Liao, Yuan; Yang, Xiye

Working Paper

Uniform inference for conditional factor models with instrumental and idiosyncratic betas

Working Paper, No. 2017-11

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Liao, Yuan; Yang, Xiye (2017) : Uniform inference for conditional factor models with instrumental and idiosyncratic betas, Working Paper, No. 2017-11, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/200268

Uniform Inference for Conditional Factor Models with Instrumental and Idiosyncratic Betas

Yuan Liao Xiye Yang*

November 12, 2017

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 a key object of interest is the effect of instruments on the factor betas. 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 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 show that a cross-sectional bootstrap procedure is essential for the uniform inference, 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

1 Introduction

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

Estimating the instruments' effects on the individual betas is one of the central econometric tasks in financial economics, because they pick up long-run patterns and fluctuations in the betas. However, there are also unmeasurable high-frequency components in betas that are more volatile. As we show in this paper, without taking into account the high-frequency movements of betas, the inference procedures of instruments' effects are not asymptotically valid. Unfotunately, 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. But as the individual factor betas demonstrate a much larger heterogeneity when the number of assets is large, it is unrealistic to require the high-dimensional factor beta matrix be fully explained by just a few instruments. This is particularly true for high frequency factor models, where assets' returns are available at a very high frequency; in the contrary, the time dynamics of observed instruments, such as the firm sizes and book-market values, often vary more smoothly and are measured at a much lower frequency, which leave large portions of stock betas' dynamics unexplained. Hence allowing for unmeasurable high-frequency beta components to be unspecified seems a natural and necessary setup. Yet, to the best of our knowledge, the literature pays little attention on this issue. 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.

The goal of this paper is to provide a *uniformly valid* inference of 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 beta's time dynamics and cross-sectional variations. These

dynamics and variations can arise from both the observed instruments and the remaining beta components that are unspecified. In the contrary, we show that the usual inference procedures only produce confidence intervals that are pointwise valid for specific DGPs, therefore potentially produce misleading inferences. In fact, benchmark methods in the financial econometric literature, which ignore the high-frequency beta dynamics, would produce *under-coveraging* confidence intervals of the instrumental effects as a consequence. On the other hand, we show that even if the unexplained beta dynamics are modeled, standard "plug-in" procedures using the estimated asymptotic variances are not uniformly valid either, because they require very strong signal strengths of the unexplained beta dynamics, otherwise would lead to *over-coveraging* confidence intervals.

The study of the effects of instruments on betas (or called "factor loadings" in the econometric literature) 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.

Modeling the beta dynamics as a fully specified function of a set of predetermined instruments goes back at least to Shanken (1990). Most of the works in the literature specify

the factor loading matrix as a parametric function (linear functions as often the case) of the instruments (e.g., Cochrane (1996); Ferson and Harvey (1999); Avramov and Chordia (2006); Gagliardini et al. (2016)), with some exceptions such as Connor and Linton (2007) and Connor et al. (2012), who modeled the loading matrix as a non-parametric function that is fully specified by the instruments associated with the risk factors. [1]

1.1 Beta Decompositions

We propose a conditional factor model in which the time-varying factor betas consist of two components: (i) a nonparametric function of the observed instruments, $\mathbf{g}_{lt}$, which we call "instrumental betas", and (ii) an unmeasurable time-varying and firm specific component, $\boldsymbol{\gamma}_{lt}$, which we call "idiosyncratic betas". Specifically, let $\boldsymbol{\beta}_{lt}$ denote the K-dimensional factor betas of the l th stock at time t. We model:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_{lt}(\mathbf{x}_{l,t-1}, \mathbf{x}_l, \mathbf{x}_{t-1}) + \boldsymbol{\gamma}_{lt}. \tag{1.1}$$

We allow the instruments $(\mathbf{x}_{l,t-1}, \mathbf{x}_l, \mathbf{x}_{t-1})$ to consist of 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. Except for being nonparametric and can be well approximated by sieve representations, we do not make any assumption on either the cross-sectional or serial structure of $\mathbf{g}_{lt}(\cdot)$. In addition, $\boldsymbol{\gamma}_{it}$ is a mean-zero random vector, representing the remaining time-varying individual factor sensitivities after conditioning on the observed instruments.

We study a continuous-time factor model, in which the factors, loadings, and idiosyncratic errors are all driven by Brownian motions through a continuous-time stochastic process. The instrumental beta $\mathbf{g}_{lt}(\mathbf{x}_{l,t-1}, \mathbf{x}_l, \mathbf{x}_{t-1})$ as well as the remaining effects are both estimated using the high-frequency data. In high frequency trading, modeling the beta in the absence of $\boldsymbol{\gamma}_{lt}$ is particularly restrictive. The restriction arises from both sides: (a) high frequency trading is often subjected to limited information, as full information cannot be measured in a frequency as high as that of the return data, leaving many effects latent. (b) Instruments may not be updated as frequently as trading occurs. So using $(\mathbf{x}_{l,t-1}, \mathbf{x}_l, \mathbf{x}_{t-1})$ itself to capture the time-varying sensitivity to the risk factors is not sufficient. The instrumental

[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.

beta $\mathbf{g}_{lt}(\mathbf{x}_{l,t-1}, \mathbf{x}_l, \mathbf{x}_{t-1})$ possesses less volatile than $\boldsymbol{\gamma}_{lt}$, which means it picks up long-run beta patterns and fluctuations, while $\boldsymbol{\gamma}_{lt}$ captures high frequency movements in beta.

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. In this case, *pointwise* inference under a fixed data generating process (DGP) is misleading. The issue here is similar to the problem of estimating parameters on a boundary. As is shown in the literature, (Andrews, 1999; Mikusheva, 2007), when a test statistic has a discontinuity in its limiting distribution, as occurs in estimating parameters on a boundary and in moment inequality models, pointwise asymptotics can be very misleading. The main difference, however, is that when estimating $\mathbf{g}_{lt}$ in the current context, the "boundary parameter" $\boldsymbol{\gamma}_{lt}$ appears to be the nuisance parameter. We provide a uniformly valid inference procedure, that is robust to various strengths of $\boldsymbol{\gamma}_{lt}$. Indeed, depending on the measurement and specification of the instruments, the strengths of $\boldsymbol{\gamma}_{lt}$ may vary both cross-sectionally and serially, and is often unknown to econometricians. We do not need to pretest or pre-know the strengths of $\boldsymbol{\gamma}_{lt}$, and provide confidence intervals that are valid uniformly over a large class of DGP. The class of DGP allows the cross-sectional variance of $\boldsymbol{\gamma}$ to vary from "weak signals" (zero or arbitrarily close to zero), all the way to variances that are bounded away from zero. In addition, the only time-varying condition on $\boldsymbol{\gamma}_{lt}$ is that it has a continuous time-path that is driven by a realized Brownian motion, and hence can be nowhere smooth.

Due to the discontinuity of the limiting distribution of the estimated $\mathbf{g}_{lt}$, we reply on a cross-sectional bootstrap procedure 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. The cross-sectional bootstrap is important 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 it leads to a correct asymptotic coverage probability and is uniformly valid over a large class of DGPs. In contrast, the usual"plug-in" method that uses the estimated asymptotic variance may be valid only pointwise, and fails to provide a uniformly valid confidence interval. This is because the estimation error of the cross-sectional variance of $\boldsymbol{\gamma}_{lt}$ can be larger than the estimand itself when it is near the boundary.

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, 2002b). 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 $\boldsymbol{\gamma}_{lt}$.

A similar decomposition to (1.1) was given by Kelly et al. (2017), where beta is decomposed into a linear function of lagged instruments as well as a unobservable loading component. They specifically require $\boldsymbol{\gamma}_t$ be strong, and obtained limiting distributions for the "instrumental betas", which are therefore, not uniformly valid. 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. There are several key differences between their works and ours. First of all, 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 between Fan et al. (2016) and ours 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 Aït-Sahalia and Xiu (2017) and references therein.

The rest of this paper is organized as follows. Section 2 describes the continuous-time conditional factor model driven by stochastic processes. Section 3 defines the estimators of the components of betas, and the unknown factors in the case of latent factors. Section 4 informally discusses the issue of uniformity and explains why it is a challenging problem for conditional models. 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 the factor risk premium. Section 8 presents a simple simulated example. In

Section 9, we present real data applications on the high-frequency stock return data of firms from S&P500. Finally, all proofs are given in the appendix.

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 The Continuous-Time Conditional Factor Model with Instruments

2.1 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 $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. In general, we allow jumps on all log-price process. But 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. In light of these, we present our model assuming both the log-price processes and factors are continuous, in order to keep notation simple. The jump-robust estimators are given in Section 3.3, where we employ a standard procedure to truncate jumps out.

In this paper, we assume the following (continuous) factor structure:

$$\mathbf{Y}_t = \mathbf{Y}_0 + \int_0^t \boldsymbol{\alpha}_s ds + \int_0^t \boldsymbol{\beta}_s d\mathbf{F}_s + \mathbf{U}_t \tag{2.1}$$

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 $\boldsymbol{\beta} = \{\boldsymbol{\beta}_s\}_{s\geq 0}$ 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

$$\begin{aligned}\mathbf{F}_t &= \int_0^t \boldsymbol{\alpha}_s^F ds + \int_0^t \boldsymbol{\sigma}_s^F d\mathbf{W}_s^F, \\ \mathbf{U}_t &= \int_0^t \boldsymbol{\sigma}_s^U d\mathbf{W}_s^U,\end{aligned} \tag{2.2}$$

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 factor $\mathbf{F}$.

We are interested in the K dimensional continuous factor process $\mathbf{F} = \{\mathbf{F}_t\}_{t\geq 0}$ and the corresponding $p \times K$ matrix process of factor loadings $\boldsymbol{\beta} = \{\boldsymbol{\beta}_t\}_{t\geq 0}$, which is adapted to the filtration $\mathcal{F}$. 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) and Li et al. (2017a)), to differentiate from the discontinuous (or jump) beta (Li et al. (2017b)).

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):

$$\boldsymbol{\beta}_{lt} = \mathbf{g}_{lt}(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l = 1, \cdots, p. \tag{2.3}$$

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. Except for being nonparametric and can be well approximated by sieve representations, we do not make any assumption on either the cross-sectional or serial structure on $\{\mathbf{g}_{lt}\}_{lt}$. On the other hand, $\boldsymbol{\gamma}_{lt}$ is a mean-zero random vector, representing the remaining time-varying individual factor risks after conditioning on the observed instruments. 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. It captures high frequency movements in beta and can be more volatile than $\mathbf{X}_{lt}$.

Let $\mathbf{G}_t$ be the $p \times K$ matrix of $\{\mathbf{g}_{lt}(\mathbf{X}_{lt})\}_{l=1}^p$ and $\mathbf{\Gamma}_t$ be the $p \times K$ matrix of $\{\boldsymbol{\gamma}_{lt}\}_{l=1}^p$. Then we have the following representation for the continuous component of $\mathbf{Y}$:

$$d\mathbf{Y}_t = \boldsymbol{\alpha}_t dt + (\mathbf{G}_t + \mathbf{\Gamma}_t) d\mathbf{F}_t + d\mathbf{U}_t, \quad \forall t \in [0, T], \tag{2.4}$$

or equivalently, in the integration form,

$$\mathbf{Y}_t = \mathbf{Y}_0 + \int_0^t \boldsymbol{\alpha}_s ds + \int_0^t (\mathbf{G}_s + \mathbf{\Gamma}_s) d\mathbf{F}_s + \mathbf{U}_t, \quad \forall t \in [0, T]. \tag{2.5}$$

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

$$y_{d+h} = \mu y_d + \boldsymbol{\rho}' \mathbf{F}_d + v_{d+h}, \quad d = 1, ..., L_n, \quad L_n \to \infty, \tag{2.6}$$

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 integrated volatility:

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

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 (2002b), we extract the latent factors from a large set of financial asset returns. But we are particularly interested in the effect on the prediction intervals of estimating the long-run factors from model (2.5) with time-varying $\mathbf{\Gamma}_t$.

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

One of the key conditions is $\mathbb{E}(\mathbf{\Gamma}_t|\mathbf{X}_t) = 0$. It implies that the instrumental and idiosyncratic beta components are orthogonal. Conditions of this type are often seen in the literature, such as the orthogonal decomposition of risks into systematic risk and idiosyncratic risk in financial economics, and the decomposition of payoffs into the projection on the discount factors and the idiosyncratic part in asset pricing theories (Cochrane, 2005). 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:

$$\boldsymbol{\beta}_{lt} = \mathbf{g}(\mathbf{X}_{lt}) + \boldsymbol{\gamma}_{lt}, \quad l \leq p \tag{2.7}$$

Consider the "ideal case" that $\boldsymbol{\beta}_{lt}$ is completely known. Then in (2.7), $\mathbf{X}_{lt}$ 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 $\mathbb{E}(\mathbf{\Gamma}_t|\mathbf{X}_t) = 0$ 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.

3 Estimation

Since our focus is on the continuous factors, not those jump ones, we shall first assume the underlying log-price processes and the factors are all continuous. In the general case with the presence of jumps, we employ the standard truncation technique to remove jumps and the corresponding estimators are given in Section 3.3. Ignoring the jumps, over the i-th sampling interval, we have the following approximation for $\Delta_i^n \mathbf{Y} := \mathbf{Y}_{i\Delta_n} - \mathbf{Y}_{(i-1)\Delta_n}$:

$$\begin{aligned} \Delta_i^n \mathbf{Y} &= \int_{(i-1)\Delta_n}^{i\Delta_n} \Big(\boldsymbol{\alpha}_t dt + (\mathbf{G}_t + \mathbf{\Gamma}_t) d\mathbf{F}_t + d\mathbf{U}_t \Big) \\ &= \boldsymbol{\alpha}_{i-1}\Delta_n + (\mathbf{G}_{i-1} + \mathbf{\Gamma}_{i-1})\Delta_i^n \mathbf{F} + \Delta_i^n \mathbf{U} + o_P(\Delta_n), \end{aligned} \tag{3.1}$$

To nonparametrically estimate $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, we assume it can be well approximated on the sieve space spanned by nonlinear transformations of $\mathbf{X}_{lt}$. Specifically, 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 $\mathbf{\Phi}_t = (\boldsymbol{\phi}_{1t}, ..., \boldsymbol{\phi}_{pt})'$ be the $p \times J$ basis matrix, and define the projection matrix:

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

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.

3.1 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 $\mathbf{\Gamma}_{i-1} := \mathbf{\Gamma}_{(i-1)\Delta_n}$. 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

$$\begin{aligned}\widehat{\mathbf{G}}_t &= \sum_{i\in I_t^n} \mathbf{P}_{i-1}\,\Delta_i^n\mathbf{Y}\,\Delta_i^n\mathbf{F}'\left(\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\,\Delta_i^n\mathbf{F}'\right)^{-1},\\ \widehat{\mathbf{\Gamma}}_t &= \sum_{i\in I_t^n}(\mathbf{I}_p-\mathbf{P}_{i-1})\Delta_i^n\mathbf{Y}\,\Delta_i^n\mathbf{F}'\left(\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\,\Delta_i^n\mathbf{F}'\right)^{-1},\end{aligned} \tag{3.2}$$

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[2]

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

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

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

[2]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.

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

$$\mathbf{P}_{i-1}\Delta_i^n\mathbf{Y} \approx \mathbf{G}_{i-1}\Delta_i^n\mathbf{F} \tag{3.3}$$

is nearly "noise-free". Hence running OLS on each local interval I_t^n for (3.3) directly leads to consistent estimator of spot $\mathbf{G}_t$. 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 (3.1) and (3.3) yields:

$$\begin{aligned}(\mathbf{I}_p - \mathbf{P}_{i-1})\Delta_i^n\mathbf{Y} = &\mathbf{\Gamma}_{i-1}\Delta_i^n\mathbf{F} + \Delta_i^n\mathbf{U} + \text{higher-order term} \\ &+ \text{projection \& sieve approx. error.}\end{aligned} \tag{3.4}$$

(3.4) 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 $\mathbf{\Gamma}_{i-1}$.

3.2 Unknown Factor Case

When factors are unknown, we employ the principal component method to estimate the latent factors first. But different from Stock and Watson (2002a); Bai (2003), Aït-Sahalia and Xiu (2017) and Pelger (2016), we employ the PCA on the projected returns. Applying PCA on the projected data is still motivated by the "noise-free" model (3.3), which preserves the factors and removes the effect of idiosyncratics. Specifically, with each local window I_t^n, we can define the following $p \times k_n$ matrix:

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

The idea is that $\mathbf{P}_{i-1}\Delta_i^n\mathbf{Y} \approx \mathbf{P}_{i-1}\mathbf{G}_{i-1}\Delta_i^n\mathbf{F}$ and we choose the local window size in such a way that the time variation of $\mathbf{P}_{i-1}\mathbf{G}_{i-1}$ is asymptotically negligible within each local window. Hence we can apply PCA onto the "idiosyncratic-free" observations $(\mathbf{P}\Delta^n\mathbf{Y})_t$. Define the estimated factors

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

whose columns equal $\sqrt{\Delta_n}$ times the eigenvectors of the $k_n \times k_n$ matrix $\frac{1}{pk_n\Delta_n}(\mathbf{P}\Delta^n\mathbf{Y})_t'(\mathbf{P}\Delta^n\mathbf{Y})_t$ [3], corresponding to the first K eigenvalues. According to Okamoto (1973), these eigenvalues are distinct almost surely. We then use the same method to estimate $\mathbf{G}_t$ and $\mathbf{\Gamma}_t$, with the estimated factors in place of $\{\Delta_i^n\mathbf{F}\}_{i\in I_t^n}$:

$$\begin{aligned}\widehat{\mathbf{G}}_t^{\text{latent}} &= \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}\mathbf{P}_{i-1}(\Delta_i^n\mathbf{Y})\widehat{\Delta_i^n\mathbf{F}}', \\ \widehat{\mathbf{\Gamma}}_t^{\text{latent}} &= \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}(\mathbf{I}_N-\mathbf{P}_{i-1})(\Delta_i^n\mathbf{Y})\widehat{\Delta_i^n\mathbf{F}}' \end{aligned} \tag{3.5}$$

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.

3.3 Jump-robust estimators

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

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

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

where $\Delta_i^n Z_{\psi_n^Z} := \Delta_i^n Z_l\, 1_{\{\|\Delta_i^n Z_l\|\le\psi_n^{Z_l}\}}$ 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)).[4]

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.

[3]This is different from the $p\times p$ covariance matrix studied by Zheng and Li (2011).

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

3.4 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 idiosyncratic components, while the ordinary PCA does not. This potentially leads to more accurate estimations when k_n is small. Secondly, the projection removes the idiosyncratic beta component $\boldsymbol{\gamma}_{lt}$, 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 $\mathbf{\Omega} = (\mathbf{\Phi}_t'\mathbf{\Phi}_t)^{-1}$, the GMM optimization is given by

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

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 $\mathbf{\Gamma}_t$.

4 Informal Discussion of Uniformity

The estimated $\mathbf{g}_{lt}$ has the following asymptotic expansion. Let $\Delta_i^n U_l$ denote the l-th component of $\Delta_i^n\mathbf{U}$. Let $h_{i,ml} = \boldsymbol{\phi}_{i,m}'(\frac{1}{p}\mathbf{\Phi}_i'\mathbf{\Phi}_i)^{-1}\boldsymbol{\phi}_{i,l}$ and $\mathbf{s}_{f,t} = \frac{1}{k_n\Delta_n}\sum_{i\in I_t^n}\Delta_i^n\mathbf{F}\Delta_i^n\mathbf{F}'$.

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

The first term on the right hand side is related to the high frequency estimation of $\boldsymbol{\beta}$. As pointed out by Mykland and Zhang (2017), it is often quite challenging to estimate the limiting variance of high frequency estimators. Even so, we have another term from cross-sectional estimation: Term (**a**) has a rate $O_P(p^{-1/2}\|\mathbf{V}_\gamma\|^{1/2})$, with $\mathbf{V}_\gamma = \operatorname{Var}(\frac{1}{\sqrt{p}}\sum_{m=1}^{p}\boldsymbol{\gamma}_{m,t}h_{t,ml}|\mathbf{X}_t)$, where $\operatorname{Var}(\cdot|\mathbf{X}_t)$ denotes the conditional variance given $\mathbf{X}_t = \{\mathbf{X}_{mt}\}_{m\le p}$. We shall assume $\boldsymbol{\gamma}_{m,t}$'s are cross-sectionally weakly dependent so that (**a**) admits a cross-sectional central limit theorem (CLT).

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

$$\sqrt{k_n p}(\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}) = O_P(1),$$

and is asymptotically normal, whose asymptotic distribution is determined by the first term

(the "U" term). 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 spot level. On the other hand, if $\mathbf{V}_\gamma$ is *strong* with all eigenvalues bounded away from zero, (**a**) becomes the dominating term, and we simply have

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

In this case, the limiting distribution is determined by the cross-sectional CLT of (**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 $\widehat{\mathbf{g}}_{lt} - \mathbf{g}_{lt}$ when the signals of cross-sectional variation of $\boldsymbol{\gamma}_{ml}$ 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.

Furthermore, for the long-run mean forecast, we forecast $y_{L_n+1|L_n}$, which is, for instance, the conditional mean of the integrated volatility over the time span $[L_nT, (L_n+1)T]$, with $L_n \to \infty$. We construct mean forecast $\widehat{y}_{L_n+1|L_n}$, and provide the forecast confidence interval using the estimated factors and lagged integrated volatility. The forecast asymptotic variance also depends on $\{\boldsymbol{\Gamma}_t\}$ through its cross-sectional and serial dynamics. In particular, the asymptotic expansion of $\widehat{y}_{L_n+1|L_n} - y_{L_n+1|L_n}$ contains, among several others, a term like:

$$\mathcal{A} := \left[\sum_{i\in S_{L_n}} \frac{1}{p}\Delta_i^n \mathbf{F}'\boldsymbol{\Gamma}'_{i-1} - \frac{1}{L_n-1}\sum_{d=1}^{L_n-1} \mathbf{w}'_n \mathbf{z}_d \sum_{j\in S_d} \frac{1}{p}\Delta_j^n \mathbf{F}'\boldsymbol{\Gamma}'_{j-1}\right] \mathcal{G}$$

where S_d denotes the observation times on the interval $[(d-1)T, dT]$; $(\mathbf{w}_n, \mathbf{z}_d, \mathcal{G})$ are low-dimensional vectors to be defined in the subsequent sections. We find that $\mathcal{A} = 0$ if and only if $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_j$, for any $i, j \in \cup_d S_d$, otherwise this term also possesses a cross-sectional CLT at the rate $O_P(p^{-1/2})$. Hence $\boldsymbol{\Gamma}_t$ plays an important role if it is time-varying. In practice,

however, econometricians do not know the degree of variations on either the time domain or the cross-sectional domain of $\mathbf{\Gamma}_t$.

We provide a uniformly valid inference procedure. More specifically, we construct a confidence interval $CI_{\tau,n,p}$ for $\mathbf{g}_{lt}$, so that at the nominal level $1-\tau$,

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

and 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.

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. *(1) The process* $\mathbf{Y}$ *is an Itô semimartingale (with its continuous component is given by* (2.1)*, where the continuous component of* $\mathbf{F}$ *and* $\mathbf{U}$ *are given by* (2.2)*. 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* (2.1) *and* (2.2)*.*

(2) Each element $\boldsymbol{\theta}_{mt}$ *of* $\boldsymbol{\theta}_t = (\mathbf{x}_t', \{\mathbf{x}_l'\}, \{\mathbf{x}_{lt}'\}, \mathbf{\Gamma}_t', \boldsymbol{\sigma}_t^{F'}, \boldsymbol{\sigma}_t^{U'})'$*, is a multivariate Itô semi-*

martingale with the form

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

Here $\{\boldsymbol{\alpha}_{ms}\}_{s\ge 0}, \{\tilde{\boldsymbol{\sigma}}_{ms}\}_{s\ge 0}$ *and* $\{\check{\boldsymbol{\sigma}}_{ms}\}_{s\ge 0}$ *are optional processes and locally bounded uniformly in* $m \le p$*. In general,* $\boldsymbol{\theta}_m = \{\boldsymbol{\theta}_{mt}\}_{t\ge 0}$ *can be driven by* $\{\mathbf{W}_t^m = (\mathbf{W}^{F'}, \mathbf{W}_m^U)_t'\}_{t\ge 0}$*, where* $\mathbf{W}_m^U$ *is the m-th element of* $\mathbf{W}^U$ *introduced in* (2.2)*, and another multi-dimensional Brownian motion* $\{\check{\mathbf{W}}_t^m\}_{t\ge 0}$ *orthogonal to* $\{\mathbf{W}_t^m\}_{t\ge 0}$*. Finally,* $\Delta\boldsymbol{\theta}_{ms}$ *represents the (possible) jump of* $\boldsymbol{\theta}_m$ *at time* s*.* [5]

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

Remark 5.1. The class of Itô semimartingale is very large and includes most stochastic processes in the literature. Constants (or constant vector), Itô processes, Poisson processes and Lévy processes are all Itô semimartingale. Also, discrete-time processes can be viewed as a pure jump process with fixed jump time points.

Recall that $\mathbf{G}_t$ is the $p \times K$ matrix of $\mathbf{g}_{lt}(\mathbf{X}_{lt})$, and $\mathbf{P}_t = \boldsymbol{\Phi}_t(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'$.

Assumption 5.2. *There are absolute constants* $c, C, \eta > 0$*, so that*
(i) $\max_{l\le p, t\in[0,T]} \mathbb{E}\,\|\mathbf{g}_{lt}(\mathbf{X}_{lt})\|^4 < C$*,* $\mathbb{E}\,\|\boldsymbol{\gamma}_{lt}\|^4 < C$*.*
(ii) $\max_{t\in[0,T], l\le p} \mathbb{E}\,\|\boldsymbol{\phi}_{lt}\|^2 \le CJ$*.*
(iii) $c < \min_{t\le T} \lambda_{\min}(\frac{1}{p}\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t) \le \max_{t\le T} \lambda_{\max}(\frac{1}{p}\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t) < C$*.*
(iv) $\max_{t\in[0,T]} \|\mathbf{G}_t - \mathbf{P}_t\mathbf{G}_t\|_\infty \le C_1 J^{-\eta}$*, where conditions (iii) (iv) hold almost surely. In addition,* $\Delta_n = o(J^{-\eta})$*.*

Recall that $\|.\|_\infty$ denotes the "max" norm of a matrix. Condition (iv) of Assumption 5.2 requires the nonparametric instrument function can be well approximated by the sieve expansion. Furthermore, the condition $\Delta_n = o(J^{-\eta})$ ensures that the sieve approximation error is first order negligible.

We now describe the asymptotic variance of $\widehat{\mathbf{g}}_{lt}$. Let $\mathbf{u}_i := \Delta_i^n \mathbf{U}/\sqrt{\Delta_n}$ and $\mathbf{f}_i = \Delta_i^n \mathbf{F}/\sqrt{\Delta_n}$. Let $\mathbf{c}_{f,t}$ $(K \times K)$ and $\mathbf{c}_{u,t}$ $(p \times p)$ be the instantaneous quadratic variation process of $\mathbf{F} = \{\mathbf{F}_t\}_{t\ge 0}$ and $\mathbf{U} = \{\mathbf{U}_t\}_{t\ge 0}$, 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

[5] 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 $\Delta Z_s := Z_s - Z_{s-}$.

asymptotic variance of $\widehat{\mathbf{g}}_{lt}$ depends on:

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

Assumption 5.3. *(i) There are absolute constants $c, C > 0$, so that $\sup_{t\in[0,T]}\|\mathbf{c}_{u,t}\| < C$, and $\inf_{t\in[0,T]}\lambda_{\min}(\mathbf{c}_{t,f}) > c$. Almost surely, $c < \inf_{t\in[0,T]}\lambda_{\min}(\mathbf{V}_{u,t}) \le \sup_{t\in[0,T]}\lambda_{\max}(\mathbf{V}_{u,t}) < C$, and $\lambda_{\max}(\mathbf{V}_{\gamma,t}) \le C\lambda_{\min}(\mathbf{V}_{\gamma,t})$.*

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

Assumption 5.3 (i) requires the conditional covariance of standardized idiosyncratic components have bounded eigenvalues. This condition holds when the idiosyncratic components are cross-sectionally weakly correlated, which is a typical assumption for the factor models of large dimensions since most of the returns' cross-sectional variations are explained by the factors. On the other hand, Assumption 5.3 (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 $\boldsymbol{\Gamma}_t$.

Assumption 5.4 (For estimated factors)**.** *(i) Define $\boldsymbol{\Sigma}_{G,t} = \frac{1}{p}\mathbf{G}_t'\mathbf{G}_t$. Almost surely, $c < \inf_{t\le T}\lambda_{\min}(\boldsymbol{\Sigma}_{G,t}) \le \sup_{t\le T}\lambda_{\max}(\boldsymbol{\Sigma}_{G,t}) < C$ for absolute constants $c, C > 0$.*
(ii) The eigenvalues of $\boldsymbol{\Sigma}_{G,t}^{1/2}\,\mathbf{c}_{f,t}\,\boldsymbol{\Sigma}_{G,t}^{1/2}$ are distinct.

Assumption 5.4 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 $J^2 = O(p)$, and $(k_n + p)\Delta_n = o(1)$. Under Assumptions 5.1-5.3, 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)**.** *Suppose $J^2 = O(p)$, and $(k_n + p)\Delta_n = o(1)$. Under Assumptions 5.1-5.4, there is a $K \times K$ rotation matrix $\boldsymbol{\Upsilon}_{nt}$, as $J, p \to \infty$, (k_n either grows or stays constant)*

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

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 $O_P((k_np)^{-1/2})$, and

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

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.

Remark 5.2. 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.

Remark 5.3. 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 5.4. 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 5.5. It is well known that $\widehat{(\Delta^n \mathbf{F})}_t$ estimates $(\Delta^n \mathbf{F})_t$ only up to a $K \times K$ transformation, which results in the rotation $\mathbf{\Upsilon}_{ni}$ for the estimated spot $\mathbf{g}$. Here the rotation is time-varying so long as $\mathbf{G}_t$ is, but is locally time-invariant on I_t^n. This means, for any $j \in I_t^n$,

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

In fact, since the effect of $\mathbf{\Gamma}_t$ is removed by the projection, the time variation of $\mathbf{\Upsilon}_{nt}$ only depends on that of $\{\mathbf{G}_t\}_{t \le T}$. This implies that in conditional factor models where only $\mathbf{\Gamma}_t$ is time-varying but $\mathbf{G}_t$ is approximately time-invariant on the entire time span $[0, T]$, the rotation is *globally* time-invariant. This feature is also appealing since $\mathbf{G}_t$ is less volatile and picks up long-run beta patterns. It is characterized by instruments that change at a much lower frequency than that of the return data (Cosemans et al., 2009).

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:

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

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 $\omega_{np} = \sqrt{\frac{J}{p}} + \sqrt{\frac{\log p}{k_n}}$.[6]

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

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 covariance 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 $\widehat{\mathbf{P}_{i-1} \Delta_i^n \mathbf{U}} = \mathbf{P}_{i-1} \Delta_i^n \mathbf{Y} - \widehat{\mathbf{G}}_{i-1} \widehat{\Delta_i^n \mathbf{F}}$

[6]Hard-threholding takes $\text{th}(s_{dl}) = s_{dl}$, while soft-thresholding takes $\text{th}(s_{dl}) = \text{sgn}(s_{dl})(|s_{dl}| - \varrho_{dl})$. 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.

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 $\max_{i\in I_t^n}\|\widehat{\mathbf{c}}_{u,i} - \mathbf{c}_{u,i}\| = o_P(\sqrt{\frac{p}{Jk_n}}\Delta_n)$[7] 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 5.1-5.4, we have*

$$\boldsymbol{\Upsilon}_{nt}^{\prime -1/2}\left(\frac{1}{k_np}\mathbf{V}_{u,t} + \frac{1}{p}\mathbf{V}_{\gamma,t}\right)^{-1/2}\boldsymbol{\Upsilon}_{nt}^{-1/2}\left(\widetilde{\mathbf{g}}_{lt}^{\text{latent}} - \boldsymbol{\Upsilon}_{nt}\mathbf{g}_{lt}\right) \xrightarrow{\mathcal{L}\text{-}s} N(0, \mathbf{I}_K).$$

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

5.3.1 Why Bootstrap?

A seemingly natural inference procedure is to plug-in the estimated asymptotic covariances for $\mathbf{V}_{ut}$ and $\mathbf{V}_{\gamma,t}$ using their sample analogues. This procedure, however, works only pointwise in the current context, and does not provide a uniformly valid confidence interval. To understand the issue, consider the estimation of $\mathbf{V}_{\gamma,t}$. We focus on the case when $\{\boldsymbol{\gamma}_{mt}\}_{m\leq p}$ are cross-sectionally uncorrelated, conditionally on $\mathbf{X}_t$. Then

$$\mathbf{V}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^2 \,\text{Var}(\boldsymbol{\gamma}_{mt}|\mathbf{X}_t).$$

If $\boldsymbol{\gamma}_{mt}$ were known, White (1980)'s heteroskedastic covariance estimator can be applied: $\widetilde{\mathbf{V}}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^{p} h_{i,ml}^2\boldsymbol{\gamma}_{mt}\boldsymbol{\gamma}_{mt}'$. Replacing $\boldsymbol{\gamma}_{mt}$ with its consistent estimator $\widehat{\boldsymbol{\gamma}}_{mt}$, we obtain

[7] We shall verify this condition in Lemma D.1 for sparse covariance estimators.

$\widehat{\mathbf{V}}_{\gamma,t} = \frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^2 \widehat{\boldsymbol{\gamma}}_{mt}\widehat{\boldsymbol{\gamma}}_{mt}'$. Then $\widehat{\mathbf{V}}_{\gamma,t} - \mathbf{V}_{\gamma,t}$ has a decomposition

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

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 in the expansion of $\widehat{\boldsymbol{\gamma}}_{mt} - \boldsymbol{\gamma}_{mt}$ is $\mathbf{r}_m := \mathbf{s}_{f,t}^{-1}\frac{1}{\Delta_n k_n}\sum_{i\in I_t^n}\Delta_i^n \mathbf{F}\Delta_i^n U_m$, which leads to

$$\gamma\text{-estimation error} \geq \frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^2 \mathbf{r}_m\mathbf{r}_m' \asymp O_P(k_n^{-1}).$$

This results in an estimation error $\|\mathbf{V}_{\gamma,t} - \widehat{\mathbf{V}}_{\gamma,t}\|$ being *lower bounded* by an order $O_P(k_n^{-1})$, which is negligible only if it is dominated by the asymptotic variance:

$$\|\frac{1}{p}(\mathbf{V}_{\gamma,t} - \widehat{\mathbf{V}}_{\gamma,t})\| = o_P(1)\lambda_{\min}[\frac{1}{k_n p}\mathbf{V}_{u,t} + \frac{1}{p}\mathbf{V}_{\gamma,t}]$$

This is however, not the case whenever $\lambda_{\min}(\mathbf{V}_{\gamma,t}) = O_P(k_n^{-1})$ (corresponding to the case of weak γ-signal). Hence estimating $\mathbf{V}_{\gamma,t}$ introduces an estimation error that is non-negligible when $\{\boldsymbol{\gamma}_{ml}\}$ is weak. Consequently, the usual plug-in covariance estimator using $\widehat{\mathbf{V}}_{\gamma,t}$ would lead to over-coverage probabilities. But ignoring $\mathbf{V}_{\gamma,t}$ would result in under-coverage probabilities when $\{\boldsymbol{\gamma}_{ml}\}$ is strong. Hence it is is not uniformly valid.

The cross-sectional bootstrap, as we shall describe in the next subsection, resolves the uniformity issue. It directly mimics the cross-sectional variations in $\{\boldsymbol{\gamma}_{mi}\}$. The bootstrap asymptotic variance is analogously $\frac{1}{k_n p}\mathbf{V}_{u,t} + \frac{1}{p}\widetilde{\mathbf{V}}_{\gamma,t}$, and hence 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. The LLN error is of a higher order than $\mathbf{V}_{\gamma,t}$. For instance, suppose $\boldsymbol{\gamma}_{mt} = a_n\bar{\boldsymbol{\gamma}}_{mt}$, where $a_n \geq 0$ is a non-random arbitrary sequence, and $c_1 < \lambda_{\min}(\mathrm{Var}(\bar{\boldsymbol{\gamma}}_{ml}|\mathbf{X}_t)) \leq \lambda_{\max}(\mathrm{Var}(\bar{\boldsymbol{\gamma}}_{ml}|\mathbf{X}_t)) < C_2$. Then LLN error $= O_P(p^{-1/2})\mathbf{V}_{\gamma,t}$ so long as $\frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^4 \mathbb{E}(\|\bar{\boldsymbol{\gamma}}_{mt}\|^4|\mathbf{X}_t) \leq C\lambda_{\min}^2[\frac{1}{p}\sum_{m=1}^{p} h_{t,ml}^2 \mathrm{Var}(\bar{\boldsymbol{\gamma}}_{mi}|\mathbf{X}_t)]$

almost surely.

Remark 5.6. A possible alternative approach is to employ the thresholding: estimate $\mathbf{V}_{\gamma,t}$ using $\widehat{\mathbf{V}}_{\gamma,t}1_{\{\|\widehat{\mathbf{V}}_{\gamma,t}\|<c_n \log n\}}$ for some sequence $c_n \asymp \min\{k_n, \sqrt{p}\}^{-1}$, so that $c_n \log n$ "just dominates" $\|\mathbf{V}_{\gamma,t} - \widehat{\mathbf{V}}_{\gamma,t}\|$. The similar approach has been employed to deal with the distribution discontinuity in Andrews (1999); 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 $\mathbf{V}_{\gamma,t}$. It also leaves a question of choosing the constant in c_n. So we do not pursue it in this paper.

5.3.2 Cross-Sectional Bootstrap and its Uniform Validity

We propose a cross-sectional bootstrap to mimic the cross-sectional CLT as in Assumption 5.3. For this purpose, we need to assume 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 cross-sectional time series $\{\Delta_i^n Y_m^*, i \in \{1,\cdots,T_n\}\}_{m=1,...,p}$ and $\left\{\mathbf{X}_{mi}^* : i \in I_t^n\right\}_{m=1,...,p}$, where

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

Here $\{m_1, ..., m_p\}$ is a simple random sample with replacement from $\{1, ..., p\}$. Since we are interested in the instrumental beta 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_d, the entire time series $\{\Delta_i^n Y_{m_d}, i \in \{1,\cdots,T_n\}\}$ and $\{\mathbf{X}_{m_d,i}, i \in \{1,\cdots,T_n\}\}$ 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.[8]

We then let $\boldsymbol{\Phi}_i^* = (\boldsymbol{\phi}_{m_1,i}, ..., \boldsymbol{\phi}_{m_p,i})'$ 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 $\Delta_i^n \mathbf{Y}^* = (\Delta_i^n Y_1^*, ..., \Delta_i^n Y_p^*)'$ and $\mathbf{G}_i^* = (\mathbf{g}_{m_1,i}, ..., \mathbf{g}_{m_p,i})'$. Define

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

[8]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

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

in the estimated factor case. We repeat the bootstrap sampling and estimation for B times, and obtain either $\{\widehat{\mathbf{g}}_{lt}^{*b}\}_{b \le B}$ or $\{\widehat{\mathbf{g}}_{lt}^{*\,\text{latent},b}\}_{b \le B}$, depending on whether factors are observable.

When $\mathbf{g}_{lt}$ is multidimensional, it is easier to present the confidence interval for a linear transformation $\mathbf{v}'\mathbf{g}_{lt}$. For any 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 $\{|\mathbf{v}'\widehat{\mathbf{g}}_{lt}^{*\,\text{latent},b} - \mathbf{v}'\widehat{\mathbf{g}}_{lt}^{\text{latent}}|\}_{b \le B}$). The confidence interval for $\mathbf{v}'\mathbf{g}_{lt}$ (or $\mathbf{v}'\boldsymbol{\Upsilon}_{nt}\mathbf{g}_{lt}$ in the estimated factor case) is given by

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

We need the following conditions for the bootstrap validity.

Assumption 5.5. *(i) Conditionally on* $\{\mathbf{X}_t\}$, $\{\{\Delta_i^n U_m\}_{i \in I_t^n}, \{\boldsymbol{\gamma}_{mi}\}_{i \in I_t^n}\}_{m \le p}$ *are cross-sectionally uncorrelated.*
(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* $C, \eta > 0$.

Remark 5.7. The bootstrap takes independent samples from the cross-sectional units. Hence Assumption 5.5 (i) is required for the bootstrap to mimic the cross-sectional variations. One of the potential ways to allow for cross-sectional correlations is to normalize the cross-sectional units using the estimated estimated error covariance matrix $\mathbf{c}_{u,t}^{-1/2}$, which can be well estimated if it is a sparse covariance matrix. By assuming that $\boldsymbol{\Gamma}$ and the increments of $\mathbf{U}$ have similar cross-sectional dependence structure, i.e., $\mathbf{c}_{u,t}^{-1/2} \operatorname{Var}(\boldsymbol{\Gamma}_t|\mathbf{X}_t)\mathbf{c}_{u,t}^{-1/2}$ is nearly diagonal for $t \in [0,T]$, we can bootstrap the transformed data and the transformed $\{\boldsymbol{\Gamma}_{i-1}\}$ and $\{\Delta_i^n \mathbf{U}\}$ would be nearly cross-sectional uncorrelated. However, we do not pursue this approach in this paper.

Finally, we require the following moment conditions on $\boldsymbol{\Gamma}_t$:

Assumption 5.6. *There is an absolute constant* $C > 0$, *almost surely,*

$$\frac{\lambda_{\max}^{1/2}(\frac{1}{p}\sum_{m=1}^p h_{i,ml}^4 \mathbb{E}(\|\boldsymbol{\gamma}_{mi}\|^4|\mathbf{X}_i))}{\lambda_{\min}(\frac{1}{p}\sum_{m=1}^p h_{i,ml}^2 \operatorname{Var}(\boldsymbol{\gamma}_{mi}|\mathbf{X}_i))} < C, \quad \frac{\lambda_{\max}(\frac{1}{p}\sum_{m=1}^p h_{t,mm}h_{t,ll} \operatorname{Var}(\boldsymbol{\gamma}_{mt}))}{\lambda_{\min}(\frac{1}{p}\sum_{m=1}^p h_{i,ml}^2 \operatorname{Var}(\boldsymbol{\gamma}_{mi}|\mathbf{X}_i))} < C$$

If $\text{Var}(\boldsymbol{\gamma}_{mi}|\mathbf{X}_i) = 0$ *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 5.5, 5.6 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* $l \leq p, t \in [0, T]$

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

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

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

Theorem 5.5 (long-run g)**.** *Consider the known factor case*[9]*. Let* $\mathcal{P}$ *be the collection of all data generating processes* $\mathbb{P}$ *for which Assumption 5.5 and assumptions of Theorem 5.1 hold. Then for any fixed vector* $\mathbf{v} \in \mathbb{R}^K \backslash \{0\}$ *such that* $\|\mathbf{v}\| > c > 0$*, for each fixed* $l \leq p$*,*

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

6 Uniform Confidence Intervals for Long-Run Forecast

The object of interest is to forecast the conditional mean $y_{L_n+h|L_n}$ of model (2.6):

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

[9]Due to the rotation discrepancy, estimating the long-run g in the presence of time-varying beta is a much harder problem, and we shall leave it for the future research.

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:

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

Of particular interest is $y_d = \mathrm{IV}_d$, the integrated volatility a single asset. Note that $\{\mathrm{IV}_d\}_{d\leq L_n}$ is not directly observable, and has to be nonparametrically estimated. Then we have two types of *estimated regressors*: estimated integrated volatility $\{\widehat{\mathrm{IV}}_d\}_{d\leq L_n}$, and the estimated integrated factors $\{\widehat{\mathbf{F}}_d\}_{d\leq L_n}$. 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 $\{\mathrm{IV}_d, \mathbf{F}_d\}_{d\leq L_n}$.

We now describe the construction of $\{\widehat{\mathrm{IV}}_d\}_{d\leq L_n}$ and $\{\widehat{\mathbf{F}}_d\}_{d\leq L_n}$. Consider a long-run forecast, where the integrated latent factors are estimated from discrete-time return data $\{\Delta_i^n \mathbf{Y}\}_{i\leq M_n}$, with $M_n = nL_n$, realized from: $d\mathbf{Y}_t = \boldsymbol{\alpha} dt + (\mathbf{G}_t + \boldsymbol{\Gamma}_t) d\mathbf{F}_t + d\mathbf{U}_t$, $\forall t \in [0, L_n T]$. Importantly, we require $\mathbf{G}_t$ and the drift part $\boldsymbol{\alpha}$ be time-invariant over the entire interval $[0, L_n T]$. On the other hand, we still allow $\boldsymbol{\Gamma}_t$ to be time-varying with a realized trajectory driven by the Brownian motion. [10]

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

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

As for the integrated volatility, we estimate the integrated volatility over $[(d-1)T, dT]$

[10] 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.

by using truncated Bi-power variation (TBPV) as follows

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

According to Theorem 13.2.4 in Jacod and Protter (2011), the asymptotic property of $\widehat{\mathrm{IV}}_d$ is given as follows:

$$\frac{1}{\sqrt{\Delta_n}} \frac{\widehat{\mathrm{IV}}_d - \mathrm{IV}_d}{\sqrt{2 \int_{(d-1)T}^{dT} \left(\mathbf{c}_{ll,t}\right)^2 dt}} \xrightarrow{\mathcal{L}\text{-}s} N(0,1),$$

where $\mathbf{c}_t = \mathbf{c}_{f,t} + \mathbf{c}_{u,t}$ (see the paragraph preceding Assumption 5.3) is the instantaneous quadratic variation matrix of $\mathbf{Y}$ at time t and $\mathbf{c}_{ll,t}$ is its (l,l) element.

Finally, we estimated $(\mu, \boldsymbol{\rho})$ by:

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

Then the forecasted conditional mean of $\mathrm{IV}_{L_n+h|L_n} := \mu \mathrm{IV}_{L_n} + \boldsymbol{\rho}' \mathbf{F}_{L_n}$ is

$$\widehat{\mathrm{IV}}_{L_n+h|L_n} = \widehat{\mu} \widehat{\mathrm{IV}}_{L_n} + \widehat{\boldsymbol{\rho}}' \widehat{\mathbf{F}}_{L_n}.$$

6.1 Effect of Time-Varying Gammas on Forecast Intervals

To describe the asymptotic property of $\mathrm{IV}_{L_n+h|L_n} - \widehat{\mathrm{IV}}_{L_n+h|L_n}$, we introduce some notation: let $\mathbf{z}_d = (\mathrm{IV}_d, (\mathbf{H}_n \mathbf{F}_d)')'$, 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 $\mathbf{w}_n$. Here $\mathbf{H}_n$ is the rotation matrix so that $\widehat{\mathbf{F}}_d$ consistently estimates $\mathbf{H}_n \mathbf{F}_d$, whose definition is given in the Appendix.

Let $\mathcal{G} = \mathbf{G}_{L_n} \left(\frac{1}{p} \mathbf{G}_{L_n}' \mathbf{G}_{L_n} \right)^{-1} \boldsymbol{\rho}$. Then we have

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

where

$$\bar{r}_1 \;=\; \mathbf{w}_n' \frac{1}{L_n - h} \sum_{d=1}^{L_n - h} \mathbf{z}_d v_{d+1},$$

$$
\begin{aligned}
\bar{r}_2 &= \sum_{i\in S_{L_n}} \frac{1}{p}\Delta_i^n \mathbf{U}'\mathbf{G}_{L_n} \\
\bar{r}_3 &= \mu(\widehat{\mathrm{IV}}_{L_n} - \mathrm{IV}_{L_n}) \quad \text{(effect of nonparametrically estimate the integrated volatility)} \\
\bar{r}_4 &= \left[w_1 \frac{1}{L_n - h}\sum_{d=1}^{L_n-h} \mathrm{IV}_d - 1\right](\frac{1}{L_n}\sum_{d=1}^{L_n}\mathbf{F}_d)'\boldsymbol{\rho} \quad \text{(effect of bias correction for estimated factors)} \\
\bar{r}_5 &= \frac{1}{p}\left[\sum_{i\in S_{L_n}} \Delta_i^n \mathbf{F}'\boldsymbol{\Gamma}_i' - \frac{1}{L_n - h}\sum_{d=1}^{L_n-h} \mathbf{w}_n'\mathbf{z}_d \sum_{j\in S_d}\Delta_j^n \mathbf{F}'\boldsymbol{\Gamma}_j'\right]\mathcal{G}. \qquad (6.2)
\end{aligned}
$$

All these terms contribute to the limiting distribution. In addition to the first two terms similar to those of the diffusion index forecast model (Bai and Ng (2006)), we also have three new leading terms, each representing a new feature of our forecast model.

Among these terms, we would like to pay a special attention to $\bar{r}_5$, which is due to the effect of idiosyncratic betas in the time-varying factor loadings. It is an interesting matter of fact that $\bar{r}_5$ equals zero if $\boldsymbol{\Gamma}_t$ is time-invariant, but is not so in general. [11] In the presence of high-frequency movements in $\boldsymbol{\Gamma}_t$, this term is not negligible. Therefore, the forecast procedure based on estimated factors would be misleading when *either* (i) ignore the $\boldsymbol{\Gamma}_t$ component in the betas, *or* (ii) treat $\boldsymbol{\Gamma}_t$ as time-invariant. The forecast interval we present below is uniformly valid across models with various strengths and degrees of time-varying in $\boldsymbol{\Gamma}_t$.

The above expansion leads to the asymptotic variance of $\mathrm{IV}_{L_n+h|L_n} - \widehat{\mathrm{IV}}_{L_n+h|L_n}$:

$$
\frac{1}{L_n}\mathcal{V}_1 + \frac{1}{p}\mathcal{V}_2 + \Delta_n\mathcal{V}_3 + \frac{1}{L_n}\mathcal{V}_4 + \frac{1}{p}\mathcal{V}_5
$$

where, for $\mathcal{G}_l$ as the l th element of $\mathcal{G}$,

$$
\begin{aligned}
\mathcal{V}_1 &= \frac{1}{L_n - h}\sum_{d=1}^{L_n-h} (\mathbf{w}_n'\mathbf{z}_d)^2 \operatorname{Var}(v_{d+1}|\{\mathrm{IV}_d, \mathbf{F}_d^c\}_{d\le L_n}) \\
\mathcal{V}_2 &= \frac{1}{p}\sum_{i\in S_{L_n}} \mathcal{G}'\operatorname{Var}(\Delta_i^n\mathbf{U}|\mathcal{F}_{i-1})\mathcal{G} \\
\mathcal{V}_3 &= 2\int_{(d-1)T}^{dT} \left(\mathbf{c}_{ll,t}\right)^2 dt
\end{aligned}
$$

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

$$
\begin{aligned}
\mathcal{V}_4 &= \left[w_1 \frac{1}{L_n - h} \sum_{d=1}^{L_n - h} \mathrm{IV}_d - 1 \right]^2 \boldsymbol{\rho}' \operatorname{Var}(\mathbf{F}_d^c) \boldsymbol{\rho} \\
\mathcal{V}_5 &= \frac{1}{p} \sum_{l=1}^{p} \mathcal{G}_l^2 \, \mathbb{E}\left[\Omega_l^2 | \{\mathrm{IV}_d\}_{d \le L_n}, \{\mathbf{X}_t, \Delta_t^n \mathbf{F}\}_{t \in [0, L_n T]} \right] \\
\Omega_l &= \sum_{i \in S_{L_n}} \Delta_i^n \mathbf{F}' \boldsymbol{\gamma}_{il} - \frac{1}{L_n - h} \sum_{d=1}^{L_n - h} \mathbf{w}_n' \mathbf{z}_d \sum_{j \in S_d} \Delta_j^n \mathbf{F}' \boldsymbol{\gamma}_{jl}.
\end{aligned}
$$

The asymptotic variance can be estimated by the plug-in method using the sample analogues. [12] The following covariance estimators are robust to heteroskedasticity: let $\widehat{\Delta_i^n \mathbf{F}} = \widehat{\Delta_i^n \mathbf{F}} - \frac{1}{L_n} \sum_{d=1}^{L_n} \widetilde{\mathbf{F}}_d$, $\widehat{\mathbf{G}} = \frac{1}{\Delta_n M_n} \sum_{d=1}^{L_n} \sum_{j \in S_d} \mathbf{P}_j \Delta \mathbf{Y}_j \widehat{\Delta_i^n \mathbf{F}}'$, $\widehat{\mathcal{G}} = \widehat{\mathbf{G}} (\frac{1}{p} \widehat{\mathbf{G}}' \widehat{\mathbf{G}})^{-1} \widehat{\boldsymbol{\rho}}$, and $\widehat{\mathcal{G}}_l$ denotes the l th element of $\widehat{\mathcal{G}}$, $\widehat{\boldsymbol{\Gamma}}_t = \frac{1}{k_n \Delta_n} \sum_{j \in I_t^n} (\mathbf{I} - \mathbf{P}_j) \Delta_j^n \mathbf{Y} \widehat{\Delta_i^n \mathbf{F}}'$. Let

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

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

$$
\widehat{\Omega}_l = \sum_{i \in S_{L_n}} \widehat{\Delta_i^n \mathbf{F}}' \widehat{\boldsymbol{\gamma}}_{il} + \frac{1}{L_n - h} \sum_{d=1}^{L_n - h} \widehat{\mathbf{w}}_n' \widehat{\mathbf{z}}_d \sum_{j \in S_d} \widehat{\Delta_i^n \mathbf{F}}' \widehat{\boldsymbol{\gamma}}_{jl}.
$$

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) $\{\Delta_t^n U_l\}_{l \le p}$ *are cross-sectionally independent, given* $\mathcal{F}_{t-1}$.

[12] 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 $\boldsymbol{\Gamma}_t$.

(iv) Let m_n denote the number of observations in time on the interval $S_d = [(d-1)T, dT]$. Assume $\{\mathbf{X}_t, \mathbf{G}_t\}$ be nearly time-invariant. Specifically, define $\mathbf{P}_t = \mathbf{\Phi}_t(\mathbf{\Phi}_t'\mathbf{\Phi}_t)^{-1}\mathbf{\Phi}_t'$, and

$$\begin{aligned}
\varrho_{1n}^2 &:= \frac{1}{L_n}\sum_{d=1}^{L_n}\|\bar{\mathbf{P}}_d - \frac{1}{L_n}\sum_{q=1}^{L_n}\bar{\mathbf{P}}_q\|^2, \quad \bar{\mathbf{P}}_d = \frac{1}{m_n}\sum_{i\in S_d}\mathbf{P}_{i-1} \\
\varrho_{2n}^2 &:= \max_{d\le L_n}\frac{1}{m_n}\sum_{i\in S_d}\|[\mathbf{P}_{i-1} - \mathbf{P}_{L_n}]\|^2 \\
\varrho_{3n}^2 &:= \max_{d\le L_n-h}\frac{1}{m_n}\sum_{i\in S_d}\|\frac{1}{\sqrt{p}}(\mathbf{G}_{L_n} - \mathbf{G}_{i-1})\|^2 \\
\varrho_{4n}^2 &:= \max_i\|\mathbf{P}_{i-1} - \frac{1}{L_n}\sum_{q=1}^{L_n}\bar{\mathbf{P}}_q\|^2.
\end{aligned}$$

Then $(\varrho_{1n}^2 + \varrho_{4n}^2 + \varrho_{3n}^2 m_n + \varrho_{2n}^2 m_n J)\min\{p, L_n\} = o(1)$.

Condition (i) is a CLT applied to the cross-sectional units of $\mathbf{\Gamma}_t$. 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 nearly constant over the entire range $[0, L_nT]$. But note that we still allow time-varying betas, thanks to the time-varying $\{\mathbf{\Gamma}_t\}$. This is still a plausible condition since the instruments mainly capture the long-run changes in beta.

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

Theorem 6.1. *Suppose Assumptions 5.1-5.2, 5.4, 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*

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

Remark 6.1. Though we study the long-time out-of-sample forecast in this section, the result provided can be also used as the basis of testing economic hypothesis that involves estimated factors from conditional factor models with large dimensions, with h set to zero. It also sheds lights on the distribution of estimated factors in the presence of both instrumental and idiosyncratic betas. The asymptotic distribution of the estimated factors would be

determined by that of two orthogonal components:

$$\begin{aligned}\bar{r}_2 &= \sum_{i\in S_{L_n}} \frac{1}{p}\Delta_i^n \mathbf{U}'\mathbf{G}_{L_n} \\ \bar{r}_5 &= \frac{1}{p}\left[\sum_{i\in S_{L_n}} \Delta_i^n \mathbf{F}'\boldsymbol{\Gamma}_i' - \frac{1}{L_n-h}\sum_{d=1}^{L_n-h} \mathbf{w}_n'\mathbf{z}_d \sum_{j\in S_d} \Delta_j^n \mathbf{F}'\boldsymbol{\Gamma}_j'\right]\mathcal{G}. \end{aligned} \tag{6.3}$$

While the term $\bar{r}_2$ is similar to the regular principal components estimators (Bai (2003)), in the presence of time-varying idiosyncratic betas, $\bar{r}_5$ is present as a new term, and also contributes to the asymptotic distribution.

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 $\boldsymbol{\Phi}_t = (\mathbf{X}_{1t}, ..., \mathbf{X}_{pt})'$ and $\mathbf{P}_t = \boldsymbol{\Phi}_t(\boldsymbol{\Phi}_t'\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t'$. Then $\widehat{\mathbf{G}}_t$ can be defined as before: (3.2) in the known factor case, and (3.5) in the unknown factor case. Then we run a cross-sectional regression to estimate $\boldsymbol{\theta}_t$:

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

As for the asymptotic analysis, the predescribed uniformity issue is still present. For instance,

in the known factor case, it can be shown that

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

The strength of the cross-sectional variations in $\boldsymbol{\Phi}_t'\boldsymbol{\Gamma}_t$ 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 $\boldsymbol{\Gamma}_t$ may stochastically dominate the strength of the asymptotic variance from $\boldsymbol{\Phi}_t'\boldsymbol{\Gamma}_t$, 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 independently resample 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 $\boldsymbol{\Phi}_i^* = (\mathbf{X}_{1,i}^*, ..., \mathbf{X}_{p,i}^*)'$, $\mathbf{P}_i^* = \boldsymbol{\Phi}_i^*(\boldsymbol{\Phi}_i^{*'}\boldsymbol{\Phi}_i^*)^{-1}\boldsymbol{\Phi}_i^{*'}$ and $\Delta_i^n\mathbf{Y}^* = (\Delta_i^n Y_1^*, ..., \Delta_i^n Y_p^*)'$. Let the bootstrap estimator be $\widehat{\boldsymbol{\theta}}_t^* = \left(\boldsymbol{\Phi}_t^{*'}\boldsymbol{\Phi}_t^*\right)^{-1}\boldsymbol{\Phi}_t^{*'}\widehat{\mathbf{G}}_t^*$, 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

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

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 5.5, 5.6 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 Estimating Factor Risk Premia using Instrumental Betas

The estimated instrumental betas, due to potentially faster rates of convergence than the regular betas, have the potential of improving the estimation of factor risk premium for nontradable factors. Consider the following model for estimating the factor risk premium: on a specific day d,

$$R_{ld} = \boldsymbol{\beta}'_{ld}\boldsymbol{\lambda}_d + \varepsilon_{ld}, \quad l = 1, ..., p, \mathbb{E}\varepsilon_{id} = 0.^{13}$$

where R_{ld} is the daily return of asset l on day d, and $\boldsymbol{\lambda}_d$ is the factor risk premium. The matrix form is given by $\mathbf{R}_d = \boldsymbol{\beta}_d\boldsymbol{\lambda}_d + \boldsymbol{\varepsilon}_d$, where $\boldsymbol{\beta}_d = (\boldsymbol{\beta}_{1d}, ..., \boldsymbol{\beta}_{pd})'$, and the betas have the similar decomposition:

$$\boldsymbol{\beta}_d = \mathbf{G}_d + \boldsymbol{\Gamma}_d,$$

respectively represent the instrumental and idiosyncratic betas on day d. The standard Fama-MacBeth procedure (Fama and MacBeth, 1973) runs a cross-sectional regression of R_{ld} onto the estimated betas $\boldsymbol{\beta}_{ld}$. This procedure has been known to be sensitive to the accuracy of the estimated beta, and may perform poorly as the estimated betas require a relatively large panel and is not applicable if $\boldsymbol{\beta}_d$ changes in high-frequency. In contrast, the instrumental betas as being considered in the current context, is much less volatile on the time domain, and can be estimated with a much better accuracy.

Here we heuristically show that one can apply the Fama-MacBeth regression using the instrumental betas. Applying the sieve-projection matrix on the matrix form, we obtain

$$\mathbf{P}_d\mathbf{R}_d \;=\; \mathbf{G}_d\boldsymbol{\lambda}_d + \underbrace{\mathbf{P}_d\boldsymbol{\varepsilon}_d + \mathbf{P}_d\boldsymbol{\Gamma}_d\boldsymbol{\lambda}_d}_{\text{projection error}} + \underbrace{(\mathbf{P}_d\mathbf{G}_d - \mathbf{G}_d)\boldsymbol{\lambda}_d}_{\text{sieve approximation error}} \approx \mathbf{G}_d\boldsymbol{\lambda}_d.$$

Hence we can modify the Fama-MacBeth procedure by running the cross-sectional regression of the "projected return" $\mathbf{P}_d\mathbf{R}_d$ on day d onto the estimated instrumental beta $\widehat{\mathbf{G}}_d$, and estimate the risk premium $\boldsymbol{\lambda}_d$ by:

$$\widehat{\boldsymbol{\lambda}}_d = (\widehat{\mathbf{G}}'_d\widehat{\mathbf{G}}_d)^{-1}\widehat{\mathbf{G}}'_d\mathbf{R}_d$$

[13] One can also include a common intercept term and have $R_{id} = \mathbf{a}_d + \boldsymbol{\beta}'_{id}\boldsymbol{\lambda}_d + \varepsilon_{id}$.

Importantly, in the "projection error" as shown above, the high-frequency beta components $\mathbf{\Gamma}_d$ is "projected off", and converges fast as $p \to \infty$. As we have shown that $\mathbf{G}_d$ can be estimated much more accurately even under time-varying factor models, this modified procedure would produce a much higher-quality estimation of the factor risk premium. Due to the space limit, we do not formally pursue theoretical property of this procedure.

8 Simulations

We conduct a simple simulation study on the estimated spot instrumental beta, to illustrate the issue of uniformity and the under/over coverages of the usual plug-in methods. The locally constant betas are generated from the following discrete time DGP:

$$\Delta_j^n \mathbf{Y} = \boldsymbol{\alpha} + (\mathbf{G} + \mathbf{\Gamma})\Delta_j^n \mathbf{F} + \Delta_j^n \mathbf{U}, \quad j = 1, ..., k_n$$

where $\mathbf{G} = 3\mathbf{X} + 1$, and $\mathbf{\Gamma} = \sqrt{w_\gamma}\mathbf{\Gamma}_0$. Here $\Delta_j^n \mathbf{U} \sim N(0, \mathbf{I})\sqrt{\Delta_n}$, $\Delta_j^n \mathbf{F} \sim N(0,1)\sqrt{\Delta_n}$, and $\alpha_m = \Delta_n$ for each $m \leq p$. The number of factors $K = 1$. The cross-sectional components of $\mathbf{X}$ are generated independently from $N(0,1)$, and the components of $\mathbf{\Gamma}_0$ are generated independently from $N(0,1)$, Here w_γ is taken as a scalar value in the range $[0.001, 0.3]$, which determines the strength of $\mathbf{\Gamma}$. The goal is to study the coverage properties of g_1, the first component of $\mathbf{G}$, with various values of w_γ. We set $\Delta_n = (pk_n)^{-1}$. As described in the paper, we use the cross-sectional bootstrap to generate critical values for the estimated g_1, and construct confidence intervals. The number of bootstrap replications is $B = 5000$.

We construct the confidence interval using three methods and compare the coverage probabilities:

(i) the bootstrap confidence interval: $\widehat{g}_1 \pm q_{\tau,bootstrap}$

(ii) the "over-coverage" confidence interval: $\widehat{g}_1 \pm 1.96\sqrt{\frac{1}{k_n p}\widehat{\mathbf{V}}_u + \frac{1}{p}\widehat{\mathbf{V}}_\gamma}$

(iii) the "under-coverage" confidence interval: $\widehat{g}_1 \pm 1.96\sqrt{\frac{1}{k_n p}\widehat{\mathbf{V}}_u}$, where

$$\begin{aligned}
\widehat{\mathbf{V}}_u &= \mathbf{s}_{f,t}^{-1}\frac{1}{pk_n}\sum_{i=1}^{k_n}\boldsymbol{\phi}_{i,l}'(\frac{1}{p}\mathbf{\Phi}_i'\mathbf{\Phi}_i)^{-1}\mathbf{\Phi}_i'\,\mathrm{diag}\{\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\widehat{\Delta_j^n\mathbf{U}\Delta_j^n\mathbf{U}}'\}\mathbf{\Phi}_i(\frac{1}{p}\mathbf{\Phi}_i'\mathbf{\Phi}_i)^{-1}\boldsymbol{\phi}_{i,l} \\
\widehat{\mathbf{V}}_\gamma &= \frac{1}{p}\sum_{m=1}^{p}h_{ml}^2\widehat{\boldsymbol{\gamma}}_m\widehat{\boldsymbol{\gamma}}_m', \quad l = 1, h_{ml} = \boldsymbol{\phi}_{i,l}'(\frac{1}{p}\mathbf{\Phi}_i'\mathbf{\Phi}_i)^{-1}\boldsymbol{\phi}_{i,m}'.
\end{aligned}$$

Note that the "over-coverage" uses the naive plug-in method to estimate the asymptotic

variance. This method can lead to over-coverage probabilities when w_γ is near zero. In addition, the "under-coverage" is the benchmark method because it ignores the variance component coming from $\mathbf{\Gamma}$, which is the common practice in the literature. Consequently, it is expected to produce substantial under-coverage probabilities when $\mathbf{V}_\gamma$ is non-negligible.

Table 1: Coverage probabilities for the spot g_1, nominal probability= 95%

			w_γ			
k_n	p		0.001	0.1	1	3
10	100	bootstrap	0.951	0.947	0.937	0.942
		over	0.995	0.990	0.950	0.943
		under	0.946	0.926	0.430	0.279
10	300	bootstrap	0.946	0.947	0.950	0.948
		over	0.993	0.990	0.962	0.953
		under	0.931	0.923	0.437	0.294
100	100	bootstrap	0.945	0.946	0.944	0.942
		over	0.993	0.982	0.944	0.941
		under	0.928	0.822	0.151	0.106
100	300	bootstrap	0.947	0.948	0.954	0.948
		over	0.994	0.986	0.946	0.950
		under	0.939	0.841	0.156	0.089
200	100	bootstrap	0.941	0.944	0.945	0.942
		over	0.990	0.968	0.943	0.943
		under	0.923	0.754	0.121	0.059
200	300	bootstrap	0.948	0.946	0.949	0.953
		over	0.991	0.973	0.948	0.953
		under	0.929	0.756	0.111	0.069

Table 1 summarizes the coverage probabilities using 2000 replications under the 95% nominal coverage. The numerical findings are consistent with what our theory predicts: (1) The bootstrap confidence interval has good coverage probabilities, uniformly over w_γ. (2) The "over-coverage" method is significantly conservative when w_γ is small, and becomes better as w_γ increases. (3) The "under-coverage" method has a fine coverage (but still with noticeable size distortions) when w_γ is near zero, but quickly has substantial under coverages

as w_γ increases. (iv) The coverage probabilities for the bootstrap is satisfactory even if k_n is small.

9 Empirical Studies

9.1 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.

9.2 Data analysis

Our estimation is based on the 5-minute frequency, and naturally the 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.

[14] We divide the assets into three categories: large, medium, and low, based on either the firms' size or the volatility characteristics. Figures 1 and 2 plot 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. As shown in Figure 1, 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 first panel of Figure 2, 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.

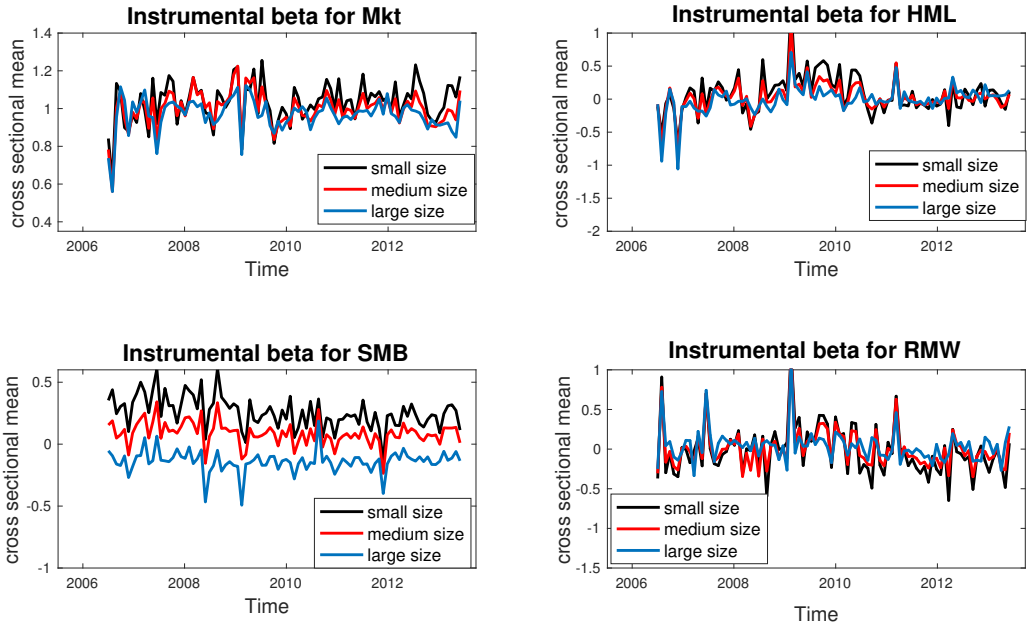

Figure 1: Cross-sectional means of instrumental beta, grouped by size. The instrumental betas are estimated on a daily basis, and this figure plots eight days' estimations for each month.

Next, we compare the standard deviations of the estimated two components in beta. The cross-sectional standard deviation is the sample standard deviation of the estimated $\mathbf{g}_{lt}$ and

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

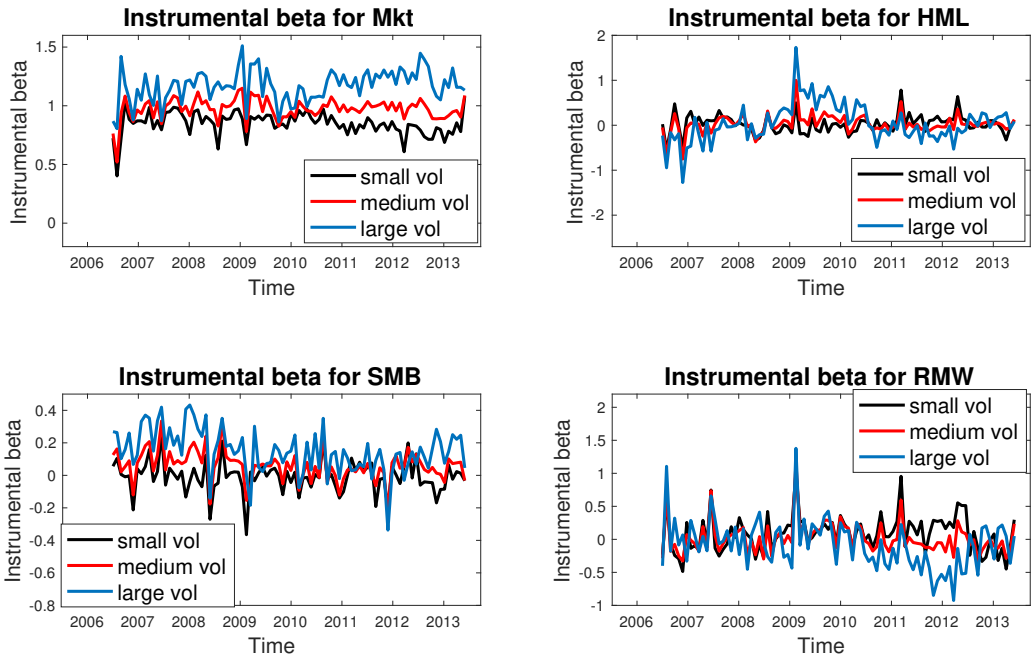

Figure 2: Cross-sectional means of instrumental beta, grouped by volatility. The instrumental betas are estimated on a daily basis, and this figure plots eight days' estimations for each month.

$\boldsymbol{\gamma}_{lt}$ among firms in each group for each fixed day. Then averaging these standard deviations over all days leads to the "averaged cross-sectional standard deviations". On the other hand, the time-series standard deviation is the sample standard deviation of the estimated $\mathbf{g}_{lt}$ and $\boldsymbol{\gamma}_{lt}$ over time for each fixed firm. Then averaging these time-series standard deviation across firms in each group leads to the "averaged time-series standard deviations". Here groups (small, medium, large) are determined by either the size or the volatility of the firms. So they respectively measure the cross-sectional and time-series variations of the two beta components. The results are given in Table 2 below, and show several interesting patterns: (1) The instrumental beta always possess significantly smaller standard deviations, in both cross-sectional and time-series, than the idiosyncratic beta. It demonstrates that there are relatively smaller cross-sectional variations in the instrumental betas among S&P500 firms. In addition, instrumental betas, as they capture long-run beta movements, are much less volatile in the time domain. (2) On average, firms with larger size and smaller volatilities tend to have smaller cross-sectional and time series variations in both components of beta. These firms have larger market values, whose betas are often more stable than the others.

(3) Even for firms with larger size and smaller volatilities, the time series standard deviations of the instrumental betas are noticeably different from zero, showing a significant degree of time-variations in betas.

Table 2: Averaged cross-sectional and time-series standard deviations of $\mathbf{G}$ and $\mathbf{\Gamma}$

	Mkt		HML		SMB		RMW	
	$\mathbf{G}$	$\mathbf{\Gamma}$	$\mathbf{G}$	$\mathbf{\Gamma}$	$\mathbf{G}$	$\mathbf{\Gamma}$	$\mathbf{G}$	$\mathbf{\Gamma}$
	Averaged (over time) cross-sectional std							
	grouped by size							
small	0.221	0.499	0.409	1.140	0.144	0.613	0.336	1.405
medium	0.179	0.446	0.352	1.063	0.112	0.524	0.281	1.290
large	0.165	0.411	0.333	0.994	0.142	0.466	0.260	1.245
	grouped by volatility							
small	0.082	0.352	0.267	0.799	0.182	0.407	0.186	0.949
medium	0.085	0.425	0.271	1.018	0.178	0.503	0.206	1.219
large	0.178	0.550	0.412	1.310	0.204	0.668	0.325	1.670
	Averaged (over firms) times-series std							
	grouped by size							
small	0.195	0.471	0.407	1.095	0.175	0.590	0.431	1.332
medium	0.152	0.416	0.324	0.993	0.140	0.511	0.353	1.214
large	0.148	0.389	0.315	0.908	0.133	0.460	0.336	1.124
	grouped by volatility							
small	0.142	0.368	0.294	0.861	0.131	0.447	0.324	1.042
medium	0.155	0.403	0.324	0.946	0.139	0.490	0.347	1.158
large	0.194	0.503	0.426	1.187	0.177	0.622	0.445	1.467

9.3 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 3 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 betas, though the pattern is not as noticeable as on the SMB betas.

Table 3: Proportion of significant $\mathbf{G}$ 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

When we group firms by the volatility, however, the pattern demonstrates noticeable variations over years. The results are given in Table 4. Note that results of 2010 are similar to 2011, and results in 2007 are similar to 2006. Firms with larger volatility tend to be more positively exposed to the HML factors than firms with smaller volatility, who are more negatively exposed to HML. This pattern appears in 2006, 2007, 2010 and 2011, but is reversed during the crisis period in 2008-2009, and European debt crisis 2012.

We now focus on two individual stocks' confidence intervals. We take the two firms that have the highest frequency to be respectively classified in the "large group" and the "small

Table 4: 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

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 3: in the long run, the smaller firm is positively exposed and the larger firm is negatively exposed to the SMB systematic risk.

We now examine the dependence of the instrumental beta on the characteristics more explicitly. First, 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,

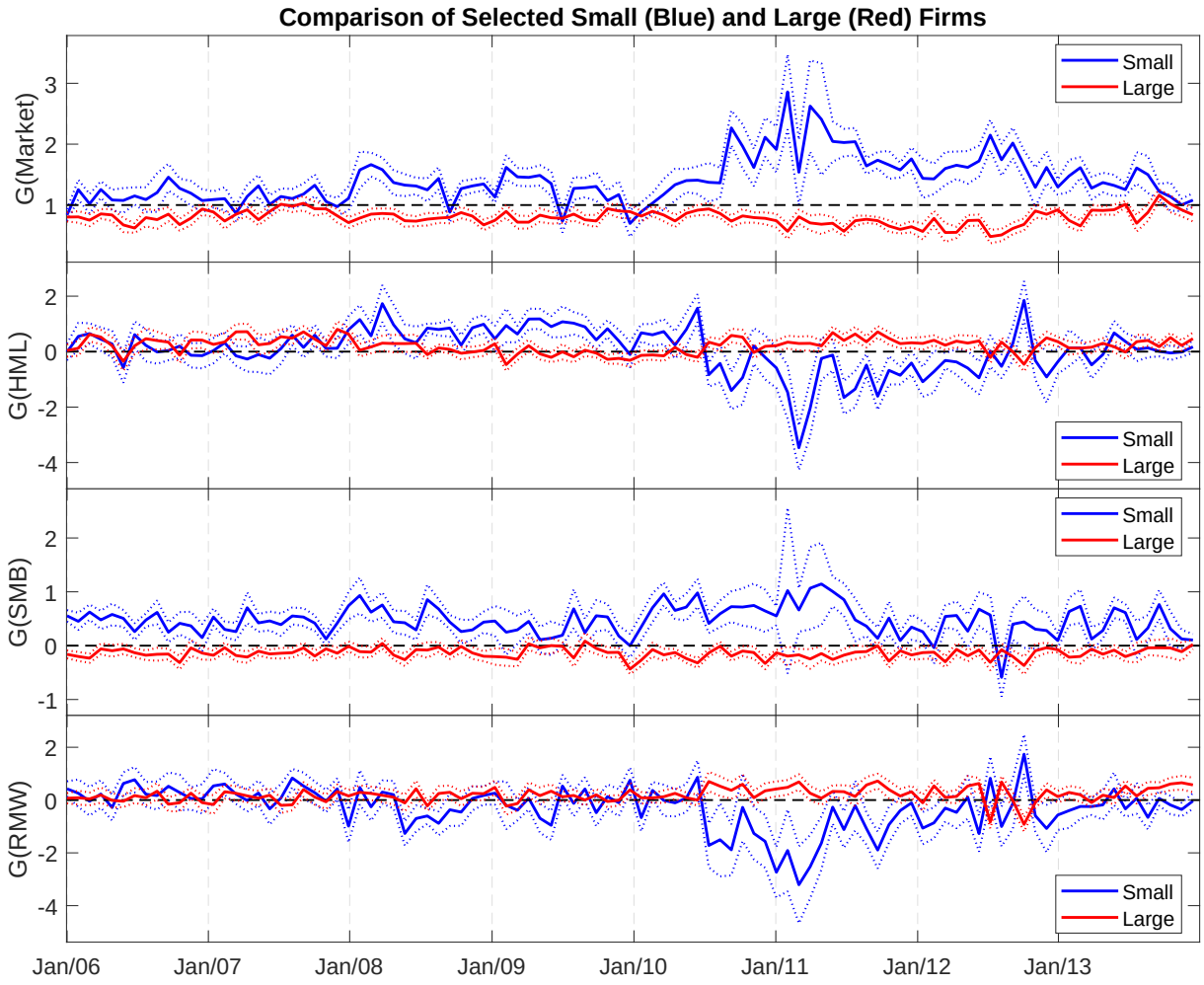

Figure 3: Two individual stocks' confidence intervals: the two firms with the largest and smallest sizes in the dataset.

and calculate the proportion of positive (and negative) significance each year. These results are reported in Table 5. 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 4-7. 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 5: 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

Second, we report the scatter plots of estimated cross-sectional G's versus cross-sectional instruments for different Fama-French factors and on four selected days in Figures 4-7 show. The black solid line is the sieve fitted G function using B-splines with degree 3 (Eilers and Marx, 1996). The plots show several interesting features. First, by comparing the subplots in the same column, one can see that the estimated G function is time-varying. Secondly, the estimated G functions have noticeable nonlinear patterns. As for the specific functions, for the market factor, there is a small downward slope, which is consistent with our finding that small-size firm slightly tends to have a G value larger than 1. A stronger downward slope can be found in the case of the SMB factor. This is also consistent with our finding that large-size firms are more likely to have negative values for the SMB factor's G. Figure 5 gives the result with size replaced by value. The second column indicates that small-value firms tend to have insignificant or negative G values for the HML factor. Figure 6 presents the result for momentum. One can see an obvious downward slope for the HML factors' in July 1, 2008. In fact, this pattern is very persistent during the 2008-2009 crisis. The result shown in Figure 7 is consistent with our finding that large-volatility stocks are very likely to have market G values larger than 1. Lastly, all those figures with non-flat G functions are

consistent with the results in Table 5.

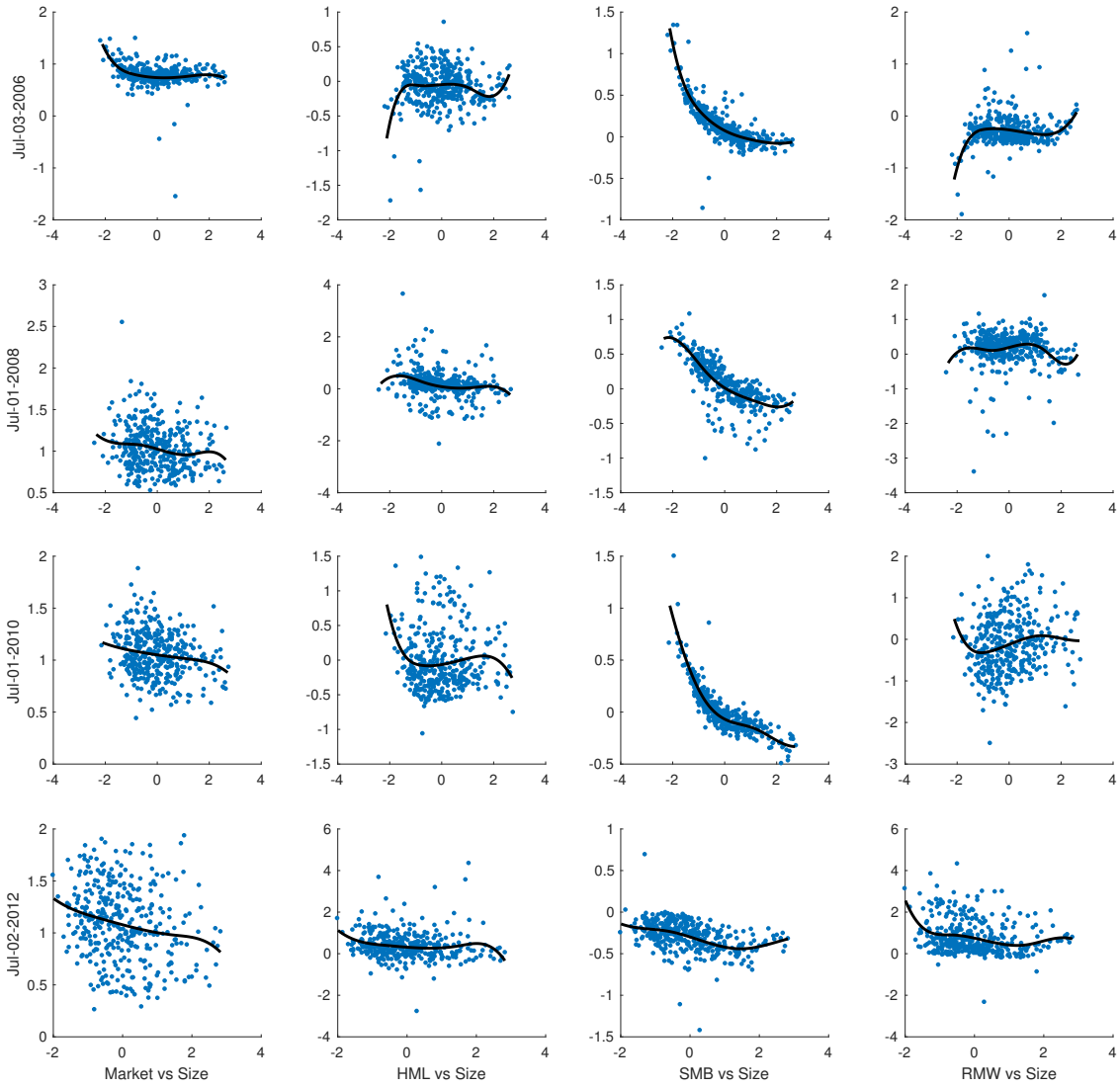

Figure 4: Different factors' G versus size at representative days

10 Conclusion

This paper studies a conditional factor model with a large number of assets for high-frequency data. It has been well known in financial economics that factor betas depend on observed instruments such as firm specific characteristics and macroeconomic variables, and a key object of interest is the effect of instruments on the factor betas. One of the key features of our

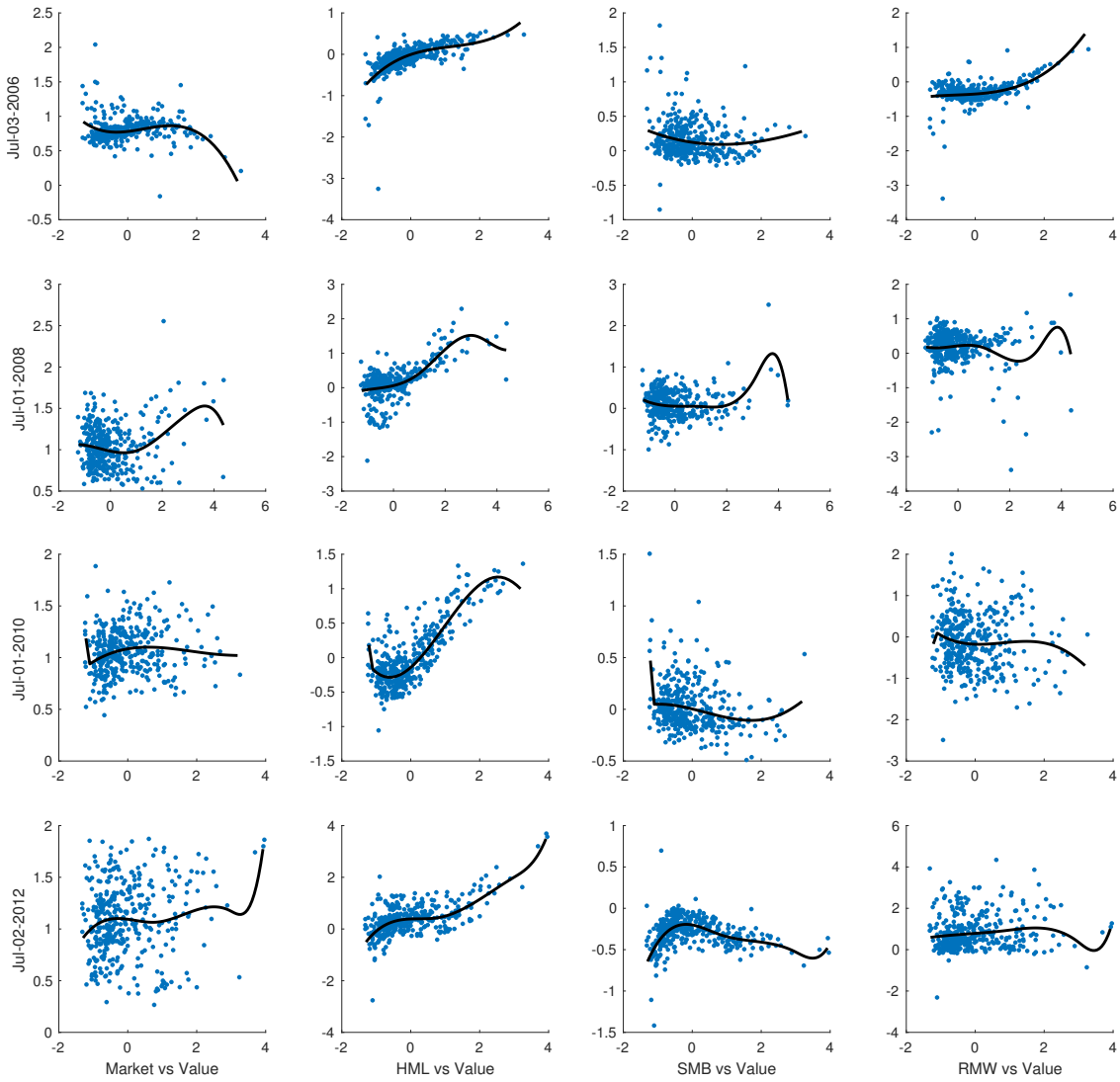

Figure 5: Different factors' G versus value at representative days

model is that we specify the factor betas as functions of time-varying observed instruments that pick up long-run beta fluctuations, plus a remaining (idiosyncratic) component that captures high-frequency movements in beta. It is found that the limiting distribution of the estimated instrument effect on the betas has a discontinuity when the strength of the idiosyncratic beta is near zero, which makes all the existing inference procedures fails to be valid and produce misleading results. 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. Our procedure allows both known and estimated factors. In addition, we employ the estimated

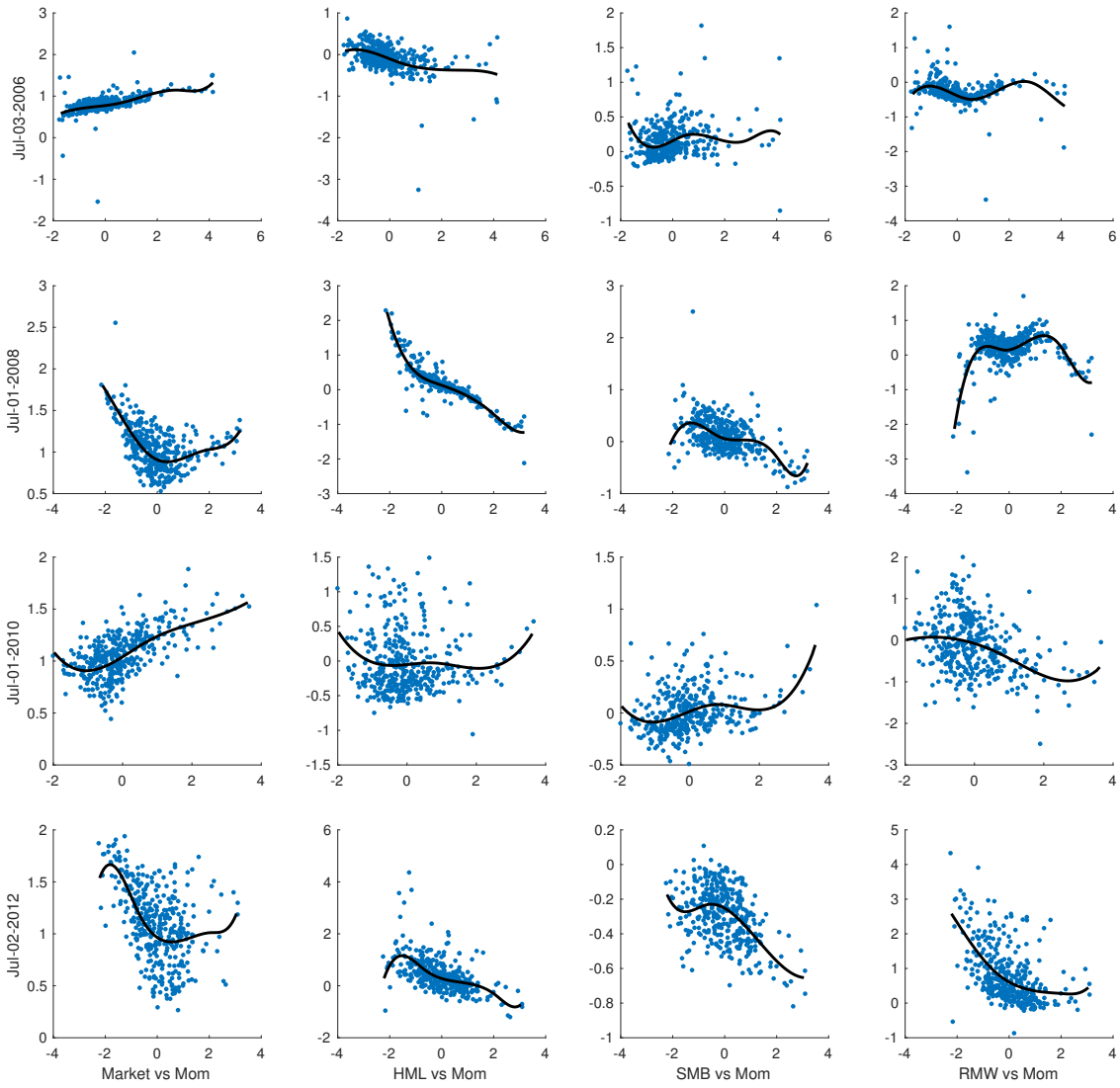

Figure 6: Different factors' G versus value at representative days

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.

Because the instrumental beta specification is a function of macroeconomic and firm variables, it captures long-run movements in beta driven by structural changes in the economic environment and in firm- or industry-specific conditions. In contrast, because the idiosyncratic beta specification is based on high-frequency returns, it picks up short-run fluctuations in beta in periods of high market volatility. The two components capture different aspects of market beta dynamics.

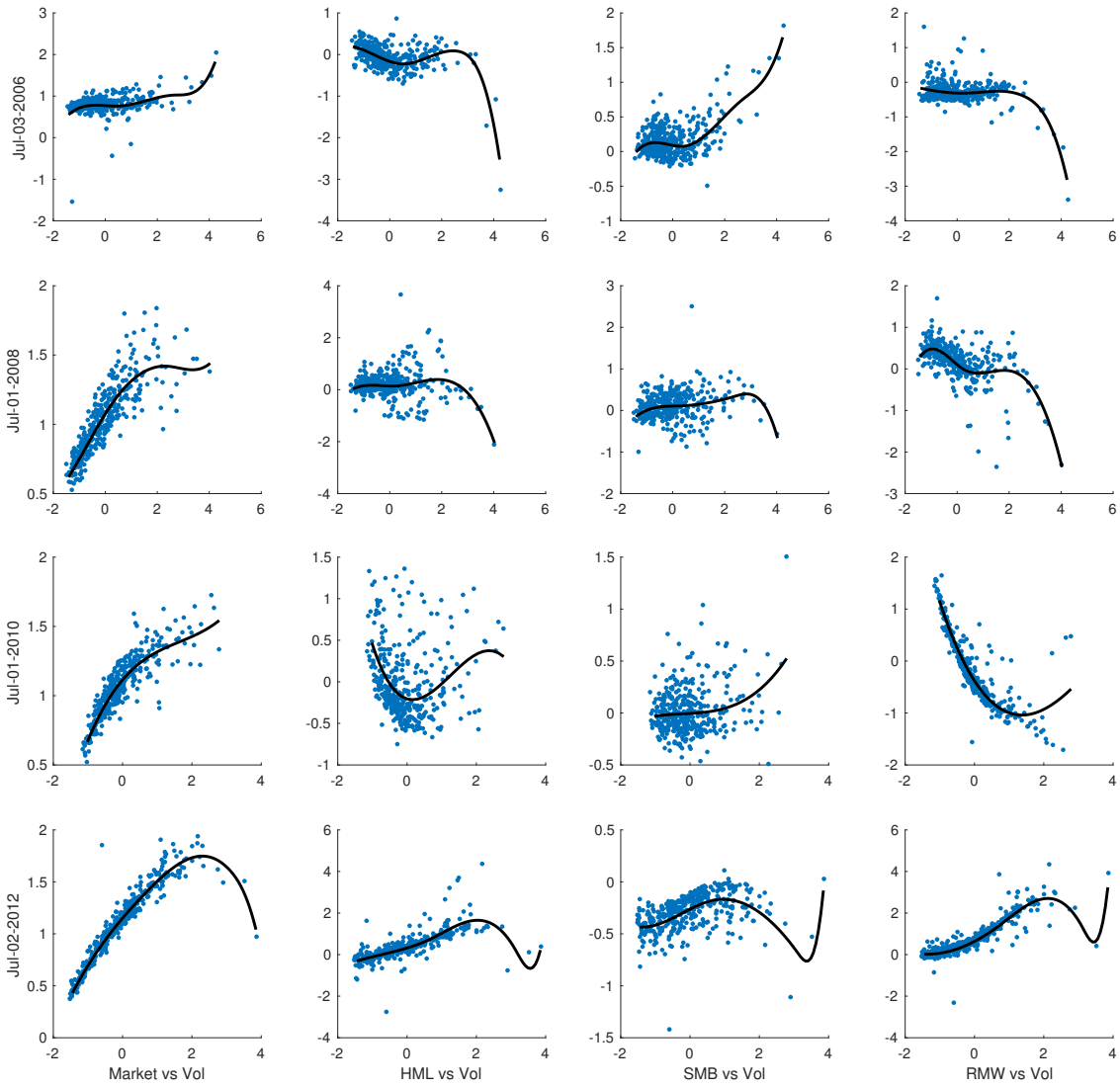

Figure 7: Different factors' G versus value at representative days

References

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

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

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. 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.

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.

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.

Cochrane, J. H. (2005). *Asest Pricing*. Princeton university press.

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.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy* **81** 607–636.

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. 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.

Kelly, B. T., Pruitt, S. and Su, Y. (2017). Instrumented principal component analysis .

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. (2017a). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics* **200** 36–47.

Li, J., Todorov, V. and Tauchen, G. (2017b). 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.

Mykland, P. A. and Zhang, L. (2017). Assessment of uncertainty in high frequency data: The observed asymptotic variance. *Econometrica* **85** 197–231.

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

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *The Annals of Statistics* 763–765.

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

Stock, J. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* **20** 147–162.

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

Zheng, X. and Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Annals of statistics* **39** 3121–3151.