e.g., Ang et al. (2009); Herskovic et al. (2016)). As evidenced by Herskovic et al. (2016), households may face common fluctuations in their idiosyncratic variances. Substituting (4.4) to (4.5), we obtain a moment condition, with

$$\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) = \boldsymbol{\beta}_{lt} c_{FF1,t} + \bar{c}_l + \mathbf{c}'_{YF,lt} \mathbf{c}^{-1}_{FF,t} \mathbf{c}_{YF,lt} - c_{YY,lt}.$$

Then the beta decomposition (4.2) expresses the idiosyncratic variance beta into the sum of a characteristic-dependent component and an orthogonal component. For a given individual l, we are interested in testing whether its idiosyncratic variance beta can be explained by its characteristics, which corresponds to the null hypothesis that $\mathbf{g}_t(\mathbf{X}_{lt}) = 0$ almost surely.

4.2 Estimation

We apply the continuous-time generalized methods of moments (GMM) based on the following moment conditions:

$$\begin{aligned}
\Psi_l(\boldsymbol{\beta}_{lt}, \mathbf{c}_{z,lt}) &= 0, \quad \forall t \in [0, T], \\
\mathbb{E}(\boldsymbol{\beta}_{lt} | \mathbf{X}_{lt}) &= \mathbf{g}_t(\mathbf{X}_{lt}), \quad l = 1, ..., p.
\end{aligned} \tag{4.6}$$

We shall first assume the underlying $\mathbf{Z}_{lt}$ is continuous. Let $\dim(\mathbf{Z}_{lt}) = K_z$ and $\dim(\Psi_l) = K_\psi$. We shall assume K_z and K_ψ are fixed constants, and $K_\psi \geq K$, so $\boldsymbol{\beta}_{lt}$ is possibly overidentified.

Ignoring the jumps, over the i-th sampling interval, we have the following discrete time observation:

$$\triangle^n_i \mathbf{Z}_l = \mathbf{Z}_{l,i\Delta_n} - \mathbf{Z}_{l,(i-1)\Delta_n} = \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\boldsymbol{\alpha}^Z_{ls} ds + \boldsymbol{\sigma}^Z_{ls} d\mathbf{W}^Z_{ls} \right) \tag{4.7}$$

---

[9]Li et al. (2016) allow unknown intercepts in (4.5). They are able to identify $\boldsymbol{\beta}_l$ as it is assumed to be time-invariant over $t \in [0, T]$ and can be estimated using all sampling intervals on the entire time span. In contrast, we allow $\boldsymbol{\beta}_{lt}$ to be time-varying, so to ensure $c_{uu,lt} > 0$, we set $\bar{c}_l = \bar{\beta}_l \bar{c}_F$, where $\bar{\beta}_l$ and $\bar{c}_F$ are respectively known upper bounds for $\max_{t \in [0,T]} |\boldsymbol{\beta}_{lt}|$ and $\max_{t \in [0,T]} |c_{FF1,t}|$. Alternative, one could allow an unknown intercept $c_{uu,lt} = \boldsymbol{\beta}_{lt} c_{FF1,t} + c_{lt}$ and set another identification restriction, e.g., $a\boldsymbol{\beta}_{lt} = c_{lt}$ for some known $a \in \mathbb{R}$.

Let $\widehat{\mathbf{c}}_{z,lt}$ be the sample quadratic variation of $\mathbf{Z}_{lt}$ on window I_t^n:

$$\widehat{\mathbf{c}}_{z,lt} = \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \Delta_i^n \mathbf{Z}_l \, \Delta_i^n \mathbf{Z}_l'.$$

We employ a simple two-step estimation:

Step 1. Let $\widehat{\boldsymbol{\beta}}_{lt}$ be the solution that satisfies:

$$\widehat{\boldsymbol{\beta}}_{lt} := \arg\min_{\boldsymbol{\beta}} \Psi_l(\boldsymbol{\beta}, \widehat{\mathbf{c}}_{z,lt})' \boldsymbol{\Omega}_{lt} \Psi_l(\boldsymbol{\beta}, \widehat{\mathbf{c}}_{z,lt}), \quad l = 1, ..., p. \tag{4.8}$$

Let $\widehat{\boldsymbol{\beta}}_t = (\widehat{\boldsymbol{\beta}}_{1t}, ..., \widehat{\boldsymbol{\beta}}_{pt})'$ be the $p \times K$ matrix. Here $\boldsymbol{\Omega}_{lt}$ is a known positive definite weight matrix. In the general case with jumps, we solve

$$\widehat{\boldsymbol{\beta}}_{lt} := \arg\min_{\boldsymbol{\beta}} \Psi_l(\boldsymbol{\beta}, \widehat{\mathbf{c}}_{z,lt}^{\psi})' \boldsymbol{\Omega}_{lt} \Psi_l(\boldsymbol{\beta}, \widehat{\mathbf{c}}_{z,lt}^{\psi}), \quad l = 1, ..., p.$$

where we replace each $\Delta_i^n \mathbf{Z}_l$ with their truncated versions: $\widehat{\mathbf{c}}_{z,lt}^{\psi} = \frac{1}{k_n} \sum_{i \in I_t^n} \Delta_i^n \mathbf{Z}_l^{\psi_n} \Delta_i^n \mathbf{Z}_l^{\psi_n'}$. Here $\Delta_i^n \mathbf{Z}_l^{\psi_n} := \Delta_i^n \mathbf{Z}_l \mathbf{1}_{\{\|\Delta_i^n \mathbf{Z}_l\| \le \psi_n\}}$ denotes the usual truncated process for the process $\Delta_i^n \mathbf{Z}_l$, with some random sequence ψ_n that converges in probability to zero as $\Delta_n \to 0$.

Step 2. Define

$$\begin{aligned} \widehat{\mathbf{G}}_t &= (\widehat{\mathbf{g}}_{1t}, ..., \widehat{\mathbf{g}}_{pt})' := \mathbf{P}_t \widehat{\boldsymbol{\beta}}_t, \\ \widehat{\boldsymbol{\Gamma}}_t &= (\widehat{\boldsymbol{\gamma}}_{1t}, ..., \widehat{\boldsymbol{\gamma}}_{pt})' := (\mathbf{I}_p - \mathbf{P}_t) \widehat{\boldsymbol{\beta}}_t. \end{aligned}$$

Therefore, the estimator (3.6) in the linear factor model with known factors is a special case of the two-step estimator presented here, while the estimator of the unknown factor case can also be considered as a special case, which replaces $\Delta_i^n \mathbf{Z}_l$ in the definition of $\widehat{\mathbf{c}}_{z,lt}$ with the "estimated regressors".[10]

4.3 General theory for two-step estimations

Our estimator belongs to the general class of two-step GMM estimators, as previously discussed in Newey (1994); Chen et al. (2003); Chen and Liao (2015); Chernozhukov et al.

---

[10]The general asymptotic theories presented in this section focus on the "known regressor case". In the case of "estimated regressors", corresponding to the latent factor case in the linear factor model, the estimated regressors should be estimated separately and plugged in, as we did for the unknown factors in Section 3. The effect of estimating the unknown regressors might introduce additional biases that need be corrected. For simplicity, we do not cover this case in the general GMM framework.

(2016): a nuisance parameter is estimated in the first step, and substituted in the second step estimation. As is shown by these authors, when the second set of moment conditions is not "Neyman orthogonal" with respect to $\boldsymbol{\beta}_{lt}$ (roughly speaking, its "directional derivative" with respect to $\boldsymbol{\beta}_{lt}$ is nonzero), the first-step estimation error $\widehat{\boldsymbol{\beta}}_{lt} - \boldsymbol{\beta}_{lt}$ is not negligible, and plays a leading role in the asymptotic distribution for $\widehat{\mathbf{g}}_{lt}$.

In the current large panel context with many moment conditions ($l = 1, ..., p$), indeed the second moment condition for $\mathbf{g}_{lt}$, given by $\mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{X}_t) = \mathbf{g}_{lt}$, is not Neyman orthogonal with respect to $\boldsymbol{\beta}_{lt}$. However, two new phenomena are present here. To see this, let $\nabla_\beta \Psi_m(\mathbf{c})$ be the $K \times K$ gradient matrix of $\Psi_m(\boldsymbol{\beta}, \mathbf{c})$ with respect to $\boldsymbol{\beta}$, which does not depend on $\boldsymbol{\beta}$ due to the linearity. Write

$$\mathbf{A}_{mt}(\mathbf{c}) := (\nabla_\beta \Psi_m(\mathbf{c})' \boldsymbol{\Omega}_{mt} \nabla_\beta \Psi_m(\mathbf{c}))^{-1} \nabla_\beta \Psi_m(\mathbf{c})' \boldsymbol{\Omega}_{mt}.$$

The first-order condition of (4.8) leads to, the $p \times K$ matrix $\widehat{\boldsymbol{\beta}}_t$ satisfies:

$$\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t \approx - \begin{pmatrix} \Psi_1(\boldsymbol{\beta}_{1t}, \widehat{\mathbf{c}}_{z,1t})' \mathbf{A}_{1t}(\mathbf{c}_{z,1t})' \\ \vdots \\ \Psi_p(\boldsymbol{\beta}_{pt}, \widehat{\mathbf{c}}_{z,pt})' \mathbf{A}_{pt}(\mathbf{c}_{z,pt})' \end{pmatrix}$$

where $\boldsymbol{\beta}_{lt}$ and $\mathbf{c}_{z,lt}$ are evaluated at the true values. Then applying Step 2 of the estimation, combined with the delta-method, yields, for each $l \leq p$,

$$\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} = \underbrace{- \frac{1}{p} \sum_{m=1}^p \mathbf{A}_{mt}(\mathbf{c}_{z,mt}) \nabla_c \Psi_m(\mathbf{c}_{z,mt}) \text{vec}(\widehat{\mathbf{c}}_{z,mt} - \mathbf{c}_{z,mt}) h_{t,ml}}_{\text{first-step effect:}(\mathbf{a})} + \underbrace{\boldsymbol{\Gamma}_t' \mathbf{P}_{t,l}}_{\text{second-step effect:}(\mathbf{b})} + o_P(\frac{1}{\sqrt{k_n p}})$$

where $\nabla_c \Psi_m(\mathbf{c}_{z,mt})$ is the gradient of $\Psi_m(\boldsymbol{\beta}_{mt}, \mathbf{c}_{z,mt})$ with respect to $\text{vec}(\mathbf{c}_{z,mt})$.

The first new phenomenon is that the first-step effect $(\mathbf{a})$ is a cross-sectional average of the estimation errors for $\{\widehat{\boldsymbol{\beta}}_{mt} - \boldsymbol{\beta}_{mt} : m = 1, ..., p\}$, whose rate of convergence is $O_P((k_n p)^{-1/2})$ and is in fact negligible when $\text{Var}(\boldsymbol{\Gamma}_t|\mathbf{X}_{lt})$ is bounded away from zero, even though the second moment condition is not Neyman orthogonal. This is essentially due to the fact that in the presence of many moment conditions, the first-step effect can be "cross-sectionally averaged out" and thus can be dominated by the second-step effects when the latter has a slower rate of convergence.

The second new phenomenon, which is also the unique feature in our model, is that

whether the second-step moment condition $\mathbf{E}(\boldsymbol{\beta}_{lt}|\mathbf{X}_t) = \mathbf{g}_{lt}$ is "noise-free" is unknown, in the sense that it is possible that the magnitude of the eigenvalues of $\mathrm{Var}(\boldsymbol{\Gamma}_t|\mathbf{X}_{lt})$ can be arbitrary in their parameter space $[0, C]$, and may vary across time. Especially, when $\mathrm{Var}(\boldsymbol{\Gamma}_t|\mathbf{X}_t)$ is near the boundary of the parameter space (zero), which occurs when $\mathbf{X}_t$ has nearly full explanatory power on $\boldsymbol{\beta}_{lt}$, the first-step effect becomes the only leading term in the asymptotic expansion.

The aforementioned two unique features of our model call for a different asymptotic analysis for the two-step estimator considered here, and similar to the discussions in the linear factor model, lead to an unknown rate of convergence and a discontinuity in the limiting distribution.

To formally present our theory, we assume that the following conditions hold uniformly over a class of DPG's: $\mathbb{P} \in \mathcal{P}$.

Assumption 4.1. *(i) For each $l \leq p$, $\mathbf{Z}_{lt}$ is a multivariate Itô semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, whose continuous component is given by*

$$\mathbf{Z}_{lt}^c = \int_0^t \boldsymbol{\alpha}_{ls}^Z ds + \int_0^t \boldsymbol{\sigma}_{ls}^Z d\mathbf{W}_{ls}^Z, \quad \forall t \in [0, T]$$

where $\{\mathbf{W}_{lt}^Z\}_{t \leq 0}$ is a Brownian motion, and $\{\boldsymbol{\alpha}_{lt}^Z\}_{t \geq 0}$ is the drift process. The jump components of $\mathbf{Z}_{lt}$ are of finite variation uniformly over $l \leq p$.

(ii) Write $\boldsymbol{\theta}_{lt} := (\mathbf{X}_{lt}, \boldsymbol{\gamma}_{lt}, \boldsymbol{\alpha}_{lt}^Z, \boldsymbol{\sigma}_{lt}^Z)$. For each $l \leq p$, $\boldsymbol{\theta}_{lt}$ is a multivariate continuous and locally bounded (uniform in $l \leq p$) Itô semimartingale with the form:

$$\boldsymbol{\theta}_{lt} = \boldsymbol{\theta}_{l0} + \int_0^t \tilde{\boldsymbol{\alpha}}_{ls} ds + \int_0^t \tilde{\boldsymbol{\sigma}}_{ls} d\mathbf{W}_s^l.$$

Here $\{\tilde{\boldsymbol{\alpha}}_{ls}\}_{s \geq 0}$ and $\{\tilde{\boldsymbol{\sigma}}_{ls}\}_{s \geq 0}$ are optional processes and locally bounded uniformly in $l \leq p$. We allow $\mathbf{W}_t^l$ to be correlated with $\mathbf{W}_{lt}^Z$.

(iii) $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_t) = 0$ for all $t \in [0, T), l \leq p$.

Next, for each $m \leq p$, $\Psi_m(\boldsymbol{\beta}, \mathbf{c})$ is linear in $\boldsymbol{\beta}$, so $\nabla_\beta \Psi(\mathbf{c})$ does not depend on $\boldsymbol{\beta}$. When $\boldsymbol{\beta}_{mt}$ is evaluated at the true value, we simply write $\Psi_m(\mathbf{c}) := \Psi_m(\boldsymbol{\beta}_{mt}, \mathbf{c})$ and $\nabla_c \Psi_m(\mathbf{c}) := \nabla_c \Psi_m(\boldsymbol{\beta}_{mt}, \mathbf{c})$ to explicitly make them as functions of $\mathbf{c}$. Here ∇_β, ∇_c respectively denote the gradient operators with respect to $\boldsymbol{\beta}$ and $\mathrm{vec}(\mathbf{c})$.

Assumption 4.2. *(i) $\Psi_m(\cdot, \cdot)$ is twice continuously differentiable for all $m \leq p$.*
(ii) $\max_{m \leq p} \|(\nabla_\beta \Psi_m(\mathbf{c}_{z,mt})' \boldsymbol{\Omega}_{mt} \nabla_\beta \Psi_m(\mathbf{c}_{z,mt}))^{-1}\| = O_P(1)$, $\max_{m \leq p} \|\boldsymbol{\Omega}_{mt}\| = O_P(1)$.

Assumption 4.3. *The sample quadratic variation satisfies: at each fixed $t \in [0, T]$,*

(i) $\quad \dfrac{1}{p} \sum_{m=1}^{p} h_{t,ml} \mathbf{A}_{mt}(\widehat{\mathbf{c}}_{z,mt})[\Psi_m(\widehat{\mathbf{c}}_{z,mt}) - \Psi_m(\mathbf{c}_{z,mt}) - \nabla_c \Psi_m(\mathbf{c}_{z,mt}) vec(\widehat{\mathbf{c}}_{z,mt} - \mathbf{c}_{z,mt})] = o_P \left(\dfrac{1}{\sqrt{pk_n}} \right)$

(ii) $\quad \dfrac{1}{p} \sum_{m=1}^{p} h_{t,ml} [\mathbf{A}_{mt}(\widehat{\mathbf{c}}_{z,mt}) - \mathbf{A}_{mt}(\mathbf{c}_{z,mt})] \nabla_c \Psi_m(\mathbf{c}_{z,mt}) vec(\widehat{\mathbf{c}}_{z,mt} - \mathbf{c}_{z,mt}) = o_P \left(\dfrac{1}{\sqrt{pk_n}} \right).$

Assumption 4.3 is a high-level assumption on the effect of estimating a large number of quadratic variations. It requires two fundamental conditions. The first line requires that the moment function $\Psi_m(\mathbf{c})$ should be well approximated by linear functions locally; the second line additionally requires the smoothness of $\mathbf{A}_{mt}(\mathbf{c})$. A loose upper bound using $\frac{1}{p} \sum_{m=1}^{p} \|\widehat{\mathbf{c}}_{z,mt} - \mathbf{c}_{z,mt}\|^2 = O_P(k_n^{-1})$ based on the Cauchy-Schwarz inequality can be achieved, so both conditions can be simply verified so long as $p = o(k_n)$. On the other hand, sharper upper bounds can also be achieved to allow for a much larger p, by noting that both conditions are regarding high-dimensional cross-sectional averages of the effects of $\widehat{\mathbf{c}}_{z,mt} - \mathbf{c}_{z,mt}$, and intuitively, these should be "averaged out" as $p \to \infty$, and can be verified case-by-case for specific models. For instance, Fan et al. (2015) and Bai and Liao (2017) provide technical arguments to verify (ii) when $\Psi_m(\mathbf{c})$ is linear in $\mathbf{c}^{-1}$. In fact, when proving Theorem 3.1 in the linear factor model, we verify these conditions directly and apply our general result theorem 4.1.

We now describe the asymptotic variance of $\widehat{\mathbf{g}}_{lt}$. For $m = 1, ..., p$, let

$$\boldsymbol{\xi}_{mt,s} := \mathbf{A}_{mt}(\mathbf{c}_{z,mt}) \nabla_c \Psi_m(\mathbf{c}_{z,mt}) vec \left(\frac{1}{s}(\mathbf{Z}_{m,t+s} - \mathbf{Z}_{mt})(\mathbf{Z}_{m,t+s} - \mathbf{Z}_{mt})' - \mathbf{c}_{z,mt} \right).$$

The asymptotic distribution of $\widehat{\mathbf{g}}_{lt}$ is jointly determined by $\boldsymbol{\xi}_{mt,s}$ and $\boldsymbol{\gamma}_{mt}$. Let

$$
\begin{aligned}
\mathbf{V}_{u,t} &= \lim_{s \to 0} \mathrm{Var} \left(\frac{1}{\sqrt{p}} \sum_{m=1}^{p} h_{t,ml} \boldsymbol{\xi}_{mt,s} \Big| \mathcal{F}_t \right) \\
\mathbf{V}_{\gamma,t} &= \mathrm{Var} \left(\frac{1}{\sqrt{p}} \sum_{m=1}^{p} h_{t,ml} \boldsymbol{\gamma}_{mt} \Big| \mathbf{X}_t \right).
\end{aligned}
\tag{4.9}
$$

Assumption 4.4. *There are absolute constants $c, C > 0$, almost surely,*

(i) $\max_{m \leq p} \sup_{t \in [0,T]} \mathrm{Var}(\boldsymbol{\gamma}_{mt} | \mathbf{X}_t) < C$, *and*

$$c < \inf_{t \in [0,T]} \lambda_{\min}(\mathbf{V}_{u,t}) \leq \sup_{t \in [0,T]} \lambda_{\max}(\mathbf{V}_{u,t}) < C, \qquad \lambda_{\max}(\mathbf{V}_{\gamma,t}) \leq C \lambda_{\min}(\mathbf{V}_{\gamma,t}).$$

(ii) $\sup_{t\in[0,T]} \max_{s\in[0,\Delta_n]} \mathbb{E}(\|\frac{1}{\sqrt{p}}\sum_{m=1}^{p} h_{t,ml}\boldsymbol{\xi}_{mt,s}\|^4|\mathcal{F}_t) = O_P(1)$.

Generally, we have the following theorem.

Theorem 4.1. *Suppose $J, p, k_n \to \infty$ and $\Delta_n \to 0$ satisfy: $J^2 = O(p)$, $k_n p J^{-2\eta} + p k_n^2 \Delta_n = o(1)$. Under Assumptions 3.2, 3.3, 4.1-4.4,*

$$\left(\frac{1}{k_n p}\mathbf{V}_{u,t} + \frac{1}{p}\mathbf{V}_{\gamma,t}\right)^{-1/2} \left(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}\right) \xrightarrow{\mathcal{L}\text{-}s} N(0, \mathbf{I}_K).$$

Therefore, when the strength of $\mathbf{V}_{\gamma,t}$ is unknown, the general two-step GMM estimation in the current context yields an unknown rate of convergence $O_P(a_{np})$, where a_{np} may vary on the range $[(pk_n)^{-1/2}, p^{-1/2}]$. This leads to a discontinuity on its limiting distribution.

Remark 4.1. The weight matrices $\{\boldsymbol{\Omega}_{mt} : m \leq p\}$ are involved in the asymptotic distribution through $\mathbf{V}_{u,t}$. Similar to the usual GMM setting, the optimal weight matrix can be determined by optimizing $\mathbf{V}_{u,t}$, and is given by

$$\boldsymbol{\Omega}_{mt} = \left(\nabla_c\Psi_m(\mathbf{c}_{z,mt})\mathbf{V}_{mt}\nabla_c\Psi_m(\mathbf{c}_{z,mt})'\right)^{-1},$$

where $\mathbf{V}_{mt} = \text{Var}\left[\text{vec}\left(\frac{1}{\Delta_n}\Delta_i^n\mathbf{Z}_m\Delta_i^n\mathbf{Z}_m'\right)\Big|\mathcal{F}_t\right]$. In the case of exact identification, as in the linear factor models, the weight matrix $\boldsymbol{\Omega}_{mt}$ no longer plays a role in the asymptotic distribution.

4.4 Cross-sectional Bootstrap

We extend the cross-sectional bootstrap described in Section 3 to the more general context, in order to achieve the uniform inference. Importantly, note that the main issue that causes the unknown limiting distribution is from $\boldsymbol{\Gamma}_t'\mathbf{P}_{t,l}$ in the expansion of $\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}$. This term comes from the second-step estimation. Hence the first-step regressions for $\boldsymbol{\beta}_t$, which only depends on time-series estimations, is not needed to be repeated in the bootstrap steps.

As such, we directly resample $\{(\widehat{\boldsymbol{\beta}}_{mt}^*, \mathbf{X}_{mt}^*) : m = 1, ..., p\}$ with replacement, where

$$\{\widehat{\boldsymbol{\beta}}_{mt}^*, \mathbf{X}_{mt}^* : m = 1, ..., p\} = \{(\widehat{\boldsymbol{\beta}}_{m_1,t}, \mathbf{X}_{m_1,t}), ..., (\widehat{\boldsymbol{\beta}}_{m_p,t}, \mathbf{X}_{m_p,t})\}$$

and $\widehat{\boldsymbol{\beta}}_{mt}$ is the estimated $\boldsymbol{\beta}_{mt}$ in step 1. Here $\{m_1, ..., m_p\}$ is a simple random sample with replacement from $\{1, ..., p\}$. We then let $\boldsymbol{\Phi}_t^* = (\boldsymbol{\phi}_{m_1,t}, ..., \boldsymbol{\phi}_{m_p,t})'$, $\mathbf{P}_t^* = \boldsymbol{\Phi}_t^*(\boldsymbol{\Phi}_t^{*'}\boldsymbol{\Phi}_t^*)^{-1}\boldsymbol{\Phi}_t^{*'}$, and

$\widehat{\boldsymbol{\beta}}_t^* = (\widehat{\boldsymbol{\beta}}_{m_1,t}^*, ..., \widehat{\boldsymbol{\beta}}_{m_p,t}^*)'$. When we are interested in $\mathbf{g}_{lt}$ for given l, we always fix the index of the l-th resampled cross-sectional unit being l, that is, $m_l = l$. Hence $\widehat{\boldsymbol{\beta}}_{lt}^* = \widehat{\boldsymbol{\beta}}_{lt}$. Define

$$\widehat{\mathbf{G}}_t^* = \mathbf{P}_t^* \widehat{\boldsymbol{\beta}}_t^*$$

and let $\widehat{\mathbf{g}}_{lt}^*$ be the l-th column of $(\widehat{\mathbf{G}}_t^*)'$.

When $\mathbf{g}_{lt}$ is multidimensional, we present the confidence interval for a linear transformation $\mathbf{v}'\mathbf{g}_{lt}$.

Algorithm 4.1. Compute the confidence interval for $\mathbf{v}'\mathbf{g}_{lt}$ as follows.

Step 1. Take a simple random sample $\{m_1, ..., m_p\}$ with replacement from $\{1, ..., p\}$. Fix $m_l = l$.

Step 2. Compute $\widehat{\mathbf{g}}_{lt}^*$ as described above.

Step 3. Repeat Step 1-2 for B times and obtain $\{\widehat{\mathbf{g}}_{lt}^{*b}\}_{b \leq B}$. For a predetermined confidence level $1 - \tau$, let q_τ be the $1 - \tau$ bootstrap quantile of $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}|\}_{b \leq B}$.

Step 4. Compute the confidence interval as:

$$CI_{nt,\tau} = [\mathbf{v}'\widehat{\mathbf{g}}_{lt} - q_\tau, \mathbf{v}'\widehat{\mathbf{g}}_{lt} + q_\tau].$$

For the first-order validity of the bootstrap, we impose a high-level assumption as Assumption 4.3 in the bootstrap sampling space. In this assumption, $\{\mathbf{A}_{mt}^*(\mathbf{c}), \Psi_m^*(\mathbf{c}), \widehat{\mathbf{c}}_{z,mt}^*, \mathbf{c}_{z,mt}^* : m \leq p\}$ denote the bootstrap samples of $\{\mathbf{A}_{mt}(\mathbf{c}), \Psi_m(\mathbf{c}), \widehat{\mathbf{c}}_{z,mt}, \mathbf{c}_{z,mt} : m \leq p\}$; and $h_{t,ml}^* = \boldsymbol{\phi}_{lt}^{*'}(\frac{1}{p}\boldsymbol{\Phi}_t^{*'}\boldsymbol{\Phi}_t^*)^{-1}\boldsymbol{\phi}_{mt}^*$.

Assumption 4.5. *At each fixed $t \in [0, T]$,*

(i) $\quad \dfrac{1}{p}\displaystyle\sum_{m=1}^{p} h_{t,ml}^* \mathbf{A}_{mt}^*(\widehat{\mathbf{c}}_{z,mt}^*)[\Psi_m^*(\widehat{\mathbf{c}}_{z,mt}^*) - \Psi_m^*(\mathbf{c}_{z,mt}^*) - \nabla_c\Psi_m^*(\mathbf{c}_{z,mt}^*)vec(\widehat{\mathbf{c}}_{z,mt}^* - \mathbf{c}_{z,mt}^*)] = o_{P^*}\left(\dfrac{1}{\sqrt{pk_n}}\right)$

(ii) $\quad \dfrac{1}{p}\displaystyle\sum_{m=1}^{p} h_{t,ml}^*[\mathbf{A}_{mt}^*(\widehat{\mathbf{c}}_{z,mt}^*) - \mathbf{A}_{mt}^*(\mathbf{c}_{z,mt}^*)]\nabla_c\Psi_m^*(\mathbf{c}_{z,mt}^*)vec(\widehat{\mathbf{c}}_{z,mt}^* - \mathbf{c}_{z,mt}^*) = o_{P^*}\left(\dfrac{1}{\sqrt{pk_n}}\right).$

Next, recall that $\boldsymbol{\xi}_{mt,s} := \mathbf{A}(\mathbf{c}_{z,mt})\nabla_c\Psi_m(\mathbf{c}_{z,mt})\mathrm{vec}\left(\frac{1}{s}(\mathbf{Z}_{m,t+s} - \mathbf{Z}_{mt})(\mathbf{Z}_{m,t+s} - \mathbf{Z}_{mt})' - \mathbf{c}_{z,mt}\right).$

Assumption 4.6. *For each $t \in [0, T]$ and $s \in [0, \Delta_n]$, there exist $\{\boldsymbol{\xi}_{mt,s}^0 : m \leq p\}$ that satisfies:*

(i) Conditionally on $\mathbf{X}_t$, $\{(\boldsymbol{\gamma}_{mt}, \boldsymbol{\xi}_{mt,s}^0) : m \leq p\}$ are cross-sectionally uncorrelated.

(ii) $\sup_{t \in [0,T]} \max_{s \in [0,\Delta_n]} \frac{1}{p}\sum_{m \leq p} \mathbb{E}(\|\boldsymbol{\xi}_{mt,s} - \boldsymbol{\xi}_{mt,s}^0\|^4|\mathcal{F}_t) = O_P((pk_n)^{-2}).$

We would like to emphasize that Assumption 4.6 does *not* assume that $\{\mathbf{Z}_{mt} : m \leq p\}$ should be cross-sectionally uncorrelated, because in many applications it doe not hold for $\mathbf{Z}_{mt}$, but in fact holds for $\boldsymbol{\xi}_{mt}^0$. In the linear regression model (4.3) for instance, $\mathbf{Z}_{mt} = (Y_{mt}, \mathbf{F}_{mt})$, $m \leq p$, then $\{Y_{mt} : m \leq p\}$ would not be uncorrelated if $\mathbf{F}_{mt}$ contains common regressors (e.g., factors). On the other hand, it can be directly verfied that in this case,

$$\boldsymbol{\xi}_{mt,s}^0 = -\mathbf{c}_{FF,mt}^{-1} \frac{1}{s}(\mathbf{F}_{m,t+s} - \mathbf{F}_{mt})(U_{m,t+s} - U_{mt})$$
$$\boldsymbol{\xi}_{mt,s} = \boldsymbol{\xi}_{mt,s}^0 + \boldsymbol{\delta}_{mt,s},$$

where $\boldsymbol{\delta}_{mt,s}$ satisfies Assumption 4.6 (ii), so is negligible; $\mathbf{c}_{FF,mt}$ denotes the quadratic variation of $\mathbf{F}_{mt}$. In the idiosyncratic volatility model (Example 4.2), it is also straightforward to verify that

$$\boldsymbol{\xi}_{mt,s}^0 = -\mathbf{c}_{FF,t}^{-1}\left(\frac{1}{s}(U_{m,t+s} - U_{mt})^2 - c_{uu,mt}\right)h_{t,ml}.$$

So $\boldsymbol{\xi}_{mt,s}^0$ are cross-sectionally uncorrelated given that $\{U_{m,t+s} - U_{mt} : m \leq p\}$ are cross-sectionally uncorrelated.

Theorem 4.2 (Uniformly valid confidence intervals). *Suppose $J^2 = o(p)$, $J^{-\eta}pk_n + pk_n^2\Delta_n = o(1)$. Let $\mathcal{P}$ be the collection of all data generating processes $\mathbb{P}$ for which Assumptions 3.2, 3.3, 4.1- 4.6 hold. Then for any fixed vector $\mathbf{v} \in \mathbb{R}^K\backslash\{0\}$ such that $\|\mathbf{v}\| > c > 0$, for each fixed $l \leq p, t \in [0,T]$*

$$\sup_{\mathbb{P}\in\mathcal{P}} |\mathbb{P}(\mathbf{v}'\mathbf{g}_{lt} \in CI_{nt,\tau}) - (1-\tau)| \to 0.$$

5   Empirical Studies

5.1   The data

We use the price data of stocks from the S&P 500 index constituents for the period from July 2006 through June 2013. We collect intraday transactions data of each stock from the TAQ database and construct returns every five minutes. We drop the overnight returns for excluding stock splits and dividend issuances, and abnormal prices that bounced back within a few seconds. Stocks with missing price data are also dropped. Therefore there are in total 380 stocks in our dataset. In addition, we construct the Fama-French four factors with five-minute frequency by first generating the five-minute returns of each common stocks on the

NYSE, the AMEX, and the NASDAQ in the CRSP database and then following the method described in Fama and French (1992). These factors are: the market factor (Mkt), the small-minus-big market capitalization (SMB) factor, high-minus-low book to market ratio (HML) factor, and the profitability factor (RMW), the difference between the returns of firms with high and low operating profitability.

We also collect fundamentals of those stocks from the Compustat database over the same period to construct firm characteristics. We consider four characteristics for each stock: size, value, momentum, and volatility as in Connor et al. (2012). The annual size and value characteristic of each stock is the logarithm of the market value and the ratio of the market value to the book value in the previous June respectively. The monthly momentum and volatility characteristic of each stock is the cumulative returns of the last twelve months including the previous month, and the standard deviation of the last twelve months, including previous month respectively.

5.2 Cross-sectional variations

Our estimation is based on the 5-minute frequency, and intervals are taken as daily windows. Hence there are $k_n = 78$ observations for each stock each day. We use the linear sieve basis, which are simply the standardized values of the four characteristics, and estimate the spot characteristic and idiosyncratic beta for each company on each trading day.[11] We divide the assets into three categories: large, medium, and low, based on either the firms' size or the volatility characteristics. Figure 1 plots the cross-sectional average of the characteristic betas corresponding to each of the four factors, classified by either the size or the volatility. Both size and volatility have noticeable effects on at least one of the factor betas. The cross-sectionally averaged characteristic betas for the SMB factor are noticeably different across three size groups, and in the long run, companies with larger size (market value) tend to be less sensitive to the SMB factor than companies with smaller size. As shown in the fifth panel of Figure 1, companies with smaller volatilities tend to be less sensitive to the market factor than companies with larger volatilities. While both phenomena have been documented in the literature, the characteristic betas, however, capture long-run movements in beta driven by structural changes in the economic environment and in firm- or industry-specific conditions, so demonstrate long-run patterns in betas from these figures.

---

[11]We also tried B-splines with degree 3 (Eilers and Marx, 1996), particularly for estimating and plotting the $\mathbf{g}_t(\cdot)$ functions. Results obtained a very similar.

In addition, we also estimate the cross-sectional variations in the idiosyncratic betas, measured by $\frac{1}{|G|} \sum_{j \in G} \widehat{\gamma}_{lt,k}^2$ for $k = $ Mkt, HML, SMB and RMW factors. In the lower panel firms are grouped by volatility: $G \in \{\text{small vol}, \text{medium vol}, \text{large vol}\}$. So the computed measure shows the cross-sectional variations among firms of small, medium and large volatilities. Figure 2 plots $\frac{1}{|G|} \sum_{j \in G} \widehat{\gamma}_{lt,k}^2$ over time. There are substantial differences on the cross-sectional variations among firms with difference sizes. In particular, the strength of $\mathbf{\Gamma}$ is the strongest and also the most volatile across time for firms of large volatilities, is the weakest but least volatile across time for firms of small volatilities. This measures different prediction power of the characteristics on betas among firms of different level of volatilities. Figure 2 also demonstrates that the strength of $\mathbf{\Gamma}$ can be represented by various asymptotic sequences across time, so the uniform inference is very essential.

5.3 Confidence Intervals

We construct 95% construct confidence intervals for each of the firms' characteristic betas on a daily base, and report and compare them among three groups (by either size or volatility). On each trading day we construct the confidence intervals and calculate the proportion of positive/negative significances among firms in each group. Then we average these (cross-sectional) proportions over all days within a fixed year, leading to the "averaged proportion of significance" for each group.

When the groups are formed by size, Table 1 reports the results of 2006, and we find that results of other years (2007 through 2012) demonstrate similar patterns: (1) All stocks have significantly positive characteristic betas loading on the market factor. In fact, most of the characteristic betas for the market factor are larger than one. (2) There is a substantial difference in the characteristic betas on the SMB factor between firms of small/medium size and firms of large size. Only 4.7% of firms of large size have positive significance, but this proportion is as high as 87% for firms of small size. On the other hand, more than fifty percent of firms of large size have negative significance, but there are less than one percent of firms of small size. This shows that the in-firm conditions and characteristics produce a long-run mechanism making small firms positively exposed and large firms negatively exposed to the SMB systematic risk. It becomes more interesting when we compare the results with the proportions of $\mathbf{\Gamma}$ and $\mathbf{\beta}$. We find that for SMB, the proportion of positive $\mathbf{\beta}$ is 37% for large firms, and 71% for small firms, while the proportion of negative $\mathbf{\beta}$ is 62% for large firms, and 28% for small firms. In contrast, these proportions respectively become 51% and

48% for positive $\boldsymbol{\Gamma}$, and 52% for negative $\boldsymbol{\Gamma}$, so the difference among firms of large and small sizes in $\boldsymbol{\Gamma}$ is much less noticeable. This suggests that the characteristic beta is the main driving horse to determine the sign of $\boldsymbol{\beta}$, while the idiosyncratic beta is more related to beta's cross-sectional variations. (3) As the size becomes larger, there is also a decreasing pattern on the negative significance of the HML beta (more noticeable on the SMB betas).

Table 1: Cross-sectional Proportion of significant $\mathbf{G}$ of groups by size, 2006

size	positive significance				negative significance			
	Mkt	HML	SMB	RMW	Mkt	HML	SMB	RMW
small	1	0.261	0.870	0.134	0	0.177	0.004	0.184
medium	1	0.234	0.450	0.162	0	0.215	0.039	0.150
large	1	0.133	0.047	0.154	0	0.229	0.544	0.062

When we group firms by the volatility, however, the pattern demonstrates noticeable variations over years. The results are given in Table 2. Results of 2010 are similar to 2011, and results in 2007 are similar to 2006 so are not presented. Firms with larger volatility tend to be more positively exposed to the HML factors than firms with smaller volatility, who are more negatively exposed to HML. This pattern appears in 2006, 2007, 2010 and 2011, but is reversed during the crisis period in 2008-2009, and European debt crisis 2012.

We now focus on two individual stocks' confidence intervals. We take the two firms that have the highest frequency to be respectively classified in the "large group" and the "small group" by size, and call them "large" and "small". Figure 3 plots the estimated characteristic betas and the associated confidence intervals of the two firms over time. As for the beta associated with the market factor, while both are positively significant, the characteristic betas of the firm with smaller size are constantly larger than one, making it more sensitive to the changes of market risks than the firm with the larger size. In addition, the pattern shown by the characteristic beta of the SMB factor is similar to Table 1: in the long run, the smaller firm is positively exposed and the larger firm is negatively exposed to the SMB systematic risk.

5.4 Testing characteristic relevance in asset pricing models

We now test the relevance of characteristics on factor betas, an important research question in asset pricing. We consider the linear specification $\mathbf{g}_{lt} = \mathbf{X}'_{lt}\boldsymbol{\theta}_t$ for a common coefficient matrix $\boldsymbol{\theta}_t$ and test the relevance of each of the characteristics $\mathbf{X}_{lt} = $ (size, value, momentum

Table 2: Cross-sectional Proportion of significant **G** of groups by volatility

volatility	positive significance				negative significance			
	Mkt	HML	SMB	RMW	Mkt	HML	SMB	RMW
				2006				
small	1	0.409	0.313	0.158	0	0.055	0.307	0.168
medium	1	0.187	0.473	0.151	0	0.159	0.214	0.153
large	1	0.034	0.583	0.142	0	0.403	0.068	0.075
				2008				
small	1	0.180	0.320	0.313	0	0.400	0.318	0.049
medium	1	0.378	0.511	0.280	0	0.170	0.178	0.092
large	1	0.644	0.565	0.195	0	0.079	0.090	0.138
				2009				
small	1	0.234	0.202	0.131	0	0.333	0.387	0.242
medium	1	0.286	0.341	0.152	0	0.296	0.230	0.223
large	1	0.346	0.506	0.184	0	0.198	0.086	0.171
				2011				
small	1	0.397	0.285	0.521	0	0.164	0.295	0.050
medium	1	0.284	0.340	0.144	0	0.315	0.273	0.232
large	1	0.175	0.278	0.017	0	0.445	0.224	0.621
				2012				
small	1	0.234	0.202	0.131	0	0.333	0.387	0.242
medium	1	0.286	0.341	0.152	0	0.296	0.230	0.223
large	1	0.346	0.506	0.184	0	0.198	0.086	0.171

and volatility) on the betas. Note that the (i, k)-th element of $\boldsymbol{\theta}_t$, denoted by $\theta_{t,ik}$, represents the effect of characteristic i on the k the factor beta. Although this specifies a linear function $\mathbf{g}_t(\cdot)$, $\mathbf{X}_{lt}$ could include nonlinear (sieve) transformations of each characteristic. Our test is uniformly valid over $\boldsymbol{\gamma}_{lt}$. To estimate $\boldsymbol{\theta}_t$, we use the linear sieve $\boldsymbol{\Phi}_t = (\mathbf{X}_{1t}, ..., \mathbf{X}_{pt})'$ and $\mathbf{P}_t = \boldsymbol{\Phi}_t(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'$. Then

$$\widehat{\boldsymbol{\theta}}_t = \left(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t\right)^{-1}\boldsymbol{\Phi}_t'\widehat{\mathbf{G}}_t.$$

We construct the bootstrap confidence intervals for each component of the estimated $\boldsymbol{\theta}_t$ on each trading day, and calculate the proportion of positive (and negative) significance each year. These results are reported in Table 3. For most of the period, the volatility has a significantly positive effect on the market factor, the value characteristic has a significantly positive effect on the HML factor, and the size characteristic has a significantly negative effect on the SMB factor. These results are consistent with the fitted $\mathbf{g}_t(\cdot)$ functions in Figures 1-4 (in the appendix). Also note that size has insignificant effects on the market

beta. We explain this from two aspects: on one hand, the market beta is mostly affected by the volatility instrument, and once it is conditioned, the size is no longer significant. On the other hand, we focus on firms that constitute to the S&P 500 index, whose sizes are relatively large, and are therefore not essential in explaining the market betas.

Table 3: Proportion of significant characteristics

characteristics	positive significance				negative significance			
	Mkt	HML	SMB	RMW	Mkt	HML	SMB	RMW
				2008				
size	0.024	0.036	0.000	0.048	0.000	0.215	0.984	0.128
value	0.008	0.892	0.008	0.048	0.000	0.000	0.219	0.076
momentum	0.004	0.044	0.159	0.315	0.139	0.450	0.187	0.060
volatility	0.534	0.598	0.295	0.092	0.000	0.048	0.040	0.167
				2011				
size	0.000	0.052	0.000	0.088	0.000	0.139	0.976	0.052
value	0.028	0.984	0.020	0.016	0.000	0.000	0.235	0.211
momentum	0.000	0.032	0.175	0.191	0.008	0.371	0.072	0.064
volatility	0.956	0.004	0.147	0.000	0.000	0.741	0.195	0.821
				2012				
size	0.000	0.048	0.000	0.155	0.000	0.171	0.920	0.044
value	0.008	0.996	0.016	0.104	0.000	0.000	0.235	0.108
momentum	0.080	0.024	0.100	0.004	0.000	0.355	0.175	0.522
volatility	0.813	0.171	0.458	0.187	0.000	0.155	0.036	0.084

Finally, addtional numerical results are presented in the appendix, where we plot the estimated $\mathbf{g}_t(\cdot)$ functions fitted by B-splines.

6   Conclusion

This paper studies a conditional factor model with a large number of individuals for high-frequency data. One of the key features of our model is that we specify the factor betas as functions of time-varying observed characteristics that pick up long-run beta movements driven by structural changes in the economic environment and in firm- or industry-specific conditions, plus a remaining (idiosyncratic) component that captures high-frequency movements in beta, which picks up short-run fluctuations in beta in periods of high market volatility. The two components capture different aspects of market beta dynamics. We show that the model can be extended to a more general continuous-time many moment conditions

setting, and estimated using two-step GMM setting.

The limiting distribution of the estimated characteristic effect on the betas has a discontinuity when the strength of the idiosyncratic beta is near zero. We provide a uniformly valid inference using a cross-sectional bootstrap procedure for the characteristic betas, and do not need to pretest to know whether or not the idiosyncratic beta exists, or their strengths.

The proposed framework can be extended for inference about general coefficients in other important econometric models, such as discrete time conditional factor models, panel data models with varying coefficients, and nonlinear panel models. In these models, the structural parameter may be decomposed into the sum of a characteristic driven component plus an orthogonal component. An important example would be linear panel data model with varying coefficients. In these models the asymptotic distribution of the estimated characteristic effect would also have a discontinuity, and thus standard "plug-in" inference fails to hold uniformly. The proposed bootstrap framework would then be very useful in these models. We shall leave these studies for future research.

References

Ait-Sahalia, Y., Kalnina, I. and Xiu, D. (2014). The idiosyncratic volatility puzzle: A reassessment at high frequency. Tech. rep., Working Paper, University of Chicago.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* **201** 384–399.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica* **67** 1341–1383.

Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405.

Andrews, D. W. (2004). The block–block bootstrap: improved asymptotic refinements. *Econometrica* **72** 673–700.

Andrews, D. W. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics* **91** 1–23.

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics* **106** 132–156.

Avramov, D. and Chordia, T. (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies* **19** 1001–1040.

BAI, J. and LIAO, Y. (2017). Inferences in panel data with interactive effects using large covariance matrices. *Journal of Econometrics* **200** 59–78.

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221.

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536.

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925.

BOLLERSLEV, T., LI, S. Z. and TODOROV, V. (2016). Roughing up beta: Continuous vs. discontinuous betas, and the cross section of expected stock returns. *Journal of Financial Economics* **120** 464–490.

CAI, T. T., YUAN, M. ET AL. (2012). Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics* **40** 2014–2042.

CHAMBERLAIN, G. and ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean-variance analyssi in large asset markets. *Econometrica* **51** 1305–1324.

CHEN, X. and LIAO, Z. (2015). Sieve semiparametric two-step gmm under weak dependence. *Journal of Econometrics* **189** 163–186.

CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 1591–1608.

CHEN, X. and POUZO, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* **80** 277–321.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. K. (2016). Double machine learning for treatment and causal parameters. Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

COCHRANE, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* **104** 572–621.

CONNOR, G. and KORAJCZYK, R. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance* **48** 1263–1291.

CONNOR, G. and KORAJCZYK, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics* **15** 373–394.

CONNOR, G. and LINTON, O. (2007). Semiparametric estimation of a characteristic-based factor model of stock returns. *Journal of Empirical Finance* **14** 694–717.

CONNOR, G., MATTHIAS, H. and LINTON, O. (2012). Efficient semiparametric estimation of the fama-french model and extensions. *Econometrica* **80** 713–754.

DANIEL, K. and TITMAN, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *the Journal of Finance* **52** 1–33.

DAROLLES, S., FAN, Y., FLORENS, J.-P. and RENAULT, E. (2011). Nonparametric instrumental regression. *Econometrica* **79** 1541–1565.

EILERS, P. H. and MARX, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science* 89–102.

FAMA, E. F. and FRENCH, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance* **47** 427–465.

FAN, J., FURGER, A. and XIU, D. (2016a). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics* **34** 489–503.

FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* **75** 603–680.

FAN, J., LIAO, Y. and WANG, W. (2016b). Projected principal component analysis in factor models. *Annals of statistics* **44** 219.

FAN, J., LIAO, Y. and YAO, J. (2015). Power enhancement in high dimensional cross-sectional tests. *Econometrica* **83** 14971541.

FERSON, W. E. and HARVEY, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance* **54** 1325–1360.

GAGLIARDINI, P., OSSOLA, E. and SCAILLET, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica* **84** 985–1046.

GHYSELS, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *The Journal of Finance* **53** 549–573.

HALL, P. and HOROWITZ, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics* **33** 2904–2929.

HERSKOVIC, B., KELLY, B., LUSTIG, H. and VAN NIEUWERBURGH, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics* **119** 249–283.

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications* **119** 2249–2276.
URL http://ideas.repec.org/a/eee/spapps/v119y2009i7p2249-2276.html

JACOD, J. and PROTTER, P. (2011). *Discretization of Processes*. Springer.

JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics* **41** 1462–1484.

JAGANNATHAN, R. and WANG, Z. (1996). The conditional capm and the cross-section of expected returns. *The Journal of finance* **51** 3–53.

KALNINA, I. (2012). Nonparametric tests of time variation in betas. *Université de Montréal* .

KELLY, B. T., PRUITT, S. and SU, Y. (2017). Instrumented principal component analysis .

KETZ, P. (2017). Testing overidentifying restrictions when the true parameter vector is near or at the boundary of the parameter space. Tech. rep., Working Paper.

KETZ, P. (2018). Subvector inference when the true parameter vector may be near or at the boundary .

KIM, S., KORAJCZYK, R. A. and NEUHIERL, A. (2018). Arbitrage portfolios in large panels. *Available at SSRN* .

KING, M., SENTANA, E. and WADHWANI, S. (1994). Volatility and links between national stock markets. *Econometrica* **62** 901–33.

LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics* 386–404.

LETTAU, M. and LUDVIGSON, S. (2001). Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of political economy* **109** 1238–1287.

LI, J., TODOROV, V. and TAUCHEN, G. (2016). Inference theory for volatility functional dependencies. *Journal of Econometrics* **193** 17–34.

LI, J., TODOROV, V. and TAUCHEN, G. (2017). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics* **200** 36–47.

LI, J., TODOROV, V. and TAUCHEN, G. (2018). Jump factor models in large cross-sections.

LI, J. and XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* **84** 1613–1633.

MANCINI, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari* **LXIV** 19–41.

MYKLAND, P. A., ZHANG, L. ET AL. (2006). Anova for diffusions and ito processes. *The Annals of Statistics* **34** 1931–1963.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* 1349–1382.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578.

Pedersen, R. S. (2017). Inference and testing on the boundary in extended constant conditional correlation garch models. *Journal of Econometrics* **196** 23–36.

Pelger, M. (2016). Large-dimensional factor modeling based on high-frequency observations. Working paper.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations.* 2nd ed. Springer.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics* **45** 99–120.

Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97** 1167–1179.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 817–838.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **100** 1394–1411.
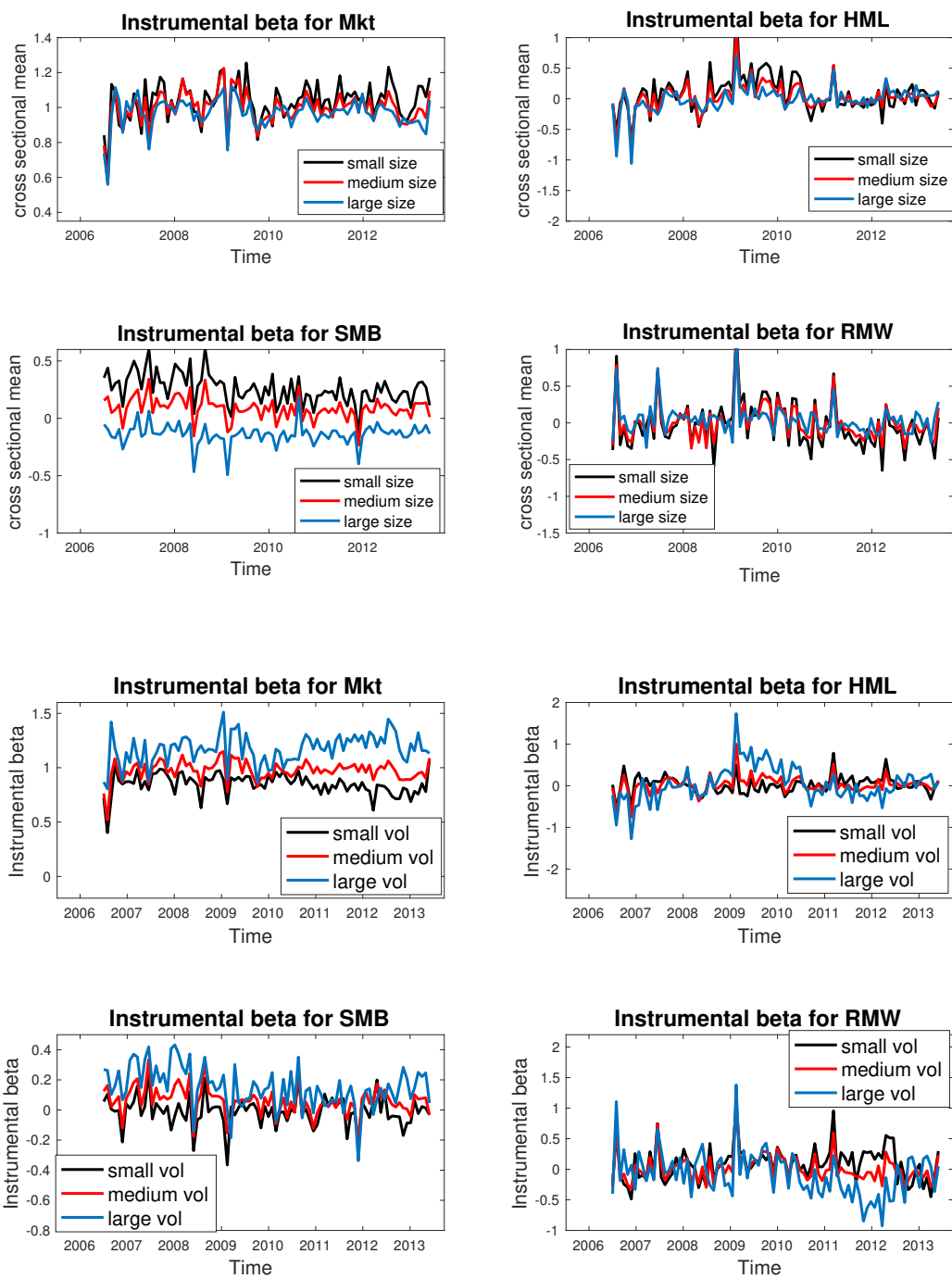
Figure 1: Cross-sectional means of characteristic beta, grouped by size (upper four) and by volatility (lower four). The characteristic betas are estimated on a daily basis, and this figure plots eight days' estimations for each month.
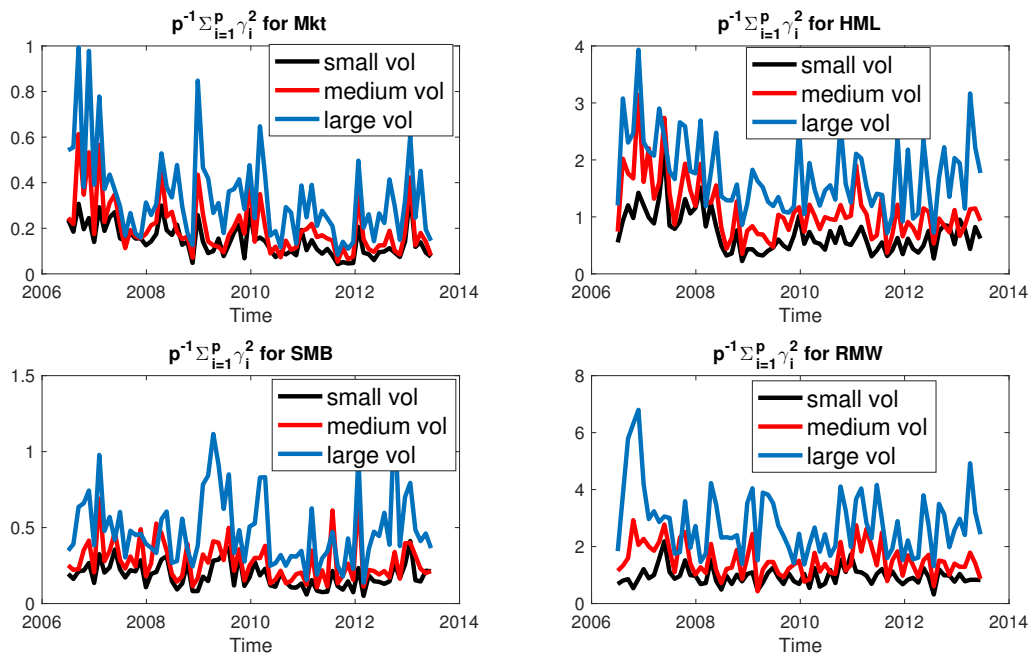
Figure 2: Cross-sectional variations of $\mathbf{\Gamma}$ across times. Firms are grouped into small, medium and large sizes.
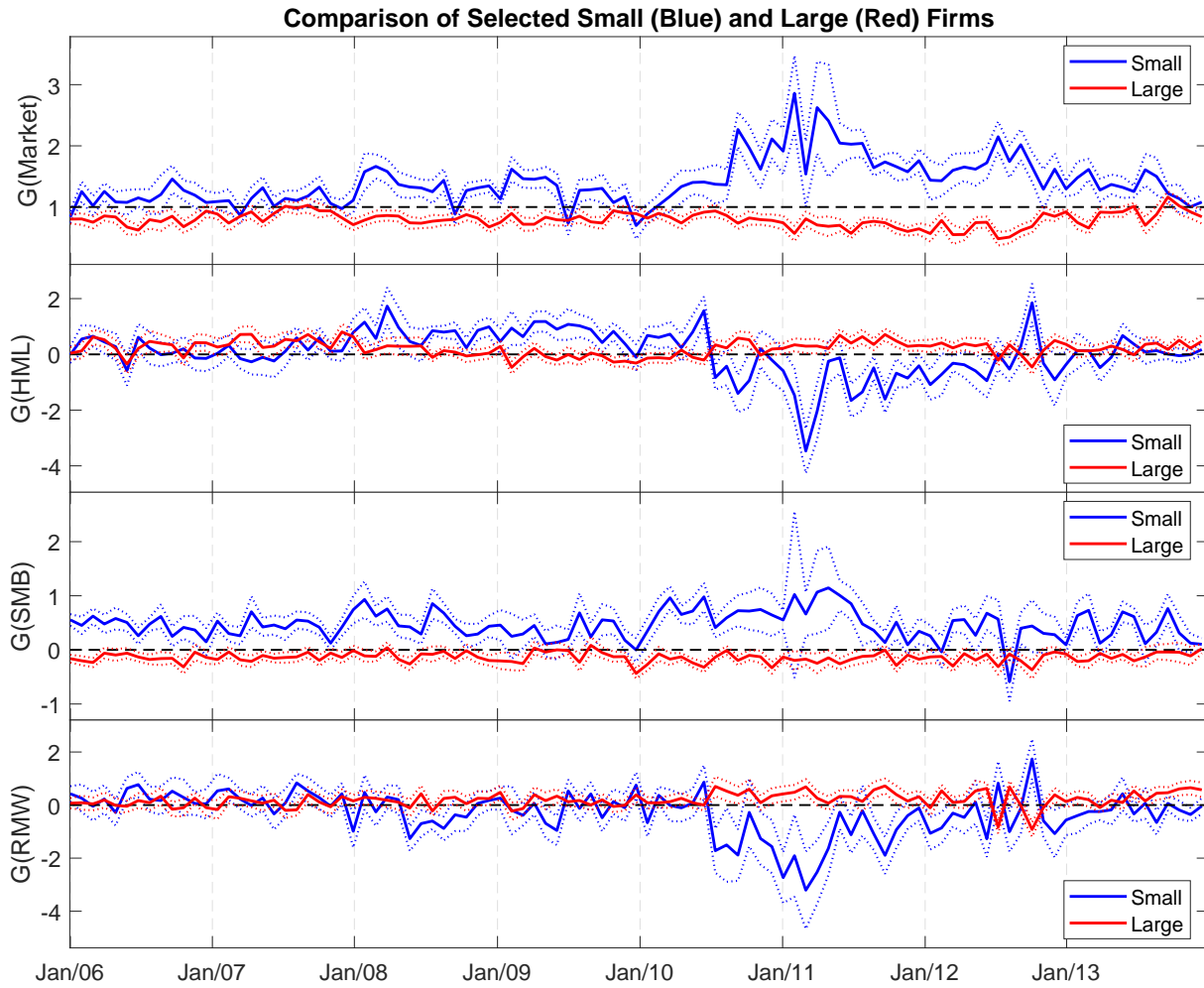
Figure 3: Two individual stocks' confidence intervals: the two firms with the largest and smallest sizes in the dataset.