Uniform Inference and Prediction for Conditional Factor Models with Instrumental and Idiosyncratic Betas*

Yuan Liao Xiye Yang

September 24, 2018

Abstract

It has been well known in financial economics that factor betas depend on observed instruments such as firm specific characteristics and macroeconomic variables, and the key object of interest is the effect of instruments on the factor betas. We specify the factor betas as functions of time-varying observed instruments that pick up long-run beta fluctuations, plus an orthogonal idiosyncratic component that captures high-frequency movements in beta. It is often the case that researchers do not know whether or not the idiosyncratic beta exists, or its strengths, and thus uniformity is essential for inferences. It is found that the limiting distribution of the estimated instrument effect has a discontinuity when the strength of the idiosyncratic beta is near zero, which makes usual inferences fail to be valid and produce misleading results. In addition, the usual "plug-in" method using the estimated asymptotic variance is only valid pointwise. The central goal is to make inference about the effect on the betas of firms' instruments, and to conduct out-of-sample forecast of integrated volatilities using estimated factors. Both procedures should be valid uniformly over a broad class of data generating processes for idiosyncratic betas with various signal strengths and degrees of time-variant. We achieve the uniformity through a cross-sectional bootstrap procedure, and our procedure also features a bias correction for the effect of estimating unknown factors.

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

*Department of Economics, Rutgers University, 75 Hamilton St. New Brunswick, NJ 08901, USA. Email: yliao@econ.rutgers.edu; xiyeyang@econ.rutgers.edu. We are grateful to the comments from the audience of the 2017 Midwest Econometric Workshop, the 12th Greater New York Metropolitan Colloquium, 2018 Econometric Society Summer Meeting, 2018 Market Microstructure and High Frequency Data, 2018 Big data on financial markets, and seminar participants at CUHK and UPenn.

1 Introduction

Conditional factor models have been playing an important role in capturing the time-varying sensitivities of individual assets to the risk factors, in which the factor betas of the assets are varying over time. Extensive empirical studies have shown that assets' individual betas can be largely explained by asset specific characteristics and instruments. These include lagged instruments that are common to all stocks, instruments specific to individual stocks, as well as observations of other firm characteristics. Estimated betas as functions of the conditioning instruments represent the effects of instruments on firm specific sensitivities to the risk factors. They pick up long-run patterns and fluctuations in the betas. Therefore, estimating the instruments' effects on the individual betas is one of the central econometric tasks in financial economics.

While instruments may fully explain the factor betas at times when they are just updated and made publicly available, there are also times when factor betas contain either unmeasurable or high-frequency components that are more volatile and cannot be captured by the instruments. In those occasions, modeling the beta as fully specified functions of observed instruments can be very restrictive. This is particularly true for high-dimensional and high-frequency factor models, where individual factor betas demonstrate large heterogeneity when the number of assets is large, and assets' returns are available at a very high frequency. In the contrary, instruments 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 instruments, the inference procedures of instruments' effects are not asymptotically valid. Unfortunately, this is often the case in the financial econometric literature, which has been dominated by modeling betas as fully specified functions of the observed instruments, 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). [1] We need to be particularly cautious when modeling betas. 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.

We propose a conditional factor model in which the time-varying factor betas are as

[1] While Connor and Linton (2007) and Connor et al. (2012) proposed to model the instruments using nonparametric functions, they require that the nonparametric function and instruments be time-invariant and have an additive-structure.

follows: let $\boldsymbol{\beta}_{lt}$ denote the K-dimensional factor betas of the l th stock at time t. Let $(\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t)$ be a vector of observed "instruments". We model:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_{lt}(\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t) + \boldsymbol{\gamma}_{lt}, \quad \mathbb{E}(\boldsymbol{\gamma}_{it}|\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t) = 0. \tag{1.1}$$

The main message from equation (1.1) is that we allow $\boldsymbol{\gamma}_{lt} := \boldsymbol{\beta}_{lt} - \mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t)$ *to be possibly nonzero with unknown strengths.* So betas may (or may not) have cross-sectional and dynamic variations after conditioning on instruments. The factor beta consists of two components: (i) a nonparametric function of the observed instruments, $\mathbf{g}_{lt} := \mathbf{g}_{lt}(\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t)$, which we call "instrumental betas", and (ii) an unmeasurable time-varying and firm specific component, $\boldsymbol{\gamma}_{lt}$, which we call "idiosyncratic betas". The instrumental 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 instruments. The goal of this paper is to provide a *uniformly valid* inference of $\mathbf{g}_{lt}$, the instrumental 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 various asymptotic sequences representing its cross-sectional variance. We study a high-dimensional continuous-time factor model, in which the factors, betas, and idiosyncratic errors are all driven by Brownian motions through a continuous-time stochastic process. Both components in the decomposition of $\boldsymbol{\beta}_{lt}$ are estimated using the high-frequency data. In addition, we allow the instruments $(\mathbf{x}_{l,t}, \mathbf{x}_l, \mathbf{x}_t)$ to be lagged common time-varying instruments and macroeconomic variables, time-invariant or change only at a lower frequency firm specific characteristics, and instruments that are both time-varying and firm specific. Depending on the measurement and specification of the instruments, the strengths of both components in $\boldsymbol{\beta}_{lt}$ may vary both cross-sectionally and serially, and is often unknown to econometricians. The class of DGPs we consider admits all these time dynamics and cross-sectional variations from both components of beta's , hence is very broad.

We find that, the strength of $\boldsymbol{\gamma}_{lt}$ plays a crucial role in the asymptotic behavior of estimated instrument effect, and affects both the rate of convergence and limiting distributions. In particular, the asymptotic distribution of $\mathbf{g}_{lt}$ has a "discontinuity" when the strength of $\boldsymbol{\gamma}_{lt}$, measured by its cross-sectional variance, is 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 instrumental 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 the unexplained beta dynamics, leading to *over-coveraging* confidence intervals. The discontinuity issue here is similar to the problem of estimating parameters on a boundary. As is shown by, e.g., (Andrews, 1999) and (Mikusheva, 2007), 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, in the sense that it resamples the cross-sectional units and keeps all the serial observations for each sampled individual asset. This procedure is essential because the discontinuity arises due to the strength of the cross-sectional variance of $\boldsymbol{\gamma}_{lt}$, and the cross-sectional bootstrap avoids the estimation error for both the unobserved $\boldsymbol{\gamma}_{lt}$ and its cross-sectional variance. 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. In this framework, our procedure also provides a new feature for random coefficient models, where the random coefficient can be decomposed into the addition of a "deterministic" and a "random" component, and the inference of the former is affected by the strength of the latter.

The strength of variations in $\boldsymbol{\gamma}_{lt}$ also plays an essential role in the long-run forecast for the integrated volatility of a fixed asset using estimated factors. We construct out-of-sample forecast confidence intervals for the conditional mean of the integrated volatility using a model similar to the diffusion index forecast (Stock and Watson, 2002). As in the diffusion index forecast, the confidence interval depends on the effect of estimating latent financial factors from a large amount of financial asset returns. We find that whether or not the strength of $\boldsymbol{\gamma}_{lt}$ plays a role in the forecast interval depends on whether $\boldsymbol{\gamma}_{lt}$ is time-varying. When it is indeed time-varying, ignoring it in the factor model, as is commonly treated in the literature, continues to produce misleading forecast confidence intervals. As before, we construct forecast intervals that are robust to the strength of $\boldsymbol{\gamma}_{lt}$, and is uniformly valid over a large class of DGPs that allows different types of time-variations in $\boldsymbol{\gamma}_{lt}$.

1.1 The literature

The study of the effects of instruments 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. Modeling the betas using these instruments is thus essential to distinguish effects for firms with different levels of book-to-market ratio. Firms' momentum is also one of the commonly used instruments, 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 instruments track variations in expected returns that is not captured by the Fama-French (Fama and French, 1992) three-factor model, and that these instruments have explanatory power on the factor loadings because they pick up betas' time-variation. In addition, the effects of common instruments such as the term spread (difference between yields on 10-year Treasury and three-month T-bill) and default spread (yield difference between Moody's Baa-rated and Aaa-rated corporate bonds) 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 instruments, a similar decomposition to (1.1) was given by Kelly et al. (2017), where betas are decomposed into a linear function of lagged instruments as well as an unobservable loading component. They specifically require $\boldsymbol{\gamma}_t$ be strong, with cross-sectional variances that are bounded away from zero. In this setting, they obtained limiting distributions for the "instrumental betas", which do not possess the discontinuity issue as we encounter in this paper, and is not "uniformly correct". This then results in a inference procedure that can be potentially severely conservative. In addition, Cosemans et al. (2009) decomposed beta into a weighted sum of firms' characteristic beta and remainders. Using a hierarchical Bayesian approach, they found a large increase in the cross-sectional explanatory power of the conditional CAPM. Moreover, Fan et al. (2016) studied a model whose betas have a similar decomposition. They did not allow time-varying conditional factor models. As we show in this paper, allowing time-varying betas make a key difference for the long-run out-of sample forecast using estimated factors. The effect of the idiosyncratic betas does play a key role in the constructed forecast interval in the conditional model, while it does not in unconditional models. The more important difference is that they did not study the uniform inference. Our paper is also related to the recently rapidly growing literature on continuous-time factor models, such as Pelger (2016); Aït-Sahalia and Xiu (2017) and references therein.

The rest of this paper is organized as follows. Section 2 informally discusses the issue of uniformity and the intuitive solutions. Section 3 describes the continuous-time conditional factor model driven by stochastic processes. Section 4 defines the estimators of the components of betas, and the unknown factors in the case of latent factors. Section 5 presents the asymptotic results of the estimators. Section 6 presents results of long-run forecasts using estimated factors, as long as inference of long-run instrumental betas. Section 7 discusses extensions on testing the instrumental relevance and estimating varying-coefficient models. In Section 8, we present real data applications on the high-frequency stock return data of firms from S&P500. Finally, in the appendix, we give a simple simulated example to examine the uniformity of the proposed inference in finite sample, as well as all the technical proofs.

Notation: We observe asset returns 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 *universe constant* if it does not depend on any pointwise DGP. 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 our model and the issue about uniform inference using a discrete-time, unconditional factor model with observed factors. Consider the following discrete-time model:

$$y_{it} = [\mathbf{g}(\mathbf{X}_i) + \boldsymbol{\gamma}_i]'\mathbf{f}_t + u_{it}, \quad i = 1, ..., p; \quad t = 1, ..., T. \tag{2.1}$$

Here $(y_{it}, \mathbf{f}_t, \mathbf{X}_i)$ are observable. We assume that $\mathbb{E}(\boldsymbol{\gamma}_i|\mathbf{X}_1, ..., \mathbf{X}_p) = 0$. In addition, $\mathbf{g}(\cdot)$ is a smooth nonparametric function that has a sieve approximation: $\mathbf{g}(\mathbf{X}_i) \approx \mathbf{B}'\boldsymbol{\phi}_i$, where $\boldsymbol{\phi}_i = (\phi_1(\mathbf{X}_i), ..., \phi_J(\mathbf{X}_i))'$ is a $J \times 1$ vector of sieve basis functions of $\mathbf{X}_i$, with $\mathbf{B}$ the corresponding coefficients. For a fixed $l \leq p$, we estimate $\mathbf{g}_l := \mathbf{g}(\mathbf{X}_l)$ by a combination of cross-sectional and time-series regression: $\widehat{\mathbf{g}}_l = \mathbf{s}_f^{-1} \frac{1}{pT} \sum_{t=1}^T \sum_{i=1}^p \mathbf{f}_t y_{it} h_{il}$, where $\mathbf{s}_f = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t'$, and $h_{il} = \boldsymbol{\phi}_i' (\frac{1}{p} \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\phi}_j')^{-1} \boldsymbol{\phi}_l$. Then ignoring the sieve approximation error, $\widehat{\mathbf{g}}_l$ has the following expansion

$$\widehat{\mathbf{g}}_l - \mathbf{g}_l = \underbrace{\frac{1}{p} \sum_{i=1}^p \boldsymbol{\gamma}_i h_{il}}_{(\mathbf{a})} + \underbrace{\mathbf{s}_f^{-1} \frac{1}{pT} \sum_{t=1}^T \sum_{i=1}^p \mathbf{f}_t u_{it} h_{il}}_{(\mathbf{b})} + \text{negligible terms}. \tag{2.2}$$

Thus the asymptotic distribution depends on the interplay of two leading terms. Term $(\mathbf{a})$ arises from the cross-sectional estimation, which has a rate $O_P(p^{-1/2} \|\mathbf{V}_\gamma\|^{1/2})$, with $\mathbf{V}_\gamma = \mathrm{Var}(\frac{1}{\sqrt{p}} \sum_{i=1}^p \boldsymbol{\gamma}_i h_{il} | \mathbf{X})$, where $\mathrm{Var}(\cdot|\mathbf{X})$ denotes the conditional variance given $\{\mathbf{X}_l\}_{l \leq p}$. We shall assume $\boldsymbol{\gamma}_i$'s are cross-sectionally independent so that $(\mathbf{a})$ admits a cross-sectional central limit theorem (CLT). In addition, term $(\mathbf{b})$ has a rate $O_P((Tp)^{-1/2})$ under standard weakly-dependent conditions.

If $\mathbf{V}_\gamma$ is *weak*, whose eigenvalues, treated as sequences, decay at rate faster than $O_P(T^{-1})$, then $(\mathbf{b})$ is the dominating term, leading to

$$\sqrt{Tp}(\widehat{\mathbf{g}}_l - \mathbf{g}_l) = O_P(1),$$

whose asymptotic distribution is determined by $(\mathbf{b})$. Intuitively, this occurs when the idiosyncratic betas have weak signals from the cross-sectional variations. As a result, the observed instruments 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, $(\mathbf{a})$ becomes the dominating term, and we simply have

$$\sqrt{p}(\widehat{\mathbf{g}}_l - \mathbf{g}_l) = O_P(1).$$

In this case, the limiting distribution is determined by the cross-sectional CLT of $(\mathbf{a})$. Intuitively, this means when the idiosyncratic betas have strong cross-sectional variations, time-domain averaging is not helpful to remove their effect on estimating $\mathbf{g}_{lt}$, and only cross-sectional projection does the job. This leads to a slower rate of convergence.

Consequently, there is a discontinuity on the limiting distribution of $\widehat{\mathbf{g}}_l - \mathbf{g}_l$ when the signals of cross-sectional variation of $\boldsymbol{\gamma}_l$ is near the "boundary". 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 economic studies take the "weak" case as the default assumption, while some other studies (e.g., Cosemans et al. (2009)) provide evidence of the latter case. Above all, to our best knowledge, all the existing inferences are pointwise, and is not robust to the strength of gamma's variations. Pointwise inferences, therefore, can be misleading.

Such a discontinuity has a fundamental impact on the uniform inference about $\mathbf{g}_{lt}$. Let $\sqrt{p}\mathbf{V}_\gamma$ and $\sqrt{Tp}\mathbf{V}_u$ respectively denote the asymptotic variances of terms (**a**) and (**b**) in the expansion of (2.2). In particular, when $\{\boldsymbol{\gamma}_i\}_{i\leq p}$ are cross-sectionally uncorrelated,

$$\mathbf{V}_\gamma = \frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \operatorname{Var}(\boldsymbol{\gamma}_i|\mathbf{X}).$$

A "standard" inference procedure is to plug-in the estimated asymptotic covariances for $\mathbf{V}_u$ and $\mathbf{V}_\gamma$ using their sample analogues. This procedure, however, works only pointwise, and does not provide a uniformly valid confidence interval.[2] To understand the issue, consider the estimation of $\mathbf{V}_\gamma$. If $\boldsymbol{\gamma}_i$ were known, White (1980)'s heteroskedastic covariance estimator can be applied: $\widetilde{\mathbf{V}}_\gamma = \frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'$. Replacing $\boldsymbol{\gamma}_i$ with its consistent estimator $\widehat{\boldsymbol{\gamma}}_i$, we obtain $\widehat{\mathbf{V}}_\gamma = \frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i'$. Then $\widehat{\mathbf{V}}_\gamma - \mathbf{V}_\gamma$ has a decomposition

$$\underbrace{\frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \left[\widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i' - \boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'\right]}_{\gamma\text{-estimation error}} + \underbrace{\frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \left[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i' - \operatorname{Var}(\boldsymbol{\gamma}_i|\mathbf{X}_t)\right]}_{\text{LLN error}} \tag{2.3}$$

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. One of the leading terms

[2]More precisely, when $\mathbf{V}_\gamma = 0$, plugging in its consistent estimator over-estimates the asymptotic variance, leading to a valid but severely conservative inferences. On the other hand, not estimating $\mathbf{V}_\gamma$ would result in severe under-coverages when it is bounded away from zero.

in the expansion of $\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i$ is $\mathbf{r}_i := \mathbf{s}_f^{-1} \frac{1}{T} \sum_{t \leq T} \mathbf{f}_t u_{it}$, which leads to

$$\gamma\text{-estimation error} \geq \frac{1}{p} \sum_{i=1}^{p} h_{il}^2 \mathbf{r}_i \mathbf{r}_i' \asymp O_P(T^{-1}).$$

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). Hence estimating $\mathbf{V}_\gamma$ introduces an estimation error that is non-negligible when $\{\boldsymbol{\gamma}_i\}$ is weak. 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-coverage probabilities when $\{\boldsymbol{\gamma}_i\}$ is strong. Hence it is not uniformly valid. We aim to provide a uniformly valid inference procedure. More specifically, we construct a confidence interval CI_τ for $\mathbf{g}_l$, so that at the nominal level $1-\tau$,

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

Here the probability measure $\mathbb{P}$ is taken uniformly over a broad DGP class $\mathcal{P}$, which admits various cross-sectional variations in $(\boldsymbol{\gamma}_i,\, \mathbf{g}(\mathbf{X}_i))$ and dynamics if they are also time-varying. Uniformity in the above sense is essential for inferences in this context, because it makes the inference valid and robust to the unknown degrees of dynamics of factor betas.

2.2 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 $\widehat{\mathbf{g}}_l^*$ using the bootstrap data. Finally, we calculate critical value of $\widehat{\mathbf{g}}_l^* - \widehat{\mathbf{g}}_l$, from a set of bootstrap estimators. This procedure is very simple, but perhaps surprisingly, leads to the desired uniform coverage for $\mathbf{g}_l$.

To prove the bootstrap validity, it is essential to show that this procedure directly mimics the cross-sectional variations in $\{\boldsymbol{\gamma}_i\}$. To see this intuitively, we note that the bootstrap asymptotic variance of $\widehat{\mathbf{g}}_l^*$ is analogously $\frac{1}{Tp}\mathbf{V}_u + \frac{1}{p}\widetilde{\mathbf{V}}_\gamma$, where $\widetilde{\mathbf{V}}_\gamma = \frac{1}{p}\sum_{i=1}^{p} h_{il}^2 \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i'$. The only approximation error for $\mathbf{V}_\gamma$ is:

$$\widetilde{\mathbf{V}}_\gamma - \mathbf{V}_\gamma = \underbrace{\frac{1}{p}\sum_{i=1}^p h_{il}^2 \left[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i' - \mathrm{Var}(\boldsymbol{\gamma}_i|\mathbf{X})\right]}_{\text{LLN error}}.$$

Consequently, the γ-estimation error component in (2.3) is avoided. The LLN error is of a higher order than $\mathbf{V}_\gamma$, regardless of the signal strength of $\boldsymbol{\gamma}_i$. For instance, suppose $\boldsymbol{\gamma}_i$ is generated from a rescaled sequence, that is, $\boldsymbol{\gamma}_i = a_{NT}\bar{\boldsymbol{\gamma}}_i$, where $a_{NT} \geq 0$ is a non-random arbitrary sequence, and $\bar{\boldsymbol{\gamma}}_i$ is a random sequence so that for $C > 0$, almost surely,

$$\frac{1}{p}\sum_{i=1}^p h_{il}^4\, \mathbb{E}(\|\bar{\boldsymbol{\gamma}}_i\|^4|\mathbf{X}) \leq C\lambda_{\min}^2[\frac{1}{p}\sum_{i=1}^p h_{il}^2\, \mathrm{Var}(\bar{\boldsymbol{\gamma}}_i|\mathbf{X})].$$

Then the LLN-error $= o_P(1)\mathbf{V}_\gamma$. Hence the approximation error for the asymptotic variance of $\widehat{\mathbf{g}}_l$ is negligible regardless of the strength of $\mathbf{V}_\gamma$. 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. We note several fundamental differences between our problem and that of Andrews (2000)'s. First of all, we have different sources of discontinuity in the asymptotic distribution of the estimator. Recall that the asymptotic expansion of $\widehat{\mathbf{g}}_l - \mathbf{g}_l$ consists of two main terms:

$$\mathbf{s}_f^{-1}\frac{1}{Tp}\sum_{t=1}^T\sum_{m=1}^p \mathbf{f}_t u_{it} h_{il}, \quad \text{and} \quad \underbrace{\frac{1}{p}\sum_{i=1}^p \boldsymbol{\gamma}_i h_{il}}_{(\mathbf{a})},$$

The "discontinuity" is a consequence of the interplay between the two terms. Depending on the strength of $\mathrm{Var}(\boldsymbol{\gamma}_i|\mathbf{X})$, the dominating term can vary. But term (**a**) itself is continuous with respect to $\boldsymbol{\gamma}_i$, because $\mathrm{Var}(\boldsymbol{\gamma}_i|\mathbf{X})$ is not explicitly modeled as an unknown parameter. In sharp contrast, in the model of Andrews (2000), $\mathrm{Var}(\boldsymbol{\gamma}_i|\mathbf{X})$ is explicitly modeled as an unknown parameter, and the asymptotic distribution is discontinuous with respect to the boundary parameter, and is not due to the issue of interplay among multiple terms in the asymptotic expansion. Secondly, the usual "plug-in" method for the estimated asymptotic variance is not uniformly valid because the effect of estimating $\boldsymbol{\gamma}_i$ dominates the asymptotic variance of interest. As we illustrated in the above, this error can be avoided by resampling the cross-sectional units, and the bootstrap variance directly estimates the asymptotic variance of term (**a**). Finally, when inferencing about $\mathbf{g}_l$, whether $\mathrm{Var}(\boldsymbol{\gamma}_i|\mathbf{X})$ is near the boundary is unknown under both the null and the alternative hypotheses. In contrast, in the literature, this parameter is on the boundary under the null.

Remark 2.1. 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 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 parameters near the boundary (e.g., Andrews (1999, 2000); 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.

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; (5) idioscynractics are cross-sectionally independent; (4) there are no drifts. In this paper we shall formally emplore this idea in a continuous-time conditional factor model with drifts using high-frequency data, and separately consider observed and estimated factors. We also describe a block cross-sectional bootstrap algorithm to allow for weakly dependent cross-sectional idiosyncratic terms.

3 The Continuous-Time Factor Model with Instruments

3.1 The model

Consider a financial market with p number of stocks. Let $\mathbf{Y}_t = (Y_{1t}, \cdots, Y_{pt})$ be the vector of log-prices of these stocks at time t. We assume $\mathbf{Y} = \{\mathbf{Y}_t\}_{t\geq 0}$ is a multivariate Itô semi-martingale 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, introducing jumps substantially complicates the notation and does not bring any new economic insights. The jump-robust estimators are given in Section 4.3, where we employ a standard procedure to truncate jumps out. [3]

[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.

In this paper, 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 (continuous) factor $\mathbf{F}_t$ and the idiosyncratic continuous risk $\mathbf{U}_t$ can be represented as

$$\begin{aligned} \mathbf{F}_t &= \int_0^t \boldsymbol{\alpha}_s^F ds + \int_0^t \boldsymbol{\sigma}_s^F d\mathbf{W}_s^F, \\ \mathbf{U}_t &= \int_0^t \boldsymbol{\sigma}_s^U d\mathbf{W}_s^U, \end{aligned} \tag{3.2}$$

where $\mathbf{W}^U$ and $\mathbf{W}^F$ are two multi-dimensional Brownian motions and are orthogonal (in the martingale sense) to each other, and $\boldsymbol{\alpha}^F = \{\boldsymbol{\alpha}_s^F\}_{s\geq 0}$ is the drift process of the K dimensional continuous factor process $\mathbf{F} = \{\mathbf{F}_t\}_{t\geq 0}$. 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)), to differentiate from the discontinuous (or jump) beta (Li et al. (2017)). In addition, for each firm $l \leq p$, we observe a set of (possibly) time-varying instruments:

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

We allow the instruments $\mathbf{X}_{lt}$ to consist of (1) time-varying instruments $\mathbf{x}_t$ that are common to stocks (such as term and default spread and macroeconomic variables); (2) firm specific instruments $\mathbf{x}_l$ that are time-invariant over the sampling period $[0, T]$ (such as size and value which change annually); and (3) instruments $\mathbf{x}_{l,t}$ that are both time-varying and firm specific.

In this paper, we consider the following decomposition of the factor loadings (continuous betas):

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

The effect of instruments on the factor loadings are represented by $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, and is called "instrumental beta". Here $\mathbf{g}_{lt}(\cdot)$ is a 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 instruments, and captures high frequency movements in beta. The two components capture different aspects of beta dynamics. For the identification purpose, we assume $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_{lt}) = 0$, which well separates the characteristic effects and remaining effects. Let $\mathbf{G}_t$ be the $p \times K$ matrix of $\{\mathbf{g}_{lt}(\mathbf{X}_{lt})\}_{l=1}^p$ and $\boldsymbol{\Gamma}_t$ be the $p \times K$ matrix of $\{\boldsymbol{\gamma}_{lt}\}_{l=1}^p$. Then we have the following representation for the continuous component of $\mathbf{Y}$:

$$d\mathbf{Y}_t = \boldsymbol{\alpha}_t dt + (\mathbf{G}_t + \boldsymbol{\Gamma}_t)d\mathbf{F}_t + d\mathbf{U}_t, \quad \forall t \in [0, T]. \tag{3.4}$$

We rely on two key assumptions to reduce the "essential number of parameters" in $\{\mathbf{g}_{lt}(\mathbf{X}_{lt})\}_{l \leq p, t \in [0,T]}$. On the time domain, we assume that components of $\mathbf{X}_{lt}$ are driven by Brownian motions, and that $t :\to \mathbf{g}_{lt}(\cdot)$ is a smooth function in $t \in [0, T]$ uniformly over $l \leq p$. Therefore, $\mathbf{g}_{lt}(\mathbf{X}_{lt})$ is approximately time-invariant on local windows. On the cross-sectional domain, we shall assume that $\{\mathbf{g}_{lt}(\mathbf{X}_{lt})\}_{l \leq p}$ can be well-approximated using sieve expansions, that is, $\max_{t \in [0,T]} \|\mathbf{G}_t - \mathbf{P}_t \mathbf{G}_t\|_\infty$ decays fast, where $\mathbf{P}_t$ is a $p \times p$ projection matrix using the sieve transformations of $\{\mathbf{X}_{lt}\}_{l \leq p}$ (Assumption 5.3 gives formal definitions). Therefore, the cross-sectional elements can be approximately considered as realizations from a smooth nonparametric function.

We separately study two cases: known and unknown factor cases. By the "known factor case", we explain the returns through a set of common factors that are observed at the same time points of the high-frequency return data. Recently, Ait-Sahalia et al. (2014) 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 number of assets (up to a locally time-invariant rotation matrix).

We also use the estimated factors for the long-run forecast. Consider

$$y_{d+h} = \mu y_d + \boldsymbol{\rho}' \mathbf{F}_d + v_{d+h}, \quad d = 1, ..., L_n, \quad L_n \to \infty, \tag{3.5}$$

where $h > 0$ is the lead time between information available and y_{d+h}, the dependent variable to forecast, and

$$\mathbf{F}_d := \int_{(d-1)T}^{dT} d\mathbf{F}_t.$$

Here μ and $\boldsymbol{\rho}$ are the unknown coefficients, and v_{d+h} is the innovation. Of interest is to construct the out-of-sample prediction confidence interval for the conditional mean $y_{L_n+h|L_n} =$

$\mu y_{L_n} + \boldsymbol{\rho}'\mathbf{F}_{L_n}$. A typical example arises from forecasting the log integrated volatility:

$$y_d = \log \mathrm{IV}_d, \quad \mathrm{IV}_d := \int_{(d-1)T}^{dT} \sigma_t^2 dt,$$

where σ_t^2 is the spot volatility of certain asset. Similar to the diffusion index forecast, the common factors are extracted from the large number of assets' high-frequency returns. While there is a large literature on forecasting volatilities (e.g., Engle and Bollerslev (1986); Andersen et al. (2006); Hansen and Lunde (2011)), motivated by Stock and Watson (2002), we extract the latent factors from a large set of financial asset returns. But we are particularly interested in the effect from the time-varying $\boldsymbol{\Gamma}_t$ on the prediction intervals.

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

One of the key conditions is $\mathbb{E}(\boldsymbol{\Gamma}_t|\mathbf{X}_t) = 0$. It implies that the instrumental and idiosyncratic beta components are orthogonal. This condition serves as a central condition to achieve the identification of the instrument effect, 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. Technically, this condition can be understood as assuming $\mathbf{g}_{lt}(\mathbf{X}_t) = \mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{X}_t)$. Hence we are estimating the instrument effects as the conditional mean of the betas. In the absence of this condition, identification is lost, and we need further exogenous variables to identify the effect of instruments. For simplicity, we assume $\mathbf{g}_{lt}(\mathbf{X}_{lt}) = \mathbf{g}(\mathbf{X}_l)$ as a time-invariant nonparametric function. Consider the decomposition:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l \leq p \tag{3.6}$$

Consider the "ideal case" that $\boldsymbol{\beta}_{lt}$ is completely known. Then in (3.6), $\mathbf{X}_{lt}$ is endogenous. To identify $\mathbf{g}(\cdot)$, consider an instrumental variable approach: we need to find an exogenous multi-dimensional process $\mathbf{Z}_{lt}$ so that $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{Z}_{lt}) = 0$. Define the operator:

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

We then have $\mathcal{T}(g) = \mathbb{E}(\boldsymbol{\beta}_{lt}|\mathbf{Z}_{lt})$. The identification of $\mathbf{g}$ depends on the invertibility of $\mathcal{T}$, and holds if and only if the conditional distribution of $\mathbf{X}_{lt}|\mathbf{Z}_{lt}$ 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 $\mathbf{g}$ 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 $\mathbb{E}(\boldsymbol{\Gamma}_t|\mathbf{X}_t) = 0$ 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.

4 Estimation

Since our focus is on the continuous factors, we shall first assume the underlying log-price processes and the factors are all continuous. Ignoring the jumps, over the i-th sampling interval, we have the following approximation for $\Delta_i^n \mathbf{Y} := \mathbf{Y}_{i\Delta_n} - \mathbf{Y}_{(i-1)\Delta_n}$:

$$\begin{aligned}\Delta_i^n \mathbf{Y} &= \int_{(i-1)\Delta_n}^{i\Delta_n} \Big(\boldsymbol{\alpha}_t dt + (\mathbf{G}_t + \boldsymbol{\Gamma}_t) d\mathbf{F}_t + d\mathbf{U}_t\Big) \\ &= \boldsymbol{\alpha}_{i-1}\Delta_n + (\mathbf{G}_{i-1} + \boldsymbol{\Gamma}_{i-1})\Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U} + o_P(\Delta_n),\end{aligned} \tag{4.1}$$

At a representative observation time $t = i\Delta_n$, 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:

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

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 window I_t^n.

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

4.1 Known Factor Case

Here we follow the standard simplified notation in the literature: $\mathbf{G}_{i-1} := \mathbf{G}_{(i-1)\Delta_n}$, $\mathbf{P}_{i-1} := \mathbf{P}_{(i-1)\Delta_n}$ and $\boldsymbol{\Gamma}_{i-1} := \boldsymbol{\Gamma}_{(i-1)\Delta_n}$. In the known factor case, we also observe $\{\Delta_i^n \mathbf{F}\}_{i \in I_t^n}$ in each interval. The key component of the estimation is the projection matrix. To estimate $\mathbf{G}_t$, we first project the high frequency returns within the above local window onto the space of sieve basis of the instruments, and obtain the "projected return" $\{\mathbf{P}_{i-1}\Delta_i^n \mathbf{Y}\}_{i \in I_t^n}$. Then we run OLS of the projected returns on the factors, leading to the estimated instrumental beta

$\widehat{\mathbf{G}}_t$ at time t. Following a similar procedure, one can estiamte $\mathbf{\Gamma}_t$. These two estimators are given by

$$\begin{aligned}\widehat{\mathbf{G}}_t &= \sum_{i\in I_t^n} \mathbf{P}_{i-1}\,\Delta_i^n\mathbf{Y}\,\Delta_i^n\mathbf{F}' \left(\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\,\Delta_i^n\mathbf{F}'\right)^{-1}, \\ \widehat{\mathbf{\Gamma}}_t &= \sum_{i\in I_t^n}(\mathbf{I}_p-\mathbf{P}_{i-1})\Delta_i^n\mathbf{Y}\,\Delta_i^n\mathbf{F}' \left(\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\,\Delta_i^n\mathbf{F}'\right)^{-1},\end{aligned} \tag{4.2}$$

Then the l-th component of $\widehat{\mathbf{G}}_t$ and $\widehat{\mathbf{\Gamma}}_t$, denoted by $\widehat{\mathbf{g}}_{lt}$ and $\widehat{\boldsymbol{\gamma}}_{lt}$, respectively represent the estimated instrumental beta and idiosyncratic beta for the l-th stock. At those discrete observational time points (when $t = i\Delta_n$), these are written as $\widehat{\mathbf{g}}_{l,i\Delta_n}$ and $\widehat{\boldsymbol{\gamma}}_{l,i\Delta_n}$. But we follow the more standard simplified notation that are commonly used literature, write $\widehat{\mathbf{g}}_{l,i} := \widehat{\mathbf{g}}_{l,i\Delta_n}$, and $\widehat{\boldsymbol{\gamma}}_{l,i} := \widehat{\boldsymbol{\gamma}}_{l,i\Delta_n}$ for discrete time estimators. The integrated beta components, i.e. $\int_0^T \mathbf{g}_{ls}ds$ and $\int_0^T \boldsymbol{\gamma}_{ls}ds$, are respectively estimated by, for example, using the overlapping spot estimates[4]

$$\sum_{i=1}^{\lfloor T/\Delta_n\rfloor-k_n} \widehat{\mathbf{g}}_{l,i}\Delta_n, \qquad \sum_{i=1}^{\lfloor T/\Delta_n\rfloor-k_n} \widehat{\boldsymbol{\gamma}}_{l,i}\Delta_n.$$

4.2 Unknown Factor Case

When factors are unknown, we employ the principal component (PCA) method to estimate the latent factors first. But different from Stock and Watson (2002); Bai (2003), Aït-Sahalia and Xiu (2017) and Pelger (2016), we employ the PCA on the projected returns. With each local window I_t^n, we can define the following $p \times k_n$ matrix:

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

[4] One can also use the non-overlapping spot estimates (for some $j \in \{1, \cdots, k_n\}$):

$$\sum_{i=0}^{\lfloor T/(k_n\Delta_n)\rfloor-1} \widehat{\mathbf{g}}_{l,ik_n+j}k_n\Delta_n, \qquad \sum_{i=0}^{\lfloor T/(k_n\Delta_n)\rfloor-1} \widehat{\boldsymbol{\gamma}}_{l,ik_n+j}k_n\Delta_n.$$

In fact, the overlapping estimator is the average of k_n different but highly correlated non-overlapping estimators (with $j = 1, \cdots, k_n$). Hence they have the same asymptotic behavior.

Define the estimated factors

$$\widehat{\Delta^n \mathbf{F}} = (\widehat{\Delta_i^n \mathbf{F}} : i \in I_t^n)' = (\widehat{\Delta_{\lfloor t/\Delta_n \rfloor+1}^n \mathbf{F}}, ..., \widehat{\Delta_{\lfloor t/\Delta_n \rfloor+k_n}^n \mathbf{F}})', \quad k_n \times K,$$

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

$$\begin{aligned} \widehat{\mathbf{G}}_t^{\text{latent}} &= \frac{1}{k_n\Delta_n} \sum_{i \in I_t^n} \mathbf{P}_{i-1}(\Delta_i^n \mathbf{Y}) \widehat{\Delta_i^n \mathbf{F}}', \\ \widehat{\mathbf{\Gamma}}_t^{\text{latent}} &= \frac{1}{k_n\Delta_n} \sum_{i \in I_t^n} (\mathbf{I}_N - \mathbf{P}_{i-1})(\Delta_i^n \mathbf{Y}) \widehat{\Delta_i^n \mathbf{F}}' \end{aligned} \tag{4.3}$$

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 instrumental beta and idiosyncratic beta for the l-th stock. The superscript "latent" indicates that the estimators are defined for the case of latent factors.

We now give an intuitive explanation on the rationale of this procedure. Apply the projection to the discretized model:

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

By the identification conditions $\mathbb{E}(\mathbf{\Gamma}_t|\mathbf{X}_t) = 0$ 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 drifts, we have

$$\mathbf{P}_{i-1}\Delta_i^n \mathbf{Y} \approx \mathbf{G}_{i-1}\Delta_i^n \mathbf{F} \tag{4.4}$$

is nearly "noise-free". Therefore in the case of known factors, running OLS on each local interval I_t^n for (4.4) directly leads to consistent estimator of spot $\mathbf{G}_t$. In the case of unknown factors, (4.4) also implies, for $\Delta^n \mathbf{F} = (\Delta_i^n \mathbf{F} : i \in I_t^n)'$, which is $k_n \times K$,

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

This implies that the columns of $\Delta^n \mathbf{F}$ are approximately the eigenvectors of the "idiosyncraticfree" 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 $(\mathbf{P}\Delta^n\mathbf{Y})'_t(\mathbf{P}\Delta^n\mathbf{Y})_t$. This is the intuition that why our procedure is very robust to the strength and variations of $\mathbf{\Gamma}_t$.

Furthermore, take the difference between (4.1) and (4.4) yields:

$$\begin{aligned}(\mathbf{I}_p - \mathbf{P}_{i-1})\Delta_i^n\mathbf{Y} =& \mathbf{\Gamma}_{i-1}\Delta_i^n\mathbf{F} + \Delta_i^n\mathbf{U} + \text{higher-order term} \\ &+ \text{projection \& sieve approx. error.}\end{aligned} \tag{4.5}$$

(4.5) shows that $\mathbf{\Gamma}_{i-1}$ represents the sensitivity to the risk factors of the remaining components of returns, after the instrument effect is conditioned. Hence a local OLS leads to the estimated $\mathbf{\Gamma}_{i-1}$.

4.3 Jump-robust estimators and Micro-structure noise

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:

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

where $\Delta_i^n Z_{\psi_n^Z} := \Delta_i^n Z_l\, 1_{\{\|\Delta_i^n Z_l\| \le \psi_n^{Z_l}\}}$ 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)).[5]

In the unknown factor case, we only need to replace each $\Delta_i^n\mathbf{Y}$ with its corresponding truncated versions. The estimates $(\Delta_i^n\mathbf{Y}_{\psi_n^Y})_{i\in I_t^n}$ will converge to the true increments of the continuous factor.

[5]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 $[0, t]$.

4.4 Selected Possible Alternative Estimation Methods

Ordinary PCA. When the factors are unknown, the ordinary PCA is a directly competing method. It would estimate latent factors using the leading eigenvectors of the $k_n \times k_n$ matrix $(\Delta^n \mathbf{Y})_t'(\Delta^n \mathbf{Y})_t$ where $(\Delta^n \mathbf{Y})_t$ denotes the $p \times k_n$ matrix of the return data $\Delta_t^n \mathbf{Y}$ for $j \in I_t$. Our method, using the PCA on the projected return data, has at least three advantages over the ordinary PCA. First of all, the projection removes the effect of $\Delta_i^n \mathbf{U}$. This potentially leads to more accurate estimations when k_n is small. Secondly, the projection removes the idiosyncratic beta $\boldsymbol{\gamma}_{lt}$, which is the key to the robustness to the strength of $\boldsymbol{\gamma}_{lt}$ and to the uniform inferences. Finally, when using the estimated factors for forecasting integrated volatilities, as we shall show later, our method allows $\boldsymbol{\gamma}_t$ to be time-varying. Instead, the ordinary PCA would require $\boldsymbol{\gamma}_t$ be time-invariant over the entire time span, which is a very restrictive condition since it captures high-frequency movements in beta.

Time series regression. When factors are known, a seemingly competing method is time series regressions. For instance, suppose we parametrize $\mathbf{g}_{li} = \boldsymbol{\theta}_l' \mathbf{X}_{li}$, and run time series regression on the fixed l th equation:

$$\arg\min_{\boldsymbol{\theta}_l, \boldsymbol{\gamma}_l} \sum_{j \in I_t^n} [\Delta_i^n Y_l - ((\Delta^n \mathbf{F} \circ \mathbf{X}_l)_j \boldsymbol{\theta}_l + \Delta_j^n \mathbf{F} \boldsymbol{\gamma}_l)]^2, \quad (\Delta^n \mathbf{F} \circ \mathbf{X}_l)_j = \Delta_j^n \mathbf{F} \mathbf{X}_{l,j-1}'.$$

In fact, this approach does not work because $\mathbf{X}_{li}$ is nearly time-invariant on each local window, resulting in nearly multicollinearity in $\{(\Delta^n \mathbf{F} \circ \mathbf{X}_l)_j, \Delta_j^n \mathbf{F}\}_{j \in I_t^n}$. Even if the degree of time-variation in $\{\mathbf{X}_{lt}\}$ is large, the rate of convergence for the estimated $\mathbf{g}_{lt}$ would be $O_P(k_n^{-1/2})$, which can be very slow when the length of the local interval is small. In fact, $\mathbf{g}_{lt}$ has to be estimated using the cross-sectional information, because the majority of source of variations on the time domain comes only from the factors, which is not sufficient to identify $\mathbf{g}_{lt}$ from $\boldsymbol{\gamma}_{lt}$. In the contrary, the proposed combination of cross-sectional and time series regressions, with $p \to \infty$, is a more appropriate method.

Generalized Method of Moments. Finally, our method is also closely related to GMM. Fix t, we construct the GMM estimator for $(\{\Delta_j^n \mathbf{F}\}_{j \in I_t^n}, \boldsymbol{\beta}_t)$ from the following moment conditions:

$$\mathbb{E} \boldsymbol{\phi}_{lt} [\Delta_j^n \mathbf{Y} - \boldsymbol{\beta}_j \Delta_j^n \mathbf{F}]' = 0, \quad j \in I_t^n, \quad l = 1, .., p.$$

With the weight matrix $\boldsymbol{\Omega} = (\boldsymbol{\Phi}_t' \boldsymbol{\Phi}_t)^{-1}$, the GMM optimization is given by

$$\min_{\boldsymbol{\beta}_t, \Delta^n \mathbf{F}} \operatorname{tr} \{ [(\Delta^n \mathbf{Y})_t' \boldsymbol{\Phi}_t - \Delta^n \mathbf{F} \boldsymbol{\beta}_t' \boldsymbol{\Phi}_t] \boldsymbol{\Omega} [(\Delta^n \mathbf{Y})_t' \boldsymbol{\Phi}_t - \Delta^n \mathbf{F} \boldsymbol{\beta}_t' \boldsymbol{\Phi}_t]' \}, \quad \text{s.t. } \Delta^n \mathbf{F}' \Delta^n \mathbf{F} = \mathbf{I} \Delta_n k_n.$$

Concentrating out $\boldsymbol{\beta}_t = \frac{1}{k_n\Delta_n}(\Delta^n\mathbf{Y})_t\Delta^n\mathbf{F}$ using the first order condition, we obtain the GMM estimator for $\Delta^n\mathbf{F}$, whose rows are $\sqrt{\Delta_n k_n}$ times the eigenvectors of the $k_n \times k_n$ matrix $(\Delta^n\mathbf{Y})_t'\mathbf{P}_t(\Delta^n\mathbf{Y})_t$ corresponding to the first K eigenvalues. This estimator is asymptotically the same as the proposed $\widehat{\Delta^n\mathbf{F}}$. But the major difference is that the GMM estimator does not take into account the local time variations in $\{\mathbf{P}_j : j \in I_t^n\}$, while our estimator does, although the time variation is small due to the properties of Brownian motions. So this can be understood as an (approximate) GMM interpretation of the proposed factor estimators. Nevertheless, we would still apply OLS on the projected return data to estimate $\mathbf{G}_t$ and $\boldsymbol{\Gamma}_t$.

5 Formal Treatments

5.1 Assumptions

We assume that the following conditions hold uniformly over a class of DPG's: $\mathbb{P} \in \mathcal{P}$. By *absolute constants*, we mean constants that are given, and do not depend the specific data generating process in $\mathcal{P}$. We apply the standard assumptions to define the stochastic processes as follows (e.g., Protter (2005) and Jacod and Protter (2011)).

Assumption 5.1 (Data Generating Process)**.** *(i) The process* $\mathbf{Y}$ *is an Itô semimartingale (with its continuous component is given by* (3.1)*, where the continuous component of* $\mathbf{F}$ *and* $\mathbf{U}$ *are given by* (3.2)*. We assume the jump components of* $\mathbf{Y}$ *and* $\mathbf{F}$ *are of finite variation. Almost surely, the processes* $\{\boldsymbol{\alpha}_{lt}\}_{t\geq 0}$, $\{((\boldsymbol{\beta}_{lt}\boldsymbol{\sigma}_t^F)', \boldsymbol{\sigma}_{lt}^{U'})'\}_{t\geq 0}$, $\{\mathbf{g}_{lt}\}_{t\geq 0}$ *and* $\{\boldsymbol{\gamma}_{lt}\}_{t\geq 0}$ *have càdlàg (right continuous with left limits) and locally bounded paths uniformly in* $l \leq p$ *(see Protter (2005) and Jacod and Protter (2011) for details), where* $\boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_{lt}', \boldsymbol{\sigma}_{lt}^{U'}$ *are the* l *th element (or row) of* $\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t, \boldsymbol{\sigma}_t^U$*, defined in* (3.1) *and* (3.2)*.*

(ii) Each element $\boldsymbol{\theta}_{mt}$ *of* $\boldsymbol{\theta}_t = (\{\mathbf{X}_{lt}'\}, \boldsymbol{\Gamma}_t', \boldsymbol{\sigma}_t^{F'}, \boldsymbol{\sigma}_t^{U'})'$*, is a multivariate Itô semimartingale with the form*

$$\boldsymbol{\theta}_{mt} = \boldsymbol{\theta}_{m0} + \int_0^t \tilde{\boldsymbol{\alpha}}_{ms}ds + \int_0^t \tilde{\boldsymbol{\sigma}}_{ms}d\mathbf{W}_s^m + \int_0^t \check{\boldsymbol{\sigma}}_{ls}d\check{\mathbf{W}}_s^m + \sum_{s\leq t}\Delta\boldsymbol{\theta}_{ms}$$

Here $\{\boldsymbol{\alpha}_{ms}\}_{s\geq 0}, \{\tilde{\boldsymbol{\sigma}}_{ms}\}_{s\geq 0}$ *and* $\{\check{\boldsymbol{\sigma}}_{ms}\}_{s\geq 0}$ *are optional processes and locally bounded uniformly in* $m \leq p$*. In general,* $\boldsymbol{\theta}_m = \{\boldsymbol{\theta}_{mt}\}_{t\geq 0}$ *can be driven by* $\{\mathbf{W}_t^m = (\mathbf{W}^{F'}, \mathbf{W}_m^U)_t'\}_{t\geq 0}$*, where* $\mathbf{W}_m^U$ *is the* m*-th element of* $\mathbf{W}^U$ *introduced in* (3.2)*, and another multi-dimensional Brownian motion* $\{\check{\mathbf{W}}_t^m\}_{t\geq 0}$ *orthogonal to* $\{\mathbf{W}_t^m\}_{t\geq 0}$*. Finally,* $\Delta\boldsymbol{\theta}_{ms}$ *represents the (possible) jump of*

$\boldsymbol{\theta}_m$ *at time* s. [6]

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

Recall that $\mathbf{G}_t$ is the $p \times K$ matrix of $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, $\boldsymbol{\phi}_{lt} = (\phi_1(\mathbf{X}_{lt}), ..., \phi_J(\mathbf{X}_{lt}))'$ is the vector of sieve transformations, and $\mathbf{P}_t = \boldsymbol{\Phi}_t(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'$. In addition, let

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

Assumption 5.2 (Smoothness with respect to time)**.** *Let* $\widetilde{\mathbf{g}}_l(t, \mathbf{x}) := \mathbf{g}_{lt}(\mathbf{x})$. *Then* $\widetilde{\mathbf{g}}_l$ *and* ϕ_j *are differentiable, satisfying: there is* $C > 0$ *so that*

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

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

The above assumption ensures that $\mathbf{g}_{lt}(\mathbf{X}_{lt})$ and $\boldsymbol{\phi}_{lt}$ are smooth transformations of $\mathbf{X}_{lt}$, so are also semimartingales, and thus the local time-variation error of $\mathbf{X}_{lt}$ carries over to $\mathbf{g}_{lt}(\mathbf{X}_{lt})$ and $\mathbf{P}_t$.

Assumption 5.3 (Moment Bounds)**.** *There are absolute constants* $c, C, \eta > 0$, *so that*
(i) $\max_{l\leq p, t\in[0,T]} \mathbb{E}\,\|\mathbf{g}_{lt}(\mathbf{X}_{lt})\|^4 < C$, $\mathbb{E}\,\|\boldsymbol{\gamma}_{lt}\|^4 < C$.
(ii) $\max_{t\in[0,T], l,m\leq p} \mathbb{E}\,h_{t,lm}^4 \leq C$.
(iii) $c < \min_{t\leq T} \lambda_{\min}(\frac{1}{p}\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t) \leq \max_{t\leq T} \lambda_{\max}(\frac{1}{p}\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t) < C$.
(iv) $\max_{t\in[0,T]} \|\mathbf{G}_t - \mathbf{P}_t\mathbf{G}_t\|_\infty \leq C_1 J^{-\eta}$. *Conditions (iii) (iv) hold almost surely.*

We now describe the asymptotic variance of $\widehat{\mathbf{g}}_{lt}$. Let $\mathbf{u}_t := \Delta_t^n\mathbf{U}/\sqrt{\Delta_n}$ and $\mathbf{f}_t = \Delta_t^n\mathbf{F}/\sqrt{\Delta_n}$. Let $\mathbf{c}_{f,t}$ $(K \times K)$ and $\mathbf{c}_{u,t}$ $(p \times p)$ be the instantaneous quadratic variation process of $\mathbf{F} = \{\mathbf{F}_t\}_{t\geq 0}$ and $\mathbf{U} = \{\mathbf{U}_t\}_{t\geq 0}$, respectively, that is, $\mathbf{c}_{f,t} = d[\mathbf{F}, \mathbf{F}]_t/dt$ and $\mathbf{c}_{u,t} = d[\mathbf{U}, \mathbf{U}]_t/dt$, $\forall t \in [0, T]$. In addition, let $\mathbf{s}_{f,t} = \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n} \Delta_i^n\mathbf{F}\Delta_i^n\mathbf{F}'$ and $h_{i,ml} := \boldsymbol{\phi}_{i,m}'(\frac{1}{p}\boldsymbol{\Phi}_i'\boldsymbol{\Phi}_i)^{-1}\boldsymbol{\phi}_{i,l}$. The asymptotic variance of $\widehat{\mathbf{g}}_{lt}$ depends on:

$$\begin{aligned}
\mathbf{V}_{u,t} &= \mathbf{s}_{f,t}^{-1}\frac{1}{pk_n}\sum_{i\in I_t^n} \mathbf{s}_{f,i}\boldsymbol{\phi}_{i-1,l}'(\frac{1}{p}\boldsymbol{\Phi}_{i-1}'\boldsymbol{\Phi}_{i-1})^{-1}\boldsymbol{\Phi}_{i-1}'\mathbf{c}_{u,i}\boldsymbol{\Phi}_{i-1}(\frac{1}{p}\boldsymbol{\Phi}_{i-1}'\boldsymbol{\Phi}_{i-1})^{-1}\boldsymbol{\phi}_{i-1,l}\,\mathbf{s}_{f,t}^{-1}, \\
\mathbf{V}_{\gamma,t} &= \mathrm{Var}(\frac{1}{\sqrt{p}}\sum_{m=1}^{p} \boldsymbol{\gamma}_{mt}h_{t,ml}|\mathbf{X}_t).
\end{aligned}$$

[6] For any stochastic process Z and any time point s, let $Z_{s-} := \lim_{u\uparrow s} Z_u$ be the left limit of Z at time s. Then $\Delta Z_s := Z_s - Z_{s-}$.

Let u_{mt} denote the mth ($m \leq p$) element of $\mathbf{u}_t$. Also recall that $\mathbf{\Gamma}_t$ is the $p \times K$ matrix of $\boldsymbol{\gamma}_{mt}$.

Assumption 5.4 (Cross-sectional Weak Dependence)**.** *There are absolute constants $c, C > 0$, so that*

(i) Almost surely,

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

(ii) If $\{\boldsymbol{\gamma}_{mt}\}_{m\leq p} \neq 0$, 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) $\sup_{t\in[0,T]}\|\mathbb{E}(\mathbf{\Gamma}_t\mathbf{\Gamma}_t'|\mathbf{X}_t)\| < C$, $\inf_{t\in[0,T]}\lambda_{\min}(\mathbf{c}_{t,f}) > c$. *Finally,*

$$\sup_{t\in[0,T]}\frac{1}{p^2}\sum_{m,l\leq p}\sum_{m',l'\leq p}|Cov(u_{mt}u_{lt}, u_{m't}u_{l't})| < C, \qquad \sup_{t\in[0,T]}\max_{m\leq p}\frac{1}{p}\sum_{l,l\leq p}|Cov(u_{mt}u_{lt}, u_{mt}u_{l't})| < C,$$

and $\sup_{t\in[0,T]}\max_{m\leq p}\sum_{l=1}^{p}|\mathbb{E}u_{mt}u_{lt}| < C$.

Assumption 5.4 (i) requires the conditional covariance of standardized idiosyncratic components have bounded eigenvalues. This condition holds when the idiosyncratic components are cross-sectionally weakly correlated. Assumption 5.4 (ii) requires the cross-sectional variations of $\{\boldsymbol{\gamma}_{mi}\}$, if nonzero, be driven by sufficiently weakly dependent random sequences, so that the cross-sectional central limit theorem (CLT) holds. We do not make any condition on the lower bound of eigenvalues of $\mathbf{V}_{\gamma,t}$, so the considered class of DGP's is robust to the strength of the cross-sectional variations of $\mathbf{\Gamma}_t$. Condition (iii) requires $\{\mathbf{u}_t, \mathbf{\Gamma}_t\}$ are cross-sectionally weakly uncorrelated.

Assumption 5.5 (For estimated factors)**.** *(i) Define $\mathbf{\Sigma}_{G,t} = \frac{1}{p}\mathbf{G}_t'\mathbf{G}_t$. Almost surely, $c < \inf_{t\leq T}\lambda_{\min}(\mathbf{\Sigma}_{G,t}) \leq \sup_{t\leq T}\lambda_{\max}(\mathbf{\Sigma}_{G,t}) < C$ for absolute constants $c, C > 0$.*
(ii) The eigenvalues of $\mathbf{\Sigma}_{G,t}^{1/2}\,\mathbf{c}_{f,t}\,\mathbf{\Sigma}_{G,t}^{1/2}$ are distinct.

Assumption 5.5 is similar to the *pervasive condition* in the approximate factor model's literature, which identifies the latent factors (up to a rotation).

5.2 Asymptotic Normality and Uniform Bias Correction

We first present the estimated spot $\mathbf{g}_{lt}$ when factors are observable.

Theorem 5.1 (known factor case)**.** *Suppose* $J^2 = O(p)$*,* $k_n p J^{-2\eta} + p k_n^2 \Delta_n = o(1)$*. Under Assumptions 5.1-5.4, as* $J, p \to \infty$*, (*k_n *either grows or stays constant)*

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

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 $\frac{1}{pk_n\Delta_n}(\mathbf{P}\Delta^n\mathbf{Y})_t'(\mathbf{P}\Delta^n\mathbf{Y})_t$. Let

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

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

Here $\mathbf{P}_{i,l}$ denotes the l-th column of $\mathbf{P}_i$. We have the following theorem.

Theorem 5.2 (unknown factor case)**.** *In addition to the assumptions for Theorem 5.1, suppose* $k_n J = o(p^2)$*. Then*

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

As in the known factor case, $\mathbf{V}_{\gamma,t}$ directly impacts on the rate of convergence and limiting distribution of $\widehat{\mathbf{g}}_{lt}$. We make several remarks.

Remark 5.1. If $\|\mathbf{V}_{\gamma,t}\| = o_P(k_n^{-1})$, then the rate of convergence is $O_P((k_n p)^{-1/2})$, and

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

Intuitively, this occurs when idiosyncratic betas have weak signals from the cross-sectional variations. As a result, the observed instruments captures almost all the beta fluctuations, leading to a fast rate of convergence on the spot level. 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 $\boldsymbol{\gamma}$. Intuitively, this means when idiosyncratic betas have strong cross-sectional variations, time-domain averaging is not helpful to remove their effect on estimating $\mathbf{g}_{lt}$, and only cross-sectional projection does the job. This leads to a slower rate of convergence.

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 $\mathbf{V}_{\gamma,t}$. In fact, its eigenvalues can be any sequences in a large range, resulting in sophisticated rate of convergence for $(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt})$.

Remark 5.2. The similar phenomena is also present in the case of estimated factors. But it is also interacting with the bias. Since the bias has an order $O_P(p^{-3/2})$, we actually have: if $\|\mathbf{V}_{\gamma,t}\| = o_P(k_n^{-1})$,

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

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*:

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

Therefore when the signals from $\boldsymbol{\gamma}$ is sufficiently strong, the rate of convergence slows down, and dominates the bias arising from the effect of estimating factors.

Remark 5.1 (Interplay between $(k_n, p, \Delta_n, J^\eta)$)**.** We require $k_n p J^{-2\eta} + p k_n^2 \Delta_n = o(1)$, which reflects two major approximation errors in our context: (i) the sieve approximation for $\mathbf{g}_{lt}$ and (ii) the use of local windows and treating $\mathbf{g}_{lt}$ to be locally time-invariant. The first condition $k_n p J^{-2\eta} = o(1)$ ensures that the sieve approximation error is first order- asymptotically negligible compared to the leading order. On the other hand, due to the Itô semimartingale assumption, and by the Burkholder-Davis-Grundy inequality (cf. Chapter 2 of **?**), they obey

$$\mathbb{E}(|\mathbf{g}_{l,i-1} - \mathbf{g}_{l,t}|^2 \,\big|\, \mathcal{F}_{i-1}) \leq L k_n \Delta_n$$

for any $i \in I_t^n$, where L is a positive finite number that is uniform in t. As a result, we could control for this time-domain smoothness error with the assumption $p k_n^2 \Delta_n = o(1)$.

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. Recall that,

$$\text{BIAS}_g = \mathbf{M}_t \frac{1}{k_n\sqrt{p}} \sum_{i \in I_t^n} \mathbf{P}_{i-1}\, \mathbf{c}_{u,i-1}\, \mathbf{P}_{i-1,l}$$

Here $\mathbf{M}_t$ 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 error covariance matrix $\mathbf{c}_{u,i-1}$, which is high-dimensional when p is large. We consider three 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 variance $\mathbf{c}_{u,i-1}$ 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:

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

CASE II: cross-sectionally weakly correlated (sparse)

In this case $\mathbf{c}_{u,i-1}$ 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" in the recent statistical literature (e.g., Bickel and Levina (2008); Fan et al. (2013)) can be applied, yielding a nearly $\min\{k_n, p\}^{1/2}$- consistent sparse covariance estimator $\widehat{\mathbf{c}}_{u,i-1}$. 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:

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

where $\text{th}(\cdot)$ is a thresholding function, whose typical choices are the hard-thresholding and soft-thresholding. Here the threshold value $\varrho_{dl} = \bar{C}(s_{dd} s_{ll})^{1/2} \omega_{np}$, with
$\omega_{np} = \sqrt{\frac{\log p}{k_n}} + \frac{J}{p} \max_{j,d} \frac{1}{J} \|\boldsymbol{\phi}(\mathbf{x}_{jd})\|^2 \sqrt{\log J}$.[7]

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

CASE III: cross-sectionally weakly correlated but not sparse

When the cross-sectional correlation is not as weak as being "sparse", it is hard to directly estimate a high-dimensional conditional covariance matrix. But note that the co-

[7]Hard-threholding takes $\text{th}(s_{dl}) = s_{dl}$, while soft-thresholding takes $\text{th}(s_{dl}) = \text{sgn}(s_{dl})(|s_{dl}| - \varrho_{dl})$. We shall justify the choice of ω_{np} in Section C.6. 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.

variance appears in the bias through the covariance of projected error $\mathbf{P}_j\Delta_j^n\mathbf{U}$, which can be directly estimated using the projection procedure: let $\widehat{\mathbf{P}_{i-1}\Delta_i^n\mathbf{U}} = \mathbf{P}_{i-1}\Delta_i^n\mathbf{Y} - \widehat{\mathbf{G}}_{i-1}\widehat{\Delta_i^n\mathbf{F}}$ and $(\widehat{\mathbf{P}_{i-1}\Delta_i^n\mathbf{U}})_l'$ denotes the transposed l-th row of $\widehat{\mathbf{P}_{i-1}\Delta_i^n\mathbf{U}}$.

$$\widehat{\text{BIAS}}_g = \widehat{\mathbf{M}}_t\frac{1}{k_n\Delta_n\sqrt{p}}\sum_{i\in I_t^n}\widehat{\mathbf{P}_{i-1}\Delta_i^n\mathbf{U}}(\widehat{\mathbf{P}_{i-1}\Delta_i^n\mathbf{U}})_l'.$$

This procedure avoids directly estimating the residuals $\Delta_j^n\mathbf{U}$, and is advantageous since $\mathbf{G}_i$ can be estimated at a potentially much faster convergence rate than the betas.

Formally, we focus on CASE I and CASE II for the bias correction in the following theorem.

Theorem 5.3 (Bias correction)**.** *Suppose $Jk_n = o(p^3)$. Consider CASE I and CASE II for estimated factors. In particular, assume that for* $\max_{i\in I_t^n}\|\widehat{\mathbf{c}}_{u,t} - \mathbf{c}_{u,i}\| = o_P(\sqrt{\frac{p}{Jk_n}})$[8] *in CASE II. Define the bias-corrected instrumental beta estimator $\widetilde{\mathbf{g}}_{lt}^{\text{latent}} = \widehat{\mathbf{g}}_{lt}^{\text{latent}} - \widehat{\text{BIAS}}_g$. Under Assumptions 5.1-5.5, we have*

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

The limiting distribution of $\boldsymbol{\gamma}_{lt}$ has a similar behavior, and features a similar bias-correction procedure in the estimated factor case. We omit the formal results for brevity.

5.3 Uniform Confidence Intervals Using Cross-Sectional Bootstrap

The presented asymptotic results display a discontinuity issue in the distribution of the estimated $\widehat{\mathbf{g}}_{lt}$. The estimated $\mathbf{g}_{lt}$ has the following asymptotic expansion.

$$\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt} = \underbrace{\frac{1}{p}\sum_{m=1}^p\boldsymbol{\gamma}_{m,t}h_{t,ml}}_{(\mathbf{a})} + \underbrace{\mathbf{s}_{f,t}^{-1}\frac{1}{k_np\Delta_n}\sum_{i\in I_t^n}\sum_{m=1}^p\Delta_i^n\mathbf{F}\Delta_i^nU_mh_{i-1,ml}}_{(\mathbf{b})} + \text{negligible terms} \quad (5.1)$$

Let $\sqrt{p}\mathbf{V}_{\gamma,t}$ and $\sqrt{k_np}\mathbf{V}_{u,t}$ respectively denote the asymptotic variances of terms ($\mathbf{a}$) and ($\mathbf{b}$) in the expansion of (5.1). In particular, when $\{\boldsymbol{\gamma}_{mt}\}_{m\le p}$ are cross-sectionally uncorrelated, $\mathbf{V}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^p h_{t,ml}^2\,\text{Var}(\boldsymbol{\gamma}_{mt}|\mathbf{X}_t)$. If we use the standard "plug-in" procedure to make

[8] We shall verify this condition in Section C.6 for sparse covariance estimators.

inference about $\mathbf{g}_{lt}$, we would estimate $\mathbf{V}_{\gamma,t}$ by White (1980)'s heteroskedastic covariance estimator $\widehat{\mathbf{V}}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^2 \widehat{\boldsymbol{\gamma}}_{mt}\widehat{\boldsymbol{\gamma}}_{mt}'$, whose estimation error is

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

But as we discussed in Section 2, the γ-estimation error is lower bounded by a rate of $O_P(k_n^{-1})$, which is not negligible when (**a**) is the dominating term, resulting in possibly very conservative inferences. As such, we propose to use cross-sectional bootstrap to achieve uniform inference.

5.3.1 Independent cross-sectional bootstrap

For simplicity, we start with assuming cross-sectional independence of $\{\Delta_i^n U_m\}_{m\leq p}$ and $\{\boldsymbol{\gamma}_{mt}\}_{m\leq p}$. Let $T_n = \lfloor T/\Delta_n \rfloor$. Independently resample time series $\{\Delta_i^n Y_m^*, i \in \{1,\cdots,T_n\}\}_{m=1,\ldots,p}$ and $\left\{\mathbf{X}_{mi}^* : i \in I_t^n\right.$ $_{m=1,\ldots,p}$, where

$$\begin{aligned}
\{\Delta_i^n Y_m^*, i \in \{1,\cdots,T_n\}\}_{m=1,\ldots,p} &= \Big\{\{\Delta_i^n Y_{m_1}, i \in \{1,\cdots,T_n\}\}, \ldots, \{\Delta_i^n Y_{m_p}, i \in \{1,\cdots,T_n\}\}\Big\} \\
\left\{\mathbf{X}_{mi}^* : i \in I_t^n\right._{m=1,\ldots,p} &= \Big\{\left\{\mathbf{X}_{m_1,i} : i \in \{1,\cdots,T_n\}\right., \cdots, \left\{\mathbf{X}_{m_p,i} : i \in \{1,\cdots,T_n\}\right.\Big\}.
\end{aligned}$$

Here $\{m_1, \ldots, m_p\}$ is a simple random sample with replacement from $\{1, \ldots, p\}$. When we are interested $\mathbf{g}_{lt}$ for the l-th specific stock, we always fix $m_1 = l$ in the resampled data. We do not need to mimic the time series variations, so for each sampled index $m \in \{m_1, \ldots, m_p\}$, the entire time series $\{\Delta_i^n Y_m, i \in \{1,\cdots,T_n\}\}$ and $\{\mathbf{X}_{m,i}, i \in \{1,\cdots,T_n\}\}$ are kept. In addition, we keep the entire time series $\{\Delta_i^n \mathbf{F} : i \in \{1,\cdots,T_n\}\}$ in the case of known factors, and $\{\widehat{\Delta_i^n \mathbf{F}} : i \in \{1,\cdots,T_n\}\}$ in the case of unknown factors.[9]

We then let $\boldsymbol{\Phi}_i^* = (\boldsymbol{\phi}_{m_1,i}, \ldots, \boldsymbol{\phi}_{m_p,i})'$ and $\mathbf{P}_i^* = \boldsymbol{\Phi}_i^*(\boldsymbol{\Phi}_i^{*'}\boldsymbol{\Phi}_i^*)^{-1}\boldsymbol{\Phi}_i^{*'}$. Let $\mathbf{P}_{i,l}^*$ be the l th column of $\mathbf{P}_i^*$. Let $\Delta_i^n \mathbf{Y}^* = (\Delta_i^n Y_1^*, \ldots, \Delta_i^n Y_p^*)'$ and $\mathbf{G}_i^* = (\mathbf{g}_{m_1,i}, \ldots, \mathbf{g}_{m_p,i})'$. Define

$$\widehat{\mathbf{g}}_{lt}^* = \left(\sum_{i\in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{F}'\right)^{-1} \sum_{i\in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{i-1,l}^* \tag{5.2}$$

[9] 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. Even if we did, its effect would be first-order negligible.

in the case of known factors, and

$$\widehat{\mathbf{g}}_{lt}^{*\,\mathrm{latent}} = \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}\widehat{\Delta_i^n\mathbf{F}}\,\Delta_i^n\mathbf{Y}^{*\prime}\mathbf{P}_{i-1,l}^{*} \tag{5.3}$$

in the estimated factor case. 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:

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

Consequently, the γ-estimation error component is avoided. This forms the fundament to the bootstrap asymptotic validity.

When $\mathbf{g}_{lt}$ is multidimensional, it is easier to present the confidence interval for a linear transformation $\mathbf{v}'\mathbf{g}_{lt}$. The following algorithm summarize the steps for computing the confidence intervals.

Algorithm 5.1. Compute the confidence interval for $\mathbf{v}'\mathbf{g}_{lt}$ (or $\mathbf{v}'\boldsymbol{\Upsilon}_{nt}\mathbf{g}_{lt}$ in the estimated factor case) as follows.

Step 1. Fix $m_1=l$. Take a simple random sample $\{m_2,...,m_p\}$ with replacement from $\{1,...,p\}$.

Step 2. Obtain $\Delta_i^n\mathbf{Y}^*=(\Delta_i^nY_{m_1},...,\Delta_i^nY_{m_p}^*)'$ and $\boldsymbol{\Phi}_i^*=(\boldsymbol{\phi}_{m_1,i},...,\boldsymbol{\phi}_{m_p,i}^*)'$, and compute $\widehat{\mathbf{g}}_{lt}^{*}$ (or $\widehat{\mathbf{g}}_{lt}^{*\,\mathrm{latent}}$ in the case of estimated factors) as in either (5.2) or (5.3).

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

Step 4. Compute the confidence interval as:

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

We need the following conditions for the bootstrap validity.

Assumption 5.6. *(i)* $\{\{\Delta_i^nU_m\}_{i\in I_t^n},\{\boldsymbol{\gamma}_{mi}\}_{i\in I_t^n}\}_{m\le p}$ *are cross-sectionally uncorrelated conditionally on* $\{\mathbf{X}_t\}$.

(ii) Almost surely in the bootstrap sampling space, $\sup_{t\in[0,T]}\|\mathbf{G}_t^* - \mathbf{P}_t^*\mathbf{G}_t^*\|_\infty \le CJ^{-\eta}$ *for absolute constants* $C, \eta > 0$.

Finally, we require the following moment conditions on $\mathbf{\Gamma}_t$. This condition is used to ensure that the LLN-error is negligible regardless of the strengh of $\boldsymbol{\gamma}_{mt}$.

Assumption 5.7. *There is an absolute constant* $C > 0$*, almost surely,*

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

If $\operatorname{Var}(\boldsymbol{\gamma}_{mt}|\mathbf{X}_t) = 0$ *for* $m = 1, ..., p$*, then the above ratios are defined to be zero.*

Theorem 5.4 (Uniformly valid confidence intervals)**.** *Let* $\mathcal{P}$ *be the collection of all data generating processes* $\mathbb{P}$ *for which Assumptions 5.6, 5.8 and assumptions of Theorems 5.1 and 5.2 hold. Then for any fixed vector* $\mathbf{v} \in \mathbb{R}^K\backslash\{0\}$ *such that* $\|\mathbf{v}\| > c > 0$*, for each fixed* $l \le p, t \in [0, T]$

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

We close this section by presenting a uniform confidence interval for the long-run inference. Consider estimating $\mathbf{v}'\int_0^T \mathbf{g}_{lt}dt$ in the case of known factors. It is estimated by $\widehat{\int_0^T \mathbf{g}_{lt}dt} := \sum_{t=1}^{[T/\Delta_n]-k_n} \widehat{\mathbf{g}}_{lt}\Delta_n$. 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 $1-\tau$ th bootstrap quantile of $\{|\mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt}^{*b} - \mathbf{v}'\widehat{\int_0^T \mathbf{g}_{lt}dt}|\}_{b\le B}$. 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 5.8, but for estimating the integrated instrumental effects.

Assumption 5.8. *There is an absolute constant* $C > 0$*, almost surely,*

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

Theorem 5.5 (long-run g). *Consider the known factor case*[10]*. Let $\mathcal{P}$ be the collection of all data generating processes $\mathbb{P}$ for which Assumption 5.6(ii) and assumptions of Theorem 5.1 hold. In addition, suppose $\{\{\Delta_t^n U_m\}_{t\in[0,T]}, \{\boldsymbol{\gamma}_{mt}\}_{t\in[0,T]}\}_{m\leq p}$ are cross-sectionally uncorrelated conditionally on $\{\mathbf{X}_t\}$. Then for any fixed vector $\mathbf{v} \in \mathbb{R}^K\backslash\{0\}$ such that $\|\mathbf{v}\| > c > 0$, for each fixed $l \leq p$,*

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

5.3.2 Block cross-sectional bootstrap

To relax Assumption 5.6(i), we can allow $\{\{\Delta_t^n U_m\}_{t\in[0,T]}, \{\boldsymbol{\gamma}_{mt}\}_{t\in[0,T]}\}_{m\leq p}$ to be cross-sectionally block-dependent, and rely on block cross-sectional bootstraps. More specifically, suppose the cross-sectional index set has a non-overlapping partition $\{1, ..., N\} = B_1 \cup ... \cup B_H$, 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) Firms' block mememberships are known; (that is, $\{B_h\}_{h\leq H}$ are known)[11]

(ii) $\mathrm{Cov}(\Delta_t^n U_{l_1}, \Delta_t^n U_{l_2}) = 0$ and $\mathrm{Cov}(\gamma_{mt,k_1}, \gamma_{mt,k_2}) = 0$ for any $t \in [0, T]$ if the two firms (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, firms are possibly correlated only within the same block. Empirically, the assumption of known blocks of finie 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).

As these blocks are known a priori, they 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 5.2. Compute the confidence interval for $\mathbf{v}'\mathbf{g}_{lt}$ as follows.

[10]Due to the rotation discrepancy, estimating the long-run g with unknown factors is a very challenging problem in the presence of time-varying beta, and we shall leave it for the future research.

[11]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, that is: $\{\Delta_i^n Y_m, i \in \{1, \cdots, T_n\}\}$, $\{\mathbf{X}_{m,i}, i \in \{1, \cdots, T_n\}\}$ for all $m \in B_h^*$. Combine all these time series, we obtain

$$\begin{aligned}\Delta_i^n \mathbf{Y}^* &= (\Delta_i^n Y_1^*, ..., \Delta_i^n Y_{p^*}^*)' = (\Delta_i^n Y_m : m \in B_h^*, h \leq H),\\ \mathbf{\Phi}_i^* &= (\boldsymbol{\phi}_{1,i}^*, ..., \boldsymbol{\phi}_{p^*,i}^*)' = (\boldsymbol{\phi}_{m,i} : m \in B_h^*, h \leq H).\end{aligned}$$

and $\mathbf{P}_i^* = \mathbf{\Phi}_i^* (\mathbf{\Phi}_i^{*'} \mathbf{\Phi}_i^*)^{-1} \mathbf{\Phi}_i^{*'}$. Here $p^* = \sum_{h=1}^H |B_h^*|_0$. Note that $p^* \asymp p$ because each block contains finite number of individuals, so H grows at the same rate as p.

Step 3: Define

$$\begin{aligned}\widehat{\mathbf{g}}_{lt}^* &= \left(\sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{F}' \right)^{-1} \sum_{i \in I_t^n} \Delta_i^n \mathbf{F} \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{i-1,l}^* \quad \text{when factors are known}\\ \widehat{\mathbf{g}}_{lt}^{*\,\text{latent}} &= \frac{1}{k_n \Delta_n} \sum_{i \in I_t^n} \widehat{\Delta_i^n \mathbf{F}}\, \Delta_i^n \mathbf{Y}^{*'} \mathbf{P}_{i-1,l}^* \quad \text{when factors are estimated.}\end{aligned}$$

Step 4: Repeat Step 1-3 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. Let q_τ (or q_τ^{latent}) be the $1 - \tau$ th bootstrap quantile of $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}|\}_{b \leq B}$ (or $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*\,\text{latent},b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}^{\text{latent}}|\}_{b \leq B}$). 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:

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

As for the confidence interval for the integrated instrumental effect $\int_0^T \mathbf{g}_{lt} dt$, the use of block cross-sectional bootstrap is very similar. Given the sampled bootstrap data, the computation of $\widehat{\int_0^T \mathbf{g}_{lt} dt}$ and its associated bootstrap quantile is the same as in Section 5.3.1. Finally, the proof of first-order bootstrap validity is also similar to that of Theorem 5.4 and 5.5, building on the well known results of the validity of block-bootstrap (Andrews, 2004; Lahiri, 1999). We omit the formal proof for technical simplicity.

6 Uniform Confidence Intervals for Long-Run Forecast

The object of interest is to forecast the conditional mean of model:

$$y_{d+h} = \mu y_d + \boldsymbol{\rho}'\mathbf{F}_d + v_{d+h}, \quad d = 1, ..., L_n, \quad L_n \to \infty, \tag{6.1}$$

where $h > 0$ is the lead time between information available and the dependent variable. Here d represents the d th "day", and we observe in total L_n days. On the d th day, we have the integrated factor: $\mathbf{F}_d := \int_{(d-1)T}^{dT} d\mathbf{F}_t$, which is extracted from a high-dimensional discrete-time return data $\{\Delta_i^n \mathbf{Y}\}_{i \le M_n}$, with $M_n = nL_n$, realized from:

$$d\mathbf{Y}_t = \boldsymbol{\alpha} dt + (\mathbf{G}_t + \boldsymbol{\Gamma}_t) d\mathbf{F}_t + d\mathbf{U}_t, \quad \forall t \in [0, L_n T].$$

Hence given the information set $\mathcal{F}_d$ on day d, we approximate $\mathbb{E}(y_{d+h}|\mathcal{F}_d)$ by a diffusion index forecast ideal for a data-rich environment, which as been used frequently in the macroeconomic-forecast literature (e.g., Stock and Watson (2002); Bai and Ng (2006)). The goal is to provide an out-of-sample forecast of $y_{L_n+h|L_n} := \mu y_{L_n} + \boldsymbol{\rho}'\mathbf{F}_{L_n}$. We construct mean forecast $\widehat{y}_{L_n+1|L_n}$, and using the asymptotic properties of the estimated factors and lagged integrated volatility, we construct a forecast interval $[\widehat{y}_{L_n+1|L_n} \pm q_{\tau,n,p}]$ for $y_{L_n+1|L_n}$ (where $q_{\tau,n,p}$ is the critical value), so that

$$\lim_{p,n\to\infty} \sup_{\mathbb{P}\in\mathcal{P}} \left| \mathbb{P}\left(y_{L_n+1|L_n} \in [\widehat{y}_{L_n+1|L_n} \pm q_{\tau,n,p}] \right) - (1-\tau) \right| = 0.$$

Here the probability measure $\mathbb{P}$ is taken uniformly over a broad DGP class $\mathcal{P}$, which admits various strengths of cross-sectional variations in $\boldsymbol{\gamma}_{lt}$, $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, as well as various dynamics on the time-domain. Uniformity in the above sense is essential for inferences in this context, because it makes the inference valid and robust to the unknown sources and degree of dynamics of factor betas.

Of particular interest is $y_d = \log \mathrm{IV}_d$, the log integrated volatility a single asset. As is documented in some of the literature, the volatilities are different from "uncertainty" in that they have predictable components in the stock market (Jurado et al., 2015). Note that model (6.1) and its associated forecasts are in low-frequency discrete time, but we shall use the high-frequency return data to estimate $\{\mathrm{IV}_d, \mathbf{F}_d\}_{d \le L_n}$. Since $\{\mathrm{IV}_d\}_{d \le L_n}$ is not directly observable, and has to be nonparametrically estimated. Then we have two types of *estimated regressors*: estimated integrated volatility $\{\widehat{\mathrm{IV}}_d\}_{d \le L_n}$, and the estimated integrated factors

$\{\widehat{\mathbf{F}}_d\}_{d\leq L_n}$. We estimate the integrated volatility over $[(d-1)T, dT]$ by using the standard truncated Bi-power variation (TBPV) (Jacod and Protter, 2011): for $d = 1, ..., L_n$,

$$\widehat{y}_d := \log \widehat{\mathrm{IV}}_d, \quad \text{where } \widehat{\mathrm{IV}}_d := \sum_{i\in S_d} (\Delta_i^n Y_l)^2 \, 1_{\{\|\Delta_i^n Y_l\| \leq \phi_n^l\}}.$$

In the presence of instrumental betas, the estimated factors are defined using the same method described in Section 3. Specifically, let $\mathbf{P}_{i-1} = \mathbf{\Phi}_{i-1}(\mathbf{\Phi}_{i-1}'\mathbf{\Phi}_{i-1})^{-1}\mathbf{\Phi}_{i-1}'$ defined by the basis $\mathbf{\Phi}_{i-1}$ for $i = 1, ..., M_n$. Let $(\mathbf{P}\Delta^n\mathbf{Y}) = [\mathbf{P}_0\Delta_1^n\mathbf{Y}, ..., \mathbf{P}_{M_n-1}\Delta_{M_n}^n\mathbf{Y}]$ be the $p \times M_n$ matrix. Let $\widetilde{\Delta^n\mathbf{F}}$ be an $M_n \times K$ matrix of estimated factors, whose columns equal $\sqrt{\Delta_n M_n}$ times the eigenvectors of $(\mathbf{P}\Delta^n\mathbf{Y})'(\mathbf{P}\Delta^n\mathbf{Y})$ corresponding to its first K eigenvalues. But for long-time estimations, the effect of accumulated drifts would introduce a biased factor estimation. Hence we use a simple de-biased integrated factor estimator: let S_d be the index of observations in the interval $[(d-1)T, dT]$. Define

$$\widehat{\mathbf{F}}_d := \widetilde{\mathbf{F}}_d - \frac{1}{L_n}\sum_{d=1}^{L_n} \widetilde{\mathbf{F}}_d, \quad \text{where } \widetilde{\mathbf{F}}_d = \sum_{i\in S_d} \widetilde{\Delta_i^n\mathbf{F}}, \quad d = 1, ..., L_n.$$

Finally, the forecasted conditional mean of $y_{L_n+h|L_n} := \mu y_{L_n} + \boldsymbol{\rho}'\mathbf{F}_{L_n}$ is

$$\widehat{y}_{L_n+h|L_n} = \widehat{\mu}\widehat{y}_{L_n} + \widehat{\boldsymbol{\rho}}'\widehat{\mathbf{F}}_{L_n},$$

where

$$(\widehat{\mu}, \widehat{\boldsymbol{\rho}}) = \arg\min_{\mu,\boldsymbol{\rho}} \sum_{d=1}^{L_n-h} [\widehat{y}_{d+h} - (\widehat{\mu}\widehat{y}_d + \widehat{\boldsymbol{\rho}}'\widehat{\mathbf{F}}_d)]^2.$$

6.1 Effect of Time-Varying Gammas on Forecasts

An important aspect of using the integrated factors as predictors is that we would like to employ a forecasting model that exploits as much available information as possible to control for the economic state. However, different from the theoretical findings in the diffusion index literature, the factors cannot be treated as known in the presence of time-varying $\mathbf{\Gamma}$ in the factor loadings. The asymptotic distribution of $\widehat{\mathbf{F}}_d - \mathbf{H}_n\mathbf{F}_d$ (where $\mathbf{H}_n$ is a rotation matrix) depends on a component:

$$\frac{1}{p}\sum_{i\in S_{L_n}} \Delta_i^n\mathbf{F}'\mathbf{\Gamma}_i'\mathcal{G}, \quad \text{where } \mathcal{G} = \mathbf{G}_{L_n}\left(\frac{1}{p}\mathbf{G}_{L_n}'\mathbf{G}_{L_n}\right)^{-1}\boldsymbol{\rho}. \tag{6.2}$$

To understand the new features that this term introduces to the forecast confidence intervals, we now describe the asymptotic property of $y_{L_n+h|L_n} - \widehat{y}_{L_n+h|L_n}$, by introducing some notation: let $\mathbf{z}_d = (y_d, (\mathbf{H}_n\mathbf{F}_d)')'$, and write $\mathbf{w}_n' = (w_1, \mathbf{w}_2') := \mathbf{z}_{L_n}'(\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathbf{z}_d\mathbf{z}_d')^{-1}$, where w_1 denotes the first element of $\mathbf{w}_n$. We show that

$$y_{L_n+h|L_n} - \widehat{y}_{L_n+h|L_n} = \bar{r}_1 + ...\bar{r}_5 + \text{negligible terms}$$

where

$$\begin{aligned}
\bar{r}_1 &= \mathbf{w}_n'\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathbf{z}_d v_{d+h}, \\
\bar{r}_2 &= \sum_{i\in S_{L_n}}\frac{1}{p}\Delta_i^n\mathbf{U}'\mathbf{G}_{L_n} \\
\bar{r}_3 &= \mu(\widehat{y}_{L_n} - y_{L_n}) \quad \text{(effect of nonparametrically estimate the integrated volatility)} \\
\bar{r}_4 &= \left[w_1\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathrm{IV}_d - 1\right](\frac{1}{L_n}\sum_{d=1}^{L_n}\mathbf{F}_d)'\boldsymbol{\rho} \quad \text{(effect of bias correction for estimated factors)} \\
\bar{r}_5 &= \frac{1}{p}\left[\sum_{i\in S_{L_n}}\Delta_i^n\mathbf{F}'\mathbf{\Gamma}_i' - \frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathbf{w}_n'\mathbf{z}_d\sum_{j\in S_d}\Delta_j^n\mathbf{F}'\mathbf{\Gamma}_j'\right]\mathcal{G}. \qquad (6.3)
\end{aligned}$$

All these terms contribute to the limiting distribution. Only the first two terms are similar to those of the discrete-time diffusion index model (Bai and Ng (2006)). We additionally have three new leading terms. Among them, we would like to pay a special attention to $\bar{r}_5$, which is due to the effect of idiosyncratic betas. Interestingly this term is the difference of two pieces. The first piece $\sum_{i\in S_{L_n}}\Delta_i^n\mathbf{F}'\mathbf{\Gamma}_i'\mathcal{G}$ arises from the effect of $\widehat{\mathbf{F}}_d - \mathbf{H}_n\mathbf{F}_d$ in (6.2), while the second piece arises from estimating $\boldsymbol{\rho}$. It is an important and interesting matter of fact that $\bar{r}_5 = 0$ if $\mathbf{\Gamma}_t$ is time-invariant, but is not so in general. [12] In the presence of high-frequency movements in $\mathbf{\Gamma}_t$, this term is not negligible. So the forecast asymptotic variance depends on $\{\mathbf{\Gamma}_t\}$ through both of its cross-sectional and serial dynamics. Therefore, the forecast procedure based on estimated factors would be misleading when *either* (i) ignore the $\mathbf{\Gamma}_t$ component in the betas, *or* (ii) treat $\mathbf{\Gamma}_t$ as time-invariant. In practice, however, econometricians do not know the degree of variations on either the time domain or the cross-sectional domain of $\mathbf{\Gamma}_t$. The forecast interval we present below is uniformly valid across models with various strengths and degrees of time-varying in $\mathbf{\Gamma}_t$.

[12] When $\mathbf{\Gamma}_t = \mathbf{\Gamma}$ for all $t \in [0, L_nT]$, it can be directly shown that $\bar{r}_5 = 0$ by verifying $(\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathbf{z}_d\mathbf{z}_d')^{-1}\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\mathbf{z}_d\sum_{j\in S_d}\Delta_j^n\mathbf{F}' = (0, \mathbf{I}_K)'$.

6.2 Forecast intervals

We require $\mathbf{G}_t$ and the drift part $\boldsymbol{\alpha}$ be nearly time-invariant over the entire interval $[0, L_n T]$. In particular, $\mathbf{G}_t = (\mathbf{g}_{1t}(\mathbf{X}_{1t}), ..., \mathbf{g}_{pt}(\mathbf{X}_{pt}))'$ is determined by a set of instruments $\{\mathbf{X}_{lt}\}_{l\leq p}$. On the other hand, we still allow $\boldsymbol{\Gamma}_t$ to be time-varying with a realized trajectory driven by the Brownian motion. [13]

Formally, we present the following assumption.

Assumption 6.1. *(i) Suppose* $\mathcal{V}_5^{-1/2}\sqrt{p}\bar{r}_5 \xrightarrow{\mathcal{L}\text{-}s} N(0,1)$ *if* $\bar{r}_5 \neq 0$.
(ii) The eigenvalues of $\mathcal{V}_1$ *and* $\mathcal{V}_2$ *are bounded below by an absolute constant* $c > 0$
(iii) $\{\Delta_t^n U_l\}_{l\leq p}$ *are cross-sectionally independent, given* $\mathcal{F}_{t-1}$.
(iv) $\{\mathbf{X}_{lt}, \mathbf{G}_t\}$ *are independent of* t.
(v) $\min\{\sqrt{L_n}, \sqrt{p}\}\Delta_n = o(1)$.
(vi) The sequence of IV_d *is second-order weakly stationary.*

Condition (i) is a CLT applied to the cross-sectional units of $\boldsymbol{\Gamma}_t$. Condition (ii) requires $\|\boldsymbol{\rho}\| > 0$ so that the factors should contain forecast information. When $\boldsymbol{\rho} = 0$ the problem reduces to the regular autoregressive forecast with estimated lagged integrated volatilities. Condition (iii) ensures a cross-sectional CLT for $\bar{r}_2$, as well as a simple diagonal error covariance estimator in $\widehat{\mathcal{V}}_2$. Sparse covariance estimator can be used in the presence of cross-sectional dependence. Condition (iv) requires that the instruments and the corresponding betas should be time-invariant on the entire range $[0, L_n T]$. But note that we still allow time-varying betas, thanks to the time-varying $\{\boldsymbol{\Gamma}_t\}$. This is still a plausible condition since the instruments mainly capture the long-run changes in beta. In the appendix, we provide a slightly more general condition that allows slightly time-varying $\{\mathbf{X}_{lt}, \mathbf{G}_t\}$, and our main theorem is proved under the more general condition.

The above expansion leads to the asymptotic distribution of $\mathrm{IV}_{L_n+h|L_n} - \widehat{\mathrm{IV}}_{L_n+h|L_n}$, whose asymptotic variance arise from five leading terms, and are given in the appendix. The following are the covariance estimators for each piece of $\bar{r}_1$ - $\bar{r}_5$, and are robust to heteroskedasticity [14]: let $\widehat{\Delta_i^n \mathbf{F}} = \widetilde{\Delta_i^n \mathbf{F}} - \frac{1}{L_n}\sum_{d=1}^{L_n}\widetilde{\mathbf{F}}_d$, $\widehat{\mathbf{G}} = \frac{1}{\Delta_n M_n}\sum_{d=1}^{L_n}\sum_{j\in S_d}\mathbf{P}_j\Delta\mathbf{Y}_j\widehat{\Delta_i^n \mathbf{F}}'$, $\widehat{\mathcal{G}} = \widehat{\mathbf{G}}(\frac{1}{p}\widehat{\mathbf{G}}'\widehat{\mathbf{G}})^{-1}\widehat{\boldsymbol{\rho}}$,

[13] Requiring $\mathbf{G}_t$ be time-invariant is due to the fact that $\widehat{\mathbf{F}}_d$ is estimating a rotated integrated factors, whose rotation matrix depends on $\mathbf{G}_t$. To remove the rotation discrepancy of in the estimated $\boldsymbol{\rho}'\mathbf{F}_d$, it is essential to require the rotation matrix be time-invariant in the long-run interval. This gives rise to our restriction to the time-invariant instrumental beta.

[14] Unlike estimating the asymptotic variance for $\mathbf{g}_{il}$, the plug-in method here produces a uniformly valid forecast interval due to the fact that the estimation error for $\mathcal{V}_5$ is dominated by $\mathcal{V}_2$ uniformly over various strengths of $\boldsymbol{\Gamma}_t$.

and $\widehat{\mathcal{G}}_l$ denotes the l th element of $\widehat{\mathcal{G}}$, $\widehat{\mathbf{\Gamma}}_t = \frac{1}{k_n\Delta_n}\sum_{j\in I_t^n}(\mathbf{I}-\mathbf{P}_j)\Delta_j^n\mathbf{Y}\widehat{\Delta_i^n\mathbf{F}}'$. Let

$$
\begin{aligned}
\widehat{\mathcal{V}}_1 &= \frac{1}{L_n-h}\sum_{d=1}^{L_n-h}(\widehat{\mathbf{w}}_n'\widehat{\mathbf{z}}_d)^2\widehat{v}_{d+h}^2, \quad \widehat{v}_{d+h} = \widehat{y}_{L_n+h|L_n} - \widehat{\mu}\widehat{y}_{L_n} + \widehat{\boldsymbol{\rho}}'\widehat{\mathbf{F}}_{L_n} \\
\widehat{\mathcal{V}}_2 &= \frac{1}{p\Delta_n}\widehat{\mathcal{G}}'\mathrm{diag}\left(\sum_{i\in S_{L_n}}\widehat{\Delta_i^n\mathbf{U}}\widehat{\Delta_i^n\mathbf{U}}'\right)\widehat{\mathcal{G}}, \quad \widehat{\Delta_i^n\mathbf{U}} = \Delta_i^n\mathbf{Y} - (\widehat{\mathbf{G}}+\widehat{\mathbf{\Gamma}}_i)\widehat{\Delta_i^n\mathbf{F}} \\
\widehat{\mathcal{V}}_3 &= \Delta_n\Big(1-\frac{2}{k_n}\Big)\sum_{i\in S_d}\widehat{\mathbf{c}}_{i,l}^2, \text{where } \mathbf{c}_{i,l} = (\boldsymbol{\sigma}_i^{\mathbf{Y}_l})^2 \\
\widehat{\mathcal{V}}_4 &= \left[\widehat{w}_1\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\widehat{y}_d - 1\right]^2\widehat{\boldsymbol{\rho}}'\frac{1}{L_n}\sum_{m=1}^{L_n}\widehat{\mathbf{F}}_d\widehat{\mathbf{F}}_d'\widehat{\boldsymbol{\rho}} \\
\widehat{\mathcal{V}}_5 &= \frac{1}{p}\sum_{l=1}^{p}\widehat{\mathcal{G}}_l^2\widehat{\Omega}_l^2, \quad \text{where } \widehat{\Omega}_l = \sum_{i\in S_{L_n}}\widehat{\Delta_i^n\mathbf{F}}'\widehat{\boldsymbol{\gamma}}_{il} + \frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\widehat{\mathbf{w}}_n'\widehat{\mathbf{z}}_d\sum_{j\in S_d}\widehat{\Delta_i^n\mathbf{F}}'\widehat{\boldsymbol{\gamma}}_{jl},
\end{aligned}
$$

for $\widehat{\mathbf{z}}_d = (\widehat{y}_d, \widehat{\mathbf{F}}_d')'$, $\widehat{\mathbf{w}}_n = (\frac{1}{L_n-h}\sum_{d=1}^{L_n-h}\widehat{\mathbf{z}}_d\widehat{\mathbf{z}}_d')^{-1}\widehat{\mathbf{z}}_{L_n}$ and $\widehat{y}_d = \widehat{\mathrm{IV}}_d$.

Let $z_{\tau/2}$ be the standard normal's $1-\tau/2$ th quantile.

Theorem 6.1. *Suppose Assumptions 5.1-5.3, 5.5, and 6.1 hold uniformly over all data generating processes $\mathbb{P}\in\mathcal{P}$. Define $\widehat{s}_n := (\frac{1}{L_n}\widehat{\mathcal{V}}_1 + \frac{1}{p}\widehat{\mathcal{V}}_2 + \Delta_n\widehat{\mathcal{V}}_3 + \frac{1}{L_n}\widehat{\mathcal{V}}_4 + \frac{1}{p}\widehat{\mathcal{V}}_5)^{1/2}$. Then*

$$
\sup_{\mathbb{P}\in\mathcal{P}}\left|\mathbb{P}\left(y_{L_n+h|L_n}\in[\widehat{y}_{L_n+h|L_n}\pm z_{\tau/2}\widehat{s}_n]\right)-(1-\tau)\right|\to 0.
$$

7 Extensions

7.1 Testing the relevance of instruments

Our framework of uniform inference is also useful for testing the relevance of included instruments. For this purpose, we consider a linear case,

$$
\boldsymbol{\beta}_{lt} = \mathbf{X}_{lt}'\boldsymbol{\theta}_t + \boldsymbol{\gamma}_{lt}, \quad l = 1, ..., p. \tag{7.1}
$$

Note that $\boldsymbol{\theta}_t$ is a $d\times K$ matrix, whose k th column, denoted by $\boldsymbol{\theta}_{t,k}$, represents the effect of the instruments on the betas of the k th risk factor. Inferencing about the time-varying coefficient $\boldsymbol{\theta}_t$ allows us to explain the dynamic importance of each of the instruments and that whether any instrument is relevant. Note that although (7.1) specifies a linear function $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, $\mathbf{X}_{lt}$ could include nonlinear (sieve) transformations of each individual instruments.

Most importantly, the inferences procedure should be uniformly valid over a broad DGPs that generate $\boldsymbol{\gamma}_{lt}$, as we did earlier.

To describe the estimator of $\boldsymbol{\theta}_t$, we use the linear sieve $\mathbf{\Phi}_t = (\mathbf{X}_{1t}, ..., \mathbf{X}_{pt})'$ and $\mathbf{P}_t = \mathbf{\Phi}_t(\mathbf{\Phi}_t'\mathbf{\Phi}_t)^{-1}\mathbf{\Phi}_t'$. Then $\widehat{\mathbf{G}}_t$ can be defined as before: (4.2) in the known factor case, and (4.3) in the unknown factor case. Then we run a cross-sectional regression to estimate $\boldsymbol{\theta}_t$:

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

As for the asymptotic analysis, the predescribed uniformity issue is still present. For instance, in the known factor case, it can be shown that

$$\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t = \left(\mathbf{\Phi}_t'\mathbf{\Phi}_t\right)^{-1}\left[\mathbf{\Phi}_t'\mathbf{\Gamma}_t + \frac{1}{k_n\Delta_n}\mathbf{\Phi}_t'\sum_{i\in I_t^n}\Delta_i^n\mathbf{U}\Delta_i^n\mathbf{F}_i'\mathbf{s}_{f,t}^{-1}\right] + \text{ negligible terms.}$$

The strength of the cross-sectional variations in $\mathbf{\Phi}_t'\mathbf{\Gamma}_t$ is still unknown and may potentially vary in a large range, leading to a discontinuity in the limiting distribution of $\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t$, and various possible rates of convergence. In addition, the same problem as we described earlier is still present, namely, the estimation error for $\mathbf{\Gamma}_t$ may stochastically dominate the strength of the asymptotic variance from $\mathbf{\Phi}_t'\mathbf{\Gamma}_t$, hence simply plugging-in estimators of the asymptotic variance for $\widehat{\boldsymbol{\theta}}_t$ would still not be uniformly valid. Hence we rely on the cross-sectional bootstrap.

We focus on the independent bootstrap, by independently resampling cross-sectional time series $\{\Delta_i^n Y_m^*, i \in I_t^n\}_{m=1,...,p}$ and $\left\{\mathbf{X}_{mi}^* : i \in I_t^n\right.{}_{m=1,...,p}$. Then let $\mathbf{\Phi}_i^* = (\mathbf{X}_{1,i}^*, ..., \mathbf{X}_{p,i}^*)'$, $\mathbf{P}_i^* = \mathbf{\Phi}_i^*(\mathbf{\Phi}_i^{*'}\mathbf{\Phi}_i^*)^{-1}\mathbf{\Phi}_i^{*'}$ and $\Delta_i^n\mathbf{Y}^* = (\Delta_i^n Y_1^*, ..., \Delta_i^n Y_p^*)'$. Let the bootstrap estimator be $\widehat{\boldsymbol{\theta}}_t^* = \left(\mathbf{\Phi}_t^{*'}\mathbf{\Phi}_t^*\right)^{-1}\mathbf{\Phi}_t^{*'}\widehat{\mathbf{G}}_t^*$, where

$$\widehat{\mathbf{G}}_t^* = \begin{cases} \sum_{i\in I_t^n}\mathbf{P}_{i-1}^*\Delta_i^n\mathbf{Y}^*\Delta_i^n\mathbf{F}'\left(\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\,\Delta_i^n\mathbf{F}'\right)^{-1}, & \text{known factor case} \\ \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}\mathbf{P}_{i-1}^*(\Delta_i^n\mathbf{Y}^*)\widehat{\Delta_i^n\mathbf{F}}', & \text{unknown factor case.} \end{cases}$$

Repeat the bootstrap sampling and estimation for B times, and obtain $\{\widehat{\boldsymbol{\theta}}_t^{*b}\}_{b\le B}$. Let $(\widehat{\boldsymbol{\theta}}_{t,k}^{*b}, \widehat{\boldsymbol{\theta}}_{t,k})$ respectively denote the k th column of $\widehat{\boldsymbol{\theta}}_t^{*b}$ and $\widehat{\boldsymbol{\theta}}_t$. For any unit vector $\mathbf{v} \in \mathbb{R}^d$, let q_τ be the $1-\tau$ th bootstrap quantile of $\{|\mathbf{v}'\widehat{\boldsymbol{\theta}}_{t,k}^{*b} - \mathbf{v}'\widehat{\boldsymbol{\theta}}_{t,k}|\}_{b\le B}$. In the case of known factors, the confidence interval for $\mathbf{v}'\boldsymbol{\theta}_{t,k}$ is given by

$$CI^{\theta}_{t,k,\tau} \;=\; [\mathbf{v}'\widehat{\boldsymbol{\theta}}_{t,k} - q_\tau, \mathbf{v}'\widehat{\boldsymbol{\theta}}_{t,k} + q_\tau].$$

In the unknown factor case, due to the effect of estimating the unknown factors, $\widehat{\boldsymbol{\theta}}_t$ needs to be debiased, but all the technical arguments would be very similar to those of treating the estimated $\widehat{\mathbf{G}}_t^{\text{latent}}$, we omit the formal treatment of the unknown factor case for brevity.

Theorem 7.1. *Suppose the factors are known. Let $\mathcal{P}$ be the collection of all data generating processes $\mathbb{P}$ for which Assumptions 5.6, 5.8 and assumptions of Theorem 5.1 hold. Then for any unit vector $\mathbf{v} \in \mathbb{R}^d$, for each fixed $t \in [0, T]$, and $k \leq K$,*

$$\sup_{\mathbb{P}\in\mathcal{P}} \left|\mathbb{P}(\mathbf{v}'\boldsymbol{\theta}_{t,k} \in CI^{\theta}_{t,k,\tau}) - (1-\tau)\right| \to 0.$$

7.2 Beta prediction for missing prices

Missing data on the stock prices are sometimes present for various reasons, including misrecording and infrequent trading. Suppose for a new out-of-sample individual l^*, whose return data $\{\Delta_i^n Y_{l^*} : i \in I_t^n\}$ is missing on the entire interval I_t^n, but the data for its instruments $\{\mathbf{X}_{l^*,i} : i \in I_t^n\}$ is not missing. In addition, the time point t is "in-sample", meaning that we have observations $\{\Delta_i^n \mathbf{Y}, \Delta_i^n \mathbf{F}, \{\mathbf{X}_{li} : l \leq p\} : i \in I_t^n\}$. For simplicity, we assume that the factors are observable. The goal is to "predict" (or estimate) the instrumental beta $\mathbf{g}_{l^*,t}$, which represents the "long-term" pattern of its beta during this interval. This type of missing data problem is often seen in high-frequency trading, where we have "complete data" for stocks $l \leq p$ on the interval I_t^n. But for a particular stock of interest, we do not observe its return data during this period, either due to the fact that .... On the other hand, as the instruments are updated at a possibly much lower frequency, we do have the observations of its instruments.

The key assumption that is needed is the following "out-of-sample smoothing". Let $\mathbf{X}_{lt}$ and $\mathbf{g}_{lt}$ respectively denote the instrumental beta for individual l, where $l \in \{1, ..., p\} \cup \{l^*\}$.

Assumption 7.1 (out-of-sample cross-sectional smoothing)**.** *There exists a time-varying function $\mathbf{g}_t : \chi \to \mathbb{R}^K$, satisfying: (i) for all $l \in \{1, ..., p\} \cup \{l^*\}$,*

$$\mathbf{g}_{lt} = \mathbf{g}_t(\mathbf{X}_{lt}).$$

(ii) For the sieve basis functions $\boldsymbol{\phi} = (\phi_1(\cdot), ..., \phi_J(\cdot))'$, there is a $J \times K$ sieve coefficient $\boldsymbol{\theta}_t$,

and $C, \eta > 0$ so that

$$\sup_{\mathbf{x}\in\chi} \|\mathbf{g}_t(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x})'\boldsymbol{\theta}_t\| < CJ^{-\eta}$$

where χ denotes the domain of the instruments $\mathbf{X}_{lt}$.

This assumption is to be compared with Assumption 5.3 (iv), which we refer to as "in-sample cross-sectional smoothing":

$$\max_{l\leq p} \|\mathbf{g}_{lt} - \boldsymbol{\phi}(\mathbf{X}_{lt})'(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'\mathbf{G}_t\| \leq CJ^{-\eta}. \tag{7.2}$$

From the nonparametric regression point of view, Assumption 7.1 simply assumes that the instrumental beta is a nonparametric smooth function of the cross-sectional instruments. On the other hand, (7.2) is a weaker version, which only requires the realized in-sample cross-sections are "smoothed". As we require out-of-sample prediction under missing data, (7.2) has to be strengthened. But the intuition is still clear: unlike the idiosyncratic betas, firms' long-term betas should be close so long as their instruments are close.

Given the in-sample data, we estimate the in-sample instrumental betas $\widehat{\mathbf{G}}_t$ as before, and $\widehat{\boldsymbol{\theta}}_t = (\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'\widehat{\mathbf{G}}_t$. We then "predict" $\mathbf{g}_{l^*,t}$ by

$$\widehat{\mathbf{g}}_{l^*,t} = \boldsymbol{\phi}(\mathbf{X}_{l^*,t})'\widehat{\boldsymbol{\theta}}_t.$$

To construct a confidence interval, we continue using the cross-sectional bootstrap method as described before. Let $\{\widehat{\mathbf{g}}^{*b}_{l^*,t} : b \leq B\}$ be the bootstrap estimators for $\mathbf{g}_{l^*,t}$. For any unit vector $\mathbf{v} \in \mathbb{R}^d$, let q_τ be the $1-\tau$ th bootstrap quantile of $\{|\mathbf{v}'\widehat{\mathbf{g}}^{*b}_{l^*,t} - \mathbf{v}'\widehat{\mathbf{g}}_{l^*,t}|\}_{b\leq B}$. Then the prediction confidence interval for $\mathbf{v}'\mathbf{g}_{l^*,t}$ is

$$CI^*_{t,\tau} = [\mathbf{v}'\widehat{\mathbf{g}}_{l^*,t} - q_\tau, \mathbf{v}'\widehat{\mathbf{g}}_{l^*,t} + q_\tau].$$

Similar to the conclusion in Theorem 7.1, the above confidence interval has a $1-\tau$ converge rate uniformly in $\mathcal{P}$.

7.3 Varying-coefficient models

While we consider high-frequency setting in this paper, the proposed uniform inference framework is directly applicable to a large class of discrete time panel data models. For instance, consider the following random-coefficient model:

$$y_{lt} = [\mathbf{g}(\mathbf{X}_l) + \boldsymbol{\gamma}_{lt}]'\mathbf{z}_{lt} + u_l, \quad t \leq T, l \leq p.$$

In a wage equation, y_{lt} denotes the logarithm of wage, and $\mathbf{z}_{lt}$ represents the years of schooling. The rate of return to education may depend on the individual characteristics $\mathbf{X}_l$ that is assumed to be time-invariant. We are interested in the characteristic-specific contributions to the rate of return to education. Importantly, the signal strength of $\boldsymbol{\gamma}_{lt}$, measured by its cross-sectional variance, is unknown, and can be any bounded (from infinity) non-negative asymptotic sequence. While there is a growing literature on time-varying coefficient panel data models (e.g., Su and Ullah (2011)), they all specifically assume $\boldsymbol{\gamma}_{lt} = 0$, and may suffer misspecification problems and result in invalid inference about $\mathbf{g}(\cdot)$. In the time-series context, Andrews (1999) studied a slightly variant random-coefficient model, in which the time-series variance of $\boldsymbol{\gamma}_{lt}$ is modeled explicitly as an unknown parameter that equals zero under the null hypothesis, and showed that the asymptotic distribution of estimated parameters (e.g., $\mathbf{g}_t$) has a discontinuity. Our approach admits a new inference procedure of the random coefficient model. An essential difference from the treatment of Andrews (1999), besides being used for panel data, is that we do not explicitly model the covariance of $\boldsymbol{\gamma}_{lt}$ as an unknown parameter, but allow it to be either "on the boundary", "arbitrarily close to the boundary", or "bounded away from the boundary". In this framework, our procedure also provides a new feature for varying-coefficient models, where the coefficient can be decomposed into the addition of a "characteristic component" and a "random component", and the inference of the former is affected by the strength of the latter. The achieved inference about $\mathbf{g}_t$ is valid uniformly over a large class of $\boldsymbol{\gamma}_{lt}$ with unknown strengths.

We assume $\mathbf{g}(\mathbf{X}_l) \approx \mathbf{B}\Phi(\mathbf{X}_l)$, where $\Phi(\cdot)$ denotes a vector of sieve basis functions, with coefficients $\mathbf{B}$. We estimate $\mathbf{g}(\mathbf{X}_l)$ by $\widehat{\mathbf{g}(\mathbf{X}_{ls})} := \widehat{\mathbf{B}}\Phi(\mathbf{X}_l)$, where

$$\begin{aligned} \widehat{\mathbf{B}} &= \frac{1}{p}\sum_{j=1}^{p}\mathbf{A}_j^{-1}\frac{1}{T}\sum_{r=1}^{T}\mathbf{z}_{jr}y_{jr}\Phi(\mathbf{X}_j)'\mathbf{D}^{-1}, \\ \mathbf{D} &= \frac{1}{p}\sum_{j=1}^{p}\Phi(\mathbf{X}_j)\Phi(\mathbf{X}_j)', \quad \mathbf{A}_j = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{jt}\mathbf{z}_{jt}'. \end{aligned}$$

Under the condition $\mathbb{E}(\boldsymbol{\gamma}_{lt}|\mathbf{X}_l) = 0$ for all t, and standard regularity conditions on the smoothness of $\mathbf{g}(\mathbf{X}_l)$, it can be shown that

$$\widehat{\mathbf{g}(\mathbf{X}_l)} - \mathbf{g}(\mathbf{X}_l) = \frac{1}{p}\sum_{j=1}^{p}\mathbf{A}_j^{-1}\frac{1}{T}\sum_{r=1}^{T}\mathbf{z}_{jr}u_{jr}\Phi(\mathbf{X}_j)'\mathbf{D}^{-1}\Phi(\mathbf{X}_l) + \frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\gamma}_{js}\Phi(\mathbf{X}_j)'\mathbf{D}^{-1}\Phi(\mathbf{X}_l) + \text{negligible terms.}$$

The first term has an order $O_P((Tp)^{-1/2})$, while the second term explicitly depends on the unknown cross-sectional variations in $\frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\gamma}_{js}\Phi(\mathbf{X}_j)'$, leading to a discontinuity in the limiting distribution. Hence this is the same uniformity issue as we described earlier, and simply plugging-in estimators of the asymptotic variance for $\widehat{\mathbf{g}(\mathbf{X}_l)}$ would produce a very conservative inference. On the other hand, our cross-sectional bootstrap based inference would lead to a valid inference for $\mathbf{g}(\mathbf{X}_l)$ uniformly over various possible strength of $\boldsymbol{\gamma}$. We focused our discussions on the case $\mathbf{X}_l$ and $\mathbf{g}(\cdot)$ are time-invariant as this is a discrete-time panel-data model. Our approach can be modified to allow slowly time-varying $(\mathbf{X}_{lt}, \mathbf{g}_t(\cdot))$ by applying time-domain smoothing.

8 Empirical Studies

8.1 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 each stock's abnormal prices that are out of the middle seventy percent range, and replace them with the previous five-minute price, which makes the abnormal return to zero. Moreover, we drop the overnight returns for excluding stock splits and dividend issuances. 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 instruments. We consider four instruments 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.

8.2 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 instruments, and estimate the spot instrumental and idiosyncratic beta for each company on each trading day. [15] 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 instrumental 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-sectional averaged instrumental 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 instrumental 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.

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 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

[15] We also tried B-splines with degree 3 (Eilers and Marx, 1996), and obtain similar results.

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 instruments 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.

8.3 Confidence Intervals

We construct 95% construct confidence intervals for each of the firms' instrumental 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 instrumental betas loading on the market factor. In fact, most of the instrumental betas for the market factor are larger than one. (2) There is a substantial difference in the instrumental 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 $\boldsymbol{\beta}$. We find that for SMB, the proportion of positive $\boldsymbol{\beta}$ is 37% for large firms, and 71% for small firms, while the proportion of negative $\boldsymbol{\beta}$ is 62% for large firms, and 28% for small firms. In contrast, these proportions respectively become 51% and 48% for positive $\mathbf{\Gamma}$, and 52% for negative $\mathbf{\Gamma}$, so the difference among firms of large and small sizes in $\mathbf{\Gamma}$ is much less noticeable. This suggests that the instrumental 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).

When we group firms by the volatility, however, the pattern demonstrates noticeable

Table 1: Cross-sectional Proportion of significant **G** of groups by size, 2006

	positive significance				negative significance			
size	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

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.

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

	positive significance				negative significance			
volatility	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

We now examine the dependence of the instrumental beta on the characteristics more explicitly. We consider the linear specification $\mathbf{g}_{it} = \mathbf{x}_{it}'\boldsymbol{\theta}_t$, and test the relevance of each of the four instruments $\mathbf{x}_{it}$ = (size, value, momentum and volatility). 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 G 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 instruments

	positive significance				negative significance			
Instruments	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

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 instrumental 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 instrumental 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 instrumental 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.

Finally, in the appendix, we report the estimated cross-sectional **G** functions for Fama-French factors.

9 Conclusion

This paper studies a conditional factor model with a large number of assets 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 instruments 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.

The limiting distribution of the estimated instrument 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 effect on the betas of firms' instruments, and do not need to pretest to know whether or not the idiosyncratic beta exists, or their strengths. In addition, we employ the estimated factors to conduct out-of-sample forecast of integrated volatility. Taking into account the time-varying idiosyncratic beta components is also necessary for the out-of-sample forecast interval to be uniformly valid.

References

Ait-Sahalia, Y., Kalnina, I. and Xiu, D. (2014). The idiosyncratic volatility puzzle: A re-assessment 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.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting* **1** 777–878.

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. 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. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171.

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

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.

BICKEL, P. and LEVINA, E. (2008). Covariance regularization by thresholding. *Annals of Statistics* **36** 2577–2604.

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.

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

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 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.

Cosemans, M., Frehen, R., Schotman, P. and Bauer, R. (2009). Effcient estimation of firm-specific betas and its benefits for asset pricing tests and portfolio choice .

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.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews* **5** 1–50.

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

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. (2016). Projected principal component analysis in factor models. *Annals of statistics* **44** 219.

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.

Hansen, P. and Lunde, A. (2011). Forecasting volatility using high frequency data. *The Oxford Handbook of Economic Forecasting, Oxford: Blackwell* 525–556.

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.

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

JURADO, K., LUDVIGSON, S. C. and NG, S. (2015). Measuring uncertainty. *American Economic Review* **105** 1177–1216.

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

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. (2017). Jump regressions. *Econometrica* **85** 173–195.

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

MIKUSHEVA, A. (2007). Uniform inference in autoregressive models. *Econometrica* **75** 1411–1452.

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

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.

SU, L. and ULLAH, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of empirical economics and finance* 455–497.

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.
URL `http://ideas.repec.org/a/bes/jnlasa/v100y2005p1394-1411.html`

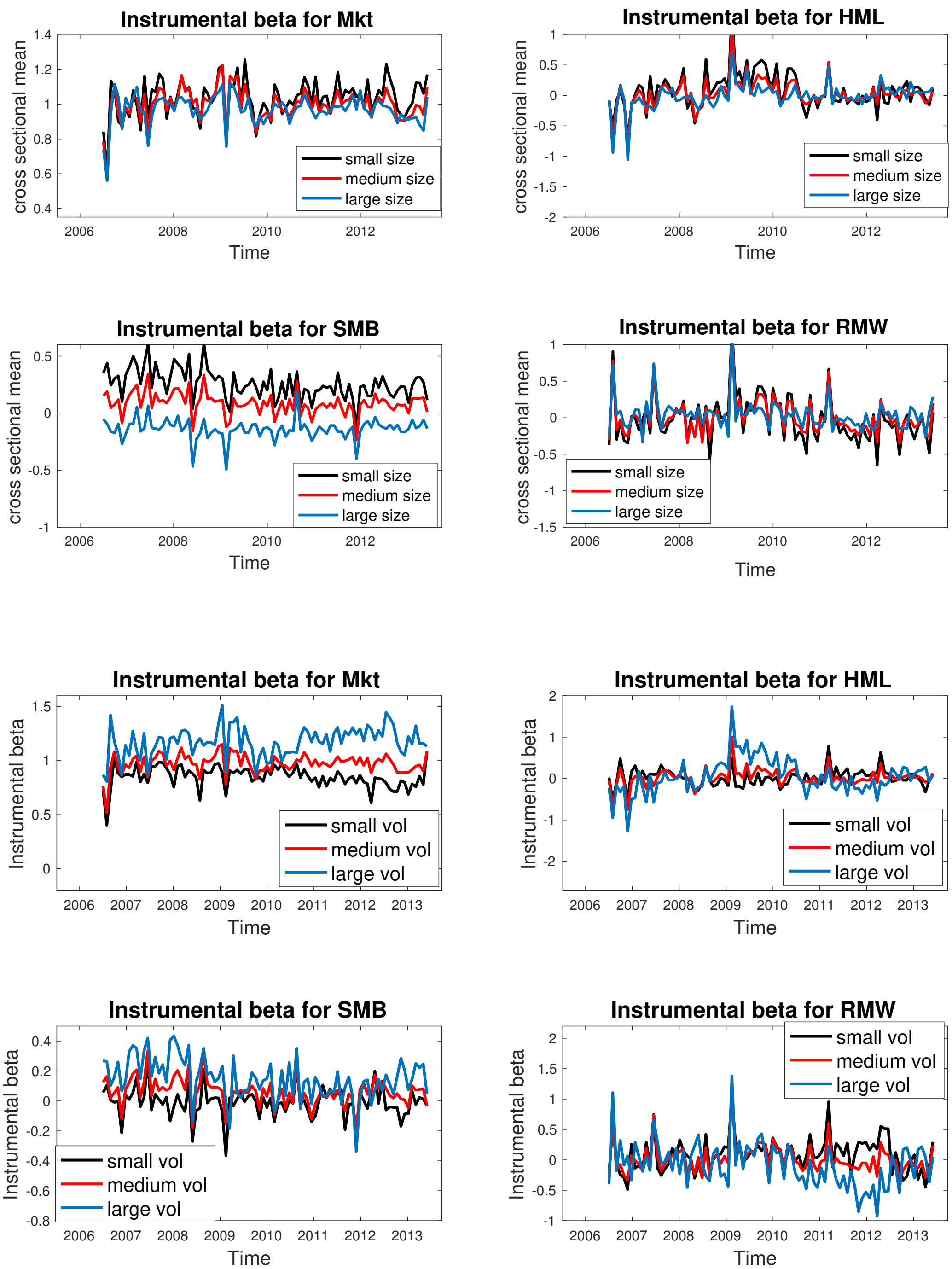

Figure 1: Cross-sectional means of instrumental beta, grouped by size (upper four) and by volatility (lower four). The instrumental betas are estimated on a daily basis, and this figure plots eight days' estimations for each month.

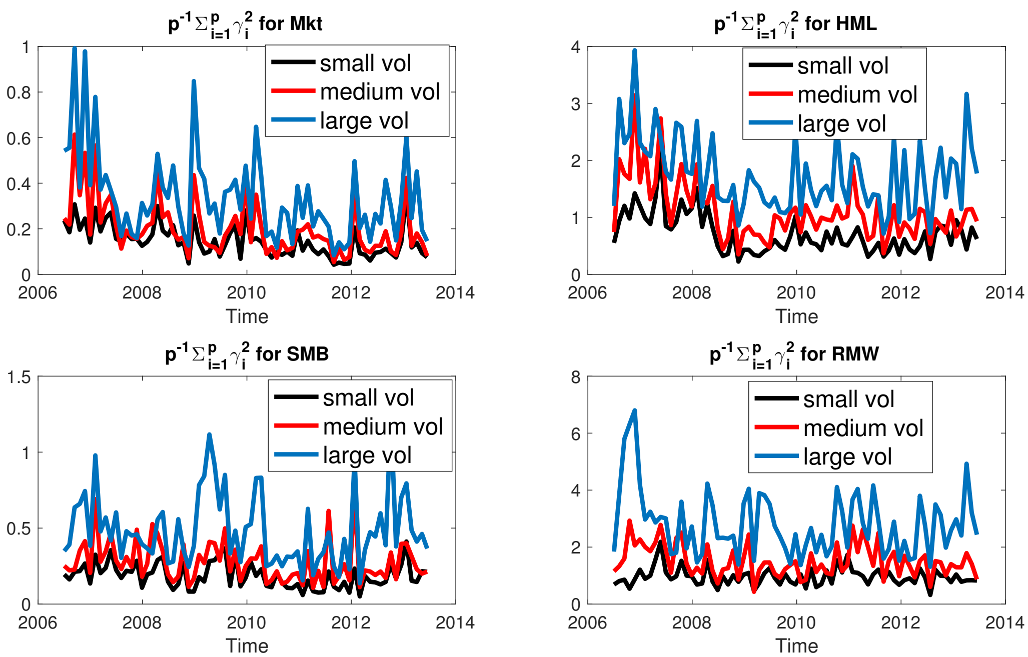

Figure 2: Cross-sectional variations of $\mathbf{\Gamma}$ across times. Firms are grouped into small, medium and large sizes.

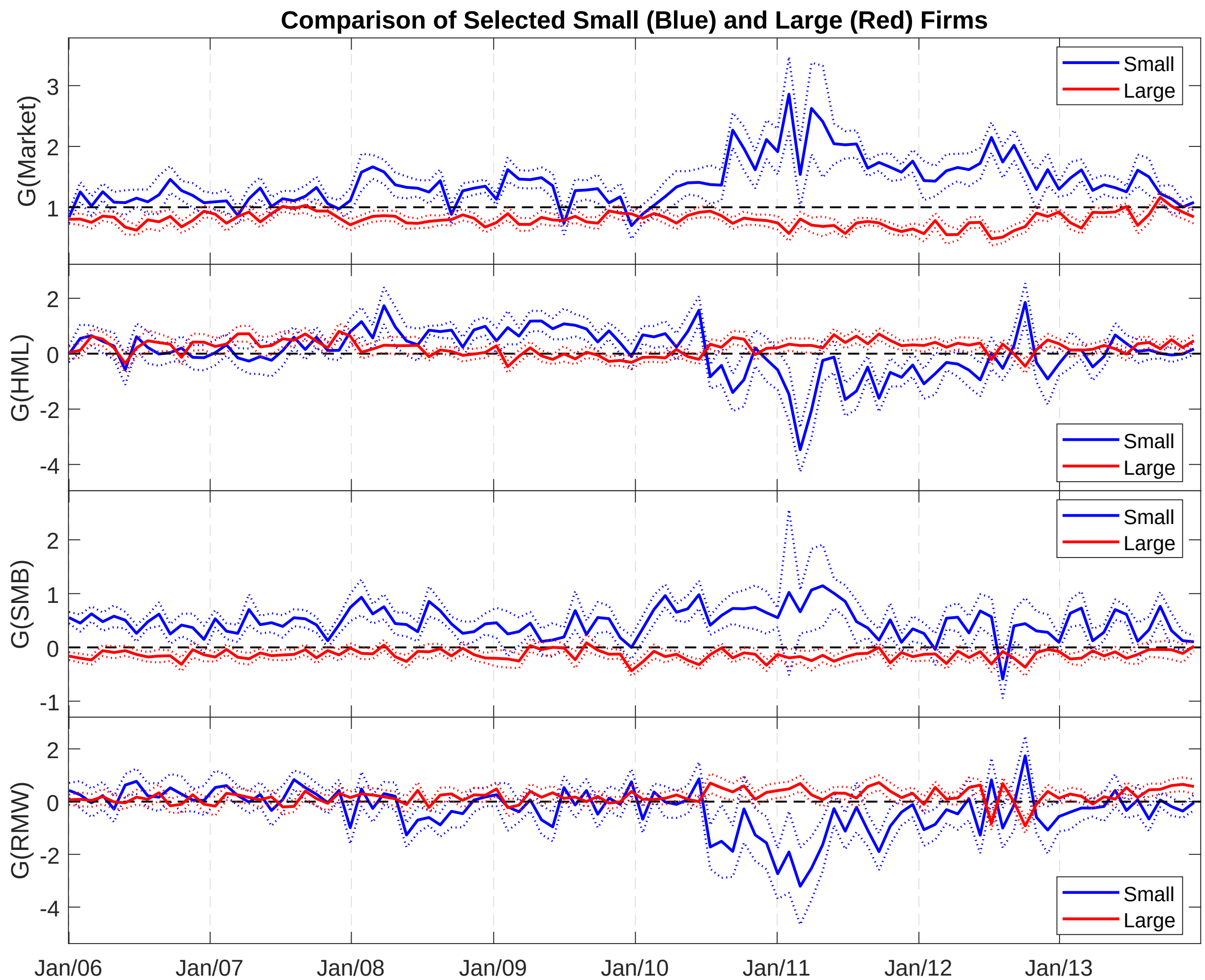

Figure 3: Two individual stocks' confidence intervals: the two firms with the largest and smallest sizes in the dataset.