6-2022

Bayesian and machine learning methods with applications in asset pricing

Yaohan CHEN
Singapore Management University, yaohan.chen.2017@phdecons.smu.edu.sg

BAYESIAN AND MACHINE LEARNING METHODS WITH APPLICATIONS IN ASSET PRICING

YAOHAN CHEN

SINGAPORE MANAGEMENT UNIVERSITY
2022

Bayesian and Machine Learning Methods with Applications in Asset Pricing

by

YAOHAN CHEN

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Economics

Dissertation Committee:

Jun YU (Supervisor / Chair)
Lee Kong Chian Professor of Economics and Finance
Singapore Management University

Jia LI
Lee Kong Chian Professor of Economics
Singapore Management University

Peter C.B. PHILLIPS
Distinguished Term Professor of Economics
Singapore Management University

Weikai LI
Assistant Professor of Finance
Singapore Management University

Singapore Management University
2022

I hereby declare that this PhD dissertation is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree in any university previously.

YAOHAN CHEN

2022

Bayesian and Machine Learning Methods with Applications in Asset Pricing

YAOHAN CHEN

Abstract

The dissertation consists of three essays on asset pricing by constructing new data set and developing new methodologies. In the first chapter, we conduct empirical studies on the volatility-managed portfolios in the Chinese stock market. Using data from the Chinese stock market, we have found that the main empirical findings in Moreira and Muir (2017) break down. Based on the empirical findings, we exploit a comprehensive set of 99 equity strategies in the Chinese stock market to analyze the value of managed portfolios. Based on these 99 equity trading strategies, we find that there exists no systematic gain from scaling the original portfolios using volatility. Our empirical results suggest that one should be careful to use volatility-managed portfolios in practice as the expected performance gains are rather limited.

In the second chapter, we review a Bayesian interpretable machine-learning method proposed by Kozak, Nagel, and Santosh (2020). We show how the method can link two strands of literature, namely the literature on empirical asset pricing and the literature on statistical learning. Based on a recently developed data-cleaning technique, we obtain 123 financial and accounting cross-sectional equity characteristics in the Chinese stock market. When applying the method of Kozak, Nagel, and Santosh (2020) to the Chinese stock market, we find that it is futile to summarize the stochastic discount factor (SDF) in the Chinese stock market as the exposure of several dominant cross-sectional equity characteristics in-sample. A cross-validated out-of-sample analysis further supports this finding.

In the third chapter, we propose several alternative parametric models for spot volatility in high frequency, depending on whether or not jumps, seasonality, and announcement effects are included. Together with these alternative parametric

models, nonlinear non-Gaussian state-space models are introduced based on the fixed-k theory of Bollerslev, Li, and Liao (2021). According to Bollerslev, Li, and Liao (2021), the log fixed-k estimator of spot volatility equals the true log spot volatility plus a non-Gaussian random variable. Bayesian methods are introduced to estimate and compare these alternative models and to extract volatility from the estimated models. Simulation studies suggest that the Bayesian methods can in general work well. Empirical studies using high-frequency market indexes and individual stock prices reveal several important results. As an application of extracting volatility, we quantify the strategic value of information.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest gratitude to my supervisor Professor Jun Yu, for his professional guidance and kind patience devoted to me over the past years during my study at Singapore Management University. I could not have undertaken my Ph.D. journey without his kind help. I am also deeply indebted to my dissertation committee members, Professor Jia Li, Professor Peter C.B. Phillips, and Assistant Professor Weikai Li. I appreciate their time spent with me for academic discussions.

I would also like to extend my sincere thanks to my friends that I have met from Singapore Management University, especially two senior friends, Assistant Professor Xiaobin Liu and Associate Professor Tao Zeng at Zhejiang University. It is their support and advice to me that make me feel not lonely during the Ph.D. journey.

I am also grateful to the librarians of the Singapore Management University library. I could not have conducted my research without database support from the Singapore Management University library.

Finally, I wish to thank my parents. Their belief in me brings me much courage in face of difficulties and has kept my spirits high during the tough times.

Chapter 1

Do Volatility-Managed Portfolios Work? Empirical Evidence from the Chinese Stock Market

1.1 Introduction

Volatility played a vital role in financial decision making, including, for instance, derivatives pricing, portfolio selection, and risk management (Engle, 2004). Early studies such as Fleming, Kirby, and Ostdiek (2001) and Fleming, Kirby, and Ostdiek (2003) document the advantage of using volatility information to improve the portfolio performance. More recent studies further document the gain associated with volatility-managed portfolios of trading strategies (for instance, Ang, 2014; Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016; Moreira and Muir, 2017, 2019; Eisdorfer and Misirli, 2020).

The basic idea of the volatility-managed portfolio is to scale the original portfolios (strategies) by taking conservative positions in the underlying factors when volatility was high and taking more aggressively levered positions following periods of low volatility. This idea can be generally understood from the global minimum variance portfolio in the conventional optimal portfolio theory (see Basak, Jagannathan, and

Ma, 2009).

Let Σ denote the variance-covariance matrix of assets and μ denote the corresponding expected returns of assets. Then the optimal portfolio theory suggests that the optimal allocation weights, $\boldsymbol{w}$, assigned to assets contained in the global minimum variance portfolio is proportional $\Sigma^{-1}\mu$. If μ is fixed, then the magnitude of each element in $\boldsymbol{w}$ is proportional to Σ^{-1}. Consequently, volatility management can be heuristically interpreted as putting smaller weights on assets with greater volatility and larger weights on those with less volatility.

However, all these documented successes of using volatility to manage portfolios are mainly restricted to one or a few strategies (factors). Specifically, those portfolios studied in Ang (2014) are mainly about conventional benchmark factors (Fama-French factors) while both the discussions in Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) are restricted to momentum-related trading strategies. By contrast, Moreira and Muir (2017) make the corresponding discussions mainly based on 10 leading factors that are widely used in empirical asset pricing literature by adding some recently proposed strategies such as the betting-against-beta strategy in Frazzini and Pedersen (2014). In this regard, those empirical findings fail to provide a broad view demonstrating how general these volatility-managed portfolios can perform improvements in comparison to the unmanaged ones.

This issue has been noticed in Cederburg, O'Doherty, Wang, and Yan (2020). In particular, Cederburg, O'Doherty, Wang, and Yan (2020) accentuate that, although using the set of leading factors as anomaly portfolios (as in Moreira and Muir, 2017) for analysis reconciles well with the leading asset pricing models, it fails to accommodate the recent findings from some machine learning methods that a larger set of anomalies (firm-level characteristics) is needed to be jointly studied (see Kelly, Pruitt, and Su, 2019; Kozak, Nagel, and Santosh, 2020; Kozak, 2020). In other words, although there exists a widely acknowledged base for volatility-managed portfolios, their performance is far from reaching a consensus.

Other than the debate between Moreira and Muir (2017) and Cederburg,

O'Doherty, Wang, and Yan (2020), Barroso and Detzel (2021) further point out that the documented extra gain from managing equity portfolios via volatility-timing vanishes once transaction costs of specific forms are accounted (for instance, the look-ahead bias considered in Liu, Tang, and Zhou, 2019). Last but not least, all the existing studies on this topic mainly focus on using data constructed from the U.S. stock market. Rarely is there any empirical analysis on whether this volatility-based strategy works in other stock markets.

We begin this chapter by applying the *market* volatility-timing strategy in Moreira and Muir (2017) to the Chinese stock market. To our surprise, we find that some documented empirical findings for the U.S. market do not hold in the Chinese stock market. As a result, our research motivation naturally stems from asking to what extent shall we support using information associated with volatility for portfolio management for gaining performance improvement? Or put it in another way, is volatility management as a portfolio management strategy still broadly applicable to other markets even without accounting for some recently proposed explanations (for instance those interpretations made from accommodating trading costs, Liu, Tang, and Zhou, 2019; Barroso and Detzel, 2021) for the controversial performance of volatility-managed portfolios?

Regarding the measurement for portfolio performance, Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) assess whether investors can improve anomaly portfolios' performances by scaling holding positions of the original portfolios based on comparing the Sharpe ratios of "volatility-managed" portfolios with those earned by the corresponding unscaled strategies. This is the so-called direct comparison. We follow this approach as in Cederburg, O'Doherty, Wang, and Yan (2020) to compare the volatility-managed portfolios with the un-managed ones directly. Specifically, we construct a relative comprehensive 99 equity (anomaly) portfolios using data collected from the Chinese stock market and the associated 99 volatility-managed anomaly portfolios. We find that for these 99 volatility-managed anomaly portfolios, only 14 of them can generate statistically significant Sharpe ratio

differences, which suggests that there exists no systematic evidence to support that investors can earn performance improvements from scaling the original anomaly portfolios using the volatility of previous period.

Apart from the direct comparison using the Sharpe ratios, we also apply another empirical method using spanning regression to check whether we can obtain performance gain by adjusting the holding positions via lagged volatility. In comparison to measuring performance gain using the Sharpe ratios directly, spanning regression was initially suggested in Moreira and Muir (2017). The essence is rooted in the appraisal ratio closely related to the asset pricing model test or comparison (see Gibbons, Ross, and Shanken, 1989; Barillas and Shanken, 2018). The main objective of this spanning regression methodology is to check whether there exists a statistically significant alpha by running univariate time-series regression using monthly excess returns (we will come back on this and discuss it more in detail both in Section 1.2 and Section 1.4). Given this objective associated with spanning regression, we can see that the major implication of spanning regression is whether investors can construct a new portfolio with higher Sharpe ratio by combining the volatility-managed portfolio with the original un-managed portfolio. This is why it is usually referred to the combination strategy in the literature. By applying spanning regression on our constructed broader sample of anomaly portfolios (99 equity trading strategies) in the Chinese stock market, we find 71 out of 99 volatility-managed anomaly portfolios earn positive alphas but with only 16 of them are statistically significant at a generally acceptable significance level. Besides, we also find another 8 volatility-managed portfolios earn significantly negative in-sample alpha generated from spanning regression. Thus, we have 24 anomaly portfolios in all that can be acceptably regarded as gaining performance improvement by combing the original ones with the ones scaled via volatility.

The rest of this chapter is summarized as follows: In Section 1.2, we review some basic concepts about volatility-managed portfolios and some technical details that have been discussed in literature. In Section 2.3, we discuss how we collect, clean,

and construct anomaly portfolios in the Chinese stock market. In Section 1.4, we conduct the empirical analysis of this chapter to check the performance of volatility-managed portfolios. Finally, Section 1.5 concludes this chapter.

1.2 Volatility-managed Portfolios: A Review

1.2.1 Construction of volatility-managed portfolios

As suggested in Moreira and Muir (2017), the basic idea for constructing volatility-managed portfolios is scaling an excess return by the inverse of its conditional variance. Thus, in each month the volatility-managed strategy increases or decreases risk exposure to the volatility-managed portfolio according to the conditional variance. The managed portfolio is then[1]

$$f_{t+1}^{\sigma} = \frac{c}{\hat{\sigma}_t^2(f)} f_{t+1}, \tag{1.1}$$

where f_{t+1} is the buy-and-hold portfolio excess return, $\hat{\sigma}_t^2(f)$ is a proxy for the portfolio's conditional variance with $\hat{\sigma}_t^2(f)$ constructed by using previous month's realized variance defined by

$$\hat{\sigma}_t^2(f) = RV_t^2(f) = \sum_{d=1/22}^{1} \left(f_{t+d} - \frac{\sum_{d=1/22}^{1} f_{t+d}}{22} \right)^2. \tag{1.2}$$

In practice, when there are no 22 trading days in a month, we may the use the alternative proxy for conditional variance suggested in Cederburg, O'Doherty, Wang, and Yan (2020) as follows

$$\hat{\sigma}_t^2(f) = \frac{22}{J_t} \sum_{j=1}^{J_t} \left(f_t^j \right)^2. \tag{1.3}$$

where $j = 1, \ldots, J_t$ index days in month t and f_t^j is the excess return for a given portfolio (factor) on day j of month t. The constant c in (1.1) controls the average

[1]Since f_{t+1} generally refers to factors, which are usually constructed as portfolios based on the cross-sectional sort on asset-specific characteristics, the volatility-managed portfolio can be alternatively interpreted as "PoP", namely the portfolio of portfolios. Besides, we also emphasize that we directly apply realized volatility to scale excess returns.

exposure of the strategy. It is selected to make the managed portfolio, f^{σ}_{t+1}, have the same unconditional standard deviation as the un-managed portfolio, f_{t+1}. In this chapter, we use the method of Cederburg, O'Doherty, Wang, and Yan (2020) (i.e. (1.3)) to calculate realized volatility.

To see the role of c in (1.1), first note that the unconditional variance of f^{σ}_{t+1} can be calculated as follows

$$\begin{aligned}\mathrm{Var}\left[f^{\sigma}_{t+1}\right] &= \mathbb{E}\left\{\mathrm{Var}_t\left[f^{\sigma}_{t+1}\right]\right\} \\ &= \mathbb{E}\left[\frac{c^2}{\hat{\sigma}^4_t(f)}\hat{\sigma}^2_{t+1}(f)\right] = \mathrm{Var}\left[f_{t+1}\right].\end{aligned}$$

Thus, if both the unconditional variances of f^{σ}_{t+1} and f_{t+1} are fixed at a specific value, say $\sigma^2(f)$, then the scaling constant c is the solution to the following equation

$$\mathbb{E}\left[\frac{c^2}{\hat{\sigma}^4_t(f)}\hat{\sigma}^2_{t+1}(f)\right] = \sigma^2(f).$$

Since the realized volatility measures the integrated volatility, by replacing $\hat{\sigma}^2_t$ with the integrated volatility σ^2_t we have

$$\mathbb{E}\left[\frac{c^2}{\sigma^4_t(f)}\sigma^2_{t+1}(f)\right] = \sigma^2(f).$$

To pin down the scaling constant c in practice, we can simply use the empirical measure as the probability measure. In particular, we can calculate the sample variance of the original factors, f_{t+1} and the sample variance of the volatility-managed counterpart (unscaled by c), $f_{t+1}/\hat{\sigma}^2_t(f)$. Then we set c to ensure the following equation hold

$$\widehat{\mathrm{Var}}\left[f_{t+1}\right] = c^2\widehat{\mathrm{Var}}\left[f_{t+1}/\hat{\sigma}^2_t(f)\right], \tag{1.4}$$

where $\widehat{\mathrm{Var}}[\cdot]$ denotes the sample variance.

In our setting, we always assume that investors have access to the risk-free asset. This is to facilitate the use of the excess returns for constructing portfolios

using arbitrary combination weights. To see this, recall the classical portfolio selection theory (Markowitz, 1952), where we usually adopt a vector denoted by $\boldsymbol{w}$ to represent the portfolio weights. Thus, If there are n assets (including both the risky and the risk-free assets), then $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$. In the literature we usually impose a restriction $\sum_{i=1}^n w_i = 1$. This restriction is not necessary if we focus on returns of the risky assets in excess of returns of the risk-free asset. Without loss of generality, we may assume that the n-th asset is the risk-free asset. Then for the remaining $(n-1)$ risky assets, we can specify the corresponding weights arbitrarily and then set the weight of the risk-free asset, w_n, to ensure the restriction (i.e. $\sum_{i=1}^n w_i = 1$) satisfied. In other words, for any risky asset indexed by i for $i = 1, \ldots, n-1$, the excess return $R_i^e = R_i - R_f$ can be combined to construct portfolios using arbitrary weights $(w_1, \ldots, w_{n-1})^\top$. In subsection 2.3.1, we will discuss more in detail how we construct zero-investment long-short portfolios based on cross-sectional characteristics (i.e. $w_n = 1$).

1.2.2 Motivation from the stylized fact about market portfolio

In this section, we use the market portfolio as an illustration of some stylized effects in the U.S. market and the Chinese stock market. Based on the data-cleaning technique of Jensen, Kelly, and Pedersen (2022), we find the empirical result, found by Moreira and Muir (2017) in the U.S market and used as the intuition for justifying volatility-managed portfolios, does not necessarily hold in the Chinese stock market. These empirical findings motivate the analysis in Section 1.4 for checking whether volatility management helps improve portfolio performance in the Chinese stock market.

Specifically, for the U.S. market, Moreira and Muir (2017) find that there is a strong (positive) relationship between the lagged volatility and the current volatility and that the mean-variance trade-off (measured as the average return divided by the variance) of the current period is negatively related to the volatility in the previous

period. We can replicate these empirical findings via the following implementation. First, for each month contained in the data sample, we calculate realized volatility associated with the market portfolio (i.e. the value-weighted return) using daily data. Then, we group months by the previous month's realized volatility and plot volatility and mean-variance trade-off over the subsequent month. This is summarized as follows,

[Place Figure 1.1 about here]

As we can see from Figure 1.1, for the U.S. market, we observe a positive relationship (as in Moreira and Muir, 2017) between the volatility of the current period and the volatility of the previous month, and the negative relationship between the mean-variance trade-off of the current period and the volatility of the previous month. However, when we apply the same procedure to the Chinese stock market we find that, while the positive relationship between the volatility of the current month and the volatility of the previous month still exists, the mean-variance trade-off of the current month is not negatively correlated with the volatility of the previous month. In addition to this, we also compare the cumulative market return of the U.S. market and that of the Chinese stock market. This is summarized in Figure 1.2. Specifically, in Figure 1.2a we plot the cumulative value-weighted return of the U.S. market from 1926; in Figure 1.2b, we plot the cumulative value-weighted return of the U.S. market from 1991, which also is the beginning of the sample period of the Chinese stock market. In Figure 1.2c we plot the cumulative value-weighted market return of the Chinese stock market. Figure 1.2a and Figure 1.2b jointly imply that for investments made in the U.S. equity market, in the long run, it pays to scale the market portfolio via volatility by decreasing the risk exposure when the market is volatile. However, Figure 1.2c implies that the advantage of the scaled market portfolio vanishes in the Chinese stock market. The value-weighted return of one-unit money invested in the Chinese stock market starting from 1991 generates a higher payoff in the long run than that from the volatility-managed counterpart.

[Place Figure 1.2 about here]

Given these empirical findings for the market portfolio, the empirical success of the volatility-managed portfolio in other stock markets is questionable.

1.3 Data

In cross-sectional asset pricing studies, it is important for researchers to carefully construct cross-sectional equity characteristics. In this section, we first briefly discuss the recent literature on constructing cross-sectional equity characteristics for asset pricing studies and explain how we use the existing methods to construct equity characteristics in the Chinese stock market. Then we discuss how characteristic-managed portfolios are constructed based on daily returns of individual assets in the Chinese stock market. We use these constructed characteristic-managed portfolios as the proxy for anomaly portfolios.

1.3.1 Individual equity characteristic data

Following Harvey and Liu (2014, 2015); Harvey, Liu, and Zhu (2016); Mclean and Pontiff (2016); Green, Hand, and Zhang (2017); Hou, Xue, and Zhang (2018); Gu, Kelly, and Xiu (2020); Demiguel, Martín, Nogales, and Uppal (2020); Freybergerk, Neuhierl, and Weber (2019); Kozak, Nagel, and Santosh (2020); Kozak (2020), we obtain firm-level equity characteristic data. Several standard data-cleaning routines are available in the literature. The method of Chen and Zimmermann (2020) is a successful response to the call for transparency and cooperation (Welch, 2019). Besides, Jensen, Kelly, and Pedersen (2022) provides a more comprehensive analysis by constructing a global dataset in response to the recent discussions on the replication crisis in empirical asset pricing studies.[2] We combine both the data cleaning routines in Chen and Zimmermann (2020) and Jensen, Kelly, and Pedersen (2022) to replicate

[2]Jensen, Kelly, and Pedersen (2022) also makes their replication procedures and data publicly available at `https://github.com/bkelly-lab/ReplicationCrisis`.

99 finance and accounting anomaly variables in the Chinese stock market from 1996 to 2020. All the data (including returns and accounting data) are obtained from the Center for Research in Security Prices (CRSP), Compustat, and the China Stock Market & Accounting Research (CSMAR) database, all of which can be downloaded from the Wharton Research Data Service (WRDS). These anomaly variables are normalized as in Freybergerk, Neuhierl, and Weber (2019) so that each characteristic is normalized over the cross-sectional dimension to take a value between 0 and 1. More precisely,

$$rc_{i,t}^{s} = \frac{\text{rank}\left(c_{i,t}^{s}\right)}{n_t + 1}, \tag{1.5}$$

where $c_{i,t}^{s}$ denotes the originally unscaled firm-level equity characteristic (indexed by superscript s) associated with stock i at time t and n_t denotes the total number of individual assets available for observations at time t. $\text{rank}(\cdot)$ denotes the cross-sectional ranking order of specific variable. Then, for each rank-transformed characteristic $rc_{s,t}^{i}$, we center it around the cross-sectional mean and divide it by the sum of average deviations from the cross-sectional mean for available stocks. Hence, we have,

$$z_{i,t}^{s} = \frac{\left(rc_{i,t}^{s} - \overline{rc}_{t}^{s}\right)}{\sum_{i=1}^{n_t}\left|rc_{i,t}^{s} - \overline{rc}_{t}^{s}\right|}, \tag{1.6}$$

where

$$\overline{rc}_{t}^{s} = \frac{1}{n_t}\sum_{i=1}^{n_t} rc_{i,t}^{s}.$$

Each column of Z_t is $\left(z_{1,t}^{s}, \ldots, z_{n_{t,1}}^{s}\right)^{\top}$. It is known in practice that individual characteristic data is imbalanced panel dat. For this reason, we exploit n_t rather than N to emphasize the time-varying cross-sectional dimension.[3]

[3]This also implicitly suggests that for each cross-section we only use those individual assets available as observations both for the corresponding returns and specific characteristics (indexed by s).

1.3.2 Characteristic-managed portfolios

Annual accounting data is realigned with monthly return data based on the following annual rebalancing rule. Returns at the monthly frequency from July of year t to June of year $t+1$ are matched to the annual accounting variables in December of $t-1$. This is also the mechanism in which we realign data to construct cross-sectional equity characteristic data. For monthly rebalancing to construct the daily characteristic-managed portfolios, a similar scheme applies. That is, to construct the daily characteristic-managed portfolios in month $t+1$ based on equity s, returns at the daily frequency are matched with the normalized characteristics $z_{i,t}^s$ in month t and $z_{i,t}^s$ are used as the weights for constructing the daily characteristic-managed portfolios. Characteristics normalized as in (2.19) ensure the managed portfolios, to some extent, mimic the long-short trading strategies so that we can use the normalized characteristics as the weights for constructing portfolios. These normalized variables are then used to construct 99 characteristic-managed portfolios. Specifically, characteristic-managed portfolio s (or factor s) is given as

$$f_{s,t+1} = \sum_{i=1}^{n_t} z_{i,t}^s R_{i,t+1}^e, \tag{1.7}$$

where $R_{i,t+1}^e$ refers to the excess return of individual stock i. Monthly portfolios will be mainly used for comparison analysis such as calculating the IS Sharpe ratios and running univariate spanning regression; while daily managed portfolios will be used for calculating realized volatility for each month. More comprehensive descriptions of these anomaly variables are listed in the appendix along with acronyms used in our replication procedure. The corresponding studies, where these anomaly variables were initially proposed, are listed in the appendix as well.

Following the cutting-edge data cleaning technique, we can approximately construct 400 anomaly variables with approximately 153 of them are regarded as the representative factor-related variables. In another paper, Chen (2022) selects 123 anomaly variables from 1995 to 2020 to construct characteristic-managed portfolios by requiring that those selected anomaly variables should overall keep at least 80% of

the sample as observations and the corresponding observations with missing anomaly variables are directly discarded. However, in this chapter, we want to keep data as informative as possible about the cross-sectional information and hence select those anomaly variables **without missing observations** from 1996 to 2020, which finally shrinks the anomaly universe from 123 anomaly variables to 99 anomaly variables. We construct the 99 characteristic-managed portfolios (or equity strategies) used for analysis in the main context with this filtered anomaly universe.

1.4 Empirical Analysis

1.4.1 Direct comparison on anomaly augmented portfolios

Using 99 equity strategies based on the cross-sectional characteristics in the Chinese stock market, we make following comparison by calculating and comparing the in-sample mean of returns and Sharpe ratios for both the original anomaly long-short portfolios and the associated volatility-managed portfolios. The results are summarized as follows

[Place Figure 1.3 about here]

[Place Figure 1.4 about here]

To assign statistical meaning to the corresponding comparison, we use the method of Wright, Yam, and Yung (2014), which improves the procedure using the Sharpe ratios for comparing portfolio performance (see Jobson and Korkie, 1981; Lo, 2002; Ledoit and Wolf, 2008; Leung and Wong, 2008) by accommodating richer statistical properties of excess returns under more general assumptions. More technical details can either be referred via the original paper of Wright, Yam, and Yung (2014) or Pav (2021, 2022). We summarize the results as follows,

[Place Table 1.1 about here]

As we can see from Table 1.1, among all the 99 anomaly-based strategies we have checked for the Chinese stock market, the performances of 60 trading strategies are seemingly improved by readjusting the holding positions via the lagged volatility given that the in-sample absolute values of the Sharpe ratios of these 60 volatility-managed portfolios increase ($\Delta SR > 0$) in comparison to those of original anomaly portfolios. However, for these 60 anomaly portfolios whose performance can be seemingly improved by scaling the holding positions using lagged volatility, only 11 of them enjoys statistically significant improvements in the Sharpe ratios (based on the methodology in Wright, Yam, and Yung, 2014). By contrast, the remaining 39 anomaly portfolios cannot directly enjoy improvements in the Sharpe ratios ($\Delta SR < 0$) via managing lagged volatility. Besides, for these 39 anomaly portfolios, we can only see statistically different performance differences (based on the increments in the absolute value of the Sharpe ratios) between the original anomaly portfolios and the volatility-managed ones. Finally, we summarize $14(= 11 + 3)$ anomalies for which the original anomaly portfolios or volatility-managed ones witness statistically significant differences in the absolute value of the Sharpe ratios in the following table.

[Place Table 1.2 about here]

1.4.2 Spanning regression approach for comparison

The empirical methodology exploited in Moreira and Muir (2017) is based on following time-series regression of the volatility-managed portfolio on the original factors,

$$f^{\sigma}_{t+1} = \alpha + \beta f_{t+1} + \epsilon_{t+1}. \tag{1.8}$$

Regression (1.8) is a straightforward empirical methodology with the empirical implication as follows: a positive intercept (α) implies that volatility timing increases

the Sharpe ratios relative to the original factors (Moreira and Muir, 2017). However, the increment in the Sharpe ratios suggested from spanning regression must correspond to a new portfolio combining both the volatility-managed portfolio and un-managed portfolio. Alpha alone does not necessarily imply the increment in the Sharpe ratio of f^{σ}_{t+1} in direct comparison to f_{t+1}. This viewpoint involves the following discussion about the connection between alpha and the Sharpe ratio of a single unscaled portfolio (i.e. f_{t+1} alone) and the connection between alpha and the Sharpe ratio of the augmented portfolio that combines both the scaled portfolio and the unscaled portfolio (i.e. f^{σ}_{t+1} and f_{t+1}).

Cederburg, O'Doherty, Wang, and Yan (2020) hold the opinion that a positive alpha in (1.8) is a lower bar for declaring success of managed portfolio relative to the Sharpe ratio difference. Recall our main target for comparison: managed portfolio f^{σ}_{t+1} and the original anomaly portfolio f_{t+1}. Intuitively, this can be interpreted as follows: a significant (positive) alpha in (1.8) only requires that $\bar{f}^{\sigma}_{t+1} > \hat{\beta}\bar{f}_{t+1}$, where $\bar{f}^{\sigma}_{t+1}$ and $\bar{f}_{t+1}$ refer to sample time-series mean of volatility-managed portfolios and sample time-series mean of original volatility portfolios respectively; $\hat{\beta}$ refers to estimation of correlation coefficient between f^{σ}_{t+1} and f_{t+1} by running OLS using (1.8). However, this requirement is not enough for guaranteeing $|\bar{f}^{\sigma}_{t+1}| > |\bar{f}_{t+1}|$, which is essentially the requirement for having an improved IS Sharpe ratio by using volatility to scale the original anomaly portfolios.[4] Specifically, suppose we obtain $\hat{\beta}$ from running spanning regression in (1.8) as $\hat{\beta} = 0.7$ while at the same time $\bar{f}^{\sigma}_{t+1} = 0.9 \times \bar{f}_{t+1}$, which suggests that

$$\hat{\alpha} = \bar{f}^{\sigma}_{t+1} - \hat{\beta}\bar{f}_{t+1} = 0.9 \times \bar{f}_{t+1} - 0.7 \times \bar{f}_{t+1} = 0.2 \times \bar{f}_{t+1},$$

then the volatility-managed portfolio still fails to generate IS increment in the Sharpe

[4]This is because by construction f^{σ}_{t+1} and f_{t+1} share the same unconditional (sample) standard deviation.

ratio. Besides, note that

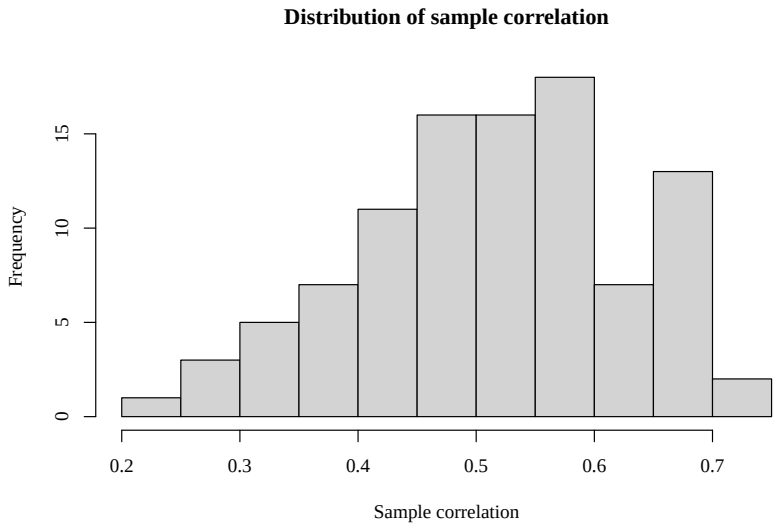
$$\hat{\beta} = \frac{\sum_{t+1}\left(f^{\sigma}_{t+1} - \bar{f}^{\sigma}_{t+1}\right)\left(f_{t+1} - \bar{f}_{t+1}\right)}{\sum_{t+1}\left(f_{t+1} - \bar{f}_{t+1}\right)^2},$$

$$\hat{\alpha} = \bar{f}^{\sigma}_{t+1} - \hat{\beta}\bar{f}_{t+1},$$

and that

$$\hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} = \frac{\sum_{t+1}\left(f^{\sigma}_{t+1} - \bar{f}^{\sigma}_{t+1}\right)\left(f_{t+1} - \bar{f}_{t+1}\right)}{\sqrt{\sum_{t+1}\left(f^{\sigma}_{t+1} - \bar{f}^{\sigma}_{t+1}\right)^2}\sqrt{\sum_{t+1}\left(f_{t+1} - \bar{f}_{t+1}\right)^2}},$$

where $\hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}}$ denotes the sample correlation between f^{σ}_{t+1} and f_{t+1}. Since, by construction, f^{σ}_{t+1} and f_{t+1} have the same sample correlation, $\hat{\beta} = \hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}}$. We calculate all the sample correlations between f^{σ}_{t+1} and f_{t+1} for the 99 anomaly portfolios and summarize the distribution of the sample correlations as follows. We can



see from the figure that for the 99 equity anomaly portfolios, the sample correlations between the original ones and the volatility-managed ones range approximately from 0.21 to 0.72, which suggests obtaining statistically significant alpha from spanning regression and having a relatively low absolute value for the IS Sharpe ratios of volatility-managed portfolios is possible.

A statistically significant alpha indicates that the optimal ex post combination of scaled and unscaled factors expands the mean-variance frontier relative to the original factor. This combination strategy allows investors to allocate wealth both in the volatility-managed portfolios and the original anomaly portfolios. Gibbons, Ross, and Shanken (1989) and Barillas and Shanken (2018) link the intercept (alpha) with the Sharpe ratio by taking the ratio of the estimated alpha to the standard error in linear regression (i.e. the so-called appraisal ratio, $AR = \hat{\alpha}/\hat{\sigma}_\epsilon$) and show that the appraisal ratio can be used to characterize the extent to which the augmented portfolios can increase the slope of the mean-variance frontier. This argument can be directly applied in the volatility-managed portfolio setting for discussing the connection between statistically significant alpha and the performance gain measured as the increment in the Sharpe ratio that is obtained from the combination strategy, as noted in Cederburg, O'Doherty, Wang, and Yan (2020). Specifically, for the investor who has the access to the risk-free security, under the standard optimal portfolio allocation theory as in Markowitz (1952), his optimal ex post allocation rule is proportional to $\hat{\Sigma}^{-1}\hat{\mu}$, where $\hat{\Sigma}$ is 2×2 matrix as the sample variance-covariance matrix of $\left[f^{\sigma}_{t+1}, f_{t+1}\right]^{\top}$ and $\hat{\mu}$ is a 2×1 vector with each entry denoting the time-series sample mean of f^{σ}_{t+1} and f_{t+1} respectively, thus $\hat{\mu} = \left[\bar{f}^{\sigma}_{t+1}, \bar{f}_{t+1}\right]^{\top}$. Since, by construction, f^{σ}_{t+1} and f_{t+1} have the same sample standard deviation, we show that

$$\begin{aligned}
\hat{\Sigma} &= \begin{bmatrix} \hat{\sigma}^2(f) & \hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} \sigma_{f^{\sigma}_{t+1}} \sigma_{f_{t+1}} \\ \hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} \sigma_{f^{\sigma}_{t+1}} \sigma_{f_{t+1}} & \hat{\sigma}^2(f) \end{bmatrix} \\
&= \hat{\sigma}^2(f) \begin{bmatrix} 1 & \hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} \\ \hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} & 1 \end{bmatrix},
\end{aligned}$$

and correspondingly

$$\hat{\Sigma}^{-1} = \left[\hat{\sigma}^2(f)\left(1 - \hat{\rho}^2_{f^{\sigma}_{t+1}, f_{t+1}}\right)\right]^{-1} \begin{bmatrix} 1 & -\hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} \\ -\hat{\rho}_{f^{\sigma}_{t+1}, f_{t+1}} & 1 \end{bmatrix}.$$

Then we have the optimal ex post allocation rule associated with the volatility-managed portfolio is

$$x_{\sigma}^{*}=\frac{\bar{f}_{t+1}^{\sigma}-\hat{\rho}_{f_{t+1}^{\sigma},f_{t+1}}\bar{f}_{t+1}}{\hat{\sigma}^{2}(f)\left(1-\hat{\rho}_{f_{t+1}^{\sigma},f_{t+1}}^{2}\right)}=\frac{\hat{\alpha}}{\hat{\sigma}^{2}(f)\left(1-\hat{\rho}_{f_{t+1}^{\sigma},f_{t+1}}^{2}\right)}. \tag{1.9}$$

Equation (1.9) has the direct implication that $\hat{\alpha}$ obtained from spanning regression determines the wealth allocated to the scaled portfolios. Besides,

$$AR^{2}=SR^{2}\left(f^{\sigma},f\right)-SR^{2}\left(f\right), \tag{1.10}$$

where $SR^{2}\left(f^{\sigma},f\right)$ refers to the IS squared Sharpe ratio of combination strategy comprising both f^{σ} (managed portfolio) and f (original portfolio).

We run time-series spanning regression of the form in (1.8) and report both the estimated coefficients and the associated Newey and West (1987) t-statistics with three lags (Kelly, Moskowitz, and Pruitt, 2021). We summarize the results from univariate spanning regression in Table 1.3 and more detailed spanning regression estimation results in Table 1.4 for those anomaly portfolios with significant estimated alpha.

[Place Table 1.3 about here]

[Place Table 1.4 about here]

Given the results summarized in Table 1.3 and Table 1.4, we find that among all the 99 anomaly portfolios, 71 volatility-managed portfolios have a positive estimate of alpha in univariate spanning regression while the remaining 28 volatility-managed portfolios have a negative estimate of alpha in univariate spanning regression. However, since all the original anomaly portfolios are constructed as the long-short portfolios based on the univariate sort on the associated equity characteristic, the sign associates with the negative alpha can readily shifted to positive by taking revere holding positions. In other words, the main implication of spanning regression is whether an investors can obtain an increment in the Sharpe ratio by combining the

volatility-managed portfolios and the original anomaly portfolios. This increment can be reflected in the appraisal ratio associated with alpha. Accordingly, whether the estimated alpha in the univariate spanning regression is statistically significant or not matters more for evaluating the corresponding performance gain. For this purpose, we see from Table 1.3 and Table 1.4 that among all the 99 anomaly portfolios we investigate, only 24 of them generate statistically significant alpha in the univariate spanning regression. This low ratio ($24/99 \approx 24\%$) suggests that scaling the holding positions of the original portfolios using the lagged volatility is not a successful strategy for improving the portfolio performance.

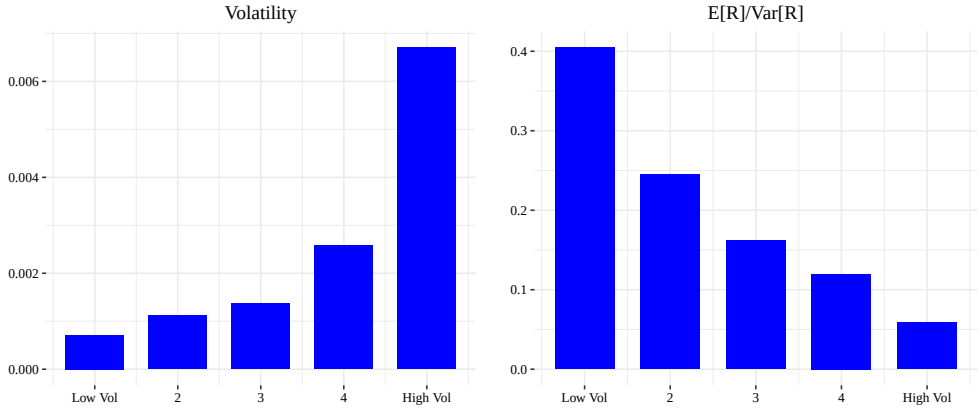
1.5 Conclusion

This chapter examines the performance of volatility-managed portfolios in the Chinese stock market. Using the standard empirical methods to collect, clean, and construct data from the Chinese stock market, we apply the standard empirical strategies to investigate whether an investor can adjust the holding positions of portfolios based on volatility to improve the performance of the original anomaly portfolios in the Chinese stock market. Our empirical results are similar to those in Cederburg, O'Doherty, Wang, and Yan (2020) for the U.S. equity market . That is, the performance of volatility-managed portfolios degrades within a broad sample of anomaly portfolios (103 trading strategies in the U.S. equity market). Based on our analysis of the Chinese stock market using 99 equity trading strategies, we also find that there exists no desired performance gain systematically by scaling anomaly portfolios using the lagged volatility as suggested in Moreira and Muir (2017).

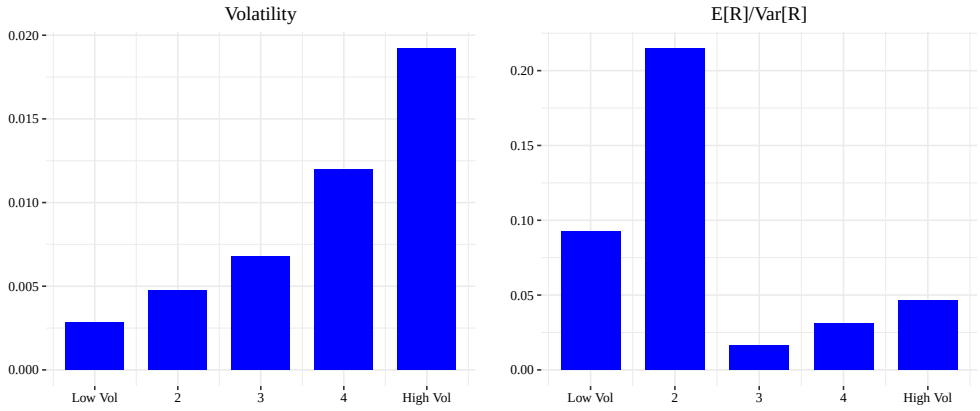
Figures and Tables

Figure 1.1

(a) The U.S. Market



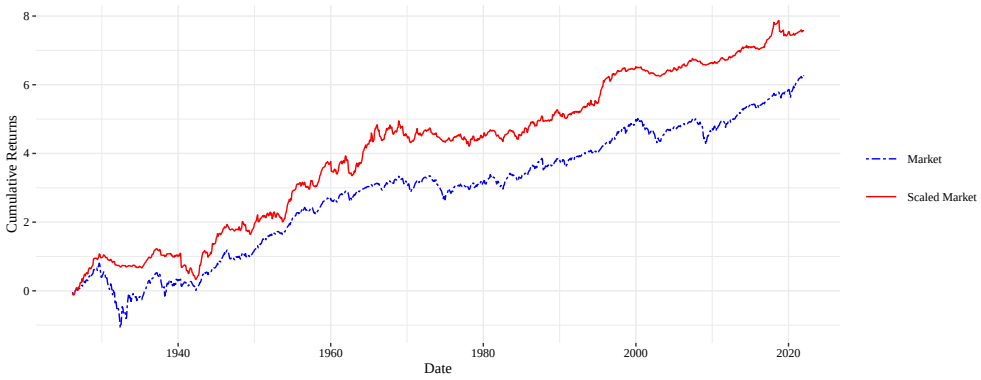
(b) The Chinese Market



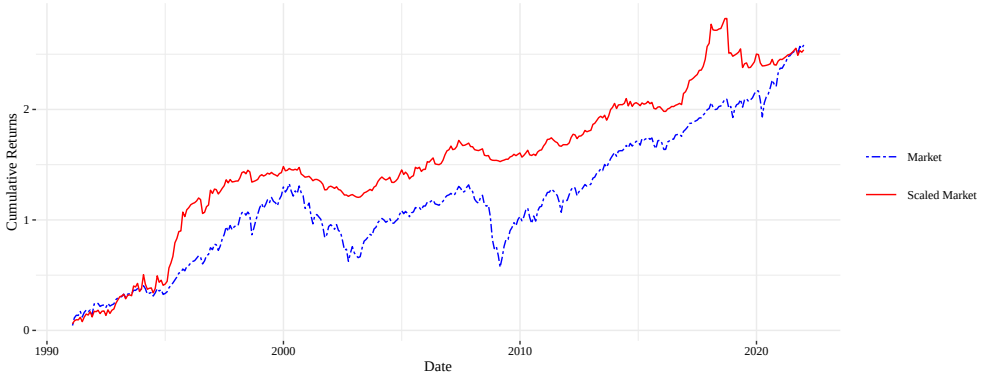
Note: In the figure above, we demonstrate results generated from sorting on the previous month's volatility both for the U.S. market (a) and the Chinese stock market (b). Specifically, we use (1.3) to calculate the realized volatility for each month. With the monthly time series of realized volatility, we sort all the months into five buckets based on realized volatility of the previous month. Then for each bucket, we calculate the average volatility (on the left for each panel) and the ratio of the average return over the average volatility as the mean-variance trade-off (on the right for each panel).

Figure 1.2

(a) The U.S. Market



(b) The U.S. Market from 1991



(c) The Chinese Market

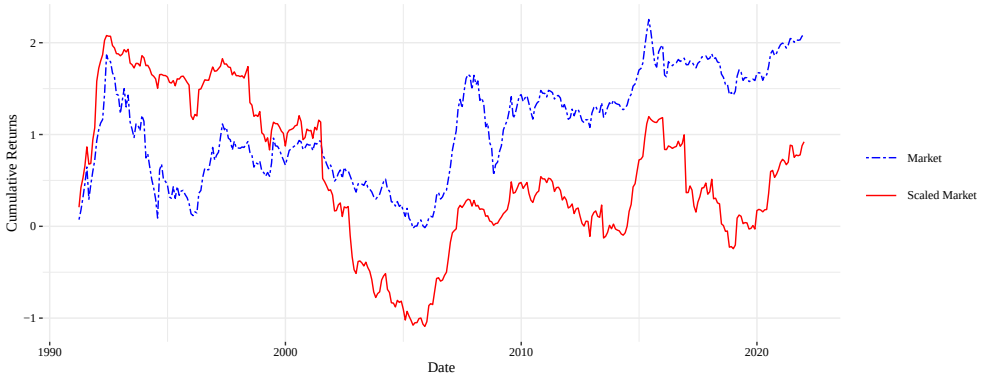
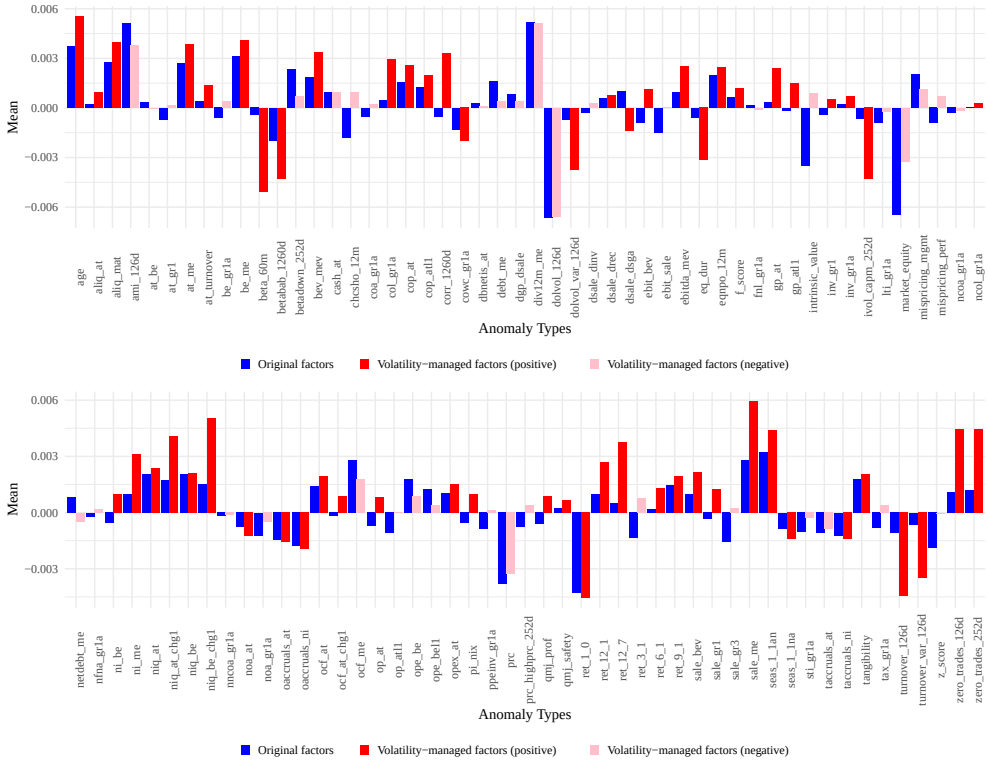
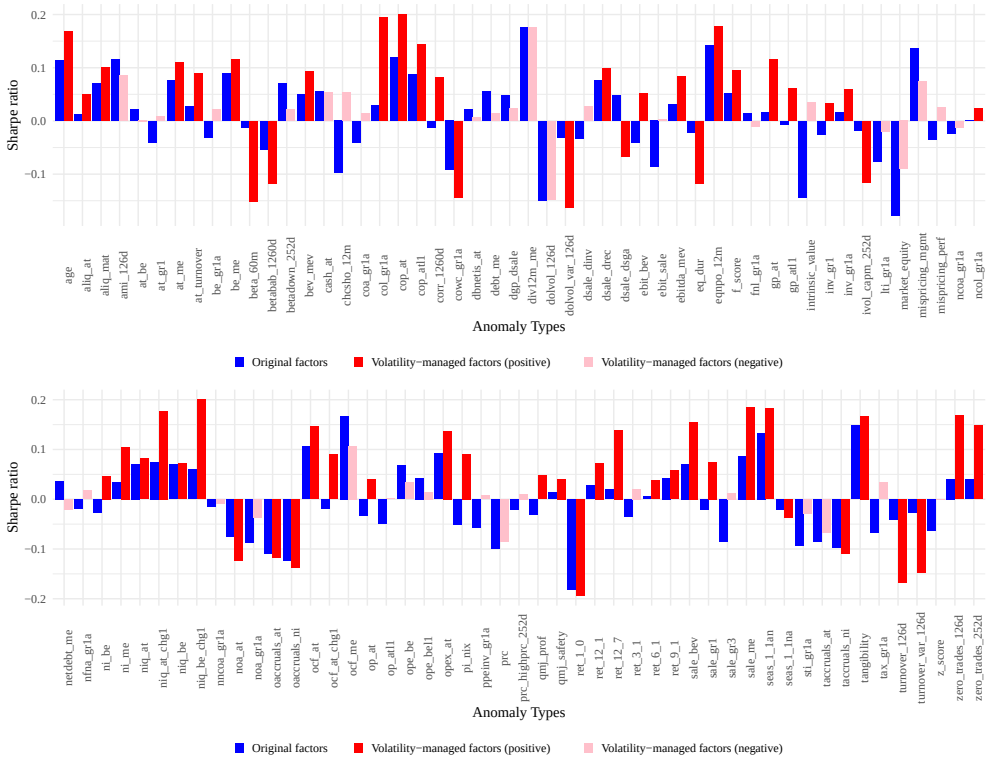


Figure 1.3



Note: In the figure above, we summarize the results of comparing IS (in-sample) mean of original anomaly portfolios and the associated volatility-managed portfolios. Both the original anomaly portfolios and the volatility-managed portfolios are at monthly frequencies and data spans from January 1996 to December 2020 (i.e. 25 years in total). As we have discussed in the main context, we use f_{t+1} to denote the original portfolio in month $t+1$ and $f^{\sigma}_{t+1} = \frac{c}{\hat{\sigma}^2_t(f)} f_{t+1}$ to denote the volatility-managed portfolio in month $t+1$. c is a constant chosen so that f_{t+1} and f^{σ}_{t+1} have the same sample unconditional standard deviation over the full sample period. There are three kinds of bars in this figure: the blue bars indicate the original factors (portfolios), the red bars indicate the volatility-managed factors (portfolios) that exhibit **larger** absolute value of IS mean in comparison to the corresponding original factors (portfolios), and the pink bars indicate the volatility-managed factors that exhibit **smaller** absolute value of IS mean in comparison to the corresponding original factors (portfolios).

Figure 1.4



Note: In the figure above, we summarize the results of comparing the IS (in-sample) Sharpe ratio of the original anomaly portfolios and the associated volatility-managed portfolios. Both the original anomaly portfolios and the volatility-managed portfolios are at monthly frequencies and data spans from January 1996 to December 2020 (i.e. 25 years in total). As we have discussed in the main context, we use f_{t+1} to denote the original portfolio in month $t+1$ and $f^{\sigma}_{t+1} = \frac{c}{\hat{\sigma}^2_t(f)} f_{t+1}$ to denote the volatility-managed portfolio in month $t+1$. c is a constant chosen so that f_{t+1} and f^{σ}_{t+1} have the same sample unconditional standard deviation over the full sample period. There are three kinds of bars in this figure: the blue bars indicate the original factors (portfolios), the red bars indicate the volatility-managed factors (portfolios) that exhibit **larger** absolute value of IS Sharpe ratio in comparison to the corresponding original factors (portfolios), and the pink bars indicate the volatility-managed factors that exhibit **smaller** absolute value of IS Sharpe ratio in comparison to the corresponding original factors (portfolios).

Table 1.1

	Sharpe ratio difference		
Sample	Total	$\Delta SR > 0$ [Signif.]	$\Delta SR < 0$ [Signif.]
All trading strategies	99	60[11]	39[3]

Note: In the table above, ΔSR refers to the difference between absolute value of the Sharpe ratios associated with the original anomaly portfolios (f_{t+1}) and the volatility-managed portfolios (f^{σ}_{t+1}). We demonstrate the number of the absolute value of the Sharpe ratios differences that are positive, negative and significant at 5% level (in square brackets).

Table 1.2

Anomaly Types	ΔSR	p-value
Market beta **[Low Risk]**	0.1382	0.0232
Net stock issues **[Value]**	-0.0449	0.0029
Change in current liabilities **[Investment]**	0.1660	0.0081
Coefficient of variation for dollar trading volume **[Profitability]**	0.1330	0.0176
Return on net operating assets **[Profitability]**	0.0097	0.0354
Profit margin **[Profit Growth]**	-0.0834	0.0395
Gross profits-to-assets **[Quality]**	0.0986	0.0358
Intrinsic-value **[Value]**	-0.1079	0.0003
Change in quarterly return on equity **[Profit Growth]**	0.1408	0.0230
Taxable income-to-book income **[Seasonality]**	0.0393	0.0105
Price momentum $t-12$ to $t-7$ **[Momentum]**	0.1196	0.0115
Share turnover **[Low Risk]**	0.1269	0.0179
Coefficient of variation for share turnover **[Profitability]**	0.1216	0.0478
Number of zero trades with turnover as tiebreaker (6 months) **[Low Risk]**	0.1273	0.0179

Table 1.3

		Univariate spanning regression	
Sample	Total	$\alpha > 0$ [Signif.]	$\alpha < 0$ [Signif.]
All trading strategies	99	71[16]	28[8]

Note: This table summarizes results from spanning regressions for 99 anomaly trading strategies in the Chinese stock market. The spanning regression is the one that we have discussed in the main context, given by $f^{\sigma}_{t+1} = \alpha + \beta f_{t+1} + \epsilon_{t+1}$, where $f^{\sigma}_{t+1}(f_{t+1})$ is the monthly return for the volatility-managed (original) portfolio. For each regression, this table reports the number of alphas that are positive, positive and significant approximately at 2.5% level (i.e. we set critical value for the corresponding t-stat as 2), negative, and negative and significant at the 2.5% level. We assess the statistical significance of alpha using Newey and West (1987) adjusted standard errors.

Table 1.4

Anomaly Types	$\hat{\alpha}$	t-stat($\hat{\alpha}$)	$\hat{\beta}$	t-stat($\hat{\beta}$)	R^2	AR^2
Firm age **[Low Leverage]**	0.0043	2.4911	0.3260	5.6823	0.1033	0.0193
Market beta **[Low Risk]**	-0.0049	-2.7247	0.3948	3.4379	0.1530	0.0252
Frazzini-Pedersen market beta **[Low Risk]**	-0.0033	-2.0269	0.4978	4.2988	0.2453	0.0109
Change in current liabilities **[Investment]**	0.0027	3.2492	0.3859	2.8768	0.1461	0.0398
Cash-based operating profits-to-book assets **[Quality]**	0.0017	2.6395	0.5866	6.6950	0.3419	0.0262
Change in current operating working capital **[Accruals]**	-0.0015	-2.2617	0.4246	2.9854	0.1775	0.0133
Dividend yield **[Value]**	0.0029	2.0264	0.4342	4.6806	0.1858	0.0121
Dollar trading volume **[Size]**	-0.0034	-2.8954	0.4768	3.9751	0.2248	0.0287
Return on net operating assets **[Profitability]**	0.0018	2.0774	0.7038	8.1265	0.4937	0.0126
Equity duration **[Value]**	-0.0029	-2.0033	0.4429	4.0892	0.1935	0.0146
Equity net payout **[Value]**	0.0014	2.2130	0.5512	5.4834	0.3015	0.0140
Gross profits-to-assets **[Quality]**	0.0022	2.5263	0.6172	7.1965	0.3788	0.0178
Idiosyncratic volatility from the CAPM (252 days) **[Low Risk]**	-0.0039	-2.4611	0.5827	7.1438	0.3373	0.0170
Change in quarterly return on equity **[Profit Growth]**	0.0036	2.9183	0.2465	2.9236	0.0576	0.0268
Change in quarterly return on equity **[Profit Growth]**	0.0046	3.3454	0.2598	2.8428	0.0644	0.0367
Taxable income-to-book income **[Seasonality]**	0.0013	2.5581	0.5083	6.9227	0.2559	0.0186
Price momentum $t-12$ to $t-7$ **[Profit Growth]**	0.0035	2.8083	0.6086	5.5528	0.3683	0.0256
Asset turnover **[Quality]**	0.0016	2.5226	0.5651	4.7138	0.3171	0.0197
Sale to market **[Value]**	0.0050	2.8300	0.3464	3.5571	0.1171	0.0273
Year 1-lagged return, annual **[Profit Growth]**	0.0029	2.4769	0.4581	3.5126	0.2072	0.0186
Share turnover **[Low Risk]**	-0.0039	-3.1728	0.5469	3.9528	0.2968	0.0303
Coefficient of variation for share turnover **[Profitability]**	-0.0032	-2.5662	0.4076	3.3853	0.1634	0.0227
Number of zero trades (6 months) **[Low Risk]**	0.0039	3.1689	0.5451	3.9584	0.2948	0.0303
Number of zero trades (12 months) **[Low Risk]**	0.0038	2.6714	0.5318	4.2990	0.2804	0.0224

Note: This table summarizes detailed estimation results from univariate spanning regression for anomaly portfolios with statistically significant alpha. R^2 refers to the adjusted R-square as the measure of IS regression fitting. AR^2 refers to the squared appraisal ratio with $AR = \hat{\alpha}/\hat{\sigma}_{\epsilon}$ and $\hat{\sigma}_{\epsilon}$ refers to the standard deviation of residuals in univariate spanning regression.

Chapter 2

Sparse Structure of Stochastic Discount Factor in the Chinese Stock Market: A Bayesian Interpretable Machine-learning Approach

2.1 Introduction

> *One of our central themes is that if assets are priced rationally, variables that are related to average returns, such as size and book-to-market equity, must proxy for sensitivity to common (shared and thus and undiversifiable) risk factors in returns.*
>
> Fama and French (1993)

> *We have a lot of questions to answer: First, which characteristics really provide independent information about average returns? Which are subsumed by others? Second, does each new anomaly variable also correspond to a new factor formed on those same anomalies? ... Third,*

how many of these new factors are really important?

Cochrane (2011)

As the two quotes cited above suggest, a formidable challenge faced within the community of financial researchers is how to handle the high-dimensionality in the potential predictors for the expected return. There are at least two difficulties associated with high-dimensionality in the potential predictors. First, whether or not there exists a sparse exposure structure of stochastic discount factor (SDF) is difficult to know. Second, what should be a reasonable functional relationship between the expected return and intrinsically useful predictors.

The first difficulty has attracted a great deal of attentions in recent years, given that a huge number of firm-level characteristics have been proposed to be the predictors in the literature. Many studies rely on the p-value of the standard test statistics (such as the t statistic) as the evidence to support or be against the use of firm-level characteristics. However, Harvey, Liu, and Zhu (2016) points out the so-called p-hacking issue in the conventional statistical test. They further propose an adjusted p-value to check the statistical evidence of the usefulness of firm-level characteristics.

To deal with the second difficulty, one way is to allow for nonlinear relationships between the expected return and predictors in the model specification. This is the exact reason why nonparametric methods have becomes increasingly popular in this literature. With the development of modern computational power and statistical algorithms, some advanced nonparametric methods have been proposed (see Freybergerk, Neuhierl, and Weber, 2019). Machine-learning methods are one of the popular nonparametric techniques.

Studies that employ machine-learning methods to study return predictability can be divided into three groups. The first group of studies aims to use and design machine-learning methods to generate good out-of-sample performance. These methods usually are flexible given the generic nonparametric feature in the methods. Gu, Kelly, and Xiu (2020) compare many machine-learning methods in terms of

their predictive power of the U.S. equity returns. It is found that neural network and regression trees perform relatively well. Other studies that use machine learning method to analyse cross-sectional returns include but not restricted to (Freybergerk, Neuhierl, and Weber, 2019; Chinco, Clark-Joseph, and Ye, 2019; Han, He, Rapach, and Zhou, 2019; Chen, Pelger, and Zhu, 2019).

The second group of studies assume that there exists a factor structure in the potentially useful predictors. The number of factors is usually much lower than the number of available characteristics. This approach has been an important part of the literature ever since the seminar works of Fama and French (1992, 1993, 1996). Generally there are two alternative ways to introduce a factor structure in this rather extensive literature. The first one uses pre-specified and observed factors based on the prior knowledge about the cross-sectional accounting information. Many factors have been established for explaining the cross-sectional variations associated with asset returns; see, for example, Fama and French (1993), Fama and French (2015), Hou, Xue, and Zhang (2015). More references can be found in two recent excellent surveys, that is, Hou, Xue, and Zhang (2018) and Chen and Zimmermann (2020). The second one assumes that factors are latent variables. In this case, statistical factor analysis techniques, such as principal component analysis (PCA), are used to extract factors and factor loadings simultaneously. Studies of this kind can be traced back at least to Connor and Korajczyk (1986) and Chamberlain and Rothschild (1983). Recently, the latent factor approach has been employed in Fan, Liao, and Wang (2016); Kozak, Nagel, and Santosh (2018); Kelly, Pruitt, and Su (2019); Kozak (2020); Lettu and Pelger (2020a,b) to study stock returns.

The third group of studies focuses directly on addressing the high-dimension problem. These studies use model selection and variable selection techniques to select useful firm-level characteristics. Since many machine-learning methodologies inherit ideas from statistical theory (Vapnik, 1998; Hastie, Tibshirani, and Friedman, 2001; Catoni, 2004, 2007), some machine-learning methodologies designed for handling the high-dimension problem are essentially statistical learning methods.

Among all the methods of this type, the most representative ones are LASSO, ridge regression and elastic net. These are also the major statistical learning methods mostly applied in economics and finance literature: Rapach, Strauss, and Zhou (2010), Messmer and Audrino (2017), Giannone, Lenza, and Primiceri (2021) and Bakalli, Guerrier, and Scaillet (2021) discuss how LASSO is applied in selecting useful predictors for making predictions in economics and finance when there are a slew of predictors; Gabauer, Gupta, Marfatia, and Miller (2020) establishes a high-dimensional vector autoregressive model with L^2-penalty (i.e. it essentially belongs to ridge regression) for estimating price network connectedness in the U.S. housing market; Kim and Swanson (2014) provides empirical evidence of how elastic net is applied in the high-dimension problem setting for forecasting financial and macroeconomic variables. Huang, Li, and Wang (2021) applies elastic net as one intermediate step for aggregating cross-sectional information of equities to construct the disagreement index in the U.S market. Bali, Goyal, Huang, Jiang, and Wen (2021) discusses the application of elastic net in addressing bond return predictability in the setting where the individual bond is cross-sectionally exposed to the high-dimensional information vector. It is known in literature (Zou and Hastie, 2005) that in comparison to LASSO as the dimension reduction method for variable selection, elastic net combines LASSO and ridge penalties and produces a model with more flexibilities and good out-of-sample prediction accuracy.

A useful addition to this group of studies is Linero (2018) where a Bayesian additive regression trees (BART) method is found to perform well. In the BART method of Linero (2018), a sparsity-inducing Dirichlet hyper-prior is used to solve the high-dimension problem. This method is empirically successful in selecting relevant variables that can yield good out-of-sample prediction accuracy, and is later theoretically justified by Ročková (2019), Ročková and Saha (2019), and Ročková and van der Pas (2019). However, it naturally inherits the major disadvantage of Chipman, George, and McCulloch (2010), which is computationally heavy in comparison to LASSO, ridge regression, or elastic net, mainly due to its underlying

MCMC sampling scheme. Besides, for all the existing machine-learning methodologies, rarely is there any discussion on whether or not these methodologies can be interpreted through the lens of existing economic theories.

Another Bayesian method to address the high-dimension problem is the Bayesian interpretable machine-learning method proposed in Kozak, Nagel, and Santosh (2020). There are a number of good features in this methods. First, the modelling framework is parsimonious but still powerful in characterizing the key asset-pricing structure. Second, it reconciles well with economic theory through the Bayesian lens (that is the reason why it is referred to as a Bayesian interpretable method) as well as with the statistical learning theory (which facilitates implementation and computation). Basically, by imposing an economically motivated prior on SDF, it is possible to show how the machine-learning methods (specifically, the penalized regression such as the ridge regression with the objective function being the Hansen-Jagannathan distance or the elastic net method with dual penalty) are related to the SDF-based asset pricing theory. Because of these attractive features, in this chapter, we apply it to analyze the returns of the Chinese stock market. In particular, we use the method to check whether or not there exists a sparse exposure structure of SDF to several dominant cross-sectional equity characteristics in the Chinese stock market.

The rest of this chapter is structured as follows: In Section 2.2, we review the theoretical modelling framework for the SDF-based linear asset pricing theory and explain how the economic theory is related to some of the machine-learning methods so that the machine-learning methods are interpretable through the Bayesian lens. In Section 2.3, we discuss how cross-sectional anomaly variables (or equivalently firm-level characteristics) are constructed. In Section 2.4, we report the main empirical findings. Finally Section 2.5 concludes this chapter.

2.2 Basic Modelling Framework

2.2.1 SDF and cross-sectional asset pricing

In much of the finance literature, the central goal is to explain the differences in returns in the cross-sectional dimension. Specifically for individual stock, denote $R_{t+1,i}$ as the return of asset i at $t+1$. The fundamental no-arbitrage condition is closely related to the existence of SDF derived from the first-order condition of the Euler equation. That is, for any return in excess of the risk-free rate $R^e_{t+1,i} = R_{t+1,i} - R^f_{t+1}$, the following key pricing formula (conditional) holds

$$\mathbb{E}_t \left[M_{t+1} R^e_{t+1,i} \right] = 0. \tag{2.1}$$

Following the convention in the literature (see Hansen and Jagannanthan, 1991; Haugen and Baker, 1996) and without loss of generality, we can assume that the SDF is of a linear functional form as

$$M_{t+1} = 1 - \omega_t^\top \left(R^e_{t+1} - \mathbb{E}_t R^e_{t+1} \right),$$

where ω_t is a $N \times 1$ vector of SDF coefficients with N being the number of firms cross-sectionally.[1] This specification implies that we normalize the excess return by the corresponding conditional mean, $\mathbb{E}_t R^e_{t+1}$.

To see how it is connected with the factor-modeling framework, considering the following construction,

$$\omega_t = Z_t \omega, \tag{2.2}$$

where Z_t is an $N \times L$ matrix of asset characteristics and ω is an $L \times 1$ vector of time-invariant coefficients. Usually the entries of matrix Z_t in (2.2) collects the information of firm-level characteristics (specifically, each row i of Z_t collects the

[1] As pointed in Kozak, Nagel, and Santosh (2018), the ground for a linear factor-based representation of SDF is essentially the law of one prices (LOP). As long as LOP holds, the factors used to represent SDF are a linear combination of asset payoffs.

characteristic information of firm i at time t). As documented in empirical asset pricing literature, usually researchers search for new measurable asset characteristics that approximately span ω_t. For example, Fama and French (1993) use two characteristics, market capitalization and the book-to-market equity ratio. Similarly, by plugging this equation into the fundamental pricing equation (2.1), we have

$$\begin{aligned} M_{t+1} &= 1 - \omega_t^\top \left(R_{t+1}^e - \mathbb{E}_t R_{t+1}^e\right) \\ &= 1 - \omega^\top Z_t^\top \left(R_{t+1}^e - \mathbb{E}_t R_{t+1}^e\right). \end{aligned}$$

We can then define L multi-factors as

$$F_{t+1} = Z_t^\top R_{t+1}^e, \tag{2.3}$$

which simply leads to the normalized representation of SDF as following

$$\begin{aligned} M_{t+1} &= 1 - \omega^\top \left(F_{t+1} - Z_t^\top \mathbb{E}_t R_{t+1}^e\right) \\ &= 1 - \omega^\top \left(F_{t+1} - \mathbb{E}_t F_{t+1}\right). \end{aligned} \tag{2.4}$$

Note that F_{t+1} is essentially assets in a portfolio form. Hence, it is possible to plug it into the key pricing formula as in (2.1). Without loss of generality we can replace the conditional mean of factors, $\mathbb{E}_t F_{t+1}$ with the unconditional mean $\mathbb{E}F_{t+1}$ (i.e. $M_{t+1} = 1 - \omega^\top \left(F_{t+1} - \mathbb{E}F_{t+1}\right)$). We can then have following unconditional pricing formula for the managed portfolios,

$$\mathbb{E}_t \left[M_{t+1} F_{t+1}^\top\right] = 0 \;\Rightarrow\; \mathbb{E}\left[M_{t+1} F_{t+1}^\top\right] = 0,$$

which implies that

$$\begin{aligned} &\mathbb{E}F_{t+1}^\top - \omega^\top \mathbb{E}\left[\left(F_{t+1} - \mathbb{E}F_{t+1}\right) F_{t+1}^\top\right] \\ =\; &\mathbb{E}F_{t+1}^\top - \omega^\top \mathbb{E}\left[\left(F_{t+1} - \mathbb{E}F_{t+1}\right)\left(F_{t+1} - \mathbb{E}F_{t+1}\right)^\top\right] \;=\; 0. \end{aligned}$$

Hence we have

$$\mathbb{E}F_{t+1} \;=\; \mathbb{E}\left[\left(F_{t+1}-\mathbb{E}F_{t+1}\right)\left(F_{t+1}-\mathbb{E}F_{t+1}\right)^{\top}\right]\omega. \tag{2.5}$$

This constant specification imposed on the managed portfolio processes implicitly suggests that we focus on unconditional asset pricing. It brings convenience using the corresponding sample moment over the time-series dimension to estimate $\mathbb{E}F_{t+1}$ and $\mathbb{E}\left[\left(F_{t+1}-\mathbb{E}F_{t+1}\right)\left(F_{t+1}-\mathbb{E}F_{t+1}\right)^{\top}\right]$, denoted by μ (managed portfolio's time-series mean) and Σ (variance-covariance matrix), respectively, for the following discussion. It will be seen in the following discussion that, the time-series analogue of the managed portfolios $\bar{\mu}$ and $\bar{\Sigma}$ can be regarded as the data used for constructing the posterior to update the prior information. This constant specification reconciles well with the empirical Bayes logic. See the corresponding discussion in **Remark** 2.3.

2.2.2 Interpretation from a Bayesian perspective

In this section, we discuss how SDF is connected with penalized cross-sectional regression from a Bayesian perspective. The discussions follow the main ideas from Kozak, Nagel, and Santosh (2020) but are more detailed than those in Kozak, Nagel, and Santosh (2020). Essentially the Bayesian prior structure is imposed on μ as follows (assuming Σ is known, and we will discuss how to obtain Σ in **Remark** 2.3).

$$\mu \sim \mathcal{N}\left(0, \frac{\kappa^2}{\tau}\Sigma^{\eta}\right), \tag{2.6}$$

where

$$\tau = \mathrm{Tr}\left[\Sigma\right],$$

and κ, η are tuning parameters to be discussed later.

Remark 2.1 *μ is $L \times 1$ vector that collects the expected return of each managed portfolio over the time-series dimension. The cross-sectional heterogeneity is captured by the prior (2.6). Thus, the prior captures investors' ex-ante belief about the*

expected return of individual managed portfolio. Integrating μ out of $\mu^\top \Sigma^{-1} \mu$ (the squared Sharpe ratio) (i.e., integrating out the ex-ante uncertainty associated with μ) yields the root expected Sharpe ratio under the prior distribution,

$$\begin{aligned}
& \mathbb{E}\left[\mu^\top \Sigma^{-1} \mu\right]^{1/2} \\
= \; & \mathbb{E}\left[\Sigma^{-1} \operatorname{Tr}\left(\mu\mu^\top\right)\right]^{1/2} \\
= \; & \left\{\Sigma^{-1} \operatorname{Tr}\left(\mathbb{E}\left[\mu\mu^\top\right]\right)\right\}^{1/2} \\
= \; & \operatorname{Tr}\left(\Sigma^{-1} \frac{\kappa^2}{\tau} \Sigma^{\eta}\right)^{1/2} \\
= \; & \left\{\frac{\kappa^2}{\tau} \operatorname{Tr}\left(\Sigma^{\eta-1}\right)\right\}^{1/2}.
\end{aligned}$$

It is κ if $\eta = 2$ so that we can use κ to capture investors' belief about the root expected Sharpe ratio of the managed portfolios.

Given the previous discussion, the prior imposed on μ as in (2.6) also implies that the prior information for ω should be

$$\omega = \Sigma^{-1}\mu \sim \mathcal{N}\left(0, \frac{\kappa^2}{\tau} \mathrm{I}_L\right), \qquad \text{with } \eta = 2,$$

where I_L refers to an identity matrix of dimension L. The matrix representation is

$$\underset{L\times 1}{F_t} = \underset{L\times 1}{\mu} + \underset{L\times 1}{\varepsilon}, \qquad \varepsilon \sim (0, \Sigma). \tag{2.7}$$

or equivalently in the stacked matrix form

$$\underset{LT\times 1}{f} = \begin{pmatrix} F_1 \\ \vdots \\ F_T \end{pmatrix} = \underbrace{(\mathbf{1}_T \otimes \mathrm{I}_L)}_{X} \underset{L\times 1}{\mu} + \underset{LT\times 1}{\tilde{\varepsilon}}, \qquad \tilde{\varepsilon} \sim (0, \Xi), \tag{2.8}$$

$$\Xi = \mathrm{I}_T \otimes \Sigma.$$

The structure of $\tilde{\Sigma}$ implies that there is no time-series correlation. Recall the usual conjugate posterior for μ under the linear model framework, denoted by $\hat{\mu}$, is

$$\hat{\mu} \;=\; \left(\Xi_0^{-1} + X^\top \Xi^{-1} X\right)^{-1} \left(\Xi_0^{-1}\mu_0 + X^\top \Xi^{-1} f\right).$$

In the case for (2.8), by construction we have

$$\mu_0 = 0, \quad \Xi_0 = \frac{\kappa^2}{\tau}\Sigma^\eta, \quad X = \mathbf{1}_T \otimes \mathrm{I}_L.$$

Hence,

$$\hat{\mu} \;=\; \left(\Xi_0^{-1} + X^\top \Xi^{-1} X\right)^{-1} X^\top \Xi^{-1} f.$$

Note that

$$\begin{aligned}
X^\top \Xi^{-1} X \;&=\; \left(\mathbf{1}_T \otimes \mathrm{I}_L\right)^\top \tilde{\Sigma}^{-1} \left(\mathbf{1}_T \otimes \mathrm{I}_L\right) \\
&=\; \left(\mathbf{1}_T \otimes \mathrm{I}_L\right)^\top \left(\mathrm{I}_T \otimes \Sigma\right)^{-1} \left(\mathbf{1}_T \otimes \mathrm{I}_L\right) \\
&=\; \left(\mathbf{1}_T^\top \otimes \mathrm{I}_L\right) \left(\mathrm{I}_T \otimes \Sigma^{-1}\right) \left(\mathbf{1}_T \otimes \mathrm{I}_L\right) \\
&=\; \left(\mathbf{1}_T^\top \otimes \Sigma^{-1}\right) \left(\mathbf{1}_T \otimes \mathrm{I}_L\right) \\
&=\; \mathbf{1}_T^\top \mathbf{1}_T \otimes \Sigma^{-1} = T\Sigma^{-1}, \\
X^\top \Xi^{-1} f \;&=\; \left(\mathbf{1}_T \otimes \mathrm{I}_L\right)^\top \left(\mathrm{I}_T \otimes \Sigma\right)^{-1} f \\
&=\; \left(\mathbf{1}_T^\top \otimes \mathrm{I}_L\right) \left(\mathrm{I}_T \otimes \Sigma^{-1}\right) f \\
&=\; \left(\mathbf{1}_T^\top \otimes \Sigma^{-1}\right) f \\
&=\; \mathrm{vec}\left(\Sigma^{-1} \tilde{f} \mathbf{1}_T\right),
\end{aligned}$$

where

$$f = \text{vec}\left(\tilde{f}\right), \qquad \tilde{f} = \begin{pmatrix} F_1 & \cdots & F_T \end{pmatrix}_{L\times T}, \qquad \tilde{f}\mathbf{1}_T = T\bar{\mu}.$$

Thus,

$$X^\top \Xi^{-1} f = \text{vec}\left(\Sigma^{-1} T\bar{\mu}\right) = \Sigma^{-1} T\bar{\mu}.$$

Finally

$$\hat{\mu} = \left(\Xi_0^{-1} + T\Sigma^{-1}\right)^{-1} T\Sigma^{-1}\bar{\mu}.$$

Consequently,

$$\begin{aligned}
\hat{\omega} &= \Sigma^{-1}\hat{\mu} \\
&= \Sigma^{-1}\left(\Xi_0^{-1} + T\Sigma^{-1}\right)^{-1} T\Sigma^{-1}\bar{\mu} \\
&= \left[T^{-1}\Sigma\left(\Xi_0^{-1} + T\Sigma^{-1}\right)\Sigma\right]^{-1}\bar{\mu} \\
&= \left[T^{-1}\Sigma\frac{\tau}{\kappa^2}\Sigma^{-\eta}\Sigma + \Sigma\right]^{-1}\bar{\mu} \\
&= \left[T^{-1}\frac{\tau}{\kappa^2}\Sigma^{2-\eta} + \Sigma\right]^{-1}\bar{\mu} \\
&= \left[\frac{\tau}{T\kappa^2}\Sigma^{2-\eta} + \Sigma\right]^{-1}\bar{\mu}. \qquad (2.9)
\end{aligned}$$

If $\eta = 2$, we have

$$\hat{\omega} = \left(\gamma \mathrm{I}_L + \Sigma\right)^{-1}\bar{\mu}, \qquad \gamma = \frac{\tau}{T\kappa^2}. \qquad (2.10)$$

Similarly, the posterior covariance of $\hat{\mu}$ is

$$\text{Var}\left(\hat{\mu}\right) = \left(\Xi_0^{-1} + X^\top\Xi^{-1}X\right)^{-1} = \left(\frac{\kappa}{\tau^2}\Sigma^{-\eta} + T\Sigma^{-1}\right)^{-1}.$$

The posterior covariance matrix can expressed as

$$\begin{aligned}\text{Var}\left(\hat{\omega}\right) &= \Sigma^{-1}\left(\frac{\tau}{\kappa^2}\Sigma^{-\eta}+T\Sigma^{-1}\right)^{-1}\Sigma^{-1}\\ &= \left[\Sigma\left(\frac{\tau}{\kappa^2}\Sigma^{-\eta}+T\Sigma^{-1}\right)\Sigma\right]^{-1}\\ &= \left[\left(\frac{\tau}{\kappa^2}\Sigma^{2-\eta}+T\Sigma\right)\right]^{-1} = \frac{1}{T}\left[\frac{\tau}{T\kappa^2}\Sigma^{2-\eta}+\Sigma\right]^{-1}.\end{aligned}$$

Since $\eta = 2$, we have

$$\text{Var}\left(\hat{\omega}\right) = \frac{1}{T}\left(\gamma \mathrm{I}_L + \Sigma\right)^{-1}, \tag{2.11}$$

where $\text{Var}\left(\hat{\omega}\right)$ can be used to construct the confidence interval or t-statistic.

Remark 2.2 (Connection with penalized estimator) *The proposed Bayesian estimator is closely related to the penalized estimator. Consider the following cases where each penalized estimator is constructed based on different objective function*

(i) The objective function is constructed to maximize the cross-sectional R^2 with the penalty imposed on the model implied Sharpe ratio,

$$\mathbb{E}(F) = \Sigma\omega \textit{ and } \left(\Sigma\omega\right)^{\top}\Sigma^{-1}\left(\Sigma\omega\right) = \omega^{\top}\Sigma\omega.$$

Then,

$$\hat{\omega} = \arg\min_{\omega}\left\{\left(\bar{\mu}-\Sigma\omega\right)^{\top}\left(\bar{\mu}-\Sigma\omega\right)+\gamma\omega^{\top}\Sigma\omega\right\}. \tag{2.12}$$

(ii) The objective function is constructed to minimize the HJ distance,

$$\hat{\omega} = \arg\min_{\omega}\left\{\left(\bar{\mu}-\Sigma\omega\right)^{\top}\Sigma^{-1}\left(\bar{\mu}-\Sigma\omega\right)+\gamma\omega^{\top}\omega\right\}. \tag{2.13}$$

(iii) The objective function is constructed as that in the ridge regression,

$$\hat{\omega} = \arg\min_{\omega}\left\{\left(\bar{\mu}-\Sigma\omega\right)^{\top}\left(\bar{\mu}-\Sigma\omega\right)+\gamma\omega^{\top}\omega\right\}. \tag{2.14}$$

(i) and (ii) share the same solution and the solution is the same as the case when $\eta = 2$*. (iii) is the same as the case* $\eta = 3$*. This is because the first order condition with respect to* ω *in (2.14) yields*

$$-\Sigma\left(\bar{\mu} - \Sigma\omega\right) + \gamma\omega = \mathbf{0}.$$

Solving this equation, we have

$$\hat{\omega} = \left(\Sigma + \gamma\Sigma^{-1}\right)^{-1}\bar{\mu},$$

which is the case implied by (2.9) when $\eta = 3$*. For (ii), when* $\eta = 2$*, we can regard (2.13) as the* L^2*-norm penalized cross-sectional regression with the HJ distance as the objective function (alternatively, it can be understood as the extension of the ridge regression with the objective function being the HJ distance). Consequently, the tuning parameter associated with the* L^2*-norm penalized cross-sectional regression,* γ *is closely related with the root expected Sharpe ratio (under the prior),* κ *(implied from (2.10)). In this regard, the imposed Bayesian prior structure (2.6) brings the corresponding economic theory to the tuning procedure.*

Remark 2.3 *The justification of the Bayesian interpretation of SDF is essentially given by the prior imposed on* μ *conditional on the fact that investors update their knowledge about the cross-sectional variance-covariance structure via the observed returns. This maps well to the robust estimator for a relatively large variance-covariance matrix in the literature (Ledoit and Wolf, 2004a,b). This connection can be easily seen from the Wishart prior imposed on the precision matrix (in general, the inverse of variance-covariance matrix, i.e.,* $P = \Sigma^{-1}$*) commonly used for Bayesian analysis. Suppose we have the following prior for the precision matrix*

$$\Sigma^{-1} \sim \mathcal{W}\left(U_0, \varpi_0\right),$$

where U_0 a $L \times L$ positive definite matrix with ϖ_0 degrees of freedoms such that $\varpi_0 > L - 1$. Let $\mathbf{x}$ follow a multivariate normal distribution with mean zero. The conditional density function is given by

$$p(\mathbf{x} \mid P) = (2\pi)^{-L/2} |P|^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}^\top P \mathbf{x}\right).$$

Since the probability density function of the Wishart distribution is

$$p(P) = \frac{|P|^{(\varpi_0 - L - 1)/2} \exp\left[-\operatorname{Tr}\left(U_0^{-1} P\right)/2\right]}{2^{\frac{\varpi_0 L}{2}} \Gamma(\varpi_0/2) |U_0|^{\varpi_0/2}},$$

the posterior distribution given $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ is

$$\begin{aligned} p(P \mid \mathbf{X}) &\propto p(P)(\mathbf{X} \mid P) \\ &\propto \prod_{i=1}^{T} \left[|P|^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}_i^\top P \mathbf{x}_i\right)\right] |P|^{(\varpi_0 - L - 1)/2} \exp\left[-\operatorname{tr}\left(U_0^{-1} P\right)/2\right] \\ &= |P|^{(T + \varpi_0 - L - 1)/2} \exp\left\{-\frac{1}{2}\operatorname{Tr}\left[\left(T\mathbf{S} + U_0^{-1}\right) P\right]\right\}, \end{aligned}$$

where

$$\mathbf{S} = \frac{1}{T}\sum_{i=1}^{T} \mathbf{x}_i \mathbf{x}_i^\top,$$

is the sample counterpart of the variance-covariance matrix. This suggests that the posterior distribution of P is also a Wishart distribution such that

$$P \mid \mathbf{X} \sim \mathcal{W}\left(\left(T\mathbf{S} + U_0^{-1}\right)^{-1}, T + \varpi_0\right).$$

To make it connected to our discussion in the main context, replacing U_0 and ϖ_0 with $\frac{1}{L}\Sigma_0^{-1}$ and L, we have

$$\Sigma^{-1} \sim \mathcal{W}\left(\frac{1}{L}\Sigma_0^{-1}, L\right).$$

Replacing $\mathbf{X}$ *with the demeaned return over the time-series dimension and the variance-covariance matrix with* $\mathbf{S} = \Sigma_T$*, we have*

$$\Sigma^{-1} \mid \mathbf{X} \sim \mathcal{W}\left(\left(T\Sigma_T + L\Sigma_0\right)^{-1}, T + L\right).$$

The expected value (posterior) is

$$\mathbb{E}\left[\Sigma^{-1} \mid \mathbf{X}\right] = (T+L)\left(T\Sigma_T + L\Sigma_0\right)^{-1} = \left[\left(\frac{L}{T+L}\right)\Sigma_0 + \left(\frac{T}{T+L}\right)\Sigma_T\right]^{-1}.$$

The typical choice for Σ_0 *is* $\Sigma_0 = \frac{1}{L}\operatorname{Tr}\left(\Sigma_T\right)\mathbf{I}_L$ *where* $\mathbf{I}_L$ *is the* $L \times L$ *identity matrix. Consequently, we use*

$$\bar{\Sigma} = \left(\frac{L}{T+L}\right)\Sigma_0 + \left(\frac{T}{T+L}\right)\Sigma_T. \tag{2.15}$$

to replace Σ *in all relevant formulas in this chapter.*

2.2.3 Dual-penalty in combination of two norms

We discussed a key insight of Kozak, Nagel, and Santosh (2020) in detail, that is, the L^2-norm penalty imposed on the cross-sectional regression has a nice Bayesian interpretation, which is grounded on economics theory. However, a more strict shrinkage penalty can also be used. In our empirical analysis, for example, we also consider the following dual L^1-L^2 penalized cross-sectional regression by adding the following L^1-norm penalty term

$$\hat{\omega} = \arg\min_{\omega}(\bar{\mu} - \Sigma\omega)^\top \Sigma^{-1}(\bar{\mu} - \Sigma\omega) + \gamma_2\omega^\top\omega + \gamma_1\sum_{i=1}^{L}|\omega_i|. \tag{2.16}$$

This choice is related to the elastic-net method proposed in Zou and Hastie (2005), with the objective function slightly modified to be the HJ distance. The objective function for cross-validation is the cross-sectional R^2 defined by

$$R^2_{\text{oos}} = 1 - \frac{\left(\bar{\mu}_{\text{o}} - \bar{\Sigma}_{\text{o}}\hat{\omega}\right)^\top\left(\bar{\mu}_{\text{o}} - \bar{\Sigma}_{\text{o}}\hat{\omega}\right)}{\bar{\mu}_{\text{o}}^\top\bar{\mu}_{\text{o}}}. \tag{2.17}$$

This is similar to the standard routine in the statistical learning literature where the whole sample is divided into K sub-samples. In each fold of cross-validation,

$K-1$ sub-samples are used as training samples to calculate the sample mean and variance-covariance matrix (over time-series dimension), denoted by $\bar{\mu}_{\text{I}}$ and $\bar{\Sigma}_{\text{I}}$, while the remained samples are used as the testing samples to calculate the sample mean and variance-covariance matrix (over time-series dimension), denoted by $\bar{\mu}_{\text{O}}$ and $\bar{\Sigma}_{\text{O}}$. For the penalized cross-sectional regression with the L^2-norm penalty, ω is estimated using (2.12) or (2.13). For the penalized cross-sectional regression with the dual L^1-L^2-norm penalty, ω is estimated using (2.16).

2.3 Data

In this section, we first briefly discuss the recent literature on constructing cross-sectional equity characteristics for asset pricing studies and explain how we use the existing methods to construct equity characteristics in the Chinese stock market. Then we discuss how characteristic-managed portfolios are constructed based on daily returns of individual assets in the Chinese stock market. Discussions contained in this section share much in common with that in Chapter 1. To make the corresponding discussions self-contained within this chapter, we still briefly summarize the key steps for cleaning and constructing data.

2.3.1 Individual equity characteristic data

As we have discussed in Chapter 1, we combine both the data cleaning routines in Chen and Zimmermann (2020) and Jensen, Kelly, and Pedersen (2022) to replicate 123 finance and accounting anomaly variables in the Chinese stock market from 1995 to 2020. All the data (including returns and accounting data) are obtained from the Center for Research in Security Prices (`CRSP`), `Compustat`, and the China Stock Market & Accounting Research (`CSMAR`) database, all of which can be downloaded from the Wharton Research Data Service (WRDS). These anomaly variables are normalized as in Freybergerk, Neuhierl, and Weber (2019) so that each characteristic is normalized over the cross-sectional dimension to take a value between 0 and 1.

More precisely,

$$rc_{i,t}^{s} = \frac{\text{rank}\left(c_{i,t}^{s}\right)}{n_t + 1}, \tag{2.18}$$

where $c_{i,t}^{s}$ denotes the originally unscaled firm-level equity characteristic (indexed by superscript s) associated with stock i at time t and n_t denotes the total number of individual assets available for observations at time t. $\text{rank}(\cdot)$ denotes the cross-sectional ranking order of specific variable. Then, for each rank-transformed characteristic $rc_{s,t}^{i}$, we center it around the cross-sectional mean and divide it by the sum of average deviations from the cross-sectional mean for available stocks. Hence, we have,

$$z_{i,t}^{s} = \frac{\left(rc_{i,t}^{s} - \overline{rc}_{t}^{s}\right)}{\sum_{i=1}^{n_t}\left|rc_{i,t}^{s} - \overline{rc}_{t}^{s}\right|}, \tag{2.19}$$

where

$$\overline{rc}_{t}^{s} = \frac{1}{n_t}\sum_{i=1}^{n_t} rc_{i,t}^{s}.$$

Each column of Z_t is $\left(z_{1,t}^{s}, \ldots, z_{n_{t,1}}^{s}\right)^{\top}$. It is known in practice that individual characteristic data is imbalanced panel dat. For this reason, we exploit n_t rather than N to emphasize the time-varying cross-sectional dimension.[2]

2.3.2 Characteristic-managed portfolios

Annual accounting data is realigned with monthly return data based on the following annual rebalancing rule. Returns at the monthly frequency from July of year t to June of year $t+1$ are matched to the annual accounting variables in December of $t-1$. This is also the mechanism in which we realign data to construct cross-sectional equity characteristic data. For monthly rebalancing to construct the daily characteristic-managed portfolios, a similar scheme applies. That is, to construct the daily characteristic-managed portfolios in month $t+1$ based on equity s, returns at

[2]This also implicitly suggests that for each cross-section we only use those individual assets available as observations both for the corresponding returns and specific characteristics (indexed by s).

the daily frequency are matched with the normalized characteristics $z^s_{i,t}$ in month t and $z^s_{i,t}$ are used as the weights for constructing the daily characteristic-managed portfolios. Characteristics normalized as in (2.19) ensure the managed portfolios, to some extent, mimic the long-short trading strategies so that we can use the normalized characteristics as the weights for constructing portfolios. These normalized variables are then used to construct 123 characteristic-managed portfolios (either in the monthly or daily frequency, and portfolios constructed at the daily frequency will be mainly used for the empirical analysis). More comprehensive descriptions of these anomaly variables are listed in the appendix along with acronyms used in our replication procedure. The corresponding papers in which these anomaly variables were initially proposed are listed in the appendix as well.

2.4 Empirical Findings

We now apply this Bayesian interpretable machine-learning method in analyzing the sparse structure of cross-sectional exposure of SDF using the data for the Chinese stock market constructed above. Our main empirical finding is that, in general, it is a futile effort to summarize the SDF as the exposure to several dominant cross-sectional characteristics in the Chinese stock market.

We first demonstrate results generated from imposing the L^2-norm penalty on cross-sectional regression (i.e., ridge regression with H-J distance as the objective function, which is also nicely interpretable from the Bayesian perspective). As we have discussed in the previous section that the tuning parameter (γ_2) associated with the L^2-penalized cross-sectional regression is closely related to the expected Sharpe ratio under the Bayesian prior (κ), we plot both IS (in-sample)/OOS (out-of-sample) R^2 across different κ values in the following figure

[Place Figure 2.1 about here]

In the following figure, we plot the coefficient path associated with the L^2-norm-

penalized regression across different root expected Sharpe ratios under the Bayesian prior (κ), that is, different strengths of the L^2-penalty imposed on the cross-sectional regression.

[Place Figure 2.2 about here]

Next, we summarize both the estimated coefficients $\hat{\omega}$ and the associated absolute value of the t-statistic calculated using (2.10) and (2.11).

[Place Table 2.1 about here]

The main implication from Table 2.1 is that although there is rarely any redundancy of the cross-sectional equity characteristics to summarize SDF in the Chinese stock market since absolute values of the SDF coefficients associated with the leading SDF factors (among 123 SDF factors) listed in Table 2.1a are not close to zero. However, according to Table 2.1b, there are 2 to 3 leading latent factors (i.e., principal components) that are statistically significant with relatively large estimated SDF coefficients. Based on the classification in Jensen, Kelly, and Pedersen (2022), it is not surprising to see that **Size**, **Value** and **Investment** related equity characteristics matter for SDF in the Chinese stock market. This empirical result is not far away from that in the U.S stock market. The t-statistics reported in Table 2.1a are calculated using (2.11). These reported t-statistics are for reference since, in general, the joint selection matters more for the penalized regression with the L^2-norm penalty than the single selection.

Finally, we discuss how this Bayesian interpretable machine-learning algorithm (i.e., single L^2-norm penalized cross-sectional regression with objective function adjusted as HJ-distance) can be extended to the cross-sectional regression with a dual-penalty by adding additional L^1-norm penalty for accommodating shrinkage purpose.

[Place Figure 2.3 about here]

Figure 2.3 essentially provides the main conclusion of this chapter. It is obvious from this figure that the optimal tuning parameter pair (γ_1, γ_2) (or equivalent (γ_1, κ)) leading to the highest OOS R^2 resides in the area with relatively smaller γ_1 and γ_2 (or equivalently reflected in the number of variables retained in the SDF, over y-axis and the root expected Sharpe ratio, over x-axis) in Figure 2.3. This cross-validated out-of-sample analysis implies that it is futile to summarize SDF in the Chinese stock market as the exposure to several dominant cross-sectional equity characteristics.

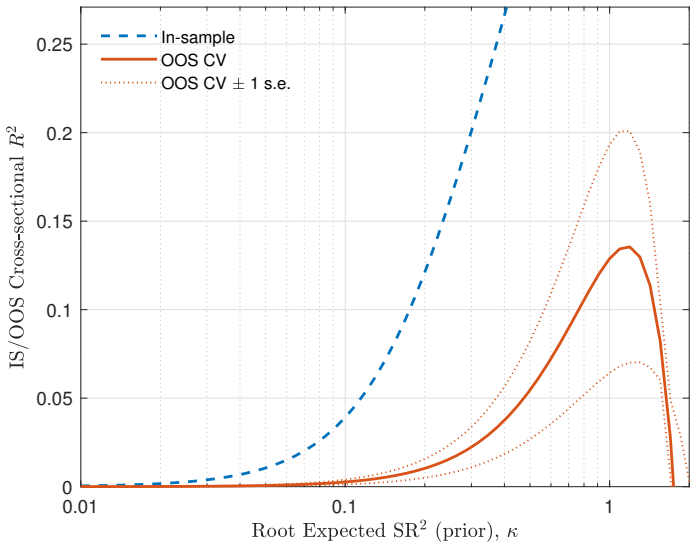
2.5 Conclusion

In this chapter, we review an interpretable machine-learning method that features an economic-theory-based foundation from a Bayesian perspective. The cross-sectional regression with the L^2-norm penalty (the ridge regression with H-J distance as the objective function) has interpretation with the economic grounds from the Bayesian perspective. Given the attractive property of the methodology proposed in Kozak, Nagel, and Santosh (2020), we apply this methodology to analyze whether there exists a sparse structure of the SDF in the Chinese stock market. From the empirical perspective, we follow the cutting-edge data cleaning routine that is in response to recent discussions about the replication crisis in the empirical cross-sectional asset pricing literature to successfully replicate and construct 123 finance and accounting characteristics of individual assets in the Chinese stock market and hence construct the corresponding characteristics (anomalies) managed portfolios. Based on these constructed characteristics (anomalies)-managed portfolios, we apply both the pure L^2-penalized cross-sectional regression (the ridge regression with the H-J distance as the objective function) and the extended L^1-L^2 penalized cross-sectional regression (elastic-net regularization) to check whether there is a sparse exposure of SDF in the Chinese stock market. Our empirical study suggests that staying within the 123 cross-sectional equity characteristic universe, it is still hard to characterize the SDF in the Chinese stock market using a few dominant characteristics, although

our empirical analysis may suggest that there exist several dominant latent factors (principal components) to summarize the SDF in the Chinese stock market.

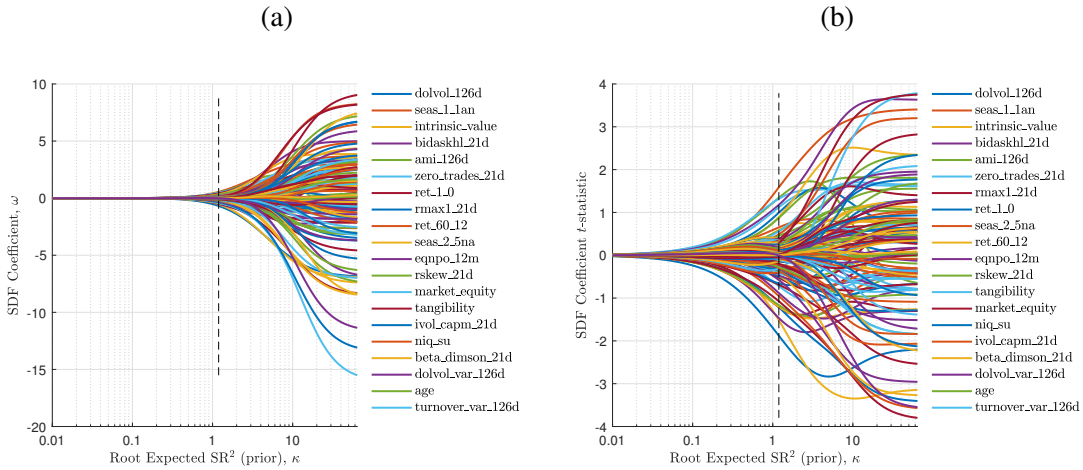
Figures and Tables

Figure 2.1



Note: In the figure above, we apply the Bayesian interpretable machine-learning method to the Chinese stock market by constructing characteristic-managed portfolios based on 123 anomaly variables. All characteristic-managed portfolio returns are constructed at the daily frequency. As we have discussed in the main context about the relationship between the root maximum squared Sharpe Ratio (κ) and the penalty parameter γ, in this figure we demonstrate cross-sectional R^2 and κ (both in-sample (dashed blue line) and out-of-sample (solid red line)). The standard-error (dotted red line) is calculated based on the split sample for cross-validation (3-folds cross-validation is used for this implementation).

Figure 2.2



Note: In the figure above, we apply the Bayesian interpretable machine-learning method in the Chinese stock market by constructing characteristic-managed portfolios based on 123 anomaly variables. This plot demonstrates coefficient paths associated with the penalized cross-sectional regression with the L^2-norm penalty: $\hat{\omega}$ as estimated SDF coefficients across different κ in (a) and corresponding t-statistics (using equation (2.10) and (2.11)) in (b). All the corresponding variables are sorted according to the absolute values. Vertical dashed lines both in (a) and (b) indicate the optimal tuning parameter based on cross-validation, i.e. κ associated with the highest OOS R^2.

Table 2.1

(a)

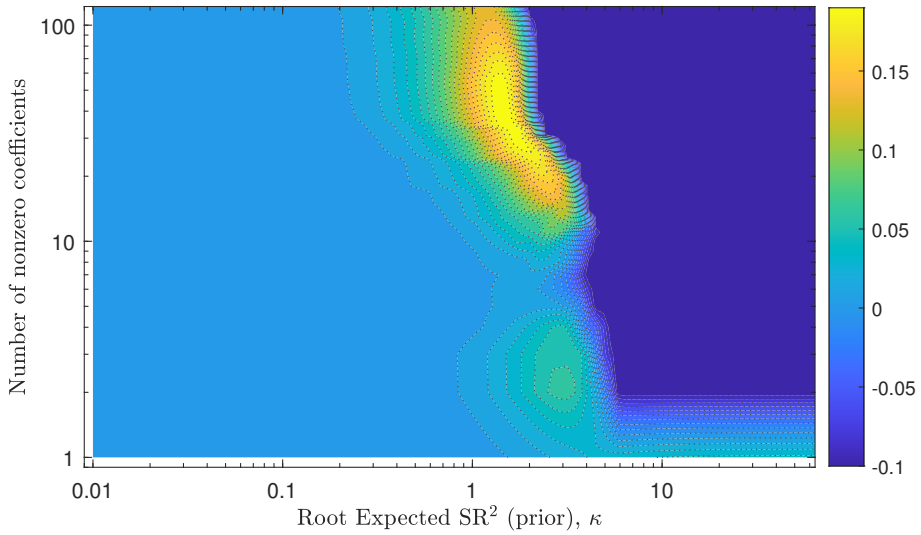
	ω	t-stat
Dollar trading volume **[Size]**	-0.6815	1.8792
Year 1-lagged return, annual **[Profit Growth]**	0.5474	1.5867
Intrinsic-value **[Value]**	-0.5456	1.5178
21 Day high-low bid-ask spread **[Low Leverage]**	-0.5164	1.4670
Amihud measure **[Size]**	0.4863	1.3422
Number of zero trades (1 month) **[Low Risk]**	0.4722	1.3180
Maximum daily return **[Low Risk]**	-0.4390	1.2135
Short-term reversal **[Size]**	-0.4446	1.2103
Years 2-5 lagged returns, nonannual **[Investment]**	-0.4271	1.2017
Long-term reversal **[Investment]**	-0.4283	1.2013

(b)

	ω	t-stat
PC 7	**1.0265**	**3.3041**
PC 8	**1.0365**	**3.2092**
PC 3	**0.4018**	**2.0319**
PC 6	0.4594	1.6777
PC 11	0.5407	1.6288
PC 24	0.5201	1.4496
PC 28	-0.4708	1.3003
PC 17	0.4283	1.2330
PC 1	-0.1212	1.1643
PC 13	-0.3400	1.0087

Note: In the table above, we summarize corresponding results obtained from applying the Bayesian interpretable-machine learning method to the Chinese stock market by constructing characteristic-managed portfolios based 123 anomaly variables. In (a) we summarize estimated coefficients $\hat{\omega}$ at the optimal L^2-norm penalty tuning parameter γ_2 (or equivalently the root expected Sharpe ratio κ under the prior distribution (based on cross-validation)). There 123 anomaly portfolios in all. In (b), anomaly portfolios returns are rotated into principal component (PC) space and corresponding estimated coefficients are demonstrated there. Coefficients are sorted descending on the absolute t-statistic values.

Figure 2.3



Note: In the figure above, we apply the Bayesian interpretable machine-learning method to the Chinese stock market by constructing characteristic-managed portfolios based 123 anomaly variables. This plot demonstrates OOS R^2 associated with the L^1-L^2-penalized cross-sectional regression discussed in the main context. L^2-penalty is tuned via γ_2, which is closely related to the root expected Sharpe ratio κ (over x-axis) under prior distribution; L^1-penalty is tuned via γ_1 and is in general proportional to the reciprocal of the number nonzero coefficients in cross-sectional regression. Hence, we use the number of nonzero coefficients (i.e. number of variables retained in SDF) to characterize the strength associated with L^1-penalty (over y-axis). Both axes are plotted on logarithmic scale. Yellow color depicts the higher OOS R^2 while the dark blue area depicts (γ_1, γ_2) pair for which the corresponding OOS R^2 is low.

Chapter 3

Alternative Parametric Models for Spot Volatility in High Frequency: A Bayesian Approach

3.1 Introduction

Financial market volatility as the measure of risk plays a vital role both in finance theory and applications of asset pricing theory in practice (Engle, 2004). Acknowledging the fact that daily volatilities are time-varying, a strand of literature focuses on modeling daily volatility parametrically based on daily returns. Examples include the ARCH model of Engle (1982), the GARCH model of Bollerslev (1986), and the stochastic volatility model of Taylor (1982). As a by-product of volatility modeling, an estimate of daily volatility can be obtained after the model is estimated.

In a more recent strand of literature, daily realized volatilities (RVs) are used to estimate daily integrated volatility (IV). Daily RV is a nonparametrical method that is based on intraday returns, usually 5-minute returns; see Andersen and Bollerslev (1997) and Andersen, Bollerslev, Christoffersen, and Diebold (2013). By exploiting intraday information, 5-minute returns can estimate daily volatility more accurately than daily returns. Subsequently, considerable efforts have been made to find a

reasonable model for daily RV, which is then used to forecast future daily RV; see Andersen, Bollerslev, Diebold, and Ebens (2001); Andersen, Bollerslev, Diebold, and Labys (2001, 2003); Gatheral, Jaisson, and Rosenbaum (2018); Wang, Xiao, and Yu (2022). Other than providing a more accurate estimate to IV, RV has been found a wide range of applications. For example, in an interesting paper, Bollerslev and Zhou (2002) use RVs, obtained from 5-minute returns, to construct GMM estimators for parameters in several parametric diffusion models.

However, most parametric models for daily volatilities (either RV or spot volatility) are not suitable for modeling spot volatilities in high frequencies. This is not surprising as spot volatilities in high frequencies have more complicated behavior than what a standard parametric diffusion model can generate.

Based on 5-minute returns on S&P500 index futures from March 11, 2007, through March 9, 2012, Stroud and Johannes (2014) proposes a high-frequency model where the total volatility has a multiplicative specification, including traditional autoregressive stochastic volatility components, seasonal components, and announcement components. They introduce a Bayesian method to estimate parameters in the model and find that all three components are important in the model.

The attempt to build a high-frequency model is important to enhance our understanding of the intraday behavior in volatility. It has potential important implications for asset pricing, volatility forecasting, trading, risk management. However, the criticism to the use of daily returns as opposed to daily RV also applies here. That is, the use of 5-minute returns is less efficient than that based on returns in a higher frequency.

Apart from the literature where the quantity of interest is the daily RV, there is another strand of growing literature that tries to estimate spot volatility from ultra-high frequency data. For example, in a recent study, Bollerslev, Li, and Liao (2021) establishes a new theory for the conduct of nonparametric inference about the latent spot volatility. Unlike the theories that assume the number of observations in local estimation blocks go to infinite, the new theory treats the estimation block

size k as fixed. As a result, the estimation error in the spot volatility estimator can be characterized by a scaled chi-square random variable. Bollerslev, Li, and Liao (2021) carry out an empirical application based on the intraday S&P 500 equity index. One of important empirical results suggest that there exist jumps at FOMC news announcement times. However, this study makes no attempt to model the dynamics of spot volatility in high frequencies.

In this chapter we propose several high-frequency models for the spot volatility based on the theory of Bollerslev, Li, and Liao (2021). All alternative specifications can be expressed as a nonlinear non-Gaussian state-space model. In particular, in all the alternative models, the observation equation, where the fixed-k estimator of spot volatility and the true spot volatility are linked, comes directly from by the theory of Bollerslev, Li, and Liao (2021). The difference of the alternative models lies in how the dynamics in the latent spot volatility is specified.

We then conduct the Bayesian analysis of all the alternative models using Markov chain Monte Carlo (MCMC), including obtaining the posterior distributions for each parameter and each latent spot volatility. In particular, the posterior mean of latent spot volatility is the smoothed estimate of spot volatility. In addition, we make a Bayesian model comparison of alternative specifications via the Deviance Information Criterion (DIC).

The rest of the chapter is organized as follows. In section 3.2 relevant preliminary mathematical concepts are introduced. In section 3.3 the fix-k theory of Bollerslev, Li, and Liao (2021) is reviewed and used to motivate our modelling strategy. Section 3.4 introduces alternative volatility models. We also discuss the Bayesian methods for parameter estimation method, volatility extraction, and model comparison method. In section 3.5 Monte Carlo experiments are designed to demonstrate that our proposed Bayesian methods in general works well. Section 3.6 contains empirical studies. Finally, section 3.7 concludes this chapter and briefly discusses the agenda for future work. More comprehensive discussions of technical details about MCMC and other additional results are contained in the appendix.

3.2 Mathematical Foundation

Before we introduce our high-frequency volatility models, we first clarify some relevant mathematical notations and related concepts. For all the following discussions, all random variables are defined on a fixed (complete) probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Besides, for two random sequences a_n and b_n, we write $a_n \asymp b_n$, if $a_n/C \leq b_n \leq Ca_n$ for some finite constant $C \geq 1$.

3.2.1 Basic mathematical results

Definition 1 (Sample Paths) *For stochastic process X defined over probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for each fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ that maps from $[0, \infty)$ into $\mathbf{R}$ is called the sample path of stochastic process X.*

Definition 2 (Hitting Time) *Let X be a stochastic process and Λ be Borel set in $\mathbf{R}$. Define*

$$T(\omega) = \inf\{t > 0 : X_t \in \Lambda\}. \tag{3.1}$$

Then, T is called the hitting time of Λ for X.

Definition 3 (Stopping Time) *Let $(\mathcal{F}_t)_{t\in\mathbf{T}}$ be a filtration on Ω, a random variable τ taking values from $\mathbf{T} \cup \{\infty\}$ is a stopping time for the filtration $(\mathcal{F}_t)_t$ if **event** $\{\tau \leq t\} \in \mathcal{F}_t$ for every $t \in \mathbf{T}$.*

Definition 4 (càdlàg and càglàd) *A stochastic process is said to be càdlàg if it has sample paths, which are right continuous with left limits almost surely. Similarly, a stochastic process X is said to be càglàd if it has sample paths, which are left continuous with right limits almost surely.*

Definition 5 *Let X be stochastic process and T be a random time. X^T is said to be the **process stopped** at T if $X_t^T = X_{t\wedge T}$. Furthermore, if X is adapted and càdlàg and if T is a stopping time, then*

$$X_t^T = X_{t\wedge T} = X_t \mathbf{1}_{\{t<T\}} + X_T \mathbf{1}_{\{t\geq T\}}. \tag{3.2}$$

Definition 6 (Local Martingale) *A process is said to be local martingale if it is locally right-continuous martingale. That is, if there is a sequence of stopping times τ_n almost surely increasing to infinity and such that the stopped processes $\mathbf{1}_{\{\tau_n > t_0\}} X^{\tau_n}$ are martingales. Equivalently, $\mathbf{1}_{\{\tau_n > t_0\}} X^{\tau_n}$ is integrable and*

$$\mathbf{1}_{\{\tau_n > t_0\}} X_{\tau_n \wedge s} = \mathbb{E}\left[\mathbf{1}_{\{\tau_n > t_0\}} X_{\tau_n \wedge t} \mid \mathcal{F}_s\right] \tag{3.3}$$

for all $s < t \in \mathbf{T}$, where $a \wedge b = min\{a, b\}$ and $\mathbf{1}_{\{\cdot\}}$ is an indicator function.

All the above definitions are prepared to define semimartingale, a concept that lays the foundation for modeling asset price in the continuous-time setting.

Definition 7 (Semimartingale) *In general, semimartingale is the stochastic process that can be decomposed as the sum of local martingale and an adapted finite-variation process. That is, in general, we have as in Revuz and Yor (2004)*

$$X_t = M_t + A_t, \tag{3.4}$$

where M_t is a local martingale and A_t is an adapted finite-variation process.

With the above mathematical concepts, Following Andersen, Bollerslev, Diebold, and Labys (2001) we adopt the assumption that logarithmic asset prices follow a univariate diffusion. In particular, for the asset indexed by i, the logarithmic return is modeled as

$$p_i(t) - p_i(t-1) \equiv r_k(t) = \int_{t-1}^{t} \mu_i(s) ds + \int_{t-1}^{t} \sigma_i(s) dW(s), \tag{3.5}$$

where $W(s)$ stands for the standard Wiener process and hence, the corresponding volatility measure is based on the quadratic variation process, denoted by $\mathrm{Qvar}_i(t)$, which yields

$$\mathrm{Qvar}_i(t) = [p_i, p_i]_t - [p_i, p_i]_{t-1} = \int_{t-1}^{t} \sigma_i^2(s) ds. \tag{3.6}$$

This is commonly referred to as the integrated volatility in the literature.

According to Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002), the integrated volatility over non-trivial time interval,

such as a day, is an important quantity of interest in finance. Many nonparametric estimators for daily IV have been proposed. Arguably, the most widely used estimator is the daily RV based on 5-minute returns.

With the increasing availability of data sampled at ultra high frequencies, how to estimate the spot volatility, that is $\sigma_i^2(t)$, has drawn a growing interest in the literature. Based on the mathematical foundation just laid out, we focus on the following model for (log) price process as in the literature,

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t. \tag{3.7}$$

This is a continuous-time Itô semimartingale process with drift, diffusion and jump.

3.3 Fixed-k Estimator of Spot Volatility

When the logarithmic price of an asset is characterized by the continuous-time Itô semimartingale process, Jacod, Li, and Liao (2020) suggests a way to estimate "spot covariance", $c_k(t) = \sigma_k(t)\sigma_k(t)^\top$, nonparametrically and uniformly as follows

$$\hat{c}_{n,j} \equiv \frac{1}{k_{n,j}\Delta_n} \sum_{i \in \mathcal{I}_{n,j}} \Delta_i^n X \Delta_i^n X^\top \mathbf{1}_{\{\|\Delta_i^n X\| \le u_n\}}, \tag{3.8}$$

where

$$\begin{aligned}
\Delta_n &: \quad T/n. \\
\Delta_i^n X &: \quad X_{i\Delta_n} - X_{(i-1)\Delta_n}. \\
u_n &: \quad \text{truncation threshold satisfying } u_n \asymp \Delta_n^{\varpi}. \\
\mathcal{I}_{n,j} &: \quad \text{set collecting indices of consecutive increments in } j\text{-th block,} \\
&\quad\ \ \text{such that } \{1,\dots,n\} = \textstyle\bigcup_{j=1}^{m_n} \mathcal{I}_{n,j} \text{ and } |\mathcal{I}_{n,j}| = k_{n,j}. \\
\mathcal{T}_{n,j} &: \quad \text{Correspondingly, } [0,T] \text{ can be dissected as } [0,T] = \textstyle\bigcup_{j=1}^{m_n} \mathcal{T}_{n,j} \\
&\quad\ \ t(n,j) \equiv (\min \mathcal{I}_{n,j} - 1)\,\Delta_n,
\end{aligned}$$

and

$$\mathcal{T}_{n,j} \equiv \begin{cases} [t(n,j), t(n,j+1)) & \text{if } 1 \le j < m_n \\ [t(n,m_n), T] & \text{if } \quad j = m_n. \end{cases}$$

$k_{n,j}$ denotes the block size. The issue of whether $k_{n,j}$ should be fixed or not is discussed in Bollerslev, Li, and Liao (2021). While it is commonly assumed $k_{n,j} \to \infty$, Bollerslev, Li, and Liao (2021) advocates a way of making inference for spot volatility with $k_{n,j} = k$ fixed. To see the link between the setting of Bollerslev, Li, and Liao (2021, henceforth BLL2021QE) and ours, note that BLL2021QE set

$$\begin{aligned} \mathcal{I}_{n,j} &\equiv \{(j-1)k+1, \ldots, jk\} \\ \mathcal{T}_{n,j} &\equiv [(j-1)k\Delta_n, jk\Delta_n)\,. \end{aligned}$$

This setting is a special case of ours with $k_{n,j} = k$.

Essentially $(\hat{c}_{n,j})_{1\le j\le m_n}$ serves as the functional estimator of $(c_t)_{t\in[0,T]}$. More specifically, $(\hat{c}_{n,j})_{1\le j\le m_n}$ is identified with t-indexed functional estimator $(\hat{c}_{n,t})_{t\in[0,T]}$ such that

$$\hat{c}_{n,t} \equiv \hat{c}_{n,j} \quad \text{for} \quad t \in \mathcal{T}_{n,j} \quad \textit{and} \quad j \in \{1, \ldots, m_n\}. \tag{3.9}$$

The following theorem, which comes from Jacod, Li, and Liao (2020), develops the properties of the estimator when $k_{n,j} \to \infty$.

Theorem 1 (Jacod, Li, and Liao (2020)) *Under the* ASSUMPTION *1 and* ASSUMPTION *2 imposed in Jacod, Li, and Liao (2020) and $k_{n,j} \asymp \Delta_n^{-\rho}$ uniformly for all $j \in \{1, \ldots, m_n\}$ and $u_n \asymp \Delta_n^{\varpi}$ such that $\rho \in (r, 1/2)$ and $\varpi \in ((1-\rho/2)/(2-\rho), 1/2)$. The following statements hold for some constant $\epsilon > 0$.*

(a) With

$$U_{n,j} \equiv k_{n,j}^{-1/2} \sum_{i\in\mathcal{I}_{n,j}} \left(\Delta_i^n W \Delta_i^n W^\top/\Delta_n - \mathbf{I}_d\right)$$

for each $1 \le j \le m_n$, we have

$$\max_{1\le j\le m_n} \sup_{t\in\mathcal{T}_{n,j}} \left\| k_{n,j}^{1/2}\left(\hat{c}_{n,t} - c_t\right) - \sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top \right\| = o_p\left(\Delta_n^\epsilon\right). \tag{3.10}$$

(b) If ASSUMPTION *2 holds, the following approximation result holds uniformly*

$$\max_{1\le j\le m_n} \sup_{t\in\mathcal{T}_{n,j}} \left| k_{n,j}^{1/2}\left(f(\hat{c}_{n,j}) - f(c_t)\right) - \mathrm{tr}[\partial f(c_{t(n,j)})\sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top] \right| = o_p\left(\Delta_n^\epsilon\right). \tag{3.11}$$

According to Theorem 9.3.2 of Jacod and Protter (2012), $k_{n,j} \to \infty$ and $k_{n,j}\Delta_n \to 0$ are needed to ensure the consistency of $\hat{c}_{n,t}$. The required conditions for the consistency is intuitive as they require the local estimation block contain an increasing number of observations (i.e. $k_{n,j} \to \infty$), while at the same time the size of local estimation block shrinks to zero asymptotically (i.e. $k_{n,j}\Delta_n \to 0$).

Although this double asymptotic scheme theoretically justifies the consistency of the nonparametric estimation of the spot volatility, it requires a carefully chosen tuning sequence $k_{n,j}$, as manifest in the above theorem. The fixed-k theory established in BLL2021QE alleviates the concern about mimicking the double-asymptotic scheme. We are now in a position to review the fixed-k theory of BLL2021QE.

3.3.1 Fixed k-inference for volatility

BLL2021QE suggests a way to nonparametrically infer latent spot volatility of asset prices characterized by continuous-time Itô semimartingale process. The main contribution of BLL2021QE lies in that the

By setting the estimation block size k fixed, the resulting spot volatility estimator of BLL2021QE is not consistent, easy-to-calculate pointwise confidence intervals are available at any given point in time.

In the univariate case, $c_t = \sigma_t^2$ is estimated by $\hat{c}_{n,j}$ where, for $t \in \mathcal{T}_{n,j}$ and $j \in \{1, \ldots, m_n\}$,

$$\hat{c}_{n,t} \equiv \hat{c}_{n,j}. \tag{3.12}$$

The only difference between (3.12) and (3.9) is that in (3.12) a fixed block size $k_{n,j} = k$ is used. Thus,

$$\hat{c}_{n,t} \equiv \hat{c}_{n,j} = \frac{1}{k\Delta_n} \sum_{i \in \mathcal{I}_{n,j}} (\Delta_i^n X)^2 \mathbf{1}_{\{|\Delta_i^n X| \leq u_n\}}. \tag{3.13}$$

The following main theorem comes from BLL2021QE.

Theorem 2 (Bollerslev, Li, and Liao (2021)) *Suppose that the* ASSUMPTION *1 imposed in BLL2021QE holds, then for any finite subset $\mathcal{M} \subseteq \{1, \ldots, m_n\}$, there exists*

a collection of independent random variables $\left(\bar{S}_j\right)_{j\in\mathcal{M}}$ *such that for any* $j \in \mathcal{M}$ *and* $t \in \mathcal{T}_{n,j}$,

$$\frac{\hat{c}_{n,t}}{c_t} - \bar{S}_j = O_p\left(\Delta_n^{(2-r)\varpi\wedge(1/2)}\right) = o_p(1), \tag{3.14}$$

where

$$\mathcal{M} \subseteq \{1,\ldots,m_n\} = (k\Delta_n)^{-1}\sum_{i\in\mathcal{I}_{n,j}}\left(W_{i\Delta_n} - W_{(i-1)\Delta_n}\right)^2,$$

is a $\bar{\chi}^2_k$*-distributed random variable. The* $\bar{\chi}^2_k$ *refers to the scaled chi-square distribution such that*

$$\bar{\chi}^2_k \equiv Z_k/k, \text{ with } Z_k \sim \chi^2_k. \tag{3.15}$$

In companion with this definition, we have the scaled inverse-chi-square distribution

$$\bar{\chi}^{-2}_k \equiv k/Z_k, \text{ with } Z_k \sim \chi^2_k. \tag{3.16}$$

With k fixed, (3.14) suggests that $\frac{\hat{c}_{n,t}}{c_t}$ can be approximated by a scaled chi-square distributed random variable, that is,

$$\frac{\hat{c}_{n,t}}{c_t} \xrightarrow{d} \bar{\chi}^2_k \tag{3.17}$$

as $n \to \infty$ (hence $\Delta_n \to 0$). Taking the log transformation, we have

$$\ln \hat{c}_{n,t} - \ln c_t = \ln \bar{\chi}^2_k = \ln Z_k - \ln k. \tag{3.18}$$

Definition 8 (Log chi-square distribution) *Let* $\mathcal{Z}_k$ *denotes the log chi-square distribution, that is*

$$\mathcal{Z}_k = \ln Z_k = \ln \chi^2_k.$$

By Lee (2012, page 379), we have the following results for the log chi-square distribution.[1]

[1]For notational simplicity, we suppress the degrees-of-freedom parameter k in the expressions.

- For the density function, we have[2]

$$p\left(\mathcal{Z}\right)=\frac{1}{2^{k/2}\Gamma(k/2)}\exp\left\{\frac{1}{2}k\mathcal{Z}-\frac{1}{2}\exp\left(\mathcal{Z}\right)\right\}\qquad\left(-\infty<\mathcal{Z}<\infty\right).$$

- For the moment generating function, we have

$$\text{MGF(t)}=2^{t}\Gamma\left(t+(k/2)\right)/\Gamma\left(k/2\right).$$

- For the mean and variance, we have

$$\begin{aligned}\mathbb{E}\left[\mathcal{Z}\right] &= \ln 2+\psi\left(k/2\right),\\ \mathbb{V}\left[\mathcal{Z}\right] &= \psi'\left(k/2\right),\end{aligned}$$

where

$$\psi(z)=\frac{\mathrm{d}}{\mathrm{d}z}\log\Gamma(z)=\frac{\Gamma'(z)}{\Gamma(z)}.$$

Remark 3.1 *It is well-known that* $\Gamma(z+1)=z\Gamma(z)$*. If we differentiate both sides of the equation, we have* $\Gamma'(z+1)=\Gamma(z)+z\Gamma'(z)$*. If we then divide both sides of the equation by* z *and substitute* $\Gamma(z)$ *by* $\Gamma(z+1)/z$*, we have*

$$\psi(z+1)=\frac{1}{z}+\psi(z).$$

This formula suggests that variance of the log chi-square distribution decreases as k *increases. In particular, it can be shown that*

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\log\Gamma(z)=\frac{\mathrm{d}}{z}\psi(z)=\psi'(z)=\sum_{j=0}^{\infty}\frac{1}{(z+j)^2}.$$

This result has the implication for the choice of the fixed local estimation block size (i.e. k*). The larger the local estimation block size is, the less the variance*

[2]It is easy to show that

$$\log p\left(\mathcal{Z}\right)=-\frac{k}{2}\log 2-\log\Gamma\left(k/2\right)+\frac{1}{2}k\mathcal{Z}-\frac{1}{2}\exp\left(\mathcal{Z}\right).$$

This formula plays an important role our acceptance-rejection sampling algorithm in the context of the algorithm of Kim, Shephard, and Chib (1998).

of the estimated spot volatility. This is consistent with the usual "bias-variance" trade-off (see Jacod, Li, and Liao, 2020; Bollerslev, Li, and Liao, 2021). According to the trade-off, a small k results in a more noisy but a less biased nonparametric estimate of the spot volatility. To make this "bias-variance" trade-off more formally, Bollerslev, Li, and Liao (2021) show that

$$\left\|\hat{c}_{n,j} - c_{(j-1)k\Delta_n}\right\|_2 \leq K\left(k^{-1/2} + (k\Delta_n)^{\kappa}\right), \tag{3.19}$$

where $\hat{c}_{n,j}$ is the nonparametric estimation of spot volatility over the jth block of length $k\Delta_n$, $\|\cdot\|_2$ is the standard L^2-norm, and κ is the "smoothness" parameter of the volatility process. (3.19) directly implies the "bias-variance" trade-off. In the following discussion, we mostly focus on the case where the local estimation window size is fixed at $k = 5$ and the price data is sampled at the frequency of every one minute within one day. Thus, $\Delta_n = 1/390 \approx 0.002$.

Although unobserved spot volatility is indexed continuously in our model, to facilitate nonparametric estimation of spot volatility, following (3.12), we assume there exists a surjective function that maps $t \in [0, T]$ to $j \in \{1, \ldots, m_n\}$.[3] To ensure our notations to be consistent wit those in the literature (such as Chernov, Ronald Gallant, Ghysels, and Tauchen (2003)), we split each day into M disjoint blocks and the size of each block is fixed as k. If T represents the total number of investigated trading days, then $n = k(MT)$ and $\Delta_n = T/n = 1/(kM)$. In this case, the total number of blocks for the T trading days is $m_n = MT$. Alternatively, we may have the following representation. For any $t \in [0, T]$ and $r \in (0, 1]$ or the corresponding discretized counterpart $t^\circ \in [0, T]$ and $r^\circ \in \{1/M, 2/M, \ldots, M/M = 1\}$,

$$t = \lfloor t- \rfloor + r,$$

and

$$t^\circ = \lfloor t- \rfloor + r^\circ,$$

[3]As implied by the fixed-k inference theory, instead of estimating $\ln c_{n,t}$ for $t \in [0, T]$, we estimate $\ln c_{n,j}$ sampled at discrete time points with $j \in \{1, \ldots, m_n\}$. In the state-space framework, therefore, we assume t is uniquely mapped to $j \in \{1, \ldots, m_n\}$ where m_n is the number of local estimation blocks.

where $\lfloor x- \rfloor$ denotes the greatest integer less than x.

Based on the fixed k-inference theory, we set up the following class of state-space model

$$\begin{cases} \ln\left(\hat{c}_{n,t^\circ}\right) &= \ln\left(c_{n,t^\circ}\right) + \epsilon_{t^\circ}, \quad \epsilon_{t^\circ} \sim \ln \bar{\chi}_k^2, & (3.20) \\ \ln\left(c_{n,t^\circ}\right) &= \text{alternative models.} & (3.21) \end{cases}$$

Clearly, the observation equation comes from the fix-k theory. Since ϵ_{t° is not a Gaussian variable, a model in this class is a nonlinear non-Gausian state-space model. Alternative model specifications will be introduced in the next section.

3.4 Alternative Model Specifications

We now specify several alternative models for the latent log spot volatility, $\ln\left(c_{n,t^\circ}\right)$. Following Stroud and Johannes (2014), in our most general specification, we assume that $\ln\left(c_{n,t^\circ}\right)$ can be decomposed into several components:

$$\ln\left(c_{n,t^\circ}\right) = \mu + h_{t^\circ} + s_{t^\circ} + a_{t^\circ}, \tag{3.22}$$

where h_{t° is stochastic volatility process, s_{t° the seasonal component, a_{t° the announcement component. By shutting down different components in $\ln\left(c_{n,t^\circ}\right)$ or having different specification for h_{t°, we end up with alternative models.

3.4.1 Alternative models

Model 1

If we shut down a_{t° and s_{t° in (3.22) and impose AR(1) structure for h_{t° with associated intercept μ, we have our first model – the benchmark model. That is,

$$\begin{aligned} \ln\left(c_{n,t^\circ}\right) &= \mu + h_{t^\circ} \\ h_{t^\circ}, &= \phi h_{t^\circ - 1/M} + e_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right). \end{aligned}$$

In this model, data is contained in $\{\ln \hat{c}_{n,t^\circ}\}$ with $\{h_{t^\circ}\}$ being latent variables. The parameters of the model are $\{\phi, \mu, \sigma_e^2\}$. For simplicity, we call this benchmark model – **Model 1**. To ensure the $\{h_{t^\circ}\}$ process to be stationary, we assume $\phi \in (0,1)$ and the distribution for the initial state is

$$\ln(c_{n,0}) \sim \mathcal{N}\left(\mu, \frac{\sigma_e^2}{1-\phi^2}\right). \tag{3.23}$$

The above model is different from the log square transformation of the lognormal stochastic volatility model widely studied in the literature; see, for example Harvey, Ruiz, and Shephard (1994). The lognormal stochastic volatility model is given by

$$\begin{aligned} r_t &= \sigma exp(h_t/2)\epsilon_t,\ \epsilon_t \sim \mathcal{N}(0,1), & (3.24)\\ h_t &= \phi h_{t-1} + \sigma_h \eta_t,\ \eta_t \sim \mathcal{N}(0,1). & (3.25) \end{aligned}$$

When applying the log square transformation to (3.24), we have

$$\ln r_t^2 = \mu + h_t + \ln \epsilon_t^2. \tag{3.26}$$

Clearly, $\ln \epsilon_t^2$ is a log chi-square random variable. This is contrast with the scaled chi-square random variable used in **Model 1**.

For the prior specification, we first introduce the auxiliary parameters ϕ^* and σ_e^{*2} respectively as in Kim, Shephard, and Chib (1998),

$$\begin{aligned} \phi &= 2\phi^* - 1 \\ \sigma_e &= \exp\left(\frac{1}{2}\ln \sigma_e^{*2}\right), \end{aligned}$$

and then use the following priors:

$$\begin{aligned} \phi^* &\sim \text{Beta}\left(\alpha_{\phi^*}, \beta_{\phi^*}\right) \\ \sigma_e^{*2} &\sim \text{I.G.}\left(\alpha_{\sigma_e^*}, \beta_{\sigma_e^*}\right) \\ \mu &\sim \mathcal{N}(0, 100) \end{aligned}$$

where $\alpha_{\phi^*}, \beta_{\phi^*}, \alpha_{\sigma_e^*}, \beta_{\sigma_e^*}$ are hype-parameters.

MCMC is applied to obtain the correlated random draws from the posterior distributions of μ, ϕ^* and $\ln(\sigma_e^{*2})$. These draws can be regarded as correlated random draws from the original parameters. Based on the MCMC draws, we can obtain the posterior mean, quantiles, variance for each parameter.

Model 2

Model 2 extends the benchmark model by combining the AR(1) structure and a discrete jump component in h_{t°. Specifically, **Model 2** is given by

$$\begin{aligned}
\ln(c_{n,t^\circ}), &= \mu + h_{t^\circ} \\
h_{t^\circ} &= \phi h_{t^\circ - 1/M} + e_{t^\circ} + J_{t^\circ}\eta_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right), \quad \eta_{t^\circ} \sim \mathcal{N}\left(\mu_\eta, \sigma_\eta^2\right),
\end{aligned}$$

where J_{t° is a jump indicator, defined by

$$J_{t^\circ} = \begin{cases} 1 & \text{with probability} \quad \kappa \\ 0 & \text{with probability} \quad 1-\kappa, \end{cases}$$

with κ being the jump probability, and η_{t° determines the jump size.

In this model, data is contained in $\{\ln \hat{c}_{n,t^\circ}\}$ with $\{h_{t^\circ}\}, \{J_{t^\circ}\}$ being latent variables. The parameters of the model are $\left\{\phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2\right\}$.

Following Chib, Nardari, and Shephard (2002), we assume the following conjugate priors for parameters in the jump component

$$\begin{aligned}
\kappa &\sim \text{Beta}(\alpha_\kappa, \beta_\kappa), \\
\mu_\eta &\sim \mathcal{N}\left(\tilde{\mu}_\eta, \tilde{\sigma}_\eta^2\right), \\
\sigma_\eta^2 &\sim \text{I.G.}\left(\alpha_{\sigma_\eta}, \beta_{\sigma_\eta}\right),
\end{aligned}$$

where I.G. denotes the Inverse-Gamma distribution.

Model 3

In **Model 3** we add the seasonal component to the benchmark model. The seasonal component for intraday volatility is used to capture the diurnal U-shaped patterns in high frequency financial data. There are several methods for modeling diurnal patterns, namely, the Fourier representation and a deterministic spline. Based on using 5-minute square return, when using a deterministic spline, Aït-Sahalia and Jacod (2014) and Christensen, Hounyo, and Podolskij (2018) document evidence of larger fluctuations near the opening and closing of the exchange than around lunch time. The model is given by

$$\begin{aligned}
\ln\left(c_{n,t^\circ}\right) &= \mu + h_{t^\circ} + s_{t^\circ}, \\
h_{t^\circ} &= \phi h_{t^\circ - 1/M} + e_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right), \\
s_{t^\circ} \equiv \tilde{s}_{r^\circ} &= 12\left(1-b\right)\left(r^\circ - \frac{1}{2}\right)^2 + b, \quad r^\circ = t^\circ - \lfloor t - \rfloor, \quad t \in [0, T].
\end{aligned}$$

The quadratic function $\tilde{s}_r = 12(1-b)\left(r - \frac{1}{2}\right)^2 + b$, is the only function within the class $f(r) = c(r-a)^2 + b$ that satisfies (i) $\int_0^1 \left(c(r-a)^2 + b\right) dr = 1$; (ii) $\operatorname{argmin}_r c(r-a)^2 + b = \frac{1}{2}$. The first condition is imposed for identification. The second condition assumes that the diurnal pattern reaches the minimum in the middle of a trading day, an empirical regularity that has been found in the literature.[4] There is a restriction in using the quadratic function. That is, it implies a symmetric diurnal pattern. In a recent study, Christensen, Hounyo, and Podolskij (2018) propose a nonparametric method to estimate the diurnal patten and find an asymmetric diurnal pattern. However, our approach can be easily extended to cover more complicated deterministic functions for diurnal pattern.

In this model, data is contained in $\{\ln \hat{c}_{n,t^\circ}\}$ with $\{h_{t^\circ}\}$ being latent variables.

[4]To satisfy the condition that $\operatorname{argmin}_r c(r-a)^2 + b = \frac{1}{2}$, we have $c > 0$ and $a = \frac{1}{2}$. Substituting $a = \frac{1}{2}$ into $\int_0^1 \left(c(r-a)^2 + b\right) dr = 1$ yields $\frac{1}{12}c + b = 1 \Rightarrow c = 12(1-b)$. Thus, the quadratic function is uniquely determined by single parameter b. The larger the value of b is, the less pronounce the quadratic volatility pattern.

The parameters of the model are $\{\phi, \mu, \sigma_e^2, b\}$.

Since **Model 3** and **Model 1** share the same AR(1) specification for $\{h_{t^\circ}\}$, we use the same priors on $\{\phi, \mu, \sigma_e^2\}$ as before. For parameter b, we assume a flat prior on $[0, 1]$, that is, $b \sim \mathcal{U}(0, 1)$.

Model 4

Different from **Model 3** that includes the component to capture the diurnal pattern, **Model 4** includes the component to capture macroeconomic news announcement effects. The motivation for incorporating announcement effects comes from recent empirical finance sdudies, for instance Lucca and Moench (2015) and Bernile, Hu, and Tang (2016). The specification of **Model 4** is given by

$$\begin{aligned}
\ln\left(c_{n,t^\circ}\right) &= \mu + h_{t^\circ} + a_{t^\circ}, \\
h_{t^\circ} &= \phi h_{t^\circ - 1/M} + e_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right), \\
a_{t^\circ} &= \sum_{q=1}^{Q} \sum_{l=0}^{L} \mathbf{1}_{t^\circ ql} \alpha_{ql},
\end{aligned}$$

where $\mathbf{1}_{t^\circ ql}$ is an indicator for news type q at time t° with $l = 0, 1, \ldots, L$ (i.e., $\mathbf{1}_{t^\circ ql} = 1$ if it is within l periods after type q announcement made at time $t^\circ - \frac{l}{M}$ and 0 otherwise), α_{ql} is the announcement effect for news type q at l periods after the announcement.

Again, since the specification for $\ln \hat{c}_{n,t^\circ}$ in **Model 4** is the same as that in **Model 1**, we use the same priors for parameters $\{\phi, \mu, \sigma_e^2\}$. Parameters $\{\alpha_{ql}\}_{q=1,\ldots,Q; l=1,\ldots,L}$ characterize announcement effects. The dimension of these parameters is $Q \times L$, and hence, it increases as Q and/or L increases. For instance, if $L = 5$ and $Q = 3$, we will have $L \times Q = 15$ parameters to determine the announcement effects. This would impose a great deal of computational challenges to the Bayesian analysis.

To alleviate the computational burden, we assume the announcement effects decay over time according to the following pattern, $\alpha_{ql} = \tilde{\alpha}_q \exp\left\{-\tilde{\beta}_q l\right\}$. This is relatively

a parsimonious specification that significantly reduces the dimension of the parameter space associated with announcement effects. The imposed decaying structure for the announcement effects is consistent with the intuition and the empirical evidence in the literature (see Stroud and Johannes, 2014; Lucca and Moench, 2015; Bernile, Hu, and Tang, 2016). Under this specification, the parameters are collected in $\left\{\tilde{\alpha}_q, \tilde{\beta}_q\right\}_{q=1}^{Q}$. When $L = 5$ and $Q = 3$, the number of parameters in connection to announcement effects reduce from 15 to $2 \times Q(= 3) = 6$. The following priors are used for the new parameters,

$$\begin{aligned}
\tilde{\alpha}_q &\sim \mathcal{N}\left(0, \tilde{\sigma}_q^2\right) \quad \text{for } q = 1, \cdots, Q, \\
\tilde{\beta}_q &\sim \mathcal{E}\left(-\tilde{\lambda}_q\right) \quad \text{for } q = 1, \cdots, Q,
\end{aligned}$$

where $\mathcal{N}(\cdot, \cdot)$ and $\mathcal{E}(\cdot)$ denote normal distribution and exponential distribution respectively. Accordingly, in this model, data is contained in $\{\ln \hat{c}_{n,t^\circ}\}$ with $\{h_{t^\circ}\}$ being latent variables. The parameters of the model are $\left\{\phi, \mu, \sigma_e^2, \tilde{\alpha}_q, \tilde{\beta}_q\right\}$.

Model 5

In **Model 5**, we consider the model specification by combining all the specifications of **Models 1-3**. Thus, we have both jumps and diurnal patterns included in the model specification. In other words, **Model 5** nests **Models 1-3**. We summarize the model structure of **Model 5** as follows with detailed explanations for notations kept in the description of **Models 1-3**.

$$\begin{aligned}
\ln\left(c_{n,t^\circ}\right) &= \mu + h_{t^\circ} + s_{t^\circ}, \\
h_{t^\circ} &= \phi h_{t^\circ - 1/M} + e_{t^\circ} + J_{t^\circ}\eta_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right), \quad \eta_{t^\circ} \sim \mathcal{N}\left(\mu_\eta, \sigma_\eta^2\right), \\
s_{t^\circ} \equiv \tilde{s}_{r^\circ} &= 12\left(1 - b\right)\left(r^\circ - \frac{1}{2}\right)^2 + b, \quad r^\circ = t^\circ - \lfloor t - \rfloor, \quad t \in [0, T].
\end{aligned}$$

The parameters of nested **Model 6** are $\left\{\phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b\right\}$.

Model 6

Recall (3.22) for our general specified functional form of latent volatility process. By shutting down (or opening up) different components, we can obtain different model specifications. **Models 1-4** discussed in the previous subsections are about adding different components (i.e. jumps, diurnal components, and announcement effect components) respectively onto the single factor volatility model, **Model 1**. Alternatively speaking, built upon the benchmark model specification in **Model 1**, by combining different specifications in **Models 2-4**, we can at most obtain $C_3^0 + C_3^1 + C_3^2 + C_3^3 = 1 + 3 + 3 + 1 = 8$ different models. **Models 1-4** are about the subset of combinations (i.e. $C_3^0 + C_3^1$). We refer to the most comprehensive model specification that includes all the components as **Model 6**. Thus **Model 6 nests** all the specifications in **Models 1-4**. We summarize the model structure of **Model 6** as follows with detailed explanations for notations kept the same as in the description of **Models 1-4**.

$$\begin{aligned}
\ln\left(c_{n,t^\circ}\right) &= \mu + h_{t^\circ} + s_{t^\circ} + a_{t^\circ} \\
h_{t^\circ} &= \phi h_{t^\circ - 1/M} + e_{t^\circ} + J_{t^\circ}\eta_{t^\circ}, \quad e_{t^\circ} \sim \mathcal{N}\left(0, \sigma_e^2\right), \quad \eta_{t^\circ} \sim \mathcal{N}\left(\mu_\eta, \sigma_\eta^2\right), \\
s_{t^\circ} \equiv \tilde{s}_{r^\circ} &= 12\left(1-b\right)\left(r^\circ - \frac{1}{2}\right)^2 + b, \quad r^\circ = t^\circ - \lfloor t - \rfloor, \quad t \in [0, T], \\
a_{t^\circ} &= \sum_{q=1}^{Q}\sum_{l=0}^{L} \mathbf{1}_{t^\circ ql}\alpha_{ql}.
\end{aligned}$$

The parameters of nested **Model 6** are $\left\{\phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b, \tilde{\alpha}_q, \tilde{\beta}_q\right\}$.

3.4.2 Bayesian analysis

MCMC sampling from posterior

The generic logic of Bayesian analysis is to make all the corresponding analyses about information that we want to learn from data based on the posterior distribution.

That is, suppose we have data denoted as $\mathbf{y}$ in general and let $\boldsymbol{\vartheta}$ denote the parameter space (it may include all the parameters associated with specific model and the corresponding latent variables if we apply data augmentation techniques), then the general Bayes' rule suggests that posterior analysis should based on the relation

$$p\left(\boldsymbol{\vartheta} \mid \mathbf{y}\right) = \frac{p\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) p\left(\boldsymbol{\vartheta}\right)}{\int p\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) p\left(\boldsymbol{\vartheta}\right) d\boldsymbol{\vartheta}}, \tag{3.27}$$

where $p\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right)$ and $p\left(\boldsymbol{\vartheta}\right)$ refer to conditional likelihood and prior distribution respectively. The main difficulty in obtaining posterior distribution as suggested in (3.27) arises from the required integration in the denominator. This integration in general is not available for closed-form solution (for instance, the nonlinear non-Gaussian state-space models in our setting summarized in (3.20) and (3.21)).

MCMC as the leading modern Bayesian technique is quite suitable for making posterior sampling from the target posterior associated with the state-space model with latent structure. The general idea of MCMC method can be understood as a combination of various sampling algorithms such as the Metropolis-Hastings (M-H) algorithm, acceptance-rejection algorithm, Gibbs sampler, and the substitution sampler (see data augmentation algorithm in Tanner and Wong, 1987) by making draws from conditional distributions associated with the target posterior. It can be theoretically justified that as long as we can let the Markov chain run long enough, those draws taken from blocks of conditional distributions constitute the target posterior distribution. Accordingly, we can use these draws to summarize the posterior mean (or mode) as the estimation of target parameters and latent variables. Gilks, Richardson, and Spiegelhalter (1995), Bolstad (2009), and Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (2013) are all good references for technical details of MCMC methods. We exploit $\boldsymbol{h}$ to denote the sequence of h_{t°, $\boldsymbol{J}$ to denote all the jump indicators, and $\boldsymbol{\eta}$ to denote all the jump sizes. For the sake of description simplicity, we let $\tilde{\boldsymbol{h}} = \boldsymbol{h} + \mu$. Meanwhile, we use $\mathbf{y}$ to denote the sequence of log transformation of nonparametric estimator of volatility, that is the sequence of $\ln\left(\hat{c}_{n,t^\circ}\right)$. The designed MCMC algorithms associated with **Models 1-6** are

summarized as follows respectively.

- **Model 1**

  (1) Initialize $\{\boldsymbol{h}, \phi, \mu, \sigma_e^2\}$.

  (2) Sample $\tilde{\boldsymbol{h}} \mid \phi, \mu, \sigma_e^2, \mathbf{y}$.

  (3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}$.

  (4) Go to (2).

- **Model 2**

  (1) Initialize $\left\{\boldsymbol{h}, \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2\right\}$.

  (2) Sample $\tilde{\boldsymbol{h}} \mid \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, \mathbf{y}$.

  (3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}, \boldsymbol{J}, \boldsymbol{\eta}$.

  (4) Sample $\boldsymbol{J} \mid \boldsymbol{h}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2$.

  (5) Sample $\boldsymbol{\eta} \mid \boldsymbol{h}, \boldsymbol{J}, \phi, \mu, \sigma_e^2, \mu_\eta, \sigma_\eta^2$.

  (6) Sample $\left\{\kappa, \mu_\eta, \sigma_\eta^2\right\} \mid \boldsymbol{J}, \boldsymbol{\eta}$.

  (7) Go to (2).

  In step (2) above, we need $\left\{\kappa, \mu_\eta, \sigma_\eta^2\right\}$ to obtain the initial condition of $\boldsymbol{h}$.

- **Model 3**

  (1) Initialize $\{\boldsymbol{h}, \phi, \mu, \sigma_e^2, b\}$.

  (2) Sample $\tilde{\boldsymbol{h}} \mid \phi, \mu, \sigma_e^2, b, \mathbf{y}$.

  (3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}$.

  (4) Sample $b \mid \boldsymbol{h}, \mu, \mathbf{y}$.

  (5) Go to (2).

  In step (4) above, we insert one M-H algorithm for sampling b.

- **Model 4**

(1) Initialize $\left\{\boldsymbol{h}, \phi, \mu, \sigma_e^2, \tilde{\alpha}_q, \tilde{\beta}_q\right\}$.

(2) Sample $\tilde{\boldsymbol{h}} \mid \phi, \mu, \sigma_e^2, \tilde{\alpha}_q, \tilde{\beta}_q, \mathbf{y}$.

(3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}$.

(4) Sample $\left\{\tilde{\alpha}_q, \tilde{\beta}_q\right\} \mid \boldsymbol{h}, \mu, \mathbf{y}$.

(5) Go to (2).

In step (4) above, we insert one M-H algorithm for sampling $\left\{\tilde{\alpha}_q, \tilde{\beta}_q\right\}$.

- **Model 5**

  This is a model nesting all the specifications from **Models 1-3**. The corresponding MCMC loop is summarized as follows

  (1) Initialize $\left\{\boldsymbol{h}, \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b\right\}$.

  (2) Sample $\tilde{\boldsymbol{h}} \mid \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b, \mathbf{y}$.

  (3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}, \boldsymbol{J}, \boldsymbol{\eta}$.

  (4) Sample $\boldsymbol{J} \mid \boldsymbol{h}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2$.

  (5) Sample $\boldsymbol{\eta} \mid \boldsymbol{h}, \boldsymbol{J}, \phi, \mu, \sigma_e^2, \mu_\eta, \sigma_\eta^2$.

  (6) Sample $\left\{\kappa, \mu_\eta, \sigma_\eta^2\right\} \mid \boldsymbol{J}, \boldsymbol{\eta}$.

  (7) Sample $b \mid \boldsymbol{h}, \mu, \mathbf{y}$.

  (8) Go to (2).

- **Model 6**

  This is the largest model nesting all the components that are expected to obtain from MCMC. The MCMC loop is summarized as follows,

  (1) Initialize $\left\{\boldsymbol{h}, \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b, \tilde{\alpha}_q, \tilde{\beta}_q\right\}$.

  (2) Sample $\tilde{\boldsymbol{h}} \mid \boldsymbol{J}, \boldsymbol{\eta}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2, b, \tilde{\alpha}_q, \tilde{\beta}_q, \mathbf{y}$.

  (3) Sample $\{\phi, \mu, \sigma_e^2\} \mid \tilde{\boldsymbol{h}}, \boldsymbol{J}, \boldsymbol{\eta}$.

  (4) Sample $\boldsymbol{J} \mid \boldsymbol{h}, \phi, \mu, \sigma_e^2, \kappa, \mu_\eta, \sigma_\eta^2$.

(5) Sample $\boldsymbol{\eta} \mid \boldsymbol{h}, \boldsymbol{J}, \phi, \mu, \sigma_e^2, \mu_\eta, \sigma_\eta^2$.

(6) Sample $\left\{\kappa, \mu_\eta, \sigma_\eta^2\right\} \mid \boldsymbol{J}, \boldsymbol{\eta}$.

(7) Sample $\left\{b, \tilde{\alpha}_q, \tilde{\beta}_q\right\} \mid \boldsymbol{h}, \mu, \mathbf{y}$.

(8) Go to (2).

DIC for model comparison

Deviance Information Criterion (DIC), proposed and well-discussed in Spiegelhalter, Best, Carlin, and van der Linde (2002, 2014), is a popular method for model selection when MCMC output is ready. There are a few nice features about DIC. First, DIC is applicable to a wide range of statistical models. Second, it does not suffer from Jeffreys-Lindley-Barlett's paradox. Third, it can be obtained even under improper priors. Finally, Li, Yu, and Zeng (2021) justify DIC by showing that DIC is an asymptotically unbiased estimator of the Kullback-Leibler divergence between the data generating process and the plug-in predictive distribution. DIC in general is given as follows

$$\mathrm{DIC} = D\left(\bar{\boldsymbol{\vartheta}}\right) + 2P_D, \tag{3.28}$$

where

$$\begin{aligned} D(\boldsymbol{\vartheta}) &= -2\ln p\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) \\ P_D &= -2\int [\ln p(\mathbf{y} \mid \boldsymbol{\vartheta}) - \ln p(\mathbf{y} \mid \bar{\boldsymbol{\vartheta}})] p(\boldsymbol{\vartheta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}. \end{aligned}$$

$\bar{\boldsymbol{\vartheta}}$ refers to the posterior mean of parameter $\boldsymbol{\vartheta}$ and $\mathbf{y}$ generically denote observable data. Specifically, for the DIC definition in (3.28), $D(\bar{\boldsymbol{\vartheta}})$ as the product of a negative number (-2) and log-observed-data likelihood evaluated at the posterior mean of parameters could be interpreted as the measure of model fit and hence it is expected to be minimized for fitting data purpose. While P_D as the second term in (3.28) could be interpreted as the product of a negative number (-2) and the posterior

expected deviance between log-observed-data likelihood evaluated at different parameter values over parameter space and log-observed-data likelihood evaluated at the posterior mean. The corresponding expectation is taken with respect to posterior distribution $p(\boldsymbol{\vartheta} \mid \mathbf{y})$. Since P_D is increasing with $\ln p\left(\mathbf{y} \mid \bar{\boldsymbol{\vartheta}}\right)$, there exists trade-off between these two terms $D\left(\bar{\boldsymbol{\vartheta}}\right)$ and P_D, therefore the objective to minimize DIC is reconciled with the goal of achieving balance between "model fit" and "model complexity". This also corresponds to the opinion that DIC can be understood as the Bayesian version of AIC. Computing DIC using conditional likelihood is straightforward based on posterior sampling from MCMC: (i) for $D\left(\bar{\boldsymbol{\vartheta}}\right)$, we just make $D(\boldsymbol{\vartheta})$ evaluated at posterior mean $\bar{\boldsymbol{\vartheta}}$; (ii) and p_D is calculated using posterior sample mean from MCMC.

3.5 Monte Carlo Experiments

In this section, we conduct several Monte Carlo experiments to check the performance of the proposed Bayesian method in estimating parameters, extracting volatility estimates, and in comparing alternative models. Several data generation processes, which match with different alternative model specifications in the previous section, have been used to simulate data.

3.5.1 Experiment 1

In the first experiment, we simulate data according to,

$$\begin{aligned} dX_t &= \exp(h_t/2)\, dW_t, \\ dh_t &= \kappa_h(\mu_h - h_t)\, dt + \sigma_h dB_t \\ \mathbb{E}[W_t B_t,] &= \rho = 0. \end{aligned}$$

Specifically we simulate discrete data by setting $dt := \Delta_i^n = 1/390$, which implies that price data is sampled every one minute, leading to $6.5 \times 60 = 390$ returns as we assume there are 6.5 trading hours. Parameter specification for this experiment is as follows: $\kappa_h = 0.2$, $\mu_h = -5$, $\rho = 0$, $\sigma_h = 0.4$. For notational simplicity, we write $\ln(c_{n,t})$ as h_t. When extracting latent volatility, we focus our attention on the one-month interval (i.e., $T = 22$, assuming that there are 22 trading days within one month). In total we have $n = 22 \times 390 = 8580$ returns.[5]

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical red dashed lines indicating the location of posterior means.

[Place Figure 3.1 about here]

Remark 3.2 (From DGP Dynamic to Block Dynamic) *Recall that in general data is generated from the following dynamic system, if we treat $dh_t = h_{t+1} - h_t$ of the second equation above, we are able to rewrite the second equation characterizing*

[5]We simulate price-level data using the described DGP each for a specific experiment. It works fine for different replicate experiments. Besides, we simulate price-level data by only accounting for diffusion process. This DGP scheme does not distort our general target since our approach relies on the distribution of the difference between the log fixed-k estimator of spot volatility and the true unobserved latent volatility. The derived distribution in BLL2021QE generally applies to continuous Itô semimartingale and therefore any fixed-k estimator of volatility associated price-level data generated from continuous Itô semimartingale does not affect the nonlinear non-Gaussian state-space model we establish. Similar DGP scheme has also been used in literature such as Xiu (2010).

latent spot volatility dynamics as follows

$$
\begin{aligned}
h_{t+1} - h_t &= \kappa_h \left(\mu_h - h_t\right) dt + \sigma_h dB_t \\
&\Leftrightarrow \\
(h_{t+1} - \mu_h) - (h_t - \mu_h) &= -\kappa_h dt (h_t - \mu_h) + \sigma_h dB_t \\
&\Rightarrow \\
h_{t+1} - \mu_h &= \underbrace{(1 - \kappa_h dt)}_{\phi_h} (h_t - \mu_h) + \underbrace{\sigma_h dB_t}_{\varepsilon_{t+1}}
\end{aligned}
$$

Given the property of Brownian motion, $\sigma_h dB_t \sim \mathcal{N}\left(0, \sigma_h^2 dt\right)$. Data is generated from this continuous setting. But to apply nonparametric estimation of spot volatility in our established framework, we have to select k consecutive time intervals ($dt \equiv \Delta_i^n$) to construct "local estimation window". This implies that we have to move from "observation" dynamics to "block" dynamics. If we simulate data from system given above, we are modeling following dynamics (let $\tilde{h}_t = h_t - \mu_h$ denote the demeaned latent spot volatility)

$$
\begin{aligned}
\tilde{\tilde{h}}_{t+k} &= \phi_h \tilde{\tilde{h}}_{t+k-1} + \varepsilon_{t+k} \\
&= \phi_h \left(\phi_h \tilde{\tilde{h}}_{t+k-2} + \varepsilon_{t+k-1}\right) + \varepsilon_{t+k} \\
&= \phi_h (\phi_h (\phi_h \tilde{\tilde{h}}_{t+k-3} + \varepsilon_{t+k-2}) + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\
&\cdots \\
&= \phi_h^k \tilde{\tilde{h}}_t + \left(\phi_h^k \varepsilon_t + \cdots + \varepsilon_{t+k}\right).
\end{aligned}
$$

Thus the "block" dynamics should be given as follows

$$
\tilde{\tilde{h}}_{j+1} = \phi_h^k \tilde{\tilde{h}}_j + e_{j+1}. \tag{3.29}
$$

This could be interpreted as that $\tilde{\tilde{h}}_j$ is regarded as constant spot volatility within the j-th block constructed by every k intervals. Variance of $e_{j+1} = \phi_h^k \varepsilon_t + \cdots + \varepsilon_{t+k}$ is

given as follows

$$\sigma_e^2 = \frac{\sigma_h^2 dt(1-\phi_h^{2(k+1)})}{1-\phi_h^2}.$$

For instance, if $dt := \Delta_i^n = 1/390$*,* $\sigma_h = 0.4$*,* $k = 5$*, and* $\kappa_h = 0.2$*, then* $\phi_h = 1 - \kappa_h dt \approx 0.9995$*,* $\phi_h^k \approx 0.9974$ *and*

$$\sqrt{\frac{\sigma_h^2 dt(1-\phi_h^{2(k+1)})}{1-\phi_h^2}} \approx 0.0496.$$

These quantities should be reasonably compared with posterior means summarized from MCMC outputs.

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.2 about here]

[Place Figure 3.3 about here]

For the results demonstrated above, the local estimation window size is fixed at $k = 5$. In this regard, we treat the unobserved latent volatility as constant every 5 minutes, and accordingly for this fixe-k scheme we have $M = 390/5 = 78$ local estimation blocks each day and $22 \times 78 = 1716$ local estimation blocks for 22 days.

3.5.2 Experiment 2

In this experiment, we simulate data according to,

$$\begin{aligned}
dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\
\tilde{h}_t &= \mu_h + h_t, \\
dh_t &= -\kappa_h h_t dt + \sigma_h dB_t + \mathbf{1}_{\{t=t^\circ\}} J_t \eta_t, \\
\mathbb{E}\left[W_t B_t\right] &= \rho = 0, \\
J_t &\sim \text{Bernoulli}(\kappa), \\
\eta_t &\sim \mathcal{N}\left(\mu_\eta, \sigma_\eta^2\right).
\end{aligned}$$

Specifically we simulate discrete data by setting $dt := \Delta_i^n = 1/390$, which implies that price data is sampled every one minute, leading to $6.5 \times 60 = 390$ returns as we assume there are 6.5 trading hours. Parameter specification for this experiment is as follows: $\kappa_h = 0.2$, $\mu_h = -5$, $\rho = 0$, $\sigma_h = 0.4$, $\kappa = 0.0025$, $\mu_\eta = 0.8$, $\sigma_\eta = 1.2$. For notational simplicity, we write $\ln(c_{n,t})$ as h_t. $\kappa = 0.0025 \approx 1/(78 \times 5)$ suggests that we assume approximately there is one jump each week. When extracting latent volatility, we focus our attention on the one-month interval (i.e., $T = 22$, assuming that there are 22 trading days within one month). In toal we have $n = 22 \times 390 = 8580$ returns. Besides, the jump component is incorporated in the transition dynamics of the volatility process using $\mathbf{1}_{\{t=t^\circ\}}$ to ensure that latent spot volatility transition dynamics in DGP are reconciled with corresponding specifications in **Model 2**. In other words, we only consider jumps that happen at the end of each local estimation block.

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical

red dashed lines indicating the location of posterior means.

[Place Figure 3.4 about here]

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.5 about here]

3.5.3 Experiment 3

In this experiment, we simulate data according to,

$$\begin{aligned} dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\ \tilde{h}_t &= \mu_h + h_t + s_t, \\ dh_t &= -\kappa_h h_t dt + \sigma_h dB_t, \\ \mathbb{E}\left[W_t B_t\right] &= \rho = 0, \\ s_t &= 12(1-b)\left(t - \lfloor t- \rfloor - \frac{1}{2}\right)^2 + b, \end{aligned}$$

where s_t takes the quadratic functional form as described for **Model 3**, thus

$$s_t = 12(1-b)\left(t - \lfloor t- \rfloor - \frac{1}{2}\right)^2 + b.$$

All the parameters except for the parameter that specifies intraday volatility pattern inherit from **Experiment 1**. We specify b for different cases: one for $b = 0.4$, which represents the case when daily volatility exhibits relatively strong diurnal U-shaped patterns; while the other for $b = 0.8$, which represents the case when daily volatility exhibits relatively weak diurnal U-shaped patterns.

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical red dashed lines indicating the location of posterior means.

[Place Figure 3.6 about here]

[Place Figure 3.7 about here]

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.8 about here]

3.5.4 Experiment 4

In this experiment, we simulate data according to,

$$\begin{aligned}
dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\
\tilde{h}_t &= \mu_h + h_t + a_{t^\circ}, \\
dh_t &= -\kappa_h h_t dt + \sigma_h dB_t, \\
\mathbb{E}\left[W_t B_t\right] &= \rho = 0, \\
a_{t^\circ} &= \sum_{q=1}^{Q} \sum_{l=0}^{L} \mathbf{1}_{t^\circ ql} \alpha_{ql}, \\
\alpha_{ql} &= \tilde{\alpha}_q \exp\left\{-\tilde{\beta}_q l\right\},
\end{aligned}$$

where a_{t° inherits the corresponding specification directly from **Model 4** such that

$$a_{t^\circ} = \sum_{q=1}^{Q} \sum_{l=0}^{L} \mathbf{1}_{t^\circ ql} \alpha_{ql}.$$

where $\mathbf{1}_{t^\circ ql}$ is indicator for news type q at t° with $l = 0, 1, \ldots, L$ (i.e., $\mathbf{1}_{t^\circ ql} = 1$ if it is l periods after type q announcement made at time $t^\circ - \frac{l}{M}$ and 0 otherwise). We currently focus on the single-type announcement effect, thus $Q = 1$ in this experiment. Besides, L is the parameter capturing the longest length of periods for which the announcement effect survives once it is made. L is set equal to 5 in this experiment. To simulate announcement indicators, we assume that the announcement happens at a rate approximately equal to $0.004 \approx 1/(78 \times 3)$. We specify the announcement effects by setting $\tilde{\alpha}_q = 0.8$, $\tilde{\beta}_q = 0.1$.

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical red dashed lines indicating the location of posterior means.

[Place Figure 3.9 about here]

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.10 about here]

3.5.5 Experiment 5

In this experiment, we simulate data by accommodating all the components in **Experiments 1-3**.

$$\begin{aligned}
dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\
\tilde{h}_t &= \mu_h + h_t + s_t, \\
dh_t &= -\kappa_h h_t dt + \sigma_h dB_t + \mathbf{1}_{\{t=t^\circ\}} J_t \eta_t, \\
\mathbb{E}\left[W_t B_t\right] &= \rho = 0, \\
J_t &\sim \text{Bernoulli}(\kappa), \\
\eta_t &\sim \mathcal{N}\left(\mu_\eta, \sigma_\eta^2\right), \\
s_t &= 12(1-b)\left(t - \lfloor t- \rfloor - \frac{1}{2}\right)^2 + b.
\end{aligned}$$

All the corresponding notations inherit directly those in **Experiments 1-3** and the values assigned to these parameters as well.

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical red dashed lines indicating the location of posterior means.

[Place Figure 3.11 about here]

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.12 about here]

3.5.6 Experiment 6

In this experiment, we simulate data by accommodating all the components in **Experiments 1-4** via following data generating process

$$\begin{aligned} dX_t &= \exp\left(\tilde{h}_t/2\right) dW_t, \\ \tilde{h}_t &= \mu_h + h_t + s_t + a_{t^\circ}, \\ dh_t &= -\kappa_h h_t dt + \sigma_h dB_t + \mathbf{1}_{\{t=t^\circ\}} J_t \eta_t, \\ a_{t^\circ} &= \sum_{q=1}^{Q} \sum_{l=0}^{L} \mathbf{1}_{t^\circ ql} \alpha_{ql}, \\ \alpha_{ql} &= \tilde{\alpha}_q \exp\left\{-\tilde{\beta}_q l\right\}. \end{aligned}$$

All the corresponding notations inherit directly those in **Experiments 1-4** and the values assigned to these parameters as well.

First, to check the performance of the Bayesian method in estimating model parameters, we report the posterior means, posterior standard errors for all parameters. The corresponding results are summarized in the following figures with the vertical red dashed lines indicating the location of posterior means.

[Place Figure 3.13 about here]

Second, to check the performance of the smoothing and filtering methods in extracting latent volatility, we plot the simulated (true) volatility, fixed-k estimation of volatility, and smoothed volatility in figures:

[Place Figure 3.14 about here]

3.5.7 Discussions on model comparison for Model 1-4

For **Experiments 1-4** associated with **Model 1-4**, we summarize model comparison results using DIC in the following table, where **Model i=1,2,3,4** and **DGP i=1,2,3,4** refer to model specifications and data generating processes discussed in the previous subsections respectively. For each data generating process **DGP i=1,2,3,4**, DIC based on conditional likelihood is reported along with the decomposed components in (3.28). In general, the model with a relatively smaller DIC (highlighted in bold) should be preferred to the model with a relatively larger DIC. It is possible to see from this Monte Carlo experiment that standard deviance information criterion can in general select the alternative model associated with the true data generating process. Since for **DGP i=1,2,3**, it is by design that there is **no observed announcement indicators**, hence we do not compare **Model 4** with **Model i=1,2,3** for **DGP i=1,2,3**.

[Place Table 3.1 about here]

3.6 Empirical Study

This section discusses the empirical applications of our proposed estimation method to spot volatility. We see from earlier discussions that a prominent feature associated with spot volatility estimation by fixing the local estimation window size is the "noise" introduced, which motivates our proposed methodology using the MCMC techniques to extract the spot volatility from the noise nonparametric estimate. We first demonstrate some applications of the proposed methodology in tracking volatilities associated with individual assets. Then we discuss an application to quantify the economic value of the private information closely connected with the return volatility in high frequency in the studies of financial microstructure.

3.6.1 Extracting spot volatility for individual asset

Data used for our empirical study is collected mainly from two sources. For the data corresponding to the U.S. equity market, we mainly collect it from the NYSE TAQ database (Trade and Quote database).[6] We follow the procedure suggested in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), which has been encompassed in `highfrequency` package maintained at `CRAN` (Boudt, Kleen, and Sjørup, 2021). This procedure aims to eliminate non-zero trades and filter for valid sales conditions.[7] Aside from that, we merge trade entries that have the same timestamp into a single one. That is, if there are multiple observations available for a specific timestamp, we take the median of these multiple observations as the corresponding observation associated with that timestamp. For the Chinese stock market, we mainly use one-minute price-level data of CSI 300 index futures.[8]

For the sake of mitigating the effect of microstructure "noise", it is a common practice to use price data sampled at the one-minute sampling frequency, (see Zhang, Mykland, and Aït-Sahalia, 2005). For the one-minute sampling scheme, we calculate the corresponding one-minute return within trading hours each day from 9:30 to 16:00 by taking close-open return (i.e. close-open log price difference) as return from the interval from 9:30 to 9:31 and close-close return (i.e. close-close log price difference) as the return for the following sampling intervals up till to 16:00, which

[6]This database contains intraday transactions data (trades and quotes). Generally there are 3 kinds of data products: **Trade & Quote Daily Product** (09/10/2003-present), **Trade & Quote Monthly Product** (01/01/1993-12/31/2014) and **NYSE Reg Sho Data** (01/01/2005-07/31/2007). In general, The TAQ Daily and Monthly data products are nearly identical whereas the key difference arises from that **Monthly Data Product** is delivered a whole month at a time, typically 60-90 days after the last trading day of the month. **Daily Data Product** is delivered one day at a time, hours after trading stops, and is available on WRDS the next day. We retain our focus on using **Trade & Quote Daily Product** as it is actively maintained and is of relatively higher quality in the sense that sampling intervals are more refined for the **Trade & Quote Daily Product** such that timestamps are provided at milliseconds (10^{-3} `secs`) granularity through March 2015, and in microseconds (10^{-6} `secs`) starting in April 2015.

[7]For more about sales conditions, readers may refer to NYSE online documentation about daily TAQ trade files at `https://www.nyse.com/publicdocs/nyse/data/Daily_TAQ_Client_Spec_v3.3.pdf`. By implementing this procedure, we essentially retain our focus on stocks exchanged in a single exchange market (for instance, `T/Q` refers to the NASDAQ exchange market).

[8]More details are summarized in `http://www.cffex.com.cn/en_new/CSI300IndexFutures.html`.

is usually the ending trading hours each day.[9] More specifically, we should expect 391 observed data at price level and accordingly $6.5 \times 60 = 390$ one-minute returns per day. We use two market indices (S&P 500 index ETF representing the U.S. market and CSI 300 index futures representing the Chinese market respectively) and one individual stock (Apple Inc.) as the data source. We demonstrate both the extracted volatility (red dashed line) and the corresponding nonparametric estimation of volatility (blue solid line) with local estimation block size fixed (i.e. fixed $k = 5$, every 5-minutes) as follows[10]

S&P 500 index ETF from TAQ in November 2015

[Place Figure 3.15 about here]

Apple Inc. Stock Price in August 2017

[Place Figure 3.16 about here]

CSI 300 Index futures in January 2020

[Place Figure 3.17 about here]

CSI 300 Index futures in August 2020

[Place Figure 3.18 about here]

For each index, the extracted volatility is based on **Model 5**, the nested model including all the specifications of **Models 1-3**. For all the applications above, we apply the truncation technique in Mancini (2001) based on the suggestion in BLL2021QE for applying fixed-k inference theory in practice. Specifically, our truncation threshold u_n is selected as satisfying $u_n \asymp \Delta_n^{\varpi}$ with $\varpi \in (0, 1/2)$. We choose $u_n = C\bar{\sigma}\Delta_n^{\varpi}$ such that $C = 3$ and $\varpi = 0.49$, where the preliminary estimator of volatility, $\bar{\sigma}$, can

[9] Timing scheme for trading hours corresponding to CSI 300 index futures. From 2010 to 2015, we have pirce-level data sampled at one-minute frequency from 9:15 to 15:15 each day; while from 2016 to 2021, we have data sampled at one-minute frequency from 9:30 to 15:00 each day.

[10] The extracted volatility is referred to as the smoothed volatility in this thesis. One may also extract volatility as the filtered volatility.

be chosen as the bipower variation estimator of Barndorff-Nielsen and Shephard (2004). Besides, for all the results summarized from Figure 3.15 to Figure 3.18, we use the blue solid line to indicate the fixed-k estimator ($k = 5$) of spot volatility and use the red dashed line to indicate the smoothed volatility using our proposed methodology.

We also make a model comparison across **Models 1-3** based on DIC to check whether we need to extend the benchmark model specification (**Model 1**) to incorporate either jumps in the volatility dynamics or the diurnal pattern of latent volatility process. Specifically for each year (S&P 500 index ETF in 2015, Apple Inc. Stock Price in 2017 and CSI 300 Index futures in 2020) and each month we apply our proposed methodology to extract volatility based on model specification corresponding to **Model 1**, **Model 2**, and **Model 3**. Results are summarized in Table 3.2, Table 3.3, and Table 3.4 respectively.

[Place Table 3.2 about here]

[Place Table 3.3 about here]

[Place Table 3.4 about here]

These results suggest that for most cases, we need to allow for the jump specifications in modeling the latent volatility process and that is why we apply **Model 5** that nests both the jump specifications and intraday diurnal pattern specification for extracting volatility. We also summarize the posterior mean and standard deviations of parameter MCMC draws as follows.

[Place Table 3.5 about here]

[Place Table 3.6 about here]

[Place Table 3.7 about here]

By comparing the posterior mean of b summarized in the last column of Table 3.5, Table 3.6, and Table 3.7, we find that: the volatility of individual stock return exhibits a relatively stronger (lower b) diurnal pattern while the volatility associated with market indices exhibits a relatively weaker (higher b) diurnal pattern in the corresponding periods.

3.6.2 Spot volatility, liquidity, and strategic value of information

In this section, we specifically focus on one application of our proposed methodology for extracting volatility to study the financial market microstructure, and hence, to obtain the economic value of information in strategic trading.

Volatility plays a vital role in modern financial market microstructure studies. Essentially all the studies about the financial market structure are about recovering insiders' private trading information in comparison to relatively uninformed liquidity traders. In other words, the central question is what the value of asset-specific information is to a strategic trader and how to practically quantify such a kind of value, which is also the amount investors would pay for information. The idea for quantifying the value of information is to use two components: (i) the extent to which specific information can offer speculator the reduction in uncertainty; (ii) liquidity associated with assets for which the acquired information can be used to trade quickly without generating adverse effects on the assets' prices. Inspired by Grossman and Stiglitz (1980) and the subsequent studies in Kyle (1985) and Back (1992), this idea has been justified using the ratio of uncertainty about the asset's fundamentals and the asset's illiquidity measure.

Given the recent finding in empirical literature (Collin-Dufresne and Fos, 2015; Kacperczyk and Pagnotta, 2019; Akey, Grégoire, and Martineau, 2022) that private information is hardly reflected by equity prices, Kadan and Manela (2021) extends the modeling framework of Kyle (1985) and Back (1992) (comprehensively summarized in Back (2017)) and proposes that in equilibrium the ex-ante dollar expected profits

of informed trader over a specific interval indexed from 0 to 1 can serve as the measure of value of information to strategic trader, which is given by[11]

$$\Omega = \frac{\sigma_v^2}{\lambda} P_0, \tag{3.30}$$

where σ_v^2 characterizes volatility associated with private information of informed traders and P_0 refers to the initial price of the specific asset over this timing interval. Besides, we follow the convention in literature as in Back (2017) assuming that private information is denoted by $\tilde{v}$ and follows log-normal distribution such that $\ln \tilde{v} \sim \mathcal{N}(\mu, \sigma_v^2)$. In (3.30), λ is widely known as Kyle's Lambda, initially proposed in Kyle (1985), as the measure of sensitivity of assets' return to share order flow (Lee and Ready, 1991; Ellis, Michaely, and O'Hara, 2000; Holden and Jacobsen, 2014). Kyle's lambda serves as an alternative measure reflecting financial market turbulence, which is usually high during periods in which the whole financial market is exposed to a systematic crisis such as the 2008 financial depression and more recent years Covid-19 global pandemic crisis. As we can see from (3.30) that σ_v^2 as the major component characterizing the value of information to strategic trader also measures the magnitude of reduction in uncertainty that speculator would have if had acquired corresponding information. More importantly, as we will see in the following discussion that although σ_v^2 is originally the measure of uncertainty associated with information (for instance, at price level), it can be directly used as the measure of volatility associated with the logarithmic return of asset.

In the literature on financial market microstructure, it is usually assumed that observed cumulative share orders in the continuous-time modeling framework, denoted by Y_t, can be decomposed into cumulative share orders of informed trader, denoted by X_t; and cumulative share orders of uninformed trader, denoted by Z_t, thus

$$Y_t = X_t + Z_t. \tag{3.31}$$

Dynamics specified for is directly characterized via Brownian motion as $dZ_t =$

[11] Alternatively, it is possible to interpret 0 as the starting point of specific interval while 1 as the ending point of specific interval.

$\sigma_z dB_t$. Then results contained in Theorem 3 and Example 2 in Back (1992) and Back (2017) suggest that in equilibrium

$$\frac{dP_t}{P_t} = \lambda dY_t \text{ and } dX_t = \frac{\frac{\ln \bar{v}-\mu}{\lambda} - Y_t}{1-t}dt, \quad \bar{v} = \mathbb{E}\left[\tilde{v}\right], \tag{3.32}$$

where P_t refers to price process associated with target asset in equilibrium. Then it is straightforward to see

$$\begin{aligned}
\frac{dP_t}{P_t} &= \lambda dY_t \\
&= \lambda dX_t + \lambda dZ_t \\
&= \lambda dX_t + \frac{\sigma_v}{\sigma_z}\sigma_z dB_t \\
&= \lambda \frac{\frac{\ln \tilde{v}-\mu}{\lambda} - Y_t}{1-t}dt + \sigma_v dB_t,
\end{aligned}$$

which yields one important implication for the empirical strategy that σ_v^2 as the measure of volatility associated with private information can be used as the proxy for volatility associated with logarithmic return. Specifically, for two consecutive asset prices P_{τ_i} and $P_{\tau_{i-1}}$, by applying log-linearization we have

$$r_i := \ln\left(\frac{P_{\tau_i}}{P_{\tau_{i-1}}}\right) \approx \frac{P_{\tau_i} - P_{\tau_{i-1}}}{P_{\tau_{i-1}}},$$

where the right-hand side of the equation above can be regarded as the discretized approximation of $\frac{dP_t}{P_t}$.

The general idea for quantifying strategic value of information is simple: suppose we have a specific way to estimate Kyle's lambda, denoted by $\hat{\lambda}$, then we can construct a spot information value associated with informed traders by using our proposed spot volatility estimation as the proxy of asset return volatility over each tiny interval using Bayesian techniques based upon the fixed-k inference theory. Thus,

$$\hat{\Omega} = \frac{\hat{\sigma}_v^2}{\hat{\lambda}} P_0. \tag{3.33}$$

This is a more desirable method than the widely used integrated volatility (i.e. annualized realized volatility) method. This is because the model developed in Back (1992) is naturally a continuous-time extension of the discrete-time model of Kyle (1985). Accordingly, spot volatility is preferred to integrated volatility. In practice, a feasible method to estimate λ is to regress the asset's return on share order flow over the interval. The theoretical justification for using regression to back out λ originated from modeling insider trading in the continuous-time setting proposed by Back (1992). Recently, this idea is discussed more comprehensively in Back (2017) and Kadan and Manela (2021), that is,

$$r_i = \lambda y_i + \varepsilon_i, \tag{3.34}$$

where $r_i = p_{\tau_i} - p_{\tau_{i-1}} = \ln P_{\tau_i} - \ln P_{\tau_{i-1}}$ (log-return of assets over specific interval indexed by i) and $y_i = Y_{\tau_i} - Y_{\tau_{i-1}}$ (share order flow over specific interval indexed by i).[12]

Construction of the proxy for share order flow is an active research area. In the literature on financial market microstructure, this is usually achieved by designing trade classification algorithms to identify trading direction. We borrow the ideas from Holden and Jacobsen (2014) by using following "order imbalance" as the proxy of share order flow,[13]

$$\text{Order Imbalance} = \frac{Buys - Sells}{Buys + Sells}, \tag{3.35}$$

where the trading classification scheme to identify the trading direction, Buys $(+1)$ and Sells (-1), is inherited from Chakrabarty, Li, Nguyen, and Van Ness (2007) and Holden and Jacobsen (2014). The estimated private information value associated insider trading is demonstrated as follows based on **Model 5**.

[Place Figure 3.19 about here]

Specifically, for each day we use univariate regression to estimate Kyle's lambda by regressing one-minute log-return of S&P 500 ETF on corresponding share order

[12] τ_i refers to the continuous timing index associated with interval i.

[13] We are grateful to Professor Craig W. Holden for kindly sharing their `SAS` codes.

flow over each active trading day; spot volatility is estimated by fixing $k = 5$ (i.e, every 5-minutes, so that would be $390/5 = 78$ locally estimated spot volatilities in each trading day and hence 78 locally quantified private information).

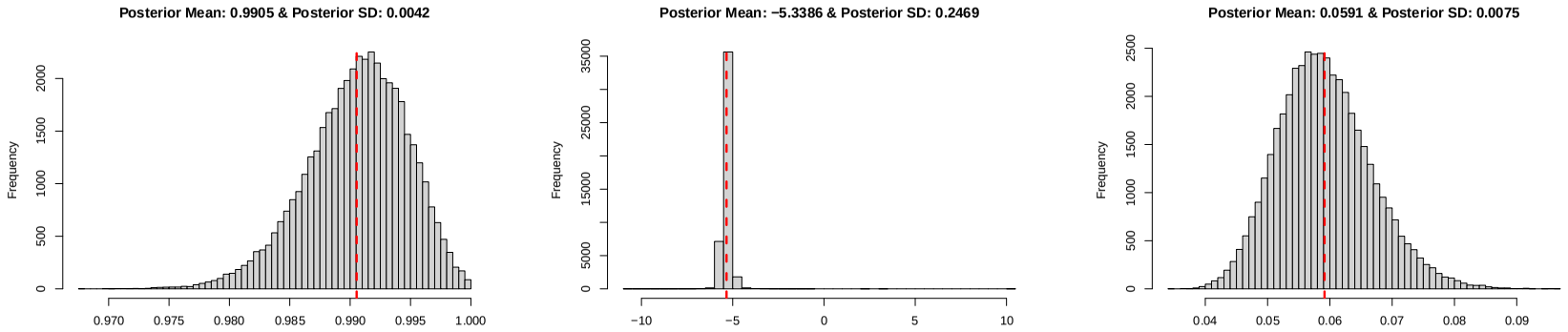
3.7 Conclusion

One of the main contributions from BLL2021QE is to establish the fixed-k inference theory for time-varying spot volatility. In this chapter, we build several parametric models for spot volatility in high frequency based on the fix-k theory. All our models can be cast into a nonlinear non-Gaussian state space form. We then develop Bayesian techniques to estimate alternative model specifications, extract spot volatility, and compare alternative models. Simulation studies show that the proposed Bayesian methods work well. To empirically demonstrate how our proposed method works in practice, we apply it to S&P 500 index ETF, Apple Inc. stock, and CSI 300 Index futures respectively. We first extract spot volatility. We then discuss how volatility is connected with the strategic value of information for the informed trader (i.e. insider) in the financial market more in detail.

Figures and Tables

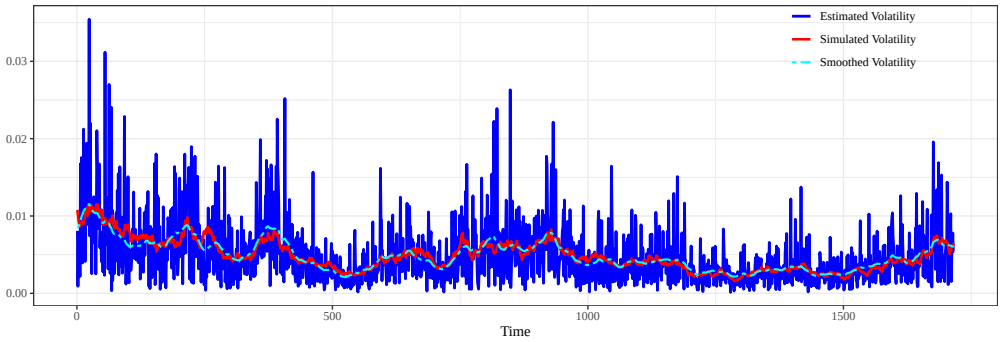
Figure 3.1

(a) ϕ (b) μ (c) $\sqrt{\sigma_e^2}$



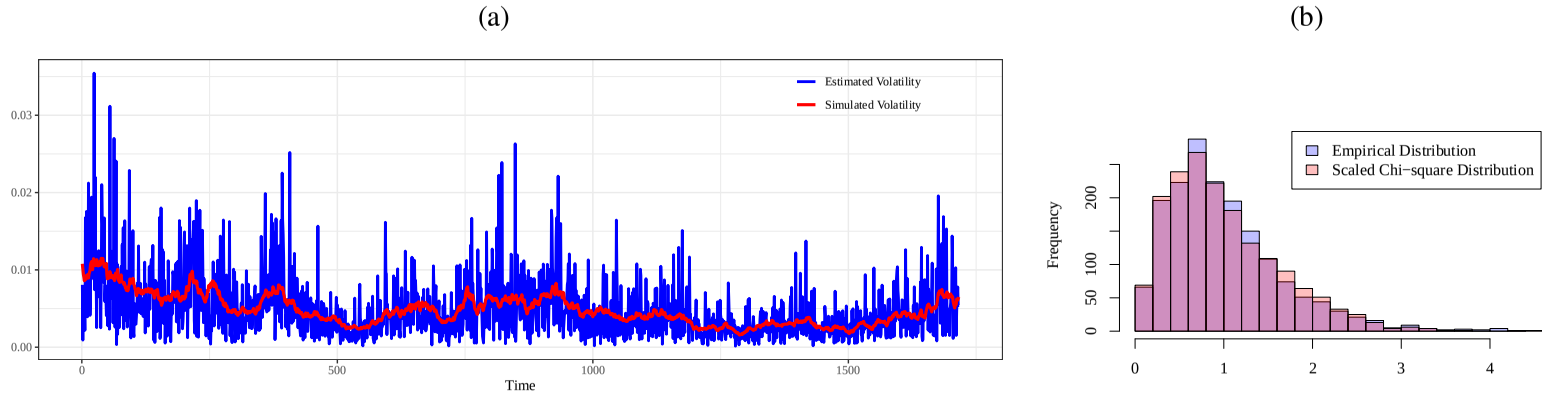
Note: Posterior Summary of Parameters for **Model 1**. Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.2



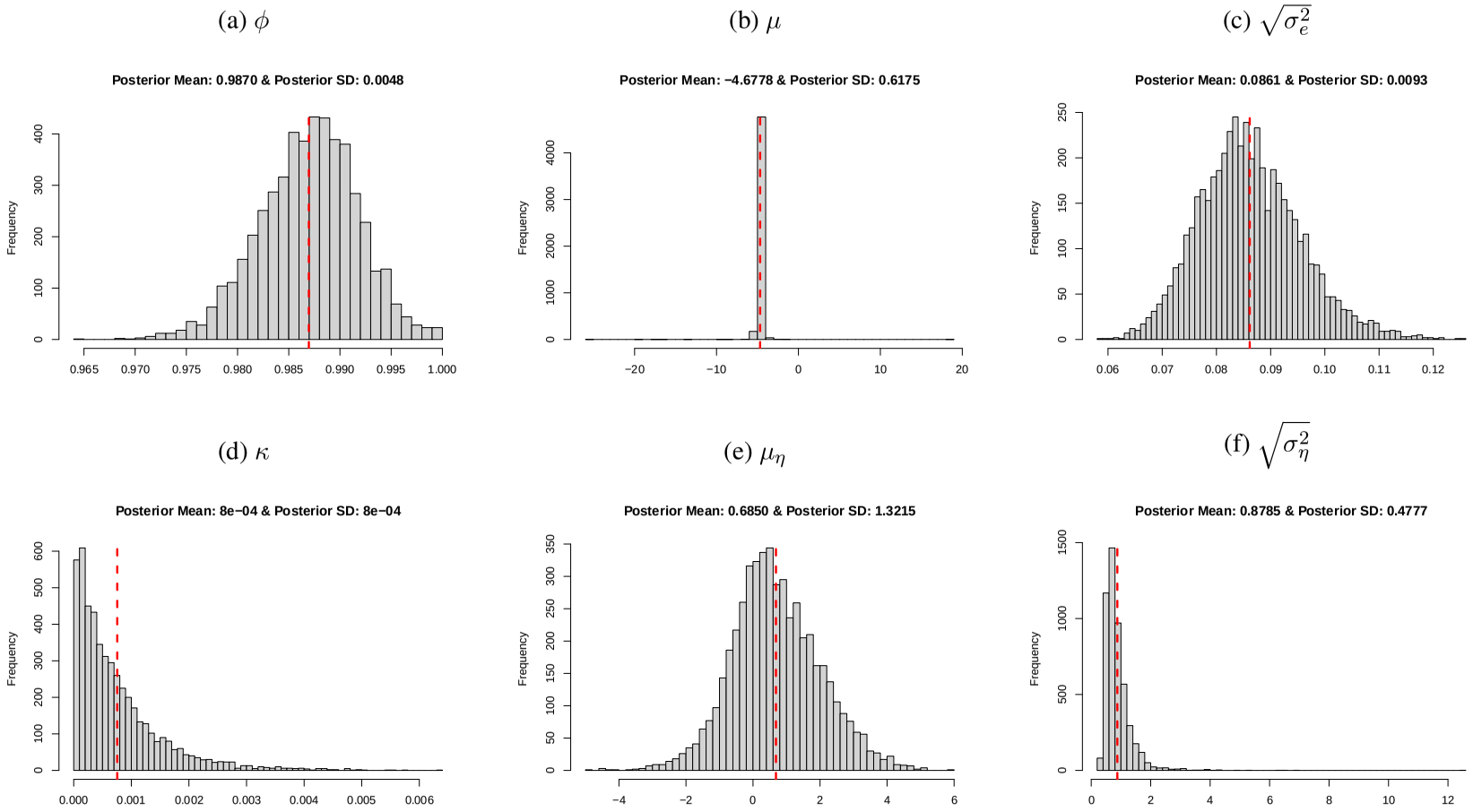
Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line), and the true simulated volatility (red solid line) based on the DGP of **Experiment 1** described in section 3.5. The sampling interval is $\Delta_n = 1/390$. The number of observations contained in each local estimation block is $k = 5$. Thus, for every 5 minutes we obtain the corresponding locally estimated volatility. Besides, for this sampling scheme we have $390/5 = 78$ local estimation blocks per day. For this experiment, we run totally 1000000 MCMC loops with the initial 100000 loops as burn-in samplings. Every 20 MCMC samples are saved for posterior analysis.

Figure 3.3



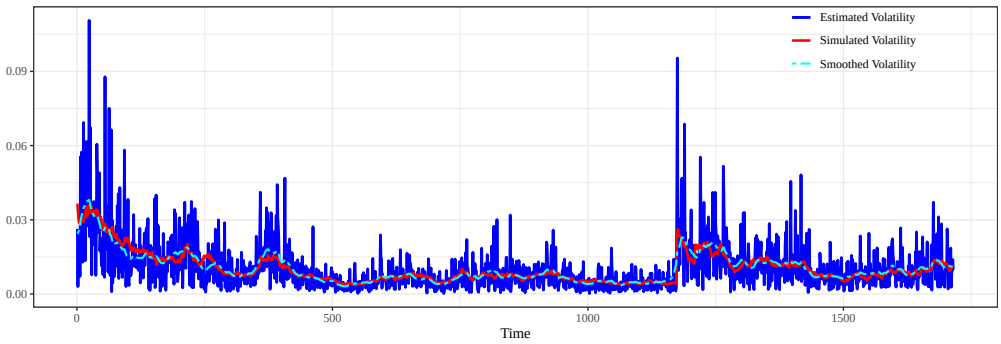
Note: This is an auxiliary figure for the demonstration purpose in companion with Figure 3.2. We jointly plot the estimated volatility from fixed-k inference (blue solid line) and the true simulated volatility (red solid line) based on DGP of **Experiment 1** in panel (a). Corresponding specifications are exactly the same as in the description of Figure 3.2. Panel (b) summarizes the histogram of gaps between the estimated volatility from fixed-k inference and the underlying true volatility process along with the histogram of random variables simulated from the scaled chi-square distribution.

Figure 3.4



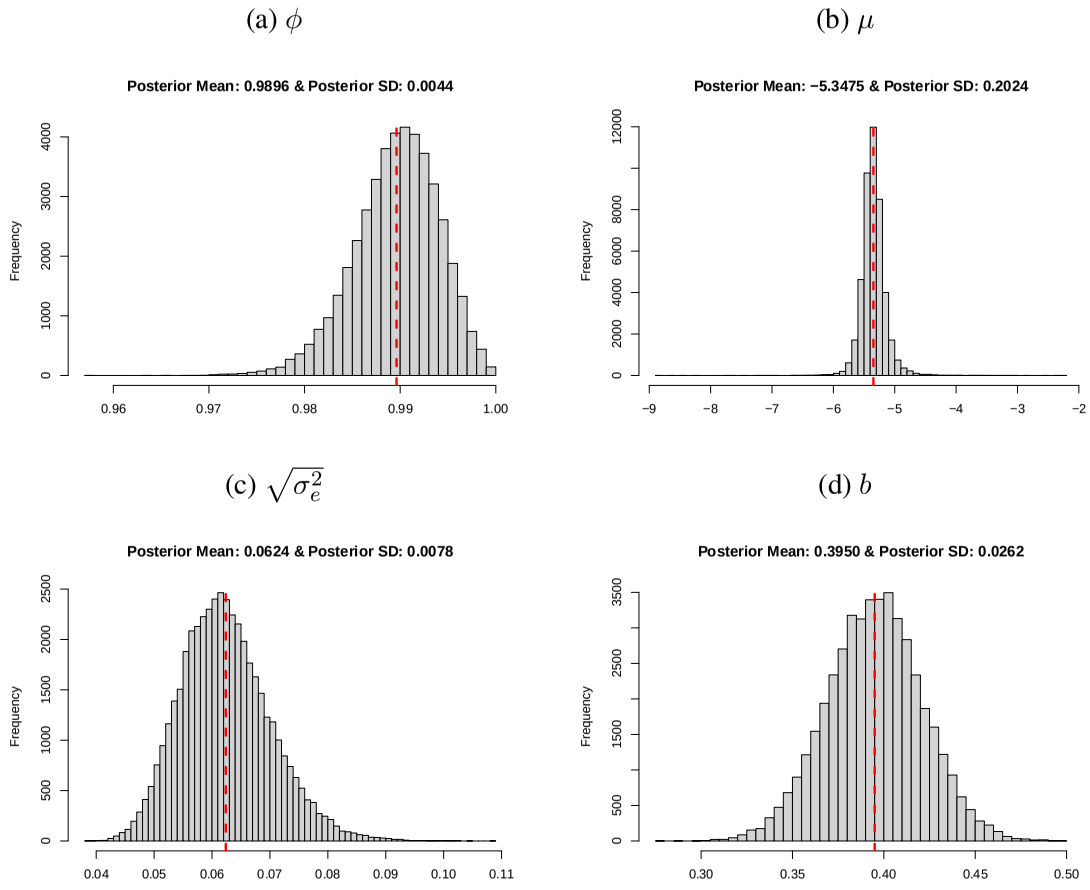
Note: Posterior Summary of Parameters for **Model 2**. Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.5



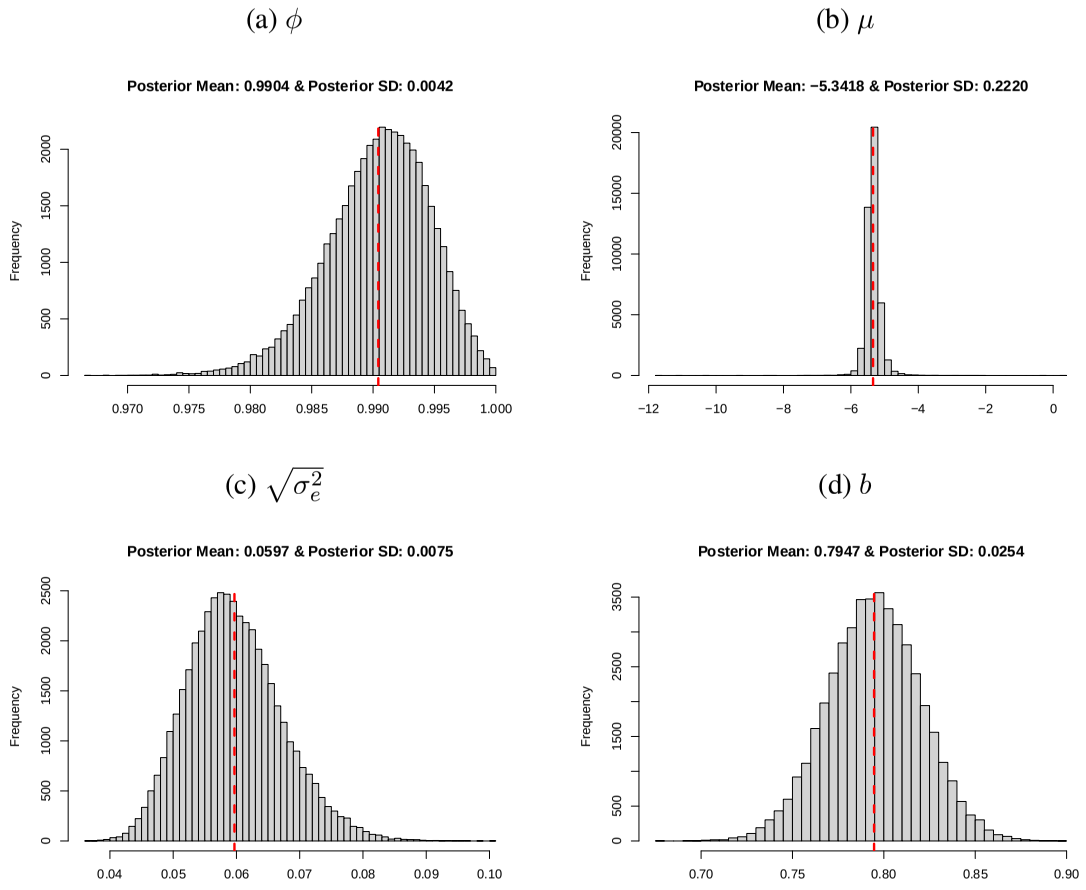
Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line). and true simulated volatility (red solid line) based on the DGP of **Experiment 2** described in section 3.5. The sampling interval is $\Delta_n = 1/390$. We use $k = 5$. For this experiment, we run totally 1100000 MCMC loops with the initial 100000 loops as burn-in samplings to be discarded. Every 100 samples are saved for posterior analysis.

Figure 3.6



Note: Posterior summary of parameters for **Model 3** (strong diurnal pattern). Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

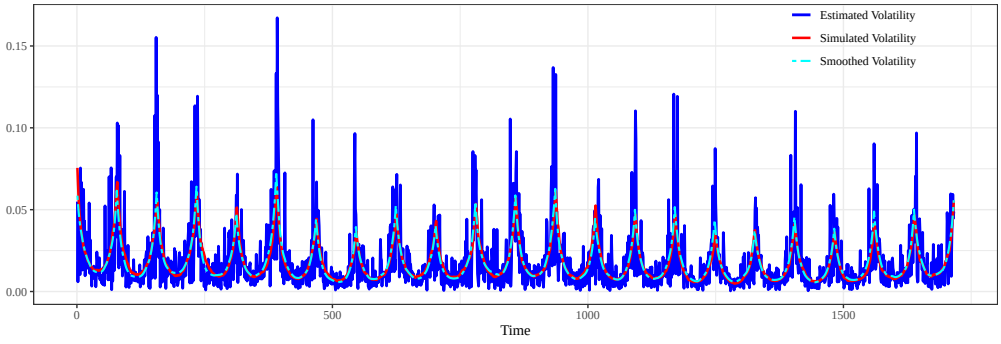
Figure 3.7



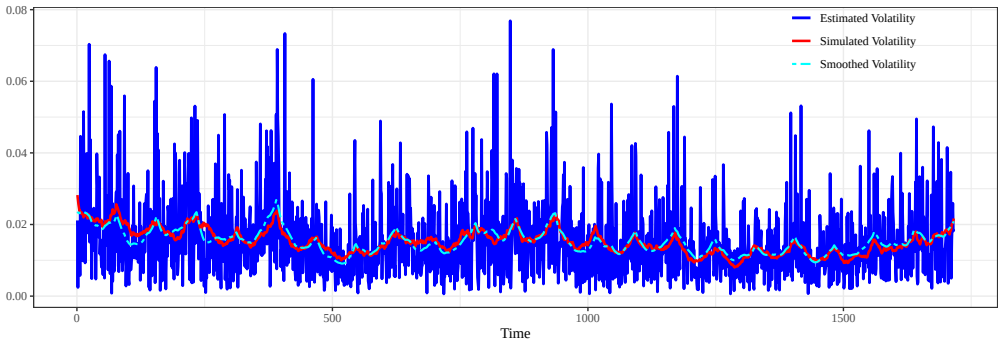
Note: Posterior summary of parameters for **Model 3** (weak diurnal pattern). Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.8

(a) Intraday-pattern parameter $b := 0.4$

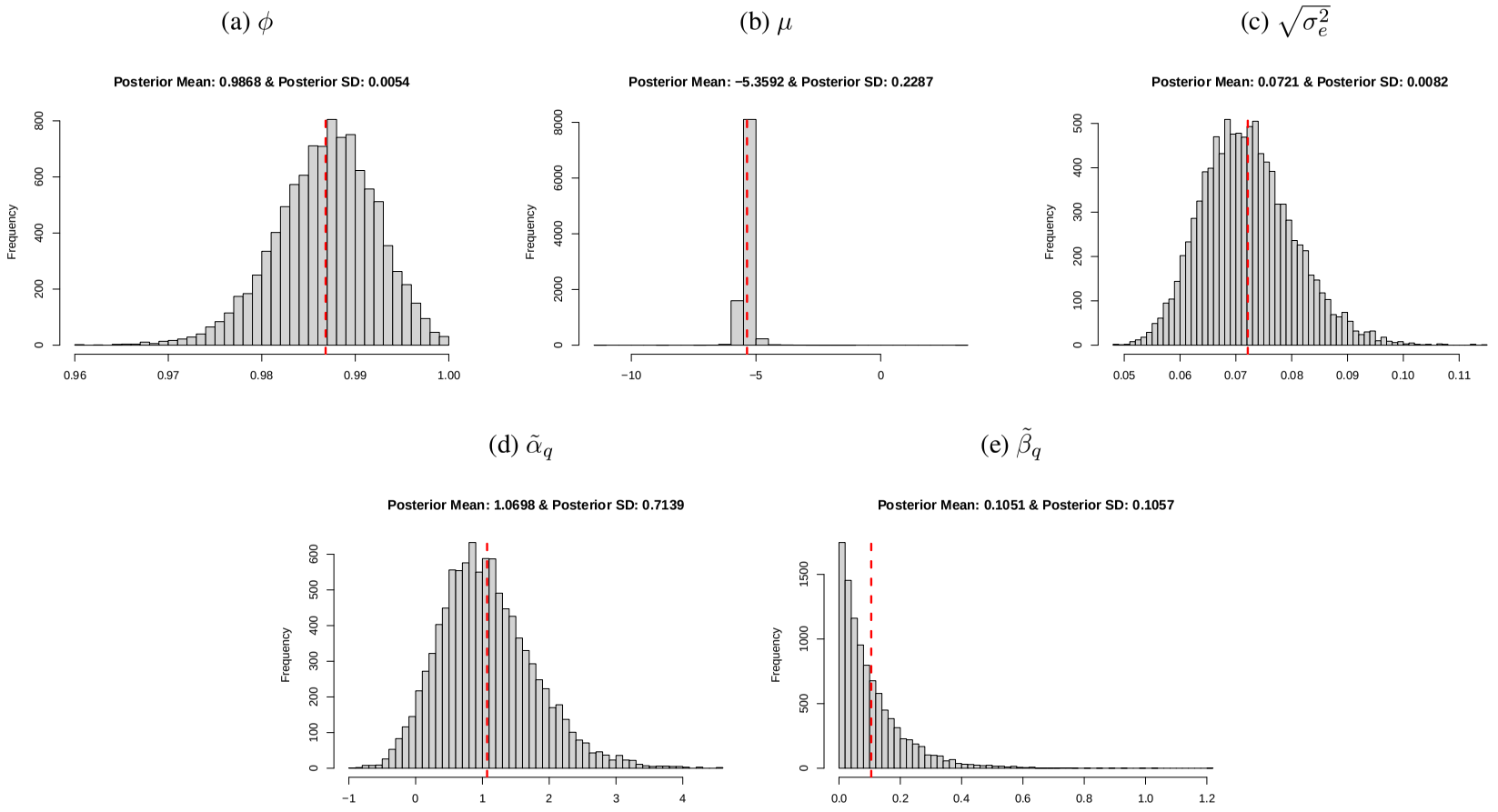


(b) Intraday-pattern parameter $b := 0.8$



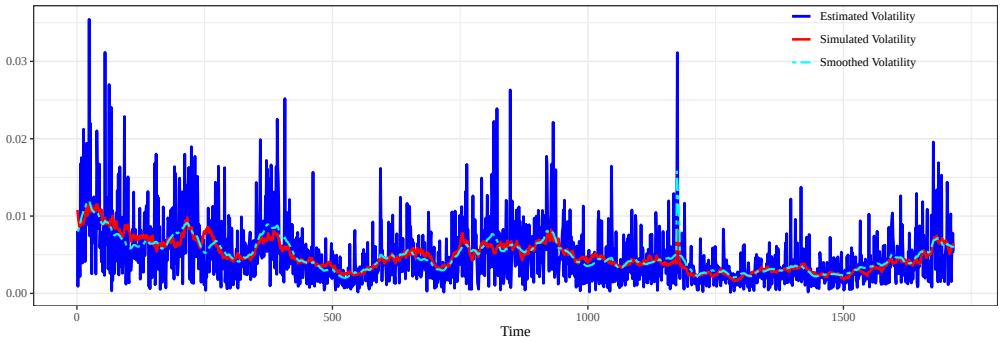
Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line), and the true simulated volatility (red solid line) based on the DGP of **Experiment 3** described in section 3.5. The sampling interval is $\Delta_n = 1/390$. We use $k = 5$.

Figure 3.9



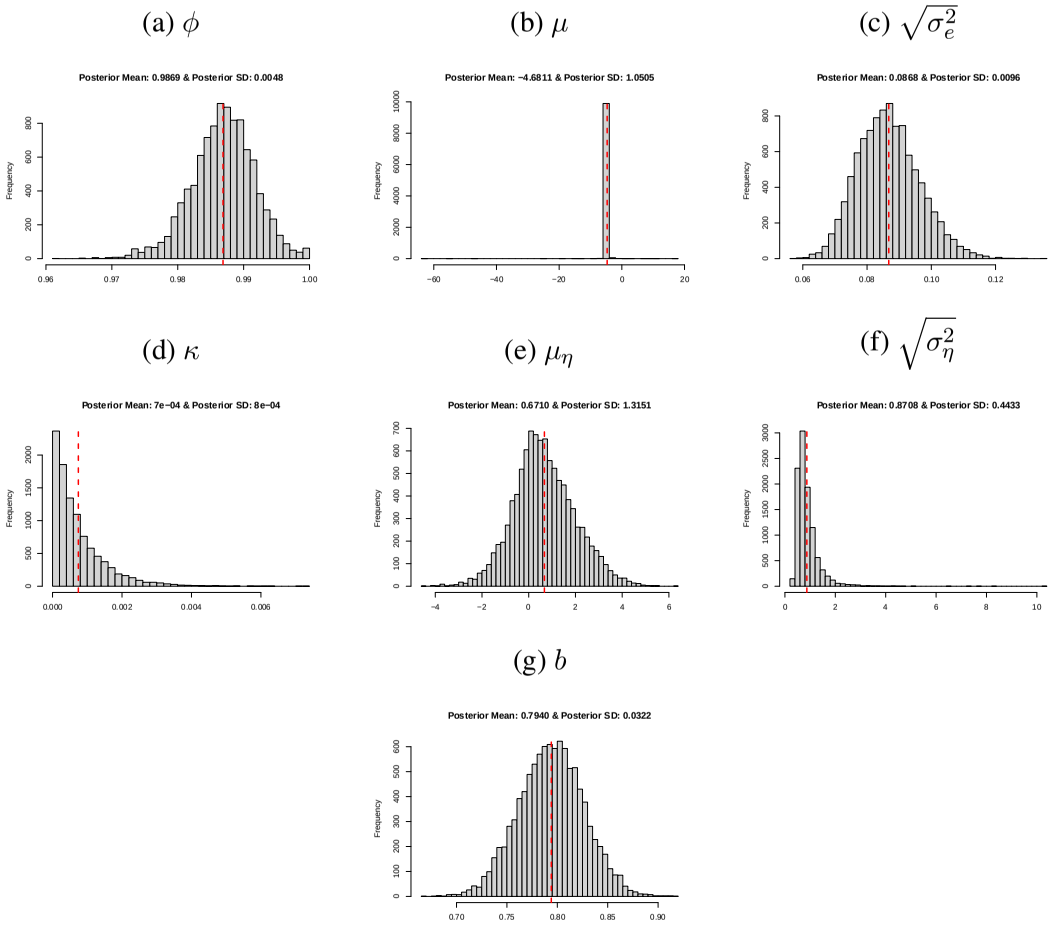
Note: Posterior summary of parameters for **Model 4**. Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.10



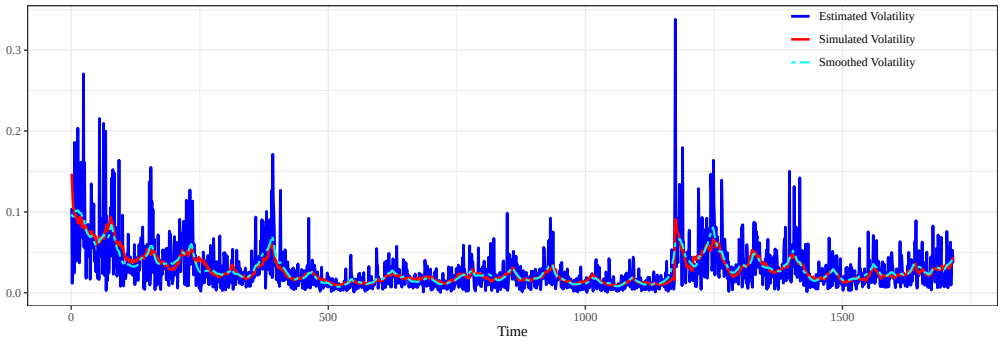
Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line), and the true simulated volatility (red solid line) based on the DGP of **Experiment 4** described in section 3.5. The sampling interval $\Delta_n = 1/390$. We use $k = 5$.

Figure 3.11



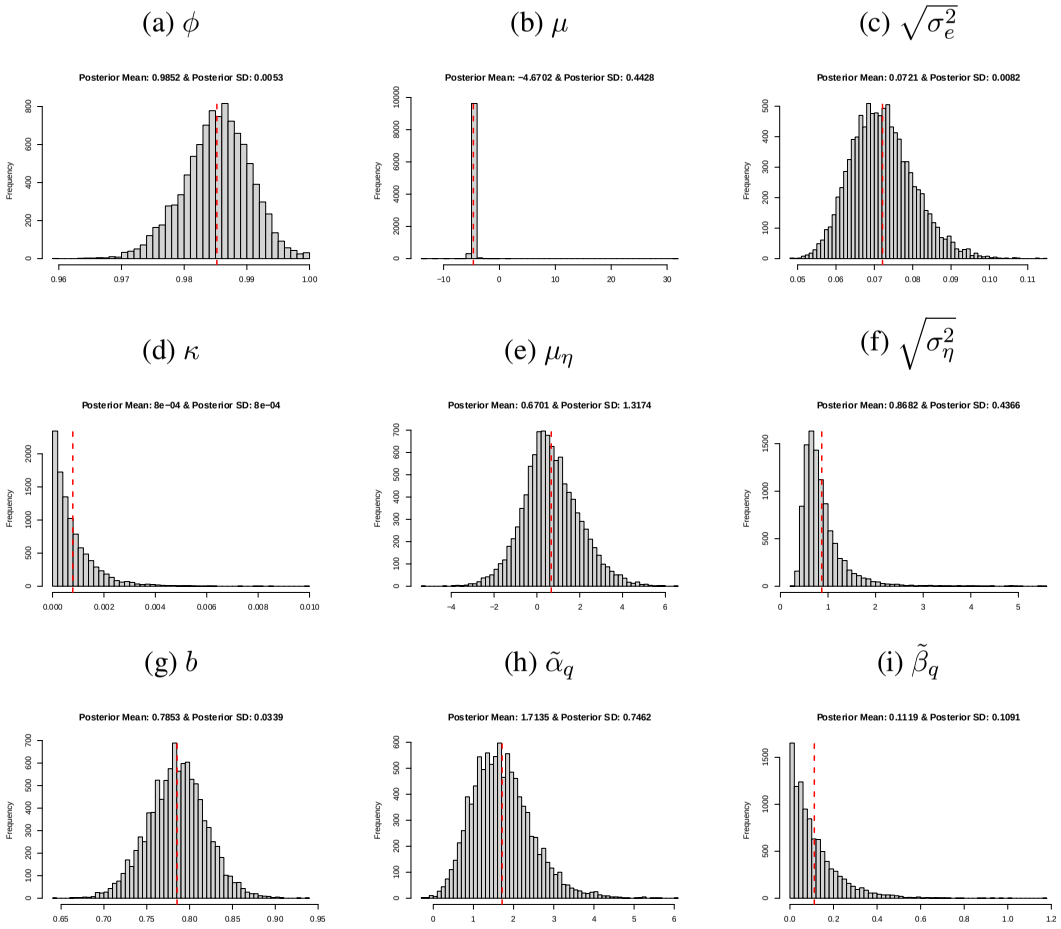
Note: Posterior summary of parameters for **Model 5**. Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.12



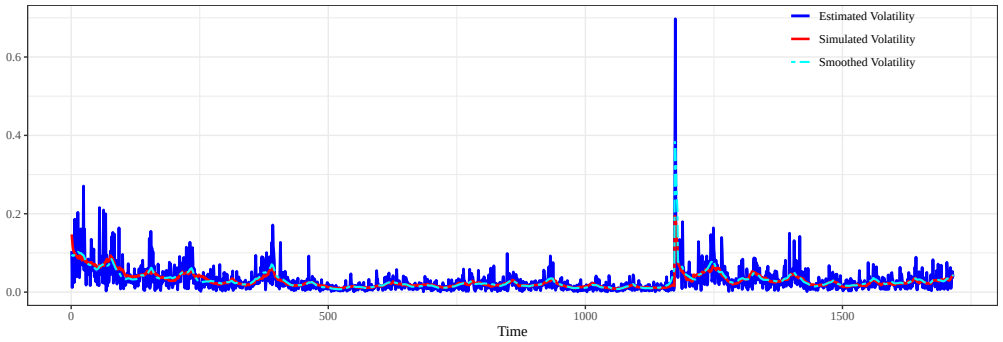
Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line), and the true simulated volatility (red solid line) based on the DGP of **Experiment 5** described in section 3.5. The sampling interval is $\Delta_n = 1/390$. We use $k = 5$.

Figure 3.13



Note: Posterior summary of parameters for **Model 6**. Vertical red dashed lines indicate the location of the posterior mean of the corresponding parameters.

Figure 3.14

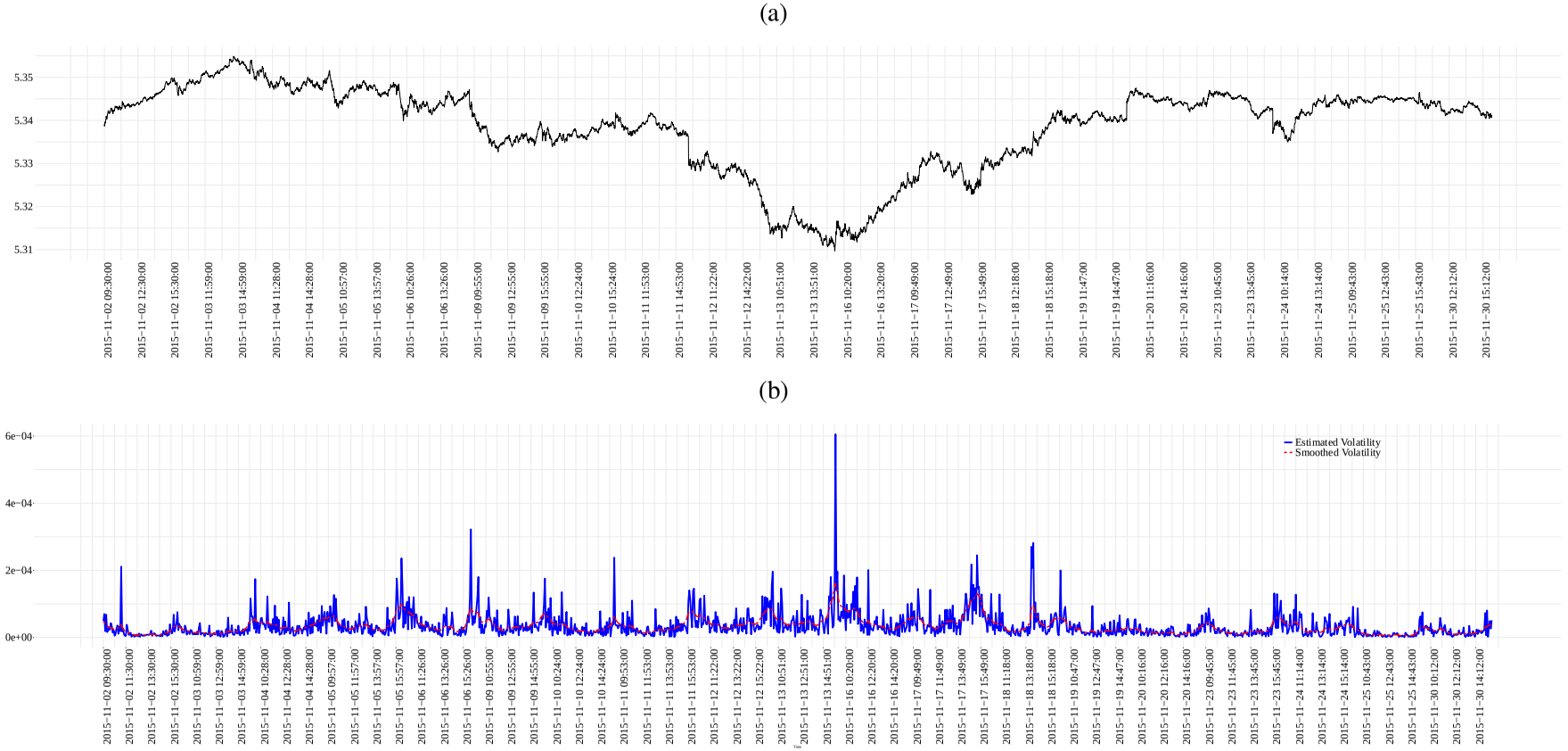


Note: In the figure above we jointly plot the estimated volatility from fixed-k inference (blue solid line), the smoothed volatility from MCMC (cyan dashed line), and the true simulated volatility (red solid line) based on the DGP of **Experiment 6** described in section 3.5. The sampling interval is $\Delta_n = 1/390$. We use $k = 5$.

Table 3.1

	DGP 1			**DGP 2**			**DGP 3**			**DGP 4**		
	DIC	$D(\hat{\boldsymbol{\vartheta}})$	p_D	DIC	$D(\hat{\boldsymbol{\vartheta}})$	p_D	DIC	$D(\hat{\boldsymbol{\vartheta}})$	p_D	DIC	$D(\hat{\boldsymbol{\vartheta}})$	p_D
Model 1	**-3310.18**	-3470.29	80.05	1222718.45	1209944.77	6386.84	-3243.91	-3745.30	250.69	-3243.12	-3444.81	100.85
Model 2	-3308.32	-3498.27	94.98	**-3210.47**	-3568.32	178.92	-3241.08	-3749.67	254.30	-3243.14	-3469.41	113.13
Model 3	-3297.71	-3473.93	88.11	56578.51	54348.87	1114.82	**-3329.11**	-3470.51	70.70	-3242.52	-3446.71	102.10
Model 4	–	–	–	–	–	–	–	–	–	**-3272.01**	-3496.99	112.49

Figure 3.15

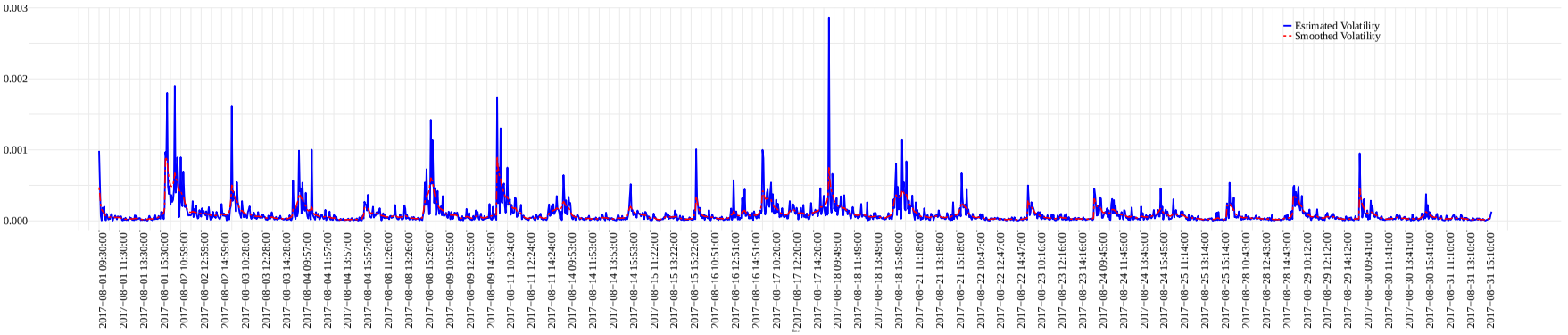


Note: In the figure above we report empirical results using price of S&P 500 index ETF sampled at the one-minute frequency. Panel (a) plots the log-price level of S&P 500 index. Panel (b) plots the nonparametric estimates of spot volatility (blue solid line) and the smoothed volatility estimate (red dashed line) using our Bayesian techniques. Results demonstrated in this figure correspond to S&P 500 index ETF within November 2015.

Figure 3.16

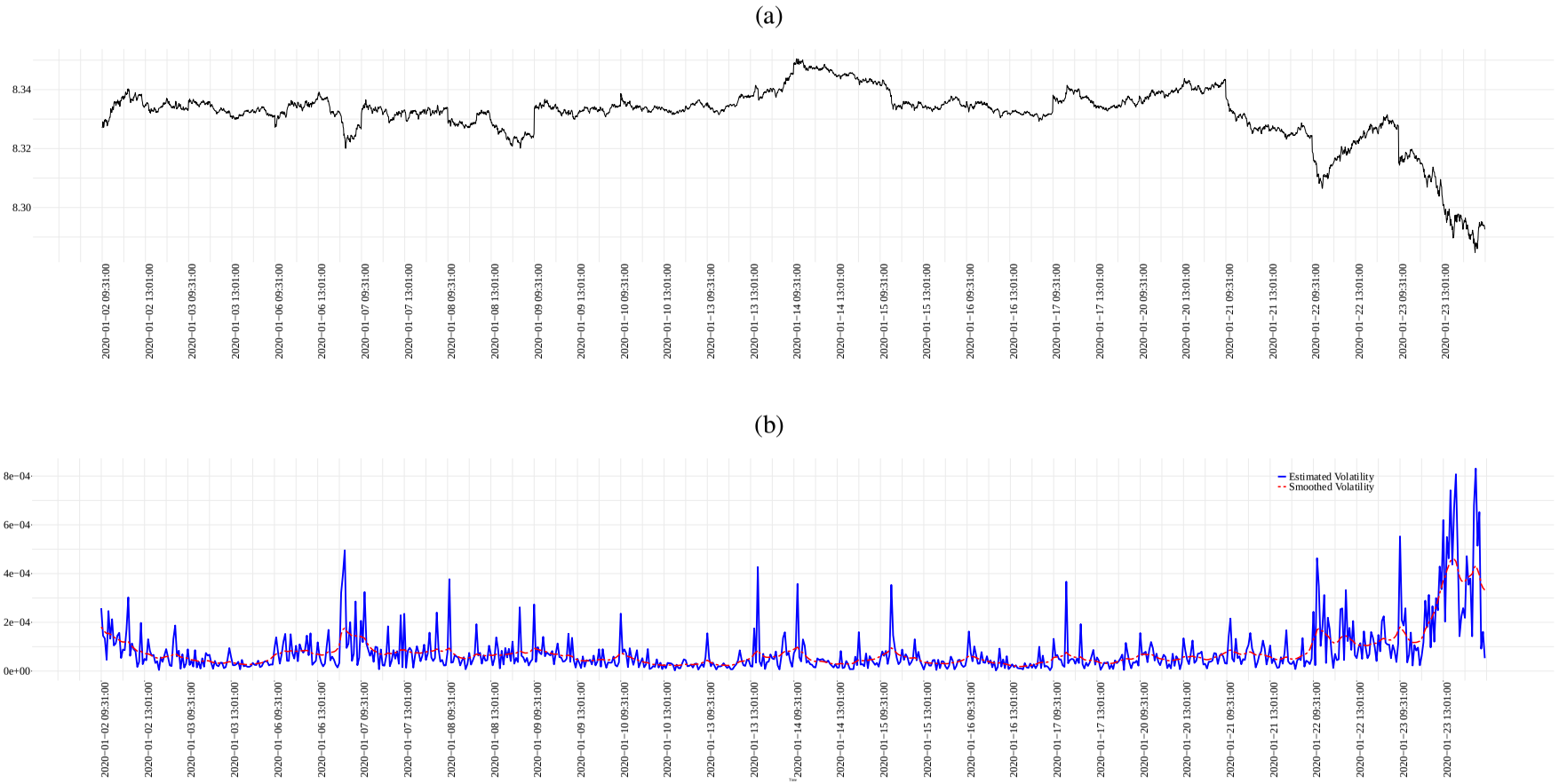
(a)

(b)



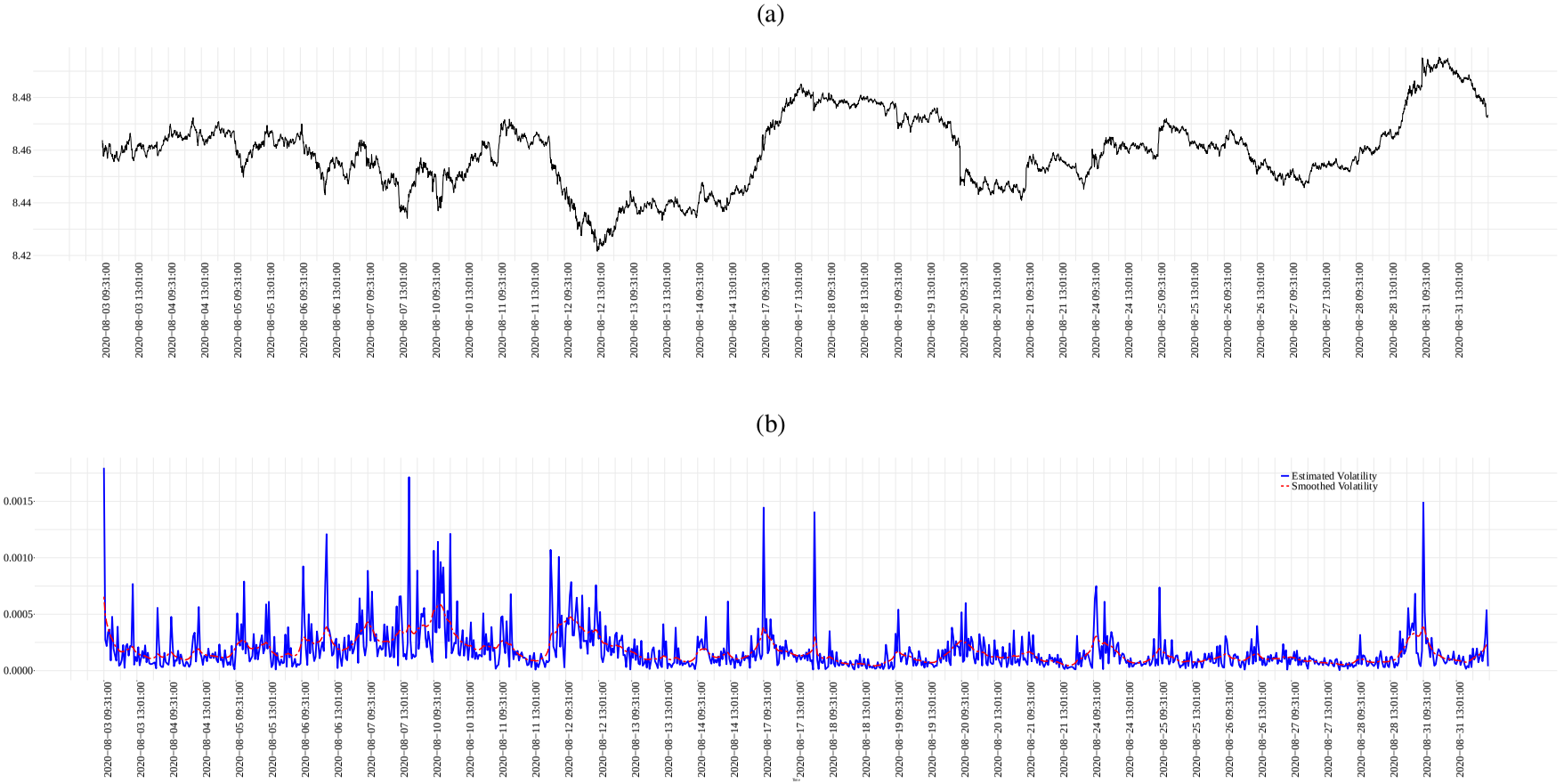
Note: In the figure above we report the empirical results using the one-minute price data of Apple Inc. Panel (a) plots the log-price of Apple Inc. Panel (b) plots the nonparametric estimates of spot volatility (blue solid line) and the smoothed volatility estimate (red dashed line) using our Bayesian techniques. Results demonstrated in this figure correspond to Apple stock within August 2017.

Figure 3.17



Note: In the figure above we report the empirical results using the one-minute price data of CSI 300 index futures in January 2020. Panel (a) plots the log-price level of CSI 300 index futures. Panel (b) plots the nonparametric estimates of spot volatility (blue solid line) and the smoothed volatility estimate (red dashed line) using our Bayesian techniques. Results demonstrated in this figure correspond to CSI 300 index futures within January 2020.

Figure 3.18



Note: In the figure above we report the empirical results using the one-minute price data of CSI 300 index futures in August 2020. Panel (a) plots the log-price level of CSI 300 index futures. Panel (b) plots the nonparametric estimates of spot volatility (blue solid line) and the smoothed volatility estimate (red dashed line) using our Bayesian techniques. Results demonstrated in this figure correspond to CSI 300 index futures within August 2020.

Table 3.2

Date	Model	DIC	$D(\boldsymbol{\vartheta})$	p_D
January 2015	1	-2727.40	-3282.86	277.73
February 2015	2	-2066.28	-2907.14	420.43
March 2015	1	-2498.42	-3200.40	350.99
April 2015	2	-2133.17	-2876.71	371.77
May 2015	2	-2250.85	-3043.06	396.10
June 2015	2	-2283.65	-3422.16	569.25
July 2015	1	-2257.16	-2885.81	314.33
August 2015	1	-2830.98	-3445.31	307.16
September 2015	3	-2982.72	-3332.93	175.10
October 2015	1	-3080.86	-3598.36	258.75
November 2015	1	-2606.25	-3096.89	245.32
December 2015	1	-2849.65	-3422.49	286.42

Note: Model selected for S&P 500 Index ETF in 2015 across **Model 1-3**.

Table 3.3

Date	Model	DIC	$D(\boldsymbol{\vartheta})$	p_D
January 2017	2	-1714.08	-2551.75	418.83
February 2017	2	-1233.57	-1793.62	280.03
March 2017	2	-1821.20	-2678.98	428.89
April 2017	2	-1631.22	-2231.49	300.14
May 2017	2	-2631.12	-3648.79	508.83
June 2017	2	-2731.98	-3771.19	519.61
July 2017	2	-2327.21	-3295.71	484.25
August 2017	2	-2709.67	-3644.26	467.30
September 2017	2	-2318.04	-3247.47	464.72
October 2017	2	-2704.97	-3748.87	521.95
November 2017	2	-2342.86	-3292.15	474.65
December 2017	1	-2377.74	-3633.71	627.98

Note: Model selected for Apple Inc. Stock Price in 2017 across **Model 1-3**.

Table 3.4

Date	Model	DIC	$D(\boldsymbol{\vartheta})$	p_D
January 2020	1	-1367.30	-1595.93	114.32
February 2020	1	-1747.86	-2071.24	161.69
March 2020	1	-1809.46	-2192.64	191.59
April 2020	1	-1711.00	-2061.63	175.32
May 2020	2	-1568.94	-1856.89	143.98
June 2020	2	-1713.23	-2274.82	280.79
July 2020	1	-2126.11	-2472.58	173.24
August 2020	2	-1785.07	-2146.15	180.54
September 2020	1	-1874.93	-2230.75	177.91
October 2020	2	-1298.32	-1720.46	211.07
November 2020	2	-1749.23	-2343.43	297.10
December 2020	2	-1797.38	-2286.43	244.53

Note: Model selected for CSI 300 Index in 2020 across **Model 1-3**.

Table 3.5

	ϕ	μ	$\sqrt{\sigma_e^2}$	κ	μ_η	$\sqrt{\sigma_\eta^2}$	b
January 2015	0.9314	-10.6814	0.2327	0.0023	0.4756	0.7925	0.6505
	0.0145	0.0920	0.0213	0.0031	1.0793	0.3836	0.0592
February 2015	0.8630	-11.8068	0.4635	0.0127	0.9162	0.7700	0.6232
	0.0250	0.1109	0.0367	0.0133	0.7617	0.2799	0.0843
March 2015	0.9205	-11.7109	0.2670	0.0029	1.2659	0.8791	0.6716
	0.0163	0.0907	0.0268	0.0031	0.8765	0.3713	0.0629
April 2015	0.8234	-11.8096	0.4126	0.0085	0.3941	0.7304	0.5651
	0.0310	0.0742	0.0348	0.0115	0.8503	0.3124	0.0658
May 2015	0.8987	-12.1515	0.3722	0.0054	1.2657	0.9002	0.5700
	0.0189	0.1060	0.0310	0.0061	0.9199	0.3801	0.0790
June 2015	0.7930	-11.9509	0.4837	0.0155	1.4376	0.8143	0.6031
	0.0321	0.0748	0.0395	0.0109	0.5847	0.2466	0.0621
July 2015	0.9416	-11.9417	0.2564	0.0025	1.5408	0.9219	0.6165
	0.0137	0.1257	0.0259	0.0024	0.9558	0.3958	0.0708
August 2015	0.9846	-10.9692	0.2260	0.0025	1.0233	0.8391	0.3343
	0.0052	0.7358	0.0179	0.0025	0.8693	0.3647	0.0694
September 2015	0.9659	-10.4311	0.1523	0.0012	0.5647	0.8383	0.5911
	0.0083	0.1180	0.0153	0.0014	1.1843	0.4297	0.0473
October 2015	0.9555	-11.1649	0.1888	0.0022	0.7109	0.7978	0.8046
	0.0099	0.1095	0.0180	0.0021	0.8601	0.3648	0.0532
November 2015	0.9414	-11.4897	0.2135	0.0018	0.4866	0.8158	0.6810
	0.0138	0.1020	0.0215	0.0024	1.1374	0.4057	0.0600
December 2015	0.9698	-11.0310	0.1868	0.0021	0.9683	0.8345	0.7201
	0.0080	0.2104	0.0187	0.0019	0.9039	0.3597	0.0570

Note: Posterior summary of parameters in **Model 5** for S&P 500 ETF

Table 3.6

	ϕ	μ	$\sqrt{\sigma_e^2}$	κ	μ_η	$\sqrt{\sigma_\eta^2}$	b
January 2017	0.8336	-11.6393	0.4606	0.0175	1.6388	0.7839	0.4773
	0.0283	0.0999	0.0374	0.0099	0.5376	0.2390	0.0799
February 2017	0.8363	-11.5400	0.4528	0.0164	1.2551	0.8021	0.5053
	0.0311	0.1226	0.0438	0.0149	0.7643	0.2800	0.1041
March 2017	0.8005	-11.4342	0.4771	0.0153	1.1766	0.7884	0.5941
	0.0296	0.0846	0.0360	0.0136	0.6884	0.2618	0.0697
April 2017	0.8148	-11.2624	0.4239	0.0132	0.6342	0.7379	0.5785
	0.0377	0.0872	0.0408	0.0149	0.7696	0.2943	0.0730
May 2017	0.8907	-11.2462	0.4094	0.0119	0.9838	0.7432	0.3568
	0.0196	0.1115	0.0307	0.0103	0.6161	0.2369	0.0768
June 2017	0.9145	-10.8191	0.3869	0.0109	1.0729	0.7394	0.3435
	0.0152	0.1269	0.0300	0.0085	0.5542	0.2337	0.0785
July 2017	0.8456	-11.1228	0.4563	0.0148	0.6327	0.6979	0.4765
	0.0241	0.0975	0.0312	0.0153	0.6697	0.2468	0.0749
August 2017	0.8947	-10.6961	0.3686	0.0083	0.7524	0.7393	0.4341
	0.0185	0.1015	0.0273	0.0088	0.7337	0.2872	0.0711
September 2017	0.9123	-10.8906	0.4057	0.0110	1.3838	0.8018	0.4142
	0.0165	0.1400	0.0305	0.0077	0.5950	0.2513	0.0876
October 2017	0.8673	-11.3134	0.4316	0.0108	0.7922	0.7307	0.4636
	0.0177	0.0951	0.0261	0.0110	0.7128	0.2624	0.0742
November 2017	0.9202	-11.0970	0.4336	0.0141	1.2448	0.7748	0.2424
	0.0172	0.1767	0.0323	0.0100	0.5673	0.2356	0.1010
December 2017	0.8127	-11.1758	0.5669	0.0078	1.9642	3.1212	0.6861
	0.0207	0.0839	0.0286	0.0038	0.9528	0.8653	0.0767

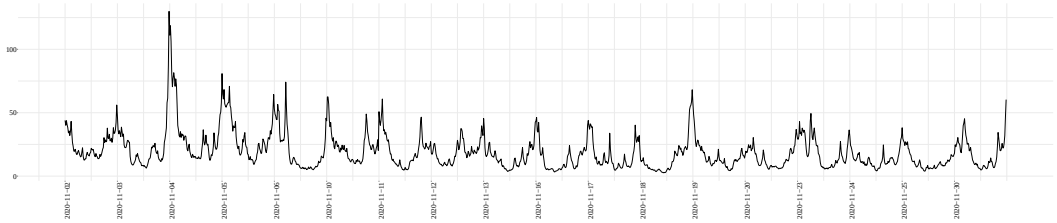
Note: Posterior summary of parameters in **Model 5** for Apple Inc. Stock.

Table 3.7

	ϕ	μ	$\sqrt{\sigma_e^2}$	κ	μ_η	$\sqrt{\sigma_\eta^2}$	b
January 2020	0.9461	-10.6461	0.2139	0.0031	0.5133	0.8219	0.8715
	0.0207	0.2011	0.0340	0.0040	1.1557	0.4347	0.0583
February 2020	0.8404	-9.9389	0.3371	0.0063	1.9077	2.9523	0.9678
	0.0318	0.0763	0.0260	0.0033	1.0855	0.9737	0.0276
March 2020	0.9532	-9.3718	0.2260	0.0047	1.1011	0.8282	0.9262
	0.0131	0.1693	0.0242	0.0040	0.7615	0.3338	0.0455
April 2020	0.8893	-10.7519	0.2662	0.0043	0.5160	0.7759	0.8996
	0.0315	0.0872	0.0370	0.0058	0.9981	0.3499	0.0505
May 2020	0.9174	-11.0502	0.2159	0.0029	0.4834	0.8201	0.8263
	0.0208	0.0978	0.0280	0.0036	1.1455	0.4057	0.0558
June 2020	0.7761	-11.0128	0.3640	0.0118	0.7256	0.7275	0.7924
	0.0339	0.0677	0.0300	0.0112	0.6980	0.2619	0.0534
July 2020	0.9430	-9.2816	0.2159	0.0023	0.5143	0.8143	0.9107
	0.0142	0.1233	0.0210	0.0030	1.1364	0.3920	0.0469
August 2020	0.9091	-9.8210	0.2396	0.0031	0.4617	0.8025	0.8289
	0.0291	0.0936	0.0350	0.0042	1.1146	0.3946	0.0547
September 2020	0.9013	-10.3231	0.2362	0.0031	0.5099	0.8049	0.8579
	0.0250	0.0814	0.0287	0.0040	1.0755	0.3928	0.0509
October 2020	0.7939	-10.6007	0.3586	0.0114	0.7909	0.7751	0.8349
	0.0485	0.0794	0.0439	0.0123	0.7867	0.3156	0.0620
November 2020	0.7933	-10.6782	0.3498	0.0116	1.1399	0.7755	0.8509
	0.0393	0.0655	0.0344	0.0087	0.5798	0.2553	0.0520
December 2020	0.8011	-10.7099	0.3386	0.0072	0.5606	0.7650	0.9509
	0.0467	0.0626	0.0405	0.0094	0.8860	0.3544	0.0352

Note: Posterior summary of parameters in **Model 5** for CSI 300 Index Futures.

Figure 3.19



Note: In the figure above, we report the estimated private information value associated with insider trading based on the estimated spot volatility from **Model 5**. Results demonstrated in this figure correspond to November 2020.

Chapter 4

Appendices

4.1 Anomaly variables used in Chinese stock market

We summarize the main cross-sectional equity characteristics (firm-level characteristics) used in the empirical analysis of Chapter 1 and Chapter 2 in this section. We follow the cutting-edge data-cleaning routine proposed in Chen and Zimmermann (2020) and Jensen, Kelly, and Pedersen (2022) to construct following 123 equity characteristics in Chinese stock market. In each item, we list the brief descriptions of corresponding anomaly variables with the acronym (in typewriter format collected in parenthesis) and in general the category (in bold collected square brackets) it belongs to in finance and accounting literature. We also list the corresponding literature that initially proposes equity characteristics. The corresponding information and description inherit directly from Jensen, Kelly, and Pedersen (2022) and readers should refer to documentation released along with Jensen, Kelly, and Pedersen (2022) for more about construction details.

1. Firm age (`age`) **[Low Leverage]**, Jiang, Lee, and Zhang (2005).

2. Liquidity of book assets (`aliq_at`) **[Investment]**, Ortiz-Molina and Phillips (2014).

3. Liquidity of market assets (`aliq_mat`) **[Low leverage]**, Ortiz-Molina and

Phillips (2014).

4. Amihud measure (ami_126d) **[Size]**, Amihud (2002).

5. Book leverage (at_be) **[Low leverage]**, Fama and French (1992).

6. Asset growth (at_gr1) **[Investment]**, Cooper, Gulen, and Schill (2008).

7. Assets-to-market (at_me) **[Value]**, Fama and French (1992).

8. Capital turnover (at_turnover) **[Quality]**, Haugen and Baker (1996).

9. Change in common equity (be_gr1a) **[Investment]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

10. Book-to-market equity be_me **[Value]**, Rosenberg, Reid, and Lanstein (1985).

11. Market beta (beta_60m) **[Low Risk]**, Fama and Macbeth (1973).

12. Dimson beta (beta_dimson_21d) **[Low Risk]**, Fowler and Rorke (1983).

13. Frazzini-Pedersen market beta (betabab_1260d) **[Low Risk]**, Frazzini and Pedersen (2014).

14. Downside beta (betadown_252d) **[Low Risk]**, Ang, Hodrick, Xing, and Zhang (2006).

15. Book-to-market enterprise value (bev_mev) **[Value]**, Penman, Richardeson, and Tuna (2007).

16. 21 Day high-low bid-ask spread (bidaskhl_21d) **[Low Leverage]**, Corwin and Schultz (2012).

17. Cash-to-assets (cash_at) **[Low Leverage]**, Palazzo (2012).

18. Net stock issues (chcsho_12m) **[Value]**, Pontiff and Woodgate (2008).

19. Change in current operating assets (coa_gr1a) **[Investment]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

20. Change in current liabilities (col_gr1a) **[Investment]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

21. Cash-based operating profits-to-book assets (cop_at) **[Quality]**, Haugen and Baker (1996).

22. Cash-based operating profits-to lagged book assets (cop_atl1) **[Quality]**, Ball, Gerakos, Linnainmaa, and Nikolaev (2016).

23. Market correlation (corr_1260d) **[Seasonality]**, Asness, Frazzini, Gormsen, and Pedersen (2020).

24. Coskewness (coskew_21d) **[Seasonality]**, Harvey and Siddique (2000).

25. Change in current operating working capital (cowc_gr1a) **[Accruals]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

26. Net debt issuance (dbnetis_at) **[Net debt issuance]**, Bradshaw, Richardson, and Sloan (2006).

27. Growth in book debt (3 years) (debt_gr3) **[Debt Issuance]**, Lyandres, Sun, and Zhang (2008).

28. Debt-to-market (debt_me) **[Value]**, Bhandari (1988).

29. Change gross margin minus change sales (dgp_dsale) **[Quality]**, Abarbanell and Bushee (1998).

30. Dividend yield (div12m_me) **[Value]**, Litzenberger and Ramaswamy (1979).

31. Dollar trading volume (dolvol_126d) **[Size]**, Brennan, Chordia, and Subrahmanyam (1998).

32. Coefficient of variation for dollar trading volume (dolvol_var_126d) **[Profitability]**, Chordia, Subrahmanyam, and Anshuman (2001).

33. Change sales minus change inventory (dsale_dinv) **[Profit Growth]**, Abarbanell and Bushee (1998).

34. Change sales minus change receivables (dsale_drec) **[Profit Growth]**, Abarbanell and Bushee (1998).

35. Change sales minus change SG&A (dsale_dsga) **[Profit Growth]**, Abarbanell and Bushee (1998).

36. Earnings variability (earnings_variability) **[Low Risk]**, Francis, LaFond, Olsson, and Schipper (2004).

37. Return on net operating assets (ebit_bev) **[Profitability]**, Soliman (2008).

38. Profit margin (ebit_sale) **[Profit Growth]**, Soliman (2008).

39. Ebitda-to-market enterprise value (ebitda_mev) **[Value]**, Loughran and Wellman (2011).

40. Equity duration (eq_dur) **[Value]**, Dechow, Sloan, and Soliman (2004).

41. Equity net payout (eqnpo_12m) **[Value]**, Daniel and Titman (2006).

42. Pitroski F-score (f_score) **[Profitability]**, Piotroski (2000).

43. Change in financial liabilities (fnl_gr1a) **[Debt Issuance]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

44. Gross profits-to-assets (gp_at) **[Quality]**, Novy-Marx (2013).

45. Gross profits-to-lagged assets (gp_atl1) **[Quality]**, Novy-Marx (2013).

46. Intrinsic-value (intrinsic_value) **[Value]**, Frankel and Lee (1998).

47. Inventory growth (inv_gr1) **[Investment]**, Belo and Lin (2012).

48. Inventory change (inv_gr1a) **[Investment]**, Thomas and Zhang (2002).

49. Idiosyncratic skewness from the CAPM (iskew_capm_21d) **[Skewness]**, Bali, Engle, and Murray (2016).

50. Idiosyncratic volatility from the CAPM (21 days) (ivol_capm_21d) **[Low Risk]**, Ali, Hwang, and Trombley (2003).

51. Idiosyncratic volatility from the CAPM (252 days) (ivol_capm_252d) **[Low Risk]** Ali, Hwang, and Trombley (2003).

52. Change in long-term net operating assetsn (lnoa_gr1a) **[Investment]**, Fairfield, Whisenant, and Yohn (2003).

53. Change in long-term investments (lti_gr1a) **[Seasonality]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

54. Market equity (market_equity) **[Size]**, Banz (1981).

55. Mispricing factor: Management (mispricing_mgmt) **[Investment]**, Stambaugh and Yuan (2016).

56. Mispricing factor: Performance (mispricing_perf) **[Quality]**, Stambaugh and Yuan (2016).

57. Change in noncurrent operating assets (nroa_gr1a) **[Investment]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

58. Change in noncurrent operating liabilities (ncol_gr1a)) **[Debt Issuance]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

59. Net debt-to-price (netdebt_me) **[Low Leverage]**, Penman, Richardeson, and Tuna (2007).

60. Change in net financial assets (nfna_gr1a) **[Debt Issuance]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

61. Earnings persistence (ni_ar1) **[Debt Issuance]**, Francis, LaFond, Olsson, and Schipper (2004).

62. Return on equity (ni_be) **[Profitability]**, Haugen and Baker (1996).

63. Number of consecutive quarters with earnings increases (ni_inc8q) **[Quality]**, Barth, Elliott, and Finn (1999).

64. Earnings volatility (ni_ivol) **[Low Leverage]**, Francis, LaFond, Olsson, and Schipper (2004).

65. Earnings-to-price (ni_me) **[Value]**, Basu (1983).

66. Quarterly return on assets (niq_at) **[Quality]**, Balakrishnan, Bartov, and Faurel (2010).

67. Change in quarterly return on assets (niq_at_chg1) **[Profit Growth]**, Abarbanell and Bushee (1998).

68. Quarterly return on equity (niq_be) **[Profitability]**, Hou, Xue, and Zhang (2015).

69. Change in quarterly return on equity (niq_be_chg1) **[Profit Growth]**, Abarbanell and Bushee (1998).

70. Standardized earnings surprise (niq_su) **[Profit Growth]**, Foster, Olsen, and Shevlin (1984).

71. Change in net noncurrent operating assets (nncoa_gr1a) **[Investment]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

72. Net operating assets (noa_at) **[Debt Issuance]**, Hirshleifer, Hou, Teoh, and Zhang (2004).

73. Change in net operating assets (noa_gr1a) **[Investment]**, Hirshleifer, Hou, Teoh, and Zhang (2004).

74. Operating accruals (oaccruals_at) **[Accruals]**, Sloan (1996).

75. Percent operating accruals (oaccruals_ni) **[Accruals]**, Hafzalla, Lundholm, and Winkle (2011).

76. Operating cash flow to assets (ocf_at) **[Profitability]**, Bouchaud, Krüger, Landier, and Thesmar (2019).

77. Change in operating cash flow to assets (ocf_at_chg1) **[Profit Growth]**, Bouchaud, Krüger, Landier, and Thesmar (2019).

78. Operating cash flow to market (ocf_me) **[Value]**, Bouchaud, Krüger, Landier, and Thesmar (2019).

79. Operating cash flow to assets (ocf_at) **[Profitability]**, Bouchaud, Krüger, Landier, and Thesmar (2019).

80. Operating profits-to-lagged book assets (op_atl1) **[Quality]**, Ball, Gerakos, Linnainmaa, and Nikolaev (2016).

81. Operating profits to book equity (ope_be) **[Profitability]**, Fama and French (2015).

82. Operating profits to lagged book equity (ope_bel1) **[Profitability]**, Fama and French (2015).

83. Operating leverage (opex_at) **[Quality]**, Novy-Marx (2010).

84. Taxable income-to-book income (pi_nix) **[Seasonality]**, Lev and Nissim (2004).

85. Change PPE and Inventory (ppeinv_gr1a) **[Investment]**, Lyandres, Sun, and Zhang (2008).

86. Price and share (prc) **[Size]**, Miller and Scholes (1982).

87. Current price to high price over last year (prc_highprc_252d) **[Momentum]**, George and Hwang (2004).

88. Quality minus Junk: Composite (qmj) **[Quality]**, Asness, Frazzini, and Pedersen (2019).

89. Quality minus Junk: Growth (qmj_growth) **[Quality]**, Asness, Frazzini, and Pedersen (2019).

90. Quality minus Junk: Profitability (qmj_prof) **[Quality]**, Asness, Frazzini, and Pedersen (2019).

91. Quality minus Junk: Safety (qmj_safety) **[Quality]**, Asness, Frazzini, and Pedersen (2019).

92. Short-term reversal (ret_1_0) **[Size]**, Jegadeesh (1990).

93. Price momentum $t-12$ to $t-1$ (ret_12_1) **[Momentum]**, Jegadeesh and Titman (1993).

94. Price momentum $t-12$ to $t-7$ (ret_12_7) **[Profit Growth]**, Novy-Marx (2012).

95. Price momentum $t-3$ to $t-1$ (ret_3_1) **[Momentum]**, Jegadeesh and Titman (1993).

96. Price momentum $t-6$ to $t-1$ (ret_6_1) **[Momentum]**, Jegadeesh and Titman (1993).

97. Long-term reversal (ret_60_12) **[Investment]**, De Bondt and Thaler (1985).

98. Price momentum $t-9$ to $t-1$ (ret_9_1) **[Momentum]**, Jegadeesh and Titman (1993).

99. Maximum daily return (rmax1_21d) **[Low Risk]**, Bali, Cakici, and Whitelaw (2011).

100. Highest 5 days of return (rmax5_21d) **[Low Risk]**, Bali, Brown, and Tang (2017).

101. Highest 5 days of return scaled by volatility (rmax5_rvol_21d) **[Skewness]**, Asness, Frazzini, Gormsen, and Pedersen (2020).

102. Total skewness (rskew_21d) **[Skewness]**, Bali, Engle, and Murray (2016).

103. Return volatility (rvol_21d) **[Low Risk]**, Ang, Hodrick, Xing, and Zhang (2006).

104. Asset turnover (sale_bev) **[Quality]**, Soliman (2008).

105. Sale growth (1 year) (sale_gr1) **[Investment]**, Lakonishok, Shleifer, and Vishny (1994).

106. Sale growth (3 years) (sale_gr3) **[Investment]**, Lakonishok, Shleifer, and Vishny (1994).

107. Sale to market (sale_me) **[Value]**, William C. Barbee, Mukherji, and Raines (1996).

108. Sale growth (1 quarter) (saleq_gr3) **[Investment]**, Lakonishok, Shleifer, and Vishny (1994).

109. Year 1-lagged return, annual (seas_1_1an) **[Profit Growth]**, Heston and Sadka (2008).

110. Year 1-lagged return, nonannual (seas_1_1na) **[Momentum]**, Heston and Sadka (2008).

111. Years 2-5 lagged returns, annual (seas_2_5an) **[Seasonality]**, Heston and Sadka (2008).

112. Years 2-5 lagged returns, nonannual (seas_2_5na) **[Investment]**, Heston and Sadka (2008).

113. Change in short-term investments (sti_gr1a) **[Seasonality]**, Heston and Sadka (2008).

114. Total accruals (taccruals_at) **[Accruals]**, Richardson, Sloan, Soliman, and İrem Tuna (2005).

115. Percent total accruals (taccruals_ni) **[Accruals]**, Hafzalla, Lundholm, and Winkle (2011).

116. Asset tangibility (tangibility) **[Low Leverage]**, Hahn and Lee (2009).

117. Tax expense surprise (tax_gr1a) **[Profit Growth]**, Thomas and Zhang (2002).

118. Share turnover (turnover_126d) **[Low Risk]**, Datar, Y. Naik, and Radcliffe (1998).

119. Coefficient of variation for share turnover (turnover_var_126d) **[Profitability]**, Chordia, Subrahmanyam, and Anshuman (2001).

120. Altman Z-score (z_score) **[Low Leverage]**, Dichev (1998).

121. Number of zero trades with turnover as tiebreaker (6 months) (zero_trades_126d) **[Low Risk]**, Liu (2006).

122. Number of zero trades with turnover as tiebreaker (1 month) (zero_trades_21d) **[Low Risk]**, Liu (2006).

123. Number of zero trades with turnover as tiebreaker (12 months) (zero_trades_252d) **[Low Risk]**, Liu (2006).

Bibliography

ABARBANELL, J. S., AND B. J. BUSHEE (1998): "Abnormal Returns to a Fundamental Analysis Strategy," *The Accounting Review*, 73(1), 19–45.

AÏT-SAHALIA, Y., AND J. JACOD (2014): *High Frequency Econometrics.* Princeton University Press.

AKEY, P., V. GRÉGOIRE, AND C. MARTINEAU (2022): "Price revelation from insider trading: Evidence from hacked earnings news," *Journal of Financial Economics*, 143(3), 1162–1184.

ALI, A., L.-S. HWANG, AND M. A. TROMBLEY (2003): "Arbitrage risk and the book-to-market anomaly," *Journal of Financial Economics*, 69(2), 355–373.

AMIHUD, Y. (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31–56.

ANDERSEN, T. G., AND T. BOLLERSLEV (1997): "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *The Journal of Finance*, 52(3), 975–1005.

ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2013): "Chapter 17 - Financial Risk Measurement for Financial Risk Management," vol. 2 of *Handbook of the Economics of Finance*, pp. 1127–1220. Elsevier.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96(453), 42–55.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

ANG, A. (2014): *Asset Management: A Systematic Approach to Factor Investing.* Oxford University Press.

ANG, A., R. J. HODRICK, Y. XING, AND X. ZHANG (2006): "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance*, 61(1), 259–299.

ASNESS, C., A. FRAZZINI, N. J. GORMSEN, AND L. H. PEDERSEN (2020): "Betting against correlation: Testing theories of the low-risk effect," *Journal of Financial Economics*, 135(3), 629–652.

ASNESS, C. S., A. FRAZZINI, AND L. H. PEDERSEN (2019): "Quality minus junk," *Review of Accounting Studies*, 24(1), 34–112.

BACK, K. (1992): "Insider Trading in Continuous Time," *The Review of Financial Studies*, 5(3), 387–409.

BACK, K. E. (2017): *Asset pricing and portfolio choice theory.* Oxford University Press.

BAKALLI, G., S. GUERRIER, AND O. SCAILLET (2021): "A penalized two-pass regression to predict stock returns with time-varying risk premia," Swiss Finance Institute Research Paper Series 21-09, Swiss Finance Institute.

BALAKRISHNAN, K., E. BARTOV, AND L. FAUREL (2010): "Post loss/profit announcement drift," *Journal of Accounting and Economics*, 50(1), 20–41.

BALI, T., A. GOYAL, D. HUANG, F. JIANG, AND Q. WEN (2021): "Different Strokes: Return Predictability Across Stocks and Bonds with Machine Learning and Big Data," Working paper.

BALI, T. G., S. J. BROWN, AND Y. TANG (2017): "Is economic uncertainty priced in the cross-section of stock returns?," *Journal of Financial Economics*, 126(3), 471–489.

BALI, T. G., N. CAKICI, AND R. F. WHITELAW (2011): "Maxing out: Stocks as lotteries and the cross-section of expected returns," *Journal of Financial Economics*, 99(2), 427–446.

BALI, T. G., R. F. ENGLE, AND S. MURRAY (2016): *Empirical Asset Pricing: The Cross Section of Stock Returns*. John Wiley & Sons, Hoboken, NJ.

BALL, R., J. GERAKOS, J. T. LINNAINMAA, AND V. NIKOLAEV (2016): "Accruals, cash flows, and operating profitability in the cross section of stock returns," *Journal of Financial Economics*, 121(1), 28–45.

BANZ, R. W. (1981): "The relationship between return and market value of common stocks," *Journal of Financial Economics*, 9(1), 3–18.

BARILLAS, F., AND J. SHANKEN (2018): "Comparing Asset Pricing Models," *The Journal of Finance*, 73(2), 715–754.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): "Realized kernels in practice: trades and quotes," *The Econometrics Journal*, 12(3), C1–C32.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (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(2), 253–280.

——— (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2(1), 1–37.

BARROSO, P., AND A. DETZEL (2021): "Do limits to arbitrage explain the benefits of volatility-managed portfolios?," *Journal of Financial Economics*, 140(3), 744–767.

BARROSO, P., AND P. SANTA-CLARA (2015): "Momentum has its moments," *Journal of Financial Economics*, 116(1), 111–120.

BARTH, M. E., J. A. ELLIOTT, AND M. W. FINN (1999): "Market Rewards Associated with Patterns of Increasing Earnings," *Journal of Accounting Research*, 37(2), 387–413.

BASAK, G. K., R. JAGANNATHAN, AND T. MA (2009): "Jackknife Estimator for Tracking Error Variance of Optimal Portfolios," *Management Science*, 55(6), 990–1002.

BASU, S. (1983): "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence," *Journal of Financial Economics*, 12(1), 129–156.

BELO, F., AND X. LIN (2012): "The Inventory Growth Spread," *The Review of Financial Studies*, 25(1), 278–313.

BERNILE, G., J. HU, AND Y. TANG (2016): "Can information be locked up? Informed trading ahead of macro-news announcements," *Journal of Financial Economics*, 121(3), 496–520.

BHANDARI, L. C. (1988): "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *The Journal of Finance*, 43(2), 507–528.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

BOLLERSLEV, T., J. LI, AND Z. LIAO (2021): "Fixed-k inference for volatility," *Quantitative Economics*, 12(4), 1053–1084.

BOLLERSLEV, T., AND H. ZHOU (2002): "Estimating stochastic volatility diffusion using conditional moments of integrated volatility," *Journal of Econometrics*, 109(1), 33–65.

BOLSTAD, W. M. (2009): *Understanding Computational Bayesian Statistics.* New Jersey: John Wiley & Sons, Inc.

BOUCHAUD, J.-P., P. KRÜGER, A. LANDIER, AND D. THESMAR (2019): "Sticky Expectations and the Profitability Anomaly," *The Journal of Finance*, 74(2), 639–674.

BOUDT, K., O. KLEEN, AND E. SJØRUP (2021): "Analyzing intraday financial data in R: The highfrequency package," Working paper.

BRADSHAW, M. T., S. A. RICHARDSON, AND R. G. SLOAN (2006): "The relation between corporate financing activities, analysts' forecasts and stock returns," *Journal of Accounting and Economics*, 42(1), 53–85, Conference Issue on Implications of Changing Financial Reporting Standards.

BRENNAN, M. J., T. CHORDIA, AND A. SUBRAHMANYAM (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49(3), 345–373.

CATONI, O. (2004): *Statistical Learning Theory and Stochastic Optimization.* Saint- Flour Summer School on Probability Theory 2001 (Jean Picard ed.), Lecture Notes in Mathematics. Springer.

——— (2007): *PAC-Bayesian supervised classification: the thermodynamics of statistical learning.* Institute of Mathematical Statistics Lecture Notes— Monograph Series, 56. Institute of Mathematical Statistics, Beachwood, OH.

CEDERBURG, S., M. S. O'DOHERTY, F. WANG, AND X. S. YAN (2020): "On the performance of volatility-managed portfolios," *Journal of Financial Economics*, 138(1), 95–117.

CHAKRABARTY, B., B. LI, V. NGUYEN, AND R. A. VAN NESS (2007): "Trade classification algorithms for electronic communications network trades," *Journal of Banking & Finance*, 31(12), 3806–3821.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51(5), 1281–1304.

CHEN, A. Y., AND T. ZIMMERMANN (2020): "Open Source Cross-Sectional Asset Pricing," Working paper, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3604626`.

CHEN, L., M. PELGER, AND J. ZHU (2019): "Deep Learning in Asset Pricing," Working paper.

CHEN, Y. (2022): "Sparse Structure of Stochastic Discount Factor in the Chinese Stock Market: A Bayesian Interpretable Machine-learning Approach," Working paper.

CHERNOV, M., A. RONALD GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116(1), 225–257, Frontiers of financial econometrics and financial engineering.

CHIB, S., F. NARDARI, AND N. SHEPHARD (2002): "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108(2), 281–316.

CHINCO, A., A. D. CLARK-JOSEPH, AND M. YE (2019): "Sparse Signals in the Cross-Section of Returns," Manuscript, forthcoming for *Journal of Finance*.

CHIPMAN, H. A., E. I. GEORGE, AND R. E. MCCULLOCH (2010): "BART: Bayesian Additive Regression Trees," *The Annals of Applied Statistics*, 4(1), 266–298.

CHORDIA, T., A. SUBRAHMANYAM, AND V. ANSHUMAN (2001): "Trading activity and expected stock returns," *Journal of Financial Economics*, 59(1), 3–32.

CHRISTENSEN, K., U. HOUNYO, AND M. PODOLSKIJ (2018): "Is the diurnal pattern sufficient to explain intraday variation in volatility? A nonparametric assessment," *Journal of Econometrics*, 205(2), 336–362.

COLLIN-DUFRESNE, P., AND V. FOS (2015): "Do Prices Reveal the Presence of Informed Trading?," *The Journal of Finance*, 70(4), 1555–1582.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15(3), 373–394.

COOPER, M., H. GULEN, AND M. J. SCHILL (2008): "Asset Growth and the Cross-Section of Stock Returns," *The Journal of Finance*, 63(4), 1609–1651.

CORWIN, S. A., AND P. SCHULTZ (2012): "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices," *The Journal of Finance*, 67(2), 719–760.

DANIEL, K., AND T. J. MOSKOWITZ (2016): "Momentum crashes," *Journal of Financial Economics*, 122(2), 221–247.

DANIEL, K., AND S. TITMAN (2006): "Market Reactions to Tangible and Intangible Information," *The Journal of Finance*, 61(4), 1605–1643.

DATAR, V. T., N. Y. NAIK, AND R. RADCLIFFE (1998): "Liquidity and stock returns: An alternative test," *Journal of Financial Markets*, 1(2), 203–219.

DE BONDT, W. F. M., AND R. THALER (1985): "Does the Stock Market Overreact?," *The Journal of Finance*, 40(3), 793–805.

DECHOW, P. M., R. G. SLOAN, AND M. T. SOLIMAN (2004): "Implied Equity Duration: A New Measure of Equity Risk," *Review of Accounting Studies*, 9, 197–228.

DEMIGUEL, V., A. MARTÍN, F. J. NOGALES, AND R. UPPAL (2020): "A Transaction-Cost Perspective on the Multitude of Firm Characteristics," *The Review of Financial Studies*, 33(5), 2180–2122.

DICHEV, I. D. (1998): "Is the Risk of Bankruptcy a Systematic Risk?," *The Journal of Finance*, 53(3), 1131–1147.

EISDORFER, A., AND E. U. MISIRLI (2020): "Distressed Stocks in Distressed Times," *Management Science*, 66(6), 2452–2473.

ELLIS, K., R. MICHAELY, AND M. O'HARA (2000): "The Accuracy of Trade Classification Rules: Evidence from Nasdaq.," *Journal of Financial & Quantitative Analysis*, 35(4), 529–551.

ENGLE, R. (2004): "Risk and Volatility: Econometric Models and Financial Practice," *The American Economic Review*, 94(3), 405–420.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

FAIRFIELD, P. M., J. S. WHISENANT, AND T. L. YOHN (2003): "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing," *The Accounting Review*, 78(1), 353–371.

FAMA, E. F., AND K. R. FRENCH (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47(2), 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— (1996): "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51(1), 55–84.

——— (2015): "A Five-factor Asset Pricing Model," *Journal of Financial Economics*, 116(1), 1 – 22.

FAMA, E. F., AND J. D. MACBETH (1973): "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

FAN, J., Y. LIAO, AND W. WANG (2016): "Projected Principal Component Analysis in Factor Models," *The Annals of Statistics*, 44(1), 219–254.

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): "The Economic Value of Volatility Timing," *The Journal of Finance*, 56(1), 329–352.

——— (2003): "The economic value of volatility timing using "realized" volatility," *Journal of Financial Economics*, 67(3), 473–509.

FOSTER, G., C. OLSEN, AND T. SHEVLIN (1984): "Earnings Releases, Anomalies, and the Behavior of Security Returns," *The Accounting Review*, 59(4), 574–603.

FOWLER, D. J., AND C. RORKE (1983): "Risk measurement when shares are subject to infrequent trading: Comment," *Journal of Financial Economics*, 12(2), 279–283.

FRANCIS, J., R. LAFOND, P. M. OLSSON, AND K. SCHIPPER (2004): "Costs of Equity and Earnings Attributes," *The Accounting Review*, 79(4), 967–1010.

FRANKEL, R., AND C. M. LEE (1998): "Accounting valuation, market expectation, and cross-sectional stock returns," *Journal of Accounting and Economics*, 25(3), 283–319.

FRAZZINI, A., AND L. H. PEDERSEN (2014): "Betting against beta," *Journal of Financial Economics*, 111(1), 1–25.

FREYBERGERK, J., A. NEUHIERL, AND M. WEBER (2019): "Dissecting Characteristics Nonparametrically," Working paper, forthcoming in *The Review of Financial Studies*.

GABAUER, D., R. GUPTA, H. A. MARFATIA, AND S. M. MILLER (2020): "Estimating U.S. Housing Price Network Connectedness: Evidence from Dynamic Elastic Net, Lasso, and Ridge Vector Autoregressive Models," Working Papers 202065, University of Pretoria, Department of Economics.

GATHERAL, J., T. JAISSON, AND M. ROSENBAUM (2018): "Volatility is rough," *Quantitative Finance*, 18(6), 933–949.

GELMAN, A., J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI, AND D. B. RUBIN (2013): *Bayesian data analysis*, Chapman & Hall/CRC Texts in Statistical Science Series. CRC, Boca Raton, Florida.

GEORGE, T. J., AND C.-Y. HWANG (2004): "The 52-Week High and Momentum Investing," *The Journal of Finance*, 59(5), 2145–2176.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2021): "Economic predictions with big data: The illusion of sparsity," ECB Working Paper 2542.

GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): "A Test Of the Efficiency of a Given Portfolio," *Econometrica (1986-1998)*, 57(5), 1121.

GILKS, W., S. RICHARDSON, AND D. SPIEGELHALTER (1995): *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis.

GREEN, J., J. R. M. HAND, AND X. F. ZHANG (2017): "The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns," *The Review of Financial Studies*, 30(12), 4389–4436.

GROSSMAN, S. J., AND J. E. STIGLITZ (1980): "On the Impossibility of Informationally Efficient Markets," *The American Economic Review*, 70(3), 393–408.

GU, S., B. KELLY, AND D. XIU (2020): "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33(5), 2223–2273.

HAFZALLA, N., R. LUNDHOLM, AND E. M. V. WINKLE (2011): "Percent Accruals," *The Accounting Review*, 86(1), 209–236.

HAHN, J., AND H. LEE (2009): "Financial Constraints, Debt Capacity, and the Cross-Section of Stock Returns," *The Journal of Finance*, 64(2), 891–921.

HAN, Y., A. HE, D. E. RAPACH, AND G. ZHOU (2019): "What Firm Characteristics Drive US Stock Returns," Working paper.

HANSEN, L. P., AND R. JAGANNANTHAN (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), 225–262.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *The Review of Economic Studies*, 61(2), 247–264.

HARVEY, C. R., AND Y. LIU (2014): "Evaluationg Trading Strategies," *Journal of Portfolio Management*, 40(5), 108–118.

——— (2015): "Backtesting," *Journal of Portfolio Management*, 42(1), 13–28.

HARVEY, C. R., Y. LIU, AND H. ZHU (2016): "...and the Cross-Section of Expected Returns," *The Review of Financial Studies*, 29(1), 5–68.

HARVEY, C. R., AND A. SIDDIQUE (2000): "Conditional Skewness in Asset Pricing Tests," *The Journal of Finance*, 55(3), 1263–1295.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2001): *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

HAUGEN, R. A., AND N. L. BAKER (1996): "Commonality in the determinants of expected stock returns," *Journal of Financial Economics*, 41(3), 401–439.

HESTON, S. L., AND R. SADKA (2008): "Seasonality in the cross-section of stock returns," *Journal of Financial Economics*, 87(2), 418–445.

HIRSHLEIFER, D., K. HOU, S. TEOH, AND Y. ZHANG (2004): "Do Investors Overvalue Firms with Bloated Balance Sheets," *Journal of Accounting and Economics*, 38, 297–331.

HOLDEN, C., AND S. JACOBSEN (2014): "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions," *The Journal of Finance*, 69(4), 1747–1785.

HOU, K., C. XUE, AND L. ZHANG (2015): "Digesting Anomalies: An Investment Approach," *The Review of Financial Studies*, 28(3), 650–705.

——— (2018): "Replicating Anomalies," *The Review of Financial Studies*, 33(5), 2019–2133.

HUANG, D., J. LI, AND L. WANG (2021): "Are disagreements agreeable? Evidence from information aggregation," *Journal of Financial Economics*, 141(1), 83–101.

JACOD, J., J. LI, AND Z. LIAO (2020): "Volatility Coupling," Working paper, forthcoming in *Annals of Statistics*.

JACOD, J., AND P. PROTTER (2012): *Discretization of Process.* Springer.

JEGADEESH, N. (1990): "Evidence of Predictable Behavior of Security Returns," *The Journal of Finance*, 45(3), 881–898.

JEGADEESH, N., AND S. TITMAN (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance*, 48(1), 65–91.

JENSEN, T. I., B. KELLY, AND L. H. PEDERSEN (2022): "Is There a Replication Crisis in Finance," Working paper, conditionally accepted in *The Journal of Finance*.

JIANG, G., C. M. LEE, AND Y. ZHANG (2005): "Information Uncertainty and Expected Returns," *Review of Accounting Studies*, 10, 185–221.

JOBSON, J. D., AND B. M. KORKIE (1981): "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *The Journal of Finance*, 36(4), 889–908.

KACPERCZYK, M., AND E. S. PAGNOTTA (2019): "Chasing Private Information," *The Review of Financial Studies*, 32(12), 4997–5047.

KADAN, O., AND A. MANELA (2021): "Liquidity and the Strategic Value of Information," Working paper.

KELLY, B. T., T. J. MOSKOWITZ, AND S. PRUITT (2021): "Understanding momentum and reversal," *Journal of Financial Economics*, 140(3), 726–743.

KELLY, B. T., S. PRUITT, AND Y. SU (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134(3), 501–524.

KIM, H. H., AND N. R. SWANSON (2014): "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence," *Journal of Econometrics*, 178, 352–367, Recent Advances in Time Series Econometrics.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), 361–393.

KOZAK, S. (2020): "Kernel Trick for the Cross Section," Working paper.

KOZAK, S., S. NAGEL, AND S. SANTOSH (2018): "Interpreting Factor Models," *The Journal of Finance*, 73(3), 1183–1223.

——— (2020): "Shrinking the Cross Section," *Journal of Financial Economics*, 135(2), 271–292.

KYLE, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), 1315–1335.

LAKONISHOK, J., A. SHLEIFER, AND R. W. VISHNY (1994): "Contrarian Investment, Extrapolation and Risk," *The Journal of Finance*, 49(5).

LEDOIT, O., AND M. WOLF (2004a): "Honey, I Shrunk the Sample Covariance Matrix," *The Journal of Portfolio Management*, 30(4), 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88(2), 365–411.

LEDOIT, O., AND M. WOLF (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15(5), 850–859.

LEE, C. M., AND M. J. READY (1991): "Inferring Trade Direction from Intraday Data.," *Journal of Finance*, 46(2), 733–746.

LEE, P. M. (2012): *Bayesian Statistics: An Introduction, 4th edition.* Wiley.

LETTU, M., AND M. PELGER (2020a): "Estimating Latent Asset-Pricing Factors," forthcoming in *The Journal of Finance*.

——— (2020b): "Factors that Fit the Time-Series and Cross-Section of Stock Returns," *Review of Financial Studies*, 33(5), 2274–2325.

LEUNG, P.-L., AND W.-K. WONG (2008): "On testing the equality of multiple Sharpe ratios, with application on the evaluation of iShares," *The Journal of Risk*, 10(3), 15–30.

LEV, B., AND D. NISSIM (2004): "Taxable Income, Future Earnings, and Equity Values.," *Accounting Review*, 79(4), 1039–1074.

LI, Y., J. YU, AND T. ZENG (2021): "Deviance Information Criterion for Model Selection: A Theoretical Justification," Working paper, Remin University of China , Singapore Management University and Zhe Jiang University.

LINERO, A. R. (2018): "Bayesian Regression Trees for High-Dimensional Prediction and Variable Selection," *Journal of the American Statistical Association*, 113(522), 626–636.

LITZENBERGER, R. H., AND K. RAMASWAMY (1979): "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence," *Journal of Financial Economics*, 7(2), 163–195.

LIU, F., X. TANG, AND G. ZHOU (2019): "Volatility-Managed Portfolio: Does It Really Work?," *The Journal of Portfolio Management*, 46(1), 38–51.

LIU, W. (2006): "A liquidity-augmented capital asset pricing model," *Journal of Financial Economics*, 82(3), 631–671.

LO, A. W. (2002): "The Statistics of Sharpe Ratios," *Financial Analysts Journal*, 58(4), 36–52.

LOUGHRAN, T., AND J. W. WELLMAN (2011): "New Evidence on the Relation between the Enterprise Multiple and Average Stock Returns," *Journal of Financial and Quantitative Analysis*, 46(6), 1629–1650.

LUCCA, D. O., AND E. MOENCH (2015): "The Pre-FOMC Announcement Drift," *The Journal of Finance*, 70(1), 329–371.

LYANDRES, E., L. SUN, AND L. ZHANG (2008): "The New Issues Puzzle: Testing the Investment-Based Explanation," *Review of Financial Studies*, 21(6), 2825–2855.

MANCINI, C. (2001): "Disentangling the jumps of the diffusion in a geometric jumping Brownian motion," .

MARKOWITZ, H. (1952): "Portfolio Selection," *The Journal of Finance*, 7(1), 77–91.

MCLEAN, R. D., AND J. PONTIFF (2016): "Does Academic Research Destroy Stock Return Predictability," *Journal of Finance*, 71(1).

MESSMER, M., AND F. AUDRINO (2017): "The (adaptive) Lasso in the Zoo-Firm Characteristic Selection in the Cross-Section of Expected Returns," .

MILLER, M. H., AND M. S. SCHOLES (1982): "Dividends and Taxes: Some Empirical Evidence," *Journal of Political Economy*, 90(6), 1118–1141.

MOREIRA, A., AND T. MUIR (2017): "Volatility Managed Portfolios," *Journal of Finance*, 72(4), 1611–1644.

——— (2019): "Should Long-Term Investors Time Volatility?," *Journal of Financial Economics*, 131(3), 507–527.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NOVY-MARX, R. (2010): "Operating Leverage," *Review of Finance*, 15(1), 103–134.

——— (2012): "Is momentum really momentum?," *Journal of Financial Economics*, 103(3), 429–453.

——— (2013): "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108(1), 1–28.

ORTIZ-MOLINA, H., AND G. M. PHILLIPS (2014): "Real Asset Illiquidity and the Cost of Capital," *Journal of Financial and Quantitative Analysis*, 49(1), 1–32.

PALAZZO, B. (2012): "Cash holdings, risk, and expected returns," *Journal of Financial Economics*, 104(1), 162–185.

PAV, S. E. (2021): *The Sharpe Ratio: Statistics and Applications.* New York: CRC Press, Taylor & Francis Group.

——— (2022): "Using the SharpeR Package," Technical report, Bank of America.

PENMAN, S., S. A. RICHARDESON, AND I. TUNA (2007): "The Book-to-Price Effect in Stock Returns: Accounting for Leverage," *Journal of Accounting Research*, 45(2), 427–467.

PIOTROSKI, J. D. (2000): "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers," *Journal of Accounting Research*, 38, 1–41.

PONTIFF, J., AND A. WOODGATE (2008): "Share Issuance and Cross-sectional Returns," *The Journal of Finance*, 63(2), 921–945.

RAPACH, D. E., J. K. STRAUSS, AND G. ZHOU (2010): "Out-of-sample Equity Premium Prediction: Combination Forecasts and links to the Real Economy," *The Review of Financial Studies*, 23(2), 821–862.

REVUZ, D., AND M. YOR (2004): *Continuous Martingales and Brownian Motion, 3rd edition.* Springer.

RICHARDSON, S. A., R. G. SLOAN, M. T. SOLIMAN, AND İREM TUNA (2005): "Accrual reliability, earnings persistence and stock prices," *Journal of Accounting and Economics*, 39(3), 437–485.

ROSENBERG, B., K. REID, AND R. LANSTEIN (1985): "Persuasive evidence of market inefficiency," *The Journal of Portfolio Management*, 11(3), 9–16.

ROČKOVÁ, V. (2019): "On Semi-parametric Bernstein-von Mises Theorems for BART," Working paper.

ROČKOVÁ, V., AND E. SAHA (2019): "On Theory for BART," Working paper, 22^{nd} International Conference on Artificial Intelligence and Statistics.

ROČKOVÁ, V., AND S. VAN DER PAS (2019): "Posterior Concentration for Bayesian Regression Trees and Forests," Manuscript, just accepted for *Annals of Statistics*.

SLOAN, R. G. (1996): "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *The Accounting Review*, 71(3), 289–315.

SOLIMAN, M. T. (2008): "The Use of DuPont Analysis by Market Participants," *The Accounting Review*, 83(3), 823–853.

SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN, AND A. VAN DER LINDE (2002): "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4), 583–639.

——— (2014): "The deviance information criterion: 12 years on," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3), 485–493.

STAMBAUGH, R. F., AND Y. YUAN (2016): "Mispricing Factors," *The Review of Financial Studies*, 30(4), 1270–1315.

STROUD, J. R., AND M. S. JOHANNES (2014): "Bayesian Modeling and Forecasting of 24-Hour High-Frequency Volatility," *Journal of the American Statistical Association*, 109(508), 1368–1384.

TANNER, M. A., AND W. H. WONG (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

TAYLOR, S. J. (1982): "Financial returns modelled by the product of two stochastic processes : a study of daily sugar prices, 1961-79," *Time series analysis : Theory and Practice*, 1, 203–226.

THOMAS, J. K., AND H. ZHANG (2002): "Inventory Changes and Future Returns," *Review of Accounting Studies*, 7, 163–187.

VAPNIK, V. (1998): *Statistical Learning Theory*. Wiley, New York.

WANG, X., W. XIAO, AND J. YU (2022): "Modeling and forecasting realized volatility with the fractional Orstein-Uhlenbeck process," Working paper, forthcoming in *Journal of Econometrics*.

WELCH, I. (2019): "Reproducing, Extending, Updating, Replicationg, Reexamining, and Reconciling," *Critical Finance Review*, 8(1-2), 301–304.

WILLIAM C. BARBEE, J., S. MUKHERJI, AND G. A. RAINES (1996): "Do Sales-Price and Debt-Equity Explain Stock Returns Better than Book-Market and Firm Size?," *Financial Analysts Journal*, 52(2), 56–60.

WRIGHT, J. A., S. C. P. YAM, AND S. P. YUNG (2014): "A test for the equality of multiple Sharpe ratios," *The Journal of Risk*, 16(4), 3–21.

XIU, D. (2010): "Quasi-maximum likelihood estimation of volatility with high frequency data," *Journal of Econometrics*, 159(1), 235–250.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales," *Journal of the American Statistical Association*, 100(472), 1394–1411.

ZOU, H., AND T. HASTIE (2005): "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.