Dependence in financial and high-dimensional time series

Li, H.

Publication date
2021
Document Version
Final published version

This thesis deals with the complex subtle and unexplored dependence structure in financial time series and high-dimensional macroeconomic time series. This thesis provides more flexible model specifications and simpler ways to account for autocorrelations in time series. First, this thesis proposes an autocorrelation-based factor model to estimate and forecast the yield curve, or the term structure of interest rates. This new model offers favorable in-sample fit and out-of-sample forecasting results. Second, it integrates more flexible distributions into the classic capital asset pricing model to better feature financial return characteristics. It develops the procedures for estimation and shows the economic value of this novel integration to institutional investors. Third, it sheds light on the dependence between the overnight return and the subsequent intraday return and examines trading opportunities and holding periods for day traders.

Hao Li graduated with B.A. in Finance (with honor in recognition of the thesis) from Nankai University, China. She obtained M.Sc. in Finance from the University of Groningen (Rijksuniversiteit Groningen), the Netherlands. While she was completing her M.Sc. in Economics at the University of Groningen, she started her Ph.D. in Financial Econometrics in the Department of Quantitative Economics at the University of Amsterdam, the Netherlands.

Dependence in Financial and High-dimensional Time Series

Hao Li

Hao Li

Dependence in Financial and High-dimensional Time Series

University of Amsterdam

DEPENDENCE IN FINANCIAL AND HIGH-DIMENSIONAL TIME SERIES

Hao Li

Typeset using LaTeX, figures generated using R and Matlab

Layout: Hao Li
Cover design: Hao Li, ProefschriftMaken, and infograph.venngage.com
Printed by: ProefschriftMaken, www.proefschriftmaken.nl
ISBN: 978–94–6423–536–4

Dependence in Financial and High-dimensional Time Series

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 11 november 2021, te 16.00 uur

door Hao Li
geboren te Xinjiang

Promotiecommissie

Promotor:	prof. dr. C.G.H. Diks	Universiteit van Amsterdam
Copromotor:	prof. dr. M. van der Wel	Erasmus Universiteit Rotterdam
Overige leden:	prof. dr. J.C. Wu	University of Notre Dame
	prof. dr. Q. Yao	London School of Economics
	prof. dr. H.P. Boswijk	Universiteit van Amsterdam
	prof. dr. C.H. Hommes	Universiteit van Amsterdam
	prof. dr. F.R. Kleibergen	Universiteit van Amsterdam

Faculteit Economie en Bedrijfskunde

Acknowledgements

It has been a long and intellectual journey to complete this thesis. Along this journey, I received tons of help and encouragements from many friendly, kind, and great people. Without their support, what I have achieved today would not be possible to make. I am so glad that I have this chance to thank them here.

First and foremost, I would like to thank my parents, who always support me financially and mentally. Without the encouragement, respect, and love that they have given me, I might not be the person I am today. They are the best parents around the world and the most beautiful and lovely people in my mind.

I am also very grateful to my supervisor Cees for his tremendous patience in answering all my questions and enormous trust that I am capable of being an independent researcher. I also benefit massively from his strong support, for examples, providing examples of cover letters for journal article submission, correcting typos in my writings, checking summary in Dutch in this thesis, and recommending me to my co-supervisor Michel. I feel extremely fortunate that I could have such a great supervisor, who is incredibly important to my personal growth.

My sincere gratitude also goes to my co-supervisor Michel, who always gave me strong confidence, many valuable comments and helpful suggestions. I deeply admire Michel for his generous, kind, open-minded, and chivalrous offer and highly appreciate him for his strong and helpful support and guidance in my life and in doing research. Michel is my brightest star in the night sky who guides me to go ahead in my growth journey.

Besides my supervisor and co-supervisor, I also wish to thank Te, who supervised my MSc thesis and recommended me to Cees to start my PhD studies. Te, like my most reliable family member, is always there whenever I need help or support and does his all best every time. I am deeply indebted to him for all his support and help without even the slightest reservation.

I am extremely grateful to Howell, and my PhD committee members Qiwei, Peter, Frank, Cynthia, and Cars for their kindness, patience, and generosity during my very difficult times or when I had questions to consult. They all mean a lot to me. Without their crucial support, I could have not completed this intellectual journey. Honestly, I have restarted to love and enjoy

doing research since I overcame these difficulties. I will forever remember and highly appreciate them for their kind and generous offer.

My many thanks also go out to our amazing and lovely econometricians and CeNDEFfers, Anita, Arturas, Bart, David, Fabian, Florian, Hans, Isabelle, Jan (econometrician), Jan (CeNDEFfer), Jonas, Kees Jan, Lina, Mario, Maurice, Noud, Roald, Sander, Steffi, and Yi, and brilliant and sweet PhD students Aad, Alessandro, Alex, Alexandre, Andrea, Daria, Ekaterina, Enrico, Eva (econometrician), Eva (CeNDEFfer), Evgenii, Frieder, Gabriela, Gavin, Gregor, Guanyu, Hao (male), Joep, Johan, Lingwei, Martina, Merrick, Moutaz, Myrna, Quinten, Ramon, Raviar, Robert, Sanna, Sebastian, Simmon, Timo, Tolga and Tommaso, who are pleasant to work with. I still remember the inspiring and delightful discussions with Hao, Lingwei, Merrick, and Yi, Kees Jan's share of conference opportunities, Jan's generous offer of giving comments and feedback, Bart's helpful comments on my presentations, Tolga's unselfish experience sharing helping me to obtain 4.5/5 from Time Series Analysis course evaluation, group's cycling on Texel and group's movie night at Johan's home, and delicious dessert made by Anita. Because of you all, I feel my PhD life is interesting and wonderful.

I also highly appreciate Xiao, Yi, Lingwei, Bo, Shuo, Kai, Yicong, Jiahua, and Yajing who helped and/or encouraged me when I was on the job market, which perhaps is the most competitive one till now. With your concern, understanding, willing to help, and encouragements, I never felt that I was lonely or helpless during that time. I also wish to thank Eduardo Horta and Sven Otto for sharing me code. Thanks to their generous sharing, I made progress faster and saved a lot of time. Lastly, I sincerely thank Sweder for his constructive feedback on my job market paper.

If I have not explicitly mentioned all the dearest people who have accompanied me during my PhD over the last years, it does not mean that I have forgotten them. I feel very happy and lucky that they were/are around me.

My intellectual journey has not ended. Equipped with the knowledge I have obtained during my PhD life, I have a strong motivation to explore and find answers to unresolved questions in term structure modeling, the marco-finance interaction, and high frequency financial time series over day and night. I am ready to start and excited about the following new journeys.

Contents

Chapter 1

Introduction

Dependence structure in time series is frequently investigated by financial economists and econometricians. In many studies, such analysis is carried out by parametric and nonparametric methods, such as maximum likelihood methods, kernel smoothing methods, and functional principal component analysis. This thesis investigates the complex subtle and novel dependence structure in financial time series and macroeconomic high-dimensional time series with more flexible model specifications and much simpler ways to account for serial dependence, compared with the methods documented in the literature.

Chapter 2 is on modeling and forecasting the yield curve.[1] Strong serial dependence typically arises, when we track the yield curve (or the term structure of interest rates) over time. To benefit from this feature, this chapter proposes a new factor model to estimate and forecast yield curves based on factors driving this serial dependence. In my semiparametric approach, factor loadings are related to the autocovariance functions of the continuous and smooth function of the yield curve subject to unobservable errors.[2] The dynamic evolution is driven by a vector autoregression for a small set of factors. The number of factors and aggregation of information over different lags are determined by the yield data. Applying this method to monthly US government bond yields from January 1985 through December 2020, I find that the dynamic structure of yield curves reduces to a vector process lying in a 3-dimensional space, with 1-month lag information. Yield curve residuals from this new model exhibit less autocorrelation than those obtained from the alternative three-factor models. Moreover, this new model provides favorable forecasting results.

[1]This chapter benefits from Michel van der Wel's valuable comments and suggestions, and discussions with the seminar participants at the UvA Econometrics lunch seminar, the Econometric Society Winter Meeting 2020, the 2020 European Conferences of the Econom[etr]ics Community, the ASSA 2021 meeting AEA poster session, the 2021 Asian Meeting of the Econometric Society, and the 2021 China Meeting of the Econometric Society on the earlier versions of my job market paper.

[2]Each yield curve observed at any given time is regarded as a function.

Chapter 3 extends the classical CAPM in the sense that the error term follows an asymmetric power distribution.[3] Financial return series at higher frequency exhibits asymmetry and fat tails and thus suggests a more flexible distributional assumption to account for these properties. This chapter offers procedures for parameter estimation by the maximum likelihood method when disturbances in the models are asymmetric power distributed, and develops portfolio strategy to apply the estimation results. Empirically, information criteria for model selection, density plots estimated by kernel density estimation, and distribution tests indicate that the asymmetric power distributional assumption is preferred for financial return series at higher frequency, compared with the commonly-used normal distribution. Moreover, this chapter shows the economic value of this more flexible model specification to long-run institutional investors.

Chapter 4 is on intraday return patterns of stock market.[4] This chapter investigates patterns of dependence between the overnight return and returns just before and after the overnight periods by examining trade and quote data at the transactional level. Except for the linear dependence, more complex nonlinear dependence is revealed by the nonparametric kernel regression technique. Accordingly, trading strategies based on these patterns of dependence are developed for day traders.

[3]This chapter is co-authored with Te Bao from the Nanyang Technological University and Cees Diks from the University of Amsterdam. I wish to thank the China Scholarship Council (CSC) for the financial support.

[4]This chapter is co-authored with Cees Diks from the University of Amsterdam and Valentyn Panchenko from the University of New South Wales. I wish to thank the China Scholarship Council (CSC) for the financial support.

Chapter 2

Modeling and forecasting serially dependent yield curves

2.1 Introduction

To model the yield curve (which is the collection of yields for different maturities), factor models are commonly used in the literature. Such factor models generally combine a model for the dynamics of such factors with a specification for how the factors relate to (or, span) the full cross-section of yields. In arbitrage-free models, such as dating back to Vasicek (1977) and Cox *et al.* (1985), the model starts from the factor dynamics while no-arbitrage restrictions determine the relation of the yields to these factors. In structural models, such as the dynamic Nelson and Siegel (1987) model as introduced by Diebold and Li (2006), first factors are extracted from the yield curve for which then factor dynamics are assumed. However, when the yield curve is spanned by a small set of common factors as in these approaches, the yield curve residuals are highly serially correlated (see, e.g., Duffee, 2011; Hamilton and Wu, 2014) and have non-zero mean (see the in-sample fit results in e.g., Diebold and Li, 2006; Christensen *et al.*, 2009; Van Dijk *et al.*, 2014; Raviv, 2015).

This chapter introduces an autocorrelation-based factor model. Besides the common factors that appear largely settled in the term structure literature, this new factor model is able to absorb the autocorrelated idiosyncratic components into factors. This is because all the dynamic elements of the observed yield curve (cross-sectionally continuous and smooth) subject to unobservable errors, including the dynamics only affect over some period of time, may be included into the factors in the model.[1] Specifically, the factors and factor loadings are directly related to the

[1] In my approach each yield curve is a continuous and smooth function across maturities subject to e.g. measurement errors.

autocovariance functions of the yield curve, by means of implementing an eigenanalysis on a non-negative operator aggregating the information over different lags.

My method also sheds light on the recent debate on the number of factors necessary for yield curve modeling. Except for many studies using three yield-curve factors either yields or unobserved latent variables (see e.g., Litterman and Scheinkman, 1991; Christensen *et al.*, 2011; Gürkaynak and Wright, 2012, for a survey), Duffee (2011) finds evidence of one hidden factor not explained by the common factors. Christensen and van der Wel (2019) find that at least three or four factors are necessary to explain the term structure movements, while Joslin *et al.* (2014) find evidence of five factors including two of them from macroeconomic data. Also, Crump and Gospodinov (2021) demonstrate the economic costs of using the first three common factors when cross-sectional dependence is exclusively addressed. Hence, it is interesting to investigate the number of factors when the serial dependence of the yield curve is also addressed in the autocorrelation-based factors for yield curve modeling.

Modeling and forecasting the yield curve have been investigated comprehensively in the bulk of literature using non/semiparametric approach.[2] Litterman and Scheinkman (1991) employ principal component analysis (PCA) to explain the term structure movements of bond yields by three factors of level, steepness, and curvature. The nonparametric kernel smoothing methods are served for yield curve modeling and forecasting as well, see e.g. Linton *et al.* (2001), Caldeira and Torrent (2017), Koo *et al.* (2021), and Liu and Wu (2021). Also, there are many studies viewing observed yield curves as realizations of function-valued than vector-valued random variables, see e.g. Bowsher and Meeks (2008), Hays *et al.* (2012), Härdle and Majer (2016), Bardsley *et al.* (2017), Almeida *et al.* (2018), Otto and Salish (2019) and Sen and Klüppelberg (2019), connecting functional data with functional factor loading curves. In their papers the yields observed at different maturities at a given time are a discrete sampling from a true underlying yield curve. Among them, natural cubic splines, tensor B-splines, and cubic B-splines functions are used to interpolate yields within and out of sample, yielding the true yield curves. By doing so, factor model's in-sample fit and out-of-sample forecasting performance can be improved, see e.g. Bowsher and Meeks (2008) and Almeida *et al.* (2018). In my semiparametric approach, I use cubic B-splines to transform yields at different maturities into the underlying yield curve functions, and then implement functional principal component analysis (FPCA) to find factor loadings and factors. After that, the dynamic evolution is modeled by a vector autoregressive (VAR) model.

This autocorrelation-based factor model, with a stable choice of three factors and one lag information in yields, is seen to improve the in-sample fit, using monthly constant-maturity zero-coupon

[2]Naturally, this is on top of an extensive literature on parametric approaches, which also includes e.g., Chambers *et al.* (1984), Nelson and Siegel (1987), Svensson (1994), Christensen *et al.* (2011) and Van Dijk *et al.* (2014).

Treasury yield curve data of Liu and Wu (2021) over the period January 1985 up to December 2020. More specifically, the yield curve residuals exhibit less autocorrelation and have zero mean. This finding is robust with respect to the choice of dimension and lag order of information over time. Compared with the closely-related dynamic functional factor model of Otto and Salish (2019), looking for both common factors and the factors absorbed from the serially correlated idiosyncratic components improves the model's in-sample fit. Root mean squared error (RMSE) and Mean absolute error (MAE) for model residuals are, on average, reduced by 14.8% and 12.2%, respectively. Compared with the dynamic Nelson–Siegel (DNS) model, without specifying functional form of factor loadings provides more flexibility; the average reductions in RMSE and in MAE reach 21.3% and 26.6%, respectively.

Moreover, using serial dependence information in the yield curve improves the out-of-sample forecasting performance. Relative to the forecasts of five three-factor benchmark models, DNS model, a functional factor model based on FPCA on the conventional covariance operator, a vector-valued PCA model, and two extensions of the DNS model with shifting endpoints, the forecasts of this new autocorrelation-based factor model have superiority including less non-zero mean, less autocorrelation in prediction errors at different maturities, and smaller root mean squared prediction error (RMSPE) across the term structure dimension and over time at the 1-month-ahead forecast horizon.

The remainder of this chapter is organized as follows. Section 2.2 introduces this new autocorrelation-based factor model and the procedures for estimating and forecasting the serially dependent yield curves. Section 2.3 presents the application of this method to the US monthly Treasury yield data. Section 2.4 concludes.

2.2 Methodology

In this section I formalize my approach. In 2.2.1 I discuss the new model specification. Next, in 2.2.2 and 2.2.3 I turn to estimation and forecasting, respectively.

2.2.1 The autocorrelation-based factor model

I denote an observed yield curve with $Y_t(u)$, for time to maturity $u \in [a, b]$ at time $t = 1, \ldots, T$, where a denotes the short end of the yield curve and b denotes the long end of the yield curve. The sequence of yield curves can be expressed as

$$Y_t(u) = X_t(u) + \varepsilon_t(u), \qquad t = 1, \ldots, T, \qquad u \in [a, b], \tag{2.1}$$

where $\varepsilon_t(\cdot)$ is a noise term, in the sense that we assume (i) $\mathbb{E}\{\varepsilon_t(u)\} = 0$ for all t and all $u \in [a, b]$; (ii) $\mathrm{Cov}\{\varepsilon_t(u), \varepsilon_{t+k}(v)\} = 0$ for all $u, v \in [a, b]$ provided that $k \neq 0$, (iii) $\mathrm{Cov}\{X_t(u), \varepsilon_{t+k}(v)\} = 0$ for all $u, v \in [a, b]$ and $k \neq 0$. Note that $Y_t(u)$ should be square-integrable on the domain $[a, b]$ in a Hilbert space $\mathcal{L}_2$ and that the curve process of particular interest $X_t(u)$ and the errors $\varepsilon_t(u)$ are not separately observable; only their sum is observed.

Note that Y_t is usually recorded on discrete grids.[3] Like other factor models viewing the yield curve as a functional time series, I convert these cross-sectional yields at different maturities to a time series observation of a continuous and smooth function. I use cubic B-splines to approximate the underlying functions, instead of employing natural cubic splines. Models using natural cubic splines rely on specifying knots, that is yields at the knots of the natural cubic spline are considered as the factors in these models (e.g., Bowsher and Meeks, 2008; Jungbacker *et al.*, 2014; Almeida *et al.*, 2018; Feng and Qian, 2018). However, since any small deviation in any segment in-between knots may change the yield curve globally, the yields not at the knots are difficult to interpolate accurately in spite of accommodating smoothness. In contrast, all cross-sectional yields at time t can be used as knots of a cubic B-spline expansion, and each segment is a weighted sum of basis functions only. Consequently, a small deviation, e.g. a small hump in the short end of the yield curve influenced by changes in monetary policy instruments, does not lead to changes in the other segments of the yield curve during the interpolation. Hence, employing cubic B-splines to transform discrete yields data to $Y_t(u)$ strikes a balance between smoothness and accuracy (see e.g., De Boor, 1978; Ramsay and Silverman, 2002; Ramsay *et al.*, 2009).

The autocorrelation-based factor model is defined as

$$Y_t(u) = \mu(u) + \sum_{j=1}^{d} \eta_{tj}\psi_j(u) + \varepsilon_t(u), \qquad t = 1, \ldots, T, \qquad u \in [a, b], \tag{2.2}$$

where $\mu(u) \equiv \mathbb{E}\{Y_t(u)\}$, η_{tj} denotes the jth factor at time t and $\psi_j(u)$ denotes the jth loading for time to maturity u. Here, $\psi_1(u)$,..., $\psi_d(u)$ are orthonormal eigenfunctions of the non-negative operator

$$K(u, v) = \sum_{k=1}^{p} \int_a^b c_k(u, z)\, c_k(v, z)\, \mathrm{d}z, \qquad k = 1, \ldots, p, \tag{2.3}$$

where $c_k(u, v) \equiv \mathrm{Cov}\{X_t(u), X_{t+k}(v)\}$ is the kth lag autocovariance function of the true centered yield curve process. Thus, in this factor model, the factor loadings are directly related to the autocovariance functions (or the lagged covariance functions after the duality operation described

[3]For example, in the popular unsmoothed Fama-Bliss zero-coupon US treasuries yield data for yield curve modeling and forecasting, the maturities are 3, 6, 9, 12, 15, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

in Section 2.2.2) of the yield curve, and the factors determine to what extent these autocovariance functions drive the yield curve at a given time.

Using the autocovariance function rather than the covariance function of the yield curve process is theoretically appealing. Let $c_Y(u,v)$ be the covariance kernel of the yield curve $Y(\cdot)$, and thus $c_Y(u,v) = c_X(u,v) + c_\varepsilon(u,v)$ instead of $c_Y(u,v) = c_X(u,v)$ due to the existence of some errors $\varepsilon(\cdot)$ e.g. numerical rounding when the yield data are sequentially recorded. Since the yield curve process of particular interest $X(\cdot)$ and $\varepsilon(\cdot)$ are both unobservable, $\widehat{c}_Y(u,v)$ using the observed yield data is not a consistent estimator of $c_X(u,v)$ due to the presence of $\varepsilon(\cdot)$. In contrast, the autocovariance kernel $c_k(u,v) : k \in \mathbb{Z}$ of $X(\cdot)$ and of $Y(\cdot)$ are equal for $k \neq 0$, making the estimator $\widehat{c}_k(u,v)$ based on observed yield data a consistent estimator for the true yield curve process, without making the assumption that the yield curves are independent or that the variance of the errors $\varepsilon(\cdot)$ goes to zero as the sample size tends to infinity. By construction $\psi_1(u),\ldots,\psi_d(u)$ are orthonormal in $\mathcal{L}_2$, where d is a positive integer called the dimension of the yield curves, implying the number of factors in the model. Note that using $K(u,v)$ allows to gather information at different lags from $1,\ldots,p$ together. Moreover, due to random fluctuations in the sample, a single non-negative definite operator $\int_a^b c_k(u,z)\, c_k(v,z)\,\mathrm{d}z$ may lead to spurious choices of the dimension, while, according to Bathia *et al.* (2010), the summation of non-negative definite operators at different lags can overcome this.

The model can be compared to two existing approaches in the literature. First, in contrast to on the DNS model proposed by Diebold and Li (2006) and variants thereof, as proposed by Van Dijk *et al.* (2014), this method does not require imposing parametric assumptions on the shape of the loadings, i.e. the eigenfunctions $\psi_1(u),\ldots,\psi_d(u)$. Specifically, in the DNS approach the number of factors d is set to three and the three factor loading functions are defined to pick up the level, slope and curvature of the yield curve, and determined by one parameter. Second, different from the term structure models employing FPCA, such as in Sen and Klüppelberg (2019) and Otto and Salish (2019), the factor loadings are eigenfunctions of a non-negative operator aggregating the product of two kth lag autocovariance functions over p lags. The autocorrelated idiosyncratic components in the yield curve will be identified as factors in the model. In contrast, in extant term structure models employing (F)PCA, the factor loadings are the eigenfunctions of the conventional covariance operator, $c_Y(u,v) = \mathrm{Cov}\{Y_t(u), Y_t(v)\}$, requiring some extra assumptions on Y_t or ε_t to find the factors. Moreover, applying this to serially dependent yield data potentially cannot generate optimal dimension reduction (see e.g. Hörmann *et al.*, 2015). This means that the dynamic factor models based on the conventional covariance operator likely prefer more factors than my novel autocorrelation-based model to achieve similar explanatory power.

2.2.2 Estimation

Let $\widehat{\mu}(u) = \frac{1}{T}\sum_{t=1}^{T} Y_t(u)$ for $u \in [a, b]$ denote the empirical mean function of $Y_t(u)$ and using this to consider the demeaned curve $Y_t(u) - \widehat{\mu}(u)$. We minimize the objective function

$$\sum_{t=1}^{T} \left\| Y_t(u) - \widehat{\mu}(u) - \sum_{j=1}^{d} \eta_{tj}\psi_j(u) \right\|^2, \tag{2.4}$$

subject to the orthonormality conditions given by assumptions made earlier. Suppose that the loading functions $\psi_1(u),\ldots,\psi_d(u)$ are orthonormal, we can minimize Eq. (2.4) with respect to the factors η_{tj} for all $j = 1,\ldots,d$ and $t = 1,\ldots,T$.

The least squares objective function

$$\min_{\eta_{tj}} \sum_{t=1}^{T} \left\| Y_t(u) - \widehat{\mu}(u) - \sum_{j=1}^{d} \eta_{tj}\psi_j(u) \right\|^2, \tag{2.5}$$

is achieved when $\eta_{tj} = \langle Y_t(u) - \widehat{\mu}(u), \psi_j(u)\rangle$ for any $j = 1,\ldots,d$ and $t = 1,\ldots,T$. That is, the sum of squares in Eq. (2.5) is minimized when projecting the demeaned yield curve at time t onto the jth loading function to obtain η_{tj}. As a result, the least squares objective function is identical to

$$\min_{\psi_1(u),\ldots,\psi_d(u)} \sum_{t=1}^{T} \left\| Y_t(u) - \widehat{\mu}(u) - \sum_{j=1}^{d} \langle Y_t(u) - \widehat{\mu}(u), \psi_j(u)\rangle \psi_j(u) \right\|^2. \tag{2.6}$$

That is to say, looking for d-dimensional orthonormal eigenfunctions $\psi_1(u),\ldots,\psi_d(u)$ and the corresponding eigenvalues in the framework of functional data analysis.

Following Bathia *et al.* (2010) to find a d-dimensional orthonormal system in a square-integrable function space, let

$$\widehat{K}(u, v) = \sum_{k=1}^{p} \int_a^b \widehat{c}_k(u, z)\, \widehat{c}_k(v, z)\, \mathrm{d}z, \qquad k = 1,\ldots,p, \qquad u, v \in [a, b], \tag{2.7}$$

where

$$\widehat{c_k}(u, v) := \frac{1}{(T-p)} \sum_{t=1}^{T-p} (Y_t(u) - \widehat{\mu}(u))(Y_{t+k}(v) - \widehat{\mu}(v)),$$

be the k lags sample autocovariance operator. Then using the observed yields data, Eq. (2.7) can

be written as

$$\widehat{K}(u,v) = \frac{1}{(T-p)^2}\sum_{t,s=1}^{T-p}\sum_{k=1}^{p}(Y_t(u)-\widehat{\mu}(u))(Y_s(v)-\widehat{\mu}(v)) \\ \times \langle Y_{t+k}(u)-\widehat{\mu}(u), Y_{s+k}(v)-\widehat{\mu}(v)\rangle, \qquad k=1,\ldots,p, \qquad u,v\in[a,b]. \quad (2.8)$$

However, $\widehat{K}(u,v)$ may be an $\infty\times\infty$ matrix if the centered yield curve $Y_t(\cdot)-\widehat{\mu}(\cdot)$ is a functional curve evaluated at a fine grid such as an $\infty\times 1$ vector on the domain [a,b]. Following Bathia *et al.* (2010), the duality property is applied here to make the eigenanalysis tractable in a $\mathcal{L}_2$.[4] That is the $\infty\times\infty$ matrix of $\widehat{K}(u,v)$ has the same $\widehat{d}$ non-zero eigenvalues of the $(T-p)\times(T-p)$ matrix

$$\widehat{\mathbf{K}}(u,v) := \frac{1}{(T-p)^2}\sum_{k=1}^{p}\mathbf{Y}_k\mathbf{Y}_0, \quad (2.9)$$

where $\mathbf{Y}_k$ is the $(T-p)\times(T-p)$ matrix with the (t,s)th element $\langle Y_{t+k}-\widehat{\mu}, Y_{s+k}-\widehat{\mu}\rangle$ for $k=0,\ldots,p$. Furthermore, let $\widehat{\boldsymbol{\gamma}}_j=(\widehat{\gamma}_{1,j},\ldots,\widehat{\gamma}_{T-p,j})$, $j=1,\ldots,\widehat{d}$, be the eigenvectors of $\widehat{\mathbf{K}}(u,v)$ corresponding to the $\widehat{d}$ largest non-zero eigenvalue $\widehat{\theta}_1,\ldots,\widehat{\theta}_{\widehat{d}}$. Then, the eigenfunctons $\widehat{\psi}_j$ of $\widehat{K}(u,v)$, that is the factor loadings in my autocorrelation-based factor model, are obtained by performing the Gram-Schmidt orthonormalization process to the functions

$$\sum_{t=1}^{T-p}\widehat{\gamma}_{t,j}\times(Y_t(\cdot)-\widehat{\mu}(\cdot)), \qquad j=1,\ldots,\widehat{d}, \quad (2.10)$$

then $\widehat{\psi}_1,\ldots,\widehat{\psi}_{\widehat{d}}$ are the orthonormal eigenfunctions of $\widehat{K}(u,v)$ satisfying the identity $\int\widehat{K}(u,v)\,\widehat{\psi}_j(v)\,\mathrm{d}u=\widehat{\theta}_j\widehat{\psi}(v)$. Thus, the fitted yield curve can be defined by

$$\widehat{Y}_t(u)=\widehat{\mu}(u)+\sum_{j=1}^{\widehat{d}}\widehat{\eta}_{tj}\widehat{\psi}_j(u), \qquad t=1,\ldots,T, \qquad j=1,\ldots,\widehat{d}, \qquad u\in[a,b], \quad (2.11)$$

where the eigenfunction $\widehat{\psi}_j$ is the jth factor loading and the factor, obtained by projecting the demeaned yield curve at time t onto the jth loading function, $\widehat{\eta}_{tj}=\langle Y_t-\mu,\widehat{\psi}_j\rangle$ is analogous to the factor $\beta_{i,t}$, $i=1,2,3$, in the DNS model.

Next we discuss the consistency of the estimators. Let $Y_1(u),\ldots,Y_T(u)$ be a sample of the yield curve sequence on the domain $[a,b]$ in a Hilbert space. Assume that $\mu(u)=\mathbb{E}\{Y_t(u)\}=$

[4]The duality property, I refer to Bathia *et al.* (2010), is that $\mathbf{AB}'$ and $\mathbf{B}'\mathbf{A}$ share the same nonzero eigenvalues for any matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes. So if γ is an eigenvector of $\mathbf{B}'\mathbf{A}$, $\mathbf{A}\gamma$ is the eigenvector of $\mathbf{AB}'$ with the same eigenvalue.

$\frac{1}{T}\sum_{t=1}^{T} Y_t(u)$, $u \in [a,b]$ and $t \in 1,\ldots,T$. Following the regularity conditions in Bathia *et al.* (2010) such as ψ-mixing condition, $\mathbb{E}\{\int_a^b Y_t(u)^2\,\mathrm{d}u\}^2 < \infty$, $\mathrm{Cov}\{X_t(u), \varepsilon_{t+k}(v)\} = 0$ for all $u, v \in [a,b]$ and $k \in \mathbb{Z}^+$, and the non-zero eigenvalues θ_j, $j \in \mathbb{Z}^+$ of operator K are monotonically decreasing as j increases. Under these assumptions, the following theorems hold.

Theorem 1. *Let the above conditions hold true. Then as* $T \to \infty$,

(a) $\sup_{u\in[a,b]} |\widehat{\mu}(u) - \mu(u)| = O_P(T^{-1/2})$,

(b) The Hilbert-Schmidt norm for the operator $\widehat{K} - K$, $\|\widehat{K} - K\|_{\mathcal{S}} = O_P(T^{-1/2})$,

(c) $|\widehat{\theta}_j - \theta_j| = O_P(T^{-1/2})$,

(d) $\left(\int_a^b \{\widehat{\psi}_j(u) - \psi_j(u)\}^2\,\mathrm{d}u\right)^{1/2} = O_P(T^{-1/2})$,

for all $j = 1,\ldots,d$ *and* $\theta_1 > \ldots > \theta_d > 0$.

Theorem 2. *Let conditions above hold true. Then as* $T \to \infty$,

(a) $|\widehat{\theta}_j - \theta_j| = O_P(T^{-1})$,

(b) $\left(\int_a^b \{\widehat{\psi}_j(u) - \sum_{j=d+1}^{\infty}\langle\psi_j, \widehat{\psi}_j\rangle\psi_j(u)\}^2\,\mathrm{d}u\right)^{1/2} = O_P(T^{-1/2})$,

for all $j > d$,*and* $j \in \mathbb{Z}^+$.

Theorem 3. *Let conditions above hold true. Assume that* d *is known. Then as* $T \to \infty$, *it holds that the discrepancy between* $\widehat{\mathcal{M}} = \mathrm{span}\{\widehat{\psi}_1,\ldots,\widehat{\psi}_d\}$ *and* $\mathcal{M} = \mathrm{span}\{\psi_1,\ldots,\psi_d\}$, $D(\widehat{\mathcal{M}}, \mathcal{M}) = \sqrt{1 - \frac{1}{d}\sum_{j,k=1}^{d}(\langle\psi_k, \widehat{\psi}_j\rangle)^2}$.

The proof of the above theorems in principle is the same as that of Theorems 1 and 2 in Bathia *et al.* (2010). Therefore I omit it here. Section 2.5 shows that the dynamic space $\mathcal{M}$ spanned by the eigenfunctions of the unobserved stochastic process of the true yield curves $\{X_t(u)\}$ is the same space spanned by $\boldsymbol{\eta}_t$ using observed yield curves $\{Y_t(u)\}$.

2.2.3 Forecasting serially dependent yield curves

Since $\widehat{\boldsymbol{\eta}}_t = (\widehat{\eta}_{t,1},\ldots,\widehat{\eta}_{t,d})'$ carries most of information on the observed yield curves, forecasting the yield curves is equivalent to forecasting these factors. According to Aue *et al.* (2015), the h-step-ahead forecast of a functional object based on a proper multivariate time series model such as a VAR model for the factors is asymptotically equivalent to the forecast of that functional

object based on the p-order functional autoregressive (FAR) model (see Theorem 3.1 in Aue *et al.*, 2015). Moreover, as stated by Aue *et al.* (2015), the bias in the mean squared prediction error (MSPE) caused by applying dimension reduction can be avoided by using a VAR model to the d-dimensional factors. Thus, the selection of the lag order p and the dimension d of the observed yield curves is associated with an appropriate VAR model for $\widehat{\boldsymbol{\eta}}_t$.

The procedure to predict the yield curve is summarized as follows:

1. Transform the observed yield curves $Y_1, \ldots, Y_t$ into a d-dimensional vector time series of factors $\widehat{\boldsymbol{\eta}}_t = (\widehat{\eta}_{t,1}, \ldots, \widehat{\eta}_{t,d})'$, where d and p are fixed.

2. Use a d-dimensional VAR model without the constant term for a stationary process of $\widehat{\boldsymbol{\eta}}_t$ to forecast the h-step-ahead out-of-sample $\widehat{\boldsymbol{\eta}}_{t+h}$.

3. Employ the h-step-ahead out-of-sample forecast $\widehat{\boldsymbol{\eta}}_{t+h}$ to yield the h-step-ahead yield curve forecast $\widehat{Y}_{t+h}$ according to the Karhunen–Loève expansion, jointly with the d-dimensional factor loading(s) $\psi_j(u)$ and the empirical mean function of the observed yield curves $\widehat{\mu}(u)$ for $u \in [a, b]$ available at time t.

Given the information set $I_t = \{Y_1, \ldots, Y_t\}$, where $t < T$, the estimated factors $\widehat{\boldsymbol{\eta}}_t$ can be obtained according to Section 2.2.2. Then the h-step-ahead out-of-sample forecast of the d-dimensional factors for $h \in \mathbb{Z}^+$ in the second step can be expressed by

$$\boldsymbol{\eta}_{t+h|t} = \mathbb{E}_t[\boldsymbol{\eta}_{t+h}|I_t] = \sum_{k=1}^{p} \Gamma_k \boldsymbol{\eta}_{t+h-k|t}. \tag{2.12}$$

Finally, the h-step-ahead out-of-sample forecast of the yield curve is obtained by

$$\begin{aligned} Y_{t+h|t}(u) &= \mathbb{E}[Y_{t+h}(u)|I_t] \\ &= \mu(u) + \Psi(u)\boldsymbol{\eta}_{t+h|t} \\ &= \mu(u) + \Psi(u)\sum_{k=1}^{p} \Gamma_k \boldsymbol{\eta}_{t+h-k|t}, \qquad u \in [a, b], \end{aligned} \tag{2.13}$$

where $\Psi(u) = (\psi_1(u), \ldots, \psi_d(u))$. Note that the constant term is excluded in the VAR model. This is because $\mathbb{E}(\eta_{tj}) = 0$ across time for $j = 1, \ldots, d$ by construction in the framework of functional principle component analysis.

Minimizing the functional Squared Prediction Error (SPE) can be used to choose dimension d in the context of forecasting functional data, see e.g. Hyndman and Ullah (2007). As the factor

loadings ψ_j are orthonormal and factors η_{tj} are uncorrelated by construction, according to Aue *et al.* (2015), the h-step-ahead functional SPE can be decomposed as

$$\begin{aligned}\mathbb{E}\{\|Y_{t+h} - \widehat{Y}_{t+h}\|^2\} &= \mathbb{E}\left\{\left\|\sum_{j=1}^{\infty}\eta_{t+h,j}\psi_j - \sum_{j=1}^{d}\widehat{\eta}_{t+h,j}\psi_j\right\|^2\right\} \\ &= \mathbb{E}\left\{\left\|\boldsymbol{\eta}_{t+h} - \widehat{\boldsymbol{\eta}}_{t+h}\right\|^2\right\} + \sum_{j=d+1}^{\infty}\theta_j \\ &\approx \frac{t+pd}{t-pd}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}) + \sum_{j=d+1}^{\infty}\widehat{\theta}_j,\end{aligned} \tag{2.14}$$

where $\|\cdot\|$ represents the Euclidean norm and $\|\psi_j\| = 1$ for orthonormal eigenfunctions, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ is obtained from a d-dimensional VAR model for a stationary process of $\widehat{\boldsymbol{\eta}}_{t+h} = \sum_{k=1}^{p}\widehat{\boldsymbol{\Gamma}}_k\widehat{\boldsymbol{\eta}}_{T+h-k} + \widehat{\boldsymbol{e}}_{t+h}$, where $\boldsymbol{\Gamma}_k$ is a $d \times d$ matrix of coefficients. The approximation above indicates that the functional SPE is associated with the dimension d and the lag order p, if the observed yield curves are autocorrelated. Hence, I propose a data-driven method based on the Mean Squared Prediction Error (MSPE) to select the dimension d and the lag order p simultaneously.

Similar to studies such as Hyndman and Ullah (2007) and Aue *et al.* (2015), I use the functional MSPE (fMSPE) in the setting of functional data analysis, based on which the fMSPE for jointly determining the dimension d and lag order p is defined as

$$\mathrm{fMSPE}(p,d) = \frac{1}{P}\sum_{t=1}^{P}\int_a^b (Y_{t+h}(u) - \widehat{Y}_{t+h}(u))^2\,\mathrm{d}u, \qquad u \in [a,b], \tag{2.15}$$

where P is the number of observations in the validation set. To choose the dimension and lag order of the yield curves, I use the information set $I_t = \{Y_1, \ldots, Y_{L+m}\}$, where L is the size of training set, which is large enough to produce reliable forecasts of autocorrelation-based factors $\widehat{\boldsymbol{\eta}}_{t+h}$ and thus forecasts of yield curves, and $m = 0, \ldots, P-1$, to obtain the next P periods of h-step-ahead yield curve forecasts $\widehat{Y}_{L+m+h}$. Therefore the dimension d and the lag order p can be simultaneously determined by the minimum of the fMSPEs produced by different combinations of d and p. Finally, using the selected combination of d and p to make the out-of-sample h-step-ahead forecasts of yield curves and evaluating this new autocorrelation-based factor model's performance over $T - P - L$ periods, where $T - P - L$ is the size of the test set.

In terms of the stability in the determination of dimension d and lag order p, I employ the time series cross-validation (CV) to examine how this combination of dimension and lag order is stable over time. This consists of the following steps. First, splitting the data set into 6 equal-

sized subsets, that is of size $L = T/6$. Note that L should be large enough to make reliable forecasts. Second, starting the h-step-ahead out-of-sample forecasting exercise from $L+1$ to $2L$ and obtaining a matrix of fMSPEs corresponding to different combinations of d and p over the periods $L+1,\ldots,2L$. Third, continuing the recursive out-of-sample forecasts for the subsequent subsets $\{2L+1,\ldots,3L\}$,..., and $\{5L+1,\ldots,6L\}$, and obtaining the corresponding matrix of fMSPEs for each subset. That is the size of each validation set is $P = L$. Fourth, computing the element-wise mean of these 5 matrices of fMSPEs and thus the dimension and the lag order are simultaneously determined according to the indices of the minimum element in this computed matrix.

The reason why I consider the MSPE, i.e. a measure of forecasting performance, rather than bootstrapping to select d and p is that the bootstrap test procedure based on resampling the estimated noise term to determine the value of d in Bathia *et al.* (2010) does not work well in the context of forecasting yield curves.[5] More importantly, Bathia *et al.* (2010) mainly focus on identification instead of estimation and prediction, and the bootstrap test to determine the value of d is proposed for the identification problem therein. In addition, the MSPE can be used to implement time series CV, which offers an alternative to the bootstrap method and is one of the most widely-used standard procedures for model evaluation/selection. Thus, considering that the goal in this section is forecasting yield curves, it is reasonable to abandon the bootstrap method and employ the MSPE instead.

2.3 Empirical illustrations

In this section, I consider the monthly constant-maturity zero-coupon Treasury yield curve data of Liu and Wu (2021) to investigate the in-sample fit and the out-of-sample forecasting performance of this new autocorrelation-based factor model. Compared with the popular monthly unsmoothed Fama-Bliss zero-coupon yields of US Treasuries, this innovative data set has less outliers when monthly yields are constructed and less average pricing errors in general. Moreover, this new data set contains representative information in the raw Treasury bond data, and allows for both important global smoothness and crucial local variation.

Like many studies on yield curve modeling and forecasting such as Diebold and Li (2006), Raviv (2015), and Otto and Salish (2019), I select 17 maturities of 3, 6, 9, 12, 15, 18, 21, 24,

[5] I implemented the bootstrap test proposed by Bathia *et al.* (2010) to determine the value of d given different values of p, ranging from 1 to 12, for the yield data. However, I obtained different values of d depending on the choice of p which are all statistically significant and thus there is no unique way to select d and p jointly based on the test results.

30, 36, 48, 60, 72, 84, 96, 108, and 120 months starting in January 1985, and the first forecast is made for the yield curve in January 1994.[6] However, unlike Diebold and Li, Bowsher and Meeks (2008), Koopman *et al.* (2010), Hays *et al.* (2012), and Raviv (2015), whose data samples up to the end of 2000, I consider a data sample up to the end of 2020 and thus the recursive out-of-sample forecasting exercise can be made up to and including December 2020. This is because some novel patterns of yield curve shapes have occurred since 2000 and more often in recent years. For instance, a small hump in the short end of the yield curve arose 4 times in the second half of the year 2020. These new patterns can be explained by the effects of the unconventional tools of monetary policies, e.g. forward guidance and large-scale asset purchases (LSAP, well-known as quantitative easing) on the shape of the yield curve (see e.g., Swanson and Williams, 2014; Wu and Xia, 2016; Swanson, 2021), since the the conventional instrument federal funds rate has been at the zero lower bound for many years after December 2008. It is worthwhile to use more recent data to examine whether this novel autocorrelation-based factor model is able to reproduce both the stylized facts of the yield curve, as documented in Diebold and Li (2006), and some new yield curve shapes since then. Therefore, I use the yield data with 17 fixed maturities from January 1985 to December 2020 ($T = 432$).

Table 2.1 provides some descriptive statistics for the yield data over two periods, 1985–2000 and 2001–2020. These summary statistics reveal the common stylized facts of the yield curve and some new information. First, the average yield curve over the period of 2001–2020 is still increasing across maturities, although it is pushed downward. However, it displays a non-parallel downward shift, suggesting steeper slope and less curvature in the average yield curve over 2001–2020. More specifically, the 2-year and 3-year yields fall more significantly than the other yields, resulting a concave downward curve for maturities from 3-month to 12-month, a concave upward curve for maturities from 12-month to 24-month, and then a concave downward curve again for longer maturities. Second, yields are persistent and stronger with the length of maturities. More persistence presents in the yields over the period of 2001–2020 for all maturities. Third, short rates are more volatile than long rates. Moreover, the short-term interest rates at 3-month, 6-month, 9-month, and 12-month maturities over the period of 2001–2020 are more volatile than those interest rates in the former period, while the longer rates over the period of 2001–2020, instead, are less volatile than those rates in the former period.

Figure 2.1 depicts a three-dimensional plot of the yield curve data set used in this section and shows some examples of two consecutive yield curves. This plot suggests that the data are prone

[6]The challenge with going beyond 10 years of maturities is that no 15, 20 or 25 year bond is issued during this sample period. The issuance of 30-year maturity Treasury bonds is irregular. A 20-year maturity bond is a 30-year maturity bond that was issued 10 years ago, and generally such bonds are less liquid.

Table 2.1: Descriptive statistics for yield curves

Maturity	Panel A: 1985–2000							Panel B: 2001–2020						
(Month)	Mean	SD	Min	Max	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$	$\widehat{\rho}(30)$	Mean	SD	Min	Max	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$	$\widehat{\rho}(30)$
3	5.647	1.473	2.789	9.090	0.979	0.576	-0.073	1.375	1.517	0.024	5.064	0.979	0.632	0.041
6	5.818	1.485	2.904	9.334	0.978	0.568	-0.031	1.459	1.524	0.049	5.131	0.980	0.651	0.066
9	5.966	1.491	3.016	9.493	0.975	0.555	-0.001	1.528	1.515	0.082	5.184	0.981	0.662	0.088
12	6.095	1.493	3.145	9.646	0.971	0.540	0.027	1.584	1.501	0.103	5.189	0.980	0.668	0.107
15	6.208	1.492	3.304	9.893	0.968	0.527	0.058	1.632	1.485	0.109	5.164	0.979	0.672	0.128
18	6.302	1.488	3.477	10.108	0.966	0.514	0.087	1.680	1.470	0.114	5.132	0.978	0.674	0.153
21	6.373	1.482	3.643	10.261	0.963	0.502	0.114	1.733	1.455	0.120	5.106	0.977	0.677	0.180
24	6.433	1.473	3.780	10.381	0.961	0.493	0.139	1.787	1.441	0.124	5.086	0.976	0.679	0.208
30	6.549	1.458	4.023	10.662	0.958	0.481	0.188	1.903	1.420	0.114	5.064	0.974	0.679	0.255
36	6.657	1.443	4.197	10.838	0.955	0.472	0.230	2.029	1.397	0.123	5.050	0.973	0.679	0.293
48	6.844	1.439	4.327	11.240	0.952	0.458	0.295	2.273	1.350	0.172	5.029	0.969	0.673	0.346
60	6.963	1.425	4.344	11.348	0.952	0.463	0.341	2.487	1.309	0.231	5.049	0.967	0.668	0.393
72	7.081	1.453	4.376	11.645	0.953	0.452	0.373	2.688	1.271	0.313	5.172	0.966	0.663	0.425
84	7.156	1.436	4.352	11.810	0.951	0.455	0.395	2.863	1.257	0.382	5.307	0.965	0.667	0.456
96	7.228	1.414	4.420	11.603	0.953	0.470	0.416	3.014	1.248	0.445	5.438	0.965	0.670	0.478
108	7.291	1.421	4.500	11.741	0.953	0.470	0.427	3.149	1.246	0.488	5.554	0.965	0.671	0.494
120 (level)	7.345	1.409	4.589	11.657	0.955	0.480	0.442	3.285	1.264	0.530	5.673	0.966	0.682	0.515
Slope	1.698	1.179	-0.605	4.088	0.964	0.406	-0.066	1.910	1.213	-0.528	4.358	0.965	0.536	-0.218
Curvature	-0.126	0.622	-1.999	1.631	0.896	0.304	-0.051	-1.085	0.830	-2.695	0.344	0.950	0.604	0.046

Notes: This table provides descriptive statistics for US monthly yields over two periods 1985–2000 and 2001–2020. I present mean, standard deviation (SD), minimum (MIN), maximum (MAX), 1 month $\widehat{\rho}(1)$ and 1 year $\widehat{\rho}(12)$ sample autocorrelation coefficients for each maturity. I also present the statistics for empirical proxies for the level, slope, and curvature of the yield curve. As defined by Diebold and Li (2006), the proxy for level factor is the 120-month yield, for slope it is the 120-month yield minus the 3-month yield, and for curvature it is two times the 24-month yield minus the 3-month and 120-month yields.

to serial dependence between two consecutive observations of yield curves, although the short-term yields are more volatile than the long-term yields. Hence, the raw yield curves seem to be serially dependent and the serial dependence among yield curves may be informative in terms of modeling and forecasting.

For the comparison of this novel method and the competing prediction methods documented in the literature such as Diebold and Li (2006), Jungbacker *et al.* (2014), Van Dijk *et al.* (2014) and Otto and Salish (2019), the first forecast is made for yield curve in January 1994. Hence, the number of factors d and the lag order of information p are determined according to the first 108 observations of the yield curves from January 1985 to December 1993. The size of training set L is set to 60 and the validation set is set to 48 to ensure an effective selection of d and p according to the calculated fMSPE for the validation set. The recursive out-of-sample forecasts are made for the yield curves over the period January 1994 up to December 2020.

In the extant literature on yield curve forecasting, it has proven to be challenging to outperform the random walk model at the 1-month-ahead horizon. Thus, I focus on the novel method's performance in out-of-sample 1-month-ahead forecasting experiment in this section, and set $h = 1$ in Eq. (2.15) when the forecast horizon is 1-month.

Then, I find that the determined dimension d and lag order p are 3 and 1 respectively, which jointly produce the smallest value of the fMSPEs, fMSPE $= \frac{1}{P}\sum_{t=1}^{P}\int_a^b (Y_{t+1}(u) - \widehat{Y}_{t+1}(u))^2 \,\mathrm{d}u$, where $u \in [3, 120]$, according to the procedure described in Section 2.2.3, with $P = 48$ periods extending from January 1990 to December 1993. Figure 2.2(a) shows the corresponding three-dimensional surface plot of the fMSPEs depending on different values of d and p and the valley (minima of fMSPEs) corresponding with the combination of $d = 3$ and $p = 1$. Figure 2.2(b) confirms that this choice is stable over time.[7] Therefore, the combination of $d = 3$ and $p = 1$ is selected for the implementation of the 1-month-ahead out-of-sample forecasting exercise, beginning in January 1994 and extending through December 2020.

[7]This 36-year yield data set is split into six 6-year subsets, 72 observations for each subset, for implementing the time series CV. I also implemented the time series CV when the size of the first 6 or 4 sets is set to 60 or 84 and obtained the same choice of d and p.

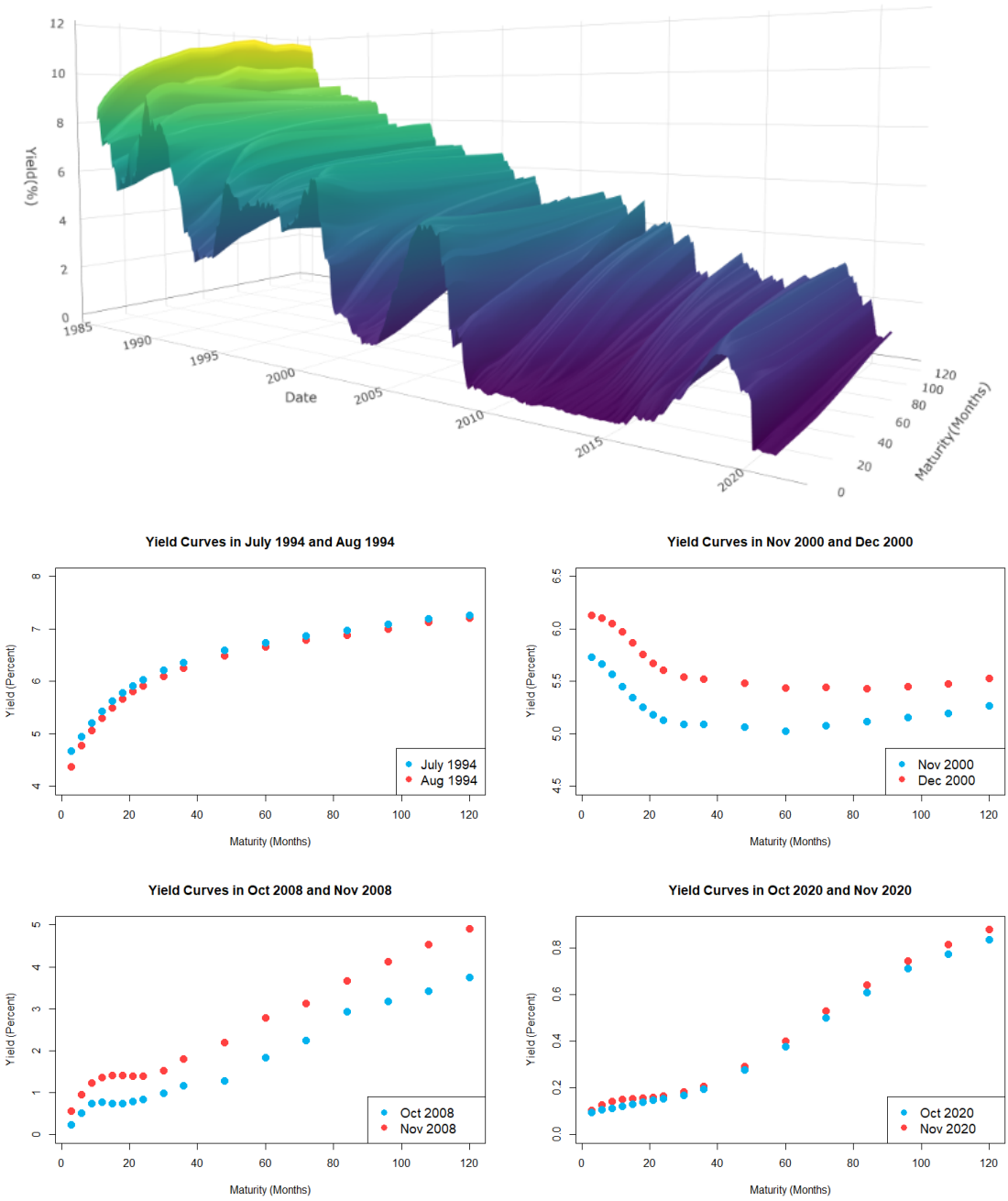
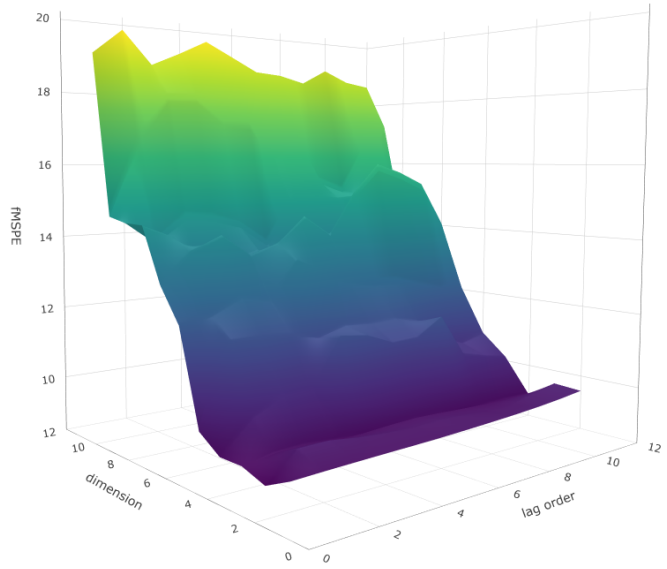


Figure 2.1: Autocorrelated yield curves (monthly constant maturity zero-coupon Treasury yield) and examples of two consecutive raw yield curve data. This data set consists of monthly yield data at maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months from January 1985 to December 2020.

(a)



(b)

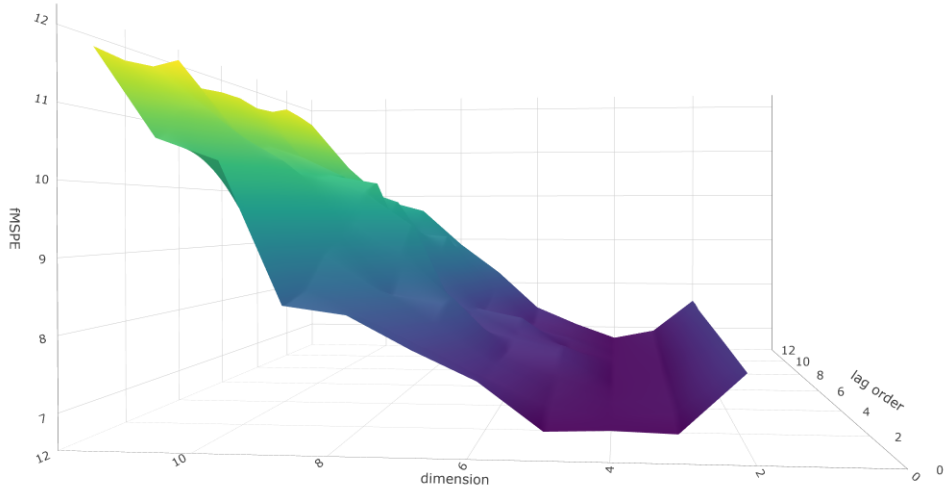


Figure 2.2: Three-dimensional surface plots of functional Mean Squared Prediction Errors (fMSPEs) depending on different values of dimension and lag order. The minimum value of fMSPEs is reached when $d = 3$ and $p = 1$ for both the validation set and the time series cross-validation data. The validation data set for subplot (a) ranges from January 1990 to December 1993. The data set for implementing time series cross-validation for subplot (b) ranges from January 1985 to December 2020 and the size of each subset is 72.

2.3.1 Fitting yield curves

Before using this new autocorrelation-based three-factor model, the discrete yield data $Y_{t,i}$, where i ranges from the maturities $\{3, \ldots, 120\}$, needs to be transformed into a functional time series $Y_t(u)$, $u \in [3, 120]$. Therefore, the cubic B-spline expansion as discussed in Section 2.2.1 is applied to create $Y_t(u)$. Like Otto and Salish (2019), the roughness penalty parameter is set to 10^{-8} to ensure twice continuous differentiability.

This new autocorrelation-based factor model in Eq. (2.2) also links with the DNS model. Let $\mu(u) = 0$, dimension $d = 3$, factor loadings $\psi_1(u) = 1$, $\psi_2(u) = \frac{1-\mathrm{e}^{-\lambda u}}{\lambda u}$, $\psi_3(u) = \frac{1-\mathrm{e}^{-\lambda u}}{\lambda u} - \mathrm{e}^{-\lambda u}$, where the maturities $u = i$ and $\lambda = 0.0609$, and use the data without transformation, the model in Eq. (2.2) becomes the DNS model. Note that the dynamic structure of the yield curve depending on 1-month lag information is reduced to a vector process lying in a 3-dimensional space, according to the data-driven method described in Section 2.2.3. This coincides with the number of factors in Litterman and Scheinkman (1991)'s seminal work, which employs PCA to the US Treasury data to find the common factors, to explain term structure movements, and the number of factors in the DNS models proposed by Diebold and Li (2006). Figure 2.3 illustrates the similarities and differences between the DNS factor loadings and the new autocorrelation-based factor loadings (given $p = 1$) based on the full data sample.

Now I assess Diebold and Li's (2006) three-factor model of level, slope and curvature and the three-factor model that driving the serial dependence of yield curves in fitting the yield data over time. Figure 2.4 displays the model-implied average yield curves of the DNS model and my autocorrelation-based factor model, as well as the actual average yield curve. Compared with the DNS model (blue curve), modeling the yield curves based on the factors driving serial dependence across yield curves (red curve) can provide more flexibility to describe the shape of yield curves and thus may improve the goodness-of-fit of model. Figure 2.5, furthermore, illustrates the improvement of the in-sample fit for some selected dates.

I also consider Otto and Salish's (2019) dynamic functional factor model for yield curves as a benchmark model for comparison. In this approach, the factor loadings are the eigenfunctions of the sample covariance operator, and the factors are the projections of the demeaned yield curve onto the corresponding factor loadings.[8] Relative to Otto and Salish's (2019) dynamic functional factor model, proposed based on FPCA as well, factors and factor loadings in my autocorrelation-based factor model are created based on the autocovariance functions. More specifically, factor loadings, as discussed in Section 2.2.2, are estimated by projecting the eigenfunctions of the operator related to the lagged covariance functions onto the demeaned yield curve after applying

[8] I obtained the factor loadings and factors by the code shared by Sven Otto.

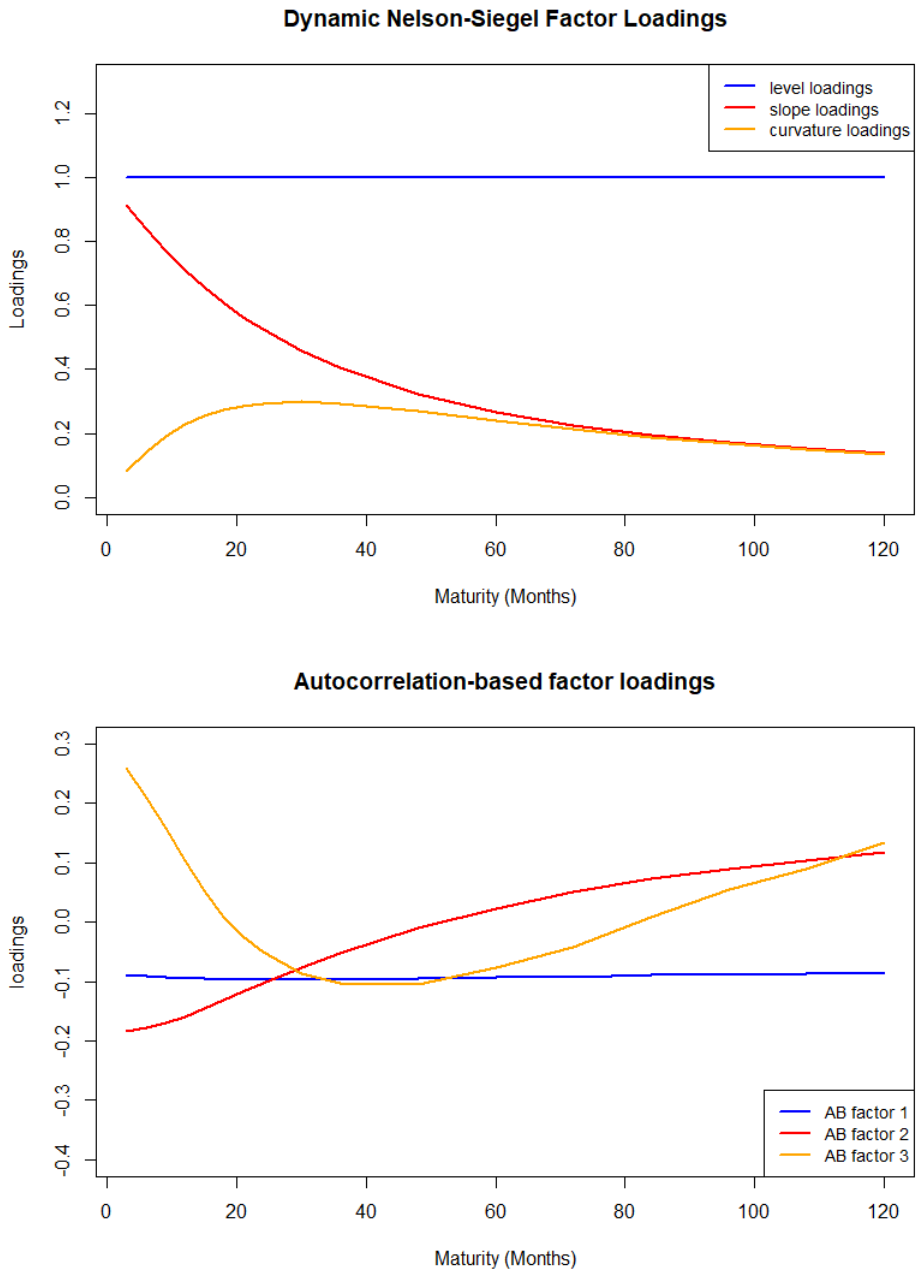


Figure 2.3: Nelson-Siegel factor loadings and autocorrelation-based factor loadings that account for serial dependence (lag order $p = 1$) across yield curves. The data set is from January 1985 to December 2020.

the duality property, and then the corresponding factors are obtained by projecting the factor loadings onto the demeaned yield curve like Otto and Salish (2019). Therefore, I can investigate the improvement in in-sample fit caused by considering factors driving the serial dependence among

yield curves in yield curve modeling.

Table 2.2 presents the descriptive statistics for the yield curve residuals from my new autocorrelation-based three-factor model with 1-month lag information. Table 2.3 describes the fit of the three-factor DNS model and the model proposed by Otto and Salish (2019) with the specification of three factors. It can be seen that the yield curve residuals from my autocorrelation-based three-factor model's fit exhibit zero mean and less autocorrelation. This is closer to the assumption of white noise in the context of factor models.

Table 2.2: Descriptive statistics for yield curve residuals from the new autocorrelation-based three-factor model ($p = 1$)

Maturity (Month)	Mean	SD	Min	Max	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
3	0.004	0.136	-0.639	0.299	0.811	0.113
6	-0.001	0.055	-0.235	0.120	0.837	0.111
9	-0.000	0.024	-0.054	0.115	0.666	0.190
12	0.000	0.039	-0.107	0.208	0.758	0.102
15	-0.000	0.048	-0.137	0.259	0.796	0.092
18	0.000	0.050	-0.139	0.224	0.816	0.106
21	0.000	0.048	-0.113	0.144	0.833	0.118
24	-0.000	0.044	-0.106	0.129	0.843	0.132
30	0.000	0.034	-0.084	0.146	0.805	0.124
36	0.000	0.029	-0.073	0.172	0.830	0.023
48	0.000	0.027	-0.202	0.091	0.761	0.095
60	-0.000	0.036	-0.209	0.112	0.845	-0.052
72	0.000	0.045	-0.167	0.213	0.877	0.112
84	-0.000	0.032	-0.145	0.167	0.750	-0.008
96	0.000	0.022	-0.134	0.103	0.757	0.181
108	0.000	0.026	-0.093	0.093	0.822	0.087
120	0.000	0.047	-0.122	0.145	0.820	0.129

Notes: I fit the autocorrelation-based three-factor model with 1-month lag information, using the yield data from January 1985 to December 2020. SD is short for standard deviation. Min and Max are the minimum and maximum values of yield curve residuals, respectively. $\widehat{\rho}(\cdot)$ represents the sample autocorrelations of yield curve residuals at lags 1 and 12 months.

Table 2.4 summarizes the in-sample fit performance according to the measures of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The last two columns in each panel present the ratios of the measures achieved by the autocorrelation-based model to those measures achieved by the DNS model or by the FPCA model. A number smaller than one indicates reductions in these measures achieved by this autocorrelation-based three-factor model. Overall, the results in this table provide evidence that the autocorrelation-based three-factor model has better in-sample fit than the three-factor DNS model and Otto and Salish's (2019) dynamic functional factor model with the specification of three factors. Compared with DNS model's fit, the average

Table 2.3: Descriptive statistics for yield curve residuals from Diebold and Li's (2006) Nelson–Siegel model (DNS) and Otto and Salish's (2019) dynamic functional factor model (FPCA)

Maturity	Panel A: DNS						Panel B: FPCA					
(Month)	Mean	SD*	Min	Max	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	Min	Max	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.070**	1.653	-0.536	0.192	0.962	0.320	**-0.042**	0.879	-0.720	0.292	0.983	1.103
6	**-0.008**	3.139	-0.053	0.047	1.340	0.424	**-0.036**	0.757	-0.303	0.109	0.981	0.840
9	**0.028**	0.580	-0.098	0.248	0.840	0.595	**-0.029**	0.661	-0.138	0.083	0.779	0.453
12	**0.043**	0.729	-0.139	0.363	0.902	0.324	**-0.025**	0.943	-0.173	0.191	0.903	0.280
15	**0.045**	1.000	-0.131	0.348	0.946	0.326	**-0.021**	0.982	-0.194	0.234	0.955	0.355
18	**0.038**	1.339	-0.090	0.246	0.977	0.378	**-0.018**	0.981	-0.176	0.190	0.975	0.474
21	**0.023**	1.685	-0.051	0.120	1.007	0.346	**-0.016**	0.972	-0.149	0.105	0.981	0.533
24	**0.005**	1.861	-0.071	0.074	1.083	0.433	**-0.016**	0.958	-0.132	0.095	0.981	0.555
30	**-0.021**	1.128	-0.169	0.153	1.091	2.001	**-0.015**	0.916	-0.103	0.109	0.966	0.523
36	**-0.035**	0.664	-0.201	0.164	1.007	0.114	**-0.014**	0.889	-0.110	0.136	0.961	0.119
48	**-0.045**	0.409	-0.471	0.144	0.878	0.222	**-0.010**	0.904	-0.228	0.071	0.946	0.398
60	**-0.056**	0.549	-0.376	0.127	0.945	-0.154	**-0.007**	0.980	-0.231	0.117	0.984	-1.369
72	**-0.037**	0.678	-0.334	0.277	0.970	0.358	**-0.005**	0.999	-0.175	0.213	0.995	0.961
84	**-0.020**	0.855	-0.157	0.193	0.961	-0.101	**-0.003**	1.007	-0.145	0.166	0.996	-2.845
96	**0.004**	0.790	-0.146	0.112	0.973	0.778	-0.002	1.006	-0.143	0.102	0.998	1.096
108	**0.033**	0.597	-0.083	0.258	0.958	0.288	-0.001	0.977	-0.090	0.092	0.989	0.654
120	**0.072**	0.607	-0.133	0.458	0.909	0.273	0.000	0.978	-0.148	0.157	0.991	0.768

Notes: Using the yield data from January 1985 to December 2020, I fit the Diebold and Li's (2006) three-factor model of level, slope, and curvature (DNS), and Otto and Salish's (2019) dynamic functional factor model with the specification of three factors (FPCA), where factors are obtained based on the functional principal components of the empirical covariance operator. The superscript of * represents a ratio calculated by dividing a measure (Standard Deviation, $\widehat{\rho}(1)$, or $\widehat{\rho}(12)$) for yield curve residuals reported in Table 2.2 by the corresponding measure for yield curve residuals from the DNS, or from the FPCA. A number smaller than 1 in the last two columns in Panels A and B implies that the new autocorrelation-based three-factor model with 1-month lag information reduces autocorrelation in yield curve residuals. Bold numbers indicate a residual mean significantly different from zero at the 5% level of significance.

Table 2.4: Summary of in-sample fit performance

Maturity	Panel A: RMSE			Panel B: MAE		
(Month)	AB	relative to DNS	relative to FPCA	AB	relative to DNS	relative to FPCA
3	0.136	1.259	0.849	0.101	1.184	0.895
6	0.055	2.854	0.679	0.044	2.803	0.725
9	0.024	0.483	0.517	0.019	0.498	0.523
12	0.039	0.568	0.813	0.028	0.513	0.777
15	0.048	0.730	0.902	0.037	0.676	0.912
18	0.050	0.943	0.925	0.039	0.900	0.941
21	0.048	1.311	0.924	0.038	1.358	0.939
24	0.044	1.812	0.907	0.035	1.926	0.920
30	0.034	0.931	0.851	0.027	0.946	0.856
36	0.029	0.517	0.820	0.021	0.459	0.782
48	0.027	0.339	0.861	0.020	0.335	0.859
60	0.036	0.415	0.962	0.026	0.399	1.006
72	0.044	0.592	0.993	0.032	0.601	1.019
84	0.032	0.749	1.002	0.022	0.750	1.015
96	0.022	0.781	1.003	0.014	0.711	1.030
108	0.026	0.471	0.976	0.019	0.481	0.984
120	0.047	0.445	0.978	0.037	0.462	1.003
Average	0.050	**0.787**	**0.852**	0.033	**0.734**	**0.878**

Notes: I present results for 3 three-factor models: autocorrelation-based model (AB), Diebold and Li's (2006) model of level, slope, and curvature (DNS), and Otto and Salish's (2019) dynamic factor model with the specification of three factors (FPCA). For each maturity, I calculate two measures, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The last two columns in each panel present the ratios of RMSEs and MAEs achieved by the AB model to those measures achieved by the DNS model or by the FPCA model. A number smaller than one indicates reductions in these measures are achieved by this autocorrelation-based three-factor model.

reductions in RMSE and MAE arrive at 21.3% and 26.6%, respectively. Specifically, the reduction in these measures is significant at maturities ranging from 9-month to 18-month and at maturities longer than 2-year. Relative to Otto and Salish's (2019) model, RMSE and MAE are, on average, reduced by 14.8% and 12.2% respectively. Specifically, considering factors driving serial dependence among yield curves significantly improves in-sample fit at shorter maturities.

To investigate how robust the choice of $d = 3$ and $p = 1$ is both over time and across maturities, I use a 5-year rolling window rather than an expanding window for estimation to achieve the RMSEs at all 17 maturities over time. Figure 2.6 suggests that the choice of d and p seems to be stable over time but the RMSEs at shorter maturities are more volatile than those at longer maturities.

Further, I set $d = 4$ or $p = 2$ as robustness checks. Panels A–B of Table 2.5 show that increasing the number of factors in this autocorrelation-based model can improve in-sample fit. Panels C–D of Table 2.5 suggest that increasing the lag order improves in-sample fit as well. The improvement associated d is compatible with some extensions of DNS models, see, e.g., Söderlind and Svensson (1997) under a parametric framework and coincides with the so-called "blessing of dimensionality" in the context of high-dimensional time series. The improvement related to p in in-sample fit is not as significant as that related to d.

Figure 2.7 shows the 5-year rolling RMSEs for the combinations of $d = 4$ and $p = 1$ and of $d = 3$ and $p = 2$ used for model estimation at all 17 maturities over time, and delivers similar patterns as in Figure 2.6 when $d = 3$ and $p = 1$.

Furthermore, Figure 2.8 presents three factors estimated by my autocorrelation-based model, DNS model, and Otto and Salish's (2019) functional dynamic factor model. Unlike $\hat{\beta}_t$ from the DNS and factors in Otto and Salish's (2019) model (i.e. the so-called FPC scores), the autocorrelation-based factors seem to be non-stationary. For instance, $\widehat{\eta}_{t1}$ looks like non-stationary with a trend and a drift. Therefore, I use an Augmented Dicky-Fuller (ADF) unit root test (Dickey and Fuller, 1979) to test for stationarity of these autocorrelation-based factors. The number of lags in the ADF regression is determined by minimizing the Akaike Information Criterion (AIC). The test results show that the first autocorrelation-based factor cannot reject the null hypothesis of unit root at the 1% significance level when both the drift term and the trend term are considered, cause the test statistic of -3.95 is slightly larger than the critical value (-3.98 at the 1% level). The second autocorrelation-based factor is nonstationary at the 5% level when the drift term is included, and at the 1% level when the drift term is excluded. The third autocorrelation-based factor is stationary at any conventional significance level. However, the test results show that I can reject the null hypothesis of unit root at any conventional significance level for the changes in

Table 2.5: Summary statistics for yield curve residuals from the autocorrealtion-based factor model

	AB factor model ($d = 4$, $p = 1$)							
Maturity	Panel A: relative to DNS				Panel B: relative to FPCA			
(Month)	MAE*	RMSE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	MAE*	RMSE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	0.467	0.480	0.815	0.462	0.353	0.323	0.834	1.592
6	0.647	0.667	0.982	1.112	0.167	0.159	0.719	2.204
9	0.489	0.481	0.835	0.591	0.514	0.515	0.774	0.450
12	0.389	0.400	0.833	0.486	0.589	0.573	0.834	0.421
15	0.362	0.385	0.876	0.606	0.488	0.476	0.884	0.661
18	0.397	0.430	0.925	0.680	0.414	0.422	0.924	0.852
21	0.532	0.539	0.914	0.317	0.368	0.380	0.891	0.487
24	0.739	0.736	0.868	-0.087	0.353	0.369	0.787	-0.112
30	0.563	0.588	0.882	0.465	0.510	0.538	0.781	0.122
36	0.410	0.470	0.966	0.072	0.699	0.745	0.923	0.075
48	0.219	0.248	0.757	0.012	0.561	0.630	0.816	0.022
60	0.238	0.280	0.875	-0.479	0.602	0.648	0.911	-4.266
72	0.386	0.399	0.961	0.451	0.655	0.669	0.986	1.211
84	0.690	0.732	0.967	-0.423	0.934	0.979	1.002	-11.960
96	0.674	0.732	0.955	0.691	0.977	0.939	0.979	0.973
108	0.284	0.317	0.944	0.168	0.582	0.657	0.974	0.381
120	0.320	0.315	0.845	0.349	0.695	0.693	0.922	0.980
	AB factor model ($d = 3$, $p = 2$)							
Maturity	Panel C: relative to DNS				Panel D: relative to FPCA			
(Month)	MAE*	RMSE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	MAE*	RMSE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	1.194	1.271	0.962	0.298	0.902	0.857	0.984	1.030
6	2.851	2.902	1.340	0.386	0.737	0.691	0.981	0.765
9	0.501	0.486	0.843	0.574	0.526	0.519	0.781	0.437
12	0.510	0.565	0.901	0.314	0.772	0.807	0.902	0.272
15	0.674	0.728	0.945	0.313	0.910	0.900	0.954	0.341
18	0.901	0.943	0.976	0.358	0.942	0.924	0.974	0.449
21	1.365	1.313	1.006	0.315	0.944	0.925	0.980	0.485
24	1.941	1.820	1.082	0.374	0.927	0.911	0.981	0.479
30	0.957	0.939	1.093	1.644	0.865	0.858	0.967	0.430
36	0.464	0.521	1.007	0.077	0.790	0.826	0.962	0.080
48	0.332	0.336	0.876	0.195	0.853	0.854	0.945	0.349
60	0.398	0.413	0.944	-0.179	1.004	0.956	0.983	-1.598
72	0.600	0.591	0.969	0.331	1.017	0.990	0.994	0.887
84	0.752	0.750	0.961	-0.081	1.018	1.003	0.996	-2.303
96	0.710	0.779	0.972	0.768	1.029	1.000	0.997	1.082
108	0.478	0.468	0.957	0.276	0.978	0.969	0.988	0.626
120	0.460	0.442	0.907	0.254	0.998	0.972	0.988	0.715

Notes: The superscript of * represents the ratio calculated by dividing a measure (MAE, RMSE, $\widehat{\rho}(1)$ or $\widehat{\rho}(12)$) for yield curve residuals from the new autocorrealtion-based (AB) factor model by the corresponding measure for yield curve residuals from Diebold and Li's (2006) factor model of level, slope and curvature (DNS) or from Otto and Salish's (2019) dynamic functional factor model with the specification of three factors (FPCA). A number smaller than 1 implies improvement in in-sample fit or reductions in autocorrelation.

autocorrelation-based factors. Neither the trend term nor the drift term is significantly different from 0 for the first autocorrelation-based factor.[9] Note that the second step in the procedure determining the lag order p and the number of factors d in Section 2.2.3 focuses on a stationary process of these autocorrelation-based factors. Hence, $\boldsymbol{\eta_t}$ is replaced by changes in $\boldsymbol{\eta_t}$ for the VAR model to select the lag order and the dimension of yield curves. And the corresponding h-step-ahead forecast of η_{t+h} is then obtained by adding back the level of η_t. Compared with other term structure models viewing the yield curve as a functional times series (see e.g., Bowsher and Meeks, 2008; Almeida *et al.*, 2018), this approach is more parsimonious to account for nonstationary factors.[10]

Recall that Hörmann *et al.* (2015) argue that the dimension reduction relying on the covariance function of the functional data may not be adequate when this approach is applied to serially dependent data. The first autocorrelation-based factor explains more than 99.96% of variation of the yield data, which is higher than the first factor in Otto and Salish's (2019) dynamic functional factor model (98.03%). The second autocorrelation-based factor explains more than 0.03% of variation. Therefore, the first two factors of my autocorrelation-based model can jointly explain more than 99.99% of variation in the serially dependent yields. In contrast, Otto and Salish's (2019) dynamic functional factor model needs at least 12 factors to achieve the same explanatory power using the same yield data.

[9]The drift term is not statistically significant at the 5% level when the trend term is excluded. When the trend term is included, the drift term is insignificant at the 10% level and the trend term is insignificant at any conventional significance level.

[10]These models adopt Error Correction Models (ECM) for factors' dynamics.

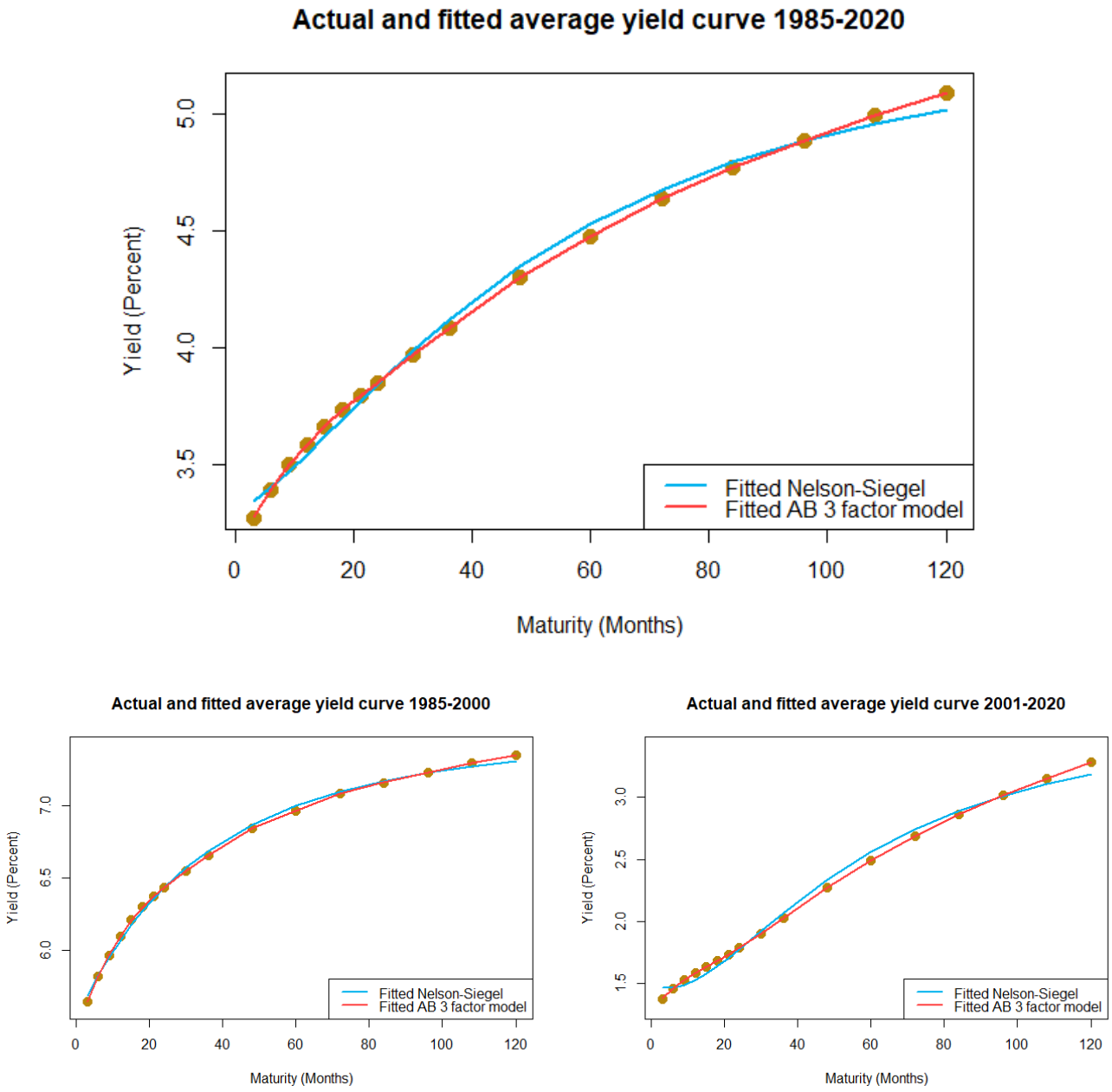


Figure 2.4: Actual and fitted yield curves. The dotted curve represents the actual data, the blue line is obtained by replicating Diebold and Li's (2006) method ($\lambda = 0.0609$), and the red line is obtained by my new autocorrelation-based (AB) three-factor model ($p = 1$).

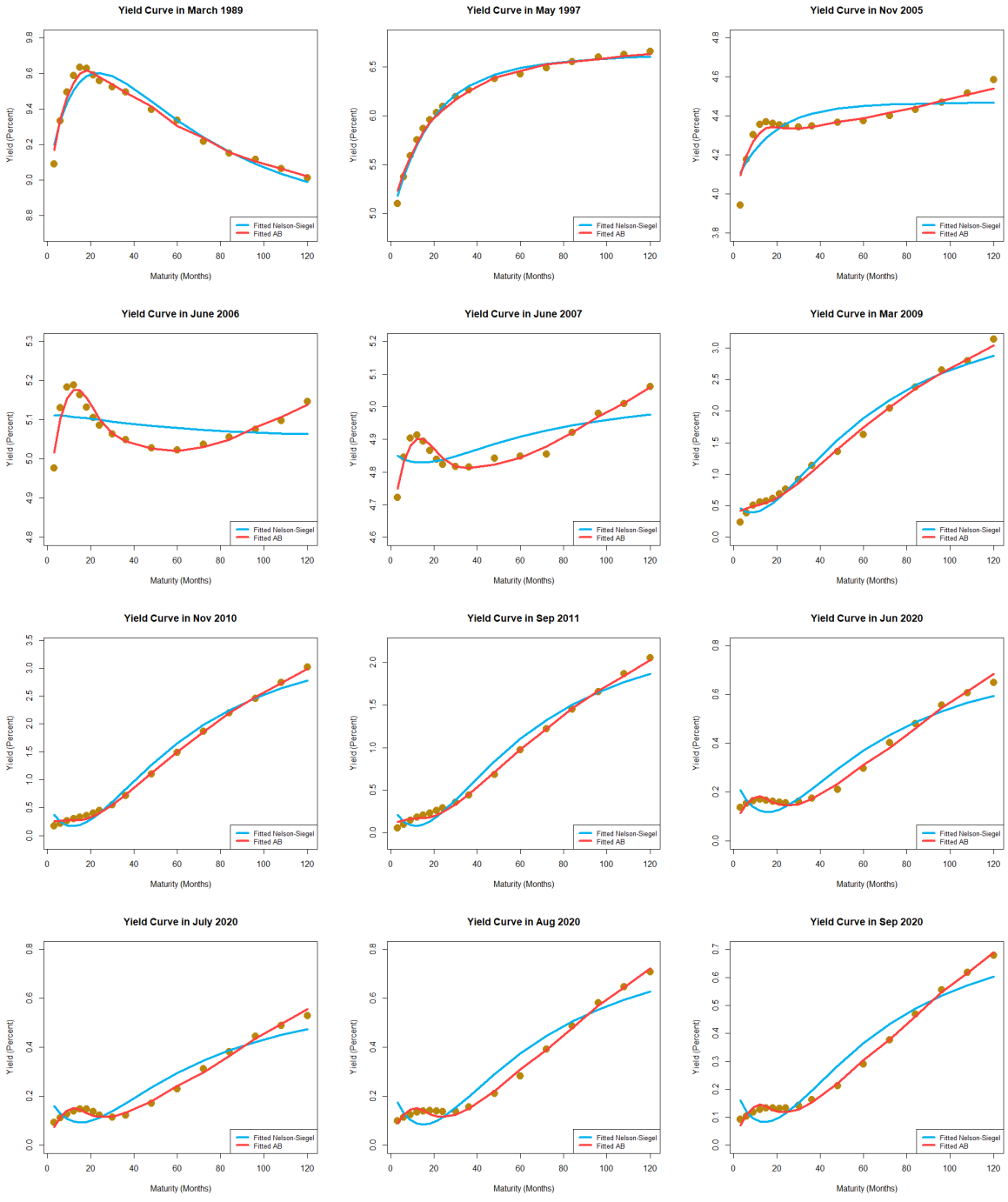


Figure 2.5: Actual and fitted yield curves. The dotted curve represents the actual data, the blue line is obtained by replicating Diebold and Li's (2006) method ($\lambda = 0.0609$), and the red line is obtained by my new autocorrelation-based (AB) three-factor model ($p = 1$).

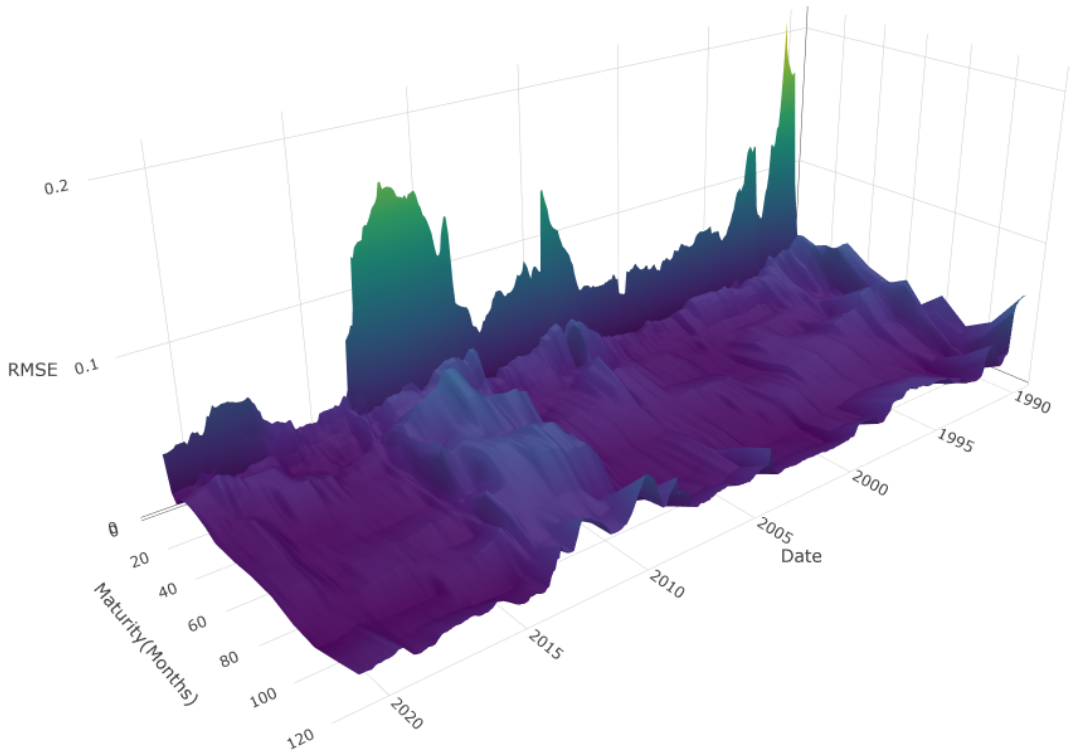
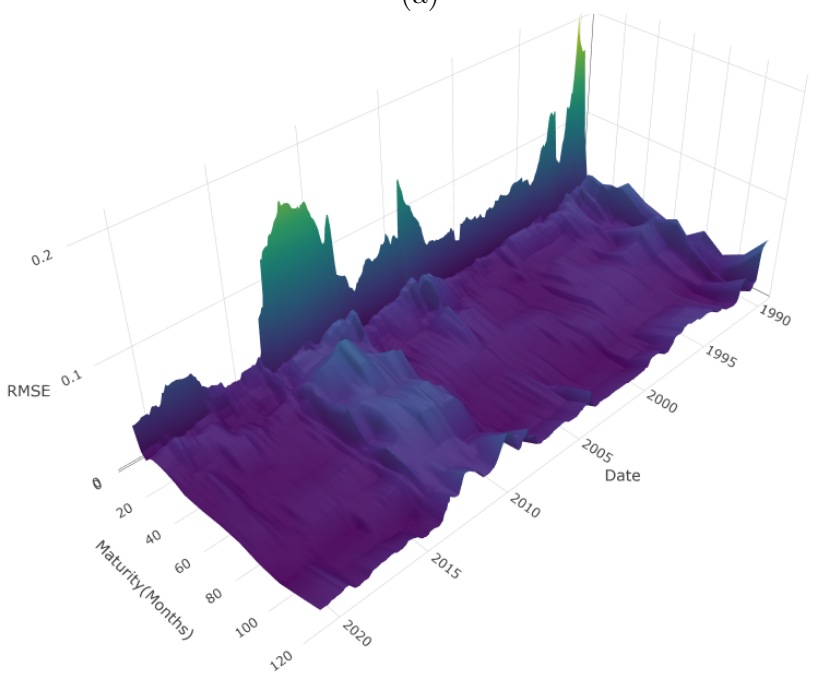


Figure 2.6: Five years rolling RMSEs, $p = 1$ and $d = 3$

(a)



(b)

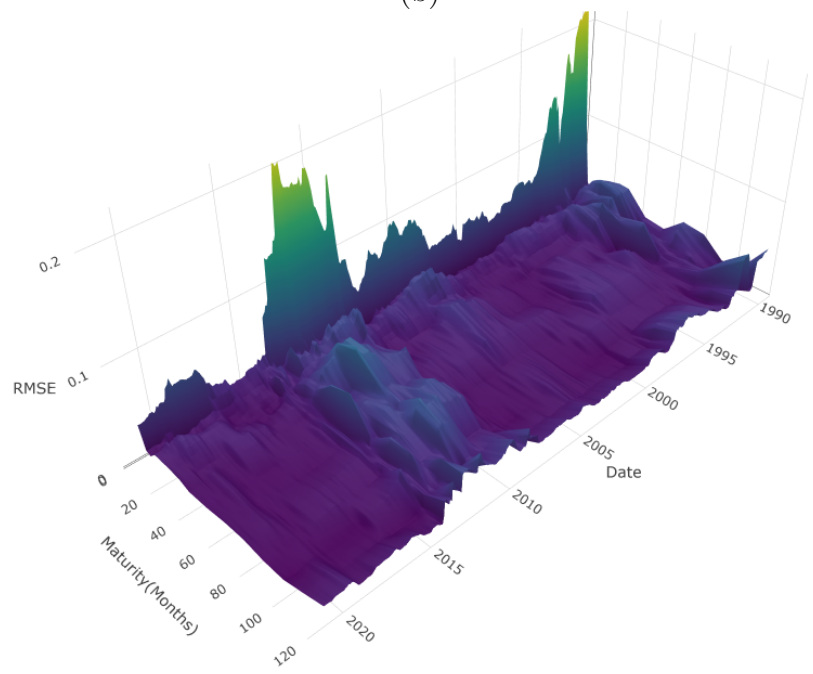


Figure 2.7: Five years rolling RMSEs. (a) $p = 1$ and $d = 4$. (b) $p = 2$ and $d = 3$.

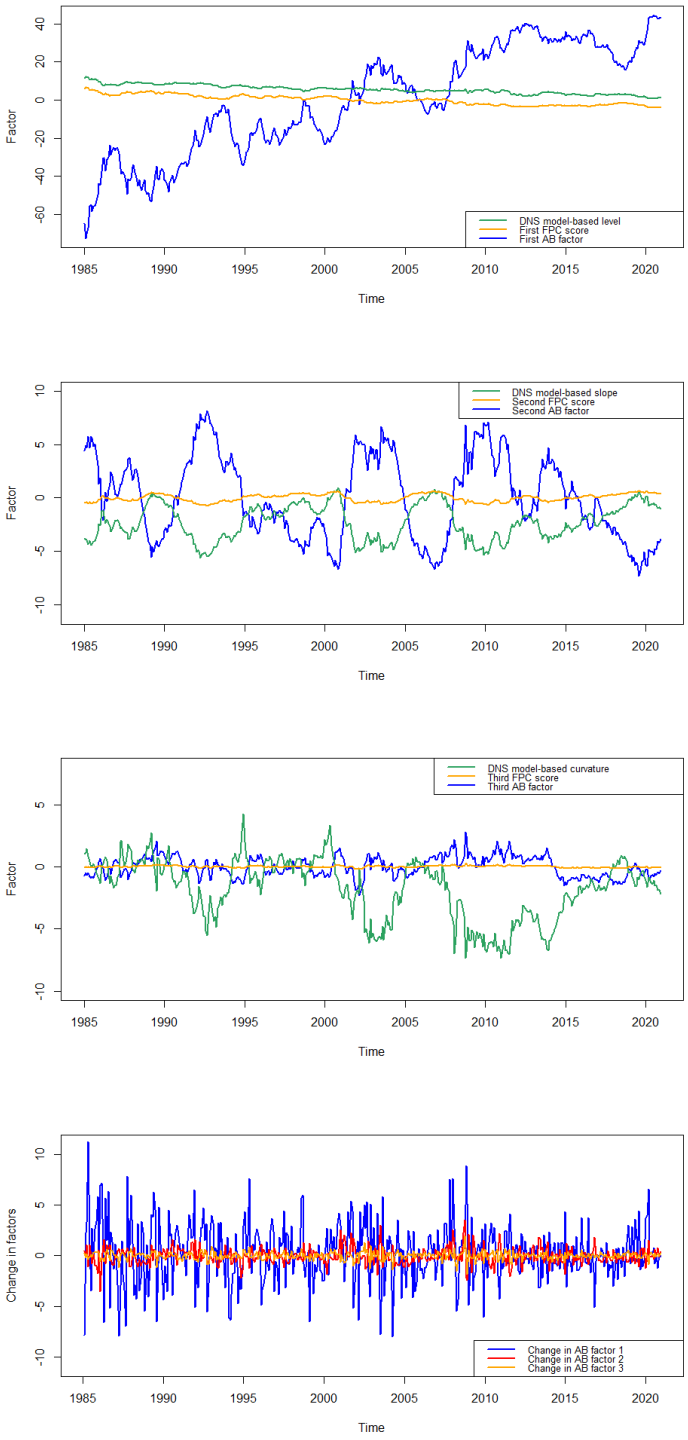


Figure 2.8: Three factors from DNS model (level, slope and curvature), autocorrelation-based (AB) factor model, and Otto and Salish's (2018) functional dynamic factor model (FPC scores). The last plot shows the changes in three AB factors.

2.3.2 Forecasting yield curves

A good in-sample fit does not guarantee good out-of-sample forecasting performance. According to the procedure proposed in Section 2.2.3, the optimal number of factors and the lag order are $d = 3$ and $p = 1$ for this new autocorrelation-based factor model for forecasting. Diebold and Li's (2006) DNS model of three factors is used as an important benchmark for comparing the out-of-sample 1-month-ahead forecasting performance. Besides, two extensions of the DNS model allowing for shifting endpoints of the yield factors are used as competitor models as well. Following Van Dijk *et al.* (2014), the first specification called ESL is based on exponential smoothing for level factor and the second specification called ESLSC is based on exponential smoothing for the three factors of level, slope and curvature obtained from the DNS model. In addition, Otto and Salish's (2019) dynamic functional factor model based on FPCA and the regular PCA for vector-valued yield curves both with the specification of three factors are also considered as competitors. Moreover, when the forecast horizon is 1 month, it seems hard to outperform the random walk (RW) forecast based on the results of Table 4 in Diebold and Li (2006). Therefore, besides these five three-factor models above, the RW with "no change" forecast is used as the last benchmark model for comparison.

Following the procedure of forecasting yield curves in Diebold and Li (2006), the AR(1) specification is employed to obtain the 1-month-ahead forecasts of the three factors in the setting of DNS model and the factor loadings are determined by the fixed value $\lambda = 0.0609$. Consistent with Van Dijk *et al.* (2014), the exponential smoothing parameter α is set to 0.1 for monthly data. The 1-month-ahead RW forecast is specified by $\widehat{Y}_{t+1,i|t} = Y_{t,i}$, where i denotes maturities and $i \in \{3, \ldots, 120\}$. In terms of the forecasts of the dynamic functional factor model of Otto and Salish (2019), the factors are forecasted by a VAR(1) model due to the lagged cross-correlations present in these factors. I also use a VAR(1) model to obtain the forecasts of the first three principle components for PCA.

The out-of-sample 1-month-ahead forecasting performance for maturity i, ranging from $\{3, \ldots, 120\}$, can be evaluated according to the Root Mean Squared Prediction Error (RMSPE), which can be obtained by

$$\text{RMSPE}_i = \sqrt{\frac{1}{P}\sum_{t \in P}(Y_{t+1,i} - \widehat{Y}_{t+1|t}(u_i))^2}, \quad u_i = i, \tag{2.16}$$

where P is the number of evaluation periods. The smaller the RMSPE, the better the forecasting performance. I use the evaluation period from January 1994 up to December 2020 and thus $P = 324$.

Tables 2.6–2.8 present the out-of-sample 1-month-ahead forecasting results. Specifically, Table 2.6 reports the forecasting results using the autocorrelation-based factor model. Tables 2.7–2.8 report the forecasting results using benchmark models relative to this new three-factor model. In Tables 2.7–2.8 I use asterisks behind the measures, SD, RMSPE, $\widehat{\rho}(1)$ and $\widehat{\rho}(12)$, to denote the ratios of these measures for the prediction errors in Table 2.6 to those measures for the prediction errors from the competitor models. An absolute value of these ratios smaller than 1 implies that the new autocorrelation-based three-factor model with 1-month lag information improves the out-of-sample 1-month-ahead forecasting performance and/or reduces autocorrelation in prediction errors. Surprisingly, the 3-dimensional autocorrelation-based factor model outperforms all competitor models with the specification of three factors: DNS, FPCA, PCA, ESL, and ESLSE at almost all maturities when the forecast horizon is 1 month. There exist occasional ratios of RMSPE smaller but very close to 1, however, I find that the Diebold and Mariano (1995) test statistics are insignificant when I test the statistical significance for them. Thus, I conclude that this new 3-dimensional autocorrelation-based factor model systematically outperforms the other 5 competitor models with 3 factors.

Table 2.6: Out-of-sample 1-month-ahead forecasting results of 3-dimensional autocorrelation-based factor model with 1-month lag information (1994–2020)

Maturity (Month)	Mean	SD	RMSPE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
3	**-0.028**	0.214	0.215	0.232	0.051
6	-0.012	0.183	0.183	0.072	0.002
9	-0.003	0.187	0.187	0.036	0.007
12	0.002	0.204	0.203	0.059	0.031
15	0.004	0.220	0.220	0.067	0.055
18	0.004	0.232	0.232	0.056	0.071
21	0.001	0.240	0.240	0.042	0.078
24	-0.004	0.246	0.246	0.027	0.079
30	-0.016	0.252	0.253	-0.011	0.076
36	-0.024	0.256	0.257	-0.034	0.087
48	-0.019	0.266	0.267	-0.027	0.118
60	-0.025	0.267	0.268	-0.011	0.089
72	0.004	0.266	0.266	-0.019	0.082
84	-0.003	0.263	0.263	-0.020	0.042
96	-0.024	0.261	0.261	-0.009	0.018
108	-0.020	0.261	0.261	-0.007	0.001
120	**-0.032**	0.263	0.265	0.020	-0.002

Notes: A bold number indicates that the mean of the prediction errors significantly different from zero at the 5% level of significance.

Figure 2.9 shows some examples of the cumulative Squared Prediction Errors (SPE) at selected maturities over time. The new autocorrelation-based factor model's forecasting performance is

Table 2.7: Out-of-sample 1-month-ahead forecasting results (1994–2020)

Maturity	Panel A: RW					Panel B: DNS				
(Month)	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	-0.009	1.062	1.070	0.897	0.720	**-0.136**	0.960	0.827	0.551	0.478
6	-0.010	0.946	0.947	0.196	0.226	**-0.065**	0.918	0.874	0.171	0.039
9	-0.010	0.929	0.928	0.100	-3.661	**-0.024**	0.912	0.906	0.085	0.133
12	-0.011	0.960	0.959	0.182	4.045	-0.008	0.929	0.928	0.143	0.463
15	-0.011	0.990	0.989	0.231	2.387	-0.011	0.947	0.946	0.171	0.725
18	-0.012	1.010	1.009	0.213	2.082	**-0.023**	0.963	0.959	0.154	0.953
21	-0.012	1.020	1.019	0.174	1.908	**-0.039**	0.976	0.964	0.126	1.137
24	-0.013	1.020	1.019	0.117	1.829	**-0.056**	0.982	0.958	0.087	1.288
30	-0.013	1.006	1.006	-0.055	1.820	**-0.085**	0.979	0.931	-0.040	1.366
36	-0.014	0.996	0.999	-0.197	1.768	**-0.103**	0.973	0.910	-0.135	1.214
48	-0.014	0.999	1.000	-0.212	1.858	**-0.121**	0.962	0.884	-0.108	1.014
60	-0.015	1.000	1.003	-0.099	1.664	**-0.129**	0.966	0.879	-0.047	0.921
72	-0.015	1.010	1.008	-0.205	2.038	**-0.114**	0.982	0.905	-0.097	1.067
84	-0.015	1.000	0.999	-0.256	-34.051	**-0.087**	0.996	0.945	-0.141	1.521
96	-0.015	1.001	1.003	-0.116	-1.215	**-0.055**	0.996	0.979	-0.076	2.865
108	-0.016	1.007	1.008	-0.096	-0.065	-0.021	1.002	1.001	-0.070	-0.153
120	-0.016	1.020	1.025	0.369	0.076	0.028	0.989	0.991	0.174	-0.531
	Panel C: FPCA					Panel D: PCA				
	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.085**	1.130	1.039	0.848	1.399	**-0.041**	1.131	1.114	2.770	3.670
6	**-0.072**	1.035	0.960	0.286	0.216	**-0.039**	1.029	1.007	2.031	0.167
9	**-0.062**	0.987	0.937	0.127	0.303	**-0.042**	0.979	0.956	0.728	0.164
12	**-0.058**	0.982	0.945	0.192	0.765	**-0.045**	0.975	0.953	0.937	0.479
15	**-0.058**	0.991	0.959	0.217	1.011	**-0.049**	0.982	0.960	1.065	0.689
18	**-0.061**	1.002	0.970	0.190	1.184	**-0.053**	0.988	0.964	1.027	0.819
21	**-0.067**	1.012	0.974	0.152	1.307	**-0.058**	0.992	0.964	0.900	0.918
24	**-0.074**	1.017	0.972	0.104	1.409	**-0.065**	0.992	0.960	0.671	1.001
30	**-0.089**	1.023	0.965	-0.051	1.637	**-0.077**	0.986	0.945	-0.345	1.055
36	**-0.098**	1.027	0.960	-0.184	1.660	**-0.085**	0.978	0.934	-1.031	0.975
48	**-0.094**	1.036	0.975	-0.177	1.409	**-0.079**	0.968	0.932	-0.444	0.904
60	**-0.098**	1.036	0.972	-0.075	1.418	**-0.085**	0.970	0.930	-0.162	0.859
72	**-0.070**	1.036	0.999	-0.152	1.391	**-0.058**	0.977	0.956	-0.432	0.852
84	**-0.076**	1.033	0.990	-0.184	1.699	**-0.069**	0.987	0.955	-1.726	0.818
96	**-0.091**	1.029	0.972	-0.087	1.196	**-0.088**	0.986	0.939	-0.477	0.517
108	**-0.089**	1.035	0.978	-0.082	0.144	**-0.085**	0.985	0.940	-0.301	0.049
120	**-0.101**	1.030	0.964	0.205	-0.122	**-0.098**	0.978	0.925	0.329	-0.052

Notes: I present the out-of-sample 1-month-ahead forecasting results of competitor models, the random walk (RW), dynamic Nelson-Siegel (DNS), functional dynamic factor model (FPCA), and the three-factor model based on the vector-valued PCA for the expanding window from January 1994 to December 2020. The superscript of * represents the ratio of measures, calculated by dividing a measure (SD, RMSPE, $\widehat{\rho}(1)$, or $\widehat{\rho}(12)^*$) for the prediction errors reported in Table 2.6 by the corresponding measure for the prediction errors from the competitor models. An absolute value of these ratios smaller than 1 implies that the proposed 3-dimensional autocorrelation-based factor model with 1-month lag information improves the out-of-sample 1-month-ahead forecasting performance or reduces autocorrelation in the prediction errors. Bold numbers indicate that the mean of the prediction errors significantly different from zero at the 5% level of significance.

Table 2.8: Out-of-sample 1-month-ahead forecasting results (1994–2020)

Maturity	Panel A: ESL					Panel B: ESLSC				
(Month)	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.120**	0.907	0.815	0.482	0.602	**-0.108**	0.887	0.816	0.445	0.422
6	**-0.050**	0.866	0.845	0.147	0.086	**-0.036**	0.839	0.829	0.134	0.030
9	-0.008	0.872	0.871	0.074	0.427	0.007	0.846	0.846	0.069	0.124
12	0.008	0.900	0.899	0.126	1.129	0.024	0.878	0.873	0.120	0.509
15	0.005	0.925	0.925	0.150	1.523	0.022	0.907	0.903	0.145	0.852
18	-0.008	0.945	0.945	0.133	2.017	0.010	0.929	0.928	0.130	1.144
21	-0.024	0.959	0.955	0.107	2.535	-0.006	0.944	0.943	0.105	1.362
24	**-0.041**	0.965	0.953	0.072	2.918	-0.023	0.950	0.947	0.070	1.504
30	**-0.070**	0.960	0.930	-0.032	2.301	**-0.052**	0.950	0.933	-0.031	1.430
36	**-0.087**	0.952	0.910	-0.102	1.493	**-0.070**	0.946	0.920	-0.103	1.210
48	**-0.105**	0.939	0.883	-0.080	1.118	**-0.090**	0.939	0.897	-0.082	1.057
60	**-0.114**	0.942	0.879	-0.034	1.116	**-0.100**	0.944	0.893	-0.035	1.052
72	**-0.098**	0.965	0.909	-0.065	1.402	**-0.087**	0.967	0.922	-0.067	1.353
84	**-0.072**	0.979	0.946	-0.081	3.557	**-0.061**	0.983	0.958	-0.085	3.531
96	**-0.040**	0.987	0.980	-0.040	-1.643	**-0.031**	0.990	0.988	-0.042	-1.609
108	-0.005	1.003	1.006	-0.037	-0.040	0.003	1.004	1.007	-0.039	-0.044
120	**0.044**	0.999	0.993	0.106	0.086	**0.051**	0.997	0.986	0.108	0.123

Notes: I present the out-of-sample 1-month-ahead forecasting results of competitor models, exponential smoothing for level factor (ESL) and exponential smoothing for level, slope, and curvature (ESLSC). The superscript of * represents a measure of ratio, calculated by dividing a measure (SD, RMSPE, $\widehat{\rho}(1)$, or $\widehat{\rho}(12)$) for the prediction errors reported in Table 2.6 by the corresponding measure for the prediction errors from the competitor models. An absolute value of these ratios smaller than 1 implies that the proposed 3-dimensional autocorrelation-based factor model with 1-month lag information improves the out-of-sample 1-month-ahead forecasting performance or reduces autocorrelation in the prediction errors. Bold numbers indicate that the mean of the prediction errors significantly different from zero at the 5% level of significance.

one of the best among these 7 parametric, semiparametric, and nonparametric methods and the superiority of this new factor model's forecasts is stable over time.

Since yields are highly persistent over time and very close to a unit root process, it is difficult for dynamic factor models to outperform a random walk at the 1-month-ahead horizon. However, some earlier studies on forecasting yield curves, such as Bowsher and Meeks (2008), Van Dijk *et al.* (2014), and Raviv (2015), show that it is possible to outperform the 1-month-ahead random walk forecasts when the forecasts are made over the period of 1994–2000. On the other hand, results in Table 2.1 indicate that persistence in yields is stronger over the period of 2001–2020, perhaps suggesting that it is more challenging to outperform the 1-month-ahead random walk forecasts over this period. Hence, I examine the out-of-sample forecasting performance for two subsample periods, 1994–2000 and 2001–2020, as a robustness check of subsample stability. Tables 2.9–2.11 report the forecasting results. Panel A in Table 2.10 shows that this novel 3-dimensional autocorrelation-based factor model's forecasts outperform the RW forecasts over the period of 1994–2000 as well and with less autocorrelation in prediction errors.[11] The forecasting results in Tables 2.10 and 2.11 suggest that my new autocorrelation-based three-factor model with 1-month lag information is favored relative to the other three-factor models in forecasting yield curves at the 1-month-ahead horizon over two sub-periods 1994–2000 and 2001–2020 as well.

[11]The superiority of this autocorrelation-based factor model in forecasting yield curves lasts up to the end of 2014 when the forecasts are made recursively. In the earlier versions of this chapter the forecasts from this autocorrelation-based three-factor model also outperform the random walk forecasts up to the end of 2016, using the popular unsmoothed Fama-Bliss zero yields for US Treasuries of maturities 3, 6, 9, 12, 15, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months over the period January 1985 up to December 2016.

Table 2.9: Out-of-sample 1-month-ahead forecasting results of 3-dimensional autocorrelation-based factor model with 1-month lag information (subsample periods: 1994–2000 and 2001–2020)

Maturity	Panel A: 1994–2000					Panel B: 2001–2020				
(Month)	Mean	SD	RMSPE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$	Mean	SD	RMSPE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
3	**0.042**	0.160	0.164	0.152	0.128	**-0.053**	0.225	0.231	0.186	0.024
6	0.027	0.159	0.161	0.179	0.013	**-0.026**	0.189	0.191	-0.004	0.028
9	0.017	0.186	0.186	0.197	-0.051	-0.010	0.187	0.187	-0.057	0.062
12	0.012	0.212	0.211	0.195	-0.079	-0.002	0.201	0.201	-0.024	0.097
15	0.010	0.232	0.231	0.184	-0.089	0.002	0.216	0.215	-0.006	0.126
18	0.009	0.247	0.246	0.175	-0.085	0.002	0.227	0.226	-0.015	0.144
21	0.009	0.259	0.257	0.173	-0.079	-0.002	0.234	0.234	-0.032	0.152
24	0.007	0.267	0.265	0.168	-0.071	-0.008	0.239	0.238	-0.050	0.154
30	0.000	0.272	0.270	0.133	-0.055	-0.021	0.246	0.246	-0.083	0.148
36	-0.002	0.271	0.270	0.104	-0.047	**-0.032**	0.251	0.252	-0.100	0.156
48	0.015	0.270	0.268	0.069	-0.041	-0.031	0.265	0.266	-0.074	0.182
60	-0.001	0.269	0.268	0.104	-0.059	-0.033	0.267	0.268	-0.059	0.150
72	0.032	0.261	0.262	0.059	-0.046	-0.006	0.268	0.268	-0.052	0.128
84	0.022	0.262	0.261	0.088	-0.064	-0.012	0.263	0.263	-0.063	0.083
96	-0.016	0.254	0.253	0.101	-0.074	-0.026	0.264	0.264	-0.046	0.053
108	-0.015	0.252	0.251	0.103	-0.075	-0.021	0.264	0.264	-0.042	0.030
120	-0.031	0.251	0.251	0.128	-0.075	-0.032	0.268	0.269	-0.012	0.025

Notes: Bold numbers indicate that the mean of prediction errors is significantly different from zero at the 5% level of significance.

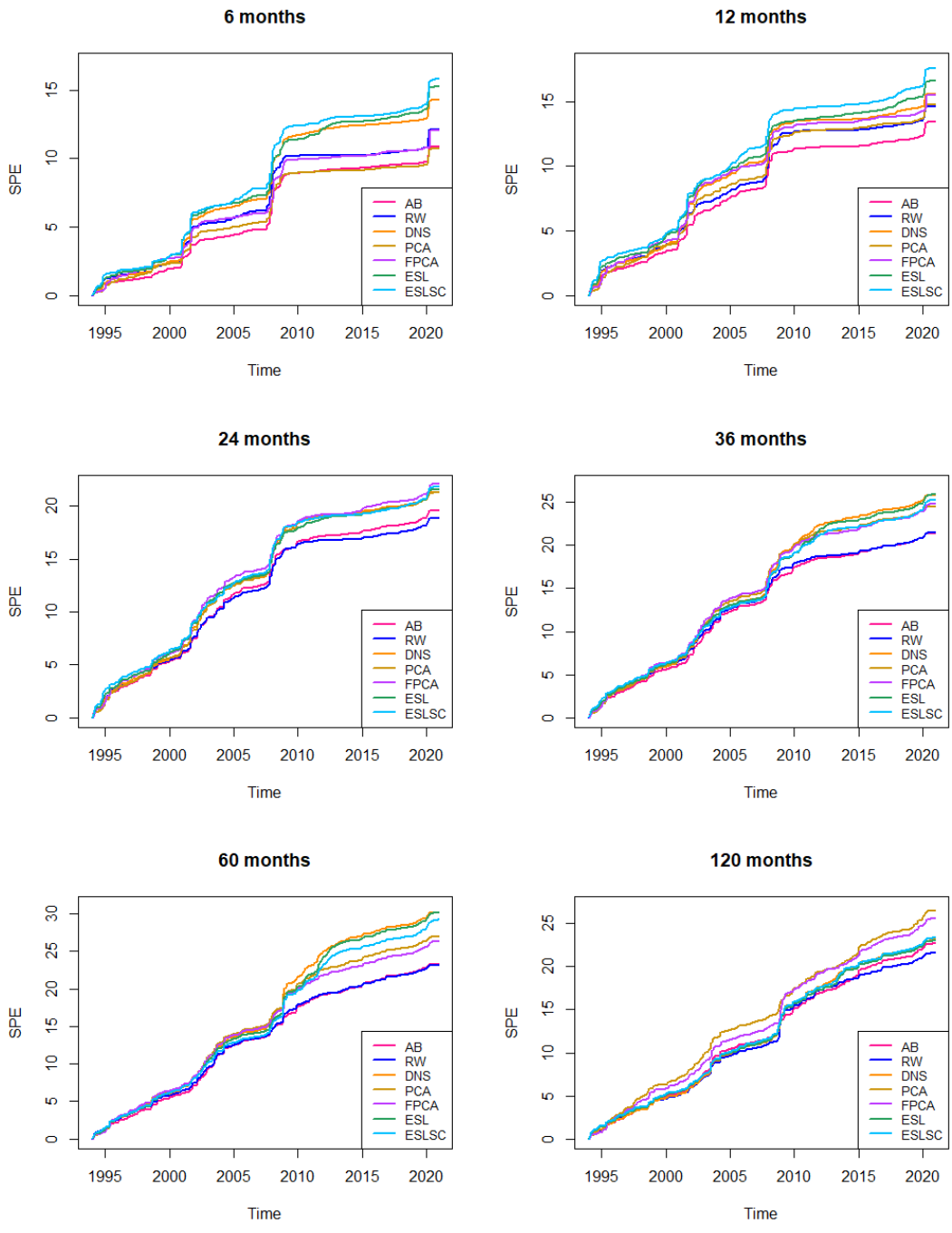


Figure 2.9: Cumulative sum of Squared Prediction Errors (SPE) of monthly yield data at maturities of 6, 12, 24, 36, 60, 120 months from January 1994 to December 2020.

Table 2.10: Out-of-sample 1-month-ahead forecasting results (subsample periods: 1994–2000 and 2001–2020)

Maturity	Panel A: RW (1994–2000)					Panel B: RW (2001–2020)				
(Month)	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	0.031	0.906	0.924	0.525	1.616	-0.024	1.081	1.104	0.910	0.282
6	0.028	0.890	0.893	0.418	-0.132	-0.023	0.960	0.962	-0.015	0.377
9	0.025	0.907	0.904	0.476	0.378	-0.023	0.941	0.936	-0.197	0.752
12	0.021	0.928	0.926	0.492	0.561	-0.022	0.978	0.973	-0.097	0.939
15	0.018	0.947	0.946	0.480	0.625	-0.022	1.014	1.009	-0.028	1.005
18	0.016	0.965	0.964	0.467	0.600	-0.021	1.034	1.029	-0.085	1.035
21	0.013	0.982	0.981	0.472	0.570	-0.021	1.041	1.036	-0.207	1.050
24	0.011	0.992	0.992	0.470	0.528	-0.021	1.035	1.032	-0.353	1.063
30	0.008	0.984	0.984	0.390	0.428	-0.021	1.016	1.016	-0.718	1.081
36	0.006	0.971	0.970	0.320	0.383	-0.021	1.007	1.012	-1.111	1.121
48	0.001	0.958	0.960	0.242	0.359	-0.020	1.012	1.015	-1.274	1.257
60	-0.002	0.969	0.969	0.376	0.495	-0.019	1.010	1.015	-1.326	1.173
72	-0.004	0.966	0.973	0.219	0.379	-0.019	1.023	1.020	-2.235	1.188
84	-0.005	0.985	0.989	0.334	0.496	-0.019	1.004	1.002	-6.945	1.584
96	-0.007	0.986	0.988	0.394	0.580	-0.018	1.005	1.008	-3.998	1.647
108	-0.008	1.005	1.006	0.413	0.587	-0.018	1.008	1.009	-2.953	1.228
120	-0.009	1.015	1.022	0.539	0.563	-0.018	1.021	1.026	8.736	1.303
	Panel C: DNS (1994–2000)					Panel D: DNS (2001–2020)				
	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.058**	0.901	0.885	0.412	16.424	**-0.163**	0.974	0.818	0.500	0.185
6	-0.017	0.874	0.883	0.380	-0.133	**-0.082**	0.932	0.872	-0.012	0.196
9	0.017	0.909	0.910	0.422	0.385	**-0.038**	0.920	0.905	-0.159	0.361
12	0.036	0.936	0.926	0.437	0.534	-0.024	0.935	0.930	-0.068	0.496
15	0.040	0.945	0.934	0.426	0.570	**-0.029**	0.958	0.951	-0.018	0.615
18	0.030	0.950	0.944	0.411	0.541	**-0.042**	0.981	0.965	-0.052	0.720
21	0.012	0.959	0.958	0.414	0.518	**-0.057**	0.995	0.967	-0.127	0.810
24	-0.008	0.967	0.967	0.414	0.495	**-0.074**	0.998	0.954	-0.223	0.891
30	-0.034	0.975	0.967	0.357	0.431	**-0.103**	0.990	0.917	-0.426	0.969
36	-0.047	0.975	0.961	0.299	0.387	**-0.122**	0.981	0.892	-0.550	0.959
48	-0.058	0.973	0.953	0.213	0.347	**-0.142**	0.966	0.862	-0.367	0.904
60	**-0.072**	0.974	0.942	0.314	0.500	**-0.149**	0.971	0.859	-0.329	0.860
72	**-0.072**	0.983	0.955	0.196	0.428	**-0.129**	0.985	0.890	-0.358	0.902
84	-0.056	0.992	0.974	0.297	0.513	**-0.098**	0.998	0.936	-0.803	0.968
96	-0.032	0.996	0.991	0.373	0.571	**-0.063**	0.998	0.975	-0.817	0.964
108	-0.014	1.006	1.006	0.387	0.540	-0.023	1.000	1.000	-0.955	0.803
120	0.008	1.008	1.015	0.483	0.577	0.035	0.985	0.984	-0.171	0.644
	Panel E: FPCA (1994–2000)					Panel F: FPCA (2001–2020)				
	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.087**	1.032	0.929	0.639	0.591	**-0.087**	1.094	1.035	0.601	1.039
6	**-0.079**	0.965	0.881	0.527	0.458	**-0.068**	1.027	0.971	-0.019	0.625
9	**-0.068**	0.949	0.900	0.514	0.836	**-0.057**	0.977	0.938	-0.235	0.672
12	**-0.058**	0.953	0.923	0.495	0.786	**-0.054**	0.963	0.932	-0.090	0.754
15	-0.052	0.956	0.936	0.469	0.736	**-0.057**	0.966	0.936	-0.022	0.830
18	-0.048	0.958	0.942	0.444	0.660	**-0.063**	0.976	0.942	-0.063	0.900
21	-0.047	0.961	0.947	0.438	0.605	**-0.072**	0.985	0.943	-0.155	0.964
24	-0.050	0.965	0.950	0.431	0.563	**-0.081**	0.990	0.938	-0.281	1.036
30	**-0.061**	0.969	0.946	0.368	0.475	**-0.097**	0.996	0.929	-0.636	1.189
36	**-0.069**	0.971	0.942	0.310	0.446	**-0.108**	0.998	0.925	-0.968	1.250
48	**-0.066**	0.980	0.955	0.238	0.453	**-0.104**	1.001	0.938	-0.788	1.197
60	**-0.086**	0.975	0.931	0.334	0.676	**-0.102**	1.004	0.945	-0.718	1.239
72	**-0.062**	0.984	0.965	0.226	0.644	**-0.072**	1.000	0.966	-0.690	1.281
84	**-0.076**	0.987	0.952	0.319	0.643	**-0.076**	0.999	0.961	-1.378	1.417
96	**-0.106**	0.991	0.917	0.377	0.688	**-0.086**	0.997	0.953	-1.136	1.397
108	**-0.106**	1.001	0.924	0.402	0.646	**-0.080**	1.001	0.961	-1.631	1.090
120	**-0.121**	1.000	0.906	0.475	0.689	**-0.089**	0.997	0.953	-0.286	0.950

Notes: Bold numbers indicate that the mean of prediction errors is significantly different from zero at the 5% level of significance.

Table 2.11: Out-of-sample 1-month-ahead forecasting results (subsample periods: 1994–2000 and 2001–2020)

Maturity	Panel A: PCA (1994–2000)					Panel B: PCA (2001–2020)				
(Month)	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	**-0.041**	1.131	1.114	2.770	3.670	**-0.039**	1.129	1.138	3.703	-12.523
6	**-0.039**	1.029	1.007	2.031	0.167	**-0.031**	1.035	1.030	0.117	1.442
9	**-0.042**	0.979	0.956	0.728	0.164	**-0.031**	0.978	0.968	1.850	0.884
12	**-0.045**	0.975	0.953	0.937	0.479	**-0.034**	0.972	0.959	2.122	0.869
15	**-0.049**	0.982	0.960	1.065	0.689	**-0.039**	0.980	0.965	0.489	0.876
18	**-0.053**	0.988	0.964	1.027	0.819	**-0.045**	0.987	0.969	0.605	0.883
21	**-0.058**	0.992	0.964	0.900	0.918	**-0.053**	0.990	0.966	0.887	0.902
24	**-0.065**	0.992	0.960	0.671	1.001	**-0.061**	0.989	0.959	1.177	0.929
30	**-0.077**	0.986	0.945	-0.345	1.055	**-0.076**	0.979	0.940	2.223	0.941
36	**-0.085**	0.978	0.934	-1.031	0.975	**-0.086**	0.969	0.927	4.255	0.898
48	**-0.079**	0.968	0.932	-0.444	0.904	**-0.081**	0.955	0.922	-2.241	0.860
60	**-0.085**	0.970	0.930	-0.162	0.859	**-0.082**	0.963	0.931	-2.246	0.836
72	**-0.058**	0.977	0.956	-0.432	0.852	**-0.053**	0.971	0.954	-3.796	0.835
84	**-0.069**	0.987	0.955	-1.726	0.818	**-0.059**	0.984	0.962	1.848	0.880
96	**-0.088**	0.986	0.939	-0.477	0.517	**-0.073**	0.987	0.957	1.612	0.855
108	**-0.085**	0.985	0.940	-0.301	0.049	**-0.069**	0.985	0.957	2.247	0.646
120	**-0.098**	0.978	0.925	0.329	-0.052	**-0.080**	0.977	0.944	-0.474	0.525
	Panel C: ESL (1994–2000)					Panel D: ESL (2001–2020)				
	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	-0.025	0.827	0.849	0.329	-1.637	**-0.153**	0.935	0.809	0.444	0.207
6	0.016	0.809	0.818	0.327	-0.081	**-0.073**	0.893	0.852	-0.011	0.232
9	**0.050**	0.854	0.835	0.370	0.284	**-0.029**	0.892	0.885	-0.138	0.432
12	**0.070**	0.887	0.853	0.385	0.416	-0.014	0.921	0.920	-0.060	0.586
15	**0.073**	0.901	0.867	0.376	0.455	-0.020	0.956	0.952	-0.016	0.723
18	**0.064**	0.909	0.885	0.363	0.436	**-0.033**	0.983	0.973	-0.046	0.849
21	0.045	0.920	0.908	0.366	0.417	**-0.048**	0.997	0.977	-0.108	0.961
24	0.026	0.931	0.927	0.365	0.396	**-0.064**	0.998	0.965	-0.181	1.054
30	-0.000	0.942	0.942	0.313	0.338	**-0.094**	0.985	0.925	-0.315	1.081
36	-0.014	0.945	0.944	0.259	0.303	**-0.113**	0.971	0.897	-0.378	0.998
48	-0.025	0.947	0.945	0.183	0.267	**-0.133**	0.951	0.864	-0.250	0.933
60	-0.038	0.946	0.938	0.268	0.374	**-0.140**	0.956	0.860	-0.210	0.927
72	-0.039	0.959	0.957	0.164	0.312	**-0.119**	0.975	0.894	-0.209	0.998
84	-0.023	0.967	0.967	0.246	0.392	**-0.089**	0.990	0.940	-0.329	1.083
96	0.002	0.972	0.974	0.302	0.448	**-0.054**	0.997	0.982	-0.281	1.194
108	0.019	0.985	0.983	0.314	0.437	-0.014	1.012	1.014	-0.317	1.325
120	0.041	0.989	0.983	0.394	0.461	**0.044**	1.002	0.996	-0.082	0.867
	Panel E: ESLSC (1994–2000)					Panel F: ESLSC (2001–2020)				
	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$	Mean	SD*	RMSPE*	$\widehat{\rho}(1)^*$	$\widehat{\rho}(12)^*$
3	-0.011	0.845	0.873	0.332	-6.309	**-0.142**	0.906	0.807	0.392	0.157
6	0.029	0.817	0.820	0.328	-0.121	**-0.059**	0.853	0.832	-0.009	0.171
9	**0.063**	0.857	0.826	0.373	0.371	-0.013	0.853	0.853	-0.121	0.348
12	**0.082**	0.890	0.842	0.393	0.508	0.004	0.886	0.886	-0.054	0.508
15	**0.086**	0.904	0.858	0.387	0.535	-0.001	0.924	0.924	-0.014	0.644
18	**0.076**	0.909	0.876	0.373	0.510	-0.013	0.954	0.952	-0.041	0.756
21	0.057	0.917	0.899	0.375	0.490	-0.028	0.971	0.964	-0.098	0.844
24	0.037	0.925	0.918	0.373	0.466	**-0.044**	0.975	0.960	-0.166	0.912
30	0.010	0.934	0.933	0.317	0.396	**-0.074**	0.969	0.934	-0.299	0.936
36	-0.004	0.936	0.936	0.263	0.346	**-0.093**	0.962	0.913	-0.373	0.902
48	-0.016	0.940	0.940	0.185	0.290	**-0.116**	0.950	0.883	-0.256	0.883
60	-0.031	0.944	0.938	0.274	0.390	**-0.125**	0.956	0.879	-0.216	0.875
72	-0.033	0.958	0.959	0.168	0.317	**-0.106**	0.976	0.910	-0.217	0.954
84	-0.017	0.965	0.966	0.250	0.404	**-0.077**	0.993	0.955	-0.354	1.059
96	0.007	0.968	0.970	0.304	0.469	**-0.044**	1.002	0.994	-0.310	1.197
108	0.023	0.979	0.976	0.312	0.460	-0.005	1.014	1.018	-0.346	1.218
120	0.045	0.983	0.975	0.392	0.483	**0.053**	1.002	0.990	-0.085	0.750

Notes: Bold numbers indicate a prediction error significantly different from zero at the 5% level of significance.

2.4 Conclusions

In conclusion, I propose an autocorrelation-based factor model for modeling and forecasting the yield curve based on the factors that drive serial dependence of the yield curve in the framework of functional principal component analysis.

Similar to other dynamic functional factor models for the yield curve, see e.g., Bowsher and Meeks (2008), Jungbacker *et al.* (2014), Almeida *et al.* (2018), and Otto and Salish (2019), my approach is based on the cross-sectionally continuous and smooth underlying yield curve subject to unobservable errors (e.g. measurement errors). The novel part in my approach is that the factor loadings are related to the autocovariance functions of the yield curve and thus the factors determine to what extent these autocovariance functions drive the yield curve at a given time. The dynamic evolution is driven by a vector autoregression for these autocorrelation-based factors. Consequently, the autocorrelated idiosyncratic components present in the model errors are absorbed into these factors, leading to less autocorrelated yield curve residuals with zero mean.

Like the largely favored specification of three factors in the bulk of term structure models, dating back to Litterman and Scheinkman (1991), three factors are also identified by my data-driven method, which simultaneously determines the number of factors and the order of the lagged information used from the yield curves. Specifically, applying my autocorrelation-based factor model to monthly constant-maturity zero-coupon Treasury yield curve data of Liu and Wu (2021) at 17 maturities from 3 to 120 months over the period of January 1985 up to December 2020, I find that the choice of 3 factors and 1-month lag information is robust over time.

My autocorrelation-based model has favorable in-sample properties. First, the yield curve residuals from this new model exhibit zero mean and less autocorrelation. Second, it improves the in-sample fit. Compared with the DNS model imposing parametric assumptions on the shape of factor loadings, my autocorrelation-based model free of functional form of factor loadings offers more flexibility in describing the shape of yield curves and reduces both RMSE and MAE by more than 20%. When the serially correlated idiosyncratic components are considered as factors, besides the common factors, the average reductions in RMSE and MAE are both more than 10%.

Also, my autocorrelation-based model is the most preferred model from an out-of-sample forecasting perspective. Among 7 parametric, semiparametric, and nonparametric approaches, the prediction errors from my model have less autocorrelation and less non-zero mean at different maturities, and the forecasts of my model deliver smaller root mean squared prediction errors. Moreover, my autocorrelation-based model's superior out-of-sample 1-month-ahead forecasting performance relative to the other 5 three-factor models is stable over time.

The much simpler way of dealing with serial dependence of functional data and the naturally

resulted functional form to account for cross-sectional dependence provide future research with many potential directions. For instance, one can formulate a test for making inference for the number of factors and the true factor space of the yield curve considering both cross-sectional dependence and time series dependence of the yield curve.[12] In addition, based on the factor structure accommodating both cross-sectional dependence and time series dependence, one can examine which factors can be designated as "unspanned" and whether these factors are same as in Joslin *et al.* (2014). One can explore the economic significance of this new autocorrelation-based factor model in, such as fixed-income products portfolio management, financial derivatives pricing, and credit portfolio risk management.

2.5 Appendix I

Let $\{Y_t(u)\}_{t\in\mathbb{Z}^+}$ be the observed yield curve time series in a Hilbert space of square integrable functions defined on the domain $u \in [a, b]$, equipped with inner product $\langle x, y\rangle = \int_a^b = x(u)\, y(u)\, \mathrm{d}u$ for all $x, y \in \mathcal{H}$ and the norm is given by $\|x\|^2 = \langle x, x\rangle$. The sequence of yield curve can be expressed as

$$Y_t(u) = X_t(u) + \varepsilon_t(u), \qquad u \in [a, b], \tag{2.17}$$

where $\varepsilon_t(\cdot)$ is assumed to be the noise term, in the sense that (i) $\mathbb{E}\{\varepsilon_t(u)\} = 0$ for all t and all $u \in [a, b]$; (ii) $\mathrm{Cov}\{\varepsilon_t(u), \varepsilon_{t+k}(v)\} = 0$ for all $u, v \in [a, b]$ provided that $k \neq 0$, (iii) $\mathrm{Cov}\{X_t(u), \varepsilon_{t+k}(u)\} = 0$ for all $u, v \in [a, b]$ and $k \neq 0$. Assume that

$$\int_a^b \mathbb{E}\{X_t(u)^2 + \varepsilon_t(u)^2\}\, \mathrm{d}u < \infty, \tag{2.18}$$

with $\mu(u) \equiv \mathbb{E}\{X_t(u)\}$, and the autocovariance kernel $c_k(u, v) \equiv \mathrm{Cov}\{X_t(u), X_{t+k}(v)\}$, and $\mathrm{Cov}\{X_t(u), \varepsilon_{t+k}(u)\} = 0$ for all integer k, $k \neq 0$. Based on the Karhunen–Loève theorem and given Eq. (2.18), the centered stochastic process of particular interest $\{X_t(u) - \mu(u)\}$ can be expressed as

$$X_t(u) - \mu(u) = \sum_{j=1}^{\infty} \xi_{tj}\phi_j(u), \tag{2.19}$$

where $\xi_{tj} = \int_a^b \{X_t(u) - \mu(u)\}\phi_j(u)\mathrm{d}u$ are pairwise uncorrelated scalar random variables with $\mathbb{E}(\xi_{tj}) = 0$, $\mathrm{Var}(\xi_{tj}) = \lambda_j$, ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, and $\mathrm{Cov}(\xi_{ti}, \xi_{tj}) = 0$ for $i \neq j$, $\phi_j(u)$ are continuous real-valued functions on the domain $[a, b]$ that are pairwise orthogonal in Hilbert space.

[12] Crump and Gospodinov (2021) exclusively focuses on the role of cross-sectional dependence for PCA.

For $c_X = c_0 = \mathrm{Cov}\{X_t(u), X_t(v)\}$, the corresponding spectral decomposition is in the form of

$$c_X(u,v) = \sum_{j=1}^{\infty} \lambda_j \phi_j(u)\phi_j(v), \quad u, v \in [a,b], \tag{2.20}$$

where λ_j is the eigenvalue corresponding to the jth orthonormal eigenfunction $\phi_j(u)$. Then, we have

$$\int_a^b c_X(u,v)\,\phi_j(v)\,\mathrm{d}v = \lambda_j \phi_j(u), \quad j \in \mathbb{Z}^+, \quad u \in [a,b]. \tag{2.21}$$

If the centered yield curve is d-dimensional, i.e. $\lambda_d \geq 0$ and $\lambda_{d+1} = 0$ for a positive and finite integer $d \geq 1$, Eq. (2.20) can be written as

$$c_X(u,v) = \sum_{j=1}^{d} \lambda_j \phi_j(u)\phi_j(v), \tag{2.22}$$

and Eq. (2.19) can be expressed by

$$X_t(u) - \mu(u) = \sum_{j=1}^{d} \xi_{tj}\phi_j(u), \tag{2.23}$$

Substituting Eq. (2.23) into Eq. (2.17), then we have

$$Y_t(u) = \mu(u) + \sum_{j=1}^{d} \xi_{tj}\phi_j(u) + \varepsilon_t(u). \tag{2.24}$$

Now consider the proof for $\sum_{j=1}^{d} \xi_{tj}\phi_j(u) = \sum_{j=1}^{d} \eta_{tj}\psi_j(u)$. Eq. (2.20) can be rewritten as

$$c_X(u,v) = \sum_{j=1}^{d} \lambda_j \phi_j(u)\phi_j(v) = \sum_{i,j=1}^{d} \sigma_{ij}^k \phi_i(u)\phi_j(v), \tag{2.25}$$

where σ_{ij}^k is the (t,s)th element in the autocovariance matrix of $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{td})$ at lag k, $\mathbb{E}(\boldsymbol{\xi}_t \boldsymbol{\xi}'_{t+k})$, say $\boldsymbol{\Sigma}_k$. Then a non-negative operator

$$\int_a^b c_k(u,z)\,c_k(v,z)\,\mathrm{d}z = \sum_{i,j=1}^{d} \omega_{ij}^k \phi_i(u)\phi_j(v), \tag{2.26}$$

where ω_{ij}^k is the (t,s)th element in the non-negative definite matrix $\boldsymbol{\Sigma}_k \boldsymbol{\Sigma}'_k$. For any $\zeta(\cdot)$ in the orthogonal complement of $\mathcal{M}$ on the domain $[a,b]$ in a Hilbert space $\mathcal{L}_2$, it holds for any lag k

that,

$$\int_a^b \int_a^b c_k(u,z)\, c_k(v,z)\, \zeta(v)\, \mathrm{d}z\, \mathrm{d}v = 0, \quad u \in [a,b], \quad \zeta(\cdot) \in \mathcal{M}^{\perp}, \tag{2.27}$$

and

$$\sum_{k=1}^{p} \int_a^b \int_a^b c_k(u,z)\, c_k(v,z)\, \zeta(v)\, \mathrm{d}z\, \mathrm{d}v = 0, \quad k = 1, \ldots, p, \quad u \in [a,b], \quad \zeta(\cdot) \in \mathcal{M}^{\perp}, \tag{2.28}$$

where $\sum_{k=1}^{p} \int_a^b c_k(u,z)\, c_k(v,z)\, \mathrm{d}z$, $k = 1, \ldots, p$, is the non-negative operator $K(u,v)$ in Eq. (2.3), holds if and only if Eq. (2.27) holds for all $1 \leq k \leq p$.

Proposition 1. *For $p \geq k_0$, $k_0 \in \mathbb{Z}^+$, if the the matrix $\mathbf{\Sigma}_{k_0}$ is full rank, the non-negative operator K_{k_0} has exactly d non-zero eigenvalues, and $\mathcal{M}$ is the linear space spanned by the corresponding d eigenfunctions.*

Note that $\mathrm{rank}(\mathbf{\Sigma}_{k_0}) = d$ is based on the assumption that the true yield curve process of particular interest $\{X_t(u)\}$ is d-dimensional. If $\mathrm{rank}(\mathbf{\Sigma}_{k_0}) < d$ for all k, the component with no autocorrelations in $\{X_t(u)\}$ should be absorbed into the error terms $\varepsilon(u)$. The autocovariance operator c_k is not non-negative definite, and thus it is not necessarily true that $\int_a^b \sum_{k=1}^{p} c_k(u,v) f(v) \neq 0$ for all $f(v) \in \mathcal{M}$. The operator K is used to gather information at different lags together and to avoid spurious selection of $\widehat{d}$.

Thus, Eq. (2.23) can be written as

$$X_t(u) - \mu(u) = \sum_{j=1}^{d} \xi_{tj} \phi_j(u) = \sum_{j=1}^{d} \eta_{tj} \psi_j(u), \tag{2.29}$$

where the functional principal component scores $\boldsymbol{\eta}_t$ and the eigenfunctions ψ_j can be estimated by the observed yield curves, while ξ_{tj} and ϕ_j cannot.

Chapter 3

A Generalized CAPM with Asymmetric Power Distributed Errors with an Application to Portfolio Construction

3.1 Introduction

Although the Gaussian (normal) distribution is widely used in capital asset pricing and portfolio selection (see e.g., Markowitz, 1952, 1959; Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black, 1972; Campbell *et al.*, 1997), there is ample empirical evidence that returns are usually not normally distributed (see, e.g., Fama, 1965a; Mandelbrot, 1967; Blattberg and Gonedes, 1974; Affleck-Graves and McDonald, 1989). According to Campbell *et al.* (1997), specifically taking asymmetry and fat-tails of financial data into account is relevant for asset pricing.

To cope with fat tails and skewness of financial data, Komunjer (2007) constructs the Asymmetric Power Distribution (APD), extending the Generalized Power Distribution (GPD) by accommodating asymmetry. Likewise, the Asymmetric Exponential Power Distribution (AEPD) proposed by Zhu and Zinde-Walsh (2009) includes two tail parameters to describe the decay of the tail densities. Table 3.1 indicates how a number of well-known distributions are nested as special cases of the APD and the AEPD.

Despite the vast empirical evidence on non-normality of returns (see Sharpe, 2007b), most existing literature on empirical asset pricing adheres to the assumption of normally distributed (log) returns. In this chapter, we consider the seminal Capital Asset Pricing Model (CAPM) under the more general assumption that the error terms are independent and identically asymmetric power distributed (IIAPD) with zero mean and definite variance σ_ε^2, but with extra skewness

Table 3.1: Special cases of the APD and the AEPD.

Distributions	Parameters in APD	Parameters in AEPD
Normal	$\alpha = 1/2, \lambda = 2$	$\alpha = 1/2, p_1 = p_2 = 2$
GPD	$\alpha = 1/2, \lambda > 0$	$\alpha = 1/2, p_1 = p_2$
Laplace	$\alpha = 1/2, \lambda = 1$	$\alpha = 1/2, p_1 = p_2 = 1$
Asymmetric Laplace	$\alpha \neq 1, \lambda = 1$	$\alpha \neq 1/2, p_1 = p_2 = 1$
Two-piece normal	$\alpha \neq 1/2, \lambda = 2$	$\alpha \neq 1/2, p_1 = p_2 = 2$
SEPD		$\alpha \neq 1/2, p_1 = p_2$

The GPD allows flexibility in modeling the tail behavior. According to Komunjer (2007), the GPD corresponds to the uniform distribution when $\lambda = \infty$, short-tailed distributions when $2 \leq \lambda < \infty$, and fat-tailed ones when $0 < \lambda < 2$. According to Zhu and Zinde-Walsh (2009), the APD is a sort of SEPD (Skew Exponential Power Distribution) due to the quantification of asymmetry; however, the AEPD corresponds with the SEPD if $p_1 = p_2$.

parameter α and tail parameter(s) to accommodate the asymmetry and fat tails of the return distribution. Our work provides a generalization of CAPM because the previously assumed normal return distributions can be considered as special cases of IIAPD. Zeckhauser and Thompson (1970) consider the estimation of linear models with power distributions as highly desirable and expect the effects on the estimated coefficients to be substantial. Thus, considering IIAPD errors potentially is empirically highly relevant to researchers and practitioners.

To examine whether two tail parameters (p_1 and p_2 in the AEPD capture the decay of the left and right tail, respectively) can better describe the distributions of asset returns than one (λ in the APD determines the decay of both tails), we develop two versions of the generalized CAPM with IIAPD errors, namely, the CAPM-IAPD (independent and identically asymmetric power distributed with one tail parameter) and the CAPM-IAEPD (independent and identically asymmetric exponential power distributed with two tail parameters). Using these generalized models as well as the regular CAPM, we address the following research questions:

(i) To what extent does the specification of the error term affect whether the CAPM is 'alive' (working) or 'dead' (rejected)?

(ii) Does the generalized CAPM with non-normal errors outperform the regular CAPM with normal errors in terms of goodness-of-fit measures?

(iii) Does the generalized CAPM with two tail shape parameters (CAPM-IAEPD) have better fit and predictive power than the CAMP with only one tail shape parameter (CAPM-IAPD)?

(iv) What does the assumption of IIAPD errors imply for the practitioners in the fields of se-

curity valuation and portfolio management? Do the portfolios constructed based on these generalized CAPM outperform those based on the regular one?

Our results show that the specification with IIAPD errors has nice properties. First, it helps to 'save' the CAPM in the sense that the CAPM with IIAPD errors is substantially less likely to be rejected than the regular one with normally distributed errors when using monthly returns data. Second, the generalized CAPM with IIAPD errors outperforms the regular CAPM in terms of goodness-of-fit measures in particular with weekly and daily data. Meanwhile, incorporating two parameters for the tail shape does not seem to further improve model's fit or predictive power of the model relative to the model with only one tail parameter.

As indicated in research question (*iv*), we also investigate the economic significance of this generalized CAPM in portfolio management. Do the portfolios constructed based on this generalized model outperform those based on the regular CAPM? Our results suggest the answer is 'yes' ; the portfolios constructed using the IIAPD errors outperform the portfolio employing normal errors, according to the Sharpe Ratios, Jensen's Alphas, and Treynor Ratios.

By assuming the possibility of lending and borrowing at the risk-free rate of interest, the Sharpe-Lintner version of the CAPM is given by

$$\mathbb{E}\left(R_i\right) = R_f + \beta_{iM}\left(\mathbb{E}\left(R_M\right) - R_f\right), \tag{3.1}$$

where $\mathbb{E}[R_i]$ is the expected return of asset i, R_f is the return on the risk-free asset, and $\mathbb{E}[R_M]$ is the expected return on the market portfolio. The compact form of Eq (3.1), in terms of excess returns (i.e. returns in excess of the risk-free rate), is expressed by

$$\mathbb{E}\left(Z_i\right) = \beta_{iM}\mathbb{E}\left(Z_M\right), \tag{3.2}$$

where Z_i represents the return on the ith asset in excess of the risk-free rate, $Z_i \equiv R_i - R_f$, Z_M represents the excess return on the market portfolio of assets, $Z_M \equiv R_M - R_f$, and $\beta_{iM} = \frac{\mathrm{Cov}(Z_i, Z_M)}{\mathrm{Var}(Z_M)}$. The data-generating process (DGP) underlying the CAPM is assumed to be given by

$$Z_{it} = \alpha_{iM} + \beta_{iM} Z_{Mt} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathrm{NID}(0, \sigma^2_{\varepsilon_i}), \tag{3.3}$$

where i denotes the asset, t denotes the time period ($t = 1, \ldots, T$), Z_{it} and Z_{Mt} are the excess returns in time period t for asset i and the market portfolio, respectively, $Z_{it} \equiv R_{it} - R_{ft}$, $Z_{Mt} \equiv R_{Mt} - R_{ft}$, α_{iM} represents the asset excess return intercept, β_{iM} is the proportionality factor that reflects the sensitivity of asset i relative to the market risk, and ε_{it} is a normally, independent

distributed (NID) random variable with mean 0 and variance $\sigma^2_{\varepsilon_i}$.

To examine the effects of the assumption of IIAPD error terms on the estimated coefficient, we consider research question (i). As in several previous papers (see e.g., Fama and French, 1996), here 'alive' implies that α_{iM} is statistically insignificant and β_{iM} is statistically significant; thus, if at least one of them does not hold, the CAPM is called 'dead'. Therefore, research question (i) will be answered by the 95% confidence intervals of α_{iM} and β_{iM}. In addition, the Akaike Information Criterion (AIC) and Bayesian information criterion (BIC), density plots of the residuals, and distribution tests are used to address research questions (ii) and (iii). Finally, an applicable portfolio strategy, the Minimum Variance Portfolio (MVP), will be presented in Section 3.6 with backtesting based on historical data. The MVP is applied because its composition is less affected the estimation of expected returns, which are difficult to estimate, and the optimization of MVP depends on the estimates of β_{iM}. Moreover, it performs well in practice and is popular among investors.

The empirical analysis is conducted based on the European market. The EURO STOXX 50 index is used as a proxy of the market portfolio. Maximum Likelihood is used to analyze its constituents. We consider daily, weekly and monthly returns in the analysis. After that, a back test on the monthly basis is used to investigate portfolio construction strategies.

To our knowledge, there is very little previous work in which IIAPD errors are applied to model asset returns conditional on market portfolios. Li and Lin (2014) checked the validity of CAPM for the French stock market based on a model called CAPM-AEPD, however, they used the location parameter μ and the scale parameter σ of the AEPD as the proxies of the mean and standard deviation of the distribution. Although μ and σ indeed correspond to the mean and standard deviation in the special case of a normal distribution, they are not suitable proxies for the mean and standard deviation of the APD or AEPD.

This chapter is also related to a broad class of studies on the application of other non-Gaussian distributions in finance. The class of symmetric stable Paretian distributions has, for instance, been used to model fat-tailed return distributions (see e.g., Mandelbrot, 1963a, 1963b; Fama, 1965b, 1971). In addition, portfolio theory under stable Paretian laws has been developed (see Rachev and Mittnik, 2000). However, Cootner (1964) argues that the evidence of Paretian distributed security returns is too casual. Moreover, the probability density functions (PDFs) of the stable Paretian distributions cannot be expressed analytically, and their higher order moments are undefined. Campbell *et al.* (1997) deem that the stable Paretian distributions are too fat-tailed. Moreover, Thomas and Gup (2010) argue that the empirical patterns of stock returns are inconsistent with the stable Paretian theory of constant 'alpha peakedness' and 'beta skewness' for returns over different

time intervals. As a result, the stable Paretian does not appear to be the ideal distribution to model the fat tails of financial returns.

A number of alternative distributions have been proposed in the literature. For the sake of flexibility in modeling fat tails, the seminal Exponential Power Distribution (EPD, also called the Generalized Laplace distribution, the Generalized Power Distribution, or the Generalized Error Distribution) family proposed by Subbotin (1923) has also been applied in finance (see, e.g., Harvey, 1981; Nelson, 1991). However, the EPD does not allow for asymmetry in financial returns data. The Skew Normal (SN) distribution and the Skew Exponential Power distribution (SEPD) proposed by Azzalini (1985, 1986) can fit data with skew distributions. Despite this attractive property, the skewness parameter may be estimated to be infinite. Moreover, consistency of the maximum likelihood estimators (MLEs) has not been established for the SN distribution and the SEPD. In contrast, Fernández *et al.* (1995), Theodossiou (2000) and Komunjer (2007) extend the EPD family in the sense of accommodating asymmetry, categorized by Zhu and Zinde-Walsh (2009) as the second method of constructing SEPD. For the second method of constructing SEPD family, Zhu and Zinde-Walsh (2009) prove consistency of MLEs when the tail shape parameter is larger than one. Besides, Komunjer (2007) also proves consistency of MLEs in the context of the APD. Since the APD can provide consistent parameter estimates and Komunjer's (2007) method to skew EPD is conceptually simpler and readily applicable to the nested densities, the APD as a sort of SEPD is used for modeling excess returns in this chapter. Furthermore, Zhu and Galbraith (2010) apply a generalized asymmetric Student-t distribution to asset pricing and Zhang *et al.* (2011) apply non-Gaussian distributions to study market volatility.

Studies on CAPM with other non-normal generating errors, e.g. errors following a Cauchy distribution, Student t-distribution and a mixture of these distributions and normal distribution are already done by Owen and Rabinovitch (1983) and Hodgson *et al.* (2002). These papers show some difference between the results based on these distributions and the normal distribution, but did not reach a clear conclusion on how the generalized distributions improve the regular CAPM. Indeed, Hodgson *et al.* (2002) even show that models assuming a non-Gaussian distribution can be less consistent. This chapter shows that the asymmetric power distribution improves the "survival rate" and the predictive power of CAMP, which is different from these papers. Besides, this chapter connects the properties of return distribution with the classic CAPM. By comparing the CAPM with the NID errors and a more generalized IIAPD errors, we investigate the effects of the NID and IIAPD error term assumptions on the estimated coefficients and then we backtest our new models in the sense of the economic significance of these effects. This in our view is novel and contributes to the literature.

This chapter is organized as follows: The models are presented in Section 3.2. Section 3.3 shows the log-likelihood functions of models. Section 3.4 describes the data. Section 3.5 presents the results and the goodness of fit of the models. Section 3.6 examines the implications of the different error distributions to portfolio constructions. Finally, Section 3.7 concludes.

3.2 The CAPM-IAPD and the CAPM-IAEPD

Under the mean-variance framework of modern portfolio theory, $\mathbb{E}(\varepsilon_{it}) = 0$ and $\mathbb{E}(\varepsilon_{it}^2) = \sigma_{\varepsilon_i}^2$. A random variable ε with mean zero and variance σ_ε^2 can be obtained by standardizing X, as

$$\varepsilon = \sigma_\varepsilon \frac{X - \omega}{\delta}, \tag{3.4}$$

where σ_ε is the standard deviation of ε and X is a random variable with mean ω and variance δ^2.

Assume that X has an arbitrary standard asymmetric power distribution. Specifically, let X be a standard APD random variable in the context of the CAPM-IAPD and a standard AEPD random variable in the context of CAPM-IAEPD. Then, X can be expressed by a function of the error term, that is,

$$\begin{aligned} X &= \omega + \frac{\delta}{\sigma_\varepsilon}\varepsilon \\ &= \omega + \frac{\delta}{\sigma_\varepsilon}\left(Z - \alpha_M - \beta_M Z_M\right), \end{aligned} \tag{3.5}$$

where Z is the random variable of excess returns and Z_M a non-stochastic predetermined term. As the density of the return variables is of interest, let $f_X(x)$ represent the density of a given asymmetric power distribution, then the PDF of the return variable Z is given by

$$\begin{aligned} f_Z(z) &= f_X(x)\left|\frac{\mathrm{d}z}{\mathrm{d}x}\right|^{-1} \\ &= f_X\left(\omega + \delta \cdot \frac{Z - \alpha_M - \beta_M Z_M}{\sigma_\varepsilon}\right) \cdot \frac{\delta}{\sigma_\varepsilon}. \end{aligned} \tag{3.6}$$

The data-generating process (DGP) of the CAPM-IAPD reads

$$\begin{aligned} Z_{it} &= \alpha_{iM} + \beta_{iM} Z_{Mt} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathrm{IAPD}(\alpha, \lambda, 0, \sigma_{\varepsilon_i}^2), \\ \varepsilon_{it} &= \sigma_{\varepsilon_i} \frac{X - \omega}{\delta}, \\ \mathbb{E}(\varepsilon_{it}) &= 0, \\ \mathbb{E}(\varepsilon_{it}^2) &= \sigma_{\varepsilon_i}^2, \end{aligned} \tag{3.7}$$

where ε_{it} denotes the IAPD error term with mean 0, variance $\sigma^2_{\varepsilon_i}$, that can be expressed by the standard APD random variable X. According to Komunjer (2007), the mean and variance of X are given by

$$\begin{aligned} \omega &= \mathbb{E}(X) = \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}(1-2a)\delta_{\alpha,\lambda}^{-1/\lambda}, \\ \delta^2 &= \mathrm{Var}(X) = \frac{\Gamma(3/\lambda)\Gamma(1/\lambda)(1-3\alpha+3\alpha^2)-\Gamma(2/\lambda)^2(1-2\alpha)^2}{\Gamma(1/\lambda)^2}\delta_{\alpha,\lambda}^{-2/\lambda}, \end{aligned} \tag{3.8}$$

respectively, where α is the skewness parameter, $\alpha \in (0,1)$ measures the degree of asymmetry, $\lambda > 0$ is the tail parameter which controls the tail decay, $\delta_{\alpha,\lambda} \equiv \frac{2\alpha^\lambda(1-\alpha)^\lambda}{\alpha^\lambda+(1-\alpha)^\lambda}$ with $\delta_{\alpha,\lambda} \in (0,1)$, and $\Gamma(\cdot)$ is the Gamma function.

According to Komunjer (2007), the density of X is given by

$$f_X(x) = \begin{cases} \left(\dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right)\exp\left(-\dfrac{\delta_{\alpha,\lambda}}{\alpha^\lambda}|x|^\lambda\right), & \text{for } x \le 0 \\ \left(\dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right)\exp\left(-\dfrac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda}|x|^\lambda\right), & \text{for } x > 0. \end{cases} \tag{3.9}$$

The DGP of the CAPM-IAEPD is

$$\begin{aligned} Z_{it} &= \alpha_{iM} + \beta_{iM}Z_{Mt} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathrm{IAEPD}(\alpha, p_1, p_2, 0, \sigma^2_{\varepsilon_i}) \\ \varepsilon_{it} &= \sigma_{\varepsilon_i}\frac{Y-\omega}{\delta}, \\ \mathrm{E}(\varepsilon_{it}) &= 0, \\ \mathrm{E}(\varepsilon^2_{it}) &= \sigma^2_{\varepsilon_i}, \end{aligned} \tag{3.10}$$

where ε_{it} denotes the IAEPD error term with mean 0 and variance variance $\sigma^2_{\varepsilon_i}$, can be expressed by the standard AEPD random variable Y. According to Zhu and Zinde-Walsh (2009), the mean and the variance of Y are given by

$$\begin{aligned} \omega &= \mathbb{E}(Y) = \frac{1}{B}\left[(1-\alpha)^2\frac{p_2\Gamma(2/p_2)}{\Gamma^2(1/p_2)} - \alpha^2\frac{p_1\Gamma(2/p_1)}{\Gamma^2(1/p_1)}\right] \\ \delta^2 &= \mathrm{Var}(Y) \\ &= \frac{1}{B^2}\left\{(1-\alpha)^3\frac{p_2^2\Gamma(3/p_2)}{\Gamma^3(1/p_2)} + \alpha^3\frac{p_1^2\Gamma(3/p_1)}{\Gamma^3(1/p_1)} - \left[(1-\alpha)^2\frac{p_2\Gamma(2/p_2)}{\Gamma^2(1/p_2)} - \alpha^2\frac{p_1\Gamma(2/p_1)}{\Gamma^2(1/p_1)}\right]\right\}, \end{aligned} \tag{3.11}$$

respectively, where α is the skewness parameter, $\alpha \in (0,1)$, $p_1 > 0$ is the left tail parameter and $p_2 > 0$ the right tail parameter. $B \equiv \alpha K_{EP}(p_1) + (1-\alpha)K_{EP}(p_2) = (\frac{\alpha}{\alpha^*})K_{EP}(p_1) = (\frac{1-\alpha}{1-\alpha^*})K_{EP}(p_2)$, where $\alpha^* = \frac{\alpha K_{EP}(P_1)}{\alpha K_{EP}(p_1)+(1-\alpha)K_{EP}(p_2)}$ to ensure continuity and if $p_1 = p_2$, $\alpha^* = \alpha$.

$K_{EP}(p) \equiv 1/\left[2p^{1/p}\Gamma(1+1/p)\right]$, where $\Gamma(\cdot)$ is the Gamma function. Based on Zhu and Zinde-Walsh (2009), the PDF of a standard AEPD random variable Y can be defined as

$$f_Y(y) = \begin{cases} \left(\frac{\alpha}{\alpha^*}\right) K_{EP}(p_1) \exp\left(-\frac{1}{p_1}\left|\frac{y}{2\alpha^*}\right|^{p_1}\right), & \text{for } y \leq 0 \\ \left(\frac{1-\alpha}{1-\alpha^*}\right) K_{EP}(p_2) \exp\left(-\frac{1}{p_2}\left|\frac{y}{2(1-\alpha^*)}\right|^{p_2}\right), & \text{for } y > 0. \end{cases} \tag{3.12}$$

According to Zhu and Zinde-Walsh (2009), the APD and the AEPD are skewed to the right if $\alpha < 1/2$ and to the left if $\alpha > 1/2$, and p_1 controls the left tail while p_2 controls the right tail. Specifically, $p_1 < 2$ leads to a fat left tail, which becomes fatter for smaller p_1 (the same applies to p_2 regarding the right tail).

3.3 Maximum Likelihood Estimation

The method of Maximum Likelihood (ML) is used and the BFGS algorithm is employed. The log-likelihood for the usual CAPM with NID errors takes the form

$$L_T(\theta) = \sum_{t-1}^{T} \log\left(\frac{1}{\sigma_{\varepsilon_i}\sqrt{2\pi}} \exp\left(-\frac{(Z_{it} - \alpha_{iM} - \beta_{iM} Z_{Mt})^2}{2\sigma_{\varepsilon_i}^2}\right)\right), \tag{3.13}$$

where 'log' denotes the natural logarithm, the parameter vector $\theta = (\alpha_{iM}, \beta_{iM}, \sigma_{\varepsilon_i})^T$ and $\alpha_{iM} \in \mathbb{R}$, $\beta_{iM} \in \mathbb{R}$, and $\sigma_{\varepsilon_i} > 0$. To maximize the log–likelihood with respect to the parameters in θ, $\hat{\theta}$ is the solution to the problem $\max_{\theta\in\Theta} L_T(\theta)$ where Θ is a given parameter set, $\Theta \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^*$.

In terms of the CAPM-IAPD, log-likelihood function, $L_T(\theta)$ reads as

$$\begin{aligned} L_T(\theta) \;=\; & T \log\left(\frac{\delta}{\sigma_{\varepsilon_i}}\left[\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)}\right]\right) \\ & + \sum_{t=1}^{T}\left[-\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda}\left|\omega + \frac{\delta(Z_{it} - \alpha_{iM} - \beta_{iM} Z_{Mt})}{\sigma_{\varepsilon_i}}\right|^\lambda 1\left(\omega + \frac{\delta(Z_{it} - \alpha_{iM} - \beta_{iM} Z_{Mt})}{\sigma_{\varepsilon_i}} \leq 0\right)\right. \\ & \left. - \frac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda}\left|\omega + \frac{\delta(Z_{it} - \alpha_{iM} - \beta_{iM} Z_{Mt})}{\sigma_{\varepsilon_i}}\right|^\lambda 1\left(\omega + \frac{\delta(Z_{it} - \alpha_{iM} - \beta_{iM} Z_{Mt})}{\sigma_{\varepsilon_i}} > 0\right)\right], \end{aligned} \tag{3.14}$$

where the notation is consistent with that of Eqs (3.7)–(3.9). Since $\alpha_{iM} \in \mathbb{R}$, $\beta_{iM} \in \mathbb{R}$, $0 < \alpha < 1$, $\lambda > 0$ and $\sigma_{\varepsilon_i} > 0$, and $\theta = (\alpha_{iM}, \beta_{iM}, \alpha, \lambda, \sigma_{\varepsilon_i})^T$ is the parameter vector, $\hat{\theta}$ will be the solution to the problem $\max_{\theta\in\Theta} L_T(\theta)$ where Θ is a parameter set, $\Theta \subset \mathbb{R} \times \mathbb{R} \times (0,1) \times \mathbb{R}_+^* \times \mathbb{R}_+^*$.

Similarly, the log-likelihood function $L_T(\theta)$ of the CAPM-IAEPD is given by

$$\begin{aligned} L_T(\theta) &= T\log\left(\frac{\delta}{\sigma_{\varepsilon_i}}B\right) \\ &+\sum_{t=1}^{T}\left[-\frac{1}{p_1}\left|\frac{\omega+\frac{\delta(Z_{it}-\alpha_{iM}-\beta_{iM}Z_{Mt})}{\sigma_{\varepsilon_i}}}{2\alpha^*}\right|^{p_1}1\left(\omega+\frac{\delta(Z_{it}-\alpha_{iM}-\beta_{iM}Z_{Mt})}{\sigma_{\varepsilon_i}}\leq 0\right)\right. \\ &\left.-\frac{1}{p_2}\left|\frac{\omega+\frac{\delta(Z_{it}-\alpha_{iM}-\beta_{iM}Z_{Mt})}{\sigma_{\varepsilon_i}}}{2(1-\alpha^*)}\right|^{p_2}1\left(\omega+\frac{\delta(Z_{it}-\alpha_{iM}-\beta_{iM}Z_{Mt})}{\sigma_{\varepsilon_i}}>0\right)\right], \end{aligned} \tag{3.15}$$

where the notation is consistent with that of Eqs (3.10)–(3.12), the parameter vector is given by $\theta = (\alpha_{iM}, \beta_{iM}, \alpha, p_1, p_2, \sigma_{\varepsilon_i})^T$ and $\alpha_{iM} \in \mathbb{R}$, $\beta_{iM} \in \mathbb{R}$, $0 < \alpha < 1$, $p_1 > 0$, $p_2 > 0$ and $\sigma_{\varepsilon_i} > 0$. Therefore, $\hat{\theta}$ will be the solution to the problem $\max_{\theta\in\Theta} L_T(\theta)$ where Θ is a parameter set, $\Theta \subset \mathbb{R}\times\mathbb{R}\times(0,1)\times\mathbb{R}_+^*\times\mathbb{R}_+^*\times\mathbb{R}_+^*$.

3.4 Data description

In studies using US stock returns a broad-based stock index, such as the S&P 500 or the Center for Research in Security Prices (CRSP) market index, is usually taken as the proxy of the market portfolio. In this chapter, the EURO STOXX 50 Index is chosen as the proxy of the market portfolio of the European Market. In fact, the EURO STOXX 50 Index is highly correlated with the S&P 500, and the correlation was pegged at 0.865.[1] In addition, the EURO STOXX 50 Index is viewed as Europe's leading Blue-chip index for the Eurozone with currency consistency. Furthermore, the constituents of this index are supersector leaders in the Eurozone with representativeness and diversification, and the time interval effect influences big firms less.[2]

As recommended by the European Money Markets Institute (EMMI), we use the Euro Interbank Offered Rate (Euribor) as proxies of the risk-free rate for data at lower frequencies to take into account the effects of tax, regulatory, and liquidity conditions for the European market. Besides, the Euro OverNight Index Average (Eonia) is used as the risk-free rate for the realized daily returns based on the argument by Hull (2011) that investors gradually perceive the overnight index swap rate as the risk-free rate for high frequency financial data after the recent financial

[1] John W. Labuszewski, "Inter-Market Stock Index Spreads." CME GROUP (August 2013): https://www.cmegroup.com/trading/equity-index/files/intermarket-stock-index-spreads.pdf.

[2] The index covers 50 stocks from Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain, which are diversified in 19 different industries.

crisis.[3] Concerning the rate of return for assets that pay dividends regularly, we use the total return index from Datastream (Thomson Reuters) to account for capital appreciation and income. As usual we define the (log) return R_t on an asset at time t as

$$R_t \equiv \log \mathrm{TR}_t - \log \mathrm{TR}_{t-1}, \tag{3.16}$$

where TR_t is the total return index value at time t and 'log' denotes the natural logarithm. The excess return of any asset Z_t,

$$Z_t \equiv 100 \times (R_t - R_{f,t-1}), \tag{3.17}$$

where R_t is as defined above and $R_{f,t-1}$ denotes the risk-free rate expressed at time $t-1$. In the empirical analysis we will use daily, weekly as well as monthly data. Therefore, a period refers to a trading day, a week or a month depending on the frequency of data used.

Table 3.2: Discriptive statistics of log returns on EURO STOXX 50 and on its 50 constituents

	Negatively skewed	Positively skewed	SES	Leptokurtic	SEK	J-B test
Panel A: Daily returns						
Whole ($T = 4003$)	33%	39%	0.0387	100%	0.0774	0%
Pre-crisis ($T = 1765$)	39%	27%	0.0583	100%	0.1166	0%
After crisis ($T = 1831$)	27%	29%	0.0572	96%	0.1145	0%
Panel B: Weekly returns						
Whole ($T = 800$)	75%	2%	0.0866	100%	0.1732	0%
Pre-crisis ($T = 353$)	61%	6%	0.1304	100%	0.2607	2%
After crisis ($T = 366$)	18%	10%	0.1280	86%	0.2561	14%
Panel C: Monthly returns						
Whole ($T = 183$)	76%	2%	0.1811	86%	0.3621	6%
Pre-crisis ($T = 80$)	57%	0%	0.2739	69%	0.5477	22%
After crisis ($T = 85$)	6%	6%	0.2657	22%	0.5314	73%

Notes: SES denotes the standard error of skewness, computed as $\sqrt{6/T}$, SEK denotes the standard error of kurtosis, computed as $\sqrt{24/T}$, leptokurtic represents that the return distribution with kurtosis > 3 is statistically significant, negatively skewed represents that the distribution is skewed left at the 5% significance level, and positively skewed represents that the distribution is skewed right at the 5% significance level. The J-B test represents the null hypothesis of Jarque-Bera test is not rejected at the 5% significance level. Detailed descriptive statistics are available from the authors upon request.

To track the total return index of the EURO STOXX 50 Index as long as possible, we considered the largest possible sample period available to us, which is from January 2001 to May 2016. The constituents of EURO STOXX 50 Index are listed in Section 3.8. Since the 2007-2009 financial

[3]Nevertheless, a well-discussed convention based on a comprehensive dimension for estimating the risk-free rate seems to be a hole in the theoretical literature.

crisis may have introduced a large structural break in the specification of models, we also divide our sample into two sub-samples: one from the beginning date (2001) of the dataset to October 2007 and the other starting in May 2009 when the market more or less stabilized up to the last observation in our data sample. Therefore, we can focus on the specification of the error terms. Table 3.2 summarizes the descriptive statistics of the sample data in the sense of asymmetry and fat tails. Most of the test statistics for the skewness are statistically significantly different from 0, except for weekly and monthly returns after the crisis. In addition, the overwhelming majority of values of the kurtosis are statistically significantly larger than 3, except for monthly returns after the crisis. Therefore, Table 3.2 shows evidence that most of the return series do not follow a normal distribution, in particular at higher data frequencies.

3.5 Empirical results and goodness-of-fit

This section reports the empirical results. Goodness-of-fit is measured by the Akaike Information Criterion (AIC) and Bayesian information criterion (BIC), density plots of the residuals and the estimated parametric distribution, and corresponding distributional tests.

3.5.1 Estimates of coefficients and confidence intervals

Since minus the inverse of the Hessian is a way to estimate the covariance matrix of the ML estimator, evaluated at the vector of ML estimates (see Davidson and MacKinnon, 2004), the standard error (SE) for the jth estimate can be calculated as

$$SE_j = \sqrt{-(e_j^T H^{-1} e_j)}, \tag{3.18}$$

where e_j is a vector and consists of zeros and a single one on the jth position and H denotes the Hessian at $\hat{\theta}$. Due to the asymptotic normality and consistency of the MLE, see Zhu and Zinde-Walsh (2009) for the MLE properties of the AEPD and Komunjer (2007) for the MLE properties of the APD, an empirical 95% confidence interval (CI) of parameter j is given by

$$CI_j = (e_j^T \hat{\theta} e_j - 1.96 \cdot SE_j, e_j^T \hat{\theta} e_j + 1.96 \cdot SE_j), \tag{3.19}$$

where $\hat{\theta}$ denotes the solution to the problem $\max_{\theta \in \Theta} L_T(\theta)$.[4]

Research question (i) relates to the effects of the assumption of IIAPD error terms on the

[4]The command 'fmincon' is used for the optimization procedure.

Table 3.3: The effects of the assumption of IIAPD error terms on the estimated coefficients for monthly returns

The number of 'alive'	Whole sample period	Pre-crisis	After the crisis
CAPM	6	15	8
CAPM-IAPD	22	25	30
CAPM-IAEPD	19	19	18

Notes: Here 'alive' means that α_{iM} is statistically insignificant and β_{iM} is statistically significant; thus, if at least one of them does not hold, the CAPM is called 'dead'.

estimated coefficients. Following the commonly-used definition, we claim that the CAPM is 'dead' or rejected for a stock unless the estimation generates both a significant $\hat{\beta}_{iM}$ and an insignificant $\hat{\alpha}_{iM}$. Table 3.3 shows that employing the assumption of IIAPD error terms has a large impact on the estimated coefficients. Specifically, it makes rejection of the CAPM substantially less frequently for the monthly data. Under the assumption of IIAPD error terms, the CAPM is 'alive' for 22 stocks in CAPM-IAPD and 19 for CAPM-IAEPD during the whole sample period, which helps 'save' the regular CAPM with only 6 stocks. Likewise, the number increases from 15 with the normally distributed error terms to 25 (and 19) with the IIAPD error terms during the pre-crisis period, and from 8 to 30 (and 18) during the period after the crisis.

3.5.2 Goodness-of-fit

Standard information criteria such as AIC and BIC are widely used to select models (see e.g., Chatfield, 1996; Peña and Rodriguez, 2005; Spanos, 2010). Therefore, we use the AIC and BIC to address research questions (*ii*) and (*iii*). As defined by Akaike (1974),

$$AIC = 2k - 2L_T(\hat{\theta}), \tag{3.20}$$

and according to Schwarz (1978), the BIC is given by

$$BIC = k\log(n) - 2L_T(\hat{\theta}), \tag{3.21}$$

where $L_T(\hat{\theta})$ is the value of the log-likelihood function, evaluated at $\hat{\theta} = \text{argmax}_{\theta\in\Theta} L_T(\theta)$ and k is the number of components of $\hat{\theta}$. The smaller the AIC or BIC value, the better the goodness-of-fit.

Table 3.4 presents the in-sample-fit results. Concerning research question (*ii*), the evidence from the AIC and BIC values indicates that the generalized CAPM with IIAPD errors performs better than the CAPM with normally distributed errors, especially for the higher frequency data. In particular, the AIC and BIC values from the CAPM are, on average, larger than those for

Table 3.4: BIC and AIC values of all models

	Mean of BIC	Mean of AIC	Selected by BIC	Selected by AIC
	Whole sample period daily returns: T=4003			
CAPM	14227.55	14208.69	0	0
CAPM-IAPD	13217.72	13186.24	47	36
CAPM-IAEPD	13223.56	13185.79	3	14
	Whole sample period weekly returns: T=800			
CAPM	4053.37	4039.34	0	0
CAPM-IAPD	3938.87	3915.45	43	30
CAPM-IAEPD	3942.64	3914.53	7	20
	Whole sample period monthly returns: T=183			
CAPM	1186.06	1176.45	27	12
CAPM-IAPD	1178.91	1162.86	19	22
CAPM-IAEPD	1182.01	1162.75	4	16
	Pre-crisis daily returns: T=1765			
CAPM	6154.59	6138.23	0	0
CAPM-IAPD	5741.08	5713.70	47	33
CAPM-IAEPD	5745.85	5712.99	3	17
	Pre-crisis weekly returns: T=353			
CAPM	1759.11	1747.58	5	0
CAPM-IAPD	1717.57	1698.23	37	29
CAPM-IAEPD	1720.49	1697.29	8	21
	Pre-crisis monthly returns: T=80			
CAPM	516.82	509.75	31	17
CAPM-IAPD	514.37	502.46	16	19
CAPM-IAEPD	517.14	502.84	3	14
	After crisis daily returns: T=1831			
CAPM	5786.40	5769.86	0	0
CAPM-IAPD	5553.09	5525.53	47	36
CAPM-IAEPD	5558.45	5527.66	3	14
	After crisis weekly returns: T=366			
CAPM	1731.84	1720.13	11	2
CAPM-IAPD	1717.87	1698.36	35	27
CAPM-IAEPD	1721.45	1698.03	4	21
	After crisis monthly returns: T=85			
CAPM	531.10	523.77	44	28
CAPM-IAPD	535.33	523.12	5	13
CAPM-IAEPD	538.23	523.57	1	9

the generalized CAPMs (the CAPM-IAPD and the CAPM-IAEPD) for daily and weekly returns. Even for monthly data, the average value of the AIC for the CAPM with normally distributed errors is largest regardless of sample period. Moreover, the CAPM-IAPD, for most of the cases, is selected by the AIC for monthly returns during the whole sample period and before the crisis.

Concerning research question (*iii*), the inclusion of two tail parameters does not seem to add much explanatory power to the model. On the contrary, the CAPM-IAPD has a smaller AIC and BIC than the CAPM-IAEPD regardless of data frequency and sample period. Therefore, the AIC and BIC values do not lend support to the statement that CAPM-IAEPD performs better than CAPM-IAPD.

Secondly, to visualize the actual and a given theoretical distribution, the probability density function (PDF) of the actual random variable can be estimated by kernel density estimation (Silverman, 1986). The Gaussian Kernel is used for the density estimation and the bandwidth h, $h = 1.06 \times \text{Std}(\hat{u}_t) \times T^{-1/5}$, where Std denotes the standard deviation and T denotes the number of time periods. The black solid line represents the empirical distribution and the red dotted line represents the given distribution. Compared with the normal distribution, the APD and the AEPD display a very close fit. Without loss of generality, we present some typical and representative density plots here.[5] Figure 3.1 suggests that the generalized CAPM under the IIAPD errors assumption is more realistic than the CAPM under the normal errors assumption. Similarly, there is no clear evidence showing that the CAPM-IAEPD is better than the CAPM-IAPD.

Thirdly, a distribution test is performed to investigate research questions (*ii*) and (*iii*). The empirical 95% confidence interval (see Section 5.1) plays an important role in determining the statistical significance as well as the distribution test. The reason is that the AEPD and the APD can be reduced to the normal distribution, provided that $\alpha = 0.5$, $p_1 = p_2 = 2$ and $\alpha = 0.5$, $\lambda = 2$ respectively. As a result, if 0.5 lies outside the CI of $\hat{\alpha}$ (i.e. $\alpha \neq 0.5$ at the 95% significant level) or 2 lies outside the CI of $\hat{\lambda}$ (for the CAPM-IAPD) or of $\hat{p}_1$ or $\hat{p}_2$ (for the CAPM-IAEPD), the null hypothesis of a normal distribution can be rejected at the 5% significance level and thus the result is statistically significant. Besides, the CIs can test some non-normal distributions nested in these two distributions, e.g. the Laplace distribution. This is because these two can be reduced to the Laplace distribution if $\alpha = 0.5$, $p_1 = p_2 = 1$ and if $\alpha = 0.5$, $\lambda = 1$. Table 3.5 summarizes the distribution test results.

Table 3.5 shows evidence that the assumption of IIAPD errors is more suitable based on the high significance of the skewness and tail parameters. First, it shows that the generalized CAPM can capture the asymmetry of return distributions due to the statistically significant skewness

[5]The others are available from the author upon request.

Table 3.5: Distribution test results of the error terms

	CAPM-IAPD Normal distribution $H_0: \alpha = 0.5, \lambda = 2$	CAPM-IAEPD Normal distribution $H_0: \alpha = 0.5, p_1 = p_2 = 2$	CAPM-IAPD Laplace distribution $H_0: \alpha = 0.5, \lambda = 1$	CAPM-IAEPD Laplace distribution $H_0: \alpha = 0.5, p_1 = p_2 = 1$
Whole sample period: daily				
Significant αs	32	45	32	45
Significant λs	50		32	
Significant p_1s		50		45
Significant p_2s		50		48
H_0 is rejected	50	50	46	50
Whole sample period: Weekly				
Significant αs	37	45	37	45
Significant λs	50		30	
Significant p_1s		50		45
Significant p_2s		49		45
H_0 is rejected	50	50	46	49
Whole sample period: Monthly				
Significant αs	24	38	24	38
Significant λs	34		26	
Significant p_1s		37		37
Significant p_2s		32		32
H_0 is rejected	36	42	40	43
Pre-crisis: daily				
Significant αs	39	47	39	47
Significant λs	50		27	
Significant p_1s		50		48
Significant p_2s		50		46
H_0 is rejected	50	50	47	50
Pre-crisis: weekly				
Significant αs	35	47	35	47
Significant λs	50		28	
Significant p_1s		46		44
Significant p_2s		47		39
H_0 is rejected	50	49	47	49
Pre-crisis: monthly				
Significant αs	25	42	25	42
Significant λs	24		16	
Significant p_1s		39		35
Significant p_2s		34		35
H_0 is rejected	28	45	28	43
After crisis: daily				
Significant αs	35	49	35	49
Significant λs	50		33	
Significant p_1s		50		46
Significant p_2s		49		44
H_0 is rejected	50	50	48	49
After crisis: weekly				
Significant αs	16	41	16	41
Significant λs	47		36	
Significant p_1s		43		40
Significant p_2s		42		38
H_0 is rejected	47	47	46	47
After crisis: monthly				
Significant αs	17	34	17	34
Significant λs	15		18	
Significant p_1s		33		30
Significant p_2s		26		28
H_0 is rejected	20	37	25	35

Notes: The skewness parameter α accounts for the asymmetry, and 'significant αs' stands for how many αs out of 50 are statistically significant, i.e., 0.5 lies out of $CI_{\hat{\alpha}}$. Zhu and Zinde-Walsh (2009) indicate that the parameter α determines to the right if $\alpha < 0.5$ and to the left if $\alpha > 0.5$, and $p < 2$ leads to a fat tail, fatter for smaller p; therefore, if 2 lies outside $CI_{\hat{\lambda}}$, $CI_{\hat{p}_1}$, or $CI_{\hat{p}_2}$, the tail parameter can capture the fat tail of the return distributions.

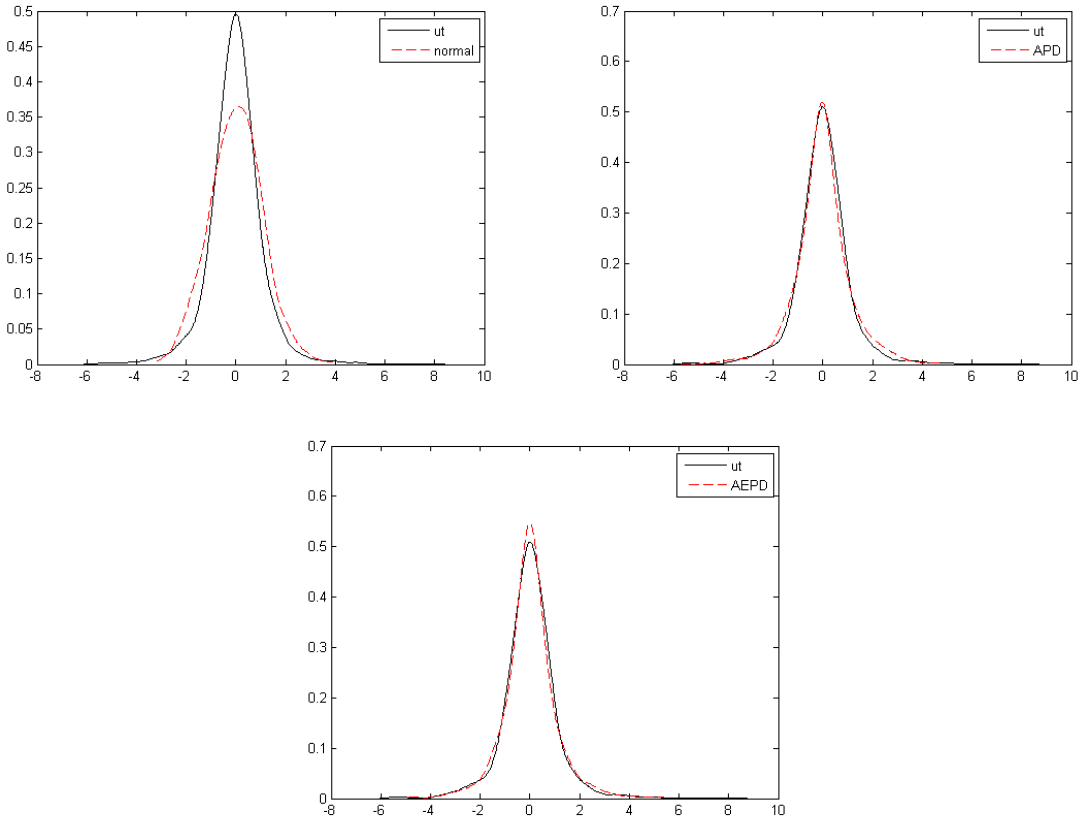


Figure 3.1: The PDF plots of actual residuals (solid line) and the given distribution (dashed line).

parameter. Table 3.5 also demonstrates that the vast majority of tail parameters are statistically significant based on the normal distribution test for daily and weekly returns. The result above is consistent with the result of descriptive statistics in Section 3.4. Therefore, we conclude that the generalized CAPM with IIAPD errors outperforms the regular CAPM in the sense of accommodating asymmetry and fat-tails.

Table 3.5 also shows that the CAPM-IAEPD with two tail parameters is more sensitive to asymmetry regardless of data frequency and sample period. As argued by Zhu and Zinde-Walsh (2009), the APD is a sort of SEPD and the skewness parameter in the SEPD is not rich enough to capture all of the asymmetry of the return distributions, especially the asymmetry in the tails, because the SEPD's left tail is always thinner than its right tail. However, there is no distinct difference between the CAPM-IAPD and the CAPM-IAEPD because the corresponding numbers of significant results are comparable in terms of weekly and daily returns.

3.6 Application to portfolio strategy and a backtest

The analysis above indicates that the IIAPD errors assumption outperforms the NID assumption in the CAPM estimation. This section studies whether the specification of the error distribution also generates different levels of profitability in portfolio construction. The minimum variance portfolio (MVP), the lowest point on the efficient frontier, is applied to construct a portfolio strategy for the following reasons: (1) The MVP is very popular among investors, because it performs well and is easy to apply. (2) Although MVPs are typically heavily concentrated in the lowest volatility stocks (e.g., DeMiguel *et al.*, 2009), low volatility stocks come with higher returns (see e.g., Ang *et al.*, 2006; Bali and Cakici, 2008). (3) The only required inputs in the optimization of the MVPs are correlations and volatilities, i.e. only depending on risk (see e.g., Amenc and Martellini, 2002); therefore, the systematic risk, quantified by β_{iM} can be utilized. (4) It is well known that estimating the expected returns is difficult, but the composition of the MVP is not affected by expected returns.

3.6.1 Portfolio strategy

From an active portfolio management point of view, allocating money into the constituents of a broad market index is a smart approach to reduce the likelihood of excessive trading and the liquidity issue of offsetting positions. As mentioned above, the constituents of the EURO STOXX 50 index are Europe's leading Blue-chip stocks and supersector leaders and the time interval effect influences big firms less. Assume that investors allocate all money into the European stock market and wish to minimize risk, namely, invest in the MVP. The portfolio construction, i.e. the asset allocation problem, involves the maximization of a quadratic function of the decision variables, subject to a set of linear constraints, some of which are inequalities (see, e.g., Markowitz, 1956; Sharpe, 1963; Sharpe, 1970; Sharpe, 1971; Elton *et al.*, 1978). However, according to Sharpe (1978), the approaches developed by Sharpe (1963, 1970, 1971) and by Elton *et al.* (1978) for a simplified model hardly accommodate a different model of covariance relationships. Therefore, the algorithm developed by Sharpe (1978) is employed to construct an MVP, exploiting the simplicity of the constraints but allowing as much complexity as desired in the covariance matrix. According to Sharpe (2007b), in terms of the MVP, an investor who wishes to minimize risk will have a risk tolerance of zero, no matter how much expected return is sacrificed in the process. Thus, the asset allocation problem of the MVP in terms of relative risk aversion can be written as

$$u = x^T e - 0.5\,\mathrm{rra}\left(x^T C x\right), \qquad \text{subject to}: \quad \mathrm{sum}(x) = 1, lb \leq x \leq ub, \tag{3.22}$$

where 'rra' denotes relative risk aversion, x is a $(n \times 1)$ vector of decision variables, the weights of holdings to represent a portfolio, e denotes a $(n \times 1)$ vector of expected returns (estimated by the mean of the historical returns), lb denotes a $(n \times 1)$ vector of lower bounds, and ub denotes a $(n \times 1)$ vector of upper bounds (all elements of x are non-negative if short sales are not allowed, $lb = 0$, and $ub = \iota$), C denotes the $(n \times n)$ covariance matrix defined as

$$C = \begin{cases} \sigma_{ij} = \sigma_i^2, & \text{if } i = j \\ \sigma_{ij} = \beta_{iM}\beta_{jM}\sigma_M^2, & \text{if } i \neq j, \end{cases} \tag{3.23}$$

where σ_i^2 denotes the variance of security i's return, σ_M^2 denotes the variance of excess return on the market portfolio. β_{iM} and β_{jM} are yielded by the CAPMs. Based on this security selection process, this portfolio strategy can, to some extent, realize the optimization according to the optimally selected securities.

3.6.2 Backtest results

Since the maximum likelihood method performs well for large sample sizes according to asymptotic theory, from the sample size point of view, more observations may be preferable. Therefore, we track the historical data of the EURO STOXX 50 index and its constituents as long as possible. There are 49 stocks that can be included into the portfolio, and the list of constituents is in Section 3.8. The investment horizon is 7 years, from 29 May 2009, when the market became more or less stabilized, to 29 April 2016. From the perspective of investment policy statement, some assumptions are made for convenience. Ideally the investment constraints exclude the investor's taxation level and liquidity needs. Moreover, there are no transaction costs and investor's net cash inflow of the portfolio is 0 (i.e. there are no funding payments and no withdrawals). Due to restrictions on short sales for most investors, short sales are not allowed here. Rebalancing is done on a monthly basis to reduce the likelihood of excessive trading. The data used for estimating β_{iM} and β_{jM} are disposed as described in Section 3.4. Assume that the market value of the portfolio at the beginning is the amount of money that the investor wishes to allocate into the MVP and there is no cash position at the beginning. The initial value V_t of the portfolio, $V_t = TR_t \times x_{t-1}$ where TR_t is the $(1 \times n)$ vector of the total return index value at period t, x_{t-1} is the vector of weights determined at period $t-1$, based on the historical data until period $t-1$.

Table 3.6 reports the backtesting results for performance evaluation. The methodology of average monthly total return, the Sharpe Ratio and corresponding annualized calculations are in line with Morningstar (2005). The Standard Deviation (a measure of risk) is estimated by

Table 3.6: Backtest report of performance analysis on the investment based on three portfolio strategies

Investment horizon: 29/05/2009 to 29/04/2016 Benchmark: EURO STOXX 50 Index	Strategy A: CAPM approach	Strategy B: CAPM-IAPD approach	Strategy C: CAPM-IAEPD approach
Return/Risk Analysis			
Total return (realized)	117.55%	125.36%	124.09%
Average monthly total return	1.42%	1.51%	1.50%
Annualized total return	18.38%	19.71%	19.49%
Standard deviation	3.65%	3.60%	3.91%
Annualized standard deviation	12.63%	12.48%	13.54%
Annualized Sharpe Ratio	0.9993	1.1088	1.0073
Best month	11.28%	11.82%	11.60%
Worst month	-9.11%	-8.89%	-9.39%
% of Up month	66.27%	66.27%	68.67%
% of Down month	33.73%	33.73%	31.33%
Avg. Monthly Gain	3.33%	3.35%	3.44%
Avg. Monthly Loss	-2.35%	-2.10%	-2.77%
Relative Performance			
Annualized excess return (relative to the benchmark portfolio)	9.72%	11.05%	10.83%
Jensen's Alpha	0.88%***	0.99%***	0.95%***
	(3.0016)	(3.2619)	(3.0788)
Beta	0.52***	0.48***	0.56***
	(8.5489)	(7.6301)	(8.9113)
Tracking Error (on monthly basis)	3.60%	3.82%	3.57%
Annualized Information Ratio	0.78	0.84	0.88
R-Squared	0.47	0.42	0.50
Treynor Ratio	0.35	0.41	0.34

Notes: A label '***' indicates significance at the 1% level. T-values are given in parentheses. The risk-free rate is the 1 month Euribor rate.

the sample standard deviation. The Sharpe Ratio, the ratio of the difference between expected return and the riskless rate to standard deviation, is another risk-adjusted performance measure. Normally, the greater the Sharpe Ratio, the better the performance (a given number greater than 1 is categorized as good). The Information Ratio is a measure of the risk-adjusted return. For quants, the annualized Information Ratio is calculated by (annualized excess return)/($\sqrt{12}\times$monthly σ_{te}), where the excess return is relative to the benchmark portfolio's annualized return, and σ_{te} denotes tracking error, a measure of how closely a portfolio follows the index to which it is benchmarked. The manager's skill is considered to be good if the Information Ratio is larger than 0.5, implying that an investment historically delivered 0.5% in excess returns for every unit of risk above the benchmark. As discussed by Jensen (1968), Jensen's Alpha is the intercept term, α_j, in $Z_{jt} = \alpha_j + \beta_j Z_{Mt} + u_{jt}$, where Z_{jt} denotes the excess return on an arbitrary actual portfolio j, which is another risk-adjusted measure to evaluate the abnormal return of a portfolio, and the higher the better. Finally, the Treynor Ratio T, $T = Z_j/\beta_j$, where Z_j represents the annualized portfolio return in excess of the risk-free rate, is a measure of the return per unit of systematic risk, the higher the better as well.

In general, the frequently-used performance measures (the Sharpe Ratio, Jensen's Alpha, and the Treynor Ratio) indicate that the portfolios constructed using the IIAPD errors (Strategy B) outperform that using normally distributed errors (Strategy A).[6] From the perspective of portfolio return/risk, Strategies B and C are superior to Strategy A due to higher annualized Sharpe Ratios, cause a Sharpe Ratio that is greater than 1 implies good risk–adjusted performance. Therefore, utilizing the IIAPD errors assumption for portfolio construction can lead to 'good' risk-adjusted performance. In addition, the positive alpha of Strategy B (0.99%) or C (0.95%) is statistically significant. Note that Strategy B, with an annualized Sharpe Ratio of 1.1088, also outperforms Strategy C. Regarding the EURO STOXX 50 as the benchmark portfolio, the portfolio in the case of Strategy B is also first-rank in terms of relative performance. It generates the highest excess return (11.05% relative to the EURO STOXX 50), statistically largest significant Jensen's Alpha (0.99%), and the highest Treynor Ratio (0.41).[7]

[6]We also performed other symmetrically distributed error terms like Student t-distributed error terms for different investment horizons, and strategy B still has superiority.

[7]As a robustness check we did backtesting on the investment when the investment horizon is 5 years, to take into account that possible different results might be caused by different investment horizons; meanwhile, we also tried other symmetric distributions like Student t-distribution using different investment horizons, and we had very similar findings.

3.7 Conclusions

The modeling of non-normally distributed returns attracts more and more attention from academia and policy makers after the recent financial crisis (Thomas and Gup, 2010). This chapter proposes modeling asymmetric and fat-tailed return distributions using a family of non-Gaussian distributions. By assuming IIAPD errors, the CAPM is generalized to two new specifications: CAPM-IAPD and CAPM-IAEPD.

The application of the generalized CAPM to the European stock market suggests that the assumption of IIAPD errors has very nice properties. First, it makes the rejection of CAPM much less likely than under the assumption of normally distributed errors for monthly returns. Second, the generalized CAPMs outperform the regular CAPM with normally distributed errors in terms of AIC and BIC values, density plots through the Kernel Density Estimation, and related distribution tests, in particular for higher frequency data. Third, the generalized CAPMs under the IIAPD errors assumption are more realistic in the sense of accommodating the asymmetry and fat tailed return distributions. Their density plots display a very close match between the residuals and the distributions considered.

Meanwhile, there is little evidence that CAPM-IAEPD outperforms the CAPM-IAPD model in terms of the Akaike Information Criterion and Bayesian information criterion, though AEPD is supposed to be able to capture a larger fraction of the asymmetry of the distributions, especially the asymmetry in the tails (Zhu and Zinde-Walsh, 2009). The advantage of CAPM-IAPED may be more pronounced with larger samples, e.g. stocks in the S&P500 index or in the FTSE Global Equity Index Series (GEIS), however, this investigation is beyond the scope of this chapter.

As stated by Zeckhauser and Thompson (1970), the incorporation of IIAPD errors causes a substantial change in the estimated coefficients of the regression. We propose a portfolio strategy to apply this notion. The backtest results show that the portfolios constructed by using the IIAPD errors can outperform the portfolio employing normal errors in the sense of commonly used performance measures: the Sharpe Ratio, Jensen's Alpha, and the Treynor Ratio. This result confirms that the IIAPD error does not only improve the modeling of asset prices, but may also help to construct more profitable trading strategies for the industry.

There are several directions for future work. Maximum likelihood estimation performs well when the sample size is large. Hence, from the sample size point of view, higher frequency returns are preferable within the same time frame. This leaves open the question what effects the frequency has on the portfolio returns. For finite samples, the bootstrap can be used to assess the statistical stability of these outcomes. Thus, some bootstrapping methods can be used to investigate the significance of coefficients. For instance, the Bias-corrected and accelerated (BC_a)

method developed by DiCiccio and Efron (1996) and the corresponding BC_a confidence intervals. Besides, these more generalized distributional assumptions can be integrated in more complex asset pricing models for e.g. financial derivatives and in volatility modeling and forecasting.

3.8 Appendix II

EURO STOXX 50 index constituents

Table 3.7: EURO STOXX 50 index constituents

AIR LIQUIDE	AIRBUS GROUP	ALLIANZ (XET)
ANHEUSER-BUSCH INBEV	ASML HOLDING	ASSICURAZIONI GENERALI
AXA	BANCO SANTANDER	BASF (XET)
BAYER (XET)	BBV.ARGENTARIA	BMW (XET)
BNP PARIBAS	CARREFOUR	DAIMLER (XET)
DANONE	DEUTSCHE BANK (XET)	DEUTSCHE POST (XET)
DEUTSCHE TELEKOM (XET)	E ON (XET)	ENEL
ENGIE	ENI	ESSILOR INTL.
FRESENIUS (XET)	IBERDROLA	INDITEX
ING GROEP GDR	INTESA SANPAOLO	L'OREAL
LVMH	MUENCHENER RUCK. (XET)	NOKIA
ORANGE	PHILIPS ELTN.KONINKLIJKE	SAFRAN
SAINT GOBAIN	SANOFI	SAP (XET)
SCHNEIDER ELECTRIC SE	SIEMENS (XET)	SOCIETE GENERALE
TELEFONICA	TOTAL	UNIBAIL RODAMCO
UNICREDIT	UNILEVER DR GDR	VINCI
VIVENDI	VOLKSWAGEN PREF.(XET)	

Notes: The total return index of ENGIE starts from 7 July 2005 and INDITEX starts from 22 May 2001. The total Return Index of EURO STOXX 50 index is available from Datastream from 2 January 2001. Data are retrieved on 24 May 2016.

Chapter 4

Predicting Intraday Return Patterns based on Overnight Returns for the US Stock Market

4.1 Introduction

Studies such as Wood *et al.* (1985), Smirlock and Starks (1986), Harris (1986), Jain and Joh (1988), Hong and Wang (2000) and Bogousslavsky (2016) show that average returns (mean returns, or market mean returns) vary over the day.[1] To explain this, Hong and Wang (2000) solve an equilibrium model for a competitive stock market in which investors trade for asset allocation as well as for informational reasons. The model indicates that time patterns in intraday stock returns are driven by time-varying hedging demand and time-varying information asymmetry. In addition, Hong and Wang (2000) find U-shaped and inverted U-shaped mean intraday stock return patterns during trading sessions.[2]

Here we propose a novel approach based on cumulative returns, referred to as cumulative regression (CumRe), to forecast intraday market return patterns conditional on overnight returns. By using this intraday pattern forecast, day traders can decide when to close out same-day trading positions to make profits. The literature already reports evidence of dependence between the overnight return and the subsequent intraday returns. For instance, Liu and Tse (2017) find that overnight returns significantly help predict the first half-hour market returns and the last half-hour market returns. Berkman *et al.* (2012) find a strong tendency for positive overnight returns

[1] A more precise statement is that market average returns vary during exchange trading sessions. We decompose returns into intraday and overnight returns based on exchange trading and non-trading periods for simplicity.

[2] See Subsection C.3 "Time Patterns of Returns" of Section V in Hong and Wang (2000) for details.

followed by negative intraday returns. In practice, traders often view the close-to-open gap (i.e. overnight return in this chapter) as an important marker; serving as either a support zone or resistance zone for that day's trading activity. Meanwhile, major macroeconomic announcements, such as GDP and CPI announcements, and the vast majority of earnings announcements are released either before the market opens or after the market closes (see e.g., Oldfield and Rogalski, 1980; Gao *et al.*, 2018; Bogousslavsky, 2021). As stated by Hong and Wang (2000), market prices cease to provide information to the uninformed investors when the market is closed, but continue to provide information after the market reopens. We propose predicting the subsequent intraday return patterns on day t based on close-to-open returns (i.e. overnight returns) observed earlier on the same day.

We employ both a parametric CumRe method, based on predictive regressions, and a non-parametric CumRe method using kernel regressions, to predict intraday patterns in mean returns. The predictive regression is widely used in the return predictability literature (see e.g., Ang and Bekaert, 2007; Goyal and Welch, 2008; Rapach and Zhou, 2013) considering linear dependence between the predictor(s) and the response variable. Non-parametric methods provide flexibility in accommodating nonlinear dependence between the overnight returns and intraday returns. Guidolin *et al.* (2009) find that capturing nonlinear effects may be the key to improving forecasts and that non-linear dynamics should be modelled for the US stock market. Maasoumi and Racine (2002) conclude that the evidence of dependence-based on linear models is somewhat inconclusive and indicate the presence of nonlinear dependence in returns. Tsay (2010) also shows that nonlinearities exist in high-frequency financial time series. Abhyankar *et al.* (1997) find nonlinear dependence and chaos in US stock market returns. Therefore, we compare the intraday return patterns predicted by using both the parametric CumRe and the more flexible non-parametric CumRe. We evaluate the parametric and non-parametric versions of CumRe in terms of out-of-sample forecasting performance as well as economic significance.

The main goal of this chapter is to identify upward/downward trends in intraday cumulative returns in response to overnight returns. We utilize these time patterns in dependence between the overnight returns and the subsequent intraday returns to formulate promising trading strategies for day traders. We examine the intraday return patterns within the first and the last half-hours separately. These two half-hours during each trading day are most popular among traders. According to, among others, Wood *et al.* (1985), Andersen and Bollerslev (1997), and Heston *et al.* (2010), volatilities during these two half-hours are usually higher than during the other trading periods. In addition, the trading volumes during these time intervals are, on average, larger than during other time intervals (see e.g., Gao *et al.*, 2018). Moreover, overnight returns

can significantly help predict returns within the first and last half-hour, compared with intraday returns in other time intervals.

Similar to studies such as Rogalski (1984), Jain and Joh (1988), Wood *et al.* (1985), we also use the return on the market index as a proxy of the market return. However, we use ETFs transaction-level data[3], instead of the underling non-tradable cash index. This is because the spot prices of the underling index may be distorted by non-synchronous trading (Cliff *et al.*, 2008; Yang *et al.*, 2010). We examine SPDR S&P 500 ETF (ticker symbol SPY), the largest ETF with the structure of Unit Investment Trust. Compared with other ETFs for the US market, the much larger trading volumes and higher daily turnover of SPY indicate that it is widely used as a trading vehicle for more active traders. Besides, SPDR S&P 500 ETF has the lowest and relatively most stable bid-ask spread.[4] According to Amihud and Mendelson (1987), the stochastic nature of stock returns is affected by two major factors: the arrival of new information and noise transactions, i.e. the smaller the bid-ask spread, the smaller the noise-induced variance. Therefore, we are likely to capture the impact of new information on returns when we examine SPY data.

Our results show that CumRe can identify the intraday patterns in returns documented in the previous literature (see e.g., Amihud and Mendelson, 1987; Hong and Wang, 2000; Liu and Tse, 2017). Specifically, the market overreaction at the beginning, the market under-reaction at the closing and the opposite trends during the first and the last half-hours of trading. Besides, non-parametric CumRe shows intraday return patterns in more detail and we find that the non-parametric CumRe improves the out-of-sample forecasting performance in terms of the Diebold and Mariano (1995) test statistic and the leave-one-*year*-out cross-validation.

Overall, the overnight returns help predict the market directions within the first and the last half-hours of the trading sessions. The timing strategies based on the signs of the overnight returns always generate higher annualized returns, and usually yield better Sharpe ratios. Besides, our results indicate that holding established positions for 15-minute or 20-minute after the market opens yields higher annualized returns and substantially higher Sharpe ratios, and always positive and higher utility gains than holding positions for the full last half-hour of trading time.

This chapter is organized as follows. Section 4.2 introduces methodology. Section 4.3 describes data and presents evidence of predictability. Section 4.4 compares the out-of-sample forecasting

[3]We use data from NYSE Trade and Quote (TAQ) database of Wharton Research Data Services. This database contains intraday transactions data in two databases, Consolidated Quote database for intraday quotes and Consolidated Trade database for intraday quotes, for all securities listed on the New York Stock Exchange (NYSE) and American Stock Exchange (AMEX), as well as Nasdaq National Market System (NMS) and SmallCap issues.

[4]The average bid/ask spread is one penny ($0.01) even when market volatility spikes and during financial crisis. Ross, J.E. (2018, January 17). Steady Under Pressure: How SPY Performs in Times of Market Stress. Retrieved from https://global.spdrs.com/blog/post/2018/jan/steady-under-pressure-how-spy-performs-in-times-of-market-stress.html. Last access date 03-12-2018.

performance of non-parametric CumRe and parametric CumRe. Section 4.5 presents economic significance in terms of market timing strategies and utility gains of mean-variance investors. Section 4.6 concludes.

4.2 Parametric and Non-parametric CumRe

We consider the case of discrete-time stock prices. Let t indicate the trading day. The regular trading hours in the United States are from 9:30 a.m. to 4 p.m. Eastern time. Consequently, the trading period on day t lasts 6.5 hours (390 minutes).

We employ the cumulative returns similar to Tsay (2010, Ch. 5) and explore their pattern forecasts. We consider day traders, who open and close their positions on the same day, trying to maximize their profits based on cumulative returns.

Cumulative returns are formulated by expanding a 5-minute interval starting from the same initial price. We focus on 5-minute interval, because it strikes a balance between market microstructure effects and the blurring price reactions over larger time intervals (see e.g., Andersen *et al.*, 2007). There are 78 cumulative 5-minute return observations per trading day.

Formally, for each i and j, satisfying $0 \leq i < j \leq 78$, let $p_{t,i}$ be the log price at which a security is first traded on day t, i.e. $i \times 5$ minutes after the market opens to establish a position on that security, and $p_{t,j}$ be the log price at which the security is executed to close out the position established earlier on day t.

A return pattern on day t can be defined as $\mathbf{y}_t = \left(y_t^{i,i+1}, y_t^{i,i+2}, \ldots, y_t^{i,j}\right)$, where $y_t^{i,j}$ are the cumulative returns on day t, given by $y_t^{i,j} \equiv p_{t,j} - p_{t,i}$. We focus on trading within the first half-hour and the last half-hour per day, that is, patterns $\mathbf{y}_t^f = \left(y_t^{0,1}, y_t^{0,2}, \ldots, y_t^{0,6}\right)$ and $\mathrm{y}_t^l = \left(y_t^{72,73}, y_t^{72,74}, \ldots, y_t^{72,78}\right)$, respectively.

Let τ represent the τ^{th} element in the return pattern vector $\mathbf{y}_t$. There are 6 elements in the vector of the first or the last half-hour return pattern. The τ^{th} element of the return pattern on day t, $\mathbf{y}_t$, can be expressed as $y_{t,\tau}$, where $\tau = 1, \ldots, 6$, $t = 1, \ldots, T$, T is the total number of trading days in the sample.

In the setting of cumulative regression, we assume the vector-valued time-series process $\{\mathbf{Y}_t\}$ under consideration (e.g. the first or last half-hour return pattern) to be strictly stationary. Given an overnight return $X_t = x$, the conditional mean of the τ^{th} element in $\mathbf{Y}_t$ is denoted by

$$m(\tau, x) \equiv \mathbb{E}(Y_{t,\tau}|x). \tag{4.1}$$

Parametric predictive regression is commonly used in the return predictability literature (see

e.g., Stambaugh,1999; Ang and Bekaert, 2007; Goyal and Welch, 2008; Rapach and Zhou, 2013). Gao *et al.* (2018) and Liu and Tse (2017) employ the predictive regressions to investigate the linear dependence between the predictor(s) and the response variables, namely the overnight returns and intraday returns.[5] We propose a parametric CumRe based on predictive regressions of intraday return patterns on the overnight returns. The parametric CumRe is specified as

$$y_{t,\tau} = \alpha_\tau + \beta_\tau x_t + \epsilon_{t,\tau}, \tag{4.2}$$

where α_τ is a constant term and β_τ is the coefficient of the predictor x_t, i.e. the observed overnight return on day t.

In nonparametric regression, Nadaraya-Watson estimators based on Gaussian kernels are among the most popular estimators. The corresponding nonparametric estimator of $m(\cdot)$ is given by

$$\hat{m}(\tau, x) \equiv \sum_{t=1}^{T} y_{t,\tau} w_t(x, h), \tag{4.3}$$

where $w_t(x,h) = K\left(\frac{x_t - x}{h}\right) / \sum_{t'=1}^{T} K\left(\frac{x_{t'} - x}{h}\right)$ with $K(u) = 1/\sqrt{2\pi}\exp(-u^2/2)$, and the bandwidth for the τ^{th} element in $\mathbf{y}_t$, h_τ, is to be selected from a pre-defined set of values $h_1, \cdots, h_q$ according to least-squares cross-validation (CV). Here we consider a range of bandwidths between 0.1 and 15.[6]

Also in time-series settings, CV can be used for bandwidth selection (Yao and Tong, 1998). For details regarding the corresponding mild regularity conditions for strictly stationary discrete-time time series, we refer to Yao and Tong (1998). According to Härdle *et al.* (2004), minimizing CV(h) is equivalent to minimizing the average squared error (ASE) while trading off bias and variance. For the purpose of prediction, we employ leave-one-out[7] CV to select a bandwidth which minimizes

[5]The overnight return is included in formulating the first half-hour return in Gao *et al.* (2018), resulting in a statistically linear dependence between the first half-hour returns and the last half-hour returns.

[6]The function for generating the bandwidth selection set is taken to be equidistant on a logarithmic scale, and given by $h_k = a \cdot \exp\left[(k-1)/(q-1) \cdot (\log b - \log a)\right]$, $k = 1, \ldots, q$, where a is the lower bound of the bandwidth selection set, b is the upper bound of the bandwidth selection set, q is the number of bandwidths in the selection set. We take $q = 80$ to account for the trade-off between computational intensity and precision of bandwidth selection. The upper bound here is determined by the extreme values of returns in percentage according to our data set.

[7]To take into account serial correlation in time series, we used the autocorrelation functions and the partial autocorrelation functions of elements in $\mathbf{y}_t^f$ and $\mathbf{y}_t^l$. The first order autocorrelation is not significantly different from 0 at the 90% significance level for all the elements in $\mathbf{y}_t^f$. Even if the first order autocorrelation for a couple of elements in $\mathbf{y}_t^f$ is significantly different from 0 at the 95% significance level, these autocorrelation values (less than $|0.18|$) are not large enough to make a dramatic difference in bandwidth selection. That is, given this level of autocorrelation, the bandwidths selected by cross-validation and a correlation-corrected method (CDPI) are comparable for a time series process. We refer to Opsomer *et al.* (2001) for details.

the mean squared prediction error (MSPE). The selected bandwidth for τ, h_τ is given by

$$h_\tau = \underset{h\in\{h_1,\ldots,h_q\}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T}(y_{t,\tau} - \hat{y}_{-t,\tau}(\tau, x))^2, \tag{4.4}$$

where $\hat{y}_{-t,\tau}(x, h)$ is the estimator of $y_{t,\tau}$ formed by leaving out the t^{th} observation when generating the prediction for observation t.

In the setting of conditional mean forecasts of return patterns, the bandwidth h_τ herein controls the amount of data which is effectively used. This approach gives more weight to the historical overnight return observations X_t such that $\|X_t - x\|$ is closer to 0, which is similar to some local weighted average methods such as the nearest-neighbors method in machine learning.

For each τ, the estimator of the conditional cumulative distribution function (CDF) $F_{Y_\tau|X}(y|x)$ of the cumulative returns $Y_{t,\tau}$ given an overnight return x is

$$\widehat{F}_{Y_\tau|X}(\mathbf{y}|x) = \sum_{t=1}^{T} I(Y_{t,\tau} \leq y) w_t(x, h_\tau), \tag{4.5}$$

where $I(\cdot)$ is a indicator function, which equals 1 if its argument is true, and 0 otherwise. The CDF of the non-parametric CumRe estimator will provide more detail in intraday return patterns, e.g., opposite trends in the first and the last half-hours, larger return volatility at the opening phase than at the closing, leverage effects in both the first and the last half-hours.

To avoid capturing spurious predictability due to in-sample overfitting, we examine the out-of-sample predictive performance. For the conditional mean forecasts of return patterns, we introduce leave-one-year-out cross-validation and the Diebold-Mariano test to compare the forecast accuracy of non-parametric CumRe with that of the parametric CumRe. We propose the squared prediction error (SPE) of a return pattern $\mathbf{y}_t$ during the first or the last half-hour to measure the expected squared distance between the pattern predicted by a specific value of the overnight return and the true intraday return pattern during the trading session. We define the distance between the realized return pattern and the predicted return pattern on day t by

$$\text{SPE}_t = \sum_{\tau=1}^{6} \left(\frac{\epsilon_{t,\tau}}{\sqrt{\tau}}\right)^2, \tag{4.6}$$

where $\epsilon_{t,\tau} = y_{t,\tau} - \hat{y}_{t,\tau}$ is the predition error when $\hat{y}_{t,\tau}$ is predicted by the overnight return. Herein the prediction error $\epsilon_{t,\tau}$ is standardized by the square root of τ, since each particular increment of this cumulative return pattern has a prediction error variance that is proportional to the time over which the return is realized.

We assume that returns on the market during the first and the last half-hours in one year are independent with the other years. We use leave-one-year-out cross-validation (LOYOCV). Let N denote the number of calendar years, then the data set is partitioned into N disjoint sets for N calendar years, say $I = \{I_1, \ldots, I_N\}$. Define $I_{n^-} = \cup_{m \neq n} I_m$, so that for a particular evaluation n in the cross-validation, I_n is used as test set, and the remaining sets, I_{n^-}, are used for model fitting. The mean squared prediction error (MSPE) for year n across all τs is then

$$\text{MSPE}_n(I_{n^-}) = \frac{1}{|I_n|} \sum_{\forall t \in I_n} \text{SPE}_t(I_{n^-}), \tag{4.7}$$

where $|\cdot|$ denotes the cardinality of a set and thus $|I_n|$ is the number of trading days in year n, and I_{n^-} denotes conditioning on information from all the other years for parameter estimation.The sum of squared prediction errors (SSPE) for all N available calendar years is then just SSPE $= \sum_{n=1}^{N} \text{MSPE}_n(I_{n^-})$. The smaller the SSPE, the better the predictive performance is.

We will use the SPEs to formulate a type of Diebold and Mariano (1995) test to compare the forecast accuracy of non-parametric and parametric CumRe for return pattern forecasts. The null hypothesis is that the two forecast methods have the same forecast accuracy. The test statistic and corresponding test results are provided in Section 4.4.

4.3 Data and Evidence of Predictability

We use SPDR S&P 500 ETF prices, sampled at the 5-minute frequency. The sample period ranges from January 1, 2003 to December 31, 2013. Following studies using the TAQ dataset on higher-frequency data such as Boffelli *et al.* (2016), Maneesoonthorn *et al.* (2017), and Perera and Koul (2017), the raw data has been cleaned using methods similar to those of Brownlees and Gallo (2006).[8]

In line with Oldfield and Rogalski (1980), Wood *et al.* (1985), and Liu and Tse (2017), the percentage overnight return is calculated as

$$\text{ovr}_t = 100 \times (p_{t,0} - p_{t-1,78}), \tag{4.8}$$

where ovr_t is the overnight return on day t, $p_{t,0}$ is the log price at 9:30 a.m.[9] on day t, and $p_{t-1,78}$

[8]The corresponding R package is TAQMNGR. We use a choice of $k = 60$, $\gamma = 0.02$ for the data cleaning algorithm, as mentioned by Brownlees and Gallo (2006).

[9]We also examined $\text{ovr}_t = 100 \times (p_{t,09:35} - p_{t-1,16:00})$ and $\text{ovr}_t = 100 \times (p_{t,09:45} - p_{t-1,16:00})$ as in Bogousslavsky (2021), and we obtained correlation values between overnight returns and the subsequent intraday returns similar to those found in the correlation matrix of the intraday returns, provided in Table 4.7.11 in Section 4.7

Table 4.1: Comparison of the dispersion and deviations for cumulative returns within the first and the last half-hours (Jan 2003 – Dec 2013). The entries are based on 2714 observations for each return series, which did not contain missing values. This table shows the ratios of the dispersion and deviation measures. The superscripts f and l denote the first and last half-hours, respectively.

	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$	Mean	S.E.	Ratio> 1
$\text{ABS}^f/\text{ABS}^l$	1.09	1.04	1.16	0.97	1.03	0.99	1.05*	0.02	4
$\text{IQR}^f/\text{IQR}^l$	1.37	1.34	1.26	1.27	1.24	1.24	1.29***	0.02	6
$\text{Min}^f/\text{Min}^l$	0.93	0.73	0.69	0.64	0.75	0.75	0.75***	0.04	0
$\text{Max}^f/\text{Max}^l$	1.25	1.30	1.81	1.37	1.37	1.32	1.40***	0.08	6
SD^f/SD^l	1.09	1.07	1.11	1.07	1.07	0.91	1.05*	0.03	5
$\text{Skew}^f/\text{Skew}^l$	-6.20	-2.65	-1.93	-2.62	-2.36	-1.13	-2.81***	0.72	0
$\text{Kurt}^f/\text{Kurt}^l$	0.80	1.09	1.26	1.02	1.03	0.83	1.00	0.07	4

Notes: ABS denotes range (the difference between the maximum and minimum observation). IQR is short for interquartile range. SD is standard deviation. Min and Max are the minimum (always negative) and maximum (always positive) values of the returns in the sample, respectively. Skew and Kurt are skewness and kurtotis, and all the values of skewness are different from 0 and all the values of Kurtotis are larger than 3. The cumulative returns during the first half-hour are always distributed skew to the right (positive skewness) while the corresponding returns during the last half-hour are always skew to the left (negative skewness), therefore all ratios are always negative. S.E. is short for standard error. *, ** and *** denote significance at the 1%, 5% and 10% significance level, respectively.

is the log price at 16:00 p.m. on day $t - 1$.

We conducted the Augmented Dickey-Fuller (ADF) test, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to examine the stationarity of the return series; all test results suggest that these return series are stationary.

According to studies such as Harris (1986), Amihud and Mendelson (1987) and Gerety and Mulherin (1994), the return volatility is larger during the opening than during the closing period. The results in Table 4.1 confirm that cumulative returns within the first half trading hours tend to be more volatile than those within the last half trading hours. Based on measures of dispersion such as the range of the return distributions (the difference between the maximum and minimum returns), and the interquartile range (the difference between 75th and 25th percentiles of return distribution, also called the midspread or middle 50%), we find that, on average, the range of the return distribution during the first half-hour is 5% larger than that during the last half-hour. The midspread is even 29% larger.

Despite its greater dispersion, we observe that the cumulative returns during the first half-hour have larger maximum values but smaller minimum values, than those returns during the last half-hour. This implies that it is possible for traders to realize higher profits and smaller losses during the first half-hour. Meanwhile, the ratios of skewness show different degrees of asymmetry in the return distributions. Notice that the ratios of skewness are all negative. This is because all the cumulative returns during the first half-hour appear to be skewed to the right (positive skewness), while the corresponding returns during the last half-hour appear to be skewed to the

left (negative skewness).

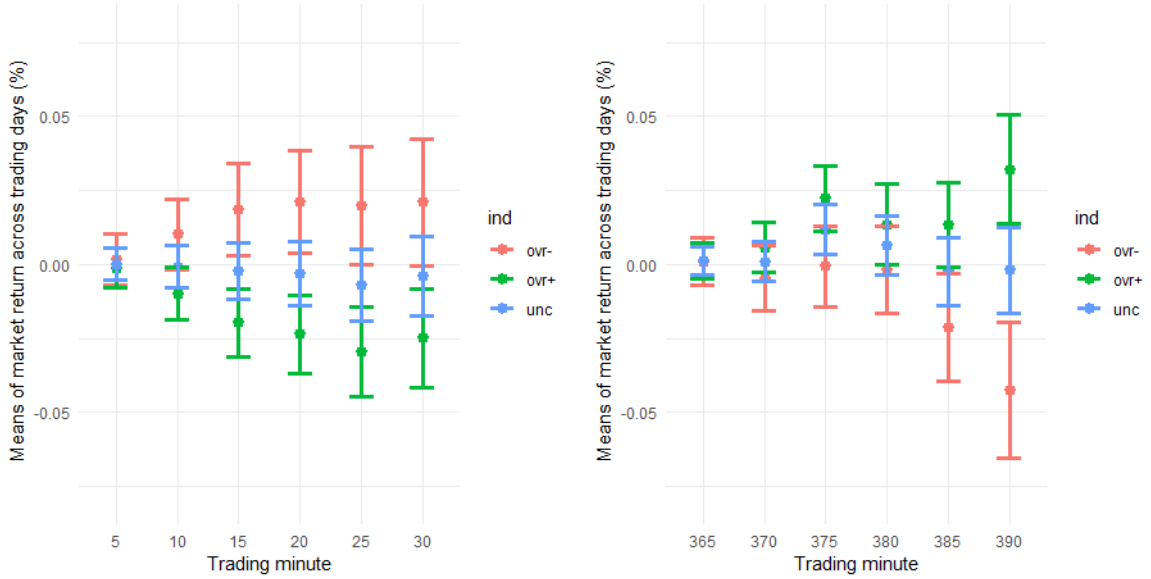


Figure 4.1: Cumulative market return (in %) within the first (left panel) and last (right panel) half trading hours conditional on positive returns, ovr+ (red dots), on negative returns, ovr− (green dots) and unconditional means, unc (blue dots). The error bars represent 95% confidence intervals.

To illustrate the behavior of the cumulative 5-minute market returns series within the first and the last half-hours, we present graphical results of the means of the market return across the trading days in Figure 4.1. The left panel shows the changing trends in the mean returns within the first half-hour. It is evident that all means conditional on negative overnight returns (the red dots), averaged at -0.46%, are positive and present a rough upward trend within the first half-hour. The mean of cumulative return peaks at around 0.02% when $\tau = 4$. After that, it declines slightly when $\tau = 5$, but goes up again to slightly above 0.02% when $\tau = 6$. In contrast, all means conditional on positive overnight returns (green dots), averaged at 0.42%, lie below 0%. It shows a downward trend until 25 minutes after the market opening, reaching a low at around -0.03% when $\tau = 5$. It is interesting to note that the means of returns when $\tau = 5$ are inflection points for both the unconditional mean returns and the conditional mean returns. The unconditional mean (in blue), present a pattern similar to means conditional on positive overnight returns. The pattern shows a slight fall, ending at about -0.07% when $\tau = 5$. Then it increases, even though the unconditional mean is still below 0.

The right-hand side panel of Figure 4.1 reflects the changing trends in the means of market returns within the last half-hour of trading time. Notice that all the unconditional and conditional

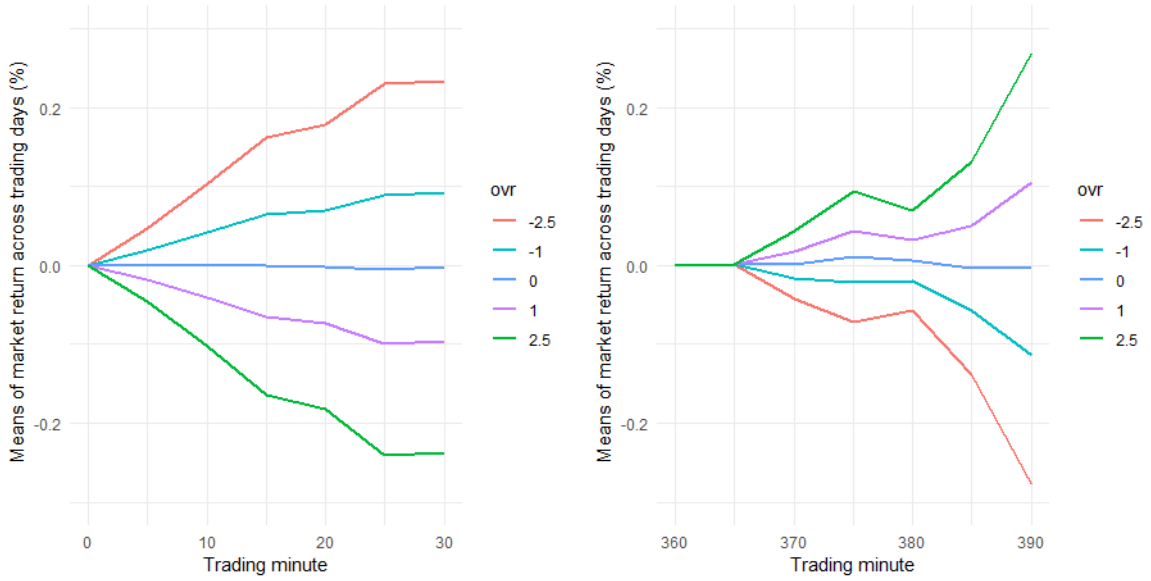


Figure 4.2: Parametric CumRe pattern forecasts depending on overnight returns within the first half trading hour (from 0 to 30-minute) and last half trading hour (from 360 to 390-minute). The symbol 'ovr' represents overnight returns. Returns are in percentages. The sample period is from January 1, 2003, through December 31, 2013.

means of the market returns are positive when $\tau = 1$ during the last half trading hour. However, the mean conditional on positive overnight returns (represented by ovr+) and that conditional on negative overnight returns (ovr−) show a completely opposite trends afterwards; ovr+ experiences an increase until $\tau = 3$ (at slightly higher than 0.02%), and then a decrease at just above 0.01% when $\tau = 5$, but jumps up to over 0.03% when $\tau = 6$. In contrast, ovr− falls to a negative number when $\tau = 2$, then goes up just below 0 when $\tau = 3$. After that, it starts to decrease and reaches a number smaller than -0.04% when $\tau = 6$ during the last half trading hour. The unconditional mean illustrates a pattern similar to ovr+. It increases until $\tau = 3$, and then decreases until $\tau = 5$ and slightly increases after, within the last half-hour.

Table 4.1 and Figure 4.1 provide evidence that the behavior of the cumulative 5-minute market mean returns within the first half-hour is different from that within the last half-hour. Besides, we find that the overnight returns are negatively (positively) correlated with the cumulative returns during the first (last) half-hour (see Table 4.7.1). Therefore, we investigate the conditional pattern forecasts depending on overnight returns during the first half-hour and the last half-hour separately.

In light of Menkveld *et al.* (2007), a short-term market overreaction is consistent with liquidity suppliers who are compensated for their services through price reversals, while market under-reaction or positive serial correlation is consistent with strategic trading by informed investors

who split their order across time to maximize profit. As observed by Amihud and Mendelson (1987), a short-term market overreaction to information is more likely during the opening phase. In that case, we may expect a negative trend pattern in market mean returns after the market opens conditional on a positive overnight return, because investors are more willing to earn profits by selling stocks after the market opens. For a negative overnight return, we may expect a positive trend pattern in market mean returns after the market opens. Amihud and Mendelson (1987) find evidence that a short-term market under-reaction to information is more likely during the closing phase of the market. Hence, we may expect a positive (negative) trend pattern in the market mean returns conditional on a positive (negative) overnight return in the closing phase.

Table 4.2: Predictability during the first and the last half-hour. Panel A shows the results of regressing the cumulative returns during the first half-hour on the overnight returns, while Panel B shows the results of regressing the cumulative returns during the last half-hour on the overnight returns. The returns are in percentages and the coefficients are scaled up by 100. Newey and West (1987) robust t–statistics are in parentheses, and significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

	Dependent variable: $y^*_{t,\tau}$											
	Panel A: first half-hour						Panel B: last half-hour					
	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$
Intercept	0.05	0.01	−0.09	−0.17	−0.51	−0.21	0.11	0.06	1.12***	0.61	−0.34	−0.41
	(0.20)	(0.03)	(-0.21)	(-0.33)	(-0.87)	(-0.35)	(0.47)	(0.18)	(2.81)	(1.35)	(-0.71)	(-0.72)
ovr	−1.88**	−4.14***	−6.50***	−7.21***	−9.44***	−9.43***	0.01	1.72	3.29*	2.55	5.37**	10.95***
	(-2.10)	(-2.90)	(-2.84)	(-3.01)	(-3.27)	(-2.67)	(0.01)	(1.05)	(1.84)	(1.35)	(2.25)	(2.71)
$R^2(\%)$	0.84	2.15	3.15	3.03	3.98	3.34	0.00	0.43	0.99	0.44	1.48	3.80
$R^2_{\text{ad}}(\%)$	0.80	2.11	3.12	2.99	3.95	3.31	-0.04	0.39	0.96	0.40	1.45	3.76

When it comes to parametric CumRe, we find that the overnight returns are significant predictors for the cumulative returns during the first half-hour, according to Table 4.2.[10] In addition, overnight returns can also be used to predict the price movements from 15:30 to 15:55 and 15:30 to 16:00, i.e., the last half-hour before closing. Furthermore, we also find the negative relationship between the overnight returns and cumulative returns through the first 30 minutes, while the positive relationship between the overnight returns and returns in the last half-hour. Figure 4.2 shows the predicted intraday return patterns depending on different values of overnight returns, namely 0%, ± 1%, ± 2.5%, during the first and the last half-hours. The opposite trends within the first and the last half-hours are consistent with existing literature such as Amihud and Mendelson (1987), Hong and Wang (2000), and Menkveld *et al.* (2007).

[10]We also examined the cumulative returns for the whole trading session i.e. $y_t^{0,1}, y_t^{0,2}, \ldots, y_t^{0,78}$. The overnight returns are significant predictors for the intraday returns from 9:30 to 14:55. However, a full discussion of longer trading periods is beyond the scope of this chapter, since we focus on the first and and last half trading hours herein.

The parametric CumRe is able to reflect only linear dependence. The Ramsey RESET test (Ramsey, 1969) suggests that there are neglected non-linearities[11] in the setting of parametric CumRe, because quadratic or cubic terms of the fitted intraday return patterns are statistically significant in the conditional pattern forecasts (except for $y^l_{t,4}$, which is significant at the 10% significance level).

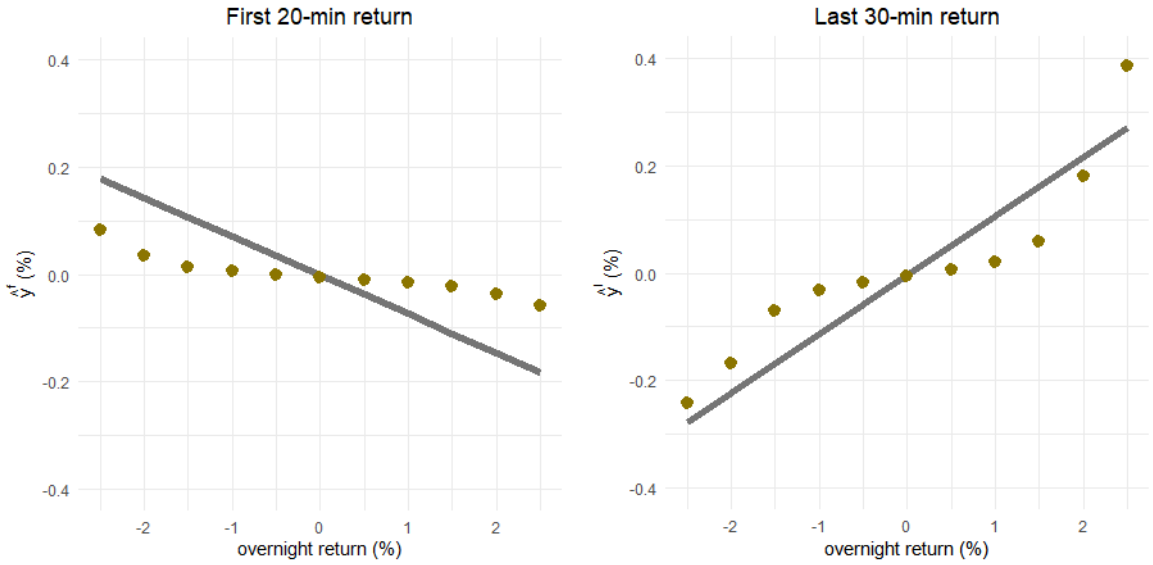


Figure 4.3: Comparison of parametric and non-parametric CumRe prediction. The left panel displays the predictions conditional on overnight returns during the first 20 trading minutes ($\tau = 4$). The right panel illustrates the predictions conditional on overnight returns for the last 30-min returns, that is, the returns from 360 to 390 trading minutes ($\tau = 6$). The overnight returns range from -2.5 to 2.5 on the figure which covers 99% of all observations. The solid line represents the predictions according to parametric CumRe. The dotted curve shows the predictions based on non-parametric CumRe. The returns are in percentages.

Next, we investigate non-parametric CumRe which may capture non-linearities. Figure 4.3 compares the forecasts provided by parametric CumRe and non-parametric CumRe for the first 20-min returns and the last 30-min returns, the periods of the highest profitability relative to risk, as we will show later. We observe substantial non-linearities in Figure 4.3. The predictions based on the non-parametric CumRe show a twisting shape, varying with magnitude of the overnight returns. The linear prediction seems to be affected by large overnight returns and may not be accurate for relatively small overnight returns. There is also a substantial asymmetry; with the same magnitude of overnight returns, the degree of reaction to a positive overnight return is

[11]We also applied the Probability Integral Transform (PIT) on the overnight returns prior to non-parametric CumRe to account for skewness and excess kurtosis/fat tails in return distributions. We found that the areas of predicted intraday returns conditional on positive and negative values of the overnight returns are asymmetric, implying that non-linearities should not be neglected. The results are available upon request.

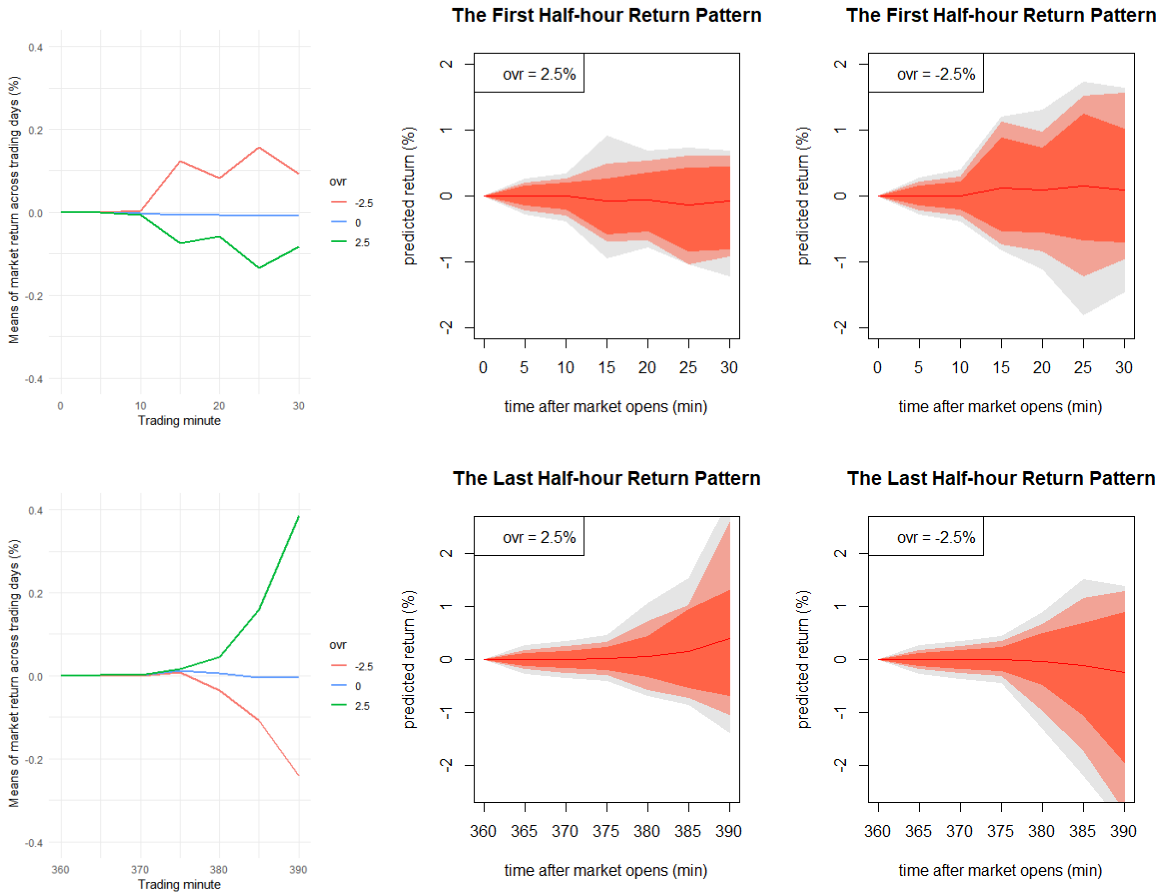


Figure 4.4: Predicted return pattern conditional on overnight return (-2.5%, 0%, and 2.5%). 80% prediction interval is shaded in red, 90% – light red and 90% – light grey. The predicted return is in percentage, and ovr is short for overnight return in the legend.

different from that to a negative overnight return. The returns for other time spans (τ-values) show similar qualitative asymmetries.

Figure 4.4 shows the predicted non-parametric return patterns during the first (top panel) and the last (bottom panel) half-hour trading periods conditional on negative and positive overnight returns of $\pm$ 2.5%.[12] We show conditional means (left side) and conditional the distribution by plotting the conditional median as well as 80%, 90% and 95% predictive intervals (right side). As in the linear case, the mean returns predicted non-parametrically given a positive value of the overnight return (green line) shows a downward trend within the first 30 trading minutes and an upward trend within the last 30 minutes; the pattern of mean returns conditional on a negative

[12]In Appendix III we show patterns conditional on $\pm$ 1% overnight returns and the patterns of the standardized returns $\tilde{y}_{t,\tau} = \frac{y_{t,\tau}}{\hat{\sigma}_{t,\tau}}$ conditional on overnight returns of 0, $\pm$ 1%, and $\pm$ 2.5%.

Table 4.3: Mean squared prediction error using leave-one-year-out cross-validation, Eq. (4.7). '$\emptyset$' indicates MSPEs from a constant 0-returns forecast used as a benchmark, 'P' indicates MSPEs from the parametric approach, 'NP' indicates MSPEs from the non-parametric approach. All numbers are scaled up by 100.

	First half-hour			Last half-hour		
	$\emptyset$	P	NP	$\emptyset$	P	NP
2003	10.97	11.00	10.95	6.58	6.75	6.67
2004	3.84	3.92	3.89	3.75	3.77	3.74
2005	2.91	3.08	2.95	3.23	3.21	3.22
2006	3.67	3.75	3.72	2.82	2.82	2.81
2007	6.03	6.07	5.97	8.13	7.92	8.06
2008	46.37	45.13	46.38	54.10	53.47	53.60
2009	26.70	27.66	27.15	18.87	19.39	18.85
2010	10.61	10.92	10.90	7.30	7.40	7.25
2011	14.72	14.86	14.67	12.86	12.84	12.88
2012	5.82	6.05	5.81	3.37	3.43	3.36
2013	4.46	4.56	4.48	2.49	2.46	2.48
Sum	136.11	137.00	136.86	123.52	123.48	122.93

overnight return (red line) shows opposite trends within the same time intervals. Moreover, the forecast distribution during the first half-hour is skewed to the left (right) conditional on a positive (negative) overnight return; while the forecast distribution for the last half-hour is skewed to the right (left) for a positive (negative) overnight return.

4.4 Forecasting Performance Evaluation

4.4.1 Leave-one-year-out cross-validation

Cross-validation (CV) is one of the most widely-used standard procedures for model evaluation in regression (Bergmeir *et al.*, 2018). Burman and Nolan (1992) propose bias correction and Burman *et al.* (1994) propose η-block CV to account for serial dependence in time series data. However, these cross-validation procedures leave out the possibly dependent observations and use data insufficiently (Bergmeir *et al.*, 2018). To reflect on time-series dependence and efficiently use the available time series data set, we propose the leave-one-year-out cross-validation to evaluate the time series pattern prediction.

Table 4.3 shows MSPEs using the leave-one-year-out evaluation for parametric CumRe (column P) and the non-parametric CumRe (column NP). These values are benchmarked against the MSPE

of a simple constant forecast of 0-returns (column $\emptyset$). While it is hard to outperform the benchmark for the first half-hour return pattern forecasts, the non-parametric CumRe overall shows smaller MSPE than the parametric CumRe in all of the years only except for year 2008. For the last half-hour return pattern forecasts, the conditional forecasts outperform the benchmark for most of the years, and overall the non-parametric CumRe shows better performance than the parametric CumRe again.

4.4.2 Comparison of predictive accuracy of patterns

Next we formally compare parametric and non-parametric CumRe-based forecasts using the Diebold and Mariano (1995) test statistic. Define the loss difference as the difference between the non-parameteric and parameteric squared prediction error on day t,

$$d_t = \mathrm{SPE}_{t,\mathrm{NP}} - \mathrm{SPE}_{t,\mathrm{P}},$$

with SPE_t as defined in Eq. (4.6). The Diebold-Mariano test statistic is defined as the average loss normalized by its standard error,

$$\text{t-stat}_{\mathrm{DM}} = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/P}},$$

where the average is taken over the evaluation period of P days. A Newey and West (1987) HAC robust estimator of σ_d^2 is used to account for serial dependence in the sequence of loss differences $\{d_t\}$.[13] The null hypothesis is that the two methods have the same forecast accuracy against the alternative of unequal forecast accuracy.

We used the estimation period from January 1, 2003 till March 9, 2009 and evaluate performance for the period from March 10, 2009 to December 31, 2013, the period right after the financial crisis. We find that for the first half-hour return patterns, the null hypothesis is rejected at the 1% significance level ($\bar{d} = -0.0054$ and $\text{t-stat}_{\mathrm{DM}} = -3.39$). For the last half-hour, the null is rejected at the 5% significance level ($\bar{d} = -0.0021$ and $\text{t-stat}_{\mathrm{DM}} = -2.22$). This gives a strong indication that the non-parametric CumRe outperforms the parametric CumRe at the forecast accuracy during the evaluation period.

[13] As in Diks *et al.* (2011), the HAC estimator of variance is calculated as $\hat{\sigma}^2 = \hat{\lambda}_0 + 2\sum_{z=1}^{Z-1} e_z \hat{\lambda}_z$, where $\hat{\lambda}_z$ denotes the lag-z sample covariance of the sequence of loss differences $\{d_t\}$ and e_z are the Bartlett weights $e_z = 1 - z/Z$ with $Z = \lfloor P^{1/4} \rfloor$, P is the number of days in the evaluation period.

4.5 Economic Significance

Next, we develop market timing strategies and compute utility gains based on CumRe for mean-variance day traders. For easier comparison with existing results we annualize returns by multiplying mean daily returns for a given strategy by 252 trading days per year. We will also compute annualized Sharpe ratios, $\sqrt{252}(\bar{r}/\hat{\sigma})$, where $\bar{r}$ and $\hat{\sigma}$ are the sample mean and the sample standard deviation of the realized daily returns of a strategy, respectively. In addition to simple standard deviation, we also compute a HAC robust estimator. To measure utility gains of portfolios constructed based on different expected returns, we use the annualized certainty equivalent rate of return (CER) which we describe in Section 4.5.2.

4.5.1 Market timing

Market timing is the practice of moving in and out of the market by attempting to predict the future direction of the market. To examine whether the overnight returns help predict the market direction in practice, we employ the overnight returns as market timing signals to trade in the market. Due to the negative relationship between the overnight returns and returns during first half-hour, on each trading day we will take a short position if the overnight return is positive at the beginning of the trading session; and then close it after $\tau \times 5$ min, $\tau \in \{1, \ldots, 6\}$; we take a long position otherwise. We examine different time intervals τ within the first half trading hour, for discovering when closing positions is most profitable for day traders.

Table 4.4 shows the results of these market timing strategies during the first and last half-hours. We compare the annualized returns based on the market time strategies I(ovr) with the returns based on simply holding long positions for the same periods.[14] When it comes to annualized returns, predicting the directions of the market during the first or the last half-hour based on the sign of overnight returns improves the performance of trading strategy. All of the annualized returns when we use the overnight return as the trading signal, on average, appear to be positive, while the simple long and short strategies yield smaller returns. The highest average return of 9.31% is obtained when we enter the market at 15:30, taking a position according to the sign of the overnight return on that day, and close the position at 16:00.

To take risk into account, we also present annualized Sharpe ratios of each timing strategy in Table 4.4. The higher the ratio, the better the performance. Conventionally, an annualized Sharpe ratio larger than 1 is considered acceptable to investors. During the first half-hour, holding a position for more than 10 minutes since 9:30 delivers a Sharpe ratio larger than 1. When it comes

[14] Holding a short position would result in an opposite return sign.

Table 4.4: Market timing strategies during the first and the last half-hours and the passive buy-and-hold strategy on a daily basis. The timing strategy denoted by I(ovr) takes a short (long) position in the market when the observed overnight return is positive (negative) during the first half-hour on each trading day t, or a long (short) position in the market when the observed overnight return is positive (negative) during the last half-hour on each trading day t. Long indicates always taking long positions for different holding periods during the first half trading hour. For each timing strategy, the average return (Avg ret) and standard deviation (Std dev) are annualized by multiplying by 252 and $\sqrt{252}$, repectively, and are expressed in percentages. We also present annualized Sharpe ratio (SRatio), skewness (Skew.), and kurtosis (Kurt.) for each strategy. Sharpe ratios calculated based on Newey-West HAC estimator of variance are in parentheses.

	First half-hour					**Last half-hour**				
Timing	Avg ret	Std dev	SRatio	Skew.	Kurt.	Avg ret	Std dev	SRatio	Skew.	Kurt.
			$\tau = 1$					$\tau = 1$		
I(ovr)	0.37	2.25	0.16 (0.16)	0.74	12.14	0.06	2.07	0.03 (0.02)	0.27	15.21
Long	0.04	2.25	0.02 (0.02)	0.21	12.17	0.29	2.07	0.14 (0.14)	-0.03	15.22
			$\tau = 2$					$\tau = 2$		
I(ovr)	2.52	3.08	0.82 (0.80)	1.32	15.35	1.31	2.88	0.45 (0.40)	1.00	14.15
Long	-0.19	3.09	-0.06 (-0.06)	1.23	15.56	0.25	2.89	0.09 (0.09)	-0.47	14.25
			$\tau = 3$					$\tau = 3$		
I(ovr)	4.85	3.99	1.22 (1.22)	1.46	17.68	3.14	3.61	0.87 (0.67)	0.82	13.91
Long	-0.57	4.00	-0.14 (-0.14)	1.29	17.97	2.98	3.61	0.83 (0.89)	-0.67	14.22
			$\tau = 4$					$\tau = 4$		
I(ovr)	5.70	4.52	1.26 (1.28)	1.56	16.25	2.08	4.22	0.49 (0.41)	0.18	16.21
Long	-0.79	4.53	-0.18 (-0.18)	1.02	16.59	1.66	4.22	0.39 (0.43)	-0.39	16.26
			$\tau = 5$					$\tau = 5$		
I(ovr)	6.32	5.16	1.23 (1.19)	1.45	14.73	4.29	4.81	0.89 (0.71)	1.14	14.51
Long	-1.76	5.17	-0.34 (-0.36)	0.66	15.07	-0.59	4.82	-0.12 (-0.14)	-0.28	14.67
			$\tau = 6$					$\tau = 6$		
I(ovr)	5.84	5.63	1.04 (1.03)	1.88	18.14	9.31	6.11	1.52 (1.16)	2.16	21.84
Long	-1.01	5.64	-0.18 (-0.20)	0.51	18.50	-0.48	6.14	-0.08 (-0.10)	-0.45	22.25
	Buy-and-hold									
	Avg ret	Std dev	SRatio	Skew.	Kurt.					
	1.30	16.23	0.08 (0.09)	-0.26	8.84					

to trades during the last half-hour, the market timing strategy conditional on the overnight return that holds the market for 30 minutes provides the highest Sharpe ratio of 1.52. All the timing strategies for the first and the last half-hours conditional on the overnight returns outperform the passive buy-and-hold strategy when we take risk into account. Moreover, the conditional timing strategies generate returns with positive skewness and much higher kurtosis, implying that these strategies often generate positive returns. We conclude that using the overnight return as the trading signal provides a good risk-return trade-off.

4.5.2 Utility Gains

Assume that a mean-variance day trader allocates funds between the market portfolio (SPY) with weight $\omega_{t,\tau}$ and a risk-free asset (3-month T-bill) with weight $(1 - \omega_{t,\tau})$. The portfolios are

constructed on a daily basis conditional on overnight returns and held for $\tau \times 5$ minutes, either during the first or last half-hour. Therefore, the risk-free rate is assumed to be 0. The weight on the market portfolio, which incorporates the investor's optimal trade-off between the expected return and risk, can be obtained by solving the following expected utility maximization problem (see e.g., Brandt, 2010),

$$\omega_{t,\tau} = \arg\max_{\omega} \left(\mathbb{E}(\omega \cdot r_{t+1,\tau}) - \frac{\gamma}{2} \mathrm{Var}(\omega \cdot r_{t+1,\tau}) \right), \tag{4.9}$$

where γ measures the level of relative risk aversion of the day trader. The solution of this maximization problem is

$$\omega_{t,\tau} = \frac{\hat{r}_{t+1,\tau}}{\gamma \hat{\sigma}^2_{t+1,\tau}}, \tag{4.10}$$

where $\hat{r}_{t+1,\tau}$ and $\hat{\sigma}^2_{t+1,\tau}$ denote investor's estimate on the mean and variance of the returns of market portfolio for time interval τ, respectively. In Campbell and Thompson (2008), Ferreira and Santa-Clara (2011), and Gao *et al.* (2018), the expected returns of the market portfolios predicted by predictors are used to construct the optimal portfolios for mean-variance investors who allocate wealth between a risky asset (a stock market index or a ETF of that index as a proxy of a market portfolio) and a risk-free asset. Similarly, we predict expected returns on day $t+1$ obtained from either parametric or non-parametric CumRe. The parameters of CumRe are estimated using an expanding window from January 1, 2003 to day t, where t ranges from March 9, 2009 to December 30, 2013. We also use the unconditional historical average returns, which is commonly used to compare the out-of-sample forecasting performance of predictors (see e.g., Campbell and Thompson, 2008), as a benchmark calculated over the expanding window. The estimate of variance is obtained by calculating the sample variance using an expanding window as well.[15]

There are different values of the risk aversion coefficient γ used in the literature: $\gamma = 2$ in Ferreira and Santa-Clara (2011), $\gamma = 3$ in Campbell and Thompson (2008), and $\gamma = 5$ in Gao *et al.* (2018). The higher the risk aversion coefficient, the lower the risk tolerance; and thus investors will invest less wealth on the market portfolio. In addition, we also consider $\gamma = 10$ to account for an even lower risk-tolerance case.

We impose the weight constraints for the market portfolios of different investment policy ranging from -4 to 4. This is because day traders in the US are allowed to use up to $4:1$ margin

[15]We also employed the GARCH model to obtain the estimate of variance. The corresponding utility gains results showed us similar qualitative and quantitative portfolio performance. For easier comparison with existing results in the literature, we use sample variance, $\hat{\sigma}^2_{t+1,\tau} = \frac{\sum_{s=1}^{t}(r_{s,\tau} - \bar{r}_\tau)^2}{t-1}$, as investor's estimate on the variance of the market portfolio on day $t+1$.

leverage according to the day-trading margin requirements.[16] In practice, we can short the SPY directly or buy the leveraged inverse ETFs (short ETFs or bear ETFs).[17]

We implement these portfolio policies according to the predetermined weights $\omega_{t,\tau}$ of Eq. (4.10) and thus the portfolio return on day $t+1$, $\rho_{t+1,\tau}$ is

$$\rho_{t+1,\tau} = \omega_{t,\tau} r_{t+1,\tau} \qquad \text{subject to} \qquad -4 \leq \omega_{t,\tau} \leq 4, \tag{4.11}$$

where $r_{t+1,\tau}$ is the realized return on the market portfolio during the time interval $\tau \times 5$ minutes on day $t+1$. We iterate this process over the out-of-sample period from March 10, 2009 to December 31, 2013 and get a time series of portfolio returns for each trading strategy. Since we employ the CER of each trading strategy to measure utility gains, we can compare the performance of the strategies that we use the unconditional historical mean (HM), and the parametric CumRe and the non-parametric CumRe both conditional on the overnight return to forecast the expected return. The CER of each portfolio is given by

$$\text{CER} = \hat{\mu}_\rho - \frac{\gamma \hat{\sigma}_\rho^2}{2}, \tag{4.12}$$

where $\hat{\mu}_\rho$ and $\hat{\sigma}_\rho^2$ are the sample mean and sample variance of the realized portfolio returns over the out-of-sample period. The higher the CER, the larger the risk-adjusted return. Similar to Gao *et al.* (2018), we use the difference between the CERs of strategies using the overnight return as the predictor and the strategies using the historical mean forecast to measure the predictability significance. We annualize the CER by multiplying by 252 and multiply the result by 100. Thus, this measure (CER dif) can be interpreted as the percentage utility gain per year of the conditional mean forecast based on the overnight return instead of the unconditional historical mean forecast. A positive CER dif indicates that the conditional forecast based on the overnight return provides higher utility gains than an unconditional mean forecast for a mean-variance investor, ceteris paribus.

[16] Day traders are allowed to have more leverage since their positions are short-term, and therefore each trade is likely to experience smaller price swings compared to positions held for days, weeks, or years. Financial Industry Regulatory Authority. Day-Trading Margin Requirements: Know the Rules. Retrieved from http://www.finra.org/investors/day-trading-margin-requirements-know-rules. Last access date 15-01-2019.

[17] For instances, Short S&P 500 (ticker symbol SH), UltraShort S&P 500 (ticker symbol SDS), and UltraPro Short S&P 500 (ticker symbol SPXU), or Daily S&P 500 Bear 1X ETF (ticker symbol SPDN), the Daily S&P 500 Bear 3X (ticker symbol SPXS) can deliver its $1\times$, $2\times$, or $3\times$ inverse exposure to the S&P 500 index. Likewise, we can go long by buying 2x and 3x Long ETFs or Bull ETFs to satisfy leverage ratios smaller than 4. We also examined the portfolio performance for different leverage ratios, which can be realized by trading these leveraged ETFs. We found that patterns in portfolio performance with lower leverage ratios are similar to the results in Table 4.5.

Table 4.5: The results of portfolio performance for different levels of risk tolerance day traders. This table shows the economic value of recursively forecasting the market return based on the overnight return during the first and the last half-hours. We apply the predicted return as the expected return and then construct optimal portfolios for mean-variance day traders with risk aversion coefficients of 2 and 10. We impose portfolio constraints from -4 to 4 according to the day-trading margin requirements in the US. The average return (Avg ret), standard deviation (Std dev), and the difference of certainty equivalent rate of return (CER dif) between the optimal mean-variance strategies (P is short for parametric and NP for non-parametric CumRe) and the corresponding historical mean forecast strategy (HM) are annualized by multiplying by 252 and reported in percentages. We also present Sharpe ratio (SRatio), skewness (Skew.), and kurtosis (Kurt.) for each strategy. Sharpe ratios calculated based on Newey-West HAC estimators of variance are given in parentheses. The one-step ahead predicted returns are determined using the expanding window from March 10, 2009.

$-4 \leq w \leq 4$, $\gamma = 2$

	First half-hour						**Last half-hour**					
	Avg ret	Std dev	SRatio	Skew.	Kurt.	CER dif	Avg ret	Std dev	SRatio	Skew.	Kurt.	CER dif
			$\tau = 1$						$\tau = 1$			
P	6.14	8.35	0.74 (0.71)	-0.15	6.36	9.15	-3.60	5.26	-0.68 (-0.64)	-1.02	11.34	-3.22
NP	-2.03	3.80	-0.54 (-0.53)	-1.42	16.20	1.53	-0.44	4.69	-0.09 (-0.09)	-0.23	6.02	0.00
HM	-3.64	2.60	-1.40 (-1.50)	-4.06	65.18		-0.44	4.70	-0.09 (-0.09)	-0.23	6.02	
			$\tau = 2$						$\tau = 2$			
P	17.29	11.59	1.49 (1.51)	0.25	3.01	18.93	-1.26	8.61	-0.15 (-0.14)	-0.08	5.45	2.23
NP	2.14	9.66	0.22 (0.23)	0.24	3.75	4.18	-3.75	6.40	-0.59 (-0.58)	0.02	8.16	0.07
HM	-2.53	6.74	-0.38 (-0.42)	-0.05	4.62		-3.79	6.59	-0.58 (-0.57)	0.11	8.24	
			$\tau = 3$						$\tau = 3$			
P	**20.43**	14.75	**1.38 (1.44)**	0.23	3.39	**20.01**	8.13	11.63	0.70 (0.63)	-0.47	7.37	-12.08
NP	6.76	13.67	0.49 (0.52)	0.32	4.52	6.64	**19.24**	11.52	**1.67 (1.60)**	0.21	7.18	-0.94
HM	-1.58	4.12	-0.39 (-0.42)	-0.35	3.62		**20.28**	11.94	**1.70 (1.66)**	0.15	6.61	
			$\tau = 4$						$\tau = 4$			
P	**22.54**	15.95	**1.41 (1.50)**	0.12	2.23	**20.72**	-1.56	13.50	-0.12 (-0.11)	-0.85	10.85	-10.02
NP	7.63	14.54	0.53 (0.58)	0.17	3.27	6.24	3.41	12.63	0.27 (0.25)	-0.24	7.88	-4.82
HM	-0.02	8.36	0.00 (0.00)	-0.11	2.59		8.10	12.11	0.67 (0.69)	0.59	10.89	
			$\tau = 5$						$\tau = 5$			
P	18.47	18.21	1.01 (1.08)	0.49	4.28	14.93	3.38	15.49	0.22 (0.20)	-0.31	8.82	6.83
NP	5.88	17.07	0.34 (0.37)	0.53	5.55	2.73	1.01	12.89	0.08 (0.07)	-0.37	17.94	5.20
HM	2.79	16.00	0.17 (0.19)	0.00	2.95		-5.38	6.81	-0.79 (-0.91)	-0.68	18.19	
			$\tau = 6$						$\tau = 6$			
P	12.13	19.94	0.61 (0.66)	0.70	4.35	12.05	**20.33**	19.25	**1.06 (0.99)**	0.36	13.83	**22.16**
NP	1.72	17.20	0.10 (0.11)	1.03	8.26	2.67	**16.09**	17.20	**0.94 (0.86)**	0.63	22.02	**18.66**
HM	-2.68	11.04	-0.24 (-0.26)	0.00	3.86		-5.16	6.14	-0.84 (-1.02)	-1.48	28.43	

$-4 \leq w \leq 4$, $\gamma = 10$

	First half-hour						**Last half-hour**					
	Avg ret	Std dev	SRatio	Skew.	Kurt.	CER dif	Avg ret	Std dev	SRatio	Skew.	Kurt.	CER dif
			$\tau = 1$						$\tau = 1$			
P	4.74	7.46	0.64 (0.60)	-0.21	10.75	2.71	-2.01	1.98	-1.02 (-0.89)	-4.74	63.77	-2.01
NP	-0.41	0.76	-0.53 (-0.53)	-1.41	16.16	0.32	-0.15	1.01	-0.14 (-0.15)	-0.32	6.51	0.00
HM	-0.74	0.53	-1.39 (-1.50)	-4.75	80.00		-0.15	1.01	-0.15 (-0.15)	-0.32	6.51	
			$\tau = 2$						$\tau = 2$			
P	14.27	10.31	1.38 (1.37)	0.35	5.47	9.59	-2.57	7.03	-0.37 (-0.35)	-0.24	11.36	-3.99
NP	0.52	2.45	0.21 (0.22)	-0.11	7.73	0.86	-0.95	1.51	-0.63 (-0.64)	-0.58	14.21	-0.02
HM	-0.55	1.38	-0.40 (-0.45)	-0.06	5.29		-0.92	1.60	-0.57 (-0.58)	-0.09	11.28	
			$\tau = 3$						$\tau = 3$			
P	**16.77**	13.02	**1.29 (1.33)**	0.32	6.14	**8.64**	5.30	9.82	0.54 (0.47)	-0.70	13.97	-6.21
NP	2.01	6.56	0.31 (0.31)	-0.85	40.73	0.21	7.71	4.46	**1.73 (1.65)**	0.47	5.80	**0.02**
HM	-0.32	0.82	-0.38 (-0.42)	-0.35	3.62		7.80	4.71	**1.65 (1.58)**	0.59	7.31	
			$\tau = 4$						$\tau = 4$			
P	**17.05**	13.91	**1.23 (1.27)**	0.23	4.66	**7.53**	-2.12	10.21	-0.21 (-0.18)	-1.19	28.62	-8.53
NP	2.31	6.92	0.33 (0.34)	-1.73	36.83	0.06	1.18	5.42	0.22 (0.19)	1.51	75.09	-1.49
HM	0.00	1.67	0.00 (0.00)	-0.11	2.59		1.58	2.75	0.58 (0.59)	0.53	9.28	
			$\tau = 5$						$\tau = 5$			
P	12.70	16.17	0.79 (0.83)	0.69	7.80	-0.39	1.30	13.10	0.10 (0.09)	-0.37	16.76	-6.11
NP	2.04	9.23	0.22 (0.23)	0.13	27.85	-2.24	1.65	8.14	0.20 (0.18)	-0.11	41.83	-0.49
HM	0.54	3.21	0.17 (0.18)	0.00	2.90		-1.08	1.36	-0.79 (-0.91)	-0.68	18.19	
			$\tau = 6$						$\tau = 6$			
P	7.54	17.40	0.43 (0.47)	0.99	8.25	-6.81	**15.45**	16.99	**0.91 (0.84)**	0.56	23.35	**2.12**
NP	-1.65	7.44	-0.22 (-0.24)	-0.35	37.49	-3.64	**12.02**	12.15	**0.99 (0.87)**	2.22	63.76	**5.74**
HM	-0.54	2.21	-0.24 (-0.26)	0.00	3.86		-1.03	1.23	-0.84 (-1.02)	-1.48	28.43	

Table 4.5 reports the results of the out-of-sample portfolio performance for mean-variance investors with different levels of risk tolerance, who enter the market either during the first or the last half-hour to open a position and then close the position after $\tau \times 5$ minutes.[18]

For the first half hour we find that portfolios constructed based on the expected returns predicted by parametric CumRe outperform those based on non-parametric CumRe or the unconditional historical mean forecasts for all holding periods. Among others, holding a position for 20 minutes ($\tau = 4$) from the beginning of each trading session delivers the highest positive annualized returns, which also yields a Sharpe ratio of over 1. In addition, the positive CERs in most cases imply that the conditional mean forecasts of expected returns depending on the overnight returns usually generate higher utility gains relative to the unconditional mean forecasts of expected returns for the mean-variance investors.

For the last half trading hour, while in the sense of constructing portfolios it is difficult for the conditional mean forecasts to consistently surpass the unconditional mean forecasts of expected returns for all the holding periods, there are two holding periods, 15-min ($\tau = 3$) and 30-min ($\tau = 6$), particularly notable. When $\tau = 3$, the portfolios depending on the non-parametric CumRe perform similar to those depending on the unconditional mean forecasts; and these portfolios always yield the highest (HAC) Sharpe ratios (larger than 1) among portfolios across all holding periods, which is even higher than the most profitable case that the weights on the market portfolios depend on the expected returns predicted by parametric CumRe with a holding period of 30-minute.

Notice that given that a day trader with a risk aversion coefficient of 2 invests on the market portfolio for the last half hour, the annualized CER dif of the portfolio depending on the overnight return, 22.16% per year, is 15.55% higher than the CER dif depending on the predictors of the first and the twelfth half-hour returns in Gao *et al.* (2018) (6.61% in Table A.7 in their online Appendix). This is because we impose the portfolio constraint that the weight on the market portfolio lies inside $[-4, 4]$, while the constraint in Gao *et al.* (2018) is $[-0.5, 1.5]$. If the constraint is the same, CER dif of the portfolio based on the overnight return is 7.57% and this portfolio yields a higher Sharpe ratio of 1.19. The unconstrained portfolio based on the overnight return generates an annualized return of 244.57% with a Sharpe ratio of 1.06 for the last half-hour.

We observe that the distributions of returns of portfolios depending on parametric and non-parametric CumRe have asymmetries and excess kurtosis. The Jarque-Bera test results suggest that these return series are not normally distributed with highly statistical significance. Autocorrelation and heteroskedasticity are present in these return series. While the measure of CER

[18]We found results of portfolio performance corresponding to different levels of risk tolerance such as $\gamma = 3, 5$ similar to Table 4.5, which are available upon request.

dif provides some qualitative evidence, it is not based on proper statistical inference procedure. In order to introduce proper inference, we use a related measure, which we refer to as Realized Utility (RU),

$$\mathrm{RU}_{t,\tau} = \omega_{t,\tau} \cdot r_{t+1,\tau} - \frac{\gamma}{2}\omega_{t,\tau}^2 \hat{\sigma}_{t+1,\tau}^2, \tag{4.13}$$

where $\hat{\sigma}_{t+1,\tau}^2$ is variance forecast for which we use the GARCH (1,1) model. For comparison with the CER dif results in Table 4.5, we use the same weights, $\omega_{t,\tau}$.

To compare the realized utilities depending on non-parametric ($\mathrm{RU_N}$) and parametric CumRe ($\mathrm{RU_P}$) for an investor who has a risk aversion coefficient of γ, we define the Diebold-Mariano test statistic as $\text{t-stat}_{\mathrm{DM}} = \frac{\bar{d}_\gamma}{\sqrt{\hat{\sigma}_{d_\gamma}^2/P}}$, where d_γ is the difference between $\mathrm{RU_N}$ and $\mathrm{RU_P}$, $\hat{\sigma}_{d_\gamma}^2$ is a HAC robust estimator of variance to account for serial correlations and heteroskedasticity in portfolio return series over the evaluation period. A positive $\bar{d}_\gamma$ indicates that the portfolio constructed based on non-parametric CumRe on average yields higher utility gains relative to that based on the parametric CumRe to investors with the risk aversion coefficient γ. Similar to CER dif, we measure the realized utility gains in percentage and annualize d_γ by multiplying by 252.

Table 4.6: Constrained and unconstrained realized utility results according to Eq. (4.13). For easier comparison, we measure realized utility gains in percentage and annualize the difference between the realized utilities depending on non-parametric and parametric CumRe by multiplying by 252. The Newey-West (1987) robust t-statistic are given in parentheses.

	$-4 \le w \le 4$											
	First half-hour						**Last half-hour**					
	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$
$\bar{d}_2$	-7.32^{**}	-14.65^{***}	-13.42^{**}	-14.42^{**}	-12.21^{*}	-9.05	3.20^{**}	-1.74	10.99	4.89	-0.57	-1.90
	(-1.99)	(-2.79)	(-2.14)	(-2.41)	(-1.87)	(-1.45)	(2.00)	(-0.36)	(1.60)	(1.08)	(-0.16)	(-0.59)
$\bar{d}_{10}$	-1.44	-7.88^{*}	-7.07	-5.66	0.08	4.08	1.99^{**}	5.20	7.21	7.48^{**}	7.67^{**}	9.14^{**}
	(-0.41)	(-1.81)	(-1.61)	(-1.24)	(0.02)	(0.72)	(2.22)	(1.55)	(1.51)	(2.11)	(2.15)	(2.34)
	Unconstrained Weights											
	First half-hour						**Last half-hour**					
	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$
$\bar{d}_2$	40.23	103.29	176.16^{**}	161.49^{**}	260.48^{***}	276.76^{***}	11.31^{*}	81.15^{**}	57.57	33.21	80.78^{*}	113.15^{**}
	(0.79)	(1.29)	(2.29)	(2.23)	(2.99)	(3.17)	(2.48)	(2.14)	(1.07)	(1.33)	(1.95)	(2.09)
$\bar{d}_{10}$	8.05	20.66	35.23^{**}	32.30^{**}	52.10^{***}	55.35^{***}	2.26^{*}	16.23^{**}	11.51	6.64	16.16^{*}	22.63^{**}
	(0.79)	(1.29)	(2.29)	(2.23)	(2.99)	(3.17)	(2.48)	(2.14)	(1.07)	(1.33)	(1.95)	(2.09)

Table 4.6 reports the realized utility difference between the portfolios constructed depending on the non-parametric CumRe and the parametric CumRe forecasts. We find that, with the day-trading margin requirements in the US, trading based on parametric CumRe predictions during the first half hour can generate significantly higher realized utilities for mean-variance investors with lower level of risk aversion ($\gamma = 2$); however, non-parametric CumRe forecasts provide significantly higher realized utilities for higher level of risk aversion ($\gamma = 10$) investors during the last half hour. Recall the twisting shape of non-parametric CumRe predictions and the linear

shape of parametric CumRe predictions in Figure 4.3, employing parametric CumRe predictions may lead to the magnitude of weights determined according to these predictions being higher than those determined according to non-parametric CumRe predictions, since 99% of the observations of overnight returns range between -2.5% and 2.5% over the evaluation period. In the absence of day-trading margin requirements for day traders, the trading strategies depending on non-parametric CumRe forecasts yield significantly higher realized utility gains relative to strategies depending on parametric CumRe forecasts.

4.6 Conclusions

Our results show that it is possible to forecast the subsequent intraday patterns in market returns based on observed market overnight returns (close-to-open returns) with both statistical and economic significance. Our forecasting approach CumRe confirms the intraday patterns in returns documented in the existing theoretical and empirical literature, such as the opposite trends during the beginning period and the closing period within the same trading day (Hong and Wang, 2000); a short-term market overreaction to information is possible in the opening, while the under-reaction effect is likely at the close (see e.g., Amihud and Mendelson, 1987); the positive or negative relationship depending on the overnight returns during the first or the last half trading hours (Liu and Tse, 2017; Gao *et al.*, 2018). Moreover, the overnight returns help predict the market directions within the first and the last half trading hours. In addition, provided that the dependence between the overnight returns and certain subsequent intraday returns is statistically significant, the magnitudes and the signs of the overnight returns can also help forecast the expected returns and thus yield higher annualized returns, Sharpe ratios, and certainty equivalent returns than the unconditional mean pattern forecasts of the expected returns—one-step-ahead (cumulative) forecasts based on a recursive approach—for mean-variance day traders.

We find non-linearities in return patterns. The non-parametric CumRe can accommodate nonlinear dependence between the overnight returns and the subsequent intraday returns, while the parametric CumRe only accounts for linear relationship. Based on transaction-level data of SPDR S&P 500 ETF from January 1, 2003 to December 31, 2013, which is an exchange-traded fund that closely tracks the S&P 500 index with the smallest bid-ask spread, we find that, statistically, non-parametric CumRe improves the out-of-sample forecasting performance in terms of the Diebold and Mariano (1995) type test statistic and the leave-one-year-out cross-validation errors.

There are a number of possible extensions to predict intraday returns based on overnight

returns. First, it might be helpful to incorporate ARMA and/or GARCH structure into the forecast methods or to take the asymmetry and fat-tails in returns into account when forecasting the intraday returns. Second, one might wonder whether overnight returns could be used as a predictor for other financial markets such as futures, currencies, and commodities. Third, we wonder whether this intraday market predictability can contribute to volatility forecasting and thus influences the performance of the portfolios constructed by both conditional mean and variance forecasts depending on the overnight returns. Fourth, as stated by Brandt (2010), the mean-variance problem ignores any preferences toward higher-order return moments, and thus we would be interested in the portfolio performance if the weights on the market portfolio are determined by a higher-order approximation of expected utility maximization. We leave these for future work.

4.7 Appendix III

4.7.1 Correlation Matrix

This table reports the correlation coefficients between the overnight returns and the subsequent intraday returns during the first and the last half hours. The variable ovr is short for overnight returns. The variables $y^f_{t,\tau}$ and $y^l_{t,\tau}$ denote the τ^{th} element, where $\tau = 1, \ldots, 6$ of the first and last half hour return pattern, respectively. The sample period is from January 1, 2003, to December 31, 2013.

Table A.1: Correlation Matrix.

	ovr	$y^f_{t,1}$	$y^f_{t,2}$	$y^f_{t,3}$	$y^f_{t,4}$	$y^f_{t,5}$	$y^f_{t,6}$	$y^l_{t,1}$	$y^l_{t,2}$	$y^l_{t,3}$	$y^l_{t,4}$	$y^l_{t,5}$	$y^l_{t,6}$
$y^f_{t,1}$		1.00	0.65	0.47	0.46	0.45	0.45						
$y^f_{t,2}$			1.00	0.76	0.69	0.65	0.63						
$y^f_{t,3}$				1.00	0.87	0.81	0.74						
$y^f_{t,4}$					1.00	0.91	0.84						
$y^f_{t,5}$						1.00	0.93						
$y^f_{t,6}$							1.00						
$y^l_{t,1}$								1.00	0.72	0.54	0.40	0.38	0.36
$y^l_{t,2}$									1.00	0.75	0.62	0.59	0.54
$y^l_{t,3}$										1.00	0.87	0.79	0.70
$y^l_{t,4}$											1.00	0.88	0.79
$y^l_{t,5}$												1.00	0.91
$y^l_{t,6}$													1.00
ovr	1.00	-0.09	-0.15	-0.18	-0.17	-0.20	-0.18	0.00	0.07	0.10	0.07	0.12	0.19

4.7.2 Intraday Return Patterns

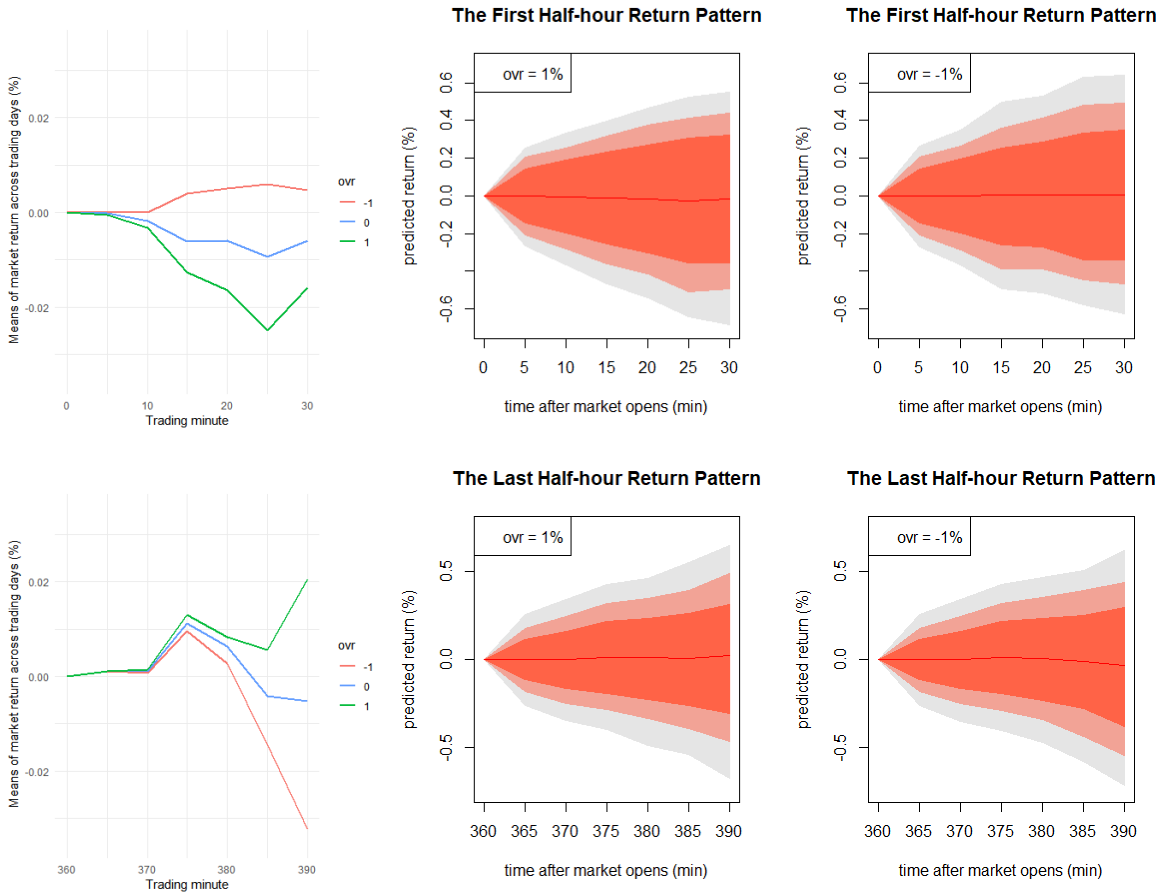


Figure B.1: Non-parametric CumRe: Predicted return pattern conditional on overnight returns $(-1\%, 0\%, +1\%)$. The outer grey shade is 80% prediction interval, the inner shades are respectively 90% and 95% prediction intervals. The predicted return is in percentage, and ovr is short for overnight return in the legend. Note that we take a relatively conservative overnight return 1% as an example. These patterns are more striking based on a larger magnitude of overnight return.

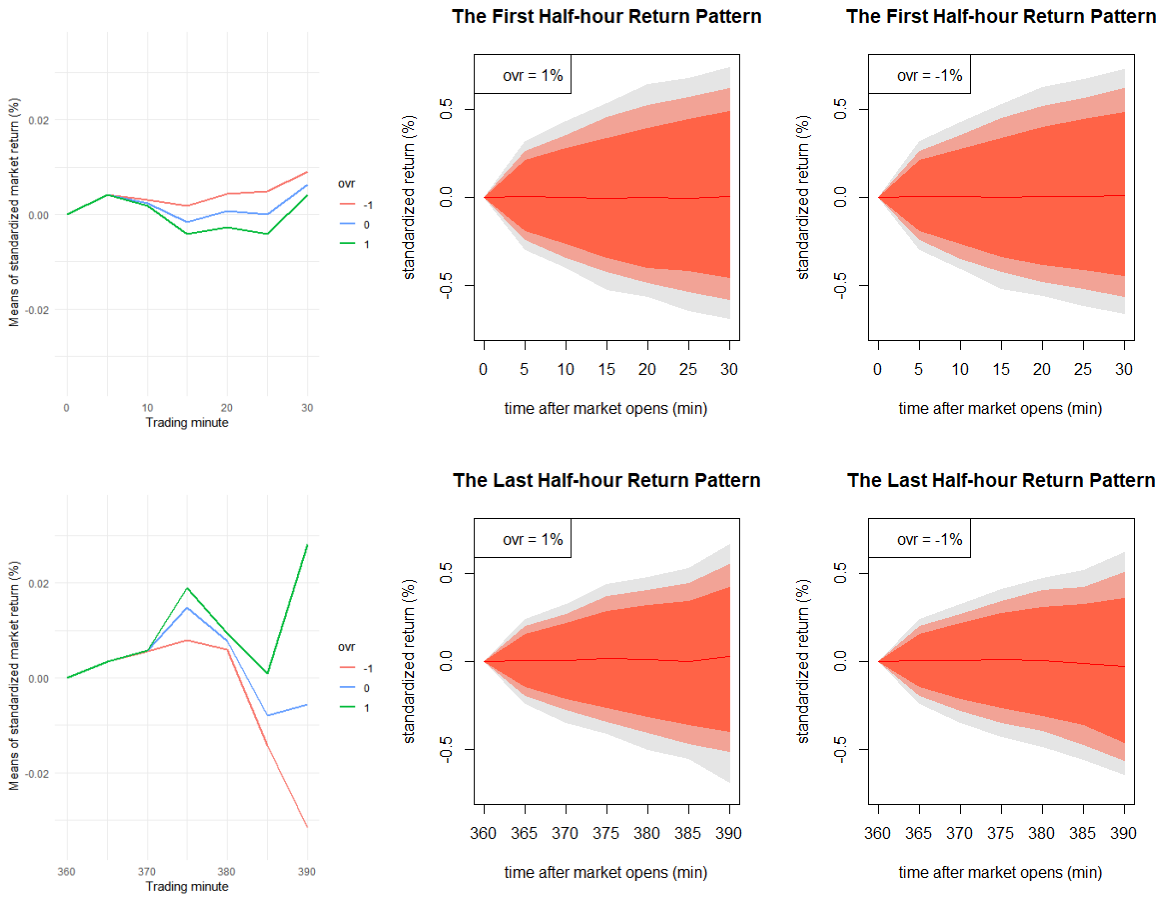


Figure B.2: Non-parametric CumRe: Predicted return pattern conditional on overnight returns $(-1\%, 0\%, +1\%)$. 80% prediction interval is shaded in red, 90% – light red and 90% – light grey. The intraday returns series are standardized by $\tilde{Y}_{t,\tau} = \frac{Y_{t,\tau}}{\hat{\sigma}_t}$, where σ_t is the conditional standard deviation of $Y_{t,\tau}$, σ_t is estimated by realized volatility (the square root of the realized variance). The predicted returns are in percentages, and ovr is short for overnight return in the legend.

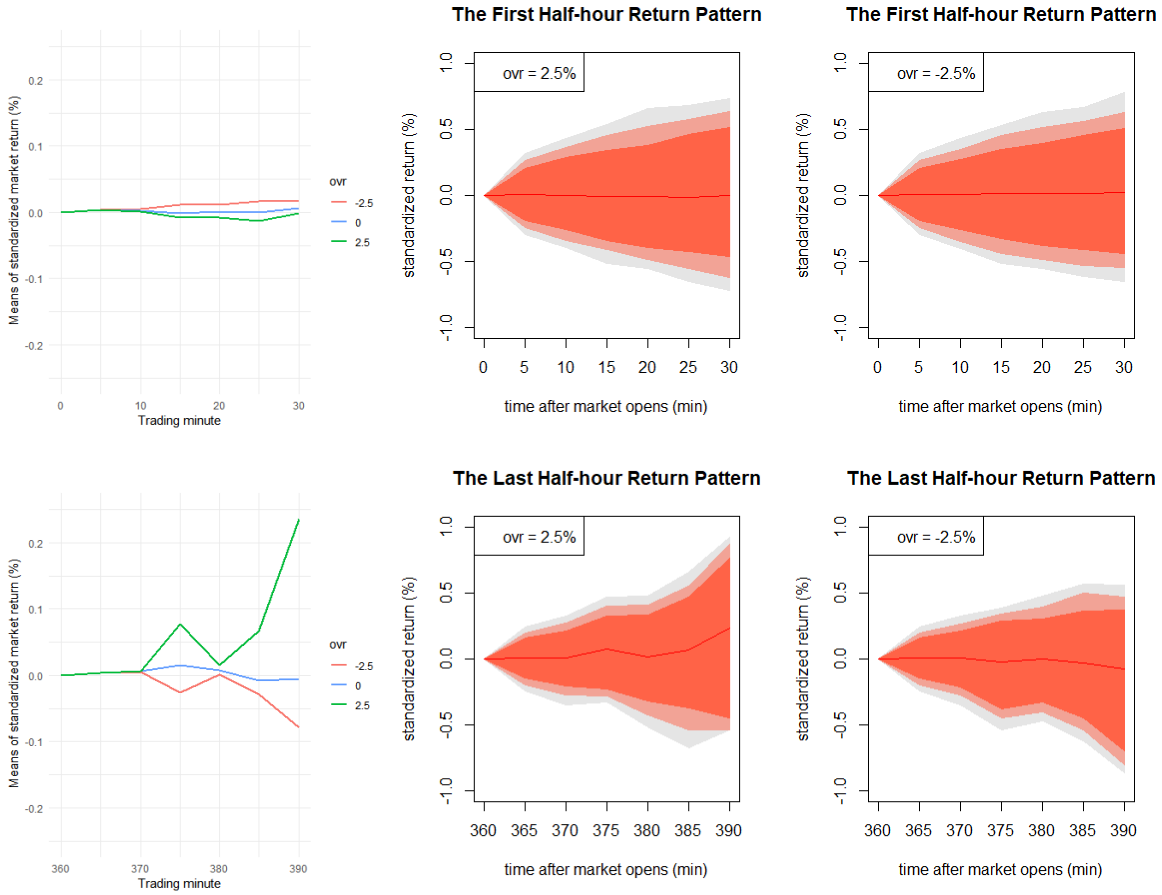


Figure B.3: Non-parametric CumRe: Predicted return pattern conditional on overnight returns $(-2.5\%, 0\%, +2.5\%)$. 80% prediction interval is shaded in red, 90% – light red and 90% – light grey. The intraday returns series are standardized by $\tilde{Y}_{t,\tau} = \frac{Y_{t,\tau}}{\hat{\sigma}_t}$, where σ_t is the conditional standard deviation of $Y_{t,\tau}$, σ_t is estimated by realized volatility (the square root of the realized variance). The predicted returns are in percentages, and ovr is short for overnight return in the legend.

Bibliography

Abhyankar, A., Copeland, L. S. and Wong, W. (1997). Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics*, **15,** number 1, 1–14.

Affleck-Graves, J. and McDonald, B. (1989). Nonnormalities and tests of asset pricing theories. *The Journal of Finance*, **44,** 889–908.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19,** 716–723.

Almeida, C., Ardison, K., Kubudi, D., Simonsen, A. and Vicente, J. (2018). Forecasting bond yields with segmented term structure models. *Journal of Financial Econometrics*, **16,** number 1, 1–33.

Amenc, N. and Martellini, L. (2002). Portfolio optimization and hedge fund style allocation decisions. *Journal of Alternative Investments*, **5,** 7–20.

Amihud, Y. and Mendelson, H. (1987). Trading mechanisms and stock returns: An empirical investigation. *The Journal of Finance*, **42,** number 3, 533–553.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, **4,** number 2, 115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, **73,** number 2, 251–277.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, **20,** number 3, 651–707.

Ang, A., Hodrick, R.J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, **61,** 259–299.

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110,** number 509, 378–392.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12,** 171–178.

Azzalini, A (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46,** 199–208.

Bali, T.G. and Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis*, **43,** 29–58.

Bardsley, P., Horváth, L., Kokoszka, P. and Young, G. (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal*, **20,** number 1, 86–117.

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, **38,** number 6, 3352–3386.

Bergmeir, C., Hyndman, R. J. and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, **120,** 70–83.

Berkman, H., Koch, P. D., Tuttle, L. and Zhang, Y. J. (2012). Paying attention: Overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, **47,** number 4, 715–741.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business*, **45,** 444–455.

Blattberg, R.C. and Gonedes, N.J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, **47,** 244–280.

Boffelli, S., Skintzi, V. D. and Urga, G. (2016). High-and low-frequency correlations in European government bond spreads and their macroeconomic drivers. *Journal of Financial Econometrics*, **15,** number 1, 62–105.

Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance*, **71,** number 6, 2967–3006.

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. *Journal of Financial Economics*, **141,** number 1, 172–194.

Bowsher, C. G. and Meeks, R. (2008). The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of the American Statistical Association*, **103,** number 484, 1419–1437.

Brandt, M. W. (2010). Portfolio choice problems, chapter 5. In *Handbook of Financial Econometrics: Tools and Techniques*, *San Diego* (eds Yacine Aït-Sahalia and Lars Peter Hansen), Handbooks in Finance, volume 1, pp. 269 – 336. North-Holland.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, **51,** number 4, 2232–2245.

Burman, P., Chow, E. and Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, **81,** number 2, 351–358.

Burman, P. and Nolan, D. (1992). Data-dependent estimation of prediction functions. *Journal of Time Series Analysis*, **13,** number 3, 189–207.

Caldeira, J. and Torrent, H. (2017). Forecasting the us term structure of interest rates using nonparametric functional data analysis. *Journal of Forecasting*, **36,** number 1, 56–73.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, **21,** number 4, 1509–1531.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *Econometrics of Financial Markets*. Princeton University Press. Princeton, NJ.

Chambers, D. R., Carleton, W. T. and Waldman, D. W. (1984). A new approach to estimation of the term structure of interest rates. *Journal of Financial and Quantitative Analysis*, **19,** number 3, 233–252.

Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting*, **15,** 495–508.

Christensen, B. J. and van der Wel, M. (2019). An asset pricing approach to testing general term structure models. *Journal of Financial Economics*, **134,** number 1, 165–191.

Christensen, J. H. E., Diebold, F. and Rudebusch, G. (2011). The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, **164,** number 1, 4–20.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2009). An arbitrage-free generalized nelson–siegel term structure model. *The Econometrics Journal*, **12,** number 3, C33–C64.

Cliff, M., Cooper, M. J. and Gulen, H. (2008). Return differences between trading and non-trading hours: Like night and day. *Available at SSRN:https://ssrn.com/abstract=1004081*.

Cootner, P. (1964). *The Random Character of Stock Prices*. MIT Press. Cambridge, MA.

Cox, J. C., Ingersoll, J. E., Jr. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, **53,** number 2, 385–407.

Crump, R. K. and Gospodinov, N. (2021). On the factor structure of bond returns. *Econometrica*, Forthcoming.

Davidson, R. and MacKinnon, J.G. (2004). *Econometric Theory and Methods*. Oxford University Press. Oxford, UK.

De Boor, C. (1978). *A practical guide to splines*, volume 27. springer-verlag New York.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, **22,** 1915–1953.

DiCiccio, T.J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, **11,** 189–228.

Dickey, D. A. and Fuller, W. A (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74,** number 366a, 427–431.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, **130,** number 2, 337–364.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13,** number 1, 253–263.

Diks, C., Panchenko, V. and Dijk, D. Van (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, **163,** number 2, 215–230.

Duffee, G. R. (2011). Information in (and not in) the term structure. *The Review of Financial Studies*, **24,** number 9, 2895–2934.

Elton, E.J., Gruber, M.J. and Padberg, M.W. (1978). Simple criteria for optimal portfolio selection: tracing out the efficient frontier. *The Journal of Finance*, **33,** 296–302.

Fama, E.F. (1965a). The behavior of stock-market prices. *Journal of Business*, **38,** 34–105.

Fama, E.F. (1965b). Portfolio analysis in a stable Paretian market. *Management Science*, **11,** 404–419.

Fama, E.F. (1971). Risk, Return, and Equilibrium. *Journal of Political Economy*, **79,** 30–55.

Fama, E.F. and French, K.R. (1996). The CAPM is wanted, dead or alive. *The Journal of Finance*, **51,** 1947–1958.

Feng, P. and Qian, J. (2018). Forecasting the yield curve using a dynamic natural cubic spline model. *Economics Letters*, **168,** 73–76.

Fernández, C., Osiewalski, J. and Steel, M.F.J. (1995). Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association*, **90,** 1331–1340.

Ferreira, M. A. and Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, **100,** number 3, 514–537.

Gao, L., Han, Y., Li, S. Z. and Zhou, G. (2018). Market intraday momentum. *Journal of Financial Economics*, **129,** number 2, 394 – 414.

Gerety, M. S. and Mulherin, J. H. (1994). Price formation on stock exchanges: The evolution of trading within the day. *Review of Financial Studies*, **7,** number 3, 609–629.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21,** 1455–1508.

Guidolin, M., Hyde, S., McMillan, D. and Ono, S. (2009). Non-linear predictability in stock and bond returns: When and where is it exploitable? *International Journal of Forecasting*, **25,** number 2, 373–399.

Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the term structure. *Journal of Economic Literature*, **50,** number 2, 331–67.

Hamilton, J. D. and Wu, J. C. (2014). Testable implications of affine term structure models. *Journal of Econometrics*, **178,** 231–242.

Härdle, W. K. and Majer, P. (2016). Yield curve modeling and forecasting using semiparametric factor dynamics. *The European Journal of Finance*, **22,** number 12, 1109–1129.

Härdle, W. K., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer.

Harris, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, **16,** number 1, 99–117.

Harvey, A.C. (1981). *The Econometric Analysis of Time Series*. Philip Allan. London.

Hays, S., Shen, H. and Huang, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, **6,** number 3, 870–894.

Heston, S. L., Korajczyk, R. A. and Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, **65,** number 4, 1369–1407.

Hodgson, D.J., Linton, O. and Vorkink, K. (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *Journal of Applied Econometrics*, **17,** 617–639.

Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, **55,** number 1, 297–354.

Hörmann, S., Kidziński, Ł. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, **77,** number 2, 319–348.

Hull, J.C. (2011). *Options, Futures, and Other Derivatives*, eighth edn. Prentice Hall. New Jersey.

Hyndman, Rob J and Ullah, Md Shahid (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, **51,** number 10, 4942–4956.

Jain, P. C. and Joh, G. (1988). The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, **23,** number 3, 269–283.

Jensen, M.C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, **23,** 389–416.

Joslin, S., Priebsch, M. and Singleton, K. J (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, **69,** number 3, 1197–1233.

Jungbacker, B., Koopman, S. J. and Van der Wel, M (2014). Smooth dynamic factor analysis with application to the us term structure of interest rates. *Journal of Applied Econometrics*, **29,** number 1, 65–90.

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, **22,** 891–921.

Koo, B., La Vecchia, D. and Linton, O. (2021). Estimation of a nonparametric model for bond prices from cross-section and time series information. *Journal of Econometrics*, **220,** number 2, 562–588.

Koopman, S. J., Mallee, M. and Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, **28,** number 3, 329–343.

Li, L. and Lin, M. (2014). Analysis of French stock market's CAPM based on asymmetric exponential power distribution. *International Journal of Applied Mathematics and Statistics*, **52,** 84–95.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, **20,** 587–615.

Linton, O., Mammen, E., Nielsen, J. P. and Tanggaard, C. (2001). Yield curve estimation by kernel smoothing methods. *Journal of Econometrics*, **105,** number 1, 185–223.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, **1,** number 1, 54–61.

Liu, Q. and Tse, Y. (2017). Overnight returns of stock indexes: Evidence from ETFs and futures. *International Review of Economics & Finance*, **48,** 440–451.

Liu, Y. and Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics* forthcoming.

Maasoumi, E. and Racine, J. (2002). Entropy and predictability of stock market returns. *Journal of Econometrics*, **107,** number 1, 291–312.

Mandelbrot, B. (1963a). New methods in statistical economics. *Journal of Political Economy*, **71,** 421–440.

Mandelbrot, B. (1963b). The variation of certain speculative prices. *Journal of Business*, **36,** 394–419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *Journal of Business*, **40,** 393–413.

Maneesoonthorn, W., Forbes, C. S. and Martin, G. M. (2017). Inference on self-exciting jumps in prices and volatility using high-frequency measures. *Journal of Applied Econometrics*, **32,** number 3, 504–532.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7,** 77–91.

Markowitz, H. (1956). The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, **3,** 111–133.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, Inc. New York.

Menkveld, A. J., Koopman, S. J. and Lucas, A. (2007). Modeling around-the-clock price discovery for cross-listed stocks using state space methods. *Journal of Business & Economic Statistics*, **25,** number 2, 213–225.

Morningstar (2005). Standard deviation and Sharpe ratio. Morningstar Methodology Paper (31 January). Retreived 7 March 2016.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, **34,** 768–783.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, **60,** number 4, 473–489.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, **59,** 347–370.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55,** number 3, 703–708.

Oldfield, G. S. and Rogalski, R. J. (1980). A theory of common stock returns over trading and non-trading periods. *The Journal of Finance*, **35,** number 3, 729–751.

Opsomer, Jean, Wang, Yuedong and Yang, Yuhong (2001). Nonparametric regression with correlated errors. *Statistical Science*, **16,** 134–153.

Otto, S. and Salish, N. (2019). A dynamic functional factor model for yield curves: identification, estimation, and prediction. Working paper.

Owen, J. and Rabinovitch, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance*, **38,** 745–752.

Peña, D. and Rodriguez, J. (2005). Detecting nonlinearity in time series by model selection criteria. *International Journal of Forecasting*, **21,** number 4, 731–748.

Perera, I. and Koul, H. L. (2017). Fitting a two phase threshold multiplicative error model. *Journal of Econometrics*, **197,** number 2, 348–367.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. John Wiley & Sons, Ltd. Chichester, UK.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies*, volume 77. Springer.

Ramsay, JO, Hooker, Giles and Graves, Spencer (2009). Introduction to functional data analysis. In *Functional data analysis with R and MATLAB*, pp. 1–19. Springer.

Ramsey, J.B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series* B, **31,** 350–371.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In *Handbook of Economic Forecasting*, volume 2, pp. 328–383. Elsevier.

Raviv, E. (2015). Prediction bias correction for dynamic term structure models. *Economics Letters*, **129,** 112–115.

Rogalski, R. J. (1984). New findings regarding day-of-the-week returns over trading and non-trading periods: A note. *The Journal of Finance*, **39,** number 5, 1603–14.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6,** 461–464.

Sen, R. and Klüppelberg, C. (2019). Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*, **35,** number 4, 1028–1043.

Sharpe, W.F. (1963). A simplified model for portfolio analysis. *Management Science*, **9,** 277–293.

Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, **19,** 425–442.

Sharpe, W.F. (1970). *Portfolio Theory and Capital Markets*. McGraw-Hill. New York.

Sharpe, W.F. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, **6,** 1263–1275.

Sharpe, W.F. (1978). An algorithm for portfolio improvement. In *Advances in Mathematical Programming and Financial Planning* (eds J.B. Lawrence. K.D.and Guerard. and G.D. Reeves), pp. 155–170. JAI Press, Inc. Stamford.

Sharpe, W.F. (2007a). Expected utility asset allocation. *Financial Analysts Journal*, **63,** 18–30.

Sharpe, W.F. (2007b). *Investors and markets: Portfolio choices, asset prices, and investment advice*. Princeton University Press. Princeton, NJ.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall. London.

Smirlock, M. and Starks, L. (1986). Day-of-the-week and intraday effects in stock returns. *Journal of Financial Economics*, **17,** number 1, 197–210.

Söderlind, P. and Svensson, L. (1997). New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics*, **40,** number 2, 383–429.

Spanos, A. (2010). Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics*, **158,** 204–220.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, **54,** number 3, 375–421.

Subbotin, M.T. (1923). On the law of frequency of error. *Matemat Sbornik*, **31,** 296–301.

Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical Report. National Bureau of Economic Research.

Swanson, E. T. (2021). Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, **118,** 32–53.

Swanson, E. T. and Williams, J. C. (2014). Measuring the effect of the zero lower bound on medium-and longer-term interest rates. *American Economic Review*, **104,** number 10, 3154–85.

Theodossiou, P. (2000). Skewed generalized error distribution of financial assets and option pricing. Technical Report. Cyprus University of Technology, Cyprus.

Thomas, R. and Gup, B.E. (2010). *The valuation handbook: valuation techniques from today's top practitioners*. John Wiley & Sons, Inc. Hoboken, NJ.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons.

Van Dijk, D., Koopman, S. J., Van der Wel, M. and Wright, J. H. (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, **29,** number 5, 693–712.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5,** number 2, 177–188.

Wood, R. A., McInish, T. H. and Ord, J. K. (1985). An investigation of transactions data for NYSE stocks. *The Journal of Finance*, **40,** number 3, 723–739.

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, **48,** number 2-3, 253–291.

Yang, J., Cabrera, J. and Wang, T. (2010). Nonlinearity, data-snooping, and stock index ETF return predictability. *European Journal of Operational Research*, **200,** number 2, 498–507.

Yao, Q. and Tong, H. (1998). A bootstrap detection for operational determinism. *Physica D: Nonlinear Phenomena*, **115,** number 1/2, 49–58.

Zeckhauser, R. and Thompson, M. (1970). Linear regression with non-normal error terms: A comment. *Review of Economics and Statistics*, **52,** 280–286.

Zhang, X., Creal, D., Koopman, S.J. and Lucas, A. (2011). Modeling dynamic volatilities and correlations under skewness and fat tails. Technical Report. Tinbergen Institute, Amsterdam. Working paper 11-078.

Zhu, D. and Galbraith, J.W. (2010). A generalized asymmetric student-t distribution with application to financial econometrics. *Journal of Econometrics*, **157,** 297–305.

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, **148,** 86–99.

Summary

This thesis, titled "Dependence in Financial and High-dimensional Time Series", contains three independent papers in Chapter 2–4, which combine state of the art techniques with research topics with increasing attention.

In Chapter 2, I propose an autocorrelation-based model to estimate and forecast yield curves. Each yield curve is treated as a continuous and smooth function. This chapter provides a full view of modeling, estimation, and forecasting of the yield curve in the framework of functional principal component analysis and lets the data determine the number of factors and the lag information used in the model.

Chapter 3 integrates more flexible distributions into the classic capital asset pricing model to better characterize return properties. This chapter provides the procedures for estimation and presents the economic significance of using this novel integration.

Intraday return patterns and predictability have drawn more and more attention recently. Chapter 4 sheds light on the dependence between the overnight return and the subsequent intraday return and examines trading opportunities and holding periods for day traders.

Samenvatting (Summary in Dutch)

Dit proefschrift, getiteld "Dependence in Financial and High-dimensional Time Series", bevat drie onafhankelijke artikelen in Hoofdstukken 2–4, die de nieuwste technieken combineren met onderzoeksonderwerpen die toenemende aandacht krijgen.

In Hoofdstuk 2 stel ik een op autocorrelatie gebaseerd model voor om rentecurves te schatten en te voorspellen. Elke rentecurve wordt behandeld als een continue en gladde functie. Dit hoofdstuk geeft een volledig beeld van het modelleren, schatten en voorspellen van de rentecurve in het kader van functionele principale componentenanalyse en laat de data het aantal factoren en de vertraagde informatie bepalen die in het model worden gebruikt.

Hoofdstuk 3 integreert meer flexibele verdelingen in het klassieke prijsmodel voor kapitaalgoederen om rendementseigenschappen beter te karakteriseren. Dit hoofdstuk geeft de schattingsprocedures en presenteert de economische significantie van het gebruik van deze nieuwe integratie.

Patronen in *intraday* rendementen en voorspelbaarheid krijgen de laatste tijd steeds meer aandacht. Hoofdstuk 4 beschijft de afhankelijkheid tussen het *overnight* rendement en het daaropvolgende *intraday* rendement en onderzoekt handelsmogelijkheden en *holding periods* voor daghandelaren.