# ESSAYS ON TIME SERIES ECONOMETRICS AND FINANCIAL ECONOMETRICS

WEN XU

LINACRE COLLEGE, UNIVERSITY OF OXFORD

SUBMITTED FOR THE DOCTOR OF PHILOSOPHY DEGREE IN ECONOMICS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF OXFORD

TRINITY TERM 2016

# Essays on Time Series Econometrics and Financial Econometrics

Wen Xu

Linacre College, University of Oxford

Submitted for the Doctor of Philosophy Degree at the Department of Economics, University of Oxford

Trinity Term 2016

**Abstract**

My DPhil thesis includes three essays on time series econometrics and financial econometrics, preceded by a brief introduction.

The first essay proposes a new class of multivariate volatility models utilizing realized measures of asset volatility and covolatility extracted from high-frequency data. Dimension reduction for estimation of large covariance matrices is achieved by imposing a factor structure with time-varying conditional factor loadings. The models are applied to modeling the conditional covariance data of large U.S. financial institutions during the financial crisis, where empirical results show that the new models have both superior in- and out-of-sample properties. We show that the superior performance applies to a wide range of quantities of interest, including volatilities, covolatilities, betas and scenario-based risk measures.

Time-varying volatility is common in macroeconomic data and has been incorporated into macroeconomic models in recent work. The second essay estimates dynamic panel data models with stochastic volatility by maximizing an approximate

likelihood obtained via Rao–Blackwellized particle filters. Monte Carlo studies reveal the good and stable performance of our particle filter-based estimator. When the volatility of volatility is high, or when regressors are absent but stochastic volatility exists, our approach can be better than the maximum likelihood estimator which neglects stochastic volatility and GMM estimators.

In the third essay, we test for the stable factor structure against considerable time variation in the factor loadings in the form of martingales. We obtain the asymptotic distribution of the test statistic by deriving the conditions under which the estimation error of the common factors is asymptotically negligible for the test statistic. Monte Carlo simulations show that the proposed test performs well. We apply the test to a panel of macroeconomic and financial variables in the UK and find the evidence of unstable factor structure during the recent financial crisis.

Acknowledgements

Foremost, I am deeply grateful to my current supervisors, Professor Bent Nielsen and Professor Kevin Sheppard, and my previous supervisor Professor Neil Shephard, for their motivation, insightful advice and continuous encouragement during my DPhil study.

Besides my supervisors, I would like to express my gratitude to Professor James Wolter, Professor Arnaud Doucet and Professor James Stock for their constructive comments.

I have also benefited from useful comments by seminar and conference participants at the University of Oxford, Lancaster University, the International Symposium on Econometric Theory and Applications 2014 (Taipei), the International Panel Data Conference 2014 (Tokyo), the European Meeting of the Econometric Society 2014 (Toulouse), the Panel Data Conference 2015 (Budapest), the conference on Frontiers of Theoretical Econometrics (Konstanz), and the 2016 Asian Meeting of the Econometric Society (Kyoto).

Financial support from the China Scholarship Council, Department of Economics, Oxford-Man Institute and University of Oxford are greatly acknowledged.

Finally, this thesis is dedicated to my parents, Xiuping Sun and Tianhong Xu, for their love and unconditional support.

# Word Count

This thesis has 163 pages in total. A typical page has 265 words. Therefore, the approximate number of words is 43195.

# Contents

# List of Tables

# List of Figures

# 1 Introduction

My DPhil thesis focuses on modeling and inference of large panels of time series data in macroeconomics and finance. The models, the method and the test I propose are based on an enlarged information set (intradaily financial data) or on more realistic specifications (stochastic volatility or structural instability). Two chapters relate to the factor approach, which has been shown to be an effective way to achieve dimension reduction. My research is especially aimed at dealing with the data from the financial crisis of 2007-2009, so two chapters involve applications to actual data of that period.

*The first essay* contributes to the recent literature of combining GARCH(Generalized AutoRegressive Conditional Heteroskedasticity)-type dynamics and realized measures to model multivariate volatility. A new class of multivariate volatility models is proposed to utilize realized measures, which are estimators of daily volatility of asset returns using intra-daily data, such as 1-minute asset prices. The simplest and most common realized measure is realized variance, which estimates the quadratic variation of the intra-daily price path using the sum of the squared high-frequency returns. If one considers the microstructure noise, the deviation from the fundamental value induced by the market specifics such as the bid-ask bounce, one might use alternative solutions such as the realized kernel estimator (Barndorff-Nielsen *et al.*, 2008), which is robust to the microstructure effect and nonsynchronous trading.

Our models are aimed for large portfolios and high-dimensional covariances. Most multivariate models are feasible when the number of assets is small, say 5 or fewer, but not feasible when the number of assets is large because of the number of parameters and the dimension of covariance matrix, due to a curse of dimensionality. The multivariate high frequency model (HEAVY) (Noureldin *et al.*, 2012), for example, suffers from it. That model is directly parametrized in terms of the covariance matrix, so the number of parameters and the dimension of realized measure are large. For example, when modeling the conditional covariance of 50 assets, a 50-dimensional realized covariance is required for the multivariate HEAVY. Our models deal with the curse of dimensionality by parameterizing the components of factor decomposition, so it only requires estimating low-dimensional realized measures regardless of the number of assets. The basic idea is that we achieve dimension reduction by imposing a factor structure on the individual asset return. Our models then assume that the three components of the factor decomposition, which are factor volatility, factor loadings and idiosyncratic volatility, are driven by the corresponding realized measure.

The models specify the dynamics of three groups of variables. The first group includes the factor variance and the conditional mean of realized measure corresponding to the factor. We assume both variables are driven by the realized measure corresponding to the factor; the second group includes the beta and the conditional mean of realized betas. The dynamic is similar to the factor covariance, but they are driven by realized betas. The third group includes the idiosyncratic variance and the conditional mean of realized idiosyncratic variance, which are driven by realized idiosyncratic variance. As a result, our models have two components of equations. The first group concerns the dynamics of conditional covariance of the low-frequency data. They are driven by corresponding realized measures including realized variance or covariance matrix of the factors, realized betas and realized idiosyncratic variance. The second group governs the dynamic of realized measures. We estimate these two components separately, using a two-step maximum likelihood estimator for each component.

We apply the models to a panel of 40 large U.S. financial institutions during the recent financial crisis. Our models dominate other competing models, such as cDCC (Aielli, 2013), beta GARCH (Braun *et al.*, 1995) and multivariate HEAVY, in terms of in-sample fit of the data. The empirical results also show that our models have superior out-of sample properties for a wide range of quantities of interest, including volatilities, covariation, betas and scenario-based risk measures such as systemic risk like marginal expected shortfall. The model's performance is particularly strong at short forecast horizons.

A version of this essay is resubmitted to *Journal of Financial Econometrics* jointly with Professor Kevin Sheppard.

*In the second essay*, we apply the particle filtering technique to estimate the dynamic panel data models with stochastic volatility. Time-varying volatility is common in macroeconomic data. Several papers (e.g. Kim & Nelson, 1999; McConnell & Perez-Quiros, 2000; Blanchard & Simon, 2001) have studied the moderation of volatility in the U.S. real GDP growth from the mid-1980s. Homescedasticity is traditionally assumed in macroeconometric models, but recently time-varying volatility has been incorporated into various models like DSGE models (Fernández-Villaverde & Rubio-Ramírez, 2007) and vector autoregressive models (Koop & Korobilis, 2010). Hamilton (2010) emphasized the importance of time-varying volatility by arguing that statistical efficiency can be improved by incorporating appropriate features of time-varying volatility into estimation of the conditional mean, which is often the direct object of interest.

Dynamic panel data models have been used to analyze macroeconomic data in which the time series dimension is far larger than that of microeconomic panels. One can incorporate time-varying volatility into dynamic panel data models when studying the impact of time variations in the volatility of business conditions or policy rules (so called uncertainty shocks) on employment, output and other economic outcomes (Bloom, 2009). The proposal is to estimate a dynamic panel data model with time-varying volatility in the innovations, which might be parameterized through stochastic volatility (Kim *et al.*, 1998; Fernández-Villaverde & Rubio-Ramírez, 2007).

In the panel models with stochastic volatility, there is no closed-form expression for the likelihood, so one needs to have recourse to simulation-based methods. Particle filters, also known as sequential Monte Carlo methods, are simulation-based techniques for state space models (Creal, 2012). In non-Gaussian and nonlinear state space models, the density of state variables often does not have the closed form expression. Particle filters approximate the density by a discrere distribution made of weighted simulation. There are many implementations of particle filters. Because of the special structure of the state space form of dynamic panel data models, we use a specific particle filter called the Rao–Blackwellized particle filter (Liu & Chen, 1998). The structure is special in that it is linear and Gaussian conditional on a subset of state variables. The idea of Rao–Blackwellized particle filter is to simulate those subset of state variables and the Kalman filter is then used for the remaining state variable given the simulations.

Specifically, we estimate the dynamic panel data models with stochastic volatility in a macroeconomic context where both cross section dimension and time series dimension can be large. We transform the first difference of the models to a state space form and then apply the Rao–Blackwellized particle filter to estimate the likelihood. Although this estimated likelihood is unbiased under some regularity conditions, direct maximization for parameter estimation would fail due to the discontinuity induced from the generalized inverse operation at the resampling stage. We use the approximation method in Olsson & Rydén (2008) to overcome this obstacle. The model parameters are estimated by maximizing the approximated likelihood. Monte Carlo studies show that the particle-filter based estimator has good and stable performance. When volatility of volatility is high or when regressors are absent, it can be better than the maximum likelihood estimator which neglects stochastic volatility (Hsiao *et al.*, 2002) and GMM estimators (e.g. Arellano & Bond 1991; Blundell & Bond 1998).

A version of this essay is accepted for publication in *Econometrics*.

*In the third essay*, we study the high-dimensional factor models and focus on the test for temporal stability in factor loadings. Factor models have been the main tool to analyze big datasets in macroeconomics for more than 15 years (Bai & Ng, 2008b; Stock & Watson, 2011). It has been very useful in macroeconomic applications as a small number of common factors can account for a large fraction of the variation in a great number of observable series. In other words, a small number of factors can summarize the comovements of a large number of series and drive the dynamic behavior of all those series. As suggested by empirical evidence, the factor models fit macroeconomic data in many countries. An important application of factor models is to make forecasts for an individual variable through the regression of that variable on the estimated factors and other predictors, which is called the factor-augmented regression.

As the number of factors and factors themselves are unobservable, a large body of research has been devoted to determining the number of latent factors and studying the asymptotic properties of the principal components estimator of the factors. Under some assumptions, the estimators are consistent and the principal components estimator is asymptotically normal. Traditionally, the factor loadings are assumed constant over time. The assumption of stable factor loadings is however not supported by empirical evidence (Stock & Watson, 2009). In theory, the estimators of the number of the factors as well as the principal components estimator of the factors can still be consistent provided that the magnitude of the temporal instability in factor loadings is below the upper bound given in Bates *et al.* (2013). If the time variation is stronger, however, the inference would be misleading and the forecasts based on the estimated factors can be inaccurate. For this reason, one needs to test whether there is considerable instability in factor loadings before every serious use of the factor models. There are generally two types of instability: discrete breaks and continuous variations in the form of martingales such as random walk. The literature on testing for stable factor models has focused on structural breaks so far. One can test against multiple breaks in factor loadings of one specific variable (Breitung & Eickmeier, 2011), or common breaks in factor loadings of all variables (Chen *et al.*, 2014; Cheng *et al.*, 2014; Corradi & Swanson, 2014; Han & Inoue, 2015).

Although the tests for structural breaks also have power to detect continuous variations in factor loadings, we aim to develop the test directly against martingale-type alternatives for the following four reasons. First, the test against martingales is more general, as the model with multiple breaks is just a special example of the models with martingale-type variations. Second, the continuous variation is a more serious issue than breaks that would affect the forecasts based on factor-augmented regressions. It is well known that a factor model with breaks can be equivalently represented by a stable factor model with a larger number of factors. Therefore, structural breaks have no theoretical influence on the forecasts based on factor-augmented regressions, if the number of factors is not ex ante imposed but is instead estimated using the methods in the literature. On the other hand, the factor models with a general martingale-type instability in loadings have no equivalent representation of stable models, and the number of factors in those models can not be consistently estimated. As a result, continuous variations in the form of martingales would theoretically affect the forecasts. Third, Del Negro & Otrok (2008) developed a factor model with random walk in loadings. If there is indeed strong instability of this type, their model might be a better candidate than stable models for the purpose of forecasting. Finally, Hansen (1992) argued that specifying the parameter process as a martingale is more appropriate if one is interested in testing whether the model captures a stable relationship, because martingales capture the notion of unstable models that gradually change over time.

To that end, we adapt the Lagrange Multiplier test of Nyblom (1989) for the high-dimensional factor models. Nyblom (1989) addressed the general issue of testing for parameter constancy against the alternative of martingales in a likelihood framework. The key challenge of extending this framework to factor models is that the accumulation of the estimation error of common factors would affect the conventional distributions of test statistics. One therefore needs to ensure that the estimation error of factors is asymptotically negligible for the test statistic. We obtain the asymptotic distributions of the proposed statistic by deriving the conditions on the relative rate of the number of cross sections and the number of time periods under which insignificant estimation error is guaranteed. Generally it requires the number of cross sections is large enough compared to the number of time periods after the change point from which the factor loadings can begin to evolve as random walk. In the benchmark model we study, common factors are stationary. We also make an extension to cover the model with integrated factors.

We study the finite sample performance of the test statistic by Monte Carlo simulations. When common factors are stationary, the test seems oversized, possibly due to the finite sample bias of the estimator of the number of factors. It turns out that the rejection frequencies become closer to the nominal size when the factors are integrated. The test tends to be consistent in that the power increases with the sample size. In the case of integrated factors, the test suffers from the issue of nonmonotonic power. The results also show that our test is more powerful than the tests designed for structural breaks to detect martingale-type instability.

The factor models are often used for forecasting macroeconomic variables such as the GDP growth. The forecasting performance varies with the target country however. The models work well for the US data or European data, but the models seem less effective for UK data especially during and after the recent finanial crisis. We suspected the temporal instability is a potential source of the poor performance. To confirm our suspicion, we apply the proposed test statistic to a panel of UK variables and find the evidence of unstable factor structure during the recent crisis period.

# 2 Factor High-Frequency Based Volatility (HEAVY) Models

## 2.1 Introduction

Conditional covariances are key inputs in risk management and portfolio optimization. Traditionally, multivariate GARCH models based on daily data are used to capture the dynamics of the second-order moments of asset returns (see Bauwens *et al.*, 2006 for a survey). While most multivariate volatility models are feasible when the number of assets is small – 5 or fewer – only a small subset remain feasible when applied to large, empirically realistic portfolios. There are a number of difficulties in high-dimension covariance modeling, including the computational effort required to invert large conditional covariance matrices when evaluating the likelihood and the high-dimensionality of the parameter space. Recent contributions to the literature have attempted to side-step these issues by using alternative estimators or carefully designed models. Engle *et al.* (2014) construct the composite likelihood by summing up the log-likelihoods of pairs of assets in order to avoid the inversion of high-dimensional matrices. The Dynamic Equicorrelation model proposed in Engle & Kelly (2012) leads to a simple analytic form of the inverse of the conditional covariance by assuming that the time-varying correlations are identical across all pairs of assets. A third, and older, approach to achieve dimension reduction is to exploit the strong factor structure of asset returns when modeling conditional covariance. Early examples include Engle *et al.* (1990), who introduced the factor-ARCH model to price Treasury bills.

Recently, intra-daily estimators of volatility – collectively known as realized measures – have been used to improve volatility models. Realized measures incorporate information from asset price paths to improve the measurement of volatility over a fixed horizon, typically one day. The simplest and most common realized measure is realized variance, which estimates the quadratic variation of the intra-daily log-price process using the sum of the squared high-frequency returns. When the prices follows a diffusion process with stochastic volatility and can be directly observed without error, realized variance converges to daily integrated variance of the underlying volatility process (Andersen *et al.*, 2001a; Barndorff-Nielsen & Shephard, 2002). However, microstructure noise such as bid-ask bounce is ubiquitous in high frequency data which limits sampling frequency. Several alternative solutions exist in the literature to control microstructure effect, including two- and multi-scale realized volatility (Aït-Sahalia *et al.* 2005; Zhang 2006), realized kernel estimators (Barndorff-Nielsen *et al.*, 2008), and the pre-averaging approach (Jacod *et al.*, 2009).

The multivariate extension of realized variance, known as realized covariance, was first introduced to econometrics in Andersen *et al.* (2001b) and the asymptotic theory was first studied in Barndorff-Nielsen & Shephard (2004). In the absence of market microstructure noise, and when prices are synchronously observed, realized covariance estimates the quadratic covariation of prices. Estimators that are robust to microstructure noise and non-synchronous trading include multivariate realized kernels (Barndorff-Nielsen *et al.*, 2011) and pre-averaging estimators (Christensen *et al.*, 2010) and realized QMLE (Shephard & Xiu, 2013). Like their low-frequency counterparts, multivariate realized measures can be transformed to estimate other quantities, such as realized correlation or realized factor loadings (also known as realized beta). Bollerslev & Zhang (2003) employ realized factor loadings constructed in the Fama–French three-factor model to improve asset pricing predictions. Barndorff-Nielsen & Shephard (2004) derive the asymptotic distribution of realized betas. Bandi & Russell (2005) study the finite-sample properties of realized betas in the presence of market microstructure noise. Andersen *et al.* (2006) find that realized betas are less persistent than realized variances and covariances and suggest modeling them as short-memory processes, and Patton & Verardo (2012) study the effect of earnings announcement on realized betas.

In this chapter, we introduce Factor HEAVY models, a new class of multivariate volatility models exploiting high frequency data and utilizing a factor approach to facilitate estimation in empirically relevant scenarios. We exploit a factor decomposition of asset prices to build the models which are feasible in high-dimensions and estimable with an imbalanced panel. Our models resemble $\beta$-GARCH, which models the factor variance, conditional $\beta$ and idiosyncratic variance each with a GARCH-type evolution. Braun *et al.* (1995) model returns on individual assets as related through a common factor, $r_{i,t} = \beta_{i,t} r_{f,t} + \varepsilon_{i,t}$, so that the covariance dynamics in the $\beta$-GARCH model are given by

$$\begin{aligned}
\sigma^2_{f,t} &= \theta_0 + \theta_1 r^2_{f,t-1} + \theta_2 \sigma^2_{f,t-1} \\
\beta_{i,t} &= \delta_{i,0} + \delta_{i,1} \frac{r_{f,t-1} r_{i,t-1}}{\sigma^2_{f,t-1}} + \delta_{i,2} \beta_{i,t-1} \\
\sigma^2_{i,t} &= \alpha_{i,0} + \alpha_{i,1} \left( r_{i,t-1} - \beta_{i,t-1} r_{f,t-1} \right)^2 + \alpha_{i,2} \sigma^2_{i,t-1},
\end{aligned} \tag{1}$$

where $\sigma^2_{f,t}$ is the variance of the factor, or in a conditional CAPM, the market return. This specification uses natural proxies of the left-hand-side variables to act as shocks – squared returns of the factor for the factor variance, standardized cross-products for the $\beta$ dynamics and squares of the innovation for the idiosyncratic variance.

We contribute to a growing literature that combines multivariate GARCH-type dynamics and realized measures, and explicitly model the closing return, not just the intra-daily return.[1] In multivariate HEAVY models (Noureldin *et al.*, 2012), the conditional covariance is modeled as a smoothed function of recent lags of realized covariance matrix. Jin & Maheu (2013) model daily returns using a Wishart-autoregressive-like structure and propose a Bayesian estimation method. Hansen *et al.* (2014) propose the realized $\beta$ GARCH model in which volatility and covolatility are separated modeled. In our models, daily returns on individual assets are driven by common factors with time-varying factor loadings.[2] The dynamics of factor volatility, conditional factor loadings and idiosyncratic volatility all follow the HEAVY structure where realized measures drive the dynamics of daily covariance (Shephard & Sheppard, 2010). Factor HEAVY models have two advantages over multivariate HEAVY models. First, multivariate HEAVY models are directly parametrized on variances and covariance and so specify common dynamics for all second moments of asset returns. Second, multivariate HEAVY models suffer from the curse of dimensionality – not only in terms of the number of parameters in the models, but also in the dimension of the realized measure required to drive the dynamics. For example, when modeling the conditional covariance of 50 assets, a 50-dimensional realized covariance is required each

[1] Initial models that included realized measures focused primarily on realized measure only modeling (Halbleib-Chiriac & Voev (2014), Golosnoy *et al.* (2012)). While the persistence of realized measures is an interesting topic, it is not appropriate for most applications in portfolio allocation or risk-management.

[2] The time-variation in conditional $\beta$s has been debated in the literature for the last two decades. Braun *et al.* (1995) use bivariate EGARCH models to find weak evidence of time-varying conditional $\beta$s. Ferson & Harvey (1993), Bali & Engle (2010), and Hansen *et al.* (2014) find significant time-series variation in the conditional $\beta$s. Bali *et al.* (2012) show the substantial time-varying conditional $\beta$s in the cross-section of daily stock returns.

day. Our models only require estimating low-dimensional realized measures irrespective of the number of assets in the models, and so can easily scale to empirically relevant dimensions. In the empirical analysis, we show that Factor HEAVY dominates other competing models in terms of in-sample performance. We also compare the out-of-sample ability of Factor HEAVY and the cDCC GARCH models (Aielli, 2013) when forecasting volatility and covolatility, $\beta$s and marginal expected shortfall (MES). The results show that Factor HEAVY outperforms cDCC models, and that the gains are particularly substantial in short term forecasting. The Factor HEAVY has greater advantage over the long memory Riskmetrics model (Zumbach, 2007) than over cDCC models when forecasting volatility and covolatility. This superior performance at short horizons is particularly useful from a regulatory point-of-view since accurate and timely detection of changes in the covariance structure of returns is required when considering interventions.

The remainder of the chapter is structured as follows. Section 2.2 introduces the Factor HEAVY models and discusses their properties. We initially focus the exposition on the 1-factor version of the models, although the extension to more factors is straight-forward. Section 2.3 discusses estimation and asymptotic properties. Section 2.4 describes the data used in the chapter, presents the result of empirical analysis including parameter estimation, in-sample performance, out-of-sample forecasting, the extension to multiple-factor models and the robustness analysis. Section 2.5 concludes the chapter. The appendix is in Section 2.6.

## 2.2 Factor HEAVY Models

### 2.2.1 Notation and Model Setup

To facilitate the exposition of the model, the initial focus is on a 1-factor specification. The full $K$-factor specification is presented in section 2.2.7. Let $r_t = (r_{f,t}, r_{1,t}, r_{2,t}, ..., r_{N,t})'$ denote a $N+1$ by 1 vector of low-frequency, typically daily, returns. The first return is a pervasive factor and the remaining $N$ are returns on individual assets assumed to be related through the factors.[3] We denote the information set formed by the history of low-frequency returns with $\mathscr{F}_t^{LF}$, the natural filtration containing all past low-frequency returns. In the standard multivariate ARCH literature, returns are typically assumed to be conditionally normally distributed so that

$$r_t|\mathscr{F}_{t-1}^{LF} \sim N(0, \Omega_t).$$

Our interest is in modeling the conditional covariance of low-frequency returns using high-frequency derived realized measures, so we augment the information set with a realized measure that estimates the quadratic covariation of the factor and individual assets, and denote the time $t$ value of this $N+1$ by $N+1$ matrix-valued random variable $RM_t$. The realized measure could be a realized covariance or a more sophisticated noise-robust measure such as a realized kernel (Barndorff-Nielsen *et al.*, 2011), pre-averaged realized variance (Christensen *et al.*, 2010) or realized QMLE covariance (Shephard & Xiu, 2013).[4] We use

[3] While mixing intra-daily information and daily information is the natural application of the Factor HEAVY models, this class of models is also useful for building models that mix daily (high frequency) and weekly or monthly (low frequency) data.

[4] While the choice of realized measure does not affect the description of the model, the steps required to estimate the model may differ depending on the estimator used. These issues are discussed in section 2.3.

$\mathscr{F}_t^{HF}$ to denote the filtration that contains both the realized measures and low-frequency returns, so that $\mathscr{F}_t^{LF} \subset \mathscr{F}_t^{HF}$.

We assume that returns, conditionally on the high-frequency information set, are normally distributed

$$r_t|\mathscr{F}_{t-1}^{HF} \sim N(0, \Sigma_t).$$

However, since the model now contains both high- and low-frequency data, it is necessary to specify a model for the process generating the realized measures. We assume that the realized measure is conditionally distributed as a Wishart with $\nu$ degrees of freedom, so that

$$RM_t|\mathscr{F}_{t-1}^{HF} \sim W_{N+1}(\nu, {}^{M_t}/_{\nu}),$$

where $M_t$ is a positive definite matrix so that $RM_t = M_t^{1/2}\Xi_t\left(M_t^{1/2}\right)'$ where $M_t^{1/2}$ is a matrix square-root of $M_t$ and $\Xi_t|\mathscr{F}_{t-1}^{HF} \sim W_{N+1}(\nu, \frac{I_{N+1}}{\nu})$. We use a partitioning of the realized measures so that

$$RM_t = \begin{bmatrix} RM_{f,t} & RM_{f1,t} & \dots & RM_{fN,t} \\ RM_{1f,t} & RM_{11,t} & \dots & RM_{1N,t} \\ \vdots & \vdots & \ddots & \vdots \\ RM_{Nf,t} & RM_{1N,t} & \dots & RM_{NN,t} \end{bmatrix}, \tag{2}$$

where $RM_{f,t}$ is a scalar measuring the quadratic variation of the factor, $RM_{ii,t}$ is a scalar realized measure estimating the quadratic variation of the individual asset and $RM_{if,t}$ is a scalar measuring the quadratic covariation between the asset and the factors.

We propose to model the conditional covariance of returns, $\Sigma_t$, as well as the conditional expectation of the realized measure, $M_t$, using a factor structure. The conditional covariance of returns and the conditional expectation of the realized measures can be written as

$$\Sigma_t = \begin{bmatrix} \sigma^2_{f,t} & \beta_{1,t}\sigma^2_{f,t} & \cdots & \beta_{N,t}\sigma^2_{f,t} \\ \beta_{1,t}\sigma^2_{f,t} & \beta^2_{1,t}\sigma^2_{f,t}+\sigma^2_{1,t} & \cdots & \beta_{1,t}\beta_{N,t}\sigma^2_{f,t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{N,t}\sigma^2_{f,t} & \beta_{N,t}\beta_{1,t}\sigma^2_{f,t} & \cdots & \beta^2_{N,t}\sigma^2_{f,t}+\sigma^2_{N,t} \end{bmatrix}, \tag{3}$$

and

$$M_t = \begin{bmatrix} \mu_{f,t} & \lambda_{1,t}\mu_{f,t} & \cdots & \lambda_{N,t}\mu_{f,t} \\ \lambda_{1,t}\mu_{f,t} & \lambda^2_{1,t}\mu_{f,t}+\mu_{1,t} & \cdots & \lambda_{1,t}\lambda_{N,t}\mu_{f,t} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N,t}\mu_{f,t} & \lambda_{N,t}\lambda_{1,t}\mu_{f,t} & \cdots & \lambda^2_{N,t}\mu_{f,t}+\mu_{N,t} \end{bmatrix}. \tag{4}$$

The model then requires specifications for the dynamics of the factor variance, $\sigma^2_{f,t}$ (and the corresponding value for the realized measure $\mu_{f,t}$), the factor loadings $\beta_{i,t}$ $(\lambda_{i,t})$ and the idiosyncratic variances $\sigma^2_{i,t}$ $(\mu_{i,t})$. Throughout the chapter, we use the notation $\mathrm{E}_{t-1}(.) = \mathrm{E}(. \mid \mathscr{F}^{\mathrm{HF}}_{t-1})$ to denote the expectation conditional on the high frequency information set. The Wishart distribution is in the linear exponential family and is the natural candidate for positive definite matrix-valued shocks. Moreover, the transformations required when building factor-based models are known in closed form and lead to simple expressions for conditional expectations of the quantities of interest.

**Proposition 1.** Define the realized $\beta$ and realized idiosyncratic variance as $R\beta_{i,t} = \frac{RM_{if,t}}{RM_{f,t}}$ and $RIV_{i,t} = RM_{ii,t} - (R\beta_{i,t})^2 RM_{f,t}$, respectively. Under the assumption

$RM_t|\mathscr{F}_{t-1}^{\mathrm{HF}} \sim W_{N+1}(\nu, \frac{M_t}{\nu})$, the conditional expectation of realized $\beta$, $\mathrm{E}_{t-1}[R\beta_{i,t}] = \lambda_{i,t}$ and that of realized idiosyncratic variance $\mathrm{E}_{t-1}[RIV_{i,t}] = \frac{\nu-1}{\nu}\mu_{i,t}$.

#### 2.2.2 Factor Variance Dynamics

The factor dynamics are assumed to follow a HEAVY structure. Suppose there is a single factor, in which case the factor returns are assumed to be conditionally normal with mean zero:

$$r_{f,t} = \sigma_{f,t}\xi_{f,t},$$

where the conditional variance of the factor return is measurable with respect to $\mathscr{F}_{t-1}^{HF}$ and $\mathrm{E}_{t-1}\left[\xi_{f,t}^2\right] = 1$ . The conditional variance of the factor is driven by the realized measure corresponding to the quadratic variation of the factor, and so

$$\sigma_{f,t}^2 = \theta_0 + \theta_1 RM_{f,t-1} + \theta_2\sigma_{f,t-1}^2. \tag{5}$$

This is not a complete model and only allows for 1-step forecasts, and so we complete the model by specifying a model for the realized measure. The dynamics of the conditional mean of the realized measure follow a similar process to the factor conditional variance,

$$\mu_{f,t} = \theta_0^M + \theta_1^M RM_{f,t-1} + \theta_2^M\mu_{f,t-1}. \tag{6}$$

These two equations correspond to the HEAVY model (Shephard & Sheppard, 2010).

### 2.2.3 Factor Loading Dynamics

Returns on individual assets are related through exposure to a common factor through a time varying loading. The factor loading of asset $i$ is

$$\beta_{i,t} = \frac{\sigma_{if,t}}{\sigma_{f,t}^2} = \frac{\mathrm{Cov}\left[r_{f,t}, r_{i,t}\right]}{\mathrm{V}\left[r_{f,t}\right]}.$$

Factor loadings are also driven by realized measures, where the natural analogue of the $\beta$ commonly referred to as the realized $\beta$ is used. Without microstructure noise, Barndorff-Nielsen & Shephard (2004) show that when the realized measure is realized covariance, then the realized $\beta$ is a consistent estimator of the ratio between integrated equity covariance with the factor and integrated variance of the factor return. The dynamic factor loadings evolve as

$$\beta_{i,t} = \delta_{i,0} + \delta_{i,1} R\beta_{i,t-1} + \delta_{i,2}\beta_{i,t-1}, \tag{7}$$

and we complete the model by specifying dynamics for the realized $\beta$:

$$\lambda_{i,t} = \delta_{i,0}^M + \delta_{i,1}^M R\beta_{i,t-1} + \delta_{i,2}^M \lambda_{i,t-1}. \tag{8}$$

There is no observation equation for the realized $\beta$ and its measurement occurs jointly with the idiosyncratic variances.

### 2.2.4 Idiosyncratic Dynamics

The final component of the model is the idiosyncratic variance, defined as $\mathrm{E}_{t-1}\left[\left(r_{i,t} - \beta_{i,t} r_{f,t}\right)^2\right]$. Since the factor and the individual asset return are assumed to be conditionally normally, we can define the idiosyncratic shock as

$$\varepsilon_{i,t} = r_{i,t} - \beta_{i,t} r_{f,t},$$

where $\varepsilon_{i,t} = \sigma_{i,t}\xi_{i,t}$ and $\sigma^2_{i,t}$ is the conditional variance of $r_{i,t}$ given the factor return. By construction, the $N$ by 1 vector $\varepsilon_t = (\varepsilon_{1,t} \ldots, \varepsilon_{N,t})'$ is contemporaneously uncorrelated, and we further assume that there are no volatility spill-overs between assets. We use the realized idiosyncratic variance to drive the dynamics of the idiosyncratic variance. We assume a univariate HEAVY-like structure for the idiosyncratic variance, so that

$$\sigma^2_{i,t} = \alpha_{i,0} + \alpha_{i,1} RIV_{i,t-1} + \alpha_{i,2}\sigma^2_{i,t-1}. \tag{9}$$

The model is completed by specifying the evolution of the realized idiosyncratic volatility as

$$\mu_{i,t} = \alpha^M_{i,0} + \alpha^M_{i,1} RIV_{i,t-1} + \alpha^M_{i,2}\mu_{i,t-1}. \tag{10}$$

#### 2.2.5 Full Specifications

The model specifications contain two distinct components. The first component, which determines the dynamics of the conditional covariance of the low-frequency data, is collectively referred to as the HEAVY-P equations,

$$
\begin{aligned}
\sigma^2_{f,t} &= \theta_0 + \theta_1 RM_{f,t-1} + \theta_2 \sigma^2_{f,t-1} \\
\beta_{i,t} &= \delta_{i,0} + \delta_{i,1} R\beta_{i,t-1} + \delta_{i,2}\beta_{i,t-1} \\
\sigma^2_{i,t} &= \alpha_{i,0} + \alpha_{i,1} RIV_{i,t} + \alpha_{i,2}\sigma^2_{i,t-1},
\end{aligned} \tag{11}
$$

where the final two equations are repeated $N$ times. We refer to the second component, which governs the dynamics of the realized measure, as the HEAVY-M equations,

$$
\begin{aligned}
\mu_{f,t} &= \theta^M_0 + \theta^M_1 RM_{f,t-1} + \theta^M_2 \mu_{f,t-1} \\
\lambda_{i,t} &= \delta^M_{i,0} + \delta^M_{i,1} R\beta_{i,t-1} + \delta^M_{i,2}\lambda_{i,t-1} \\
\mu_{i,t} &= \alpha^M_{i,0} + \alpha^M_{i,1} RIV_{i,t-1} + \alpha^M_{i,2}\mu_{i,t-1},
\end{aligned} \tag{12}
$$

where the final two equations are also repeated for each individual asset.

#### 2.2.6 Forecasting

We focus on forecasting conditional factor variances, $\beta$s and idiosyncratic variances. One-step forecasts are directly given in the HEAVY-P equations. To compute multi-step forecasts, we utilize the recursive structure of the HEAVY-M equations.

**Proposition 2.** The $s$-step forecasts in the Factor HEAVY model is given by

$$
\begin{aligned}
\sigma^2_{f,t+s|t} &= \mathrm{E}_t(\sigma^2_{f,t+s}) = \theta_2^{s-1}\sigma^2_{f,t+1} + \theta_0 \frac{1-\theta_2^{s-1}}{1-\theta_2} \\
&\quad + \theta_1 \sum_{i=1}^{s-1} \theta_2^{i-1} (\theta^M_0 \frac{1-(\theta^M_1+\theta^M_2)^{s-1-i}}{1-(\theta^M_1+\theta^M_2)} + (\theta^M_1+\theta^M_2)^{s-1-i}\mu_{f,t+1})
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
\beta_{i,t+s|t} &= \mathrm{E}_t(\beta_{i,t+s}) = \delta_{i,2}^{s-1}\beta_{i,t+1} + \delta_{i,0}\frac{1-\delta_{i,2}^{s-1}}{1-\delta_{i,2}} \\
&\quad + \delta_{i,1}\sum_{i=1}^{s-1}\delta_{i,2}^{i-1}(\delta_{i,0}^{M}\frac{1-(\delta_{i,1}^{M}+\delta_{i,2}^{M})^{s-1-i}}{1-(\delta_{i,1}^{M}+\delta_{i,2}^{M})} + (\delta_{i,1}^{M}+\delta_{i,2}^{M})^{s-1-i}\lambda_{i,t+1})
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{i,t+s|t}^{2} &= \mathrm{E}_t(\sigma_{i,t+s}^{2}) = \alpha_{i,2}^{s-1}\sigma_{i,t+1}^{2} + \alpha_{i,0}\frac{1-\alpha_{i,2}^{s-1}}{1-\alpha_{i,2}} \\
&\quad + \alpha_{i,1}\sum_{i=1}^{s-1}\alpha_{i,2}^{i-1}(\frac{\nu-1}{\nu}\alpha_{i,0}^{M}\frac{1-(\frac{\nu-1}{\nu}\alpha_{i,1}^{M}+\alpha_{i,2}^{M})^{s-1-i}}{1-(\frac{\nu-1}{\nu}\alpha_{i,1}^{M}+\alpha_{i,2}^{M})} + (\frac{\nu-1}{\nu}\alpha_{i,1}^{M}+\alpha_{i,2}^{M})^{s-1-i}\mu_{i,t+1}).
\end{aligned}
$$

The proof is presented in the appendix. While these are the forecasts for the components of the conditional covariance, they do not lead to direct forecasts of the conditional covariance. An unbiased estimate could be computed using simulation, although we find that this error is sufficiently small so that simulation is not warranted.

#### 2.2.7 Multiple Factors

The 1-factor model can be directly extended to include multiple factors. In the multi-factor HEAVY model, returns on individual assets are determined by $K$ factors and an idiosyncratic shock,

$$
r_{i,t} = \sum_{k=1}^{K}\beta_{i,k,t}r_{fk,t} + \varepsilon_{i,t} = \sum_{k=1}^{K}\beta_{i,k,t}\sigma_{fk,t}e_{fk,t} + \sigma_{i,t}e_{i,t},
$$

in which $r_{fk,t}$ and $\sigma_{fk,t}$ denote the daily return on the $k^{\text{th}}$ factor and its volatility respectively; $\beta_{i,k,t}$ represents the conditional loading of $r_{i,t}$ with respect to $r_{fk,t}$; $\{e_{fk,t}\}$ is an i.i.d. innovation sequence with zero mean and unit variance. $\varepsilon_{i,t}$ is uncorrelated with each factor return given $\mathscr{F}_{t-1}^{\mathrm{HF}}$.

We now assume that the return vector $r_t$ contains $K$ factors and $N$ individual assets where the $K$ factors are in the first $K$ positions, and the realized measure is $K+N$ by $K+N$ where the upper $K$ by $K$ block measures the quadratic covariation of the factors. We use block-partitions of the conditional covariance and the realized measure

$$\Sigma_t = \begin{bmatrix} \Sigma_{f,t} & \Sigma_{f1,t} & \dots & \Sigma_{fN,t} \\ \Sigma_{1f,t} & \Sigma_{11,t} & \dots & \Sigma_{1N,t} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{Nf,t} & \Sigma_{N1,t} & \dots & \Sigma_{NN,t} \end{bmatrix}, RM_t = \begin{bmatrix} RM_{f,t} & RM_{f1,t} & \dots & RM_{fN,t} \\ RM_{1f,t} & RM_{11,t} & \dots & RM_{1N,t} \\ \vdots & \vdots & \ddots & \vdots \\ RM_{Nf,t} & RM_{1N,t} & \dots & RM_{NN,t} \end{bmatrix},$$

where the upper left block is $K$ by $K$, the $if$ blocks are 1 by $K$ and the $ii$ blocks are scalars. We allow for a general dependence structure of factors, and so the conditional loadings for asset $i$ are

$$\beta_{i,t} = \Sigma_{f,t}^{-1}\Sigma_{f1,t}.$$

Similarly, define the multi-factor realized $\beta$s for asset $i$ as

$$R\beta_{i,t} = RM_{f,t}^{-1}RM_{fi,t},$$

where $RM_{f,t}$ is $K$ by $K$ and the $RM_{fi,t}$ is $K$ by 1, and the realized idiosyncratic variance as

$$RIV_{i,t} = RM_{ii,t} - R\beta_{i,t}'RM_{f,t}R\beta_{i,t}.$$

In the $K$-factor HEAVY model, the multiple factors follow a multivariate HEAVY model (Noureldin *et al.*, 2012), while each of the factor loadings and idiosyncratic volatility follow the same dynamics as in the 1-factor model. The HEAVY-P equations are then

$$\begin{aligned}
\Sigma_{f,t} &= CC' + ARM_{f,t-1}A' + B\Sigma_{f,t-1}B' \\
\beta_{i,t} &= \delta_{i,0} + \delta_{i,1}R\beta_{i,t-1} + \delta_{i,2}\beta_{i,t-1} \\
\sigma^2_{i,t} &= \alpha_{i,0} + \alpha_{i,1}RIV_{i,t-1} + \alpha_{i,2}\sigma^2_{i,t-1},
\end{aligned} \tag{14}$$

where $\beta_{i,t} = (\beta_{i,1,t},...,\beta_{i,K,t})$, $\delta_{i,0}$ is a $K$ by 1 vector, $\delta_{i,1}$ and $\delta_{i,2}$ are diagonal matrices and $A$,$B$ and $C$ are parameter matrices which satisfy the assumptions given in Noureldin *et al.* (2012). The HEAVY-M equations are similarly modified, and are

$$\begin{aligned}
M_{f,t} &= C^M \left(C^M\right)' + A^M RM_{f,t-1} \left(A^M\right)' + B^M M_{f,t-1} \left(B^M\right)' \\
\lambda_{i,t} &= \delta^M_{i,0} + \delta^M_{i,1}R\beta_{i,t-1} + \delta^M_{i,2}\lambda_{i,t-1} \\
\mu_{i,t} &= \alpha^M_{i,0} + \alpha^M_{i,1}RIV_{i,t-1} + \alpha^M_{i,2}\mu_{i,t-1},
\end{aligned} \tag{15}$$

where $M_{f,t} = \mathrm{E}_{t-1}(RM_{f,t})$. As before, the final two equations in (14) and (15) are repeated for each asset. Equations (14) and (15) constitute the $K$-factor HEAVY model.

## 2.3 Estimation and Inference

### 2.3.1 Estimation

This discussion focuses on the 1-factor model, before turning to the $K$-factor models. The conditional likelihood of the returns and the realized measures is the natural method to estimate the parameters. The parameters in the HEAVY-P equations are estimated by maximizing conditional likelihood

$$L = \sum_{t=1}^{T} l_t(\psi; r_t), \tag{16}$$

where $\psi = (\theta', \phi_1', \ldots, \phi_N')'$, $\theta = (\theta_0, \theta_1, \theta_2)'$, $\phi_i = (\delta_0, \delta_1, \delta_2, \alpha_0, \alpha_1, \alpha_2)'$,

$$l_t(\psi; r_t) = -\frac{1}{2}\left(\ln \mid \Sigma_t \mid + r_t' \Sigma_t^{-1} r_t\right) + c, \tag{17}$$

and $c$ is a term which does not depend on the model parameters.[5] The model structure can be directly exploited to simplify estimation by expressing the log-likelihood in two components – one which measures the likelihood of the common factor, and one which measures the likelihood of the idiosyncratic errors, $r_{i,t} - \beta_{i,t} r_{f,t}$.

**Proposition 3.** The joint log-likelihood of the daily returns can be equivalently expressed

$$\begin{aligned} l_t &= \underbrace{-\frac{1}{2}\left(\ln\left(\sigma_{f,t}^2\right) + \frac{r_{f,t}^2}{\sigma_{f,t}^2}\right)}_{l_{f,\mathrm{t}}:\ \text{factor}} + \\ & \sum_{i=1}^{N} \underbrace{-\frac{1}{2}\left(\ln(\sigma_{i,t}^2) + \frac{\left(r_{i,t} - \beta_{i,t} r_{f,t}\right)^2}{\sigma_{i,t}^2}\right)}_{l_{i,t}:\ \text{idiosyncratic}\ i} + c. \end{aligned} \tag{18}$$

This decomposition leads to a natural 2-step estimator, where the parameters of the $l_{f,t}$ are first maximized, and then the parameters governing the conditional factor loadings and idiosyncratic volatilities are estimated. The second step of the estimation process involves

---

5

The parameter dynamics are all recursive and so depend on values for initial observations. We use a backward exponentially weighted moving average based on the first $\tau = \lceil T^{0.25} \rceil$ observations of the realized measure to initialize the process so that $M_0 = \sum_{t=1}^{\tau} w_t M_t$ where $w_t = .06\,(.94)^{t-1} / \sum_{j=1}^{\tau} .06\,(.94)^{j-1}$.

$N$ optimizations, although we refer to this is a single step since the ordering of the assets does not matter.

The parameters of the HEAVY-M equations are estimated by maximizing the standardized Wishart log-likelihood,

$$L^M = \sum_{t=1}^{T} l_t^M \left(\psi^M; RM_t\right),$$

where $\psi^M = \left[\left(\theta^M\right)', \left(\phi_1^M\right)', \ldots, \left(\phi_N^M\right)'\right]'$, $\theta^M = \left(\theta_0^M, \theta_1^M, \theta_2^M\right)'$, $\phi_i^M = \left(\delta_0^M, \delta_1^M, \delta_2^M, \alpha_0^M, \alpha_1^M, \alpha_2^M\right)'$,

$$l_t^M = -\frac{\nu}{2}\left(\ln|M_t| + \operatorname{tr}\left(M_t^{-1}\left(RM_t\right)\right)\right) + c_\nu^M, \tag{19}$$

and $c_\nu^M$ is a constant conditional on the shape parameter of the standardized Wishart, $\nu$. Our interest is in the parameters of the dynamics, and so we do not estimate this parameter. Using the structure of $M_t$, this log-likelihood can be similarly decomposed into two components,

$$l_t^M = \underbrace{-\frac{\nu}{2}\left(\ln\left(\mu_{f,t}\right) + \frac{RM_{f,t}}{\mu_{f,t}}\right)}_{l_{f,t}^M\text{: factor}} + \sum_{i=1}^{N} \underbrace{-\frac{\nu}{2}\left(\ln\left(\mu_{i,t}\right) + \frac{\left(\lambda_{i,t}^2 RM_{f,t} - 2\lambda_{i,t} RM_{fi,t} + RM_{i,t}\right)}{\mu_{i,t}}\right)}_{l_{i,t}^M\text{: idiosyncratic } i} + c_\nu^M. \tag{20}$$

This decomposition also leads to a natural 2-step estimation strategy, where the parameters governing the factor dynamics are first estimated and then the parameters of the idiosyncratic volatility are estimated. This decomposition comes directly from the model and does not require correct specification or other restrictions on the realized measure. This is particularly useful since estimation of the model parameters only requires storing the realized measure for the factor, the assets and between the factor and assets, and not the complete $N+1$ by $N+1$ realized measure. The likelihood structure is particularly useful when using noise-robust realized measures which are known to suffer from data attrition due to refresh-time sampling when the number of assets, $N$, is large.

The joint likelihoods of the $K$-factor model can be similarly decomposed so that

$$l_t = \underbrace{-\frac{1}{2}\left(\ln\left|\Sigma_{f,t}\right| + r'_{f,t}\Sigma_{f,t}^{-1}r_{f,t}\right)}_{l_{f,t}\text{: factor}} + \sum_{i=1}^{N}\underbrace{-\frac{1}{2}\left(\ln\left(\sigma_{i,t}^2\right) + \frac{\left(r_{i,t} - \sum_{k=1}^{K}\beta_{i,k,t}r_{fk,t}\right)^2}{\sigma_{i,t}^2}\right)}_{l_{i,t}\text{: idiosyncratic } i} + c$$

and

$$\begin{aligned} l_t^M \quad = \quad & \underbrace{-\frac{\nu}{2}\left(\ln\left|M_{f,t}\right| - \operatorname{tr}\left(M_{f,t}^{-1}\left(RM_{f,t}\right)\right)\right)}_{l_{f,t}^M\text{: factor}} \\ & +\sum_{i=1}^{N}\underbrace{-\frac{\nu}{2}\left(\ln(\mu_{i,t}) - \frac{\lambda'_{i,t}RM_{f,t}\lambda_{i,t} - 2\lambda'_{i,t}RM_{fi,t} + RM_{ii,t}}{\mu_{i,t}}\right)}_{l_{i,t}^M\text{: idiosyncratic } i} + c_\nu^M, \end{aligned}$$

where $\lambda_{i,t} = (\lambda_{i,1,t}, ..., \lambda_{i,K,t})'$. The Factor-HEAVY structure preserves the variation-free nature of the HEAVY framework in the sense that the lagged terms from one set of equations do not appear in the other.

### 2.3.2 Quasi-likelihood Based Asymptotic Inference

Similar to the HEAVY models, the parameter estimators in Factor HEAVY equations have the common asymptotic properties of quasi-maximum likelihood estimators. Since the parameters have no link between factor and equity equations, and between HEAVY-P and HEAVY-M equations, we can consider their asymptotic properties separately. Here we focus on the 1-factor HEAVY model for simplicity. The score equations for the corresponding likelihood, evaluated at the estimated parameters are

$$\begin{array}{ll} \sum_{t=1}^{T} S_{f,t}\left(\hat{\theta}\right) = 0 & \sum_{t=1}^{T} S_{f,t}^{M}\left(\widehat{\theta^{M}}\right) = 0 \\ \sum_{t=1}^{T} S_{i,t}\left(\hat{\phi}_i\right) = 0 & \sum_{t=1}^{T} S_{i,t}^{M}\left(\widehat{\phi_i^{M}}\right) = 0, \end{array}$$

where $S_{f,t}\left(\theta\right) = \partial l_{f,t}/\partial\theta$, $S_{i,t}\left(\phi_i\right) = \partial l_{i,t}/\partial\phi_i$, $S_{f,t}^{M}\left(\theta^{M}\right) = \partial l_{f,t}^{M}/\partial(\theta^{M})$, $S_{i,t}^{M}\left(\phi_i^{M}\right) = \partial l_{i,t}^{M}/\partial(\phi_i^{M})$. Let $\theta_o$, $\phi_{i,o}$, $\theta_o^M$ and $\phi_{i,o}^M$ indicate the true parameter values. The scores evaluated at these values are martingale difference sequences with respect to $\mathscr{F}_{t-1}^{\mathrm{HF}}$. Under certain regularity conditions (see Bollerslev & Wooldridge (1992), Newey & McFadden (1994), White (1996), *inter alia.*), we have

$$\begin{array}{ll} \sqrt{T}\left(\hat{\theta}-\theta_o\right) \to N\left(0, \mathscr{I}_{\theta}^{-1}\mathscr{J}_{\theta}\mathscr{I}_{\theta}^{-1}\right) & \sqrt{T}\left(\widehat{\theta^{M}}-\theta_o^{M}\right) \to N\left(0, \left(\mathscr{I}_{\theta^M}^{M}\right)^{-1}\mathscr{J}_{\theta^M}^{M}\left(\mathscr{I}_{\theta^M}^{M}\right)^{-1}\right) \\ \sqrt{T}\left(\hat{\phi}_i-\phi_{i,o}\right) \to N\left(0, \mathscr{I}_{\phi_i}^{-1}\mathscr{J}_{\phi_i}\mathscr{I}_{\phi_i}^{-1}\right) & \sqrt{T}\left(\widehat{\phi_i^{M}}-\phi_{i,o}^{M}\right) \to N\left(0, \left(\mathscr{I}_{\phi_i^M}^{M}\right)^{-1}\mathscr{J}_{\phi_i^M}^{M}\left(\mathscr{I}_{\phi_i^M}^{M}\right)^{-1}\right), \end{array}$$

where

$$\mathscr{I}_{\theta} = -\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(\left.\frac{\partial S_{f,t}(\theta)}{\partial\theta'}\right|_{\theta=\theta_o}\right) \quad \mathscr{J}_{\theta} = \mathrm{avar}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}S_{f,t}\left(\theta_o\right)\right)$$

$$\mathscr{I}_{\phi_i} = -\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(\left.\frac{\partial S_{i,t}\left(\phi_i\right)}{\partial\phi_i'}\right|_{\phi_i=\phi_{i,o}}\right) \quad \mathscr{J}_{\phi_i} = \mathrm{avar}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}S_{i,t}\left(\phi_{i,o}\right)\right)$$

$$\mathscr{I}^{M}_{\theta^M} = -\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(\left.\frac{\partial S^{M}_{f,t}(\theta^M)}{\partial\left(\theta^M\right)'}\right|_{\theta^M=\theta^M_o}\right) \quad \mathscr{J}^{\mathscr{M}}_{\theta} = \mathrm{avar}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} S^{M}_{f,t}\left(\theta^M_o\right)\right)$$

$$\mathscr{I}^{M}_{\phi^M_i} = -\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(\left.\frac{\partial S^{M}_{i,t}(\phi^M_i)}{\partial\left(\phi^M_i\right)'}\right|_{\phi^M_i=\phi^M_{i,o}}\right) \quad \mathscr{J}^{M}_{\phi^M_i} = \mathrm{avar}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} S^{M}_{i,t}\left(\phi^M_{i,o}\right)\right).$$

We have omitted calculation of the cross-terms in the variance covariance between components of the model. While these are generally not of interest, calculation of these is straight-forward. Inference in the Factor HEAVY model is simplified when compared to typical multivariate volatility models since the estimating equations have a natural block structure due to the factor structure.

## 2.4 Empirical Analysis and Model Evaluation

### 2.4.1 Data and Descriptive Statistics

We apply the model to a sample containing 40 large U.S. financial firms from July 1, 2000 to June 30, 2010 so that a maximum of 2,511 daily observations are available. The panel is slightly heterogeneous since some of the included firms stop trading during the financial crisis. High frequency quote and trade data are extracted from the Trade and Quote (TAQ) database, and daily price data are extracted from Center for Research in Security Prices (CRSP) stock database. High-frequency data typically contains mis-recordings and other erroneous data, and so all price data was cleaned using a slightly modified set of rules to those used in Barndorff-Nielsen *et al.* (2009).[6] Our initial focus is on a 1-factor model

6

The sole modification is to use the market each day that has the highest trading volume rather than always selecting NYSE.

similar to a conditional CAPM (Jagannathan & Wang 1996; Lewellen & Nagel 2006). We choose the S&P 500 as the market proxy, and use the SPDR S&P 500 (SPY), a highly liquid ETF that tracks the S&P 500 index.

In most of the application we use realized covariance as the realized measure. We sample all prices using last-price interpolation and use a sparse sampling scheme. Andersen *et al.* (2000) argue that the optimal choice of sampling frequency for realized volatility and correlation is some intermediate value. On the one hand, the effect of microstructure noise is serious when sampling frequency is higher; on the other hand, sampling variation is large when sampling frequency is lower. In Section 2.4.4, we show that the in-sample performance of 10-minute realized covariance is slightly better than that of 5-minute realized covariance. Consequently, our preferred estimator is based on 10-minute sampling with subsampling every minute. Suppose $p_{j,t}$ is the $j^{\text{th}}$ log price vector on day $t$ containing the log prices of the factor and the 40 firms. Then the sub-sampled realized covariance is defined as

$$RC_t^{SS} = \frac{m-1}{s\,(m-s)} \sum_{i=1}^{m-s} \bar{r}_{i,t}\bar{r}'_{i,t},$$

where $\bar{r}_{i,t} = \sum_{j=i}^{s} r_{i+j,t}$ , $s$ is the length of the block (e.g. 10), $r_{j,t} = p_{j,t} - p_{j-1,t}$ are the high-frequency returns, $m$ is the number of price samples (e.g. 390 when using 1-minute returns in U.S. equity data). We consider alternative specifications where we vary $s \in \{5, 10, 15, 30\}$, as well as a realized kernel where we use the non-flat Parzen kernel and the bandwidth selection procedure outlined in Barndorff-Nielsen *et al.* (2011).

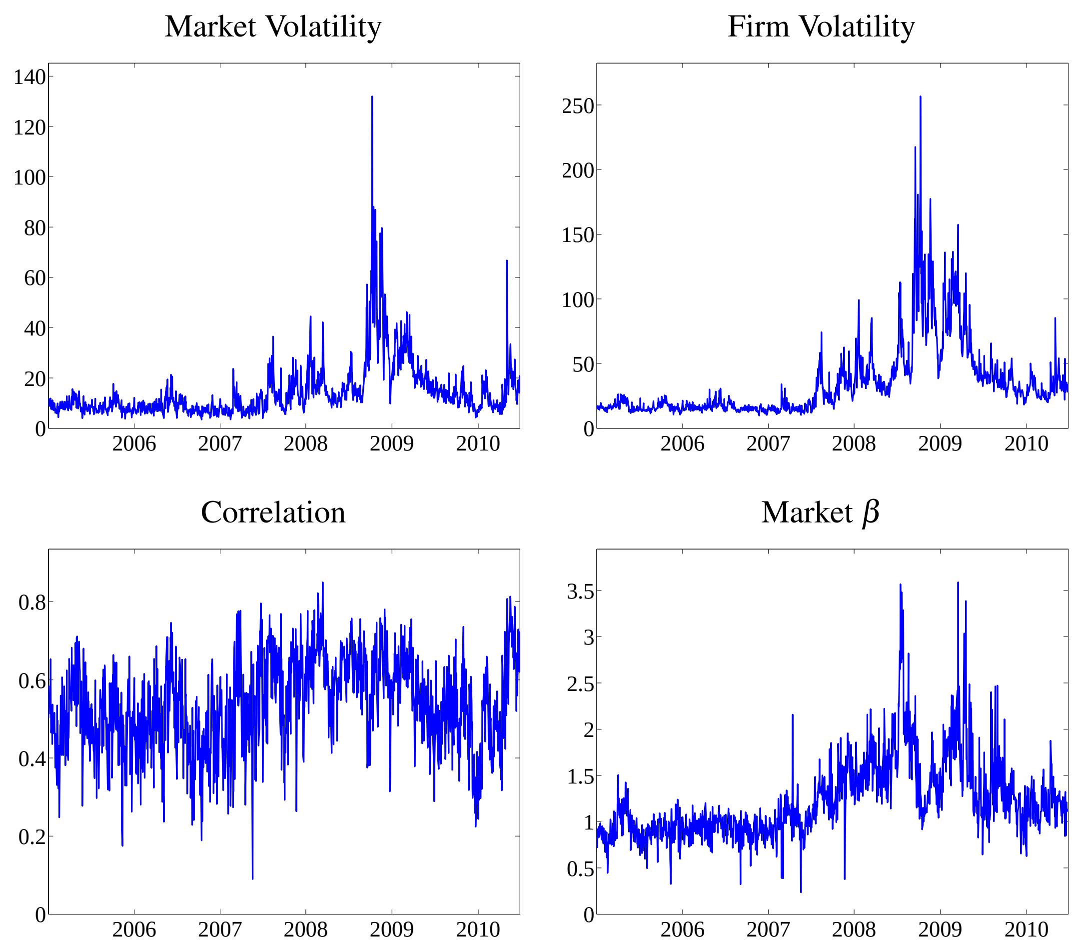


Figure 1: Top Left: Annualized realized volatility of SPY, the square root of the mean of 252 times 1-minute realized variance; Top Right: average annualized realized volatility of all firm equity returns; Bottom Left: average realized correlations between SPY and firm equity returns; Bottom Right: average realized $\beta$s.

Figure 1 shows annualized realized volatility of the factor, average annualized realized volatility of all firm equity returns, average realized correlations between SPY and equity returns, and average realized $\beta$s. It is obvious that from summer 2007, the beginning of the financial crisis, daily returns become more volatile than before the crisis. The returns on the factor are less volatile than those on equity returns. During the crisis, the average realized correlations increased relative to their pre-crisis values, although the changes in dependence are far more striking in terms of the average realized $\beta$s.

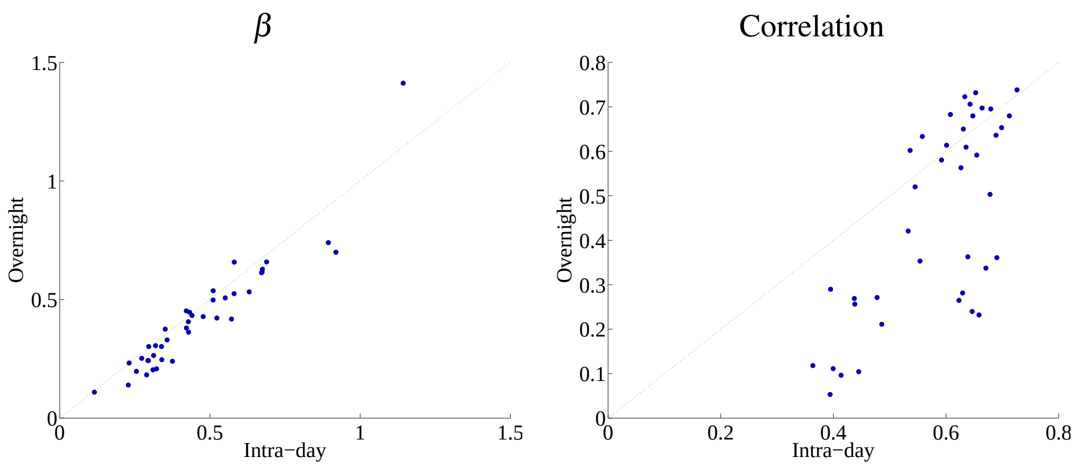


Figure 2: The left panel shows intra-daily and overnight market $\beta$s for the 40 firms in the applicaiton. The right panel shows the intra-daily and overnight correlations between the firm returns and the market for the same firms.

One of the primary difficulties encountered when incorporating realized measures into a model of the conditional covariance of low-frequency data is the missing overnight data. In models of the conditional variance, it is common to assume that the full-day variance can be simply scaled from the within trading-day variance. When studying multivariate quantities, there are more possibilities, and in particular the covariance and the variances may not scale with a common factor. We examine the choice of modeling space in Figure 2, where we compare the intra-daily and overnight $\beta$s and as well as the intra-daily and overnight correlations between the factor and equity returns. We compute these using only the open-to-close and close-to-open returns (not the other intra-daily data). The result shows $\beta$ behaves very differently from correlation – most $\beta$s lie near the 45-degree line, indicating that $\beta$s are very stable throughout the entire day. The relationship between intra-daily and overnight correlations appears to be more complicated, and many of the firms have stronger correlations with the factor during the market opening hours than during the overnight period. This is consistent with systematic news arriving during market hours and idiosyncratic news arriving after the market closes. This difference indicates that models incorporating high-frequency data that parameterize the conditional $\beta$ are better suited than models which use other transformations of the conditional covariance.

We are primarily interested in the covariance of returns over the entire day, including both the period where markets are active and the overnight return. Realized measures can only be computed for the period where the market is actively traded, so we use a transformation to ensure that the average value of the realized measure is the same as the average value of the outer product of returns (See the appendix for details).

Finally, a particular difficulty arises when using noise-robust estimators constructed using only the $K$ factors and the individual assets. This occurs since these estimators typically use some form of refresh time sampling and so will not produce identical estimates of the quadratic covariation of the factors. We propose to first transform these so that they have the same estimate of the quadratic covariation of the factors, and then standardize these modified realized measures (Refer to Section 2.6.2).

### 2.4.2 In-Sample Performance

| | $\beta$ | | | | Idiosyncratic | | | |
|---|---|---|---|---|---|---|---|---|
| | $\delta_{i,1}$ | $\delta_{i,2}$ | $\delta^M_{i,1}$ | $\delta^M_{i,1}+\delta^M_{i,2}$ | $\alpha_{i,1}$ | $\alpha_{i,2}$ | $\alpha^M_{i,1}$ | $\alpha^M_{i,1}+\alpha^M_{i,2}$ |
| Min | 0.01 | 0.01 | 0.07 | 0.96 | 0.07 | 0.06 | 0.08 | 0.90 |
| $Q_{.25}$ | 0.08 | 0.84 | 0.12 | 0.98 | 0.42 | 0.29 | 0.39 | 0.99 |
| Mean | 0.12 | 0.82 | 0.15 | 0.98 | 0.51 | 0.44 | 0.44 | 0.99 |
| Median | 0.10 | 0.89 | 0.15 | 0.99 | 0.49 | 0.42 | 0.45 | 0.99 |
| $Q_{.75}$ | 0.13 | 0.92 | 0.17 | 0.99 | 0.64 | 0.54 | 0.50 | 0.99 |
| Max | 0.53 | 0.99 | 0.26 | 0.99 | 0.92 | 0.93 | 0.63 | 1.00 |

Table 1: Cross-sectional statistics of full-sample parameter estimates for the 40 financial firms. The left panel contains estimates from the $\beta$ component of the HEAVY-P and HEAVY-M equations, and the right panel contains estimates from the idiosyncratic volatility component.

The Factor HEAVY model is estimated using the complete sample available for each firm. Tables 1 presents the minimum, the 25th percentile, the mean, the median, the 75th percentile and the maximum of the parameter estimates across all firms. The estimates of the dynamics of the conditional $\beta$ lie in a narrow range, with most sensitivity parameters estimated to be close to 0.10. The estimates of the parameters in the HEAVY-M equations fall into a particularly narrow band. The estimated dynamics of the conditional idiosyncratic volatilities are more dispersed, although mode series are highly responsive to idiosyncratic news. The estimates from the idiosyncratic component of the HEAVY-M are also more similar than the HEAVY-P, with both high sensitivity to news and large persistence.

In traditional ARCH-type models, both volatilities and $\beta$s are driven by daily shocks. We consider an augmented model of the HEAVY-P equations which allow for the natural daily shock to enter the model. The modified equations are

$$
\begin{aligned}
\sigma^2_{f,t} &= \theta_0 + \theta_1 RM_{f,t-1} + \theta_2 \sigma^2_{f,t-1} + \theta_3 r^2_{f,t-1} \\
\beta_{i,t} &= \delta_{i,0} + \delta_{i,1} \mathrm{R}\beta_{i,t-1} + \delta_{i,2}\beta_{i,t-1} + \delta_{i,3} r_{f,t-1} r_{i,t-1} / \sigma^2_{f,t-1} \\
\sigma^2_{i,t} &= \alpha_{i,0} + \alpha_{i,1} RIV_{i,t-1} + \alpha_{i,2}\sigma^2_{i,t-1} + \alpha_{i,3}\varepsilon^2_{i,t-1},
\end{aligned}
$$

where $\theta_3$ allows the daily factor return to influence the factor volatility, $\delta_{i,3}$ allows a daily $\beta$ shock to enter the $\beta$ dynamics and $\alpha_{i,3}$ allows for an idiosyncratic shock where $\varepsilon_{i,t} = r_{i,t} - \beta_{i,t} r_{f,t}$. These are tested one-at-a-time, and all are conducted at the 5% level. Estimation is conducted using the 2-step estimator described in section 2.3.1, and inference is conducted using robust standard errors similar to those described in section 2.3.2. The daily shock in the factor volatility equation is not significant. In the $\beta$ equations, $\delta_{i,3}$ is significant for 8 firms, but has a much smaller coefficient than $\delta_{i,1}$ indicating the changes in the fit value are typically small. In the idiosyncratic equations, $\alpha_{i,3}$ is significant for 7 firms.

We conduct a number of other experiments to assess whether all components of the models are necessary. We first compare the in-sample fit with that of the pairwise fit of the cDCC model, which has recently been used to fit dynamic $\beta$ models (Engle & Kelly 2012; Bali *et al.* 2012). The cDCC model specifies dynamics for the conditional variances of the market and the individual asset as standard GARCH models, and the conditional correlation is modeled using the standardized residuals.

We also compare the Factor HEAVY with the daily $\beta$-GARCH model (eq. (1)), as well as the multivariate HEAVY estimated on the market and each individual asset. Finally, we compare the Factor HEAVY to two nested specifications. The first assumes that the conditional factor loading is constant, which corresponds to restrictions that $\delta_{i,1} = \delta_{i,2} = 0$, and the second which assumes that the idiosyncratic volatility is constant, corresponding to the restrictions $\alpha_{i,1} = \alpha_{i,2} = 0$.

| | Daily | | High-Frequency Based | | |
|---|---|---|---|---|---|
| | cDCC | $\beta$ GARCH | Constant Factor Loading | Constant Idiosyncratic Variance | Multivariate HEAVY |
| Min | 85.5 | 82.48 | 4.277 | 211.6 | -23.340 |
| 25% | 137.9 | 135.8 | 31.77 | 483.9 | 1.220 |
| Median | 157.4 | 164.6 | 65.39 | 956.6 | 6.000 |
| Mean | 169.6 | 169.8 | 67.33 | 1004 | 6.356 |
| 75% | 201.0 | 190.7 | 101.7 | 1206 | 9.074 |
| Max | 307.0 | 309.7 | 149.0 | 2441 | 42.05 |

Table 2: In-sample log-likelihood comparison between Factor HEAVY and competing models for all equities. All comparisons are performed using the pair of the market and one of the financial firms.

Table 2 contains the value of the difference in the log-likelihood of these models,

$$\sum_{t=1}^{T} (l_{t,\text{Factor HEAVY}} - l_{t,\text{Alternative}}),$$

evaluated using only daily data in the case of HEAVY models. Since all models assume that returns are conditionally normal, these are directly comparable, and positive values indicate that the Factor HEAVY produced a superior in-sample fit. The Factor HEAVY produces much larger log-likelihoods than either of the daily-only models, and the closest daily model differs by over 80 log-likelihood points. Moreover, the typical difference is more than 150 points, indicating that there is substantial information in the high-frequency measures. The models nested in the Factor HEAVY – either with a constant factor loading or a constant idiosyncratic volatility – are also considerably worse, although the assumption of constant factor loading does not always lead to large changes in the log-likelihood. On the other hand, idiosyncratic volatilities appear to be time-varying with typical log-likelihood differences of 1,000 points. Finally, the results comparing the Factor HEAVY with bivariate multivariate HEAVY models indicate some preference for the Factor HEAVY, with better performance in more than 75% of the series examined.

We also examine the fit of the model using residual-based tests. We are primarily concerned with misspecification of the conditional covariance between the returns on a pair of equities since the model does not explicitly attempt to fit these. We are interested in testing the null that

$$H_0 : \mathrm{E}_{t-1}\left(r_{i,t} r_{j,t}\right) = \beta_{i,t}\beta_{j,t}\sigma^2_{f,t},$$

where we have suppressed the dependence of the factor loadings and market variance on the model parameters,

$$\zeta = \left(\theta_o', \delta_{i,o}', \delta_{j,o}'\right)',$$

which includes the parameters of the conditional variance of the factor as well as the conditional factor loadings of assets $i$ and $j$. Since the parameter estimates are variation-free across models, it is not necessary to include the parameters for the other components in $\zeta$.

We use the robust regression-based specification tests developed in Wooldridge (1990). We use $z_{i,j,t} = (1, r_{i,t-1}r_{j,t-1}, r_{i,t-2}r_{j,t-2})$ as the vector of misspecification indicator variables. These indicators will have power against static misspecification (through the constant term) as well as persistence in the cross-products. The test is implemented in two steps. First, we regress each element of $z_{i,j,t}$ on the gradient vector $\nabla_\zeta \eta_t$ of $\eta_t = r_{i,t}r_{j,t} -$

$\beta_{i,t}\beta_{j,t}\sigma^2_{f,t}$ with respect to $\zeta$ evaluated at the estimate $\hat{\zeta}$. We define $\hat{z}_{i,j,t}$ as the residual vector obtained from this regression. Second, we regress unit on the vector $\hat{\eta}_t\hat{z}_{i,j,t}$ where $\hat{\eta}_t$ is the value of $\eta_t$ evaluated at $\hat{\zeta}$. The test statistic is then computed as $T \times R^2$ where the $R^2$ comes from the second regression. The test statistics is asymptotically distributed as $\chi^2_3$. Figure 3 contains a histogram of the test statistics. We fail to reject the null of correct conditional specification in 93% of the pairs tested (the 5% critical value of a $\chi^2_3$ is 7.81), and so the rejection rate is close to size.

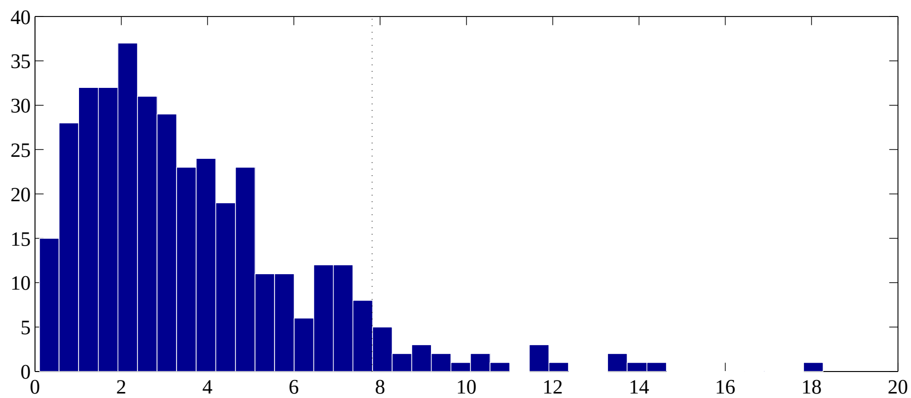


Figure 3: Histogram of misspecification test statistics.

#### 2.4.3 Out-of-Sample Forecasting Results

We next turn attention to assessing the out-of-sample performance of the Factor HEAVY, and focus the out-of-sample comparisons on the cDCC model, as a leading example of low-frequency models. All models are fitted using a recursive scheme where parameters are updated once a week (on Fridays) starting from July 2005, which allows for a minimum of 1,250 days in the smallest models. The cDCC models are always estimated pairwise with only the market and one of the financial firms. We also present a volatility forecasting comparison with the long memory Riskmetrics model (Zumbach, 2007). We will evaluate the performance of the model using 1-day, 1-week and 2-week forecast horizons, which are all important for risk management.[7]

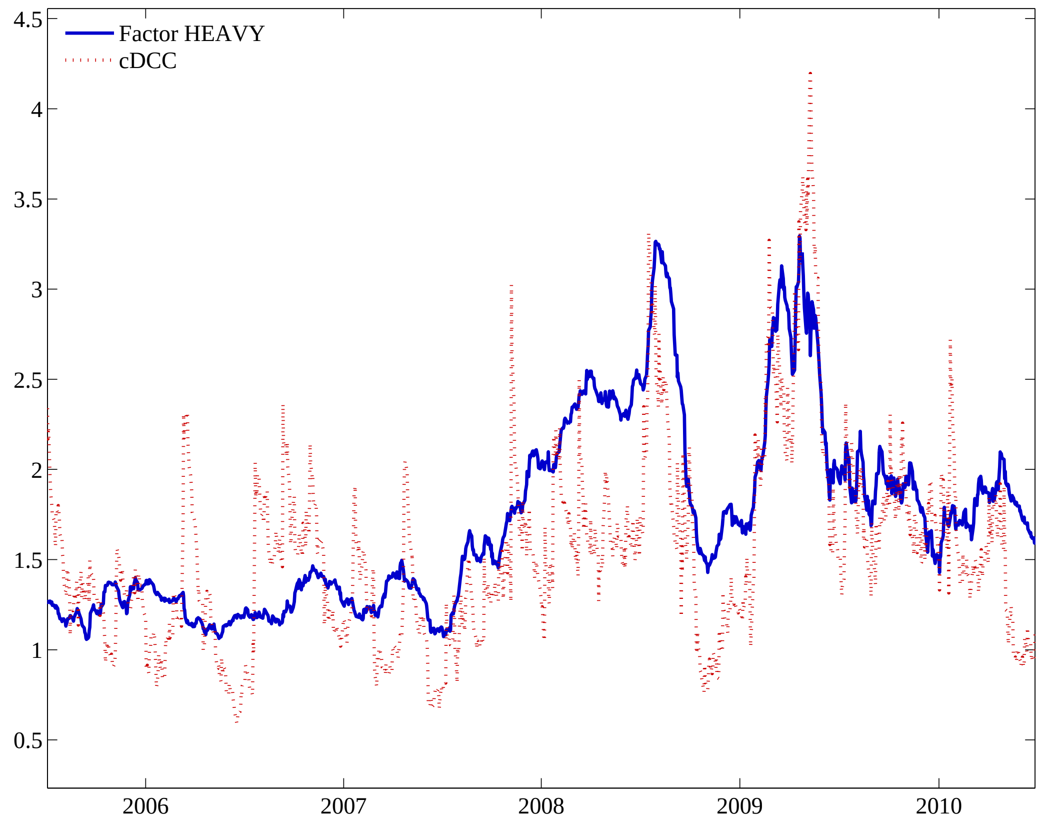


Figure 4: 1-step $\beta$ forecasts of cDCC and Factor HEAVY models for COF. Solid lines denote the Factor HEAVY forecasts; while dash lines represent the cDCC forecasts.

Figure 4 shows the 1-step forecasts of cDCC and Factor HEAVY models for COF. The conditional $\beta$ forecasts from the Factor HEAVY are more persistent than those from

7

Results for longer horizons, up to 1 months, indicate there is little difference between the models.

the cDCC. This higher persistence originates from the direct modeling of conditional $\beta$ in eq. (7) where the coefficients indicate considerable persistence and a slow response to news. Both series of conditional $\beta$s show large increases during the financial crisis and the continuing turmoil of early 2009, although they differ markedly in the pre-crisis period of 2007 until mid-2008, where the HEAVY model indicates substantial increases in conditional $\beta$ while the cDCC does not.

**Statistical Accuracy** We compare the statistical accuracy of all models using out-of-sample comparisons. The QLIK loss function has recently emerged as a sensible method to evaluate variance and covariance forecasts in the presence of noisy proxies (see Patton & Sheppard (2009), Patton (2011) and Laurent *et al.* (2013)). The QLIK loss function uses the kernel of the Gaussian log-likelihood to evaluate forecasts,

$$l_t^s\left(\hat{\Sigma}_{t+s|t}\right) \quad = \quad \ln\left|\hat{\Sigma}_{t+s|t}\right| + \hat{\Sigma}_{t+s|t}^{-1}\operatorname{tr}\left(C_{t+s}\right), \tag{21}$$

where $s$ is the number of steps ahead and $\hat{\Sigma}_{t+s|t}$ is a model-based time-$t$ conditional forecast of the covariance of $r_{t+s}$ and $C_{t+s}$ is the proxy for the unobserved covariance at time $t+s$. The natural proxy is $r_t r_t'$, although others, such as a realized measure, can be used.[8] Suppressing the explicit dependence on $\hat{\Sigma}_{t+s|t}$, the QLIK loss function can be decomposed to isolate various component losses,

[8] The proxy should satisfy the condition that $\mathrm{E}_t\left(C_{t+s}\right) = \Sigma_{t+s}$, so that it is unbiased. When forecasting the covariance over the entire day, most realized measures will fail to meet this criteria.

$$
\begin{aligned}
l_t^s &= \underbrace{\left(\ln\left(\sigma^2_{f,t+s|t}\right) - \frac{V_{f,t+s}}{\sigma^2_{f,t+s|t}}\right)}_{\text{factor}} + \sum_{i=1}^{N} \underbrace{\left(\ln\left(\sigma^2_{i,t+s|t}\right) - \frac{\beta^2_{i,t+s|t}V_{f,t+s} - 2\beta_{i,t+s|t}C_{fi,t+s} + V_{i,t+s}}{\sigma^2_{i,t+s|t}}\right)}_{\text{idiosyncratic } i} \\
&= \underbrace{\left(\ln\left(\sigma^2_{f,t+s|t}\right) - \frac{V_{f,t+s}}{\sigma^2_{f,t+s|t}}\right)}_{\text{factor}} + \sum_{i=1}^{N} \underbrace{\left(\ln(\beta^2_{i,t+s|t}\sigma^2_{f,t+s|t} + \sigma^2_{i,t+s|t}) - \frac{V_{i,t+s}}{\beta^2_{i,t+s|t}\sigma^2_{f,t+s|t} + \sigma^2_{i,t+s|t}}\right)}_{\text{individual equity } i} \\
&\quad + \text{term corresponding to the dependence structure} \\
&= l^s_{f,t} + \sum_{i=1}^{N} l^s_{i,t} + l^s_{C,t}, \qquad (22)
\end{aligned}
$$

where $V_{f,t+s}$, $V_{i,t+s}$ and $C_{fi,t+s}$ are the proxies for the factor variance, equity variance and their covariance on day $t+s$. While this decomposition is generally only applicable to factor models, it also holds for any bivariate model which contains the factor, such as the daily cDCC models.

All comparisons are implemented using Diebold & Mariano (1995) (DM) tests of equal predictive accuracy where the covariance forecasts come from the Factor HEAVY and bivariate cDCC models. The null in DM test statistics is of equal expected loss,

$$
H_0 : \mathrm{E}\left[l_t^s\left(\hat{\Sigma}^{\text{HEAVY}}_{t+s|t}\right)\right] = \mathrm{E}\left[l_t^s\left(\hat{\Sigma}^{\text{cDCC}}_{t+s|t}\right)\right],
$$

and the alternative is two-sided. Define the loss differential as $\delta_t = l_t^s\left(\hat{\Sigma}^{\text{cDCC}}_{t+s|t}\right) - l_t^s\left(\hat{\Sigma}^{\text{HEAVY}}_{t+s|t}\right)$, then positive values correspond to superior performance of the Factor HEAVY while negative values correspond to superior performance of the cDCC. The test statistic is implemented as a simple t-statistic

$$DM = \frac{1}{\sqrt{T-\tau+1}} \frac{\sum_{t=\tau}^{T} \delta_t}{\sqrt{\text{avar}\left(\delta_t\right)}},$$

where $\text{avar}\left(\delta_t\right)$ is a consistent estimator of the long-run variance of $\delta_t$. This is estimated using a Newey-West covariance estimator with $\min\left(10, \lfloor 1.2T^{1/3} \rfloor\right)$ lags.

| | 1 day | | | 1 week | | | 2 weeks | | |
|---|---|---|---|---|---|---|---|---|---|
| | Joint | Firm | Copula | Joint | Firm | Copula | Joint | Firm | Copula |
| AET | **4.42** | **3.25** | 1.44 | **2.40** | **2.64** | -0.39 | **2.82** | **2.39** | -0.53 |
| AFL | **3.14** | **1.79** | **3.30** | 1.37 | 1.35 | 0.36 | 1.62 | 1.47 | -0.05 |
| AIG | -0.09 | -1.00 | 1.20 | -0.81 | -1.27 | 1.60 | -1.17 | -1.42 | -0.07 |
| ALL | **3.21** | **2.25** | 0.98 | **1.99** | 1.61 | 1.45 | **2.39** | **1.90** | 1.09 |
| AXP | **3.88** | **2.18** | 1.48 | 1.55 | 0.99 | 1.42 | 1.04 | 0.42 | 0.69 |
| BAC | 1.37 | 0.81 | -0.24 | -0.27 | -0.67 | 0.41 | -0.70 | -1.07 | 0.34 |
| BBT | **4.22** | **2.61** | 1.06 | 1.05 | 0.36 | 0.99 | **2.45** | **2.33** | 0.84 |
| BEN | **3.70** | **2.10** | 0.72 | **1.86** | 1.24 | 1.14 | **2.11** | 1.43 | 1.08 |
| BK | **5.14** | **2.65** | **2.77** | **2.15** | **1.91** | 0.58 | **2.28** | **2.05** | 0.72 |
| C | **2.45** | 1.02 | 1.41 | 0.28 | -0.73 | 1.37 | -1.11 | *-1.77* | -0.19 |
| CB | **5.56** | **2.96** | **4.69** | **2.73** | 1.63 | **1.89** | **3.40** | **1.99** | 0.47 |
| CMA | **4.00** | **2.06** | **1.86** | **2.36** | 1.64 | **1.68** | **3.16** | **2.91** | 1.45 |
| COF | **5.90** | **3.70** | **2.35** | **2.93** | **2.86** | 0.95 | **1.83** | **2.24** | 0.14 |
| ETFC | **1.88** | **2.08** | 0.01 | 1.36 | **1.66** | 0.76 | **1.81** | **1.65** | 1.22 |
| EV | **2.87** | 0.83 | 0.34 | 0.79 | -0.70 | 1.61 | **1.69** | 0.97 | 0.47 |
| FITB | **4.07** | **2.69** | **2.01** | 0.98 | 0.29 | 1.58 | **1.86** | 1.28 | 1.03 |
| FNM | 1.37 | 1.12 | 0.44 | 1.38 | 1.11 | 1.31 | **1.70** | 1.46 | 0.55 |
| FRE | 0.88 | 0.99 | 0.00 | -0.38 | -0.59 | 1.36 | -0.49 | -0.73 | 1.50 |
| GS | 1.44 | 1.51 | -1.30 | 0.65 | **1.76** | -1.09 | 0.15 | 0.12 | -0.48 |
| HIG | 0.85 | **1.99** | -1.21 | -0.41 | -0.15 | -0.73 | 1.13 | 1.30 | 0.59 |
| HRB | **2.73** | **2.34** | -0.08 | 0.70 | 0.65 | -0.04 | -0.27 | -0.20 | -0.91 |
| JNS | **4.02** | **2.27** | **2.64** | 1.32 | 1.12 | 0.73 | 1.02 | 0.91 | 0.81 |
| JPM | **4.74** | **2.81** | 1.39 | 0.77 | 0.48 | 0.76 | 0.15 | -0.36 | 0.66 |
| KEY | **2.24** | -0.05 | **1.88** | -0.29 | -1.23 | 0.85 | 0.02 | -1.05 | 1.03 |
| LEH | **2.05** | **1.94** | 0.17 | 1.20 | **1.87** | 0.29 | -0.93 | -0.79 | *-1.72* |
| LM | **3.08** | 1.28 | 0.92 | **2.04** | **2.28** | -0.10 | **3.16** | **2.54** | 0.29 |
| MER | **4.34** | **3.99** | **1.75** | 1.08 | 0.83 | 0.76 | -1.21 | -1.16 | -1.40 |
| MET | **5.00** | **1.72** | **3.10** | **2.15** | 1.10 | **1.99** | 0.92 | 0.66 | 0.30 |
| MMC | **5.34** | **3.63** | 0.25 | 0.61 | **3.93** | -1.16 | 0.95 | **3.05** | -1.12 |
| MS | **3.01** | **2.60** | -0.61 | 0.47 | 0.94 | -0.62 | -0.98 | -0.87 | -1.30 |
| PGR | **2.17** | **2.43** | *-1.66* | 0.35 | **1.75** | -1.61 | 0.52 | 1.49 | *-2.14* |
| PNC | **4.74** | **2.82** | **2.31** | **1.82** | 0.47 | 1.46 | 0.36 | -0.13 | **1.85** |
| PRU | **3.13** | **2.71** | -1.07 | -0.46 | 0.57 | *-1.91* | 0.49 | 0.32 | -0.19 |
| SCHW | **3.22** | **2.55** | -0.54 | 1.36 | 1.52 | 0.11 | 0.89 | 0.64 | 0.00 |
| SLM | 0.01 | -0.51 | 0.53 | -0.16 | -0.24 | -0.02 | 0.91 | 1.42 | -1.28 |
| STI | **3.56** | **2.64** | 0.69 | 0.78 | 0.48 | 0.49 | 0.11 | 0.00 | -0.23 |
| STT | **3.78** | **2.12** | **3.14** | 0.22 | 0.70 | -0.27 | 0.70 | 0.93 | 0.00 |
| TROW | **2.53** | 0.44 | 0.62 | -0.09 | -0.80 | 0.45 | 0.98 | -0.08 | 0.97 |
| UNH | **2.76** | 1.23 | 0.88 | 1.24 | 0.77 | 1.14 | 1.41 | 0.86 | 0.49 |
| USB | **2.01** | 0.69 | 0.10 | -0.27 | -0.70 | 0.22 | -0.59 | -0.71 | -0.39 |

Table 3: Diebold-Mariano statistics for the null hypothesis that Factor HEAVY and cDCC forecasts are equally accurate when forecasting 1-day, 1-week and 2-week ahead volatility of firm equity returns, the dependence structure and the covariance matrix. Numbers in bold font indicate that Factor HEAVY forecasts are more accurate than cDCC forecasts at the 10% significance level. Italic numbers indicate the converse.

Table 3 contains the value of the DM test statistics for three distinct comparisons. The first compares the bivariate loss (eq. (21)) applied to the market and the individual firm, the firm only loss ($l^s_{i,t}$) and the copula ($l^s_{C,t}$). All comparisons are conducted for 1-day, 5-day and 2-week horizons, and are implemented on point-wise forecasts and not cumulative – that is, predicting $\Sigma_{t+s|t}$ not $\sum_{j=1}^{s}\Sigma_{t+j|t}$. The out-performance of these models is so strong at short horizons that cumulative comparisons are not useful. At the 1-day horizon, the Factor HEAVY model is dominant in both the joint and firm tests, with strong rejections for almost all pairs. The copula based comparisons are more mixed, although the null is rejected in favor of the Factor HEAVY for 12 of the 40 series, and only once in favor of cDCC. The results using 1-week forecasts are unsurprisingly weaker, although the Factor HEAVY is still preferred. Finally at the 2-week horizon, a similar pattern is found, with many statistically significant rejections both in the joint and firm volatility. The DM tests for the factor variance also uniformly prefer the HEAVY model, and the test statistic at the 1-day horizon is 3.72. The Factor HEAVY has bigger advantage over the long memory Riskmetrics model than over cDCC. In particular, at the 1-day horizon, Factor HEAVY forecasts are more accurate than Riskmetrics forecasts for 21 firms in the copula based comparison at the 10% significance level. Furthermore, there are still 23 firms for which the Factor HEAVY is preferred significantly in the joint test at the 1-week horizon.[9]

The ability of models to forecast conditional $\beta$ is also examined using the estimated idiosyncratic variance as the loss function. The conditional $\beta$ represents the optimal hedge ratio and so an accurate forecast should lead to a small tracking error. This leads to the loss

[9] The detailed result is omitted to save space, but available on request.

function

$$l_{i,t}^{s} = (r_{i,t+s} - \beta_{i,t+s|t} r_{f,t+s})^2.$$

| | 1 day | 1 week | 2 weeks | | 1 day | 1 week | 2 weeks |
|---|---|---|---|---|---|---|---|
| AET | 1.56 | **1.71** | **1.83** | HRB | **1.75** | **1.79** | **1.73** |
| AFL | **2.10** | **2.33** | **2.70** | JNS | **1.85** | **1.76** | 0.89 |
| AIG | 1.07 | 1.24 | 1.14 | JPM | -0.27 | 1.09 | 1.17 |
| ALL | -0.07 | 0.73 | 0.76 | KEY | 0.64 | 1.04 | 1.16 |
| AXP | -0.01 | 0.37 | 0.96 | LEH | -1.08 | -0.91 | -0.96 |
| BAC | -0.12 | 0.30 | 0.57 | LM | **1.92** | **1.92** | **2.23** |
| BBT | 1.12 | 1.20 | 1.36 | MER | **1.66** | **1.70** | 1.23 |
| BEN | 1.35 | **2.04** | **2.18** | MET | 0.60 | 0.66 | 1.24 |
| BK | 0.77 | 2.81 | **2.68** | MMC | 1.07 | 0.92 | 1.13 |
| C | 0.79 | 0.80 | 0.83 | MS | 1.35 | 1.13 | 0.61 |
| CB | 0.01 | 1.71 | **2.27** | PGR | 1.00 | 1.10 | 1.22 |
| CMA | 1.06 | 1.65 | **1.99** | PNC | -1.27 | -0.97 | 0.40 |
| COF | **1.70** | **2.31** | **2.60** | PRU | -0.91 | -0.62 | -0.65 |
| ETFC | -0.73 | 0.17 | 1.61 | SCHW | -0.65 | -0.61 | -0.78 |
| EV | -0.18 | 0.62 | 1.28 | SLM | **1.98** | **2.54** | **2.25** |
| FITB | -0.35 | 0.72 | 1.48 | STI | 0.87 | 1.55 | **1.83** |
| FNM | -0.76 | -0.72 | -0.29 | STT | 1.09 | 1.33 | **2.15** |
| FRE | -1.12 | -0.14 | 1.06 | TROW | -0.91 | -0.59 | 0.91 |
| GS | 0.82 | 0.06 | 0.10 | UNH | 1.11 | 0.88 | 1.33 |
| HIG | 1.30 | 0.85 | 1.17 | USB | -0.99 | -0.10 | 0.99 |

Table 4: Diebold-Mariano statistics for the null hypothesis that Factor HEAVY and cDCC forecasts are equally accurate when forecasting 1-day, 1-week and 2-week ahead $\beta$s. Numbers in bold font indicate that Factor HEAVY forecasts are more accurate than cDCC forecasts at the 10% significance level. Italic numbers indicate the converse.

Table 4 contains the result of DM tests of the null of equal idiosyncratic variance of the residual, comparing the Factor HEAVY to the cDCC. In most cases, the null cannot be rejected. However, when it is, the rejections indicate superior performance of the

Factor HEAVY. These results are weaker than in the QLIK tests, which may be due to heteroskedasticity of the idiosyncratic error. The statistical tests all indicate that the out-of-sample performance of the Factor HEAVY is superior to the leading low-frequency model. Moreover, the gains are particularly striking in terms of predicting variance. The gains in terms of predicting dependence, whether from the copula term which arises from the QLIK loss function, or from measuring the idiosyncratic variance, are smaller. We hypothesize that the gains to forecasting volatility may be larger since the range of conditional variances is much larger than that of conditional correlations or $\beta$s, which should aid in finding the superior model.

**Marginal Expected Shortfall Forecasting** Risk forecasting, and in particular systemic risk forecasting, is an important application of multivariate volatility models. Marginal expected shortfall, introduced by Acharya *et al.* (2012), measures the expected loss conditional on a factor being in a state of stress, and is defined as

$$\mathrm{MES}_{i,t} = -\mathrm{E}_t\left(r_{i,t+1} | r_{f,t+1} < c\right),$$

where $c$ is a threshold which determine whether the factor is showing signs of stress. The longer term $s$-step ahead MES is defined similarly but based on cumulative returns:

$$\mathrm{MES}^s_{i,t} = -\mathrm{E}_t\left(R_{i,t+1:s} | R_{f,t+1:s} < c_s\right),$$

where $R_{i,t+1:s} = \exp(\sum_{\tau=1}^{s} r_{i,t+\tau}) - 1$ is the cumulative return on firm equity from $t+1$ to $t+s$, $R_{f,t+1:s}$ is similarly defined and $c_s$ allows for explicit dependence between the horizon and the threshold for a stress event. We follow a similar procedure to that introduced in Brownlees & Engle (2012) to estimate MES (Please refer to the appendix for details). MES prediction accuracy is evaluated using Relative Mean Square Error (RMSE), which is used in place of MSE to control for strong heteroskedasticity of time-varying losses. The RMSE for $s$-step ahead MES is defined as

$$\mathrm{RMSE}_{i,t_c}^{s} = \left( \frac{-R_{i,t_c+1:s} - \mathrm{MES}_{i,t_c}^{s}}{\mathrm{MES}_{i,t_c}^{s}} \right)^2,$$

where the period from $t_c + 1$ to $t_c + s$ represents the event period of length $s$ during which the market experiences a loss larger than a pre-determined threshold. Since MES can only be evaluated when the market suffers a relatively large loss, few data points are available to evaluate the models.

Relative predictive accuracy is assessed using Diebold-Mariano tests where the loss differential is $\delta_{i,t_c} = \mathrm{RMSE}_{i,t_c}^{cDCC} - \mathrm{RMSE}_{i,t_c}^{HEAVY}$. The test statistic is

$$\sqrt{T_c} \frac{\overline{\delta}}{\sqrt{\mathrm{avar}\left(\delta_t\right)}},$$

where $T_c$ denotes the number of event days.

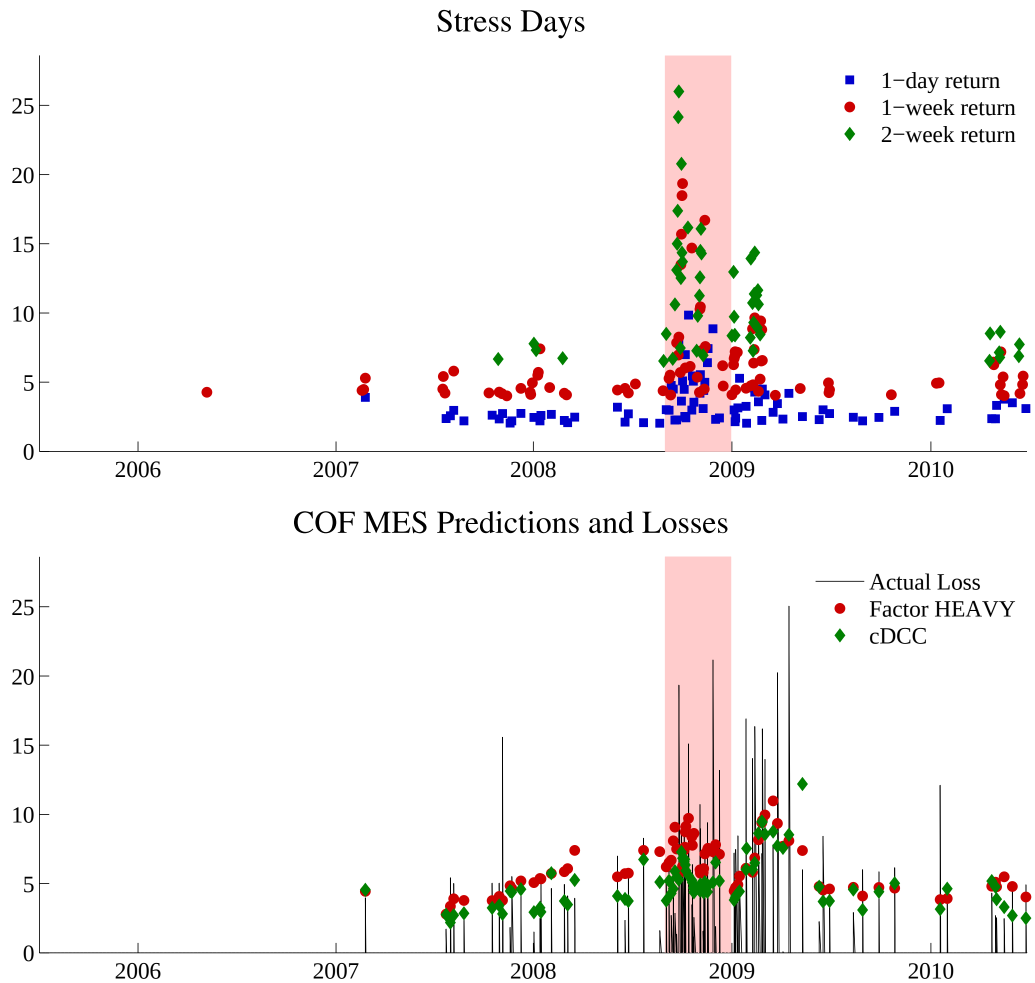


Figure 5: Event days and the beginning days of event periods from July 2005 and the corresponding market index losses. Event days are the days on which the daily market index return declines by at least 2%. Event periods are those during which the market index experiences more than the corresponding threshold loss. Actual equity losses of Capital One (COF) and 1-day MES forecasts on event days from July 2005. Solid lines denote the actual losses; circle and star markers represent Factor HEAVY and cDCC forecasts, respectively.

Figure 5 shows the actual event days and event periods and the corresponding market index losses. The corresponding threshold market index losses are 2%, 4% and 6.5%, respectively.[10] Under these settings, there are 84 event days, 90 1-week event periods and 51 2-week event periods in the sample. Figure 5 compares MES forecasts with the losses for COF on event days.[11] While it is not clear from the figure which model performs better on average, Factor HEAVY forecasts are more accurate at the 10% significance level. The figure shows that each model can track actual losses well on some event days. At the crisis peak around autumn 2008, however, Factor HEAVY forecasts outperform cDCC forecasts where cDCC forecasts tend to seriously underestimate the actual losses. Forecasts from Factor HEAVY models respond quickly to the shock while forecasts from cDCC models adjust slowly. This fact may provide insight into the source of the forecasting gains of Factor HEAVY models.

---

[10] These values are chosen since larger values of the threshold index require extremely large bootstrap samples, especially in the period when volatility is low.

[11] The bandwidth in kernel estimation is set equal to 0.05. This value is chosen to give stable results.

| | 1 day | 1 week | 2 weeks | | 1 day | 1 week | 2 weeks |
|---|---|---|---|---|---|---|---|
| AET | **3.80** | **2.34** | **2.92** | HRB | 0.04 | **1.98** | **2.74** |
| AFL | **2.83** | 0.81 | 1.41 | JNS | **2.58** | **2.02** | **2.00** |
| AIG | *-3.17* | -1.48 | 0.31 | JPM | 0.89 | 0.29 | -1.01 |
| ALL | 0.20 | -0.73 | *-1.84* | KEY | *-2.22* | -1.58 | *-2.38* |
| AXP | 0.56 | -0.99 | *-1.83* | LEH | -0.96 | -1.21 | -1.39 |
| BAC | *-2.82* | *-1.72* | *-1.75* | LM | **2.71** | **2.08** | 1.19 |
| BBT | 1.31 | -0.51 | 0.78 | MER | -1.51 | -1.19 | -0.44 |
| BEN | 0.79 | -1.14 | -0.70 | MET | -1.28 | -1.19 | *-2.05* |
| BK | **2.72** | 0.73 | 0.04 | MMC | **2.95** | **2.70** | **2.82** |
| C | -1.60 | *-1.87* | *-2.44* | MS | 0.65 | -1.07 | *-2.32* |
| CB | 1.52 | 1.33 | 0.18 | PGR | 0.89 | 0.76 | 1.10 |
| CMA | **1.95** | 0.94 | 1.54 | PNC | 0.13 | -1.16 | -1.01 |
| COF | **3.16** | **2.19** | **2.26** | PRU | -1.34 | -1.64 | *-2.17* |
| ETFC | -0.43 | -0.21 | -1.61 | SCHW | **3.16** | **1.95** | **2.27** |
| EV | -0.67 | -1.09 | -1.62 | SLM | 1.00 | -1.03 | **3.36** |
| FITB | -1.61 | *-1.72* | *-1.94* | STI | -1.57 | *-1.98* | *-2.36* |
| FNM | -1.36 | -0.92 | 1.22 | STT | 1.16 | 0.24 | -1.05 |
| FRE | *-2.98* | 0.36 | -1.12 | TROW | -0.35 | -1.00 | -0.66 |
| GS | 0.98 | 0.71 | 0.46 | UNH | **1.66** | **2.81** | **2.47** |
| HIG | -1.18 | -1.10 | 0.99 | USB | -0.14 | *-2.05* | -1.48 |

Table 5: Diebold-Mariano statistics associated with the 1-day, 1-week and 2-week MES forecasts. Numbers in bold font indicate that Factor HEAVY forecasts are more accurate than cDCC forecasts at the 10% significance level for this firm. Italic numbers indicate the superior performance by cDCC forecasts.

Table 5 presents the Diebold-Mariano statistics associated with the 1-day, 1-week and 2-week MES forecasts.[12] 1-day Factor HEAVY forecasts are more accurate than the cDCC forecasts for 23 firms, and 10 of these are significant at the 10% level. On the other hand, cDCC forecasts are significantly more accurate for 4 firms. When looking at the 1-week

[12] When doing bootstrapping for $s$-step forecasting, we fix the number of remaining simulation paths to 1000 which satisfy $R_{f,t+1:s} < c_s$ . Under this setting, the MES simulation result was found to be stable.

MES, Factor HEAVY still enjoys an advantage over cDCC in terms of the number of significant forecasts, but the advantage is somewhat smaller than for 1-day-ahead forecasts. The gains from using Factor HEAVY models, however, are further reduced at the 2-week horizon, although for some firms like COF, Factor HEAVY forecasts are significantly more accurate than cDCC forecasts at all forecast horizons of interests.

It is noted that COF has consistently good performance when forecasting volatility, covolatility, $\beta$ and MES. Since MES is the expected equity loss of the firm conditional on a threshold market index loss, we suspect that these all play a role in the 1-day MES gains, although the performance in MES forecasting is stronger in models where the Factor HEAVY performed better in terms of the copula. Factor HEAVY models are significantly more accurate when forecasting both the MES and copula for the firms such as AET, AFL, CMA, COF, JNS, LM, MMC and UNH. This relation between MES and copula forecasting remains but weakens at longer horizons.

### 2.4.4 Extensions

Three extensions to the one-factor HEAVY model are explored. The first examines the gains to including multiple factors. The second examines the effect of changing the realized measure. The third examines the evidence for including asymmetries in the model.

**Multiple Factors** We consider two specifications for the multiple-factor models. In the first specification, we have prior knowledge of the number of factors $K$, and one of the factors is an observable market proxy. The one-factor specification is therefore augmented to a $K$-factor structure where the additional factors are constructed industry portfolios. The portfolio weights are computed using principle component analysis on CAPM residuals on the daily data. Define $\hat{\varepsilon}_{i,t} = r_{i,t} - \hat{\beta}_i r_{f,t}$ to be CAPM residuals where $\hat{\beta}_i$ is the usual OLS estimator. Portfolios are then formed using the $N \times (K-1)$ weight matrix $W$ which corresponds to eigenvectors associated with the $K-1$ largest eigenvalues in the outer-product of $\hat{\Omega} = T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$. The new factors have daily returns $r_{PCA,t} = \sum W_i \hat{\varepsilon}_{i,t}$ where $W_i$ is the $i$th row of $W$. High-frequency returns are similarly constructed using the intra-daily data. Finally, realized covariances are computed using the high frequency returns of the constructed portfolios, along with the market and each individual firm.

Alternatively, we can assume that the number of factors is unknown and all factors are unobserved. In this specification, various methods in the literature of approximate factor models can be employed to estimate the number of factors (See Choi & Jeong (2013) for a review). In the empirical discussion below, we will apply Onatski (2010)'s estimator based on the differenced eigenvalues, which is found to be more accurate than Bai & Ng (2002)'s information criteria when the idiosyncratic errors are substantially correlated. Let $R$ indicate the $T \times N$ matrix of daily returns. The estimator of the number of factors is specifically $\hat{K} = \max\{K \leq K_{max}, \varpi_k - \varpi_{k+1} \geq c\}$ where $\varpi_k$ is the $k$th largest eigenvalue of $R'R/T$, $c$ is a threshold value estimated from the empirical distribution of eigenvalues (see Onatski (2010) for the calibration algorithm), and $K_{max}$ is a given upper bound which is set equal to 10 in our empirical discussion. As in the first specification, the factors for daily returns are portfolios formed using principal component analysis on the daily returns. Specifically, the $N \times K$ weight matrix $\tilde{W}$ corresponds to eigenvectors associated with the $K$ largest eigenvalues of $R'R/T$. The daily returns of the factors are $\tilde{r}_{PCA,t} = \sum \tilde{W}_i r_{i,t}$. High-frequency returns of the factors are constructed using the intra-daily returns of the firms in the similar way, given the previously estimated number of factors.

Empirically, we present the results of a multiple-factor model with the second specification for the 40 firms.[13] The estimated number of factors is 2 by Onatski (2010)'s method. Table 6 shows that the dispersion of the estimates in the $\beta$ and idiosyncratic volatility equations is similar to that in the one-factor model in Table 1. The fit is also examined based on the marginal log-likelihood of the firm. We find that the model with 2 factors improves univariate fit for 26 out of the 40 firms. When forecasting 1-day ahead volatility of firm equity returns, 2-factor model forecasts are more accurate than 1-factor model forecasts for 31 firms, and 14 of these are significant at the 10% level, while 1-factor model forecasts are significant for only 1 firm. In sum, the multiple-factor models can have more successful in-sample and out-of-sample performance.

**Alternative Realized Measures** As a robustness check, the model is fit using alternative realized measures. The top panel of table 7 reports parameter estimates and the (joint) log-likelihood across 5 choices of realized measure: realized covariance sampled between 5 and 30 minutes, all using 1-minute subsampling, and multivariate realized kernels. The second panel reports DM test statistics from 1-day forecasts for COF against the cDCC model using the same measures. All results are stable across alternative realized measures, and realized measures which sample more often perform somewhat better both in sample and out-of sample, with realized kernel producing the best in-sample fit. All key parameters which measure the sensitivity to the realized measure in the HEAVY-P equations,

[13] If we still use S&P 500 and SPY as the market proxy, the results of the first model specification is close to those of the second model specification, as the returns of S&P 500 and SPY are highly correlated with the corresponding returns of the first estimated factors.

$\theta_1$ (market) , $\delta_{i,1}$ (factor loading) and $\alpha_{i,1}$ (idiosyncratic variance), decline as the realized measure deteriorates, which is consistent with additional measurement noise when sampling less-frequently. All estimators consistently out-perform cDCC forecasts, and a similar pattern appears in terms of using more precise estimators.

| | Factor | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $A_{11}$ | $A_{22}$ | $B_{11}$ | $B_{22}$ | $A^M_{11}$ | $A^M_{22}$ | $B^M_{11}$ | $B^M_{22}$ | | | | |
| | 0.42 | 0.35 | 0.90 | 0.94 | 0.62 | 0.36 | 0.79 | 0.93 | | | | |
| | $\beta$ | | | | | | | | Idiosyncratic | | | |
| | $\delta_{i,1,1}$ | $\delta_{i,2,1}$ | $\delta_{i,1,2}$ | $\delta_{i,2,2}$ | $\delta^M_{i,1,1}$ | $\delta^M_{i,1,1}+\delta^M_{i,2,1}$ | $\delta^M_{i,1,2}$ | $\delta^M_{i,1,2}+\delta^M_{i,2,2}$ | $\alpha_{i,1}$ | $\alpha_{i,2}$ | $\alpha^M_{i,1}$ | $\alpha^M_{i,1}+\alpha^M_{i,2}$ |
| Min | 0.01 | 0.14 | 0.00 | 0.00 | 0.02 | 0.97 | 0.00 | 0.72 | 0.03 | 0.00 | 0.23 | 0.92 |
| $Q_{.25}$ | 0.07 | 0.82 | 0.01 | 0.91 | 0.10 | 0.99 | 0.01 | 0.92 | 0.47 | 0.22 | 0.39 | 1.00 |
| Median | 0.10 | 0.90 | 0.02 | 0.97 | 0.12 | 0.99 | 0.02 | 0.94 | 0.60 | 0.37 | 0.47 | 1.00 |
| Mean | 0.12 | 0.84 | 0.05 | 0.92 | 0.11 | 0.99 | 0.02 | 0.97 | 0.58 | 0.38 | 0.46 | 1.00 |
| $Q_{.75}$ | 0.16 | 0.92 | 0.08 | 0.99 | 0.13 | 0.99 | 0.03 | 0.98 | 0.70 | 0.47 | 0.52 | 1.00 |
| Max | 0.39 | 0.98 | 0.19 | 1.00 | 0.19 | 1.00 | 0.09 | 1.00 | 1.00 | 0.97 | 0.66 | 1.00 |

Table 6: Cross-sectional statistics of full-sample parameter estimates of a two-factor HEAVY model with unobserved factors for the 40 financial firms.

The bottom three panels of table 7 show the relative distribution of parameters using the 25th quantile, median and 75th quantile of the estimates across all 40 assets, as well as the distribution of likelihood differences relative to the likelihood of the 10-minute realized covariance-based model (negative indicates better performance of the 10-minute model). These are broadly consistent with the pattern for COF, where less frequent sampling leads to smaller sensitivity to news as well as smaller likelihoods. The performance of the 5-minute and 10-minute realized covariance is extremely similar.

| | In-Sample | | | | |
|---|---|---|---|---|---|
| | 5-minute | 10-minute | 15-minute | 30-minute | Kernel |
| $\theta_1$ | 0.28 | 0.26 | 0.26 | 0.24 | 0.28 |
| $\delta_{i,1}$ | 0.12 | 0.10 | 0.10 | 0.09 | 0.13 |
| $\alpha_{i,1}$ | 0.64 | 0.62 | 0.60 | 0.52 | 0.63 |
| Likelihood | -9125 | -9138 | -9163 | -9229 | -9115 |

| | Out-of-sample | | | | |
|---|---|---|---|---|---|
| | 5-minute | 10-minute | 15-minute | 30-minute | Kernel |
| Joint | 5.93 | 5.90 | 6.07 | 6.56 | 6.03 |
| Factor | 3.57 | 3.72 | 3.74 | 3.99 | 3.74 |
| Firm | 3.67 | 3.70 | 3.75 | 3.89 | 3.77 |
| Copula | 2.36 | 2.35 | 2.54 | 3.09 | 2.42 |
| $\beta$ | 1.81 | 1.70 | 1.54 | 1.19 | 1.78 |
| MES | 2.89 | 3.16 | 3.12 | 2.96 | 2.71 |

5-miunte realized covariance

| | $\delta_{i,1}$ | $\delta_{i,2}$ | $\delta_{i,1}^M$ | $\alpha_{i,1}$ | $\alpha_{i,2}$ | $\alpha_{i,1}^M$ | $\Delta L$ |
|---|---|---|---|---|---|---|---|
| $Q_{.25}$ | 0.07 | 0.74 | 0.15 | 0.41 | 0.25 | 0.41 | -2.94 |
| Median | 0.11 | 0.86 | 0.18 | 0.54 | 0.38 | 0.48 | -1.00 |
| $Q_{.75}$ | 0.17 | 0.91 | 0.21 | 0.69 | 0.48 | 0.52 | 3.78 |

15-minute realized covariance

| | $\delta_{i,1}$ | $\delta_{i,2}$ | $\delta_{i,1}^M$ | $\alpha_{i,1}$ | $\alpha_{i,2}$ | $\alpha_{i,1}^M$ | $\Delta L$ |
|---|---|---|---|---|---|---|---|
| $Q_{.25}$ | 0.06 | 0.85 | 0.10 | 0.35 | 0.33 | 0.35 | -4.27 |
| Median | 0.08 | 0.90 | 0.12 | 0.45 | 0.49 | 0.41 | -1.99 |
| $Q_{.75}$ | 0.11 | 0.92 | 0.14 | 0.59 | 0.59 | 0.47 | 1.46 |

30-minute realized covariance

| | $\delta_{i,1}$ | $\delta_{i,2}$ | $\delta_{i,1}^M$ | $\alpha_{i,1}$ | $\alpha_{i,2}$ | $\alpha_{i,1}^M$ | $\Delta L$ |
|---|---|---|---|---|---|---|---|
| $Q_{.25}$ | 0.05 | 0.90 | 0.06 | 0.24 | 0.46 | 0.29 | -19.33 |
| Median | 0.06 | 0.92 | 0.08 | 0.34 | 0.61 | 0.34 | -14.79 |
| $Q_{.75}$ | 0.08 | 0.94 | 0.10 | 0.53 | 0.71 | 0.40 | -5.97 |

Table 7: Parameter estimates, in-sample likelihood and Diebold-Mariano statistics of 1-day forecasts for COF using different realized measures, i.e. 5-minute, 10-minute, 15-minute and 30-minute realized covariance matrices with 1-minute subsampling, as well as realized kernels.

**Conditional Asymmetries in Volatility** The third extension is to examine whether asymmetries are needed in the model. GJR-GARCH and TARCH models have been broadly found to fit conditional variances better than symmetric GARCH models (Glosten *et al.* 1993; Zakoian 1994). A GJR-like specification can be constructed in the HEAVY-P using an indicator variable based on the daily return,

$$\sigma_{f,t}^2 = \theta_0 + \theta_1 RM_{f,t-1} + \theta_2 \sigma_{f,t-1}^2 + \theta_3 RM_{f,t-1} I(r_{f,t-1} < 0). \tag{23}$$

We test the null that $\theta_3 = 0$ which is strongly rejected. Without the asymmetry (imposing $\theta_3 = 0$), $\hat{\theta}_1 = 0.26$. When the asymmetry is introduced, $\hat{\theta}_1 = 0.12$ and $\hat{\theta}_3 = 0.19$. These coefficient changes are large, and the full-sample marginal log-likelihood increases from -3674 to -3611.

Following Braun *et al.* (1995), asymmetries are added to the conditional factor loading through three terms,

$$\begin{aligned}\beta_{i,t} &= \delta_{i,0} + \delta_{i,1} R\beta_{i,t-1} + \delta_{i,2}\beta_{i,t-1} + \delta_{i,3} R\beta_{i,t-1} I(r_{f,t-1} < 0)\\ &\quad + \delta_{i,4} R\beta_{i,t-1} I(\varepsilon_{i,t-1} < 0) + \delta_{i,5} R\beta_{i,t-1} I(r_{f,t-1}\varepsilon_{i,t-1} < 0),\end{aligned}$$

where $\delta_{i,3}$ allows for asymmetries due to the market return, $\delta_{i,4}$ allows asymmetries through the idiosyncratic shock and $\delta_{i,5}$ allows asymmetries through the interaction of the market return and the idiosyncratic shock. This specification is fit to all 40 firms and the null $\delta_{i,3} = \delta_{i,4} = \delta_{i,5} = 0$ is tested using a Wald test with a sandwich variance-covariance for the parameters. The null is rejected for 10 out of the 40, indicating mild evidence of asymmetries in $\beta$. Finally, conditional asymmetries are added to the idiosyncratic variance

$$\sigma_{i,t}^2 = \alpha_{i,0} + \alpha_{i,1} RIV_{i,t-1} + \alpha_{i,2}\sigma_{i,t-1}^2 + \alpha_{i,3} RIV_{i,t-1} I(r_{f,t-1} < 0) + \alpha_{i,4} RIV_{i,t-1} I(\varepsilon_{i,t-1} < 0),$$

where $\alpha_{i,3}$ allows for asymmetries due to the factor return and $\alpha_{i,4}$ allows for asymmetries due to the sign of the idiosyncratic shock. When tested with a Wald test, only 5 of the 40 series reject the null $\alpha_{i,3} = \alpha_{i,4} = 0$, indicating that models for the idiosyncratic variance do not require conditional asymmetries. These results contrast sharply with what is typically found in standard low-frequency models, such as the conditional variance models in the cDCC. Low-frequency GJR-GARCH models,

$$\tilde{\sigma}_{i,t}^2 = \omega_i^D + \alpha_i^D r_{i,t-1}^2 + \beta_i^D \tilde{\sigma}_{i,t-1}^2 + \gamma_i^D r_{i,t-1}^2 I\left(r_{i,t-1} < 0\right),$$

are fit to the market return and the individual assets. In 35 of the 41 models fit, the estimate of $\gamma_i^D$ is statistically significant. Moreover, in the market model, the estimate of $\alpha_f^D$ drops from 0.08 to 0.00 $\left(\hat{\gamma}_f^D = 0.14\right)$ when the asymmetric term is introduced. The median estimate of $\alpha_i^D$ is 0.09 when no asymmetry is included in the 40 individual firms; the introduction of asymmetries produces median estimates of 0.02 for $\alpha_i^D$ and 0.11 for $\gamma_i^D$. This difference between the low-frequency models and the high-frequency models is due to the decomposition which separates the market, which has a strong asymmetry from the idiosyncratic volatility, which does not. This additional parsimony is not applicable in standard low- (or high-) frequency cDCC-type models since the conditional variance estimated includes both components.

## 2.5 Conclusions

The chapter introduces a new class of multivariate high-frequency based volatility models utilizing a factor structure for both the high- and low-frequency conditional covariances. This structure allows for parsimonious models to be fit while simplifying the task of incorporating high-frequency realized measures. The factor volatility, $\beta$ and idiosyncratic volatility are modeled as separate HEAVY-type processes and the daily shocks in low-frequency models are replaced with the corresponding realized measures. This modification is supported by the empirical evidence that realized measures dominate daily measures for most equities. The Factor HEAVY is also validated as a covariance model using the specification tests. We show that this model improves on existing multivariate HEAVY models, and are are especially useful for a large panels of assets.

In an empirical analysis of 40 systematically important U.S. financial firms, we find a one-factor model is adequate to capture the dynamics of the conditional covariance. The Factor HEAVY models perform better both in- and out-of-sample when compared to a leading low-frequency model. We find that Factor HEAVY models have a large advantage over cDCC in covariance matrix forecasting, and that the advantage appears in a variety of dimensions including the forecast copula, $\beta$ and MES. The out-of-sample forecast ability is more pronounced at shorter forecast horizons which are the most relevant from a risk-management point-of-view. We consider extensions to multiple-factor models, which is shown to be empirically more successful than the one-factor model. Furthermore, we check the robustness of the performance to some selected realized measures. We also examine a number of extensions to the base model including the role of asymmetries in the conditional variance, $\beta$ and idiosyncratic variance, where we find that asymmetries are only needed for the conditional variance of the market.

## 2.6 Appendix

### 2.6.1 The Sample of Financial Firms

| | | | |
|---|---|---|---|
| AET | Aetna | HRB | H&R Block |
| AFL | Aflac | JNS | Janus Capital |
| AIG | American International Group | JPM | JP Morgan Chase |
| ALL | Allstate Corp | KEY | Keycorp |
| AXP | American Express | LEH | Lehman Brothers |
| BAC | Bank of America | LM | Legg Mason |
| BBT | BB&T | MER | Merrill Lynch |
| BEN | Franklin Resources | MET | Metlife |
| BK | Bank of New York Mellon | MMC | Marsh & McLennan |
| C | Citigroup | MS | Morgan Stanley |
| CB | Chubb Corp | PGR | Progressive |
| CMA | Comerica inc | PNC | PNC Financial Services |
| COF | Capital One Financial | PRU | Prudential Financial |
| ETFC | E-Trade Financial | SCHW | Schwab Charles |
| EV | Eaton Vance | SLM | SLM Corp |
| FITB | Fifth Third Bancorp | STI | Suntrust Banks |
| FNM | Fannie Mae | STT | State Street |
| FRE | Freddie Mac | TROW | T. Rowe Price |
| GS | Goldman Sachs | UNH | Unitedhealth Group |
| HIG | Hartford Financial Group | USB | US Bancorp |

### 2.6.2 Data Transformation

**Transformed Realized Measures over the Entire Day** The modified realized measure is constructed as

$$\widetilde{RM}_t = \hat{\Lambda} RM_t \hat{\Lambda}',$$

where $\hat{\Lambda} = \overline{\Sigma}^{1/2}\overline{M}^{-1/2}$, $\overline{\Sigma} = T^{-1}\sum_{t=1}^{T} r_t r_t'$ and $\overline{M} = T^{-1}\sum_{t=1}^{T} RM_t$. This transformation has some limitations, and in particular the matrix square-root of $\overline{\Sigma}$ may not be precisely estimated. Since our preferred estimator only uses the realized measure of the factor(s) and the individual assets – but not the realized measure between the individual assets – we can define a similar transformation using the $K+1$ by $K+1$ realized measures required by the estimator,

$$\widetilde{RM}_{i,t} = \hat{\Lambda}_i RM_{i,t} \hat{\Lambda}_i', \tag{24}$$

where $\hat{\Lambda}$ is estimated using only the $K$ factors and asset $i$, and

$$RM_{i,t} = \begin{bmatrix} RM_{f,t} & RM_{fi,t} \\ RM_{if,t} & RM_{ii,t} \end{bmatrix}.$$

**Transformed Realized Measures with the Same Estimate of the Quadratic Covariation of the Factors** Begin by partitioning the individual realized measure

$$RM_{i,t} = \begin{bmatrix} RM^{i}_{f,t} & RM_{fi,t} \\ RM_{if,t} & RM_{ii,t} \end{bmatrix},$$

where the notation $RM^i_{f,t}$ is used to indicate that this estimate of the quadratic covariation of the factors is specific to the noise-robust measure that includes asset $i$. Our solution is to use a common estimator of the quadratic covariation of the factors, and then to enforce that this common estimator is the same in all component realized measures. A natural estimator would be the noise robust estimator (e.g. realized kernel) applied only to the factors on day $t$. Denote that estimator $RM^C_{f,t}$ where the $C$ denotes common. The modified estimators of quadratic covariation are then

$$\widehat{RM}_{i,t} = \Pi_{i,t} RM_{i,t} \Pi'_{i,t},$$

where

$$\Pi_{i,t} = \begin{bmatrix} \left(RM^C_{f,t}\right)^{1/2} \left(RM^i_{f,t}\right)^{-1/2} & 0 \\ 0 & 1 \end{bmatrix}.$$

When there is a single factor, the transformation matrix takes a simple form where

$$\begin{aligned} \widehat{RM}_{i,t} &= \begin{bmatrix} \sqrt{\frac{RM^C_{f,t}}{RM^i_{f,t}}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} RM^i_{f,t} & RM_{fi,t} \\ RM_{if,t} & RM_{ii,t} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{RM^C_{f,t}}{RM^i_{f,t}}} & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} RM^C_{f,t} & RM_{fi,t}\sqrt{\frac{RM^C_{f,t}}{RM^i_{f,t}}} \\ RM_{if,t}\sqrt{\frac{RM^C_{f,t}}{RM^i_{f,t}}} & RM_{ii,t} \end{bmatrix}. \end{aligned}$$

Note that if a single realized measure is used across all assets, such as a single multivariate realized kernel (or realized covariance) to estimate the integrated covariance, then

$\left(RM^{C}_{f,t}\right)^{1/2}\left(RM^{i}_{f,t}\right)^{-1/2} = I_K$ and the estimator is unmodified.

### 2.6.3 Stationarity

The model structure leads to a relatively simple condition to determine whether the model is covariance stationary since both the HEAVY-P and the HEAVY-M components are driven exclusively by the realized measures. In particular, the coefficients on the realized measures in the HEAVY-P play no role in the stationarity since these simply act as sensitivities and there is no feed-back from the low-frequency observations into the realized measures.

From the properties of the Wishart distribution, all of the elements can be decomposed into their conditional expectation and a white noise error, so that $RM_{f,t} = \mu_{f,t} + \tilde{\eta}_{f,t}$, $R\beta_{i,t} = \lambda_{i,t} + \tilde{\tau}_{i,t}$, and $RIV_{i,t} = \frac{\nu-1}{\nu}\mu_{i,t} + \tilde{\iota}_{i,t}$ . Letting $v_t = (0,0,0,\tilde{\eta}_{f,t},\tilde{\tau}_{i,t},\tilde{\iota}_{i,t})'$, a mean-zero white-noise vector, the standard Factor HEAVY have a VARMA(1,1) representation. Specifically,

$$\tilde{a}_t = c + B\tilde{a}_{t-1} + v_t - Dv_{t-1}, \tag{25}$$

where

$$\tilde{a}_t = \begin{bmatrix} \sigma^2_{f,t} \\ \beta_{i,t} \\ \sigma^2_{i,t} \\ RM_{f,t} \\ R\beta_{i,t} \\ RIV_{i,t} \end{bmatrix}, c = \begin{bmatrix} \theta_0 \\ \delta_{i,0} \\ \alpha_{i,0} \\ \theta_0^M \\ \delta_{i,0}^M \\ \frac{\nu-1}{\nu}\alpha_{i,0}^M \end{bmatrix}, D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_2^M & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{i,2}^M & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{i,2}^M \end{bmatrix},$$

and

$$B = \begin{bmatrix} \theta_2 & 0 & 0 & \theta_1 & 0 & 0 \\ 0 & \delta_{i,2} & 0 & 0 & \delta_1 & 0 \\ 0 & 0 & \alpha_{i,2} & 0 & 0 & \alpha_1 \\ 0 & 0 & 0 & \theta_1^M + \theta_2^M & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{i,1}^M + \delta_{i,2}^M & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\nu-1}{\nu}\alpha_{i,1}^M + \alpha_{i,2}^M \end{bmatrix}.$$

The covariance stationarity for VARMA(1,1) requires that the eigenvalues of $B$ are all less than one in modulus. That is, the standard Factor HEAVY with a single factor is covariance stationary if $0 \leq \theta_2 < 1$, $0 \leq \delta_{i,2} < 1$, $0 \leq \alpha_{i,2} < 1$, $0 \leq \theta_1^M + \theta_2^M < 1$, $0 \leq \delta_{i,1}^M + \delta_{i,2}^M < 1$, $0 \leq \frac{\nu-1}{\nu}\alpha_{i,1}^M + \alpha_{i,2}^M < 1$.

### 2.6.4 Estimation of Marginal Expected Shortfall

The 1-step MES is particularly simple to compute, and is

$$\begin{aligned}
\mathrm{MES}_{i,t} &= -\mathrm{E}_t\left(\beta_{i,t+1} r_{f,t+1} + \varepsilon_{i,t+1} \middle| r_{f,t+1} < c\right) \\
&= -\beta_{i,t+1}\sigma_{f,t+1}\mathrm{E}_t\left(\xi_{f,t+1} \middle| \xi_{f,t+1} < {}^{c}/_{\sigma_{f,t+1}}\right) - \sigma_{i,t+1}\mathrm{E}_t\left(\xi_{i,t+1} \middle| \xi_{f,t+1} < {}^{c}/_{\sigma_{f,t+1}}\right).
\end{aligned}$$

In the cDCC model, $\beta_{i,t} = \tilde{\sigma}_{i,t}\rho_t/\tilde{\sigma}_{f,t}$, $\sigma_{i,t} = \tilde{\sigma}_{i,t}\sqrt{1-\rho_t^2}$ and $\xi_{i,t} = \left(r_{i,t} - \beta_{i,t} r_{f,t}\right)/\sigma_{i,t}$. $\mathrm{E}_t\left(\xi_{f,t+1} \middle| \xi_{f,t+1} < {}^{c}/_{\sigma_{f,t+1}}\right)$ and $\mathrm{E}_t\left(\xi_{i,t+1} \middle| \xi_{f,t+1} < {}^{c}/_{\sigma_{f,t+1}}\right)$ are estimated using a nonparametric kernel estimation similar to that in Brownlees & Engle (2012),

$$\widehat{\mathrm{E}}_t\left(x_{t+1}|\xi_{f,t+1} < {}^{c}/_{\sigma_{f,t+1}}\right) = \frac{\sum_{\tau=1}^{t} K_l\left(\frac{c}{\sigma_{f,t+1}} - \xi_{f,\tau}\right)x_\tau}{\sum_{\tau=1}^{t} K_l\left(\frac{C}{\sigma_{f,t+1}} - \xi_{f,\tau}\right)}, \tag{26}$$

where $x$ is either $\xi_f$ or $\xi_i$, $K_l(z) = \int_{-\infty}^{z/l} k(u)du$ , $k(\cdot)$ is the Gaussian kernel and $l$ is the bandwidth.

Longer-horizon MES forecasting is more complex and requires the use of a bootstrap to simulate returns. The return simulation is constructed using standardized residuals, and, in the Factor HEAVY, standardized realized measures. The $s$-step MES requires an estimate of the expected (cumulative) loss conditional on the factor suffering a large (cumulative) loss. In the Factor HEAVY, we construct the tuple of standardized innovations

$$\nu_t = \left(\xi_{f,t}, \xi_{i,t}, \mathrm{vech}\left(\Xi_{i,t}\right)'\right)',$$

where

$$\Xi_t = M_{i,t}^{-1/2} \left(RM_{i,t}\right) M_{i,t}^{-1/2}.$$

The values of the tuple $\nu_t$ are then used to simulate from the Factor HEAVY for a longer horizon, and simulated cumulative returns are constructed for both the factor and the asset. The $s$-step MES is then constructed as

$$\text{MES}_{i,t}^{s} = -\frac{\sum_{b=1}^{B} \hat{R}_{i,t+1:s}^{b} I\left[\hat{R}_{f,t+1:s}^{b} < c_s\right]}{\sum_{b=1}^{B} I\left[\hat{R}_{f,t+1:s}^{b} < c_s\right]}, \tag{27}$$

where $I[x] = 1$ if condition $x$ is satisfied and equal to zero otherwise and $\hat{R}_{i,t+1:s}$ is used to denote that these are simulated returns. This is repeated across all assets using only pair-wise measures. In the cDCC, the standardized innovations include only $\left(\xi_{f,t}, \xi_{i,t}\right)$ where the standardization is computed using the cDCC-fit values of the conditional covariance. The simulation procedure is otherwise identical. The entire procedure is recursively implemented so that all estimates including both the parameters of the conditional covariance models as well as the other inputs to the MES are “real-time”.

### 2.6.5 Proofs

**Proof of Proposition 1** A $k \times k$ random symmetric positive definite matrix $S$ is said to follow a Wishart distribution $W_k(\nu, \Sigma)$ with the degree of freedom $\nu$ and scale matrix $\Sigma$ if the density is given by

$$\frac{|S|^{\frac{\nu-k-1}{2}}}{2^{\frac{\nu k}{2}}\Gamma_k(\frac{\nu}{2})|\Sigma|^{\frac{\nu}{2}}}\exp(-\frac{\mathrm{tr}(\Sigma^{-1}S)}{2}),$$

where $\Gamma_k(\frac{\nu}{2})=\pi^{k(k-1)/4}\Pi_{i=1}^{k}\Gamma(\frac{\nu+1-i}{2})$ is the multivatiate gamma function.

As we assume $RM_t|\mathscr{F}_{t-1}^{\mathrm{HF}}\sim W_{N+1}(\nu,{}^{M_t}/\nu)$, the $2\times 2$ reailzed covariance matrix

$$RM_{i,t}|\mathscr{F}_{t-1}^{\mathrm{HF}}=\begin{bmatrix} RM_{f,t} & RM_{fi,t} \\ RM_{if,t} & RM_{ii,t}\end{bmatrix}|\mathscr{F}_{t-1}^{\mathrm{HF}}\sim W_2(\nu,\frac{M_{i,t}}{\nu}),$$

where $M_{i,t}=\begin{bmatrix} \mu_{f,t} & \lambda_{i,t}\mu_{f,t} \\ \lambda_{i,t}\mu_{f,t} & \lambda_{i,t}^2\mu_{f,t}+\mu_{i,t}\end{bmatrix}$. It follows that (Gupta & Nagar, 2000)

$$R\beta_{i,t}|\mathscr{F}_{t-1}^{\mathrm{HF}}=\frac{RM_{fi,t}}{RM_{f,t}}|\mathscr{F}_{t-1}^{\mathrm{HF}}\sim T_{1,1}(\nu,\lambda_{i,t},\mu_{f,t}^{-1},\mu_{i,t})$$

$$RIV_{i,t}|\mathscr{F}_{t-1}^{\mathrm{HF}}=(RM_{ii,t}-\frac{RM_{fi,t}^2}{RM_{f,t}})|\mathscr{F}_{t-1}^{\mathrm{HF}}\sim W_1(\nu-1,\frac{\mu_{i,t}}{\nu}),$$

where $T\sim T_{p,m}(\nu,M,\Sigma,\Omega)$ is the matrix variate $t$ distribution with the density written as

$$\frac{\Gamma_p(\frac{\nu+m+p-1}{2})}{\pi^{\frac{mp}{2}}\Gamma_p(\frac{\nu+p-1}{2})}|\Sigma|^{-\frac{m}{2}}|\Omega|^{-\frac{p}{2}}|I_p+\Sigma^{-1}(T-M)\Omega^{-1}(T-M)'|^{-\frac{\nu+m+p-1}{2}}.$$

Since $\mathrm{E}[\mathrm{S}]=\nu\Sigma$ and $\mathrm{E}[\mathrm{T}]=M$ if $S\sim W_k(\nu,\Sigma)$ and $T\sim T_{p,m}(\nu,M,\Sigma,\Omega)$,

$$\mathrm{E}_{t-1}[R\beta_{i,t}]=\lambda_{i,t}$$

$$\mathrm{E}_{t-1}[RIV_{i,t}]=({}^{\nu-1}/\nu)\,\mu_{i,t}.$$

**Proof of Proposition 2** The forecasts of $\sigma^2_{f,t+s}$, $\beta_{i,t+s}$ and $\sigma^2_{i,t+s}$ are derived by the same procedure, so we would just present for $\sigma^2_{f,t+s}$ for simplicity. By (5), we have

$$\mathrm{E}_t\left[\sigma^2_{f,t+s}\right] = \theta_0 + \theta_1 \mathrm{E}_t\left[RM_{f,t+s-1}\right] + \theta_2 \mathrm{E}_t\left[\sigma^2_{f,t+s-1}\right]. \tag{28}$$

The recursive equation (6) of the conditional mean of realized factor volatility gives rise to

$$\mathrm{E}_t\left[\mu_{f,t+s}\right] = \theta_0^M + \theta_1^M \mathrm{E}_t\left[RM_{f,t+s-1}\right] + \theta_2^M \mathrm{E}_t\left[\mu_{f,t+s-1}\right]. \tag{29}$$

Since $\mathrm{E}_t(\mu_{f,t+s}) = \mathrm{E}_t(\mathrm{E}_{t+s-1}(RM_{f,t+s})) = \mathrm{E}_t(RM_{f,t+s})$ by the law of iterated expectations, (29) can be rewritten as

$$\mathrm{E}_t\left[RM_{f,t+s}\right] = \theta_0^M + (\theta_1^M + \theta_2^M)\mathrm{E}_t\left[RM_{f,t+s-1}\right]$$

when $s \geq 2$. Thus,

$$\mathrm{E}_t\left[RM_{f,t+s}\right] = \theta_0^M \frac{1-(\theta_1^M+\theta_2^M)^{s-1}}{1-(\theta_1^M+\theta_2^M)} + (\theta_1^M+\theta_2^M)^{s-1}\mu_{f,t+1}. \tag{30}$$

By Substituting (30) for $\mathrm{E}_t\left[RM_{f,t+s-1}\right]$ into (28) and forward iterating, we finally have the close-form of $s$-step forecast of (13).

**Proof of Proposition 3** Here we only present the proof of (18), since (20) can be similarly proved. We partition $\Sigma_t$ as follows

$$
\begin{aligned}
\Sigma_t &= \begin{bmatrix}
\sigma_{f,t}^2 & \beta_{1,t}\sigma_{f,t}^2 & \beta_{2,t}\sigma_{f,t}^2 & \cdots & \beta_{N,t}\sigma_{f,t}^2 \\
\beta_{1,t}\sigma_{f,t}^2 & \beta_{1,t}^2\sigma_{f,t}^2+\sigma_{1,t}^2 & \beta_{1,t}\beta_{2,t}\sigma_{f,t}^2 & \cdots & \beta_{1,t}\beta_{N,t}\sigma_{f,t}^2 \\
\beta_{2,t}\sigma_{f,t}^2 & \beta_{2,t}\beta_{1,t}\sigma_{f,t}^2 & \beta_{2,t}^2\sigma_{f,t}^2+\sigma_{2,t}^2 & \cdots & \beta_{2,t}\beta_{N,t}\sigma_{f,t}^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta_{N,t}\sigma_{f,t}^2 & \beta_{N,t}\beta_{1,t}\sigma_{f,t}^2 & \beta_{N,t}\beta_{2,t}\sigma_{f,t}^2 & \cdots & \beta_{N,t}^2\sigma_{f,t}^2+\sigma_{N,t}^2
\end{bmatrix} \\
&= \begin{bmatrix}
\sigma_{f,t}^2 & \sigma_{f,t}^2\beta_t' \\
\sigma_{f,t}^2\beta_t & \sigma_{f,t}^2\beta_t\beta_t'+\sigma_t^2
\end{bmatrix},
\end{aligned}
\tag{31}
$$

where $\beta_t$=$(\beta_{1,t},...,\beta_{N,t})'$ and

$$
\sigma_t^2 = \begin{bmatrix}
\sigma_{1,t}^2 & 0 & \cdots & 0 \\
0 & \sigma_{2,t}^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_{N,t}^2
\end{bmatrix}.
$$

Using the partitioned inverse theorem,

$$
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix}
A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\
-(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1}
\end{bmatrix},
$$

the inverse of (31) can be written as

$$\Sigma_t^{-1} = \begin{bmatrix} \sigma_{f,t}^2 & \sigma_{f,t}^2 \beta_t \\ \sigma_{f,t}^2 \beta_t' & \sigma_{f,t}^2 \beta_t' \beta_t + \sigma_t^2 \end{bmatrix}^{-1} = \begin{bmatrix} \left(\sigma_{f,t}^2\right)^{-1} + \beta_t' \left(\sigma_t^2\right)^{-1} \beta_t' & -\beta_t' \left(\sigma_t^2\right)^{-1} \\ -\left(\sigma_t^2\right)^{-1} \beta_t & \left(\sigma_t^2\right)^{-1} \end{bmatrix}. \tag{32}$$

Define $\tilde{r}_t = (r_{1,t}, ..., r_{N,t})$ as the returns on the $N$ assets, then

$$\begin{aligned} & \begin{bmatrix} r_{f,t} & r_{1,t} & \dots & r_{N,,t} \end{bmatrix} \Sigma_t^{-1} \begin{bmatrix} r_{f,t} & r_{1,t} & \dots & r_{N,,t} \end{bmatrix}' \\ = \quad & \begin{bmatrix} r_{f,t} & \tilde{r}_t \end{bmatrix} \begin{bmatrix} \left(\sigma_{f,t}^2\right)^{-1} + \beta_t' \left(\sigma_t^2\right)^{-1} \beta_t' & -\beta_t' \left(\sigma_t^2\right)^{-1} \\ -\left(\sigma_t^2\right)^{-1} \beta_t & \left(\sigma_t^2\right)^{-1} \end{bmatrix} \begin{bmatrix} r_{f,t} \\ \tilde{r}_t' \end{bmatrix} \\ = \quad & \left( \left(\sigma_{f,t}^2\right)^{-1} + \beta_t \left(\sigma_t^2\right)^{-1} \beta_t' \right) r_{f,t}^2 - 2 r_{f,t} \beta_t \left(\sigma_t^2\right)^{-1} \tilde{r}_t' + \tilde{r}_t (\sigma_t^2)^{-1} \tilde{r}_t' \\ = \quad & \frac{r_{f,t}^2}{\sigma_{f,t}^2} + \sum_{i=1}^N \frac{\left(r_{i,t} - \beta_{i,t} r_{f,t}\right)^2}{\sigma_{i,t}^2}. \end{aligned} \tag{33}$$

The determinant of the invertible partitioned matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is $|A||D - CA^{-1}B|$, and so

$$|\Sigma_t| = \sigma_{f,t}^2 \prod_{i=1}^N \sigma_{i,t}^2. \tag{34}$$

Substituting (33) and (34) into (17), one can finally get (18).

# 3 Estimation of Dynamic Panel Data Models with Stochastic Volatility using Particle Filters

## 3.1 Introduction

Dynamic panel data models characterize the dynamic adjustment processes which are common in economic relationships. For estimation and inference, the literature has focused on the generalized method of moments (GMM) approach (e.g. Arellano & Bond 1991; Blundell & Bond 1998) and the likelihood approach (e.g. Hsiao *et al.* 2002; Hayakawa & Pesaran 2015). In this chapter, we estimate dynamic panel data models with stochastic volatility by particle filters in a likelihood approach.

It has been well documented that there is time-varying volatility in macroeconomic data. Kim & Nelson (1999), McConnell & Perez-Quiros (2000) and Blanchard & Simon (2001), for instance, studied the moderated volatility in the U.S. real GDP growth. Stock & Watson (2003) found the decline in volatility was common among many U.S. macroeconomic time series. They argued that the moderation was associated more with a decrease in the magnitude of unforecastable disturbances than with the propagation mechanism of those disturbances. Fernández-Villaverde & Rubio-Ramírez (2013) provided an updated documentation of the great moderation in the U.S. economy. Besides, they showed the presence of time-varying volatility of the Emerging Markets Bond Index+ spread reported by J.P. Morgan. Blanchard & Simon (2001) and Stock & Watson (2003) revealed that the time variation in volatility also existed in other developed countries. Traditionally homoscedasticity is assumed for the innovations of macroeconomic time series. Recently, motivated by the studies mentioned above, some papers incorporated time-varying volatility in various models. Fernández-Villaverde & Rubio-Ramírez (2007) used the particle filter to estimate DSGE models with stochastic volatility on the structural shocks. Koop & Korobilis (2010) discussed the time-varying parameter vector autoregressive models with multivariate stochastic volatility. Hamilton (2010) argued that time-varying volatility should be considered even when the conditional mean is the direct object of interest in that statistical efficiency gains can be obtained by incorporating appropriate features of time-varying volatility into estimation of the conditional mean.

Dynamic panel data models have become increasingly popular in macroeconomics to study common relationships across countries or regions, such as growth convergence (e.g. Islam, 1995), purchasing power parity (e.g. Frankel & Rose, 1996) and mean reversion of interest rates (Wu & Chen, 2001). One example of incorporating time-varying volatility into dynamic panel models is the study of uncertainty shocks, which are often modeled as time variations in the volatility of business conditions or policy rules and can influence individual firm's decisions and macroeconomic aggregates (Bloom, 2009; Basu & Bundick, 2012; Fernández-Villaverde *et al.*, 2013). To study the impact of such shocks on employment, output and other economic outcomes, one might use firm-level panel data in a standard model of the firm extended by introducing time-varying volatility of demand and productivity (Bloom, 2009). The model's solution involves estimating a dynamic panel data model with time-varying volatility in the innovations, which might be parameterized through stochastic volatility (Kim *et al.*, 1998; Fernández-Villaverde & Rubio-Ramírez, 2007). In the panel models with stochastic volatility, there is no closed-form expression for the likelihood. Consequently, some simulation-based method is required for efficient estimation.

Particle filters, also known as sequential Monte Carlo methods, are simulation-based techniques to estimate the posterior density of state variables for nonlinear and non-Gaussian state space models, see Creal (2012). The methods approximate the continuous distribution by a discrete distribution made of weighted draws called particles. There are various algorithms which differ mainly in the choices of the incremental importance distribution and the resampling algorithm aimed to improve the level of statistical efficiency in terms of Monte Carlo variation. We apply the Rao–Blackwellized particle filter (e.g. Chen & Liu 2000; Doucet *et al.* 2000; Andrieu & Doucet 2002; Karlsson *et al.* 2005) in particular to the state-space representation of dynamic panel data models in first differences and maximize the simulated likelihood using the approximation method in Olsson & Rydén (2008). When estimating the coefficients in the panel equation, Monte Carlo studies show that our estimator can be more precise than the maximum likelihood estimator which neglects stochastic volatility (Hsiao *et al.*, 2002) and GMM estimators (e.g. Arellano & Bond 1991; Blundell & Bond 1998) in the presence of stochastic volatility especially in the case of high volatility of volatility. The advantage of the particle filter-based estimator is larger in dynamic panel data models with no regressors.

The chapter is organized as follows. Section 3.2 presents the dynamic panel data models with stochastic volatility and the state-space representation. Section 3.3 introduces the particle filter-based estimator. Section 3.4 reports some Monte Carlo simulations to study finite sample properties. Section 3.5 concludes the chapter. Section 3.6 gives the appendix.

## 3.2 The Models and State Space Representation

In general, we study the dynamic panel data models with time-varying volatility given by

$$y_{it} = B(L)y_{i,t-1} + \gamma \mathbf{x}_{it} + \mu_i + \sigma_t \upsilon_{it}, \tag{35}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$; $B(L)$ denotes a polynomial of lag operators; $\mathbf{x}_{it}$ denotes a vector of strictly exogenous regressors; $\mu_i$ is the individual-specific effect on which we impose no restriction under the fixed effects specification; $\nu_{it}$ is independently, identically distributed (i.i.d.) across both $i$ and $t$ and is independent of the volatility $\sigma_t$. We parameterize $\sigma_t^2$ through the stochastic volatility process to model conditional heteroskedasticity. The specification is

$$\log(\sigma_t^2) = (1-\phi)\kappa + \phi \log(\sigma_{t-1}^2) + \eta_t. \tag{36}$$

Furthermore, $\nu_{it}$ and $\eta_t$ follow a joint normal distribution

$$\begin{pmatrix} \upsilon_{it} \\ \eta_t \end{pmatrix} \sim N(0, \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & \theta^2 \end{bmatrix}), \tag{37}$$

in which $\sigma_\nu^2$ is set to 1 for identifiability reasons. Various extensions of the basic specification can be made. For example, it is possible to incorporate dependence between $\nu_{it}$ and $\eta_t$, which is called the leverage effect suggested by the evidence of stock returns; it is also possible to model error terms using a fat-tailed distribution.

The focus of the chapter is on estimation of the coefficients in (35)-(37) in a likelihood approach. Because of stochastic volatility, the tractable expression for the exact likelihood function is unknown. Consequently, we will compute the likelihood by simulation in the state space form of the transformed model. For ease of exposition, we consider as the benchmark model the AR(1) dynamic panel model with an exogenous regressor $x_{it}$:

$$y_{it} = \beta y_{i,t-1} + \gamma x_{it} + \mu_i + \sigma_t \upsilon_{it}. \tag{38}$$

The state space form of the first difference of the model can be written as

$$\begin{aligned}
\Delta y_{it} &= \begin{bmatrix} 1 & 1 & 0 & \Delta x_{it} \end{bmatrix} \alpha_{it} &\quad (39)\\
\alpha_{it} &= \begin{bmatrix} \beta & \beta & 0 & \beta \Delta x_{i,t-1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \alpha_{i,t-1} + \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \sigma_t \nu_{it}, &\quad (40)
\end{aligned}$$

where the state vector is $\alpha_{it} = \begin{bmatrix} \beta \Delta y_{i,t-1} & \sigma_t \nu_{it} - \sigma_{t-1} \nu_{i,t-1} & -\sigma_t \nu_{it} & \gamma \end{bmatrix}'$. (39) is the observation equation, while (36) and (40) are the state equations. This form is not the only state space version of the model, but all versions have the same joint likelihood of $\Delta y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})$ given $\Delta y_{i1}$ and $\Delta x_i = (\Delta x_{i2}, \ldots, \Delta x_{iT})$, which we will use for parameter estimation. In the special case where there is no regressor, the state space form is simplified to

$$
\begin{aligned}
\Delta y_{it} &= \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \tilde{\alpha}_{it} \\
\tilde{\alpha}_{it} &= \begin{bmatrix} \beta & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tilde{\alpha}_{i,t-1} + \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \sigma_t \nu_{it},
\end{aligned}
$$

where the state vector is $\tilde{\alpha}_{it} = \begin{bmatrix} \beta \Delta y_{i,t-1} & \sigma_t \nu_{it} - \sigma_{t-1} \nu_{i,t-1} & -\sigma_t \nu_{it} \end{bmatrix}'$.

## 3.3 Estimation by Particle Filters

We aim to estimate the model in a likelihood approach. Since both $\sigma_t$ and $\upsilon_{it}$ are stochastic, the model in first difference is non-Gaussian and a closed-form expression for the exact likelihood of $(\Delta y_1, \ldots, \Delta y_N)$ given $(\Delta y_{11}, \ldots, \Delta y_{N1})$ and $(\Delta x_1, \ldots, \Delta x_N)$ does not exist. Conditional on the volatility process $\{\sigma_t\}$, however, the resulting system including (37), (39) and (40) is a linear Gaussian state space model. It follows that the likelihood of $(\Delta y_1, \ldots, \Delta y_N)$ given $(\Delta y_{11}, \ldots, \Delta y_{N1})$, $(\Delta x_1, \ldots, \Delta x_N)$ and $(\sigma_2, \ldots, \sigma_T)$ can be evaluated analytically by the Kalman filter. This constitutes the basic idea of Rao–Blackwellized particle filters (e.g. Chen & Liu 2000; Doucet *et al.* 2000; Andrieu & Doucet 2002; Karlsson *et al.* 2005).

Particle filters can be regarded as the extension of the Kalman filter to address nonlinear and non-Gaussian state space models in which the posterior density of state variables seldom has the closed-form expression. They are simulation-based techniques to obtain filtered estimates of the states as well as an unbiased estimate of the likelihood. Particle filters can be implemented in several ways, which vary with the choices of the incremental importance distribution and the resampling algorithm. The model containing (36), (37), (39) and (40) belongs to a class of models suitable for Rao–Blackwellized particle filters, which integrate out a subset of state variables ($\alpha_{it}$ in our case) in order to reduce the Monte Carlo variation of the simulation-based estimators.

We apply the algorithm of Rao–Blackwellized particle filters to the panel data models. Although we focus on the AR(1) dynamic panel model, the algorhtim can be adapted for a general model straightforwardly. Let $\mathscr{I}_{it}$ denote the information set containing all observations of $y_{it}$ and $x_{it}$ up to $t$. $m_{it|t-1} = \mathrm{E}(\alpha_{it}|\mathscr{I}_{i,t-1}, \sigma_t^2)$ and $\Sigma_{it|t-1} = \mathrm{Var}(\alpha_{it}|\mathscr{I}_{i,t-1}, \sigma_t^2)$ represent prediction mean and variance of $\alpha_{it}$, respectively. The algorithm produces $M$ simulated state variables and corresponding weights $\{\sigma_t^{2(j)}, m_{it|t-1}^{(j)}, \Sigma_{it|t-1}^{(j)}, \hat{w}_{t|t-1}^{(j)}\}_{j=1}^{M}$ for $t = 2, \ldots, T$ and $i = 1, \ldots, N$. Given the values of parameters $\omega = (\beta, \gamma, \kappa, \phi, \theta)$, the algorithm is:

1) For $j = 1, \ldots, M$, draw $\log(\sigma_2^{2(j)})$ from the stationary distribution

$$\log(\sigma_2^{2(j)}) \sim N(\kappa, \frac{\theta^2}{1-\phi^2}),$$

and for $i = 1, ..., N$, set the initial values of prediction means and variances equal to expectations and variances respectively conditional on $\Delta y_{i1}$, $\sigma_1^{2(j)}$ and $\sigma_2^{2(j)}$:

$$m_{i2|1}^{(j)} = \begin{bmatrix} \beta \Delta y_{i1} \\ 0 \\ 0 \\ \gamma \end{bmatrix}$$

$$\Sigma_{i2|1}^{(j)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2\sigma_2^{2(j)} & -\sigma_2^{2(j)} & 0 \\ 0 & -\sigma_2^{2(j)} & \sigma_2^{2(j)} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Compute the conditional log-likelihood of $(\Delta y_{12}, \ldots, \Delta y_{N2})$ as $l_2^{(j)} = \sum_{i=1}^{N} l_{i2}^{(j)}$ where

$$l_{i2}^{(j)} = -0.5 \log |V_{i2}^{(j)}| - 0.5 v_{i2}^{(j)} (V_{i2}^{(j)})^{-1} v_{i2}^{(j)},$$

with the forecast error $v_{i2}^{(j)} = \Delta y_{i2} - \begin{bmatrix} 1 & 1 & 0 & \Delta x_{i2} \end{bmatrix} m_{i2|1}^{(j)}$ and forecast variance

$$V_{i2}^{(j)} = \begin{bmatrix} 1 & 1 & 0 & \Delta x_{i2} \end{bmatrix} \Sigma_{i2|1}^{(j)} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{i2} \end{bmatrix}.$$

The log importance weight is $w_{2|1}^{(j)} = 0$ and the normalized importance weight is $\hat{w}_{2|1}^{(j)} = \frac{1}{M}$.

For $t = 3, ..., T$, $i = 1, ..., N$ and $j = 1, ..., M$,

2) Update the weight by $w_{t|t-1}^{(j)} = w_{t-1|t-2}^{(j)} + l_{t-1}^{(j)}$ to incorporate the new likelihood (Shephard, 2013) and $\hat{w}_{t|t-1}^{(j)} = \frac{\exp(w_{t|t-1}^{(j)})}{\sum_{j=1}^{M} \exp(w_{t|t-1}^{(j)})}$.

3) Draw

$$\log(\sigma_t^{2(j)}) = (1-\phi)\kappa + \phi \log(\sigma_{t-1}^{2(j)}) + \theta \eta_t^{(j)},$$

where $\eta_t^{(j)} \sim N(0,1)$ is serially independent.

4) Run the Kalman filter as follows to obtain $m_{it|t-1}^{(j)}$ and $\Sigma_{it|t-1}^{(j)}$. Let $m_{it|t-1,2}^{(j)}$ denote the entry in the 2nd row of $m_{it|t-1}^{(j)}$ and $\Sigma_{it|t-1,22}^{(j)}$ denote the entry in the 2nd row and 2nd column of $\Sigma_{it|t-1}^{(j)}$. After some algebra (see the appendix), we have

$$m_{it|t-1}^{(j)} = \begin{bmatrix} \beta \Delta y_{i,t-1} \\ m_{it|t-1,2}^{(j)} \\ 0 \\ \gamma \end{bmatrix}, \tag{41}$$

and

$$\Sigma_{it|t-1}^{(j)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_{it|t-1,22}^{(j)} & -\sigma_t^{2(j)} & 0 \\ 0 & -\sigma_t^{2(j)} & \sigma_t^{2(j)} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{42}$$

such that

$$m^{(j)}_{it|t-1,2} = -\frac{\sigma^{2(j)}_{t-1}}{\Sigma^{(j)}_{i,t-1|t-2,22}}(\Delta y_{i,t-1} - \beta \Delta y_{i,t-2} - \gamma \Delta x_{i,t-1} - m^{(j)}_{i,t-1|t-2,2}), \tag{43}$$

and

$$\Sigma^{(j)}_{it|t-1,22} = \sigma^{2(j)}_{t} + \sigma^{2(j)}_{t-1} - \frac{\sigma^{4(j)}_{t-1}}{\Sigma^{(j)}_{i,t-1|t-2,22}}. \tag{44}$$

5) Compute the conditional log-likelihood of $(\Delta y_{1t}, \ldots, \Delta y_{Nt})$ by $l^{(j)}_t = \sum_{i=1}^{N} l^{(j)}_{it}$ such that

$$l^{(j)}_{it} = -0.5\log|V^{(j)}_{it}| - 0.5 v^{(j)}_{it} (V^{(j)}_{it})^{-1} v^{(j)}_{it},$$

where $v^{(j)}_{it} = \Delta y_{it} - \begin{bmatrix} 1 & 1 & 0 & \Delta x_{it} \end{bmatrix} m^{(j)}_{it|t-1}$ and

$$V^{(j)}_{it} = \begin{bmatrix} 1 & 1 & 0 & \Delta x_{it} \end{bmatrix} \Sigma^{(j)}_{it|t-1} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}.$$

6) Resample with replacement (discussed later) $M$ particles $\sigma^{2(j)}_{t-1}$, $\sigma^{2(j)}_{t}$, $m^{(j)}_{it|t-1}$ and $\Sigma^{(j)}_{it|t-1}$ with the weight $\hat{w}^{(j)}_{t|t-1}$ every three increments.[14] Then reset $w^{(j)}_{t|t-1} = 0$ and $\hat{w}^{(j)}_{t|t-1} = \frac{1}{M}$.

[14] It is an ad-hoc choice for stability of the algorithm, see Shephard (2013).

The simulation-based estimate of the joint log-likelihood of $(\Delta y_1, \ldots, \Delta y_N)$ given $(\Delta y_{11}, \ldots, \Delta y_{N1})$ and $(\Delta x_1, \ldots, \Delta x_N)$ is

$$\hat{l}(\omega) = \sum_{i=1}^{N} \sum_{t=2}^{T} \log[\sum_{j=1}^{M} \hat{w}_{t|t-1}^{(j)} \exp(l_{it}^{(j)})]. \tag{45}$$

Note that this algorithm is appropriate when $x_{it}$ is strictly exogenous. If it is predetermined or endogenous, the likelihood function would have a different form rather than the one given above, because the conditioning set assuming strict exogeneity includes all values of $x_{it}$ from $t = 1$ to $T$.

In practice, one particle's normalized importance weight converges to one while the others converge to zero over time. In other words, the discrete distribution made of weighted draws would become degenerate. Resampling is crucial to stabilize the algorithm by eliminating the particles which have low importance weights and multiplying the heavily weighted particles. The simplest unbiased algorithm is multinomial resampling introduced in Gordon *et al.* (1993). It draws new particles $\{\tilde{\sigma}_{t-1}^{2(j)}, \tilde{\sigma}_t^{2(j)}, \tilde{m}_{it|t-1}^{(j)}, \tilde{\Sigma}_{it|t-1}^{(j)}\}_{j=1}^{M}$ from the point mass distribution $\{\sigma_{t-1}^{2(j)}, \sigma_t^{2(j)}, m_{it|t-1}^{(j)}, \Sigma_{it|t-1}^{(j)}, \hat{w}_{t|t-1}^{(j)}\}_{j=1}^{M}$. We employ multinomial resampling as it is a requirement for good asymptotic performance of the approximation method in Olsson & Rydén (2008) used later.

Although (45) is an unbiased estimator of the exact log-likelihood under some regularity conditions, direct maximization for parameter estimation suffers from discontinuity induced from the generalized inverse operation at the resampling stage, which makes invalid the common gradient-based optimization methods. Olsson & Rydén (2008) approximated the likelihood by means of step functions or B-spline interpolation and maximized the approximate likelihood. They showed consistency and asymptotic normality of the estimators under some assumptions. This seems the only work that studies asymptotic properties of parameter estimators in the particle filter literature. One of the assumptions requires a compact state space which is obviously not true in our case. We however still use this method for two reasons. One is that the compactness assumption can be potentially released given new results of uniform convergence properties in time dimension, although the full proof of the extension is beyond the scope of this chapter; the other reason is the good finite sample performance shown below.

Specifically, Olsson & Rydén (2008) discretized the parameter space $\Omega$ by a grid $\bar{\Omega} = \{\omega_g\}_{g=1}^{G} \subseteq \Omega$. Let $[\omega]$ denote the closest point in the grid to $\omega \in \Omega$.[15] The grid-based approximation of the likelihood using piecewise constant functions is given by

$$\hat{l}(\omega) \simeq \hat{l}([\omega]). \tag{46}$$

The approximation can also be made via spline interpolation, which is more efficient than piecewise constant functions but suffers from higher computational costs as the dimension of parameter space grows. Consequently, we maximize (46) to obtain the parameter estimates.

## 3.4 Monte Carlo Studies

In this section, we investigate the finite sample performance of particle filter-based estimators. The baseline data generating process in our Monte Carlo studies is an AR(1) dynamic

---

15

If there is more than one point having the smallest distance from $\omega$, the point with lowest index $g$ will be chosen.

panel model with an exogenous regressor

$$y_{it} = \beta y_{i,t-1} + \gamma x_{it} + (1-\beta)\mu_i + \sigma_t \upsilon_{it}$$

$$x_{it} = b x_{i,t-1} + \lambda \varepsilon_{it}$$

$$\mu_i = \sqrt{\tau}(\frac{q_i - 1}{\sqrt{2}})\varsigma_i$$

$$\log(\sigma_t^2) = (1-\phi)\log(\mu) + \phi\log(\sigma_{t-1}^2) + \eta_t,$$

where $q_i \sim \chi_1^2$; $\upsilon_{it}, \varepsilon_{it}, \varsigma_i \sim N(0,1)$ and $\eta_t \sim N(0, \theta^2)$; $q_i$, $\upsilon_{it}$, $\varepsilon_{it}$, $\varsigma_i$ and $\eta_t$ are all i.i.d. within series and independent of each other.[16]

We experiment with $\beta = 0.8$, $\gamma = 0.7$, $b = 0.5$, $\lambda = 2$ and $\tau = 1$. $N = 50, 100$ and $T = 50, 100$ are typical in macroeconomic applications. According to the empirical results of Fernández-Villaverde & Rubio-Ramírez (2007), we consider $\mu = 0.002$, $\phi = 0.99$ and $\theta = 0.5$. We also check the performance of the estimators in the absence of stochastic volatility ($\phi = \theta = 0$) as well as in the case of higher volatility of volatility ($\phi = 0.99$, $\theta = 1$). We set $y_{i0} = 0$ and discard the first 100 observations of simulated data in order that the series is long enough to eliminate the initial effect. The number of the particles is set to 400, as we experimented with more particles up to 1000 but found few extra benefits. We carry out 1000 replications for each experiment. The quality of the estimators are evaluated

16

$\tau$, which measures the degree of cross-section to time-series variation, can influence the finite sample properties of GMM-type estimators.

by biases and root mean square errors (RMSE).

When estimating $\beta$ and $\gamma$, we compare our estimator (PF) with some popular estimators in the literature, including the maximum likelihood estimator based on the model in first differences (FDML) (Hsiao *et al.*, 2002), GMM (Arellano & Bond, 1991) and system GMM (SGMM) (Blundell & Bond, 1998). Those estimators are still consistent under stochastic volatility, as consistency does not require the correct specification of the conditional second moments of the error term. However, they might be less efficient than the one which takes account of time-varying volatility. Let $\varepsilon_{it}$ denote the disturbance in the level equation. GMM exploits a set of linear orthogonality conditions $\mathrm{E}(y_{i,t-s}\Delta\varepsilon_{it}) = 0$ $(t \geq 2;\, s \geq 2)$ and $\mathrm{E}(x_{is}\Delta\varepsilon_{it}) = 0(t \geq 2;\, s \geq 1)$ (strict exogeneity) for the equation in first differences. We select a subset of these conditions often used in practice to improve the finite sample performance (Hayakawa & Pesaran, 2015). That is, $\mathrm{E}(y_{i,t-s-2}\Delta\varepsilon_{it}) = 0$ for $t = 2$, $s = 0$ and $t \geq 3$, $s = 0, 1$. SGMM uses extra moment conditions $\mathrm{E}(\Delta y_{i,t-1}((1-\beta)\mu_i + \varepsilon_{it})) = 0$ $(t \geq 2)$ and $\mathrm{E}(\Delta x_{it}((1-\beta)\mu_i + \varepsilon_{it}) = 0(t \geq 2)$ for the level equation. We use one-step GMM instead of two-step GMM for its better finite sample performance. We use the pattern search method in Matlab to find the maximum of the approximate likelihood function. PF is implemented at a quite low computational cost, although still slower than the other methods considered here.

Table 8 lists the estimation results for $\beta$ and $\gamma$. The results seem quite mixed and no single method is dominant in all designs. While FDML or GMM is most precise when stochastic volatility is absent, PF can be preferred in the presence of stochastic volatility especially when the volatility of volatility is high. As the volatility of volatility grows, FDML, GMM and SGMM become much worse at estimating $\gamma$. This is not surprising as they neglect the information contained in stochastic volatility. Table 9 reports the estimates of $\phi$, $\theta$ and $\mu$ in the stochastic volatility equation. The estimator $\hat{\phi}$ in the models without stochastic volatility has large bias. The results also show that the bias of $\hat{\mu}$ increases in the volatility of volatility. We also estimate an AR(1) panel model with no exogenous regressor $x_{it}$ (the data generating processes for the other variables are same as above). The estimates of $\beta$ are listed in Table 10. The absence of the regressor makes PF favoured when stochastic volatility exists. The estimates of $\phi$, $\theta$ and $\mu$ in the model with no regressors reported in Table 11 are similar to those in Table 9. In sum, as PF takes correct specification into account, its performance is generally good in finite samples and relatively stable across designs.

| $T=50$ | | | | | | | $T=100$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\phi$ | $\theta$ | FDML | GMM | SGMM | PF | $N$ | $\phi$ | $\theta$ | FDML | GMM | SGMM | PF |
| 50 | 0 | 0 | **0.000** | **0.000** | 0.001 | 0.003 | 50 | 0 | 0 | **-0.000** | **0.000** | 0.002 | 0.003 |
| | | | **(0.000)** | (0.001) | (0.002) | (0.003) | | | | **(0.000)** | **(0.000)** | (0.002) | (0.003) |
| | | | **-0.000** | **0.000** | **0.000** | 0.003 | | | | **0.000** | **0.000** | **0.000** | 0.003 |
| | | | **(0.000)** | **(0.000)** | (0.001) | (0.003) | | | | **(0.000)** | **(0.000)** | (0.001) | (0.003) |
| | 0.99 | 0.5 | -0.003 | -0.002 | **0.000** | 0.002 | | 0.99 | 0.5 | -0.001 | -0.002 | **-0.000** | 0.003 |
| | | | (0.009) | (0.012) | (0.012) | **(0.007)** | | | | **(0.005)** | (0.009) | (0.010) | **(0.005)** |
| | | | **0.001** | **0.001** | **0.001** | 0.004 | | | | **0.001** | **0.001** | **0.001** | 0.004 |
| | | | (0.017) | (0.017) | (0.016) | **(0.013)** | | | | **(0.009)** | (0.013) | (0.013) | (0.011) |
| | 0.99 | 1 | -0.011 | -0.007 | -0.002 | **0.000** | | 0.99 | 1 | -0.006 | -0.006 | **-0.002** | **0.002** |
| | | | **(0.022)** | (0.035) | (0.023) | (0.037) | | | | **(0.015)** | (0.023) | (0.023) | (0.029) |
| | | | 0.078 | **0.002** | -0.037 | -0.008 | | | | -0.010 | -0.042 | -0.062 | **-0.001** |
| | | | (1.444) | (0.983) | (1.909) | **(0.106)** | | | | (1.664) | (1.800) | (2.418) | **(0.082)** |
| 100 | 0 | 0 | 0.016 | **0.000** | 0.001 | 0.003 | 100 | 0 | 0 | **-0.000** | **0.000** | 0.001 | 0.003 |
| | | | (0.028) | **(0.000)** | (0.001) | (0.003) | | | | **(0.000)** | **(0.000)** | (0.001) | (0.003) |
| | | | -0.007 | **0.000** | **-0.000** | 0.003 | | | | **0.000** | **-0.000** | **-0.000** | 0.003 |
| | | | (0.012) | **(0.000)** | (0.001) | (0.003) | | | | **(0.000)** | **(0.000)** | **(0.000)** | (0.003) |
| | 0.99 | 0.5 | 0.016 | -0.001 | **-0.000** | 0.002 | | 0.99 | 0.5 | -0.001 | -0.001 | **0.000** | 0.003 |
| | | | (0.033) | **(0.008)** | **(0.008)** | (0.016) | | | | **(0.004)** | (0.006) | (0.006) | **(0.004)** |
| | | | -0.007 | **0.001** | **0.001** | 0.003 | | | | 0.001 | **0.000** | **0.000** | 0.003 |
| | | | (0.018) | **(0.017)** | **(0.017)** | (0.027) | | | | (0.017) | (0.010) | **(0.008)** | (0.018) |
| | 0.99 | 1 | 0.009 | -0.004 | **-0.002** | **-0.002** | | 0.99 | 1 | -0.007 | -0.003 | **-0.001** | 0.003 |
| | | | (0.039) | (0.024) | **(0.019)** | (0.048) | | | | **(0.013)** | (0.019) | (0.017) | (0.037) |
| | | | 0.050 | -0.054 | -0.057 | **-0.007** | | | | **0.004** | 0.007 | -0.005 | -0.008 |
| | | | (1.134) | (2.978) | (3.046) | **0.103** | | | | (0.942) | (1.510) | (1.591) | **(0.103)** |

Table 8: Estimates of $\beta$ and $\gamma$. In each design, the upper two rows give the biases and RMSE (in brackets) of $\hat{\beta}$, while the lower two rows give those of $\hat{\gamma}$. Numbers in bold font indicate the smallest biases and RMSE.

| $T = 50$ | | | | | | $T = 100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\phi$ | $\theta$ | $\hat{\phi}$ | $\hat{\theta}$ | $\hat{\mu}$ | $N$ | $\phi$ | $\theta$ | $\hat{\phi}$ | $\hat{\theta}$ | $\hat{\mu}$ |
| 50 | 0 | 0 | 0.968 | 0.242 | 0.077 | 50 | 0 | 0 | 0.969 | 0.160 | 0.019 |
| | | | (0.968) | (0.275) | (0.172) | | | | (0.969) | (0.167) | (0.055) |
| | 0.99 | 0.5 | -0.052 | 0.082 | 0.254 | | 0.99 | 0.5 | -0.029 | 0.063 | 0.233 |
| | | | (0.080) | (0.184) | (0.383) | | | | (0.038) | (0.149) | (0.366) |
| | 0.99 | 1 | -0.058 | -0.091 | 0.447 | | 0.99 | 1 | -0.040 | -0.086 | 0.454 |
| | | | (0.091) | (0.186) | (0.571) | | | | (0.062) | (0.185) | (0.576) |
| 100 | 0 | 0 | 0.968 | 0.317 | 0.155 | 100 | 0 | 0 | 0.968 | 0.182 | 0.054 |
| | | | (0.969) | (0.357) | (0.266) | | | | (0.968) | (0.204) | (0.149) |
| | 0.99 | 0.5 | -0.051 | 0.112 | 0.324 | | 0.99 | 0.5 | -0.031 | 0.076 | 0.261 |
| | | | (0.081) | (0.208) | (0.441) | | | | (0.042) | (0.164) | (0.381) |
| | 0.99 | 1 | -0.060 | -0.096 | 0.476 | | 0.99 | 1 | -0.046 | -0.089 | 0.505 |
| | | | (0.091) | (0.191) | (0.588) | | | | (0.066) | (0.181) | (0.609) |

Table 9: Estimates of $\phi$, $\theta$ and $\mu$. In each circumstance, the first row gives the bias and the second row gives RMSE in brackets.

| $T=50$ | | | | | | | $T=100$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\phi$ | $\theta$ | FDML | GMM | SGMM | PF | $N$ | $\phi$ | $\theta$ | FDML | GMM | SGMM | PF |
| 50 | 0 | 0 | **-0.002** | -0.090 | 0.183 | -0.035 | 50 | 0 | 0 | **0.000** | -0.038 | 0.181 | -0.011 |
| | | | **(0.015)** | (0.123) | (0.184) | (0.042) | | | | **(0.009)** | (0.053) | (0.182) | (0.014) |
| | 0.99 | 0.5 | **0.020** | -0.057 | 0.085 | -0.028 | | 0.99 | 0.5 | 0.026 | -0.032 | 0.076 | **-0.008** |
| | | | (0.059) | (0.110) | (0.121) | **(0.042)** | | | | (0.066) | (0.066) | (0.111) | **(0.022)** |
| | 0.99 | 1 | 0.030 | -0.068 | 0.065 | **-0.009** | | 0.99 | 1 | 0.035 | -0.034 | 0.048 | **0.004** |
| | | | (0.090) | (0.280) | (0.115) | **(0.060)** | | | | (0.089) | (0.084) | (0.099) | **(0.054)** |
| 100 | 0 | 0 | **-0.001** | -0.058 | 0.178 | -0.033 | 100 | 0 | 0 | **0.000** | -0.021 | 0.173 | -0.009 |
| | | | **(0.010)** | (0.083) | (0.179) | (0.041) | | | | **(0.007)** | (0.032) | (0.174) | (0.016) |
| | 0.99 | 0.5 | 0.025 | -0.038 | 0.079 | **-0.021** | | 0.99 | 0.5 | 0.033 | -0.021 | 0.076 | **-0.003** |
| | | | (0.063) | (0.083) | (0.112) | **(0.040)** | | | | (0.074) | (0.053) | (0.109) | **(0.030)** |
| | 0.99 | 1 | 0.036 | -0.037 | 0.061 | **-0.001** | | 0.99 | 1 | 0.045 | -0.028 | 0.053 | **0.009** |
| | | | (0.093) | (0.096) | (0.107) | **(0.073)** | | | | (0.097) | (0.081) | (0.098) | **(0.059)** |

Table 10: Estimates of $\beta$ in an AR(1) panel with no regressors. In each design, the first row gives the biases and the second row gives the RMSE in brackets. Numbers in bold font indicate the smallest biases and RMSE.

| $T=50$ | | | | | | $T=100$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\phi$ | $\theta$ | $\hat{\phi}$ | $\hat{\theta}$ | $\hat{\mu}$ | $N$ | $\phi$ | $\theta$ | $\hat{\phi}$ | $\hat{\theta}$ | $\hat{\mu}$ |
| 50 | 0 | 0 | 0.955 | 0.292 | 0.022 | 50 | 0 | 0 | 0.968 | 0.176 | 0.014 |
| | | | (0.956) | (0.367) | (0.034) | | | | (0.968) | (0.197) | (0.022) |
| | 0.99 | 0.5 | -0.109 | 0.227 | 0.167 | | 0.99 | 0.5 | -0.073 | 0.211 | 0.151 |
| | | | (0.136) | (0.303) | (0.309) | | | | (0.089) | (0.280) | (0.278) |
| | 0.99 | 1 | -0.092 | -0.012 | 0.321 | | 0.99 | 1 | -0.067 | 0.007 | 0.353 |
| | | | (0.123) | (0.101) | (0.488) | | | | (0.086) | (0.082) | (0.512) |
| 100 | 0 | 0 | 0.935 | 0.431 | 0.033 | 100 | 0 | 0 | 0.960 | 0.231 | 0.021 |
| | | | (0.936) | (0.530) | (0.047) | | | | (0.960) | (0.303) | (0.032) |
| | 0.99 | 0.5 | -0.128 | 0.251 | 0.151 | | 0.99 | 0.5 | -0.084 | 0.273 | 0.149 |
| | | | (0.165) | (0.316) | (0.279) | | | | (0.104) | (0.328) | (0.272) |
| | 0.99 | 1 | -0.104 | -0.022 | 0.313 | | 0.99 | 1 | -0.075 | 0.001 | 0.330 |
| | | | (0.137) | (0.109) | (0.476) | | | | (0.098) | (0.090) | (0.497) |

Table 11: Estimates of $\phi$, $\theta$ and $\mu$ in an AR(1) panel with no regressors. In each circumstance, the first row gives the bias and the second row gives RMSE in brackets.

## 3.5 Conclusions

Motivated by time-varying volatility in macroeconomic data and the growing popularity of dynamic panel data models in macroeconomics, we propose a particle filter-based method to estimate the dynamic panel data models with stochastic volatility. Specifically, we represent the transformed model in the state space form and compute the simulated likelihood by Rao–Blackwellized particle filters. The parameter estimates are obtained by maximizing the likelihood approximated by piecewise constant functions. Monte Carlo results show our estimator is relatively stable across scenarios and has good performance in finite samples, especially when the volatility of volatility is high or when regressors are absent.

## 3.6 Appendix

Here we derive the equations (41), (42), (43) and (44). For ease of notations, we omit the superscript $(j)$. We apply the standard Kalman filter recursive equations (see e.g. Durbin & Koopman 2012) to our model:

$$
\begin{aligned}
\begin{bmatrix} m_{i,t+1|t,1} \\ m_{i,t+1|t,2} \\ m_{i,t+1|t,3} \\ m_{i,t+1|t,4} \end{bmatrix}
&= \begin{bmatrix} \beta & \beta & 0 & \beta\Delta x_{it} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} m_{it|t-1,1} \\ m_{it|t-1,2} \\ m_{it|t-1,3} \\ m_{it|t-1,4} \end{bmatrix} \\
&\quad + \begin{bmatrix} K_{t,1} \\ K_{t,2} \\ K_{t,3} \\ K_{t,4} \end{bmatrix} (\Delta y_{it} - m_{it|t-1,1} - m_{it|t-1,2} - \Delta x_{it} m_{it|t-1,4})
\end{aligned}
$$

$$
\begin{bmatrix}
\Sigma_{i,t+1|t,11} & \Sigma_{i,t+1|t,12} & \Sigma_{i,t+1|t,13} & \Sigma_{i,t+1|t,14} \\
\Sigma_{i,t+1|t,21} & \Sigma_{i,t+1|t,22} & \Sigma_{i,t+1|t,23} & \Sigma_{i,t+1|t,24} \\
\Sigma_{i,t+1|t,31} & \Sigma_{i,t+1|t,32} & \Sigma_{i,t+1|t,33} & \Sigma_{i,t+1|t,34} \\
\Sigma_{i,t+1|t,41} & \Sigma_{i,t+1|t,42} & \Sigma_{i,t+1|t,43} & \Sigma_{i,t+1|t,44}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\beta & \beta & 0 & \beta\Delta x_{it} \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\Sigma_{it|t-1,11} & \Sigma_{it|t-1,12} & \Sigma_{it|t-1,13} & \Sigma_{it|t-1,14} \\
\Sigma_{it|t-1,21} & \Sigma_{it|t-1,22} & \Sigma_{it|t-1,23} & \Sigma_{it|t-1,24} \\
\Sigma_{it|t-1,31} & \Sigma_{it|t-1,32} & \Sigma_{it|t-1,33} & \Sigma_{it|t-1,34} \\
\Sigma_{it|t-1,41} & \Sigma_{it|t-1,42} & \Sigma_{it|t-1,43} & \Sigma_{it|t-1,44}
\end{bmatrix}
$$

$$
(\begin{bmatrix}
\beta & 0 & 0 & 0 \\
\beta & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
\beta\Delta x_{it} & 0 & 0 & 1
\end{bmatrix}
- \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}
\begin{bmatrix} K_{t,1} & K_{t,2} & K_{t,3} & K_{t,4} \end{bmatrix})
$$

$$
+ \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \sigma^2_{t+1} & -\sigma^2_{t+1} & 0 \\
0 & -\sigma^2_{t+1} & \sigma^2_{t+1} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}, \tag{47}
$$

where

$$\begin{bmatrix} K_{t,1} \\ K_{t,2} \\ K_{t,3} \\ K_{t,4} \end{bmatrix} = \begin{bmatrix} \beta & \beta & 0 & \beta\Delta x_{it} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Sigma_{it|t-1,11} & \Sigma_{it|t-1,12} & \Sigma_{it|t-1,13} & \Sigma_{it|t-1,14} \\ \Sigma_{it|t-1,21} & \Sigma_{it|t-1,22} & \Sigma_{it|t-1,23} & \Sigma_{it|t-1,24} \\ \Sigma_{it|t-1,31} & \Sigma_{it|t-1,32} & \Sigma_{it|t-1,33} & \Sigma_{it|t-1,34} \\ \Sigma_{it|t-1,41} & \Sigma_{it|t-1,42} & \Sigma_{it|t-1,43} & \Sigma_{it|t-1,44} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}$$

$$[\begin{bmatrix} 1 & 1 & 0 & \Delta x_{it} \end{bmatrix} \begin{bmatrix} \Sigma_{it|t-1,11} & \Sigma_{it|t-1,12} & \Sigma_{it|t-1,13} & \Sigma_{it|t-1,14} \\ \Sigma_{it|t-1,21} & \Sigma_{it|t-1,22} & \Sigma_{it|t-1,23} & \Sigma_{it|t-1,24} \\ \Sigma_{it|t-1,31} & \Sigma_{it|t-1,32} & \Sigma_{it|t-1,33} & \Sigma_{it|t-1,34} \\ \Sigma_{it|t-1,41} & \Sigma_{it|t-1,42} & \Sigma_{it|t-1,43} & \Sigma_{it|t-1,44} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}]^{-1}.$$

If

$$\begin{bmatrix} \Sigma_{it|t-1,11} & \Sigma_{it|t-1,12} & \Sigma_{it|t-1,13} & \Sigma_{it|t-1,14} \\ \Sigma_{it|t-1,21} & \Sigma_{it|t-1,22} & \Sigma_{it|t-1,23} & \Sigma_{it|t-1,24} \\ \Sigma_{it|t-1,31} & \Sigma_{it|t-1,32} & \Sigma_{it|t-1,33} & \Sigma_{it|t-1,34} \\ \Sigma_{it|t-1,41} & \Sigma_{it|t-1,42} & \Sigma_{it|t-1,43} & \Sigma_{it|t-1,44} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_{it|t-1,22} & -\sigma_t^2 & 0 \\ 0 & -\sigma_t^2 & \sigma_t^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

such as when $t = 2$,

$$\Sigma_{i2|1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2\sigma_2^2 & -\sigma_2^2 & 0 \\ 0 & -\sigma_2^2 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

then

$$
\begin{bmatrix} K_{t,1} \\ K_{t,2} \\ K_{t,3} \\ K_{t,4} \end{bmatrix} = \begin{bmatrix} \beta & \beta & 0 & \beta\Delta x_{it} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_{it|t-1,22} & -\sigma_t^2 & 0 \\ 0 & -\sigma_t^2 & \sigma_t^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}
$$

$$
[\begin{bmatrix} 1 & 1 & 0 & \Delta x_{it} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_{it|t-1,22} & -\sigma_t^2 & 0 \\ 0 & -\sigma_t^2 & \sigma_t^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \Delta x_{it} \end{bmatrix}]^{-1}
$$

$$
= \begin{bmatrix} \beta \\ -\sigma_t^2/\Sigma_{it|t-1,22} \\ 0 \\ 0 \end{bmatrix}. \tag{48}
$$

Substituting (48) into (47) yields

$$
\begin{bmatrix} \Sigma_{i,t+1|t,11} & \Sigma_{i,t+1|t,12} & \Sigma_{i,t+1|t,13} & \Sigma_{i,t+1|t,14} \\ \Sigma_{i,t+1|t,21} & \Sigma_{i,t+1|t,22} & \Sigma_{i,t+1|t,23} & \Sigma_{i,t+1|t,24} \\ \Sigma_{i,t+1|t,31} & \Sigma_{i,t+1|t,32} & \Sigma_{i,t+1|t,33} & \Sigma_{i,t+1|t,34} \\ \Sigma_{i,t+1|t,41} & \Sigma_{i,t+1|t,42} & \Sigma_{i,t+1|t,43} & \Sigma_{i,t+1|t,44} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{t+1}^2 + \sigma_t^2 - \sigma_t^4/\Sigma_{it|t-1,22} & -\sigma_{t+1}^2 & 0 \\ 0 & -\sigma_{t+1}^2 & \sigma_{t+1}^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},
$$

which implies (42) and (44).

Further, if

$$\begin{bmatrix} m_{it|t-1,1} \\ m_{it|t-1,2} \\ m_{it|t-1,3} \\ m_{it|t-1,4} \end{bmatrix} = \begin{bmatrix} \beta\Delta y_{i,t-1} \\ m_{it|t-1,2} \\ 0 \\ \gamma \end{bmatrix},$$

such as when $t = 2$,

$$m_{i2|1} = \begin{bmatrix} \beta\Delta y_{i1} \\ 0 \\ 0 \\ \gamma \end{bmatrix},$$

then we will have

$$\begin{bmatrix} m_{i,t+1|t,1} \\ m_{i,t+1|t,2} \\ m_{i,t+1|t,3} \\ m_{i,t+1|t,4} \end{bmatrix} = \begin{bmatrix} \beta\Delta y_{it} \\ -\sigma_t^2(\Delta y_{it} - \beta\Delta y_{i,t-1} - m_{it|t-1,2} - \Delta x_{it}\gamma)/\Sigma_{it|t-1,22} \\ 0 \\ \gamma \end{bmatrix},$$

which implies (41) and (43).

# 4 Testing for the Stable Relationship in High-Dimension Factor Models

## 4.1 Introduction

In the past decade, approximate factor models have been increasingly important in the analysis of high-dimensional macroeconomic datasets, as a small number of common factors can summarize the comovements of a large number of series and the models are found useful in forecasting and policy applications (Bai & Ng, 2008b; Stock & Watson, 2011). Traditionally, factor loadings of each variable are assumed constant over time. Under this assumption, various estimators of the number of factors as well as the principal components estimator of the factors can be consistent within the framework of large observations in both the time and cross-section dimensions. The assumption of loading constancy is however not supported by empirical evidence (Stock & Watson, 2009). Given temporal instability in the factor loadings, the estimators can still be consistent provided that the magnitude of instability is not exceeding some upper bound shown by Bates *et al.* (2013). If time variation is stronger, however, the inference would be misleading and the forecasts based on the principal components estimator of the factors and the estimator of the number of factors might be affected. Testing for the constancy of factor loadings is therefore necessary before every serious application of factor models. The literature on testing for loading stability has focused on structural breaks, including Breitung & Eickmeier (2011), Yamamoto & Tanaka (2013), Chen *et al.* (2014), Corradi & Swanson (2014), Cheng *et al.* (2014) and Han & Inoue (2015). Breitung & Eickmeier (2011) and Yamamoto & Tanaka (2013) tested whether the factor loadings of one specific variable have breaks, while the others concerned the shifts in the factor loadings of all variables.

In this chapter, we test against a martingale-type alternative which allows for the gradual shifts in the factor loadings over time. We focus on martingale-type instability rather than structural breaks for four reasons. Firstly, Hansen (1992) argued that specifying the parameter process as a martingale is more appropriate if one is interested in testing whether the model captures a stable relationship. Secondly, it is well known that a factor model with structural breaks in the factor loadings can be equivalently represented by a stable factor model with a larger number of factors. It follows that structural breaks in factor loadings have no theoretical influence on the forecasts based on factor-augmented regressions, if the number of factors is not ex ante imposed but is instead estimated by the information criteria of Bai & Ng (2002), for example. Factor models with martingale-type loadings, however, have no equivalent representations of stable models and the number of factors can not generally be consistently estimated using the existing methods (Takongmo & Stevanovic, 2014). Martingale-type time variation, in this sense, is a more serious issue than structural breaks that would potentially affect the forecasts based on the estimators of the factors. Thirdly, Del Negro & Otrok (2008) estimated a factor model with time-varying factor loadings using the Bayesian technique. If there is indeed strong instability of the type we aim to test against, their model might be a better candidate than stable models for the purpose of forecasting. Finally, it is known that a model with multiple breaks is just a special case of the models with time-varying parameters in the form of martingales.

Testing parameter stability against the martingale-type alternative was generally addressed in a likelihood framework in Nyblom (1989), which was extended to the models with integrated regressors in Hansen (1992) and to panel data models in Yamazaki &

Kurozumi (2014). In this chapter, different from the literature on time-varying parameters, the alternative of time variation in the form of martingales is assumed to start at a known change point, which is not necessarily at the beginning of the sample. The main challenge of adapting the general framework for high-dimensional factor models is the accumulation of estimation error of common factors over time that might change the conventional distributions of the test statistics. As in Breitung & Eickmeier (2011), in order to obtain the asymptotic distribution of the test statistic, we derive the conditions on the relative rates of the number of cross sections and the number of time periods under which the estimation error of common factors is asymptotically negligible for the test statistic. Generally the number of cross sections is required to be large enough compared to the size of time periods after the change point. Our test aims to have nontrivial power against the instabilities that are not tolerated by the principal components estimator of the factors and the estimators of the number of factors. Simulation studies show that the proposed test has good performance in finite samples and is more powerful than the tests designed for structural breaks to detect unstable factor structure. We apply our test to a panel of macroeconomic and financial data in the UK and find considerable time variation during the recent financial crisis.

The rest of the chapter is structured as follows. Section 4.2 develops the test statistic, introduces the assumptions and derives the asymptotic distributions of the test statistic. Section 4.3 concerns the model with integrated factors. Section 4.4 presents the Monte Carlo results of the proposed test, and compares the power of our test with that of the tests designed for structural breaks. We apply the test to a UK dataset in Section 4.5. Section 4.6 concludes the chapter. The appendix is given in Section 4.7.

## 4.2 The Model and Test Statistic for Stationary Factors

### 4.2.1 The model and test statistic

We study a factor model with individual-specific time-varying factor loadings given by

$$X_{jt} = \lambda_{jt}' F_t + u_{jt},$$

where $X_{jt}$ is observed for $j = 1, \ldots, N$, $t = 1, \ldots, T$; $F_t$ is a $k-$dimensional vector of unobserved common factors; $\lambda_{jt}$ is the vector of factor loadings of the $j$th variable at time $t$; and $u_{jt}$ is the idiosyncratic error. The number of factors $k$ is unobserved. The factor loadings of a given variable $i$ can evolve as the random walk after some known change point $T_a$, while those of the other variables are constants over time. Specifically,

$$\begin{aligned} \lambda_{it} &= \begin{cases} \lambda_{i1} & t \leq T_a \\ \lambda_{i,t-1} + e_{it} & t > T_a \end{cases} \\ \lambda_{jt} &= \lambda_{j1} \quad t = 1, \ldots, T;\ j = 1, \ldots, i-1, i+1, \ldots, N. \end{aligned}$$

We assume $\mathrm{E}(e_{it} e_{it}') = h_{NT} \Sigma_e$ where the scalar $h_{NT}$ indicates the intensity of temporal instability and might depend on $N$ and $T$, and $\Sigma_e$ is known and positive definite. Under this assumption, the null hypothesis is that all factor loadings are constant over time:

$$H_0 : h_{NT} = 0,$$

while the alternative hypothesis is that the factor loadings of the $i$th variable evolve as a martingale process after a change point:

$$H_1 : h_{NT} > 0.$$

Under the null hypothesis, the estimators of the number of factors in Bai & Ng (2002) and the principal components estimator of the factors (Bai, 2003) are consistent under some assumptions as $N, T \to \infty$. Under the alternative hypothesis, Bates *et al.* (2013) showed in general $\sqrt{h_{NT}} = O(1/\min(T, (NT)^{1/2}))$ is needed for consistency of Bai & Ng (2002)'s estimators and $\sqrt{h_{NT}} = o(T^{-1/2})$ is required for mean square consistency of the principal components estimator. Thus, the desired tests should have non-trivial power against the instabilities which are not tolerated by the estimators, such as $h_{NT} = a/T$ with $a$ fixed. Our alternative hypothesis differs from the one in the literature of time-varying parameters in that $T_a$ is not necessarily equal to 1. When $t > T_a$, the model can be written in the vector form as

$$X_t = \Lambda_1 F_t + v_t + u_t,$$

where $X_t = (X_{1t}, \ldots, X_{Nt})'$, $\Lambda_1 = (\lambda_{11}, \ldots, \lambda_{N1})'$, $v_t = (0, \ldots F_t' \sum_{\tau=T_a+1}^{t} e_{i\tau}, \ldots, 0)'$, and $u_t = (u_{1t}, \ldots, u_{Nt})'$ is independently and identically distributed (i.i.d.) over $t$ with $\mathrm{E}(u_t u_t') = V$.

We adapt the Lagrange Multiplier (LM) test of Nyblom (1989) for factor models. Define $\tilde{T} = T - T_a$, $\underline{X} = (X_{T_a+1}, \ldots, X_T)'$, $\underline{F} = (F_{T_a+1}, \ldots, F_T)'$, $D_F = \mathrm{diag}(I_N \otimes F_{T_a+1}', \ldots, I_N \otimes$

$F_T')$ and $L$ is a $\tilde{T}\times\tilde{T}$ lower triangular matrix with all non-zero elements equal to 1. Let

$$\Sigma = I_{\tilde{T}}\otimes V + h_{NT}D_F(L\otimes I_{kN})(I_{\tilde{T}}\otimes\tilde{\Sigma}_e)(L'\otimes I_{kN})D_F'$$

denote the conditional covariance matrix of $\mathrm{vec}(\underline{X}')$ given $\underline{F}$ where $\tilde{\Sigma}_e$ is a $kN\times kN$ sparse matrix with the only non-zero $k\times k$ block from the $ik+1$ th element to the $(i+1)k$ th element on the diagonal being equal to $\Sigma_e$. Under normality, the joint log-likelihood of $(X_{T_a+1}',\ldots,X_T')$ given $(F_{T_a+1}',\ldots,F_T')$ is

$$\begin{aligned} l_{\mathrm{vec}(\underline{X}')|\underline{F}}(\Lambda_1,h_{NT},V,\Sigma_e) &= -\frac{N\tilde{T}}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma(h_{NT})| \\ &\quad -\frac{1}{2}[\mathrm{vec}(\underline{X}')-\mathrm{vec}(\Lambda_1\underline{F}')]'\Sigma^{-1}[\mathrm{vec}(\underline{X}')-\mathrm{vec}(\Lambda_1\underline{F}')]. \end{aligned}$$

The maximum likelihood estimator of $\Lambda_1$ under the null hypothesis $h_{NT}=0$ is $\hat{\Lambda}_1 = \underline{X}'\underline{F}(\underline{F}'\underline{F})^{-1}$. Let

$$f(X_t|F_t;\hat{\Lambda}_1,V) = -\frac{N}{2}\log(2\pi)-\frac{1}{2}\log|V|-\frac{1}{2}[X_t-\hat{\Lambda}_1F_t)]'V^{-1}[X_t-\hat{\Lambda}_1F_t]$$

denote the marginal density of $X_t$ given $F_t$ under the null and let $C_i$ be a $N\times N$ diagonal matrix with the $i$th diagonal element being 1 and the others being 0. Define $G=C_i\otimes\Sigma_e=(C_i\otimes\Sigma_{ec})(C_i\otimes\Sigma_{ec}')$ where $\Sigma_e=\Sigma_{ec}\Sigma_{ec}'$ is the Cholesky decomposition of $\Sigma_e$. The self-normalized statistic adapted from Nyblom (1989) can be expressed as

$$\mathrm{LM}_i = \frac{1}{\tilde{T}^2V_{ii}^{-1}}\sum_{\tau=1}^{\tilde{T}}\sum_{t=T_a+\tau}^{T}\frac{\partial f(X_t|F_t)'}{\partial\mathrm{vec}(\hat{\Lambda}_1')}G\sum_{t=T_a+\tau}^{T}\frac{\partial f(X_t|F_t)}{\partial\mathrm{vec}(\hat{\Lambda}_1')},$$

where $V_{ii}^{-1}$ denote the $i$th diagonal element of $V^{-1}$. Let $V_{i.}^{-1}$ denote the $i$th row of $V^{-1}$. It follows that

$$\frac{\partial f(X_t|F_t)}{\partial \text{vec}(\hat{\Lambda}_1')} \quad = \quad \text{vec}(F_t X_t' V^{-1}) - \text{vec}(F_t F_t' (\underline{F}'\underline{F})^{-1}\underline{F}'\underline{X}V^{-1}),$$

and

$$\begin{aligned}\sum_{t=T_a+\tau}^{T}(C_i \otimes \Sigma_{ec}')\frac{\partial f(X_t|F_t)}{\partial \text{vec}(\hat{\Lambda}_1')} &= \Sigma_{ec}'[\sum_{t=T_a+\tau}^{T} F_t V_{i.}^{-1} X_t - \\ &\quad \sum_{t=T_a+\tau}^{T} F_t F_t' (\sum_{t=T_a+1}^{T} F_t F_t')^{-1} \sum_{t=T_a+1}^{T} F_t V_{i.}^{-1} X_t].\end{aligned}$$

From these results, we obtain

$$\text{LM}_i = \frac{1}{\tilde{T}^2 V_{ii}^{-1}} \sum_{\tau=1}^{\tilde{T}} s_{i\tau}' \Sigma_e s_{i\tau},$$

where

$$\begin{aligned}s_{i\tau} &= \sum_{t=T_a+1}^{T_a+\tau} F_t V_{i.}^{-1}(u_t + v_t) - \sum_{t=T_a+1}^{T_a+\tau} F_t F_t' (\sum_{t=T_a+1}^{T} F_t F_t')^{-1} \sum_{t=T_a+1}^{T} F_t V_{i.}^{-1}(u_t + v_t) \\ &= \sum_{t=T_a+1}^{T_a+\tau} F_t V_{i.}^{-1} X_t - \sum_{t=T_a+1}^{T_a+\tau} F_t F_t' (\sum_{t=T_a+1}^{T} F_t F_t')^{-1} \sum_{t=T_a+1}^{T} F_t V_{i.}^{-1} X_t.\end{aligned}$$

Note that $\text{LM}_i$ contains unobserved common factors and unknown model parameters including the number of factors $k$ and the covariance matrix $V$. Given a consistent estimator of the number of factors, $\hat{k}$, the principal components estimator of the factors $\hat{F}$ is $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{k}$ largest eigenvalues of $XX'$ where $X = (X_1, \ldots, X_T)'$. Let $\hat{V}_{i.}^{-1}$ denote the $i$th row of $\hat{V}^{-1}$, where $\hat{V}$ is a consistent estimator of $V$. After replacing

$\Sigma_e$ with $\hat{\Sigma}_F^{-1}$ where $\hat{\Sigma}_F = \frac{1}{T}\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'$ in order to remove the nuisance parameters in the asymptotic distribution (see more discussions in Section 4.2.2), the computable test statistic $\widehat{\mathrm{LM}}_i$ can be written as

$$\widehat{\mathrm{LM}}_i = \frac{1}{\tilde{T}^2 \hat{V}_{ii}^{-1}} \sum_{\tau=1}^{\tilde{T}} \hat{s}_{i\tau}' \hat{\Sigma}_F^{-1} \hat{s}_{i\tau},$$

where

$$\hat{s}_{i\tau} = \sum_{t=T_a+1}^{T_a+\tau} \hat{F}_t \hat{V}_{i.}^{-1} X_t - \sum_{t=T_a+1}^{T_a+\tau} \hat{F}_t \hat{F}_t' ( \sum_{t=T_a+1}^{T} \hat{F}_t \hat{F}_t')^{-1} \sum_{t=T_a+1}^{T} \hat{F}_t \hat{V}_{i.}^{-1} X_t.$$

### 4.2.2 Asymptotics for the test statistic

In order to derive the asymptotic distribution of the test statistic, we make the following assumptions on common factors, factor loadings and idiosyncratic errors. Let $\| A \| = \sqrt{\mathrm{tr}(A'A)}$ be the Frobenius norm, and let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a symmetric matrix $A$, respectively.

**Assumptions**

1. (Common factors) $\{F_t\}$ is stationary and ergodic, $\mathrm{E}(F_t F_t') = \Sigma_F$ for some positive definite $\Sigma_F$ and $\mathrm{E} \| F_t \|^4 < \infty$ for $t = 1, \dots, T$.

2. (Initial factor loadings) $\| \lambda_{j1} \| \leq \overline{\lambda} < \infty$ for $j = 1, \dots, N$ and $||\Lambda_1' \Lambda_1 / N - \Sigma_\Lambda|| \to 0$ as $N \to \infty$ for some positive definite $\Sigma_\Lambda$.

3. (Factor loading innovations of the $i$th variable )

   (a) $e_{it}$ is i.i.d. across $t$, $\mathrm{E}(e_{it}) = 0$ and $\mathrm{E}(e_{it} e_{it}') = h_{NT} \Sigma_e$ where $\Sigma_e$ is known and positive definite.

(b) $e_{it}$ is independent of both $F_s$ and $u_{js}$ for all $t, s, j$.

4. (Idiosyncratic errors) $M < \infty$ is a positive constant. For all $N$, $T$ and $\tilde{T}$,

   (a) $u_t$ is i.i.d., $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t u_t') = V$ and $\mathrm{E}(u_{jt}^8) < \infty$ for $j = 1, \ldots, N$.

   (b) $\max_j \sum_{k=1}^N |V_{kj}| \leq M$ for all $N$.

   (c) $\mathrm{E}|N^{-1/2} \sum_{j=1}^N [u_{js} u_{jt} - \mathrm{E}(u_{js} u_{jt})]|^4 \leq M$ for every $s$ and $t$.

   (d) For every $t$, $\mathrm{E}||\frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{j=1}^N F_s [u_{js} u_{jt} - \mathrm{E}(u_{js} u_{jt})]||^2 \leq M$.

   (e) For every $t$, $\mathrm{E}||\frac{1}{\sqrt{N}} \sum_{j=1}^N \lambda_{j1} u_{jt}||^4 \leq M$ .

   (f) $u_{jt}$ and $F_s$ are independent for all $j, t, s$.

   (g) Let $V_{ij}^{-1}$ denote the $ij$th element of $V^{-1}$.

      i. The number of non-zero elements in $V_{i.}^{-1}$ is $r_{i,N}$ which is finite when $N \to \infty$.

      ii. The number of non-zero elements in $\hat{V}_{i.}^{-1}$ is $\hat{r}_{i,N}$ which is finite almost surely when $N \to \infty$.

      iii. $\hat{V}_{ij}^{-1} \xrightarrow{p} V_{ij}^{-1}$ for each $i, j, N$.

   (h) $\frac{1}{\tilde{T}} \sum_{\tau=1}^{\tilde{T}} ||\frac{1}{\sqrt{N\tilde{T}}} \sum_{s=1}^{\tau} \sum_{j=1}^N \lambda_{j1} (u_{js} u_{is} - \mathrm{E}(u_{js} u_{is}))||^2 = O_p(1)$ for each $i$.

**Comments on the assumptions:**

**Part of Assumption 1, Assumptions 2 and 4(b)-(e)**: These assumptions are borrowed from Bai (2003) and are aimed at consistent estimation of the common factors and the number of factors under the null hypothesis. Generally idiosyncratic errors are allowed to be weakly dependent and heteroskedastic in the cross-section dimension.

**Assumption 4(a)**: The assumption of i.i.d. idiosyncratic errors in the time dimension is made for convenience only. It can be weakened to allow for limited time serial dependence and heteroskedasticity.

**Assumption 4(gi)**: Examples include cross-section heteroskedasticity with no dependence and cross-section heteroskedasticity with block-diagonal correlation structure of fixed block size (Choi, 2012), among others. The block-diagonal structure is motivated by the fact that in macroeconomic or financial applications the idiosyncratic components often represent industry-specific shocks and are almost uncorrelated among the variables across different industries (Bai & Liao, 2013).

**Assumptions 4(gii) and 4(giii)**: Let $\hat{u}_t$ denote the residual vector from the principal components estimation. The sample error covariance matrix $\hat{V} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ can satisfy these conditions provided we have the prior knowledge on the structure of the idiosyncratic covariance matrix (e.g. block diagonal)(Choi, 2012). .

**Assumptions 4(h)**: This condition is not so strict as all sums include variables with zero means.

Under the null hypothesis and Assumptions 1 and 4, as $\tilde{T} \to \infty$, $\frac{1}{\sqrt{\tilde{T}}}\sum_{t=T_a+1}^{T_a+[\tilde{T}r]} F_t V_{i.}^{-1} u_t \xrightarrow{\text{D}} \sqrt{V_{ii}^{-1}}\Sigma_F^{1/2} W(r)$ where $W(\cdot)$ stands for a $k$-dimensional standard Brownian motion. As a result,

$$\frac{1}{\sqrt{\tilde{T} V_{ii}^{-1}}} s_{i,[\tilde{T}r]} \xrightarrow{\text{D}} \Sigma_F^{1/2}(W(r) - rW(1)),$$

and

$$\mathrm{LM}_i \xrightarrow{\mathrm{D}} \int_0^1 (W(r) - rW(1))' \Sigma_F^{1/2} \Sigma_e \Sigma_F^{1/2} (W(r) - rW(1)) dr.$$

As is well known, one can only consistently estimate the $k-$dimensional space spanned by the true factors. Let $V_{NT}$ denote the $k \times k$ diagonal matrix of the first $k$ largest eigenvalues of $XX'/(TN)$. The principal components estimator $\hat{F}$ is a consistent estimator of the transformed factors $\tilde{F} = FH$ where $F = (F_1, \ldots, F_T)'$ and $H = (\Lambda_1' \Lambda_1 / N)(F' \hat{F} / T) V_{NT}^{-1}$ (Bai, 2003). Rewriting $\mathrm{LM}_i$ by replacing $F_t$ with $\tilde{F}_t$ yields $\underset{\sim}{LM}_i = \frac{1}{\tilde{T}^2 V_{ii}^{-1}} \sum_{\tau=1}^{\tilde{T}} \tilde{s}_{i\tau}' \Sigma_e \tilde{s}_{i\tau}$ where

$$\tilde{s}_{i\tau} = \sum_{t=T_a+1}^{T_a+\tau} \tilde{F}_t V_{i.}^{-1} X_t - \sum_{t=T_a+1}^{T_a+\tau} \tilde{F}_t \tilde{F}_t' ( \sum_{t=T_a+1}^{T} \tilde{F}_t \tilde{F}_t' )^{-1} \sum_{t=T_a+1}^{T} \tilde{F}_t V_{i.}^{-1} X_t.$$

As $N, \tilde{T} \to \infty$,

$$\underset{\sim}{LM}_i \xrightarrow{\mathrm{D}} \int_0^1 (W(r) - rW(1))' \Sigma_F^{1/2} \tilde{H} \Sigma_e \tilde{H}' \Sigma_F^{1/2} (W(r) - rW(1)) dr,$$

where $\tilde{H} = \operatorname*{plim}_{N,T \to \infty} H$. The limit distribution of $\underset{\sim}{LM}_i$ can be free from nuisance parameters if $\Sigma_e$ is replaced with $\tilde{\Sigma}_F^{-1}$ where $\tilde{\Sigma}_F = \tilde{H}' \Sigma_F \tilde{H} = \operatorname*{plim}_{N,T \to \infty} (\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t')$. After replacement, the new statistic is indicated by $\widetilde{\mathrm{LM}}_i = \frac{1}{\tilde{T}^2 V_{ii}^{-1}} \sum_{\tau=1}^{\tilde{T}} \tilde{s}_{i\tau}' \tilde{\Sigma}_F^{-1} \tilde{s}_{i\tau}$. As $N, \tilde{T} \to \infty$,

$$\widetilde{\mathrm{LM}}_i \xrightarrow{\mathrm{D}} \int_0^1 (W(r) - rW(1))' (W(r) - rW(1)) dr,$$

which is the generalized Von Mises distribution with $k$ degrees of freedom (parts of the critical values are listed in Table 12). We are interested in the condition on the relative rates of $N, T, \tilde{T}$ under which the estimation error of factors and related variables is negligible for

| | Significance Level | | |
|---|---|---|---|
| $k$ | 1% | 5% | 10% |
| 1 | 0.748 | 0.470 | 0.353 |
| 2 | 1.070 | 0.749 | 0.610 |
| 3 | 1.350 | 1.010 | 0.846 |
| 4 | 1.600 | 1.240 | 1.070 |
| 5 | 1.880 | 1.470 | 1.280 |
| 6 | 2.120 | 1.680 | 1.490 |
| 7 | 2.350 | 1.900 | 1.690 |
| 8 | 2.590 | 2.110 | 1.890 |
| 9 | 2.820 | 2.320 | 2.100 |
| 10 | 3.050 | 2.540 | 2.290 |

Table 12: Parts of the critical values of the generalized Von Mises distribution with $k$ degrees of freedom (Source: Hansen (1990), Table 1)

the asymptotic distribution of $\widehat{\mathrm{LM}}_i$.

**Theorem 1.** Under the null hypothesis and Assumptions 1-4, if $\frac{\sqrt{\tilde{T}}}{\sqrt{\min(N^2, N\tilde{T}, T^2)}} = o(1)$, then $\widehat{\mathrm{LM}}_i = \widetilde{\mathrm{LM}}_i + o_p(1)$ as $N, \tilde{T} \to \infty$. Thus, $\widehat{\mathrm{LM}}_i \xrightarrow{\mathrm{D}} \int_0^1 (W(r) - rW(1))'(W(r) - rW(1))dr$.

**Comments on Theorem 1:** If $\tilde{T} = T$, the condition reduces to $\frac{\sqrt{T}}{\sqrt{\min(N^2, T^2)}} = o(1)$, which is same as the condition for the break tests in Breitung & Eickmeier (2011) if $N \leq T$. Generally, $N$ is required to be large enough for the estimator error of factors to become asymptotically negligible.

As mentioned before, $\sqrt{h_{NT}} = o(T^{-1/2})$ is required for mean square consistency of the principal components estimator. Thus, we are interested in investigating the asymptotic distribution of the test statistic under the local alternative. Specifically, in order for the limit distribution under the alternative to be well defined, we choose $h_{NT} = a/\tilde{T}^2$ with $a$ fixed. When the order of $h_{NT}$ is larger, the test statistic will diverge. Under the assumptions, as $\tilde{T} \to \infty$, $\frac{1}{\sqrt{\tilde{T}}} \sum_{t=T_a+1}^{T_a+[\tilde{T}r]} F_t V_{i.}^{-1} v_t \xrightarrow{\mathrm{D}} \sqrt{a} V_{ii}^{-1} \Sigma_F \Sigma_e^{1/2} W_2(r)$ where $W_2(r) = \int_0^r W(s)ds$. It follows that

$$\text{LM}_i \xrightarrow{\text{D}} \frac{1}{V_{ii}^{-1}} \int_0^1 W_3'(r)\Sigma_e W_3(r)dr,$$

where $W_3(r) = \sqrt{a}V_{ii}^{-1}\Sigma_F\Sigma_e^{1/2}(W_2(r) - rW_2(1)) + \sqrt{V_{ii}^{-1}}\Sigma_F^{1/2}(W(r) - rW(1))$. Furthermore,

$$\widetilde{\text{LM}}_i \xrightarrow{\text{D}} \frac{1}{V_{ii}^{-1}} \int_0^1 W_3'(r)\Sigma_F^{-1} W_3(r)dr.$$

Note that the alternative distribution depends on $V_{ii}$, $\Sigma_F$ and $\Sigma_e$. The condition under which the estimation error of factors is insignificant for the asymptotic alternative distribution of the test statistic is same as that stated in Theorem 1.

**Theorem 2.** Given Assumptions 1-4, under the local alternative $h_{NT} = a/\tilde{T}^2$ with $a$ fixed, if $\frac{\sqrt{\tilde{T}}}{\sqrt{\min(N^2, N\tilde{T}, T^2)}} = o(1)$, then $\widehat{\text{LM}}_i = \widetilde{\text{LM}}_i + o_p(1)$ as $N, \tilde{T} \to \infty$. Therefore we have $\widehat{\text{LM}}_i \xrightarrow{\text{D}} \frac{1}{V_{ii}^{-1}} \int_0^1 W_3'(r)\Sigma_F^{-1} W_3(r)dr$.

## 4.3 The Model and Test Statistic for Integrated Factors

In this section, common factors are assumed to be integrated of order one with no cointegration among them. Specifically, $F_t = F_{t-1} + \varepsilon_t$ for $t = 1, \ldots, T$ where $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$ and $\varepsilon_t$ is independent of $e_{is}$ and $u_{js}$ for all $t, s, j$. As a result of integration, Assumption 4(d) does not hold any longer. Now the statistic is defined as a function of the relative change point $\pi$:

$$\text{LM}_i(\pi) = \frac{1}{\tilde{T}T^2\hat{V}_{ii}^{-1}} \sum_{\tau=1}^{T-[T\pi]} s_{i\tau}'(\pi)\Sigma_e s_{i\tau}(\pi),$$

where $s_{i\tau}(\pi) = \sum_{t=[T\pi]+1}^{[T\pi]+\tau} F_t V_{i.}^{-1} X_t - \sum_{t=[T\pi]+1}^{[T\pi]+\tau} F_t F_t' (\sum_{t=[T\pi]+1}^{T} F_t F_t')^{-1} \sum_{t=[T\pi]+1}^{T} F_t V_{i.}^{-1} X_t$.

Under Assumptions 4, $\frac{1}{T^2}\sum_{t=[T\pi]+1}^{[T\pi]+[\tilde{T}r]} F_t F_t' \xrightarrow{\text{D}} \Sigma_\varepsilon^{1/2} \int_\pi^{r\tilde{T}/T+\pi} W(\tilde{r})W'(\tilde{r})d\tilde{r}\Sigma_\varepsilon^{1/2}$ and $\frac{1}{T}\sum_{t=[T\pi]+1}^{[T\pi]+[\tilde{T}r]} F_t V_{i.}^{-1} u_t \xrightarrow{\text{D}} \int_\pi^{r\tilde{T}/T+\pi} \sqrt{V_{ii}^{-1}} \Sigma_\varepsilon^{1/2} W(\tilde{r})dW_4(\tilde{r})$ where $W_4(\cdot)$ stands for a standard Brownian motion and is independent of $W(\cdot)$, so

$$\frac{1}{T\sqrt{V_{ii}^{-1}}} s_{i,[\tilde{T}r]}(\pi) \xrightarrow{\text{D}} \Sigma_\varepsilon^{1/2}\tilde{W}(r,\pi),$$

where

$$\begin{aligned}
\tilde{W}(r,\pi) &= \int_\pi^{r\tilde{T}/T+\pi} W(\tilde{r})dW_4(\tilde{r}) \\
&\quad - \int_\pi^{r\tilde{T}/T+\pi} W(\tilde{r})W'(\tilde{r})d\tilde{r}\Big(\int_\pi^1 W(\tilde{r})W'(\tilde{r})d\tilde{r}\Big)^{-1}\int_\pi^1 W(\tilde{r})dW_4(\tilde{r}).
\end{aligned}$$

Consequently,

$$\text{LM}_i(\pi) \xrightarrow{\text{D}} \int_0^1 \tilde{W}(r,\pi)'\Sigma_\varepsilon^{1/2}\Sigma_e\Sigma_\varepsilon^{1/2}\tilde{W}(r,\pi)dr.$$

In the case of integrated factors, the principal components estimator $\hat{F}_I$ becomes $T$ times the eigenvectors corresponding to the $k$ largest eigenvalues of the matrix $XX'$, as in Bai (2004). Let $V_{NT}$ denote the $k\times k$ diagonal matrix of the first $k$ largest eigenvalues of $XX'/(T^2N)$. As in the case of stationary factors, one can transform the factors as $\tilde{F}_{I,t} = H_I' F_t$ where $H_I = (\Lambda_1'\Lambda_1/N)(F'\hat{F}_I/T^2)V_{NT}^{-1}$. We define $\widetilde{\text{LM}}_i(\pi) = \frac{1}{\tilde{T}T^2 V_{ii}^{-1}} \sum_{\tau=1}^{T-[T\pi]} \tilde{s}_{i\tau}(\pi)\tilde{\Sigma}_\varepsilon^{-1}\tilde{s}_{i\tau}(\pi)$ where $\tilde{s}_{i\tau}(\pi) = \sum_{t=[T\pi]+1}^{[T\pi]+\tau} \tilde{F}_{I,t}V_{i.}^{-1}X_t - \sum_{t=[T\pi]+1}^{[T\pi]+\tau} \tilde{F}_{I,t}\tilde{F}_{I,t}'(\sum_{t=[T\pi]+1}^{T} \tilde{F}_{I,t}\tilde{F}_{I,t}')^{-1}\sum_{t=[T\pi]+1}^{T} \tilde{F}_{I,t}V_{i.}^{-1}X_t$ and

$$\widetilde{\text{LM}}_i(\pi) \xrightarrow{\text{D}} \int_0^1 \tilde{W}(r,\pi)'\tilde{W}(r,\pi)dr.$$

The distribution of $\int_0^1 \tilde{W}(r,\pi)'\tilde{W}(r,\pi)dr$ depends on the relative change point $\pi$ and the di-

| | $k=1$ | | | $k=2$ | | | $k=3$ | | | $k=4$ | | | $k=5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\pi$ | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| 0.4 | 0.559 | 0.242 | 0.149 | 0.644 | 0.308 | 0.207 | 0.637 | 0.323 | 0.228 | 0.618 | 0.318 | 0.234 | 0.546 | 0.313 | 0.232 |
| 0.45 | 0.554 | 0.234 | 0.143 | 0.628 | 0.299 | 0.200 | 0.615 | 0.312 | 0.219 | 0.576 | 0.307 | 0.222 | 0.528 | 0.296 | 0.221 |
| 0.5 | 0.532 | 0.224 | 0.139 | 0.611 | 0.285 | 0.190 | 0.594 | 0.296 | 0.208 | 0.545 | 0.290 | 0.209 | 0.499 | 0.280 | 0.208 |
| 0.55 | 0.494 | 0.216 | 0.129 | 0.558 | 0.268 | 0.179 | 0.541 | 0.278 | 0.193 | 0.507 | 0.270 | 0.195 | 0.482 | 0.262 | 0.194 |
| 0.6 | 0.476 | 0.204 | 0.123 | 0.526 | 0.252 | 0.167 | 0.519 | 0.257 | 0.178 | 0.475 | 0.249 | 0.180 | 0.441 | 0.241 | 0.177 |
| 0.65 | 0.429 | 0.185 | 0.113 | 0.488 | 0.228 | 0.151 | 0.461 | 0.235 | 0.163 | 0.427 | 0.227 | 0.162 | 0.399 | 0.217 | 0.161 |
| 0.7 | 0.384 | 0.166 | 0.100 | 0.426 | 0.204 | 0.137 | 0.412 | 0.208 | 0.144 | 0.383 | 0.200 | 0.145 | 0.350 | 0.194 | 0.142 |
| 0.75 | 0.336 | 0.145 | 0.089 | 0.379 | 0.177 | 0.118 | 0.362 | 0.180 | 0.124 | 0.337 | 0.172 | 0.124 | 0.298 | 0.165 | 0.121 |
| 0.8 | 0.285 | 0.123 | 0.074 | 0.316 | 0.148 | 0.098 | 0.296 | 0.150 | 0.103 | 0.271 | 0.143 | 0.103 | 0.255 | 0.137 | 0.100 |

Table 13: Critical values of the distribution $\int_0^1 \tilde{W}(r,\pi)'\tilde{W}(r,\pi)dr$ by simulation.

mension $k$. The critical values listed in Table 13 are obtained by simulating the distribution of $\int_0^1 \tilde{W}(r,\pi)'\tilde{W}(r,\pi)dr$ for the dimension $k$ from 1 to 5 and a selection of $\pi$ from 0.4 to 0.8. We obtain a single realization from the distribution by simulating a $k$-dimensional Brownian motion and a standard Brownian motion which are independent of each other. We compute the quantiles from 500 realizations and then average them over 100 repetitions.

As for the case of stationary factors, we estimate $\widetilde{\text{LM}}_i(\pi)$ by $\widehat{\text{LM}}_i(\pi)$ such that

$$\widehat{\text{LM}}_i(\pi) = \frac{1}{\tilde{T}T^2\hat{V}_{ii}^{-1}} \sum_{\tau=1}^{T-[T\pi]} \hat{s}_{i\tau}(\pi)\hat{\Sigma}_\varepsilon^{-1}\hat{s}_{i\tau}(\pi),$$

where $\hat{s}_{i\tau}(\pi) = \sum_{t=[T\pi]+1}^{[T\pi]+\tau} \hat{F}_{I,t}\hat{V}_{i.}^{-1}X_t - \sum_{t=[T\pi]+1}^{[T\pi]+\tau} \hat{F}_{I,t}\hat{F}_{I,t}'(\sum_{t=[T\pi]+1}^{T} \hat{F}_{I,t}\hat{F}_{I,t}')^{-1} \sum_{t=[T\pi]+1}^{T} \hat{F}_{I,t}\hat{V}_{i.}^{-1}X_t$ and $\hat{\Sigma}_\varepsilon = \frac{1}{T-1}\sum_{t=1}^{T-1}(\hat{F}_{I,t+1} - \hat{F}_{I,t})(\hat{F}_{I,t+1} - \hat{F}_{I,t})'$.

**Theorem 3.** Under the null hypothesis, $\widehat{\text{LM}}_i(\pi) = \widetilde{\text{LM}}_i(\pi) + O_p(\frac{1}{\sqrt{\min(N,T^2)}})$ for any $\pi$ as $N, \tilde{T} \to \infty$. Thus, $\widehat{\text{LM}}_i(\pi) \xrightarrow{D} \int_0^1 \tilde{W}(r,\pi)'\tilde{W}(r,\pi)dr$.

## 4.4 Monte Carlo Results

The Monte Carlo results are presented in this section. We consider a two-factor model with $N = 50, 100, 150, 200$ and $T = 100, 150, 200, 250$, which are typical in empirical applications. $\{F_t\}$ is either i.i.d. $N(0, I_2)$ or random walk with $F_0 = [0, 0]'$ and $N(0, I_2)$ error terms. In all cases, we estimate the number of factors using Onatski (2010)'s edge distribution estimator. That is, $\hat{k} = \max\{k \leq k_{max}, \varpi_k - \varpi_{k+1} \geq c\}$ where $k_{max}$ is a given upper bound (set to 10 in our study), $\varpi_k$ is the $k$th largest eigenvalue of $X'X/T$ and $c$ is a threshold value estimated from the empirical distribution of eigenvalues. The idiosyncratic structure is block diagonal, and the size of each block is 5 (Choi, 2012). We generate each block by i.i.d. normal distribution. The idiosyncratic covariance matrix is estimated by the sample error covariance matrix $\hat{V} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$. It is a consistent estimator under the assumption of prior knowledge of the block diagonal structure.

Factor loadings for the variable $i$ are generated by the random walk

$$\lambda_{it} = \begin{cases} \lambda_{i1} & t \leq T_a \\ \lambda_{i,t-1} + e_{it} & t > T_a, \end{cases}$$

where $\lambda_{i1}$ is i.i.d.$N(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, I_2)$; $\mathrm{E}(e_{it}e_{it}') = \frac{a}{T}\begin{bmatrix} 1 & \sigma_\lambda^2 \\ \sigma_\lambda^2 & 1 \end{bmatrix}$ such that $a = 0, 1, 5, 10, 20$ and $\sigma_\lambda^2$ follows $\texttt{Uniform}(0, 0.9)$;[17] $T_a$ is set equal to $\frac{T}{2}$.

We use 5000 replications for Monte Carlo simulation. Table 14 shows the results of

17

$i$ is fixed in each experiment and is taken across $1, \ldots, N$.

| | I(0) Factors | | | I(1) Factors | | |
|---|---|---|---|---|---|---|
| | $T = 100$ | $T = 150$ | $T = 200$ | $T = 100$ | $T = 150$ | $T = 200$ |
| $N = 50$ | 0.069 | 0.077 | 0.081 | 0.062 | 0.057 | 0.066 |
| $N = 100$ | 0.067 | 0.071 | 0.082 | 0.055 | 0.055 | 0.053 |
| $N = 150$ | 0.070 | 0.068 | 0.071 | 0.051 | 0.050 | 0.051 |
| $N = 200$ | 0.063 | 0.070 | 0.071 | 0.052 | 0.049 | 0.049 |

Table 14: Empirical sizes (nominal size is 0.05).

empirical sizes. When common factors are stationary, the test tends to be oversized, due to the finite sample bias of the estimator of the number of factors. If the number of factors is set equal to the true value, the empirical sizes would become closer to the nominal size of 0.05. In the case of integrated factors, the rejection frequencies turn out to be close to the nominal size. Tables 15 and 16 report the results on the power of the tests for the models with stationary factors and integrated factors respectively. In both cases, it seems that the tests are consistent in that the power grows as $N$ and $T$ increase. When factors are integrated, the test suffers from the issue of nonmonotonic power, which is not unusual in the tests for structural stability.

The tests designed for structural breaks, such as likelihood ratio (LR), Wald and Lagrange multiplier statistics in Breitung & Eickmeier (2011), also have power for martingale-type time variation. In Figure 6, we compare our LM statistic with them, which assume there is a break at $T_a$. The result shows that the performance of the tests for breaks is inferior to that of LM for the purpose of detecting continuous variation. One should note that our LM test is specifically designed to detect the temporal variation after the change point, while break tests can not tell what happens afterwards. As we said, martingale-type time variation is more harmful in the sense that the model can not be transformed into a stable one with constant factor loadings.

| | $a=1$ | $a=5$ | $a=10$ | $a=20$ | $a=30$ |
|---|---|---|---|---|---|
| $N=50$, $T=100$ | 0.448 | 0.688 | 0.778 | 0.831 | 0.850 |
| $N=50$, $T=150$ | 0.562 | 0.805 | 0.871 | 0.906 | 0.915 |
| $N=50$, $T=200$ | 0.620 | 0.858 | 0.914 | 0.939 | 0.942 |
| $N=100$, $T=100$ | 0.518 | 0.769 | 0.827 | 0.871 | 0.893 |
| $N=100$, $T=150$ | 0.640 | 0.860 | 0.915 | 0.943 | 0.959 |
| $N=100$, $T=200$ | 0.720 | 0.908 | 0.946 | 0.971 | 0.974 |
| $N=150$, $T=100$ | 0.544 | 0.781 | 0.842 | 0.880 | 0.903 |
| $N=150$, $T=150$ | 0.662 | 0.874 | 0.924 | 0.948 | 0.962 |
| $N=150$, $T=200$ | 0.742 | 0.917 | 0.953 | 0.975 | 0.980 |
| $N=200$, $T=100$ | 0.565 | 0.789 | 0.852 | 0.894 | 0.904 |
| $N=200$, $T=150$ | 0.682 | 0.869 | 0.927 | 0.953 | 0.962 |
| $N=200$, $T=200$ | 0.736 | 0.924 | 0.958 | 0.974 | 0.982 |

Table 15: Power (stationary factors).

| | $a=1$ | $a=5$ | $a=10$ | $a=20$ | $a=30$ |
|---|---|---|---|---|---|
| $N=50$, $T=100$ | 0.579 | 0.594 | 0.569 | 0.527 | 0.476 |
| $N=50$, $T=150$ | 0.680 | 0.653 | 0.633 | 0.586 | 0.552 |
| $N=50$, $T=200$ | 0.726 | 0.691 | 0.651 | 0.608 | 0.578 |
| $N=100$, $T=100$ | 0.626 | 0.646 | 0.627 | 0.554 | 0.526 |
| $N=100$, $T=150$ | 0.731 | 0.702 | 0.674 | 0.606 | 0.561 |
| $N=100$, $T=200$ | 0.778 | 0.734 | 0.699 | 0.641 | 0.593 |
| $N=150$, $T=100$ | 0.639 | 0.647 | 0.638 | 0.597 | 0.564 |
| $N=150$, $T=150$ | 0.745 | 0.729 | 0.699 | 0.647 | 0.593 |
| $N=150$, $T=200$ | 0.796 | 0.774 | 0.721 | 0.681 | 0.624 |
| $N=200$, $T=100$ | 0.652 | 0.671 | 0.659 | 0.621 | 0.580 |
| $N=200$, $T=150$ | 0.751 | 0.745 | 0.717 | 0.672 | 0.640 |
| $N=200$, $T=200$ | 0.804 | 0.779 | 0.740 | 0.707 | 0.654 |

Table 16: Power (integrated factors).

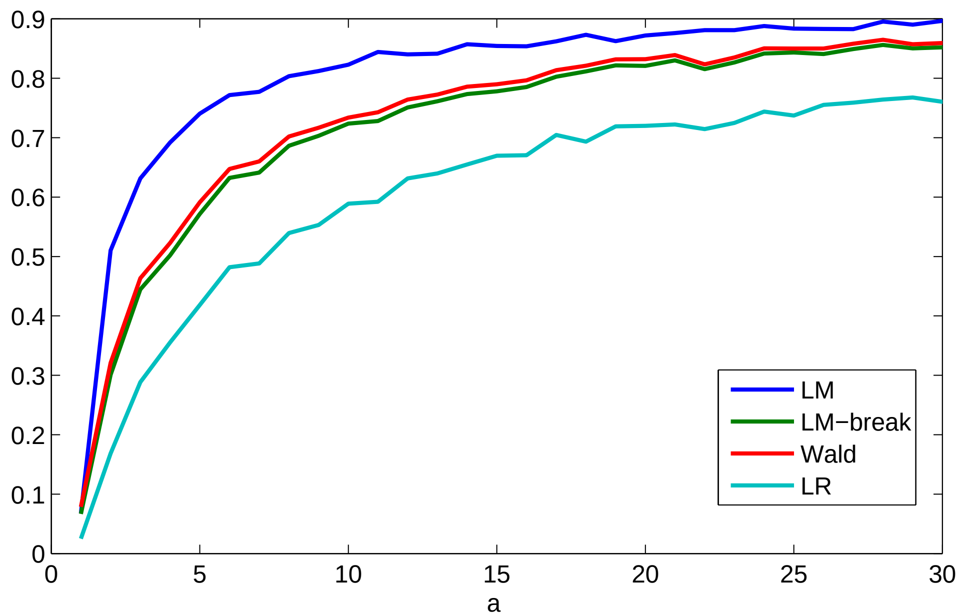


Figure 6: Power comparison for testing against martingale-type variation, $T = 100$, $N = 100$, $I(0)$ factors, block diagonal idiosyncratics. The horizontal axis gives the indicator of the strength of instability in factor loadings, while the vertical axis gives the power.

## 4.5 An Application

One application of factor models is nowcasting quarterly GDP growth using hundreds of predictors with different frequencies (Giannone *et al.* (2008) for Fed and Angelini *et al.* (2011) for ECB, for example). In practice, the performance is mixed and varies with the specific country to predict. It has good performance in the US and Euro Erea, but fails to work well in the UK during and after the recession (Xu, 2014). One can use various refinements, such as variable pre-selection prior to factor extraction and downweighting past information, but the improvement is limited after the recent financial crisis. It implies that structural instability might be a major concern. One alternative solution is the factor model with time-varying loadings and stochastic volatility proposed in Del Negro & Otrok (2008). Therefore, it would be helpful to test whether there has been temporal instability since the crisis.

The raw dataset contains 134 UK series including survey and financial indicators, measures of demand, output, labor and housing activities, prices, and 18 US and Germany variables from 1990 to 2014. The series are either monthly or daily, and need to be transformed into a quarterly and stationary quantity. The details on the composition and transformation method of the dataset are listed in the Appendix. We use only the series for which there are less the 1/3 of missing data and adjust for outliers and missing data to create a balanced dataset.[18] Although the tradition is to use as many series as possible, one reason behind some unsatisfactory performance is perhaps the adverse influence of uninformative predictors for the target variable to forecast. Boivin & Ng (2006) showed that as few as 40 pre-selected predictors can yield better forecasts than 147 predictors. Recently several

[18]Specifically, we replace the outliers and missing data with the median of a series.


Figure 7: Relative rejection rates from 2007:01 to 2009:12.

attempts have been made to improve the forecasts based on estimated factors, including pre-selection of the predictors (Bai & Ng, 2008a) and boosting (Bai & Ng, 2009), among others. We pre-select 50 variables using the elastic net, which is known as an effective tool to perform variable selection and shrinkage simultaneously (Bai & Ng, 2008a).

We test whether there was temporal instability from 2007:01 to 2009:12. Onatski (2010)'s criterion indicates that the number of factors is 13, which is a little larger than that extracted from the US dataset. Figure 7 reports the relative rejection rates, the shares of the 50 variables for which instability in factor loadings is found. The rejection rate can be as high as 55%, which implies the existence of instability during the financial crisis. It can be seen that the time variation was most serious around the middle of 2007 and gradually faded away afterwards.

## 4.6 Conclusions

In the chapter, we adapt Nyblom (1989)'s statistic for high-dimensional factor models to test for the stable factor structure against considerable time variation in factor loadings in the form of martingales, which might affect the consistency of the factor estimator and therefore the forecasting performance of factor-augmented regressions. Under plausible conditions, the estimation error of factors and related variables is negligible for the asymptotic distribution of the test statistic. Although the tests designed for the alternatives of multiple breaks also have power for time-varying parameters, our statistic is more powerful to detect martingale-type time variation. We apply the test statistic to a UK dataset and find that the time variation was most serious around the middle of 2007 but gradually faded away afterwards.

One might be interested in the joint null hypothesis that there is no time variation in the loadings of all variables. The pooled test statistic that combines $\widehat{\mathrm{LM}}_i$ is however not normally distributed, because the rate at which $\widehat{\mathrm{LM}}_i$ converges to $\widetilde{\mathrm{LM}}_\mathrm{i}$ is not sufficiently high. We leave this for future research.

## 4.7 Appendix

In the appendix, we will prove the theorems stated in the chapter. The proofs make extensive use of the Cauchy-Schwarz Inequality. For expositional simplicity, we only show the proof for the model with no cross-sectional dependence. Due to Assumption 4(g), it is straightforward to extend to the more general case.

### 4.7.1 Proof of Theorem 1

Following Bai (2003),

$$\hat{F}_t - H'F_t = \frac{1}{T}V_{NT}^{-1}(\sum_{s=1}^{T}\hat{F}_s\gamma_{st} + \sum_{s=1}^{T}\hat{F}_s\zeta_{st} + \sum_{s=1}^{T}\hat{F}_s\eta_{st} + \sum_{s=1}^{T}\hat{F}_s\xi_{st}), \qquad (49)$$

where $\gamma_{st} = \mathrm{E}(u_s'u_t)/N$, $\zeta_{st} = [u_s'u_t - \mathrm{E}(u_s'u_t)]/N$; $\eta_{st} = F_s'\Lambda_1'u_t/N$; $\xi_{st} = F_t'\Lambda_1'u_s/N$. For ease of notation, we set $T_a = 0$ but distinguish between $\tilde{T}$ and $T$ to generalize the results. $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. We have the following lemmas under the null hypothesis and assumptions for all $N$, $T$, and $\tilde{T}$.

**Lemma 1.** $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - H'F_t)F_t'||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$.

**Lemma 2.** $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - H'F_t)u_{it}||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, \tilde{T}T^2)})$.

**Lemma 3.** $\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t - H'F_t)F_t' = O_p(\frac{1}{\min(N, \sqrt{N\tilde{T}}, T)})$.

**Lemma 4.** $\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t - H'F_t)u_{it} = O_p(\frac{1}{\min(N, \sqrt{N\tilde{T}}, T)})$.

**Lemma 5.** $\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}||\hat{F}_t - H'F_t||^2 = O_p(\frac{1}{\min(N,T)})$.

**Lemma 6.** $\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t\hat{F}_t' - H'F_t(H'F_t)') = O_p(\frac{1}{\min(N, \sqrt{N\tilde{T}}, T)})$. Therefore, $\hat{\Sigma}_F^{-1} = \tilde{\Sigma}_F^{-1} + o_P(1)$.

**Proof of Lemma 1**

By (49),

$$\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - H'F_t)F_t'||^2 \leq 4||V_{NT}^{-1}||(\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}F_t'||^2}_{I} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}F_t'||^2}_{II}$$
$$+ \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}F_t'||^2}_{III} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\xi_{st}F_t'||^2}_{IV}).$$

First consider part $I$. By adding and subtracting terms,

$$\sum_{\tau=1}^{\tilde{T}} ||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}F_t'||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\gamma_{st}F_t'||^2 + \frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t'||^2).$$

For the first term,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\gamma_{st}F_t'||^2 &\leq \frac{1}{N^2T}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s - H'F_s||^2)\sum_{\tau=1}^{\tilde{T}}\sum_{s=1}^{T}||\sum_{t=1}^{\tau}F_t'\sum_{i=1}^{N}\mathrm{E}(u_{is}u_{it})||^2 \\
&\leq \frac{\tilde{T}}{T}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s - H'F_s||^2)(\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{it}^2))^2(\sum_{t=1}^{\tilde{T}}||F_t'||)^2 \\
&= O_p(\frac{\tilde{T}^3}{T}\delta_{NT}^{-2}),
\end{aligned}$$

by Theorem 1 in Bai & Ng (2002) and Assumptions 1 and 4(a).

Consider the second term,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t'||^2 &\leq \frac{1}{N^2T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}\sum_{i=1}^{N}F_sF_t'\mathrm{E}(u_{is}u_{it})||^2||H'||^2 \\
&\leq \frac{\tilde{T}}{T^2}||\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{i1}^2)||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2||H'||^2 \\
&= O_p(\frac{\tilde{T}^3}{T^2}),
\end{aligned}$$

since $||H|| = O_p(1)$ from (Bai, 2003). Therefore $I$ is $O_p(\frac{\tilde{T}^3}{T}\delta_{NT}^{-2})$.

Consider part $II$,

$$\sum_{\tau=1}^{\tilde{T}} ||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}F_t'||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\zeta_{st}F_t'||^2 + \frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\zeta_{st}F_t'||^2).$$

By Assumption 4(c),

$$
\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\zeta_{st}F_t'||^2 &\leq \frac{1}{NT}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)\sum_{\tau=1}^{\tilde{T}}\sum_{s=1}^{T}||\sum_{t=1}^{\tau}F_t'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_{is}u_{it}-\mathrm{E}(u_{is}u_{it}))||^2 \\
&\leq \frac{\tilde{T}}{NT}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)\sum_{t=1}^{\tilde{T}}||F_t||^2\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_{is}u_{it}-\mathrm{E}(u_{is}u_{it}))]^2 \\
&= O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2}).
\end{aligned}
$$

By Assumption 4(d),

$$
\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\zeta_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT}\sum_{t=1}^{\tilde{T}}||\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s(u_{is}u_{it}-\mathrm{E}(u_{is}u_{it}))||^2\sum_{t=1}^{\tilde{T}}||F_t||^2||H'||^2 \\
&= O_p(\frac{\tilde{T}^3}{NT}).
\end{aligned}
$$

Thus, $II$ is $O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2})$.

Consider part $III$.

$$
\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}F_t'||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\eta_{st}F_t'||^2+\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\eta_{st}F_t'||^2).
$$

For the first expression in the bracket,

$$
\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\eta_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)\sum_{s=1}^{T}||F_s||^2\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{N\tilde{T}}}\sum_{t=1}^{\tau}\Lambda_1'u_tF_t'||^2 \\
&= O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2}).
\end{aligned}
$$

For the second expression,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\eta_{st}F_t'||^2 &\leq \frac{\tilde{T}^2}{NT^2}(\sum_{s=1}^{T}||F_s||^2)^2\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{N\tilde{T}}}\sum_{t=1}^{\tau}\Lambda_1'u_tF_t'||^2||H'||^2 \\
&= O_p(\frac{\tilde{T}^2}{N}).
\end{aligned}$$

As a result, $III$ is $O_p(\frac{\tilde{T}^2}{N})$.

Finally consider part $IV$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\xi_{st}F_t'||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\xi_{st}F_t'||^2+\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\xi_{st}F_t'||^2).$$

For the first term,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\xi_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)\sum_{s=1}^{T}||\frac{1}{\sqrt{N}}\Lambda_1'u_s||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2 \\
&= \frac{\tilde{T}}{NT}O_p(\delta_{NT}^{-2})O_p(T)O_p(\tilde{T}^2) \\
&= O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2}).
\end{aligned}$$

Consider the second term,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\xi_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT}||\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}F_su_s'\Lambda_1||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2||H'||^2 \\
&= O_p(\frac{\tilde{T}^3}{NT}).
\end{aligned}$$

Thus, $IV$ is $O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2})$. As $||V_{NT}|| = O_p(1)$ from Bai (2003), the proof is finished by combining these results.

**Proof of Lemma 2**

$$\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t-H'F_t)u_{it}||^2 \leq 4||V_{NT}^{-1}||(\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}u_{it}||^2}_{I}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}u_{it}||^2}_{II}$$
$$+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}u_{it}||^2}_{III}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\xi_{st}u_{it}||^2}_{IV}).$$

We begin with part $I$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}u_{it}||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\gamma_{st}u_{it}||^2+\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\gamma_{st}u_{it}||^2).$$

The first expression on the right hand side is bounded by

$$\frac{\tilde{T}^2}{T}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)[\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}(\frac{1}{\sqrt{\tilde{T}}}\sum_{t=1}^{\tau}u_{it}\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}(u_{jt}^2))^2],$$

which is $O_p(\frac{\tilde{T}^2}{T}\delta_{NT}^{-2})$. The second term is bounded by

$$\frac{\tilde{T}^2}{T^2}\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{\tilde{T}}}\sum_{t=1}^{\tau}F_t u_{it}\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}(u_{jt}^2)||^2||H'||^2 = O_p(\frac{\tilde{T}^2}{T^2}).$$

Thus, $I$ is $O_p(\frac{\tilde{T}^2}{T}\delta_{NT}^{-2})$.

Part $II$ is similar to the one in Lemma 1, which is $O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2})$.

Consider part $III$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}u_{it}||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\eta_{st}u_{it}||^2+\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\eta_{st}u_{it}||^2).$$

The first term is bounded by $\frac{\tilde{T}}{NT}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s - H'F_s||^2)\sum_{s=1}^{T}||F_s||^2\sum_{t=1}^{\tilde{T}}u_{it}^2\sum_{t=1}^{\tilde{T}}||\frac{1}{\sqrt{N}}\Lambda_1'u_t||^2$, which is $O_p(\frac{\tilde{T}^3}{N}\delta_{NT}^{-2})$. For the second term,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\eta_{st}u_{it}||^2 &\leq \frac{1}{NT^2}(\sum_{s=1}^{T}||F_s||^2)^2\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{N}}\sum_{t=1}^{\tau}\Lambda_1'u_tu_{it}||^2||H'||^2 \\
&\leq \frac{2\tilde{T}^2}{NT^2}(\sum_{s=1}^{T}||F_s||^2)^2\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{N\tilde{T}}}\sum_{t=1}^{\tau}\sum_{j=1}^{N}\lambda_{j1}(u_{jt}u_{it}-\mathrm{E}(u_{jt}u_{it}))||^2||H'||^2 \\
&\quad + \frac{2}{N^2T^2}(\sum_{s=1}^{T}||F_s||^2)^2\sum_{\tau=1}^{\tilde{T}}(\sum_{t=1}^{\tau}\sum_{j=1}^{N}||\lambda_{j1}||||\mathrm{E}(u_{jt}u_{it})||)^2||H'||^2 \\
&= O_p(\frac{\tilde{T}^2}{N}) + O_p(\frac{\tilde{T}^3}{N^2}),
\end{aligned}$$

by Assumption 2, 4(b) and 4(h). $III$ is thus $O_p(\frac{\tilde{T}^3}{N\min(N,\tilde{T})})$.

Consider part $IV$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\xi_{st}u_{it}||^2 \leq 2(\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\xi_{st}u_{it}||^2 + \frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\xi_{st}u_{it}||^2).$$

For the first expression in the bracket,

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\xi_{st}u_{it}||^2 &\leq \frac{\tilde{T}^2}{N}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s - H'F_s||^2)(\frac{1}{T}\sum_{s=1}^{T}||\frac{1}{\sqrt{N}}\Lambda_1'u_s||^2)\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{\tilde{T}}}\sum_{t=1}^{\tau}F_tu_{it}||^2 \\
&= O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2}).
\end{aligned}$$

The second term is bounded by $\frac{\tilde{T}^2}{NT}||\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\Lambda_1'u_sF_s'||^2\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{\sqrt{\tilde{T}}}\sum_{t=1}^{\tau}F_tu_{it}||^2||H'||^2$, which is $O_p(\frac{\tilde{T}^2}{NT})$. Thus, $IV$ is $O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2})$.

As a result, $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - H'F_t)u_{it}||^2 = 4||V_{NT}^{-1}||(I+II+III+IV) = O_p(\frac{\tilde{T}^3}{\min(N^2,N\tilde{T},\tilde{T}T^2)})$.

**Proof of Lemma 3**

The proof of Lemma 3 is straightforwardly adapted from that of Lemma B.2 of Bai

(2003). The difference between Lemma 3 and Lemma B.2 is that Lemma 3 is a mean over $\tilde{T}$ periods while Lemma B.2 is a mean over $T$ periods. The factors are estimated using the data of $T$ periods in both cases, however. As a result, they can be of different orders in probability. Specifically,

$$\begin{aligned}
\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t - H'F_t)F_t' &= \underbrace{\frac{V_{NT}^{-1}}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}F_t'}_{I} + \underbrace{\frac{V_{NT}^{-1}}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}F_t'}_{II} \\
&\quad + \underbrace{\frac{V_{NT}^{-1}}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\eta_{st}F_t'}_{III} + \underbrace{\frac{V_{NT}^{-1}}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\xi_{st}F_t'}_{IV}.
\end{aligned}$$

First consider $I$.

$$\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}F_t' = \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\gamma_{st}F_t' + \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t',$$

where

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\gamma_{st}F_t'|| &\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||\sum_{t=1}^{\tilde{T}}\frac{1}{N}\mathrm{E}(u_s'u_t)F_t'||^2)^{\frac{1}{2}} \\
&\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}\sum_{t=1}^{\tilde{T}}||\frac{1}{N}\mathrm{E}(u_s'u_t)||^2)^{\frac{1}{2}}(\sum_{t=1}^{\tilde{T}}||F_t'||^2)^{\frac{1}{2}} \\
&= O_p(\frac{1}{\sqrt{T}}\frac{1}{\min(\sqrt{N},\sqrt{T})}),
\end{aligned}$$

and

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t'|| &= ||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}H'F_tF_t'\gamma_{tt}|| \\
&\leq \frac{||H'||}{\tilde{T}T}(\sum_{t=1}^{\tilde{T}}||F_tF_t'||^2)^{\frac{1}{2}}(\sum_{t=1}^{\tilde{T}}||\gamma_{tt}||^2)^{\frac{1}{2}} \\
&= O_p(\frac{1}{T}).
\end{aligned}$$

Thus, $I$ is $O_p(\frac{1}{\sqrt{T}}\frac{1}{\min(\sqrt{N},\sqrt{T})})$.

Consider $II$.

$$\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}F_t' = \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\zeta_{st}F_t' + \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\zeta_{st}F_t',$$

where

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\zeta_{st}F_t'|| &\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s-H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||\sum_{t=1}^{\tilde{T}}\frac{1}{N}[u_s'u_t-\mathrm{E}(u_s'u_t)]F_t'||^2)^{\frac{1}{2}} \\
&\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s-H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}\sum_{t=1}^{\tilde{T}}||\frac{1}{N}[u_s'u_t-\mathrm{E}(u_s'u_t)]||^2)^{\frac{1}{2}}(\sum_{t=1}^{\tilde{T}}||F_t'||^2)^{\frac{1}{2}} \\
&= O_p(\frac{1}{\sqrt{N}}\frac{1}{\min(\sqrt{N},\sqrt{T})}),
\end{aligned}$$

by Assumption 4(c), and

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\zeta_{st}F_t'|| &\leq \frac{||H'||}{\tilde{T}T}(\sum_{t=1}^{\tilde{T}}||\sum_{s=1}^{T}F_s\frac{1}{N}[u_s'u_t-\mathrm{E}(u_s'u_t)]||^2)^{\frac{1}{2}}(\sum_{t=1}^{\tilde{T}}||F_t'||^2)^{\frac{1}{2}} \\
&= O_p(\frac{1}{\sqrt{NT}}),
\end{aligned}$$

by Assumption 4(d). As a result, $II$ is $O_p(\frac{1}{\sqrt{N}}\frac{1}{\min(\sqrt{N},\sqrt{T})})$.

Then consider $III$.

$$\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\eta_{st}F_t' \quad = \quad \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\eta_{st}F_t' + \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\eta_{st}F_t'.$$

The first term is

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\eta_{st}F_t'|| &\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||\sum_{t=1}^{\tilde{T}}\frac{1}{N}F_s'\Lambda_1'u_tF_t'||^2)^{\frac{1}{2}} \\
&\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(||\sum_{t=1}^{\tilde{T}}\frac{1}{N}\Lambda_1'u_tF_t']||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||F_s'||^2)^{\frac{1}{2}} \\
&= O_p(\frac{1}{\sqrt{N\tilde{T}}}\frac{1}{\min(\sqrt{N},\sqrt{T})}).
\end{aligned}$$

The second term is

$$\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\eta_{st}F_t'|| &= ||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\frac{1}{N}F_s'\Lambda_1'u_tF_t'|| \\
&= \frac{||H'||}{N\tilde{T}T}||\sum_{s=1}^{T}F_sF_s'||||\sum_{t=1}^{\tilde{T}}\Lambda_1'u_tF_t'|| \\
&= O_p(\frac{1}{\sqrt{N\tilde{T}}}).
\end{aligned}$$

$III$ is therefore $O_p(\frac{1}{\sqrt{N\tilde{T}}})$.

Finally consider $IV$.

$$\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}\hat{F}_s\xi_{st}F_t' \quad = \quad \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\xi_{st}F_t' + \frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\xi_{st}F_t'.$$

The first term is

$$
\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)\xi_{st}F_t'|| &\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||\sum_{t=1}^{\tilde{T}}\frac{1}{N}F_t'\Lambda_1'u_sF_t'||^2)^{\frac{1}{2}} \\
&\leq \frac{1}{\tilde{T}T}(\sum_{s=1}^{T}||(\hat{F}_s - H'F_s)||^2)^{\frac{1}{2}}(\sum_{s=1}^{T}||\frac{1}{N}\Lambda_1'u_s||^2)^{\frac{1}{2}}||\sum_{t=1}^{\tilde{T}}F_tF_t'|| \\
&= O_p(\frac{1}{\sqrt{N}}\frac{1}{\min(\sqrt{N},\sqrt{T})}).
\end{aligned}
$$

The second term is

$$
\begin{aligned}
||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\xi_{st}F_t'|| &= ||\frac{1}{\tilde{T}T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}H'F_s\frac{1}{N}F_t'\Lambda_1'u_sF_t'|| \\
&= \frac{||H'||}{N\tilde{T}T}||\sum_{s=1}^{T}F_su_s'\Lambda_1||||\sum_{t=1}^{\tilde{T}}F_tF_t'|| \\
&= O_p(\frac{1}{\sqrt{NT}}).
\end{aligned}
$$

Thus $IV$ is $O_p(\frac{1}{\sqrt{N}}\frac{1}{\min(\sqrt{N},\sqrt{T})})$. As a result, $I+II+III+IV = O_p(\frac{1}{\sqrt{T}}\frac{1}{\min(\sqrt{N},\sqrt{T})}) + O_p(\frac{1}{\sqrt{N}}\frac{1}{\min(\sqrt{N},\sqrt{T})}) + O_p(\frac{1}{\sqrt{N\tilde{T}}}) = O_p(\frac{1}{\min(N,\sqrt{N\tilde{T}},T)})$.

**Proofs of Lemmas 4 and 5**

The proofs of Lemmas 4 and 5 are straightforwardly adapted from those of Lemma B.1 of Bai (2003) and Theorem 1 of Bai & Ng (2002), respectively. The difference is still that the average is taken over $\tilde{T}$ periods in my case, but is taken over $T$ periods in Bai & Ng (2002) and Bai (2003). The adaptation is therefore very similar to that in the proof of Lemma 3, so we omit it here.

**Proof of Lemma 6**

$$\begin{aligned}
\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t\hat{F}_t' - H'F_t(H'F_t)') &= \frac{1}{\tilde{T}}\sum_{t=T_a+1}^{T}(\hat{F}_t - H'F_t)(\hat{F}_t' - (H'F_t)') + \frac{1}{\tilde{T}}\sum_{t=T_a+1}^{T}(\hat{F}_t - H'F_t)(H'F_t)' \\
&\quad + \frac{1}{\tilde{T}}\sum_{t=T_a+1}^{T}H'F_t(\hat{F}_t' - (H'F_t)') \\
&= O_p(\frac{1}{\min(N, \sqrt{N\tilde{T}}, T)}),
\end{aligned}$$

by Lemmas 3 and 5.

**Proof of Theorem 1**

Let $\hat{d}_{i\tau} = \hat{\Sigma}_F^{-\frac{1}{2}}\hat{s}_{i\tau}$ where $\hat{\Sigma}_F^{-1} = (\hat{\Sigma}_F^{-\frac{1}{2}})'\hat{\Sigma}_F^{-\frac{1}{2}}$ by Cholesky decomposition. Similarly, $\tilde{d}_{i\tau} = \tilde{\Sigma}_F^{-\frac{1}{2}}\tilde{s}_{i\tau}$. As $\widehat{\mathrm{LM}}_i - \widetilde{\mathrm{LM}}_i = \frac{1}{\tilde{T}^2\hat{V}_{ii}^{-1}}\sum_{\tau=1}^{\tilde{T}}\hat{d}_{i\tau}'\hat{d}_{i\tau} - \frac{1}{\tilde{T}^2 V_{ii}^{-1}}\sum_{\tau=1}^{\tilde{T}}\tilde{d}_{i\tau}'\tilde{d}_{i\tau}$ and $\hat{V}_{ii}^{-1} \xrightarrow{p} V_{ii}^{-1}$ by Assumption 4(giii), we need to study the order of $\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}\hat{d}_{i\tau}'\hat{d}_{i\tau} - \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}\tilde{d}_{i\tau}'\tilde{d}_{i\tau}$. We have

$$\begin{aligned}
\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}\hat{d}_{i\tau}'\hat{d}_{i\tau} - \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}\tilde{d}_{i\tau}'\tilde{d}_{i\tau} &= \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}(\hat{d}_{i\tau} - \tilde{d}_{i\tau})'(\hat{d}_{i\tau} + \tilde{d}_{i\tau}) \\
&= \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}(\hat{d}_{i\tau} - \tilde{d}_{i\tau})'(\hat{d}_{i\tau} - \tilde{d}_{i\tau}) + \frac{2}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}(\hat{d}_{i\tau} - \tilde{d}_{i\tau})'\tilde{d}_{i\tau} \qquad (50) \\
&\leq \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\hat{d}_{i\tau} - \tilde{d}_{i\tau}||^2 + 2(\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\hat{d}_{i\tau} - \tilde{d}_{i\tau}||^2)^{\frac{1}{2}}\underbrace{(\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}\tilde{d}_{i\tau}'\tilde{d}_{i\tau})^{\frac{1}{2}}}_{=O_P(1)},
\end{aligned}$$

where

$$\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\hat{d}_{i\tau}-\tilde{d}_{i\tau}||^2 = \frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\hat{\Sigma}_F^{-\frac{1}{2}}\hat{s}_{i\tau}-\tilde{\Sigma}_F^{-\frac{1}{2}}\tilde{s}_{i\tau}||^2 \tag{51}$$

$$\leq 2||\hat{\Sigma}_F^{-\frac{1}{2}}||^2\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||(\hat{s}_{i\tau}-\tilde{s}_{i\tau})||^2+\underbrace{2||\hat{\Sigma}_F^{-\frac{1}{2}}-\tilde{\Sigma}_F^{-\frac{1}{2}}||^2\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\tilde{s}_{i\tau}||^2}_{=o_p(1)O_p(1)=o_p(1)}.$$

Thus, the key is the order of $\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||(\hat{s}_{i\tau}-\tilde{s}_{i\tau})||^2$. For expositional simplicity, we only show the remaining proof for the models with no cross-sectional dependence. Due to Assumption 4(g), it is straightforward to extend to the more general case. In the models with cross-sectional independence, $\hat{V}_{i.}^{-1}X_t=\hat{V}_{ii}^{-1}X_{it}$ and $V_{i.}^{-1}X_t=V_{ii}^{-1}X_{it}$. Therefore,

$$\begin{aligned}\hat{s}_{i\tau}-\tilde{s}_{i\tau} &= (\sum_{t=1}^{\tau}\hat{F}_t\hat{V}_{ii}^{-1}X_{it}-\sum_{t=1}^{\tau}\hat{F}_t\hat{F}_t'(\sum_{m=1}^{\tilde{T}}\hat{F}_m\hat{F}_m')^{-1}\sum_{m=1}^{\tilde{T}}\hat{F}_m\hat{V}_{ii}^{-1}X_{im})-(\sum_{t=1}^{\tau}\tilde{F}_tV_{ii}^{-1}X_{it}-\sum_{t=1}^{\tau}\tilde{F}_t\tilde{F}_t'(\sum_{m=1}^{\tilde{T}}\tilde{F}_m\tilde{F}_m')^{-1}\sum_{m=1}^{\tilde{T}}\tilde{F}_mV_{ii}^{-1}X_{im})\\ &= \underbrace{\sum_{t=1}^{\tau}(\hat{F}_t\hat{V}_{ii}^{-1}-\tilde{F}_tV_{ii}^{-1})X_{it}}_{I}-\underbrace{\sum_{t=1}^{\tau}\hat{F}_t\hat{F}_t'[(\sum_{m=1}^{\tilde{T}}\hat{F}_m\hat{F}_m')^{-1}\sum_{m=1}^{\tilde{T}}\hat{F}_m\hat{V}_{ii}^{-1}X_{im}-(\sum_{m=1}^{\tilde{T}}\tilde{F}_m\tilde{F}_m')^{-1}\sum_{m=1}^{\tilde{T}}\tilde{F}_mV_{ii}^{-1}X_{im}]}_{II}\\ &\quad \underbrace{-\sum_{t=1}^{\tau}(\hat{F}_t\hat{F}_t'-\tilde{F}_t\tilde{F}_t')(\sum_{m=1}^{\tilde{T}}\tilde{F}_m\tilde{F}_m')^{-1}\sum_{m=1}^{\tilde{T}}\tilde{F}_mV_{ii}^{-1}X_{im}}_{III}.\end{aligned}$$

Therefore,

$$\frac{1}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||(\hat{s}_{i\tau}-\tilde{s}_{i\tau})||^2\leq\frac{3}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||I||^2+\frac{3}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||II||^2+\frac{3}{\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||III||^2. \tag{52}$$

First consider$\sum_{\tau=1}^{\tilde{T}}||I||^2$. Since $\hat{V}_{ii}^{-1}\xrightarrow{p}V_{ii}^{-1}$, we only need to compute the order of

$\tilde{I} = \sum_{\tau=1}^{\tilde{T}} \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) X_{it}$. Note that

$$\sum_{\tau=1}^{\tilde{T}} ||\tilde{I}||^2 \leq 2 \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) F_t' \lambda_{i1} ||^2 + 2 \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) u_{it} ||^2. \tag{53}$$

By Lemmas 1 and 2, $\sum_{\tau=1}^{\tilde{T}} ||I||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, \tilde{T}T^2)}) + O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)}) = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$.

Consider $\sum_{\tau=1}^{\tilde{T}} ||II||^2$. Because of $\hat{V}_{ii}^{-1} \xrightarrow{p} V_{ii}^{-1}$, the key point is the order of $(\sum_{m=1}^{\tilde{T}} \hat{F}_m \hat{F}_m')^{-1} \sum_{m=1}^{\tilde{T}} \hat{F}_m X_{im} - (\sum_{m=1}^{\tilde{T}} \tilde{F}_m \tilde{F}_m')^{-1} \sum_{m=1}^{\tilde{T}} \tilde{F}_m X_{im}$. By Lemmas 3, 4 and 6,

$$(\sum_{m=1}^{\tilde{T}} \hat{F}_m \hat{F}_m')^{-1} \sum_{m=1}^{\tilde{T}} \hat{F}_m X_{im} - (\sum_{m=1}^{\tilde{T}} \tilde{F}_m \tilde{F}_m')^{-1} \sum_{m=1}^{\tilde{T}} \tilde{F}_m X_{im} = O_p(\frac{1}{\min(N, \sqrt{N\tilde{T}}, T)}).$$

Consequently, $\sum_{\tau=1}^{\tilde{T}} |II||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$.

Finally consider $\sum_{\tau=1}^{\tilde{T}} ||III||^2$. As

$$\begin{aligned}
\sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t \hat{F}_t' - \tilde{F}_t \tilde{F}_t') ||^2 \leq & \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t)(\hat{F}_t' - \tilde{F}_t') + \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) \tilde{F}_t' + \sum_{t=1}^{\tau} \tilde{F}_t (\hat{F}_t' - \tilde{F}_t') ||^2 \\
& \leq 3 \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t)(\hat{F}_t' - \tilde{F}_t') ||^2 + 6 \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) \tilde{F}_t' ||^2 \\
& \leq 3\tilde{T} (\sum_{t=1}^{\tilde{T}} ||\hat{F}_t - \tilde{F}_t||^2)^2 + 6 \sum_{\tau=1}^{\tilde{T}} || \sum_{t=1}^{\tau} (\hat{F}_t - \tilde{F}_t) \tilde{F}_t' ||^2.
\end{aligned}$$

By Lemma 1 and 5, $\sum_{\tau=1}^{\tilde{T}} ||III||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$. As a result, $\frac{1}{\tilde{T}^2} \sum_{\tau=1}^{\tilde{T}} ||(\hat{s}_{i\tau} - \tilde{s}_{i\tau})||^2 = O_p(\frac{\tilde{T}}{\min(N^2, N\tilde{T}, T^2)})$ which leads to $\widehat{\mathrm{LM}}_i - \widetilde{\mathrm{LM}}_i = O_p(\frac{\sqrt{\tilde{T}}}{\sqrt{\min(N^2, N\tilde{T}, T^2)}})$.

### 4.7.2 Proof of Theorem 2

Define $\tilde{e}_{it} = \tilde{T} e_{it}$ such that $\mathrm{E}(\tilde{e}_{it}\tilde{e}_{it}') = a\Sigma_e$. Under the local alternative,

$$\begin{aligned}\hat{F}_t - H'F_t &= \frac{1}{T}V_{NT}^{-1}(\sum_{s=1}^{T}\hat{F}_s\gamma_{st} + \sum_{s=1}^{T}\hat{F}_s\zeta_{st} + \sum_{s=1}^{T}\hat{F}_s\eta_{st} + \sum_{s=1}^{T}\hat{F}_s\xi_{st} \\ &\quad + \frac{1}{N}(\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda_1'v_t + \sum_{s=1}^{T}\hat{F}_sv_s'\Lambda_1F_t + \sum_{s=1}^{T}\hat{F}_sv_s'v_t + \sum_{s=1}^{T}\hat{F}_su_s'v_t + \sum_{s=1}^{T}\hat{F}_sv_s'u_t)).\end{aligned} \tag{54}$$

**Lemma 7.** $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - \tilde{F}_t)v_{it}||^2 = O_p(\frac{\tilde{T}^2}{min(N,T^2)})$.

**Lemma 8.** $\frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}}(\hat{F}_t - H'F_t)v_{it} = O_p(\frac{1}{\sqrt{\tilde{T}}min(N,\sqrt{N\tilde{T}},T)})$.

**Proof of Lemma 7**

By (54),

$$\begin{aligned}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_t - H'F_t)v_{it}||^2 \leq\ & 9||V_{NT}^{-1}||(\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}v_{it}||^2}_{I} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}v_{it}||^2}_{II} \\ & + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}v_{it}||^2}_{III} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\xi_{st}v_{it}||^2}_{IV} \\ & + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda_1'v_tv_{it}||^2}_{V} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_sv_s'\Lambda_1F_tv_{it}||^2}_{VI} \\ & + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_sv_s'v_tv_{it}||^2}_{VII} + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_su_s'v_tv_{it}||^2}_{VIII} \\ & + \underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_sv_s'u_tv_{it}||^2}_{IX}).\end{aligned}$$

We begin with part $I$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\gamma_{st}v_{it}||^2 \leq 2(\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\gamma_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 + \frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2).$$

Consider the first term,

$$\begin{aligned}\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\gamma_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 &\leq \frac{1}{T\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2) \\ &\quad \frac{1}{N^2}\sum_{s=1}^{T}||\sum_{t=1}^{\tau}F_t'\sum_{m=1}^{t}\tilde{e}_{im}\sum_{i=1}^{N}\mathrm{E}(u_{is}u_{it})||^2 \\ &\leq \frac{1}{T\tilde{T}}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)(\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{it}^2))^2(\sum_{t=1}^{\tilde{T}}||F_t'\sum_{m=1}^{t}\tilde{e}_{im}||)^2 \\ &= O_p(\frac{\tilde{T}^2}{T}\delta_{NT}^{-2}),\end{aligned}$$

as $(\sum_{t=1}^{\tilde{T}}||F_t'\sum_{m=1}^{t}\tilde{e}_{im}||)^2 \leq \sum_{t=1}^{\tilde{T}}||F_t'||^2\sum_{t=1}^{\tilde{T}}||\sum_{m=1}^{t}\tilde{e}_{im}||^2 = O_p(\tilde{T}^3)$.

For the second term,

$$\begin{aligned}\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\gamma_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 &\leq \frac{1}{T^2\tilde{T}^2}||\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{it}^2)||^2\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}F_tF_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2||H'||^2 \\ &= O_p(\frac{\tilde{T}^2}{T^2}),\end{aligned}$$

because $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}F_tF_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 \leq \tilde{T}\sum_{t=1}^{\tilde{T}}||F_t||^4\sum_{t=1}^{\tilde{T}}||\sum_{m=1}^{t}\tilde{e}_{im}||^2 = O_p(\tilde{T}^4)$.

It follows that $I = O_p(\frac{\tilde{T}^2}{T}\delta_{NT}^{-2})$.

Part $II$ is bounded by $O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2})$ proved in a similar manner. Consider part $III$.

$$\begin{aligned}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_s\eta_{st}v_{it}||^2 &\leq 2(\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\eta_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 \\ &\quad +\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_s\eta_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2).\end{aligned}$$

As $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\frac{1}{\sqrt{N}}\Lambda_1'u_tF_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 \leq \tilde{T}\sum_{t=1}^{\tilde{T}}||\frac{1}{\sqrt{N}}\Lambda_1'u_tF_t'||^2\sum_{t=1}^{\tilde{T}}||\sum_{m=1}^{t}\tilde{e}_{im}||^2 = O_p(\tilde{T}^4)$, the first term is

$$\begin{aligned}\frac{1}{T^2\tilde{T}^2}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)\eta_{st}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 &\leq \frac{1}{NT\tilde{T}^2}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2) \\ &\quad \sum_{s=1}^{T}||F_s||^2\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\frac{1}{\sqrt{N}}\Lambda_1'u_tF_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 \\ &= O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2}).\end{aligned}$$

Similarly, the second term is $O_p(\frac{\tilde{T}^2}{N})$. As a result, $III = O_p(\frac{\tilde{T}^2}{N})$. Part $IV$ is similarly $O_p(\frac{\tilde{T}^2}{N}\delta_{NT}^{-2})$.

Now study part $V$.

$$\begin{aligned}\sum_{\tau=1}^{\tilde{T}}||\frac{1}{NT}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda_1'v_tv_{it}||^2 &\leq 2(\frac{1}{N^2T^2\tilde{T}^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)F_s'\Lambda_{1i}F_t'\sum_{m=1}^{t}\tilde{e}_{im}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 \\ &\quad +\frac{1}{N^2T^2\tilde{T}^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_sF_s'\Lambda_{1i}F_t'\sum_{m=1}^{t}\tilde{e}_{im}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2).\end{aligned}$$

The first term is bounded by

$$\frac{1}{N^2T\tilde{T}^4}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_s-H'F_s||^2)\sum_{s=1}^{T}(F_s'\Lambda_{1i})^2\sum_{\tau=1}^{\tilde{T}}(\sum_{t=1}^{\tau}(F_t'\sum_{m=1}^{t}\tilde{e}_{im})^2)^2=O_p(\frac{\tilde{T}}{N^2}\delta_{NT}^{-2}),$$

because

$$\begin{aligned}\sum_{\tau=1}^{\tilde{T}}(\sum_{t=1}^{\tau}(F_t'\sum_{m=1}^{t}\tilde{e}_{im})^2)^2 &\leq \tilde{T}(\sum_{t=1}^{\tilde{T}}||F_t'||^2||\sum_{m=1}^{t}\tilde{e}_{im}||^2)^2\\ &\leq \tilde{T}\sum_{t=1}^{\tilde{T}}||F_t'||^4\sum_{t=1}^{\tilde{T}}||\sum_{m=1}^{t}\tilde{e}_{im}||^4\\ &= O_p(\tilde{T}^5).\end{aligned}$$

For the second term,

$$\begin{aligned}\frac{1}{N^2T^2\tilde{T}^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H'F_sF_s'\Lambda_{1i}F_t'\sum_{m=1}^{t}\tilde{e}_{im}F_t'\sum_{m=1}^{t}\tilde{e}_{im}||^2 &\leq \frac{1}{N^2T^2\tilde{T}^4}||\sum_{s=1}^{T}F_sF_s'\Lambda_{1i}||^2\sum_{\tau=1}^{\tilde{T}}(\sum_{t=1}^{\tau}(F_t'\sum_{m=1}^{t}\tilde{e}_{im})^2)^2||H'||^2\\ &= O_p(\frac{\tilde{T}}{N^2}).\end{aligned}$$

Thus, $V=O_p(\frac{\tilde{T}}{N^2})=o_p(\frac{\tilde{T}^2}{min(N,T^2)})$. Part $VI-IX$ are all $o_p(\frac{\tilde{T}^2}{min(N,T^2)})$ in a similar manner. The lemma is finally proved by combining the results of these parts.

**Proof of Lemma 8**

The proof of Lemma 8 is similar to that of Lemma 7 and is omitted here.

**Proof of Theorem 2**

In a similar manner, it can be shown that Lemmas 1-6 still hold under the local alternative $h_{NT}=a/\tilde{T}^2$ with $a$ fixed (The proofs are available on request). The proof is similar to that of Theorem 1. (52) still works. Now (53) changes to

$$\sum_{\tau=1}^{\tilde{T}} ||I||^2 \leq 3\sum_{\tau=1}^{\tilde{T}} ||\sum_{t=1}^{\tau}(\hat{F}_t - \tilde{F}_t)F_t'\lambda_{i1}||^2 + 3\sum_{\tau=1}^{\tilde{T}} ||\sum_{t=1}^{\tau}(\hat{F}_t - \tilde{F}_t)v_{it}||^2 + 3\sum_{\tau=1}^{\tilde{T}} ||\sum_{t=1}^{\tau}(\hat{F}_t - \tilde{F}_t)u_{it}||^2.$$

By Lemma 7, $\sum_{\tau=1}^{\tilde{T}} ||I||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$ is same as before. By Lemma 8, $\sum_{\tau=1}^{\tilde{T}} |II||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, T^2)})$. The order of $III$ is also unchanged. It follows that Theorem 2 is proved.

### 4.7.3 Proof of Theorem 3

The proof is similar to that for Theorem 1, so we only give the key points. Let $\delta_{I,NT} = \min(\sqrt{N}, T)$. Following Bai (2004),

$$\hat{F}_{I,t} - H_I'F_t = V_{NT}^{-1}(\frac{1}{T^2}\sum_{s=1}^{T}\hat{F}_{I,s}\zeta_{st} + \frac{1}{T^2}\sum_{s=1}^{T}\hat{F}_{I,s}\eta_{st} + \frac{1}{T^2}\sum_{s=1}^{T}\hat{F}_{I,s}\xi_{st} + \frac{1}{T^2}\sum_{s=1}^{T}\hat{F}_{I,s}\gamma_{st}) \quad (55)$$

**Lemma 9.** $\sum_{\tau=1}^{\tilde{T}} ||\sum_{t=1}^{\tau}(\hat{F}_{I,t} - H_I'F_t)F_t'||^2 = O_p(\frac{\tilde{T}T^2}{\min(N, T^2)})$.
**Lemma 10.** $\sum_{\tau=1}^{\tilde{T}} ||\sum_{t=1}^{\tau}(\hat{F}_{I,t} - H_I'F_t)u_{it}||^2 = O_p(\frac{\tilde{T}^3}{\min(N^2, N\tilde{T}, \tilde{T}^2T^2)})$ for each $i$.
**Lemma 11.** $\frac{1}{T^2}\sum_{t=1}^{\tilde{T}}(\hat{F}_{I,t} - H_I'F_t)F_t' = O_p(\frac{1}{T\min(\sqrt{N}, T)})$.
**Lemma 12.** $\frac{1}{T^2}\sum_{t=1}^{\tilde{T}}(\hat{F}_{I,t} - H_I'F_t)u_{it} = O_p(\frac{1}{T\min(\sqrt{N}T/\sqrt{\tilde{T}}, T^{5/2}/\tilde{T})})$.
**Lemma 13.** $\frac{1}{T^2}\sum_{t=1}^{\tilde{T}} ||\hat{F}_{I,t} - H_I'F_t||^2 = O_p(\frac{\tilde{T}}{T^2\min(N\tilde{T}/T, T^2)})$.
**Lemma 14.** $\frac{1}{T^2}\sum_{t=1}^{\tilde{T}}(\hat{F}_{I,t}\hat{F}_{I,t}' - H_I'F_t(H_I'F_t)') = O_p(\frac{1}{T\min(\sqrt{N}, T)})$. Therefore, $\hat{\Sigma}_F^{-1} = \tilde{\Sigma}_F^{-1} + o_P(1)$.

**Proof of Lemma 9**

By (55),

$$\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_{I,t}-H_I'F_t)F_t'||^2 \leq 4||V_{NT}^{-1}||(\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\gamma_{st}F_t'||^2}_{I}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\zeta_{st}F_t'||^2}_{II}$$
$$+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\eta_{st}F_t'||^2}_{III}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\xi_{st}F_t'||^2}_{IV}).$$

First consider part $I$. By adding and subtracting terms,

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\gamma_{st}F_t'||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\gamma_{st}F_t'||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\gamma_{st}F_t'||^2).$$

For the first term,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\gamma_{st}F_t'||^2 &\leq \frac{\tilde{T}}{T^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)(\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{it}^2))^2(\sum_{t=1}^{\tilde{T}}||F_t'||)^2\\ &\leq O_p(\frac{\tilde{T}^2}{T}\delta_{I,NT}^{-2}),\end{aligned}$$

by Lemma B.1 in Bai (2004) and $(\sum_{t=1}^{\tilde{T}}||F_t'||)^2 \leq \tilde{T}\sum_{t=1}^{\tilde{T}}||F_t'||^2 = O_p(\tilde{T}T^2)$.

Consider the second term,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\gamma_{st}F_t'||^2 &\leq \frac{\tilde{T}}{T^4}||\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(u_{i1}^2)||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2||H_I'||^2\\ &\leq O_p(\tilde{T}),\end{aligned}$$

since $||H_I|| = O_p(1)$ from Bai (2004). Therefore $I$ is $O_p(\tilde{T})$.

Consider part $II$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\zeta_{st}F_t'||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\zeta_{st}F_t'||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\zeta_{st}F_t'||^2),$$

where

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\zeta_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)\sum_{t=1}^{\tilde{T}}||F_t||^2 \\ &\quad \sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_{is}u_{it}-\mathrm{E}(u_{is}u_{it}))]^2 \\ &\leq O_p(\frac{\tilde{T}^2}{N}\delta_{I,NT}^{-2}),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\zeta_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT^4}\sum_{s=1}^{T}||F_s||^2\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{T}||\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_{is}u_{it}-\mathrm{E}(u_{is}u_{it}))||^2\sum_{t=1}^{\tilde{T}}||F_t||^2||H_I'||^2 \\ &= O_p(\frac{\tilde{T}^2T}{N}).\end{aligned}$$

Thus, $II$ is $O_p(\frac{\tilde{T}^2T}{N})$.

Consider part $III$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\eta_{st}F_t'||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\eta_{st}F_t'||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\eta_{st}F_t'||^2).$$

For the first expression in the bracket,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\eta_{st}F_t'||^2 &\leq \frac{1}{N^2T^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)\sum_{s=1}^{T}||F_s||^2\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\Lambda_1'u_tF_t'||^2\\ &= O_p(\frac{\tilde{T}T}{N}\delta_{I,NT}^{-2}),\end{aligned}$$

as $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\Lambda'u_tF_t'||^2 = O_p(N\tilde{T}T^2)$ by FCLT.

For the second expression,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\eta_{st}F_t'||^2 &\leq \frac{1}{N^2T^4}(\sum_{s=1}^{T}||F_s||^2)^2\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\Lambda_1'u_tF_t'||^2||H_I'||^2\\ &= O_p(\frac{\tilde{T}T^2}{N}).\end{aligned}$$

As a result, $III$ is $O_p(\frac{\tilde{T}T^2}{N})$.

Finally consider part $IV$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\xi_{st}F_t'||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\xi_{st}F_t'||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\xi_{st}F_t'||^2).$$

For the first term,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\xi_{st}F_t'||^2 &\leq \frac{\tilde{T}}{NT^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)\sum_{s=1}^{T}||\frac{1}{\sqrt{N}}\Lambda_1'u_s||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2\\ &= O_p(\frac{\tilde{T}T^2}{N}\delta_{I,NT}^{-2}).\end{aligned}$$

Consider the second term,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\xi_{st}F_t'||^2 &\leq \frac{\tilde{T}}{N^2T^4}||\sum_{s=1}^{T}F_su_s'\Lambda_1||^2(\sum_{t=1}^{\tilde{T}}||F_t||^2)^2||H_I'||^2\\ &= O_p(\frac{\tilde{T}T^2}{N}),\end{aligned}$$

because $||\sum_{s=1}^{T}F_su_s'\Lambda_1||=O_p(\sqrt{N}T)$. Thus, $IV$ is $O_p(\frac{\tilde{T}T^2}{N})$. As $||V_{NT}||=O_p(1)$ from Bai (2004), the proof is finished by combining these results.

**Proof of Lemma 10**

$$\begin{aligned}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_{I,t}-H_I'F_t)u_{it}||^2 &\leq 4||V_{NT}^{-1}||(\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\gamma_{st}u_{it}||^2}_{I}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\zeta_{st}u_{it}||^2}_{II}\\ &\quad+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\eta_{st}u_{it}||^2}_{III}+\underbrace{\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\xi_{st}u_{it}||^2}_{IV}).\end{aligned}$$

We begin with part $I$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\gamma_{st}u_{it}||^2\leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\gamma_{st}u_{it}||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\gamma_{st}u_{it}||^2).$$

The first expression on the right hand side is bounded by

$$\frac{\tilde{T}^2}{T^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)[\frac{1}{\tilde{T}}\sum_{\tau=1}^{\tilde{T}}(\frac{1}{\sqrt{\tilde{T}}}\sum_{t=1}^{\tau}u_{it}\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}(u_{jt}^2))^2],$$

which is $O_p(\frac{\tilde{T}^2}{T^3}\delta_{I,NT}^{-2})$. The second term is bounded by

$$\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}F_t u_{it}\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}(u_{jt}^2)||^2||H_I'||^2 = O_p(\frac{\tilde{T}}{T^2}),$$

by FCLT. Thus, $I$ is $O_p(\frac{\tilde{T}}{T^2})$.

Now Part $II$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\zeta_{st}u_{it}||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\zeta_{st}u_{it}||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\zeta_{st}u_{it}||^2).$$

The first is similar to that in Lemma 2, which is $O_p(\frac{\tilde{T}^3}{NT^2}\delta_{I,NT}^{-2})$. The second is $O_p(\frac{\tilde{T}^3}{NT})$. Thus, $II$ is $O_p(\frac{\tilde{T}^3}{NT})$.

Consider part $III$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\eta_{st}u_{it}||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\eta_{st}u_{it}||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\eta_{st}u_{it}||^2).$$

The first term is bounded by $\frac{\tilde{T}}{NT^3}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)\sum_{s=1}^{T}||F_s||^2\sum_{t=1}^{\tilde{T}}u_{it}^2\sum_{t=1}^{\tilde{T}}||\frac{1}{\sqrt{N}}\Lambda_1'u_t||^2$, which is $O_p(\frac{\tilde{T}^3}{NT}\delta_{I,NT}^{-2})$. For the second term, similar to that in Lemma 2,

$$\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\eta_{st}u_{it}||^2 \quad \leq \quad O_p(\frac{\tilde{T}^2}{N})+O_p(\frac{\tilde{T}^3}{N^2}).$$

$III$ is thus $O_p(\frac{\tilde{T}^3}{N\min(N,\tilde{T})})$.

Finally part $IV$.

$$\sum_{\tau=1}^{\tilde{T}}||\frac{1}{T^2}\sum_{t=1}^{\tau}\sum_{s=1}^{T}\hat{F}_{I,s}\xi_{st}u_{it}||^2 \leq 2(\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\xi_{st}u_{it}||^2+\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}H_I'F_s\xi_{st}u_{it}||^2).$$

For the first expression in the bracket,

$$\begin{aligned}\frac{1}{T^4}\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}\sum_{s=1}^{T}(\hat{F}_{I,s}-H_I'F_s)\xi_{st}u_{it}||^2 &\leq \frac{1}{NT^2}(\frac{1}{T}\sum_{s=1}^{T}||\hat{F}_{I,s}-H_I'F_s||^2)(\frac{1}{T}\sum_{s=1}^{T}||\frac{1}{\sqrt{N}}\Lambda_1'u_s||^2)\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}F_tu_{it}||^2\\ &= O_p(\frac{\tilde{T}}{N}\delta_{I,NT}^{-2}).\end{aligned}$$

The second term is bounded by $\frac{1}{NT^2}||\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\Lambda_1'u_sF_s'||^2\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}F_tu_{it}||^2||H_I'||^2$, which is $O_p(\frac{\tilde{T}}{N})$. Thus, $IV$ is $O_p(\frac{\tilde{T}}{N})$.

As a result, $\sum_{\tau=1}^{\tilde{T}}||\sum_{t=1}^{\tau}(\hat{F}_{I,t}-H_I'F_t)u_{it}||^2 = 4||V_{NT}^{-1}||(I+II+III+IV) = O_p(\frac{\tilde{T}^3}{\min(N^2,N\tilde{T},\tilde{T}^2T^2)})$.

**Proofs of Lemmas 11-13**

The proofs are straightforwardly adapted from Lemma B.4 of Bai (2004), Lemma B.1 of Bai (2003) and Theorem 1 of Bai & Ng (2002). The difference is mainly that the sum is taken over $\tilde{T}$ periods in my case, but is taken over $T$ periods in Bai & Ng (2002), Bai (2003) and Bai (2004). The adaptation is very similar to that in the proof of Lemma 3 and is thus omitted.

**Proof of Lemma 14**

It is similar to Lemma 6 and is thus omitted.

**Proof of Theorem 3**

As the proof is similar to that of Theorem 1, we start with (52) which now becomes

$$\frac{1}{\tilde{T}T^2}\sum_{\tau=1}^{T-[T\pi]}||(\hat{s}_{i\tau}(\pi)-\tilde{s}_{i\tau}(\pi))||^2 \leq \frac{3}{\tilde{T}T^2}\sum_{\tau=1}^{\tilde{T}}||I||^2+\frac{3}{\tilde{T}T^2}\sum_{\tau=1}^{\tilde{T}}||II||^2+\frac{3}{\tilde{T}T^2}\sum_{\tau=1}^{\tilde{T}}||III||^2.$$

By Lemma 9 and 10, $\sum_{\tau=1}^{\tilde{T}} ||I||^2 = O_p(\frac{\tilde{T}T^2}{\min(N,T^2)})$. By Lemma 11,12 and 14, $\sum_{\tau=1}^{\tilde{T}} ||II||^2 = O_p(\frac{\tilde{T}T^2}{\min(N,T^2)})$. By Lemma 10, $\sum_{\tau=1}^{\tilde{T}} ||III||^2 = O_p(\frac{\tilde{T}T^2}{\min(N,T^2)})$. As a result, $\widehat{\text{LM}}_i(\pi) = \widetilde{\text{LM}}_i(\pi) + O_p(\frac{1}{\sqrt{\min(N,T^2)}})$.

#### 4.7.4 Dataset

We apply the following transformations to the raw data:

1. Take logs if Logs=1;

2. Take the first difference if Diff=1 or the second difference if Diff=2;

3. Apply an one-sided moving average filter with a 3-period window if Filter=3 or with a 12-period window if Filter=12.

| Release | Series | Logs | Diff | Filter |
|---|---|---|---|---|
| GFK consumer confidence | Aggregate balance | 0 | 1 | 3 |
| GFK consumer confidence | How does the financial situation of your household now compare with what it was 12 months ago | 0 | 1 | 3 |
| GFK consumer confidence | How do you think the financial position of your household will change over the next 12 months | 0 | 1 | 3 |
| GFK consumer confidence | How do you think the general economic situation in this country has changed over the last 12 months | 0 | 1 | 3 |
| GFK consumer confidence | How do you think the general economic situation in this country will develop over the next 12 months | 0 | 1 | 3 |
| GFK consumer confidence | How do you think the level of unemployment will change over the next 12 months? | 0 | 1 | 3 |
| GFK consumer confidence | Do you think that there is an advantage for people to make major purchases at the present time | 0 | 1 | 3 |
| GFK consumer confidence | Over the next 12 months how do you think the amount of money you'll spend on major purchases will compare with what you spent over the last 12 months | 0 | 1 | 3 |
| ONS retail sale | All Retailers (Volume seasonally adjusted) | 1 | 1 | 3 |
| ONS retail sale | All Retailers (Value seasonally adjusted) | 1 | 1 | 3 |
| ONS retail sale | Predominantly food stores (Volume seasonally adjusted) | 1 | 1 | 3 |
| ONS retail sale | Predominantly food stores (Value seasonally adjusted) | 1 | 1 | 3 |
| ONS retail sale | Non-specialised stores | 1 | 1 | 3 |
| ONS retail sale | Textiles: clothing: footwear | 1 | 1 | 3 |
| ONS retail sale | Household goods stores | 1 | 1 | 3 |
| ONS retail sale | Non-store retailing & repair | 1 | 1 | 3 |
| CBI distributive trades | Retailing Sales | 0 | 1 | 3 |
| CBI distributive trades | Retailing Orders | 0 | 1 | 3 |
| CBI distributive trades | Retailing Sales for Time of Year | 0 | 1 | 3 |
| CBI distributive trades | Retailing Stocks | 0 | 1 | 3 |
| CBI distributive trades | Wholesaling Sales | 0 | 1 | 3 |
| CBI distributive trades | Wholesaling Orders | 0 | 1 | 3 |
| CBI distributive trades | Wholesaling Sales for Time of Year | 0 | 1 | 3 |
| CBI distributive trades | Wholesaling Stocks | 0 | 1 | 3 |
| CBI distributive trades | Motor Traders Sales | 0 | 1 | 3 |
| CBI distributive trades | Motor Traders Orders | 0 | 1 | 3 |
| CBI distributive trades | Motor Traders Sales for Time of Year | 0 | 1 | 3 |

| Release | Series | Logs | Diff | Filter |
|---|---|---|---|---|
| CBI distributive trades | Motor Traders Stocks | 0 | 1 | 3 |
| ONS travel | UK visits abroad: Expenditure abroad | 1 | 1 | 3 |
| ONS travel | OS visits to UK: Earnings | 1 | 1 | 3 |
| ONS trade | BOP: Balance, sa, Total Trade in Goods | 0 | 1 | 3 |
| ONS trade | BOP: Balance, Manufactures | 0 | 1 | 3 |
| ONS trade | BOP: Balance, Intermediate goods | 0 | 1 | 3 |
| ONS trade | BOP: IM: CVM: sa: Total Trade in Goods | 1 | 1 | 3 |
| ONS trade | BOP: EX: CVM: sa: Total Trade in Goods | 1 | 1 | 3 |
| ONS trade | BOP: Balance, Capital goods | 0 | 1 | 3 |
| ONS trade | BOP: IM, price index, Finished manufactures | 1 | 1 | 3 |
| ONS trade | Balance of Payments: Trade in Services: Total balance: CP | 0 | 1 | 3 |
| FTSE | All Share Dividend Yield | 0 | 1 | 3 |
| FTSE | All Share Price / Earnings Ratio | 0 | 1 | 3 |
| FTSE | All Share Price Index | 1 | 1 | 3 |
| FTSE | FTSE 100 | 1 | 1 | 3 |
| Exchange rate | Japanese Yen /£ | 1 | 1 | 3 |
| Exchange rate | United States Dollar /£ | 1 | 1 | 3 |
| Exchange rate | Effective exchange rate index, Sterling (Jan 2005=100) | 1 | 1 | 3 |
| Interest rate | Bank Of England Repo Rate | 0 | 1 | 3 |
| Interest rate | Overnight £ Inter-Bank Rate (Mean Libid/Libor)–8.30am | 0 | 1 | 3 |
| Interest rate | ICE LIBOR GBP 3 Month | 0 | 1 | 3 |
| Interest rate | 3 Month £ Inter-Bank Rate (Mean Libid/Libor)–10.30am | 0 | 1 | 3 |
| Interest rate | 6 Month £ Inter-Bank Rate (Mean Libid/Libor)–8.30am | 0 | 1 | 3 |
| Interest rate | Monthly average yield from British Government Securities, 20 years | 0 | 1 | 3 |
| Interest rate | End month level of discount rate, 3 month Treasury bills | 0 | 1 | 3 |
| Volatility | LIFFE FTSE 100 3 Months Constant Maturity: Implied Volatility | 0 | 1 | 3 |
| VRP | VRPSPOT(NOM,UK,5) | 0 | 1 | 3 |
| VRP | VRPSPOT(NOM,UK,10) | 0 | 1 | 3 |
| VRP | VRPSPOT(REL,UK,5) | 0 | 1 | 3 |
| VRP | VRPSPOT(REL,UK,10) | 0 | 1 | 3 |
| VRP | VRPSPOT(INF,UK,5) | 0 | 1 | 3 |
| VRP | VRPSPOT(INF,UK,10) | 0 | 1 | 3 |
| ONS labor | Total Claimant count | 0 | 1 | 3 |
| ONS labor | Claimant count rate | 0 | 1 | 3 |
| ONS labor | AEI (including bonuses), whole economy | 1 | 1 | 3 |

| Release | Series | Logs | Diff | Filter |
|---|---|---|---|---|
| ONS labor | AEI (including bonuses), private sector | 1 | 1 | 3 |
| ONS labor | In employment: UK: All: Aged 16+ | 1 | 1 | 3 |
| ONS labor | Unemployed: UK: All: Aged 16+ | 1 | 1 | 3 |
| ONS labor | Total actual weekly hours worked (millions): UK: All | 1 | 1 | 3 |
| ONS labor | Unemployed up to 6 months: UK: All: Aged 16+ | 1 | 1 | 3 |
| ONS labor | Unemployed over 6 and up to 12 months: UK: All: Aged 16+ | 1 | 1 | 3 |
| ONS labor | Unemployed over 12 months: UK: All: Aged 16+ | 1 | 1 | 3 |
| ONS labor | Unemployed over 24 months: UK: All: Aged 16+ | 1 | 1 | 3 |
| MST | Monthly amounts outstanding of monetary financial institutions' sterling M4 liabilities to private sector | 1 | 1 | 3 |
| MST | Monthly average amount outstanding of total sterling notes and coin in circulation out-side the Bank of England | 1 | 1 | 3 |
| MST | VQWK.M | 1 | 1 | 3 |
| MST | Money Stock: Retail Deposits and Cash in M4: NSA | 1 | 1 | 3 |
| MST | Monthly amounts outstanding of monetary financial institutions' sterling net lending to private sector | 1 | 1 | 3 |
| MST | Monthly value of total sterling approvals for secured lending to individuals | 1 | 1 | 3 |
| CIPS manufacturing | Consumer Goods Industries - Total New Orders | 1 | 1 | 3 |
| CIPS manufacturing | Supplier's Delivery Times | 1 | 1 | 3 |
| CIPS manufacturing | Employment | 1 | 1 | 3 |
| CIPS manufacturing | Stocks Of Finished Goods | 1 | 1 | 3 |
| CIPS manufacturing | Investment Goods Industries | 1 | 1 | 3 |
| CIPS manufacturing | Total New Orders | 1 | 1 | 3 |
| CIPS manufacturing | Output | 1 | 1 | 3 |
| CIPS manufacturing | Quantity Of Purchases | 1 | 1 | 3 |
| CIPS manufacturing | Stocks Of Purchases | 1 | 1 | 3 |
| CIPS manufacturing | Input Prices | 1 | 1 | 3 |
| ONS output | Industry C,D,E: All production industries | 1 | 1 | 3 |
| ONS output | Industry C: Mining & quarrying | 1 | 1 | 3 |
| ONS output | Industry D: Manufacturing | 1 | 1 | 3 |
| ONS output | Industry E: Electricity, gas and water supply | 1 | 1 | 3 |
| ONS output | Industry DA: Manuf of food, drink & tobacco | 1 | 1 | 3 |
| ONS output | Industry DB: Manuf of textile & textile products | 1 | 1 | 3 |
| ONS output | Industry DC: Manuf of leather & leather products | 1 | 1 | 3 |
| ONS output | Industry DD: Manuf of wood & wood products | 1 | 1 | 3 |

| Release | Series | Logs | Diff | Filter |
|---|---|---|---|---|
| ONS output | Industry DE: Pulp/paper/printing/publishing industries | 1 | 1 | 3 |
| ONS output | Industry DF: Manuf coke/petroleum prod/nuclear fuels | 1 | 1 | 3 |
| ONS output | Industry DG: Manuf of chemicals & man-made fibres | 1 | 1 | 3 |
| ONS output | Industry DH: Manuf of rubber & plastic products | 1 | 1 | 3 |
| ONS output | Industry DI: Manuf of non-metallic mineral products | 1 | 1 | 3 |
| ONS output | Industry DJ: Manuf of basic metals & fabricated prod | 1 | 1 | 3 |
| ONS output | Industry DK: Manuf of machinery & equipment | 1 | 1 | 3 |
| ONS output | Industry DL: Manuf of electrical & optical equipment | 1 | 1 | 3 |
| ONS output | Industry DM: Manuf of transport equipment | 1 | 1 | 3 |
| CBI monthly trends | Do you consider that in vo-lume terms, your present total order book is above normal? | 0 | 1 | 3 |
| CBI monthly trends | Do you consider that in volume terms, your present export order book is above normal? | 0 | 1 | 3 |
| CBI monthly trends | Adequacy of Stocks of Finished Goods | 0 | 1 | 3 |
| CBI monthly trends | What is the expected trend over the next 4 months with regards to your volume of output? | 0 | 1 | 3 |
| CBI monthly trends | What is the expected trend over the next 4 months with regards to average prices for domestic orders? | 0 | 1 | 3 |
| Experian Construction | United Kingdom: Experian Construction Activity (Index) | 1 | 1 | 3 |
| Brent crude | 1 Month Fwd, fob U$/BBL | 1 | 1 | 3 |
| Brent crude | Physical Del., fob U$/BBL | 1 | 1 | 3 |
| ONS PPI | NSI: All manufacturing: Materials only | 1 | 2 | 3 |
| ONS PPI | Prod of man ind excl.f,b, p & t sa | 1 | 2 | 3 |
| ONS PPI | Fuels Purchased by Man Ind Excl CCL | 1 | 2 | 3 |
| ONS PPI | NSI: M & F purchased by Man: Excl FBPT Excl CCL NSA | 1 | 2 | 3 |
| ONS PPI | Output of manufactured products | 1 | 2 | 3 |
| ONS PPI | NSO: All Manufacturing excl duty: sa | 1 | 2 | 3 |
| ONS CPI | Food And Non-Alcoholic Beverages | 1 | 2 | 12 |
| ONS CPI | Alcoholic Beverages, Tobacco & Narcotics | 1 | 2 | 12 |
| ONS CPI | Clothing And Footwear | 1 | 2 | 12 |
| ONS CPI | Housing, Water And Fuels | 1 | 2 | 12 |
| ONS CPI | Furn, Hh Equip & Routine Repair Of House | 1 | 2 | 12 |
| ONS CPI | Health | 1 | 2 | 12 |
| ONS CPI | Transport | 1 | 2 | 12 |
| ONS CPI | Communication | 1 | 2 | 12 |

| Release | Series | Logs | Diff | Filter |
|---|---|---|---|---|
| ONS CPI | Recreation & Culture | 1 | 2 | 12 |
| ONS CPI | Education | 1 | 2 | 12 |
| ONS CPI | Hotels, Cafes And Restaurants | 1 | 2 | 12 |
| ONS CPI | Miscellaneous Goods And Services | 1 | 2 | 12 |
| HBF | Change in net house prices during the month | 0 | 1 | 3 |
| HBF | Site Visits, compared with a year ago | 0 | 1 | 3 |
| HBF | Net Reservations, compared with a year ago | 0 | 1 | 3 |
| HAC | RICS Housing Market Survey, Prices, England and Wales, Net Balance | 0 | 1 | 3 |
| HAC | Ratio Of RICS Sales Series To RICS Stock Series | 0 | 1 | 3 |
| US | VIX | 0 | 1 | 3 |
| US | TED | 0 | 1 | 3 |
| US | Manufacturing PMI | 1 | 1 | 3 |
| US | IP manufacturing | 1 | 1 | 3 |
| US | EXPORTS F.A.S. CURA | 1 | 1 | 3 |
| US | IMPORTS F.A.S. CURA | 1 | 1 | 3 |
| US | Retail sales (current prices) | 1 | 1 | 3 |
| US | S&P/CASE-SHILLER HOME PRICE INDEX - 10-CITY COMPOSITE SADJ | 1 | 2 | 3 |
| US | Employment | 1 | 1 | 3 |
| US | Unemployment rate | 0 | 1 | 3 |
| US | M2 | 1 | 1 | 3 |
| US | S&P monthly average | 0 | 1 | 3 |
| Germany | Assessment of business | 1 | 1 | 3 |
| Germany | Bussiness expectation | 1 | 1 | 3 |
| Germany | Consumer confidence | 0 | 1 | 3 |
| Germany | Export | 1 | 1 | 3 |
| Germany | Import | 1 | 1 | 3 |
| Germany | IP manufacturing | 1 | 1 | 3 |

# References

Acharya, Viral, Pedersen, Lasse, Philippon, Thomas, & Richardson, Matthew. 2012. Measuring Systemic Risk.

Aielli, Gian Piero. 2013. Dynamic Conditional Correlation: on Properties and Estimation. *Journal of Business & Economic Statistics*, **31**(3), 282–299.

Aït-Sahalia, Yacine, Mykland, Per A, & Zhang, Lan. 2005. How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial studies*, **18**(2), 351–416.

Andersen, T.G., Bollerslev, T., Diebold, F.X., & Labys, P. 2000. Great Realizations. *Risk*, **13**, 105–108.

Andersen, T.G., Bollerslev, T., Diebold, F.X., & Labys, P. 2001a. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, **96**, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., & Wu, J. 2006. Realized beta: Persistence and Predictability. in T. Fomby and D. Terrell, eds, Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series, Vol. 20, Elsevier Science, pp. 1-39.

Andersen, Torben G, Bollerslev, Tim, Diebold, Francis X, & Ebens, Heiko. 2001b. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics*, **61**(1), 43–76.

Andrieu, C., & Doucet, A. 2002. Particle Filtering for Partially Observed Gaussian State Space Models. *Journal of the Royal Statistical Society: Series B*, **64**, 827–836.

Angelini, Elena, Camba-Mendez, Gonzalo, Giannone, Domenico, Reichlin, Lucrezia, & Rünstler, Gerhard. 2011. Short-term Forecasts of Euro Area GDP Growth. *The Econometrics Journal*, **14**, C25–C44.

Arellano, M., & Bond, S. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58**, 277–297.

Bai, Jushan. 2003. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, **71**, 135–171.

Bai, Jushan. 2004. Estimating Cross-section Common Stochastic Trends in Nonstationary Panel Data. *Journal of Econometrics*, **122**, 137–183.

Bai, Jushan, & Liao, Yuan. 2013. Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood. manuscript.

Bai, Jushan, & Ng, Serena. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, **70**, 191–221.

Bai, Jushan, & Ng, Serena. 2008a. Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics*, **146**, 304–317.

Bai, Jushan, & Ng, Serena. 2008b. Large Dimensional Factor Analysis. *Foundations and Trends (R) in Econometrics*, **3**, 89–163.

Bai, Jushan, & Ng, Serena. 2009. Boosting Diffusion Indices. *Journal of Applied Econometrics*, **24**, 607–629.

Bali, T.G., & Engle, R.F. 2010. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. *Journal of Monetary Economics*, **57**, 377–390.

Bali, T.G., Engle, R.F., & Tang, Y. 2012. Dynamic Conditional Beta is Alive and Well in the Cross-Section of Daily Stock Returns. Working Paper.

Bandi, F.M., & Russell, J.R. 2005. Realized Covariation, Realized beta and Microstructure Noise. Unpublished paper, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O.E., & Shephard, N. 2002. Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 253–280.

Barndorff-Nielsen, O.E., & Shephard, N. 2004. Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica*, **72**, 885–925.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., & Shephard, N. 2008. Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica*, **76**, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., & Shephard, N. 2009. Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal*, **12**, C1–C32.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., & Shephard, N. 2011. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of

Equity Prices with Noise and Non-synchronous Trading. *Journal of Econometrics*, **162**, 149–169.

Basu, S., & Bundick, B. 2012. Uncertainty Shocks in a Model of Effective Demand.

Bates, Brandon J, Plagborg-Møller, Mikkel, Stock, James, & Watson, Mark. 2013. Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, **177**, 289–304.

Bauwens, L., Laurent, S., & Rombouts, J.V.K. 2006. Multivariate GARCH Models: a Survey. *Journal of Applied Econometrics*, **21**, 79–109.

Blanchard, O., & Simon, J. 2001. The Long and Large Decline in US Output Volatility. *Brookings Papers on Economic Activity*, 135–174.

Bloom, N. 2009. The Impact of Uncertainty Shocks. *Econometrica*, **77**, 623–685.

Blundell, R., & Bond, S. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, **87**, 115–143.

Boivin, Jean, & Ng, Serena. 2006. Are More Data Always Better for Factor Analysis? *Journal of Econometrics*, **132**, 169–194.

Bollerslev, T., & Wooldridge, J.M. 1992. Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric reviews*, **11**, 143–172.

Bollerslev, T., & Zhang, B.Y.B. 2003. Measuring and Modeling Systematic Risk in Factor Pricing Models using High-frequency Data. *Journal of Empirical Finance*, **10**, 533–558.

Braun, P.A., Nelson, D.B., & Sunier, A.M. 1995. Good News, Bad News, Volatility, and Betas. *Journal of Finance*, **50**(December), 1575–1603.

Breitung, Jörg, & Eickmeier, Sandra. 2011. Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics*, **163**, 71–84.

Brownlees, C.T., & Engle, R.F. 2012. *Volatility, Correlation and Tails for Systemic Risk Measurement*. NYU Stern School of Business Working Paper.

Chen, Liang, Dolado, Juan J, & Gonzalo, Jesús. 2014. Detecting Big Structural Breaks in Large Factor Models. *Journal of Econometrics*, **180**, 30–48.

Chen, R., & Liu, J. 2000. Mixture Kalman Filters. *Journal of the Royal Statistical Society B*, **62**, 493–508.

Cheng, Xu, Liao, Zhipeng, & Schorfheide, Frank. 2014. Shrinkage Estimation of High-dimensional Factor Models with Structural Instabilities. manuscript.

Choi, In. 2012. Efficient Estimation of Factor Models. *Econometric Theory*, **28**, 274–308.

Choi, In, & Jeong, HanBat. 2013. Model Selection for Factor Analysis: Some New Criteria and Performance Comparisons. Research Institute for Market Economy Working Paper, Sogang University.

Christensen, Kim, Kinnebrock, Silja, & Podolskij, Mark. 2010. Pre-averaging Estimators of the ex-post Covariance Matrix in Noisy Diffusion Models with Non-Synchronous Data. *Journal of Econometrics*, **159**(1), 116–133.

Corradi, Valentina, & Swanson, Norman R. 2014. Testing for Structural Stability of Factor Augmented Forecasting Models. *Journal of Econometrics*, **182**, 100–118.

Creal, D. 2012. A Survey of Sequential Monte Carlo Methods for Economics and Finance. *Econometric Reviews*, **31**, 245–296.

Del Negro, Marco, & Otrok, Christopher. 2008. Dynamic Factor Models with Time-varying Parameters: Measuring Changes in International Business Cycles. Staff Report No. 326, Federal Reserve Bank of New York.

Diebold, F.X., & Mariano, R.S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, **13**, 253–263.

Doucet, A., Godsill, S., & Andrieu, C. 2000. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistics and Computing*, **10**, 197–208.

Durbin, James, & Koopman, Siem Jan. 2012. Time series analysis by state space methods. Oxford University Press.

Engle, R.F., & Kelly, B. 2012. Dynamic Equicorrelation. *Journal of Business & Economic Statistics*, **30**, 212–228.

Engle, R.F., Ng, V.K., & Rothschild, M. 1990. Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics*, **45**, 213–237.

Engle, R.F., Pakel, C., Shephard, N., & Sheppard, K. 2014. Fitting Vast Dimensional Time-Varying Covariance Models. Working Paper.

Fernández-Villaverde, J., & Rubio-Ramírez, J. 2007. Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies*, **74**, 1059–1087.

Fernández-Villaverde, J., & Rubio-Ramírez, J. 2013. Macroeconomics and Volatility: Data, Models, and Estimation. in Daron Acemoglu, Manuel Arellano and Eddie Dekel, eds, Advances in Economics and Econometrics: Tenth World Congress, Vol. 3, Cambridge University Press, 137-183.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., & Rubio-Ramírez, J. 2013. Fiscal Volatility Shocks and Economic Activity.

Ferson, W.E., & Harvey, C.R. 1993. The Risk and Predictability of International Equity Returns. *Review of Financial Studies*, **6**, 527–566.

Frankel, Jeffrey A, & Rose, Andrew K. 1996. A Panel Project on Purchasing Power Parity: Mean Reversion within and between Countries. *Journal of International Economics*, **40**(1), 209–224.

Giannone, Domenico, Reichlin, Lucrezia, & Small, David. 2008. Nowcasting: the Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, **55**, 665–676.

Glosten, Lawrence R, Jagannathan, Ravi, & Runkle, David E. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, **48**(5), 1779–1801.

Golosnoy, Vasyl, Gribisch, Bastian, & Liesenfeld, Roman. 2012. The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility. *Journal of Econometrics*, **167**(1), 211–223.

Gordon, N. J, Salmond, D. J, & Smith, A. FM. 1993. Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation. *IEE Proceedings F*, **140**, 107–113.

Gupta, Arjun K, & Nagar, Daya K. 2000. *Matrix Variate Distributions*. Vol. 104. Chapman & Hall/CRC.

Halbleib-Chiriac, Roxana, & Voev, Valeri. 2014. Forecasting Covariance Matrices: a Mixed Frequency Approach. *Journal of Financial Econometrics*, **doi: 10.1093/jjfinec/nbu031**, 1–35.

Hamilton, James D. 2010. Macroeconomics and ARCH. in Tim Bollerslev, Jeffry R. Russell and Mark Watson, eds, Volatility and Time Series Econometrics: Essays in Honor of Robert Engle, Oxford University Press, 79-96.

Han, Xu, & Inoue, Atsushi. 2015. Tests for Parameter Instability in Dynamic Factor Models. *Econometric Theory*, **31**, 1117–1152.

Hansen, Bruce. 1990. Lagrange Multiplier Tests for Parameter Instability in Non-linear Models. unpublished manuscript, University of Rochester, Department of Economics.

Hansen, Bruce E. 1992. Tests for Parameter Instability in Regressions with I(1) Processes. *Journal of Business & Economic Statistics*, **10**, 45–59.

Hansen, Peter Reinhard, Lunde, Asger, & Voev, Valeri. 2014. Realized beta GARCH: a Multivariate GARCH Model with Realized Measures of Volatility. *Journal of Applied Econometrics*, **29**(5), 774–799.

Hayakawa, K., & Pesaran, M. 2015. Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models with Cross-Sectional Heteroskedasticity. *Journal of Econometrics*, **188**, 111–134.

Hsiao, C., Pesaran, M., & Tahmiscioglu, K. 2002. Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods. *Journal of Econometrics*, **109**, 107–150.

Islam, Nazrul. 1995. Growth Empirics: a Panel Data Approach. *The Quarterly Journal of Economics*, **110**(4), 1127–1170.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., & Vetter, M. 2009. Microstructure Noise in the Continuous Case: the Pre-averaging Approach. *Stochastic Processes and their Applications*, **119**, 2249–2276.

Jagannathan, Ravi, & Wang, Zhenyu. 1996. The Conditional CAPM and the Cross-section of Expected Returns. *Journal of Finance*, **51**(1), 3–53.

Jin, Xin, & Maheu, John M. 2013. Modeling Realized Covariances and Returns. *Journal of Financial Econometrics*, **11**(2), 335–369.

Karlsson, R., Schon, T., & Gustafsson, F. 2005. Complexity Analysis of the Marginalized Particle Filter. *IEEE Transactions on Signal Processing*, **53**, 4408–4411.

Kim, Chang-Jin, & Nelson, Charles R. 1999. Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-switching Model of the Business Cycle. *Review of Economics and Statistics*, **81**(4), 608–616.

Kim, S., Shephard, N., & Chib, S. 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, **65**, 361–393.

Koop, G, & Korobilis, D. 2010. Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends in Econometrics*, **3**(4), 267–358.

Laurent, Sébastien, Rombouts, Jeroen VK, & Violante, Francesco. 2013. Consistent Ranking of Multivariate Volatility Models. **173**, 1–10.

Lewellen, Jonathan, & Nagel, Stefan. 2006. The Conditional CAPM does not Explain Asset-Pricing Anomalies. *Journal of Financial Economics*, **82**(2), 289–314.

Liu, Jun S, & Chen, Rong. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93**(443), 1032–1044.

McConnell, Margaret M., & Perez-Quiros, Gabriel. 2000. Output Fluctuations in the United States: What Has Changed since the Early 1980's? *American Economic Review*, **90**(5), 1464–1476.

Newey, Whitney K., & McFadden, Daniel L. 1994. Large Sample Estimation and Hypothesis Testing. In Robert F. Engle and Daniel L. McFadden(Eds.), Handbook of Econometrics, Volume 4, 2111-2245. North Holland.

Noureldin, Diaa, Shephard, Neil, & Sheppard, Kevin. 2012. Multivariate High-frequency-based Volatility (HEAVY) Models. *Journal of Applied Econometrics*, **27**(6), 907–933.

Nyblom, Jukka. 1989. Testing for the Constancy of Parameters Over Time. *Journal of the American Statistical Association*, **84**, 223–230.

Olsson, J., & Rydén, T. 2008. Asymptotic Properties of Particle Filter-based Maximum Likelihood Estimators for State Space Models. *Stochastic Processes and their Applications*, **118**, 649–680.

Onatski, Alexei. 2010. Determining the Number of Factors from Empirical Distribution of Eigenvalues. *The Review of Economics and Statistics*, **92**, 1004–1016.

Patton, A.J. 2011. Volatility Forecast Comparison using Imperfect Volatility Proxies. *Journal of Econometrics*, **160**(1), 246–256.

Patton, A.J., & Sheppard, K. 2009. Evaluating Volatility and Correlation Forecasts. in T.G. Andersen, R.A. Davis, J.P. Kreib and T. Mikosch, eds, Handbook of Financial Time Series, Springer, pp. 801–838.

Patton, A.J., & Verardo, M. 2012. Does beta Move with News? Firm-specific Information Flows and Learning about Profitability. *Review of financial studies*, **25**, 2789–2839.

Shephard, N. 2013. Martingale Unobserved Component Models.

Shephard, N., & Sheppard, K. 2010. Realising the Future: Forecasting with High-frequency-based Volatility (HEAVY) Models. *Journal of Applied Econometrics*, **25**, 197–231.

Shephard, Neil, & Xiu, Dacheng. 2013. Econometric Analysis of Multivariate Realised QML: efficient positive semi-definite estimators of the covariation of equity prices. Unpublished manuscript, University of Oxford.

Stock, James, & Watson, Mark. 2009. Forecasting in Dynamic Factor Models Subject to Structural Instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173–205.

Stock, James, & Watson, Mark. 2011. Dynamic Factor Models. *Oxford Handbook of Economic Forecasting*, **1**, 35–59.

Stock, James H, & Watson, Mark W. 2003. Has the Business Cycle Changed and Why? *NBER: Macroeconomics Annual 2002*, 159–230.

Takongmo, Charles Olivier Mao, & Stevanovic, Dalibor. 2014. Selection of the Number of Factors in Presence of Structural Instability: a Monte Carlo Study. manuscript.

White, Halbert. 1996. *Estimation, Inference and Specification Analysis*. Cambridge university press.

Wooldridge, J.M. 1990. A Unified Approach to Robust, Regression-Based Specification Tests. *Econometric Theory*, **6**, 17–43.

Wu, Jhy-Lin, & Chen, Show-Lin. 2001. Mean Reversion of Interest Rates in the Eurocurrency Market. *Oxford Bulletin of Economics and Statistics*, **63**(4), 459–473.

Xu, Wen. 2014. Nowcasting UK GDP Growth with Refined Factor Models. manuscript.

Yamamoto, Yohei, & Tanaka, Shinya. 2013. Testing for Factor Loading Structural Change under Common Breaks. manuscript.

Yamazaki, Daisuke, & Kurozumi, Eiji. 2014. Testing for Parameter Constancy in the Time

Series Direction in Fixed-effect Panel Data Models. *Journal of Statistical Computation and Simulation*.

Zakoian, Jean-Michel. 1994. Threshold Heteroskedastic Models. *Journal of Economic Dynamics and control*, **18**(5), 931–955.

Zhang, L. 2006. Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. *Bernoulli*, **12**, 1019–1043.

Zumbach, Gilles. 2007. The RiskMetrics 2006 Methodology.