

January 2017

Volatility Modelling with Applications to Equity and Foreign Exchange Markets

Sergii Pypko

The University of Western Ontario

Supervisor

Dr. Lars Stentoft

The University of Western Ontario

Joint Supervisor

Dr. Timothy Conley

The University of Western Ontario

Graduate Program in Economics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Sergii Pypko 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Econometrics Commons](#), and the [Finance Commons](#)

Recommended Citation

Pypko, Sergii, "Volatility Modelling with Applications to Equity and Foreign Exchange Markets" (2016). *Electronic Thesis and Dissertation Repository*. 4314.

<https://ir.lib.uwo.ca/etd/4314>

Abstract

My thesis consists of three chapters describing volatility forecasting during periods of financial booms and busts, the economic and statistical benefits of flexible data generating process of index returns, and multivariate model of exchange rate returns and their options.

In the first chapter, I propose a non-linear threshold model for realized volatility of S&P 500 index, allowing us to obtain a more accurate volatility forecast, especially during periods of financial crisis. The changes in volatility regimes are driven by negative past returns, where the threshold equals approximately -1% . This finding remains robust to different functional forms of volatility and different set of indices from both developing and developed countries. The additional flexibility of the model allows me to produce a more accurate one and multiple-days-ahead forecasts compared to the linear specification and GARCH family models. Finally, I derive an approximated closed form solution for multiple-step-ahead forecast, which is based on the normal-inverse Gaussian conditional distribution of returns.

In the second chapter, I develop a novel discrete-time model for the asset return based on the high-frequency data and mixture of normal (MN) distributions of the latent volatility. This model accurately replicates distributions of both returns and realized volatility under the objective measure. To compute option prices, I specify a Radon-Nikodym derivative, which includes both Gaussian and non-normal innovations, correspondingly. Crucially, my approach avoids calibration of all model's parameters. I price European Put options using Monte Carlo simulations and assess pricing performance of MN and nested Gaussian models during turbulent financial markets in 2008-2011 years. MN model does not only substantially reduce option pricing errors compared with Gaussian model, but provides an appealing econometric framework to assess evolution of investors' risk. Next, I

show a novel approach for predicting returns distribution by exploiting informational content of option prices and MN model. Finally, I build a simple quantitative strategy, which substantially outperforms returns of S&P 500 index (76% compared with 2%) during turbulent 2008-2011 years, while remaining market-neutral and had the same volatility as a benchmark returns.

In the third chapter, which is a joint work with Chang, Feunou and Fontaine, we propose a new multivariate factor model of exchange rate returns and their option-implied variances. This model documents a tight factor structure in the variance of exchange rate returns and then relate it to the economic factors. In particular, we show that the common factors driving variances of exchange rate returns include the variances of global factors and the common factors driving variances of country-specific shocks. We build a tractable multivariate asset pricing model based on these stylized facts for the underlying exchange rate returns. Our multivariate model provides a reasonable fit compared with performance of univariate models estimated for each series separately. Crucially, our model has a number of appealing benefits which are not attainable in the univariate framework. For example, this model can be used to devise a better portfolio construction or hedging strategy for a portfolio containing both currencies and currency options.

keywords: volatility modelling, derivatives pricing, risk management, forecasting

Co-Authorship Statement

This thesis contains Chapter 3 co-authored with Bo Young Chang, Bruno Feunou and Jean-Sébastien Fontaine.

Acknowledgements

I am very grateful to John Knight for his support, guidance and motivation. He was a person who inspired me to start my research.

I am grateful to my supervisors Lars Stentoft and Timothy Conley for their valuable comments and support during my studies. I want to thank Lars Stentoft for the opportunity to be engaged in the research area of derivatives pricing as his research assistant, which became the fundamental part of my dissertation. I also grateful to his comments and discussions, which were important factors for me to develop both as economist and econometrician.

I like to thank to my committee member Salvador Navarro for his support and suggestions. I am also thankful to my co-authors from the Bank of Canada, Bo Young Chang, Jean-Sébastien and Bruno Feunou, for the opportunity to work with them on my third chapter during my time at the Bank of Canada. I am grateful to Galyna Grynkviv and Miguel Cardoso for their comments and recommendations during writing of this dissertation. I would also to thank participants, referees and discussants for their insightful comments and recommendations at the number of conferences and workshops, including Midwest Finance Association 2015 Annual Meeting, 3rd Annual Doctoral Workshop in Applied Econometrics, Western University 2nd Financial Econometrics and Risk Management conference, Western University, Bank of Canada and 50th Canadian Economic Association.

Finally, I would like to thank my wife, Galyna and my son, Roman for their unconditional support, tremendous help and huge inspiration to me during the writing of this dissertation. I also want to thank my mother-in-law, Olga and father-in-law, Petro for taking care of my son, while I was working on my thesis. I also want to thank my parents, Anatolii and Alevtyna and my sister, Olena for their support. I dedicate this work to them.

Contents

Certificate of Examination	ii
Abstract	iii
Acknowledgement	vi
List of Tables	xii
List of Figures	xiv
List of Appendices	xvi
1 Volatility Forecast in Crises and Expansions	1
1.1 Introduction	1
1.2 Model	4
1.2.1 HAR-RV Model with Regime Switching	5
1.2.2 Econometric Framework for the Non-Linear Model	8
Estimation	8
Testing for Non-Linearity	10
Testing for Remaining Non-Linearity	11
Asymptotic Distribution of the Threshold Parameter	12

Stationarity	13
1.2.3 Forecasting	14
One-Step-Ahead Forecast	14
Conditional Distribution of Returns	15
1.3 Empirical Analysis	20
1.3.1 Data	20
1.3.2 Preliminary Data Analysis	20
1.3.3 Benchmark HAR Model	22
1.3.4 The TAR(2) Model	25
1.4 Forecast	28
1.4.1 One-Day-Ahead Forecast	29
1.4.2 Multiple-Step-Ahead Forecast	31
1.5 Conclusions	34
Bibliography	36
2 Uncovering a Data Generating Process of Returns with Application to the Derivatives' Pricing, Risk Management and Strategy Development	36
2.1 Introduction	36
2.2 Model	40
2.2.1 Model of DGP for returns under the objective measure P	40
2.2.2 Estimation of MN	45
2.2.3 Risk neutralization	47
2.2.4 Identification of ν_{1t} sequence	51
2.3 Empirical analysis	53
2.3.1 Data	53

	Returns and realized volatility data	53
	Option data	54
2.3.2	Estimation results	55
	Estimation of MN models	55
	Analysis of innovations	57
	MN(4) model simulations	59
2.4	Option valuation	61
2.4.1	Option pricing in the existent markets	63
2.4.2	Analysis of model-implied higher order moments	66
2.4.3	Option pricing in new markets	69
2.5	Forecasting volatility and returns distributions	70
2.5.1	Forecasting Value at Risk	71
2.5.2	Forecasting conditional mean of realized volatility	77
2.6	Simple quantitative trading strategy	79
2.7	Conclusion	82
	Bibliography	84
3	Global Factors and Common Idiosyncratic Variance	
	in Exchange Rates Volatility	84
3.1	Introduction	84
3.2	Global and idiosyncratic variance factors	86
3.2.1	Data	87
3.2.2	Common variance of country-specific Innovations	88
3.2.3	Dollar and carry conditional variance	88
	Common factors in other option-implied measures of risk	90

Robustness	92
3.3 No-arbitrage dynamic exchange rate model	93
3.3.1 Individual exchange rate dynamics	93
3.3.2 Exchange rate factor dynamics	94
3.4 Estimation	94
3.4.1 Benchmark univariate model	94
3.4.2 Exchange rate likelihood	95
3.4.3 Targeting unconditional moments	96
Factors mean	96
Country-specific variance	97
Factors variance	98
3.5 Results	98
3.5.1 Volatility dynamics	98
3.5.2 Persistence	104
3.5.3 Diagnostic checks	105
Bibliography	109
A Chapter 2 Appendix	109
A.1 m -step-ahead forecast	109
A.2 Proof of Theorem 1.2.1	110
A.3 Comparison of HAR and SETAR(2) models.	113
Bibliography	109
B Chapter 3 Appendix	114
B.1 Expectation Maximization algorithm	114

B.2	Proof of Theorem 2.2.1	117
B.3	Proof of Theorem 2.2.2	119
B.4	Proof of Theorem 2.2.3	121

List of Tables

1.1	Parameters of normal and inverse Gaussian distributions for standardized returns and volatility.	17
1.2	Descriptive statistics	21
1.3	Heterogeneous autoregressive model (HAR) estimation	23
1.4	Comparison of the TAR(1) (or HAR) and TAR(2) models	26
1.5	TAR(2) estimation	28
1.6	One-day-ahead out-of-sample forecast	29
1.7	Multiple-days-ahead out-of-sample forecast	33
2.1	Summary of the S&P 500 index options data	55
2.2	Estimations results	56
2.3	Comparison of actual and simulated moments for returns and realized volatility	59
2.4	Option valuation I	65
2.5	Option valuation II	69
2.6	Unconditional coverage property	74
2.7	Comparison of volatility forecasts	78
3.1	First component explains most of the total and country-specific variances	89
3.2	Factor Structure in Equity and Exchange Rate Markets	92

3.3	Comparison of multivariate and univariate variances	106
3.4	Ljung-Box test for remaining autocorrelation in squared exchange rate re- turns and standardized returns	107
A.1	Comparison of the TAR(1) (or HAR) and SETAR(2) models	113

List of Figures

1.1	Comparison of parametric and non-parametric distributions of standardized returns, realized volatility and returns	18
1.2	Time series dynamics of daily standardized returns, returns, realized variance, realized volatility and the logarithm of the realized variance	22
1.3	Sample autocorrelations and partial autocorrelations of returns and realized volatility	23
1.4	Comparison of actual realized volatility and model-implied volatility recovered from the HAR model	24
1.5	The dynamics of returns in two regimes	27
1.6	The confidence interval of threshold parameter	27
1.7	One-step-ahead forecast in 2008-2014	30
1.8	One-step-ahead forecast in 2008-2009	32
1.9	Multiple-step-ahead forecast in 2008-2014	34
2.1	Analysis of innovations	58
2.2	Comparison of actual and simulated returns	61
2.3	Comparison of actual and simulated realized volatility	62
2.4	Comparison of MN(4) and Gaussian model-implied moments of returns under the risk-neutral measure	67

2.5	Value at Risk for 10 days at 5%	75
2.6	Hit functions over 10 days at 5%	76
2.7	Trading strategy based on a risk neutral skewness	81
3.1	PCA1 of exchange rate returns variances, and variances of global and domestic factors	90
3.2	PCA1 of implied variances of exchange rate returns, and variances of global and domestic factors	91
3.3	Comparison of multivariate and univariate variances. Part I	100
3.4	Comparison of multivariate and univariate variances. Part II	101
3.5	Comparison of multivariate and univariate variances. Part III	102
3.6	Comparison of multivariate and univariate variances. Part IV	103
3.7	Share of individual country-specific variance attributed to common factor	104
3.8	The link between univariate model's persistence and correlation between model-implied and univariate variances	105

List of Appendices

A Chapter 2 Appendix	109
B Chapter 3 Appendix	114

Chapter 1

Volatility Forecast in Crises and Expansions

1.1 Introduction

Volatility plays an important role in financial econometrics. Measuring, modelling and forecasting financial volatility are essential for risk management purposes, portfolio allocation and option pricing. Although returns remain unpredictable, their second moment can be forecasted quite accurately, which generated a lot of research during the last thirty years motivated by Engle's seminal paper Engle (1982). The existing literature aiming to model and forecast financial volatility can be divided into two distinct groups: parametric and non-parametric models. The former assumes a specific functional form for volatility and models it as a function of observable variables, such as ARCH or GARCH models Engle (1982), Bollerslev (1986), Bollerslev et al. (1994), or as a known function of latent variables resulting in stochastic volatility models Hull and White (1987), Melino and

Turnbull (1990).

The second class defines financial volatility without imposing any parametric assumptions hence called realized volatility models Andersen et al. (2003). The main idea of the latter models is to construct consistent estimators for the unobserved integrated volatility by summing the squared returns over a very short period within a fixed time span, typically one day. The availability of high-frequency data allows high precision estimation of the continuous time pure diffusion processes given the large datasets of discrete observations. As a result, volatility essentially becomes observable and, in the absence of microstructure noise, can be consistently estimated by a realized volatility measure. This approach has two main benefits compared with GARCH and stochastic volatility models. First, researchers can treat volatility as observable and model it by applying a time series technique, for example ARFIMA or autoregressive fractionally integrated moving average models Andersen et al. (2003). Second, realized volatility models significantly outperform models based on lower frequency (daily data) in terms of forecasting power; see, e.g., Maheu and McCurdy (2011), Andersen et al. (2007), McAleer and Medeiros (2008). Indeed, the latter models adapt new information and update the volatility forecast at a slower daily frequency, while the former models can incorporate changes in volatility faster due to the more frequent arrival of intraday information.

Although the literature proposes many different approaches for modelling volatility, there is still no unique model that explains all of the stylized facts simultaneously. In particular, there is no consensus on how to model long memory, since there are at least four approaches: the non-linear model with regime switching McAleer and Medeiros (2008); the linear fractionally-integrated process Andersen et al. (2001); the mixture of heterogeneous run information arrivals Andersen and Bollerslev (1997); and the aggregation of

short memory stationary series Granger and Ding (1996). Numerous methods have been developed, since it is hard to distinguish between unit root and structural break data generating processes Perron (1989), Zivot and Andrews (1992). Choi et al. (2010) show that structural break models can outperform the long memory model if the timing and sizes of future breaks are known. Although few academics and practitioners accurately predicted the timing of the recent financial crises and European sovereign debt turmoil, a model with structural breaks seems to be more economically plausible than a fractionally-integrated long memory model. In addition, Choi et al. (2010) recommend relying on economic intuition to choose between smooth transition auto regressive models (STAR) and abrupt structural break models.

In this chapter, we extend the heterogeneous autoregressive model proposed by Corsi (2009) to take into account different regimes of volatility. The resulting model is called a non-linear threshold autoregression model, where regimes are governed by an exogenous trigger variable. This model provides a better fit of the robust measure of realized volatility for both in-sample data and out-of-sample forecasting. In addition to an improved performance in particular samples, a non-linear model also produces superior multiple-step-ahead forecasts in population according to the Giacomini and White test (Giacomini and White (2006)). We also show that the superior performance of a non-linear model is achieved during periods of high volatility. This is especially important during times of financial crises, when investors are in particular need of more accurate forecasts. Finally, we derive an approximated closed form expression for multiple-step-ahead forecast, where the past returns govern changes in volatility regimes.

This chapter finds that changes in the volatility regimes occur when return exceeds

a -1% threshold, which is in line with previous findings McAleer and Medeiros (2008), Scharth and Medeiros (2009). However, our model differs in terms of the estimation procedure and the most recent dataset that includes financial crises. In fact, the superior performance of a non-linear model becomes particularly significant during periods of elevated volatility, such as recent financial crises. More importantly, we derive an approximated closed-form expression of multiple-step-ahead forecasts, whereas other authors either focus on one-step ahead forecasts McAleer and Medeiros (2008) or using conditional simulations Scharth and Medeiros (2009).

The remainder of this chapter is organized as follows. The non-linear threshold model for realized volatility is defined in Section 1.2. Section 1.3 describes preliminary data analysis and estimation results for the S&P 500 index. Section 1.4 describes one and multiple-step-ahead forecasts. Finally, Section 1.5 concludes and provides directions for future work.

1.2 Model

In this section, we introduce two building blocks: the heterogeneous autoregressive model and the regime switching model. Then, we describe the econometric framework designed for the estimation and inference of our threshold autoregressive model. Finally, we discuss the forecasting of our model and how to derive an approximated closed form expression for its multiple-days-ahead forecasts.

1.2.1 HAR-RV Model with Regime Switching

In this section, we discuss extensions of the heterogeneous autoregressive model (HAR) of realized volatility proposed in Corsi (2009). First, let us assume that returns follow a continuous diffusion process:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \quad (1.1)$$

where $p(t)$ is the logarithm of instantaneous price, $\mu(t)$ is continuous with a finite variation mean process, $\sigma(t)$ is instantaneous volatility and $W(t)$ is standard Brownian motion. Given the process in (1.1), the integrated variance corresponding to day t is defined as:

$$IV_t^d = \int_{t-1}^t \sigma^2(\omega)d\omega. \quad (1.2)$$

Several authors show that as sampling frequency increases, integrated volatility IV_t^d can be approximated by realized variance defined as a sum of the intraday squared returns Andersen et al. (2003), Barndorff-Nielsen and Shephard (2002a,b). In essence, volatility becomes observable and can be forecasted using time series techniques.

The presence of market microstructure noise makes realized variance inconsistent and is a biased estimator of true integrated volatility. Therefore, we use the realized kernel estimator developed in Barndorff-Nielsen et al. (2008), which remains consistent under the presence of market microstructure noise. The realized kernel $RK_{K,\delta}$ is an estimator of latent realized variance and is defined as follows:

$$RK_{K,\delta} = \gamma_0(p_t) + \sum_{h=1}^H k\left(\frac{h-1}{H}\right)(\gamma_h(p_t) + \gamma_{-h}(p_t)), \quad (1.3)$$

where $\gamma_h(p_t) = \sum_{i=1}^{n(\delta)} (p_{i,t} - p_{i-1,t})(p_{i-h,t} - p_{i-h-1,t})$, $k(\cdot)$ is a weight function and $p_{i,t}$ is i -th intra-daily log price sampled at frequency δ and recorded at day t . In other words, $i = 1, \dots, n(\delta)$ and $n(\delta) = n_{seconds}/\delta$, where $n_{seconds}$ is the number of seconds during the trading day. Thus, the realized kernel is similar to the HAC (heteroskedasticity and autocorrelation consistent covariance matrix) estimator of the variance-covariance matrix for some stationary time series. Throughout this chapter, realized variance will equal the realized kernel measure defined in (1.3).

The realized kernel has several advantages over other high-frequency proxies of latent volatility. First, Brownlees and Gallo (2009) show that the realized kernel performs better (in terms of forecasting Value-at-Risk) than other high-frequency measures, including realized volatility, bi-power realized volatility, two-scales realized volatility and daily range. Second, the realized kernel is a consistent estimator of latent variance, which is robust to the market microstructure noise.

The heterogeneous autoregressive model is able to replicate the majority of stylized facts observed in data: fat tails, volatility clustering and long memory. In particular, HAR is able to generate hyperbolic decays in the autocorrelation function in a parsimonious way due to the volatility cascade property, despite the fact that this model does not belong to the class of long memory models. This model is based on the heterogeneous market hypothesis Muller et al. (14-15 October, 1993), which implies that lower frequency volatility (weekly) affects higher frequency volatility (daily), but not *vice versa*:

$$RV_{t+1}^d = c + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \epsilon_{t+1}^d, \quad (1.4)$$

where RV_t^d , RV_t^w and RV_t^m are daily, weekly and monthly realized variance, respectively, at

period t . The lower frequency, for example weekly, realized variance is computed as:

$$RV_t^w = \frac{RV_t^d + \dots + RV_{t-4}^d}{5}. \quad (1.5)$$

Similarly, the monthly realized variance is computed as the average of daily variances over 22 days. Although the HAR model is able to capture long memory and volatility clustering, it cannot explain abrupt changes in regimes. Indeed, recent subprime mortgage crises, European debt turmoil and a number of other financial calamities led to significantly different behaviour in the dynamics of the realized variance during “good” and “bad” times, as we will discuss in Section 1.3. Therefore, we propose to extend the benchmark HAR model and allow the possibility of multiple regimes, governed by either endogenous or exogenous variables. We define the threshold HAR model with two regimes as follows:

$$RV_{t+1}^d = \begin{cases} c_1 + \beta_1^d RV_t^d + \beta_1^w RV_t^w + \beta_1^m RV_t^m + \epsilon_{t+1}, & \text{if } T_{t-l} < \tau \\ c_2 + \beta_2^d RV_t^d + \beta_2^w RV_t^w + \beta_2^m RV_t^m + \epsilon_{t+1}, & \text{if } T_{t-l} \geq \tau \end{cases}, \quad (1.6)$$

where T_{t-l} is a trigger variable with some lag l and τ is the value of a threshold. Recall, that the threshold model has a very flexible structure and can be defined with different candidates for the trigger variable. However, we consider only observable trigger in this chapter based on the empirical analysis discussed in the subsection 1.3.3. In particular, we build a model with past returns governing the changes in the variance regimes.

1.2.2 Econometric Framework for the Non-Linear Model

Estimation

Next, we present the econometric techniques designed to model non-linear dynamics of time series: the self-exciting threshold autoregressive (SETAR) model and the threshold autoregressive (TAR) model introduced by Tong (1978) and Tong and Lim (1980). The main difference between these models is that the trigger variable can be either exogenous (TAR model) or endogenous (SETAR model). The TAR(m) model, where m denotes the number of regimes, is defined as follows:

$$Y_{t+1} = \theta'_1 X_t \mathbb{1}_{1,t}(\tau, l) + \dots + \theta'_m X_t \mathbb{1}_{m,t}(\tau, l) + \epsilon_{t+1}, \quad (1.7)$$

where Y_{t+1} is a univariate time series, $X_t = (1, Y_t, \dots, Y_{t-p})'$ ($p+1$) \times 1 vector, $\tau = (\tau_1, \dots, \tau_{m-1})$ and $\tau_1 < \tau_2 < \dots < \tau_{m-1}$, $\mathbb{1}_{j,t}(\tau, l) = \mathbb{1}(\tau_{j-1} \leq T_{t-l} < \tau_j)$, $\mathbb{1}(\cdot)$ is an indicator function and T_{t-l} is a threshold variable. Let us assume that $\tau_0 = -\infty$ and $\tau_m = \infty$, while the error term ϵ_{t+1} is conditionally independent on information set I_t and has a finite second moment:

$$E[\epsilon_{t+1}^2] = \sigma^2 < \infty \quad (1.8)$$

$$E[\epsilon_{t+1}|I_t] = 0.$$

In particular, if variable Y_{t+1} follows the TAR(2) process, then the model (1.7) becomes:

$$Y_{t+1} = \begin{cases} \theta'_1 X_t + \epsilon_{t+1}, & \text{if } T_{t-l} < \tau \\ \theta'_2 X_t + \epsilon_{t+1}, & \text{if } T_{t-l} \geq \tau \end{cases}. \quad (1.9)$$

Recall that Model (1.9) nests a non-linear HAR specification (1.6) if we put constraints

on the corresponding AR(22) model in each regime. Now, define the vector of all parameters of Model (1.9) as $\theta = (\theta'_1, \theta'_2, \dots, \theta'_m, \tau', l)'$. Under Assumption (1.8), the estimation of the TAR(m) model is performed using a non-linear least squares approach:

$$\hat{\theta} = \arg \min_{\theta} \sum_{t=1}^T (Y_{t+1} - \theta'_1 X_t \mathbb{1}_{1,t}(\tau, l) - \dots - \theta'_m X_t \mathbb{1}_{m,t}(\tau, l))^2. \quad (1.10)$$

Here, the minimization can be done sequentially. In particular, $\theta = (\theta'_1, \dots, \theta'_m)'$ can be computed through OLS regression of Y on $X(\tau, l)$ for fixed parameters d and τ :

$$\theta(\tau, l) = \left(X(\tau, l)' X(\tau, l) \right)^{-1} X(\tau, l)' Y, \quad (1.11)$$

where Y is the $T \times 1$ vector consisting of observations of Y_{t+1} , while $X(\tau, l)$ is the $T \times 4m$ matrix with t -th row $X_t(\tau, l)$:

$$X_t(\tau, l) = (X_t \mathbb{1}_{1,t}(\tau, l), X_t \mathbb{1}_{2,t}(\tau, l), \dots, X_t \mathbb{1}_{m,t}(\tau, l))$$

Now, let us assume for simplicity that the non-linear model has only two regimes or $m = 2$. Thus, two parameters τ and l can be estimated through minimization of the residual sum of squared errors $S(\tau, l)$:

$$(\hat{\tau}, \hat{l}) = \arg \min_{\tau, l} S(\tau, l), \quad (1.12)$$

where $S(\tau, l) = \left(Y - X(\tau, l) \hat{\theta}(\tau, l) \right)' \left(Y - X(\tau, l) \hat{\theta}(\tau, l) \right)$.

The minimization can be performed through a grid search, while noting that l is discrete. We follow Hansen (1999) approach, which allows speeding up the minimization algorithm.

In particular, he recommends eliminating the smallest and largest quantiles for the threshold variable in the grid search. This elimination does not only reduce the computational time, but also serves as a necessary condition for having enough observations in each regime. Indeed, asymptotic theory places additional constraints on the optimal threshold level, such that $\frac{n_j}{T} \geq \tau$ as $n \rightarrow \infty$. Although, there is no clear procedure for how to optimally choose τ , Hansen (1999) recommends to use a 10% quantile for the cut-off procedure.

Testing for Non-Linearity

We start by discussing the testing of the linear model or TAR(1) against the non-linear model or TAR(m), where $m > 1$. Under the null hypothesis, all parameters $\theta_1, \dots, \theta_m$ should be the same:

$$\theta_1 = \theta_2 = \dots = \theta_m. \quad (1.13)$$

Since the threshold parameter is not identified under the null hypothesis, the classical tests have a non-standard distribution. This problem is called ‘‘Davies’ problem’’ due to Davies (1977, 1987). Hansen (1999, 2000) overcomes this problem by using empirical process theory and derived the limiting distribution of the main statistics of interest F_{jk} :

$$F_{jk} = T \left(\frac{S_j - S_k}{S_k} \right), \quad (1.14)$$

where S_j and S_k are the sum of squared residuals and $k > j$. Computation of the asymptotic distribution is not straightforward, but might be faster than a bootstrap calculation. Although the literature does not assess the performance of the asymptotic against the bootstrap distribution in the context of SETAR models, Diebold and Chen (1996) show that the bootstrap technique performs better in the AR(1) context with Andrews structural change

test Andrews (1993). Thus, we use the following bootstrap algorithm for testing the linear model against the non-linear TAR(2) model:

1. Draw residuals with replacement from the linear TAR(1) model.
2. Generate a recursively simulated dataset using initial conditions Y_0, \dots, Y_p and estimates of the TAR(1) model, where p equals 22.
3. Estimate the TAR(1) and TAR(2) models on the simulated dataset.
4. Compute S_1^b and S_2^b on the simulated dataset, where b refers to specific bootstrap replication.
5. Compute statistics F_{12}^b from (1.14).
6. Repeat Steps (1)–(5) a large number of times.
7. The bootstrap p -value ($p_{bootstrap}$) equals the percentage of times that F_{12}^b exceeds the actual statistic F_{12} .

The algorithm in (1)–(7) can be used to evaluate the distribution of F_{12} under the assumption of either homoscedastic or heteroscedastic errors. We compute the bootstrap p -value under the latter assumption, since the residuals of Model (1.4) are heteroscedastic. This is in line with the literature Corsi et al. (2008).

Testing for Remaining Non-Linearity

The testing for remaining non-linearity is an important diagnostic check for the TAR (m) model. One way to address this question is to test whether the presence of the additional regime is statistically significant or not. This test relies on the aforementioned algorithm, while the bootstrap p -value is computed for statistics F_{jj+1} , where $j > 1$.

Asymptotic Distribution of the Threshold Parameter

The existing literature documents that the distribution of the parameter τ is non-standard if the threshold effect is significant Chan (1993), Hansen (1999). Hansen (1997, 2000) derives an asymptotic distribution of likelihood ratio statistics:

$$LR_1(\tau) = \frac{S_1(\tau) - S_1(\hat{\tau})}{\hat{\sigma}^2}, \quad (1.15)$$

where $S_1(\tau)$ is the residual sum of squares given parameter τ and $\hat{\sigma}^2$ is the variance of residuals of the TAR(2) model and equals $\frac{S_1(\hat{\tau})}{T-4}$. Moreover, Hansen (1997, 2000) shows that the confidence interval for the threshold parameter is obtained by inverting the distribution function of a limiting random variable. In other words, the null hypothesis $H_0 : \tau = \tau_0$ is rejected if the likelihood ratio $LR_1(\tau_0)$ exceeds the function of confidence level α :

$$c(\alpha) = -2\log(1 - \sqrt{1 - \alpha}). \quad (1.16)$$

Alternatively, the confidence interval for the threshold parameter is formed as an area where $LR_1(\tau) \leq c(\alpha)$ and is called the “no-rejection region”. We have to interpret the confidence interval for threshold parameter τ with caution, since it is typically conservative Hansen (1999, 2000). However, the ultimate test of our non-linear model is the ability to produce superior out-of-sample forecasts, which requires a tight confidence interval for the threshold parameter. We provide more discussion on page 27.

Although estimates $\hat{\theta}_1, \dots, \hat{\theta}_m$ depend on the threshold parameter τ , the asymptotic distribution remains the same as in the linear model case, since estimate $\hat{\tau}$ is super-consistent Franses and Dijk (2000). Chan (1993) and Hansen (1999) prove that dependency on the

threshold parameter is not of first order asymptotic importance, thus the confidence interval for $\hat{\theta}$ can be constructed as if $\hat{\tau}$ is a known parameter.

Stationarity

The stationarity conditions for our TAR(2) model are not easily derived, and in general, not much is known about this property for non-linear models with heteroskedastic errors – see the discussion in Franses and Dijk (2000)(pp. 79-80). The literature does propose sufficient conditions for a restricted class of non-linear models and typically for models with homoscedastic errors. In particular, Chan et al. (1985) consider SETAR(2) specification with the AR(1) model in both regimes, while Knight and Satchell (2011) establish necessary and sufficient conditions for the existence of a stationary distribution for TAR(2) and SETAR(2) models with the AR(1) process.

In contrast, our model has a richer structure within each regime, since the HAR model is a restricted version of the AR(22) process. Because of this richer structure within each regime and because neither self-exciting nor exogenous thresholds are used, it is not possible to use the results from Chan et al. (1985) and Knight and Satchell (2011) to prove stationarity. In addition, our residuals exhibit volatility clustering, and because of the heteroscedastic errors, it is not possible to exploit the necessary and sufficient conditions for strict stationarity, even for the simple HAR model derived by McAleer and Medeiros (2008).

In conclusion, as is the case in much empirical work, we have to make a trade-off between the flexibility of the model and the analytical tractability of stationarity conditions. In this chapter, we choose to design a model aiming at providing more accurate volatility forecasts, and we leave the question of stationarity for future work.

1.2.3 Forecasting

One-Step-Ahead Forecast

We assess the forecasting performance of various models by computing the one-step-ahead forecast of the realized volatility measured by the square root of the realized kernel. These forecasts are computed through rolling window estimation. First, the parameters of the model are estimated using an in-sample set, and then the one-step-ahead forecast is computed. Second, the rolling window is moved by one period ahead; the most distant observation is dropped, and the parameters of the model are re-estimated, while the threshold parameter τ and optimal lag l are kept time invariant. Finally, the one-step-ahead forecast is computed again.

We use the root mean square error (RMSE) and the mean absolute error (MAE) to compare the forecast performance of four models:

$$\begin{aligned}
 e_{t+1|t} &= Y_{t+1} - Y_{t+1|t} \\
 RMSE &= \sqrt{\frac{\sum_{j=t+1}^{t+N} e_{j+1|j}^2}{N}} \\
 MAE &= \frac{\sum_{j=t+1}^{t+N} |e_{j+1|j}|}{N},
 \end{aligned} \tag{1.17}$$

where $Y_{t+1|t}$ is the one-step-ahead conditional forecast of the daily realized volatility computed based on the rolling window for one of the four models and Y_{t+1} is the daily realized volatility at period $t + 1$. In addition, we compute R^2 of the following Mincer–Zarnowitz regression:

$$Y_{t+1} = d_0 + d_1 Y_{t+1|t} + v_t. \tag{1.18}$$

Finally, we investigate the forecasting performance of different models in population using the Giacomini and White (GW) test (Giacomini and White (2006)). The GW test fits nicely in our framework due to the following reasons. First, it does not favour models that overfit in-sample, but have high estimation errors. Second, this test is designed to compare not only unconditional, but conditional forecasts, as well. Finally, the GW test works with rolling window forecasts, where in-sample size is fixed, while out-of-sample size is growing.

Conditional Distribution of Returns

In this section, we discuss multiple-step-ahead forecasts for aggregate volatility over periods of five and 10 days. The extension of the multiple-step-ahead forecast to the linear model is straightforward, while the non-linear model has one important problem. We describe formulas used to compute the multiple-step-ahead forecast for the HAR, GARCH(1,1) and GJR-GARCH(1,1) (proposed by Glosten et al. (1993)) models in Appendix A.1. In particular, the one-step-ahead forecast remains the same for both non-linear and linear cases, while the two-step-ahead forecast is different:

$$\begin{aligned}
 Y_{t+1} &= F(Y_t, \theta) + \epsilon_{t+1} \\
 E[Y_{t+1}|I_t] &= F(Y_t, \theta) \\
 E[Y_{t+2}|I_t] &= E[F(Y_{t+1}, \theta)|I_t] \neq F(E[Y_{t+1}, \theta]|I_t),
 \end{aligned}
 \tag{1.19}$$

where I_t is the information set available at period t , F is a non-linear function, θ is a vector of estimates and Y_t is the realized volatility at period t . Equation (1.19) illustrates the main problem related to non-linear model: the expected value of a non-linear function differs from the value of a non-linear function evaluated at the expected value. In the

literature, several methods have been proposed for the computation of the multiple-step-ahead forecast, including conditional simulations in Scharth and Medeiros (2009). However, we choose a different strategy and derive an approximated closed form solution for the multiple-step forecast. Specifically, we follow an approach similar to Forsberg and Bollerslev (2002) and Stentoft (2008) to derive the conditional distribution of returns. Given the diffusion process (1.1), the returns should follow a normal distribution:

$$\begin{aligned} r_{t+1}|Y_{t+1}, I_t &\sim N(\mu_N Y_{t+1}, \sigma_N^2 Y_{t+1}^2) \\ Y_{t+1} &= \sqrt{RV_{t+1}^d}, \end{aligned} \quad (1.20)$$

where $I_t = \mathcal{F}(r_t, r_{t-1}, \dots)$ is information at the period t set generated by the history of returns and μ_N is the mean of standardized returns, and μ_N and σ_N^2 should be close to zero and one, correspondingly. See Table 1.1 for details. Meanwhile, the conditional distribution of realized volatility is closely approximated by the inverse Gaussian distribution with the following density function:

$$\begin{aligned} Y_{t+1}|I_t &\sim IG(\sigma_{t+1}, \alpha_{IG}), \\ pdf_{IG}(z, \sigma_{t+1}, \alpha_{IG}) &= \frac{\left(\frac{1}{\alpha_{IG}\sigma_{t+1}}\right)^{-0.5} z^{-1.5}}{(2\pi)^{0.5}} \exp\left(\alpha_{IG} - 0.5 \left[\frac{\alpha_{IG}\sigma_{t+1}}{z} + \frac{\alpha_{IG}z}{\sigma_{t+1}}\right]\right) \\ \alpha_{IG} &= \frac{\lambda_{IG}}{\sigma_{t+1}}, \end{aligned} \quad (1.21)$$

where σ_{t+1} is a conditional mean and λ_{IG} is a shape parameter of the inverse Gaussian distribution. The conditional mean is assumed to be filtered from the non-linear TAR(2) model as follows:

$$\sigma_{t+1} = \mathbb{1}(r_t < \tau) \cdot X_t' \theta_1 + \mathbb{1}(r_t \geq \tau) \cdot X_t' \theta_2. \quad (1.22)$$

Table 1.1: Parameters of normal and inverse Gaussian distributions for standardized returns and volatility.

Parameters	All Sample	In-Sample
μ_N	0.0840	0.0488
σ_N	1.0907	1.0937
μ_{IG}	0.0093	0.0087
λ_{IG}	0.0296	0.0369

The first column corresponds to the in-sample period, while the second to the whole sample. μ_{IG} is a scale parameter for the unconditional inverse Gaussian (IG) distribution of realized volatility.

Combining Equations (1.20) and (1.21), the conditional distribution of returns becomes a normal-inverse Gaussian distribution (NIG) with the probability density function computed as:

$$pdf(r_{t+1}|I_t) = \int_{Y_{t+1}} pdf(r_{t+1}|Y_{t+1}, I_t) \cdot pdf(Y_{t+1}|I_t) dY_{t+1} \quad (1.23)$$

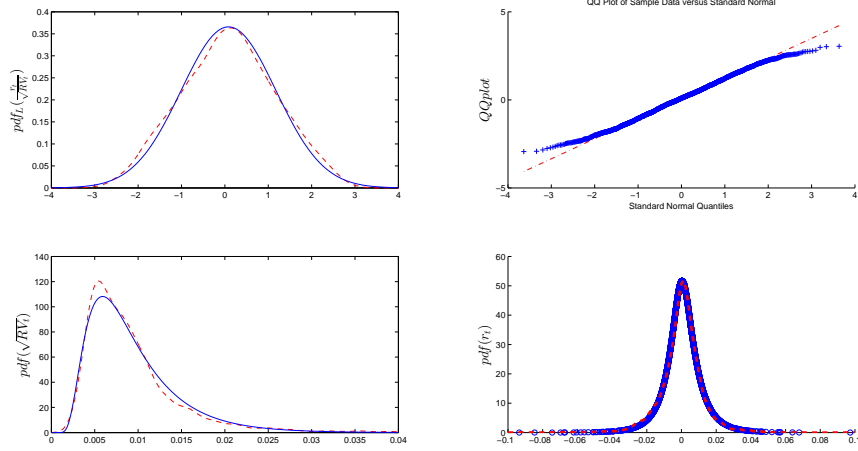
$$r_{t+1}|I_t \sim NIG(\mu_N, \sigma_N, \sigma_{t+1}, \alpha_{IG}).$$

The NIG distribution provides a relatively accurate fit of the unconditional distribution of returns. The first three graphs in Figure 1.1 demonstrate the very close match between parametric and non-parametric unconditional distributions of standardized returns and realized volatility, respectively. Table 1.1 shows the corresponding parameters of normal and inverse Gaussian distributions.

Having the distributional assumption for returns, Theorem 1.2.1 demonstrates how to obtain the approximated closed form expression for the multiple-step ahead forecast of the realized volatility.

Theorem 1.2.1 *Let $\{Y_t\}$ follow the TAR(2) process defined in (1.6), while returns follow*

Figure 1.1: Comparison of parametric and non-parametric distributions of standardized returns, realized volatility and returns



Comparison of parametric (solid blue line) and non-parametric kernel distributions (red dashed line) of standardized returns, realized volatility and returns. The first graph compares the normal distribution of standardized returns with the non-parametric distribution, while the second plots the corresponding QQ plot. The third graph illustrates the comparison between the IG distribution and the non-parametric distribution for realized volatility. The final graph shows the normal-inverse Gaussian (NIG) distribution for returns and the corresponding non-parametric distribution.

the NIG distribution with the conditional probability density function defined in (1.23), and r_{t+h-1} ($h \geq 2$) are independent of $\epsilon_1, \dots, \epsilon_{t+h-2}$. Then, the approximated h -step-ahead forecast ($h \geq 2$) is obtained as follows:

$$\begin{aligned} \hat{Y}_t(h) = E[Y_{t+h}|I_t] \approx & c_1\pi_t + c_2(1 - \pi_t) + (\beta_1^d\pi_t + \beta_2^d(1 - \pi_t))\hat{Y}_t(h-1) + \\ & + (\beta_1^w\pi_t + \beta_2^w(1 - \pi_t))\hat{Y}_t^w(h-1) + (\beta_1^m\pi_t + \beta_2^m(1 - \pi_t))\hat{Y}_t^m(h-1), \end{aligned} \quad (1.24)$$

where:

$$\begin{aligned}
\pi_t &= Pr[r_{t+1} < \tau | I_t] = \int_{-\infty}^{\tau} pdf(r_{t+1} | I_t) dr_{t+1} \\
\theta &= (c_1, \beta_1^d, \beta_1^w, \beta_1^m, c_2, \beta_2^d, \beta_2^w, \beta_2^m)' \\
\hat{Y}_t^w(h-1) &= \left[\frac{\hat{Y}_t(h-1) + \dots + \hat{Y}_t(h-5)}{5} \right] \\
\hat{Y}_t^m(h-1) &= \left[\frac{\hat{Y}_t(h-1) + \dots + \hat{Y}_t(h-22)}{22} \right],
\end{aligned} \tag{1.25}$$

Proof See Appendix A.2.

In essence, Formula (1.24) is similar to the multiple-step-ahead forecast of the GJR-GARCH(1,1) model — see Appendix A.1 for details. However, the TAR model has an additional flexibility, since probability π_t is time varying, while GJR-GARCH assumes that the corresponding probability equals to 0.5. To facilitate comparison between these two models, we compute the unconditional probability of a high volatility regime occurring based on the NIG distribution (1.23) and from returns data. Here, the probability equals the frequency of returns occurring, which is lower than the threshold value. The results show a close match between these two methods: 11.3% (NIG) vs. 13.2% (historical returns) for in-sample data.

Finally, we describe the multiple-step-ahead forecast using the rolling window approach. First, the parameters of the model are estimated using in-sample data, and probability π_t is computed. Second, multiple-step-ahead forecasts for the TAR model are calculated based on Expression (1.24), while π_t remains constant. Probability π_t can be computed for each step of forecast, as well, but this will add additional computational burden, while the results should change only marginally. In other words, we assume that $\pi_{t+h|t} = \pi_t \forall h$, where $\pi_{t+h|t} = Pr[r_{t+h} < \tau | I_t]$. We compute h -step-ahead forecasts for the HAR, GARCH(1,1)

and GJR-GARCH(1,1) models based on the formulas presented in Appendix A.1. Finally, the rolling window is moved by one period ahead; the first observation is dropped, and the parameters of the model, including π_{t+1} , are re-estimated.

1.3 Empirical Analysis

1.3.1 Data

The empirical analysis is based on high-frequency data for the S&P 500 index obtained through the Realized Library of Oxford-Man Institute of Quantitative Finance (Library Version 0.2), which is freely available:

“Researchers may use this library freely without restrictions so long as they quote in any work which uses it: Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009) “Oxford-Man Institute’s realized library”, Oxford-Man Institute, University of Oxford.”

The sample covers the period from 3 January of 2000 to 12 June of 2014, overall 3603 trading days. We exclude all days from the sample when the market was closed. Heber et al. (2009) have created the Realized Library database, which provides daily data for about 11 realized measures for 21 assets. The authors clean the raw data obtained through Reuters Data Scope Tick History and compute high-frequency estimators from cleaned data. We use a realized kernel Barndorff-Nielsen et al. (2008) as a proxy for integrated variance.

1.3.2 Preliminary Data Analysis

We start with data analysis of five main time series of interest: standardized returns, returns, realized variance, realized volatility and the logarithm of realized variance. Table 1.2

Table 1.2: Descriptive statistics

	$\frac{r_t}{\sqrt{RV_t}}$	r_t	RV_t	$\sqrt{RV_t}$	$\log(\sqrt{RV_t})$
Mean	0.08	8.0E-05	1.2E-04	9.3E-03	-9.65
Variance	1.19	1.5E-04	7.5E-08	3.8E-05	1.08
Skewness	-3.3E-03	-0.15	14.26	3.32	0.50
Kurtosis	2.57	10.24	381.25	24.58	3.47
D-F test	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.06$
Normality test (J-Btest)	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$
L-Btest 5 lags	$p = 0.01$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$
L-B test 10 lags	$p = 0.08$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$
L-B test 15 lags	$p = 0.07$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$
ARCH effect	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$	$p = 0.00$

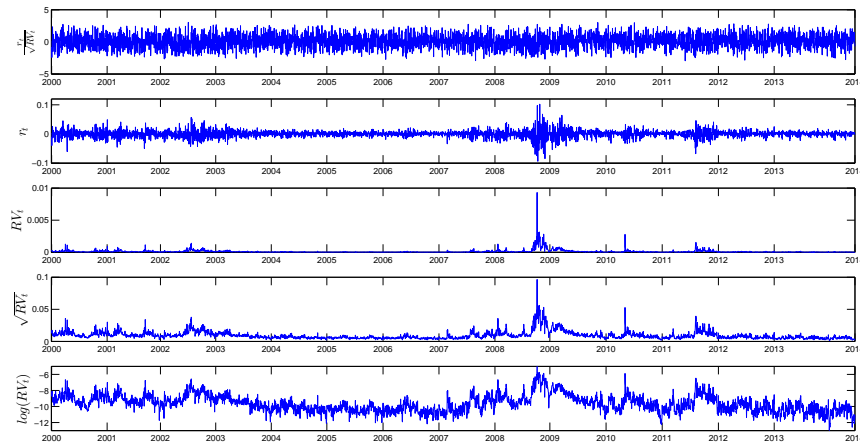
First four rows show unconditional sample mean, standard deviation, skewness and kurtosis of daily standardized returns, returns, realized variance, realized volatility and the logarithm of the realized variance of the S&P500 index. Remaining rows depict p-values obtained from Dickey-Fuller, JarqueBera, LjungBox and Engle ARCH tests for these series.

presents the descriptive statistics, while Figure 1.2 illustrates the time series dynamics of these variables.

Four of the variables are stationary at 5% according to the augmented Dickey–Fuller test, while $\log(\sqrt{RV_t})$ is stationary at 6%. The recent financial crises and European sovereign debt turmoil affected the volatility pattern and led to several spikes in the realized variance series. Although these spikes look less pronounced in the logarithm of realized variance, they remain very distinct from the volatility behaviour observed during calm times. This observation motivates the introduction of the regime switching model for volatility process.

Daily returns are weakly correlated and follow a leptokurtic and negative skewed distribution. By contrast, the distribution of the standardized returns is much closer to Gaussian, which is in line with previous empirical findings: Andersen et al. (2001, 2010). Figure 1.3 documents the long memory observed in realized volatility as the autocorrelation function

Figure 1.2: Time series dynamics of daily standardized returns, returns, realized variance, realized volatility and the logarithm of the realized variance



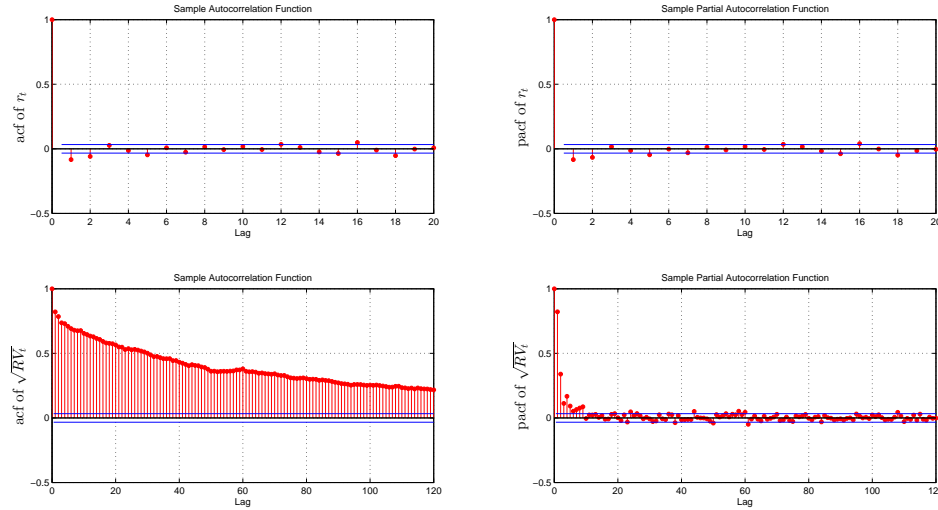
Daily standardized returns, returns, realized variance, realized volatility and the logarithm of the realized variance of the S&P500 index. The sample period goes from January 2000 till June 2014 (3603 observations).

decays at a hyperbolic rate. This result is also consistent with the literature: Andersen et al. (2003), Corsi et al. (2008), Choi et al. (2010).

1.3.3 Benchmark HAR Model

We start with the estimation of the benchmark linear Model (4) for the three specifications of dependent variable RV , \sqrt{RV} and $\log(\sqrt{RV})$, correspondingly. Table 1.3 presents the estimation results with the standard errors computed based on the HAC variance-covariance matrix. Despite relatively high R^2 for \sqrt{RV} and $\log(\sqrt{RV})$, the benchmark model fails to model spikes in volatility during turbulent times on financial markets. Figure 1.4 illustrates this point and depicts a comparison between the in-sample forecast and the actual realized kernel.

Figure 1.3: Sample autocorrelations and partial autocorrelations of returns and realized volatility



Sample autocorrelations and partial autocorrelations of returns and realized volatility of the S&P500 index. The sample period goes from January 2000 till June 2014 (3603 observations).

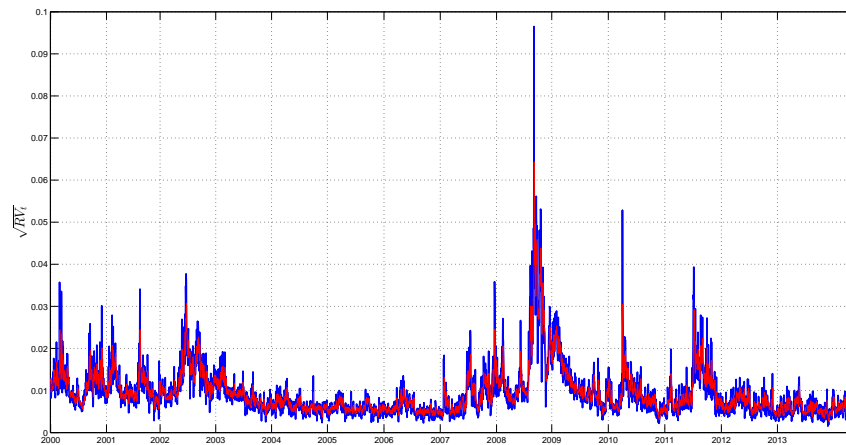
Table 1.3: Heterogeneous autoregressive model (HAR) estimation

	\mathbf{RV}_t		$\sqrt{\mathbf{RV}_t}$		$\mathbf{log} \sqrt{\mathbf{RV}_t}$	
	Estimate	SE	Estimate	SE	Estimate	SE
c	1.3E-05 ***	4.5E-06	4.6E-04 ***	2.0E-04	-0.44 ***	0.108
β^d	0.223	0.146	0.395 ***	0.058	0.336 ***	0.025
β^w	0.461 ***	0.165	0.384 ***	0.081	0.440 ***	0.036
β^m	0.216 ***	0.073	0.171 ***	0.048	0.178 ***	0.029
R^2	50.4%		72.6%		73.2%	

Reported are in-sample estimation results of the linear HAR model and corresponding standard errors computed based on the HAC variance-covariance matrix. The in-sample covers the period from February 2000 to June 2014 (3582 observations). Here, *** means that the corresponding p -value is lower than 0.01.

In particular, benchmark Model (1.24) underestimates volatility by around 40% during financial crises in 2007–2009. A similar pattern is observed during spikes in volatility in

Figure 1.4: Comparison of actual realized volatility and model-implied volatility recovered from the HAR model



In-sample comparison of actual realized volatility (blue line) and volatility recovered from the HAR model (red line). The in-sample covers the period from February 2000 to June 2014 (3582 observations).

2010 and 2011. One of the explanations of the poor performance of the HAR model during turbulent volatility periods is that it fails to take into account changes in volatility regimes. Indeed, if volatility reacts to negative returns more than to positive returns, then the arrival of the consequent negative shocks and volatility persistence can substantially increase the future volatility level. On the other hand, different economic regimes might affect volatility differently. We choose the TAR over SETAR model based on the higher value of the F_{12} statistics or, alternatively, the lower value of $p_{bootstrap}$ defined in Subsection 1.2.2.¹

¹See Appendix A.3 for details.

1.3.4 The TAR(2) Model

Next, we estimate the TAR(2) model (Table A.1 and Table 1.5), where past returns govern changes in the volatility regimes.

Table A.1 shows that regression R^2 improves substantially if regimes are driven by past returns. As a result, high values of the F_{12} statistics lead to the rejection of the null hypothesis (1.13) for all specifications at a 5% significance level. In addition, the optimal value of the threshold parameter remains the same for two specifications: RV_t and $\sqrt{RV_t}$. The τ that corresponds to logarithm specification is closely related to the second threshold of the TAR(3) model. However, the confidence interval for this parameter is very wide, which leads to the imprecise estimate of the threshold parameter. Not surprisingly, this model produces a less accurate one-step forecast than TAR(2). In particular, Dacco and Satchell (1999) document that the imprecise estimate of the threshold parameter leads to the poor forecasting performance of the simple switching model compared to the random walk model. In both cases, changes in regimes are driven not only by negative returns (leverage effect), but by significantly negative returns: -1.3% on a daily scale. McAleer and Medeiros (2008) also show that the transition between volatility regimes is governed not by negative past returns, but by “very bad news” or very negative past returns.

The fact that changes in regimes are triggered by “very negative returns” can be explained by the volatility persistence and higher intensity of shocks during bad times. Although the value of the threshold is not very large (it corresponds to the 11th percentile of the returns distribution), the increasing number of negative returns can generate a spike in the volatility. This explanation is similar to the option pricing literature, where researchers modelled volatility by adding infinite activity jumps to the return’s process Ornathanalai (2014). Even though the appearance of one small or medium jump is not enough to gener-

Table 1.4: Comparison of the TAR(1) (or HAR) and TAR(2) models

	\mathbf{RV}_t	$\sqrt{\mathbf{RV}_t}$	$\mathbf{log}(\mathbf{RV}_t)$
R^2 of TAR(1)	50.4%	72.6%	73.2%
R^2 of TAR(2)	58.0%	74.9%	74.7%
τ	-0.013	-0.013	0.001
l	0	0	0
F_{12}	649.6	318.3	214.0
$p_{bootstrap}$	0.00	0.03	0.00

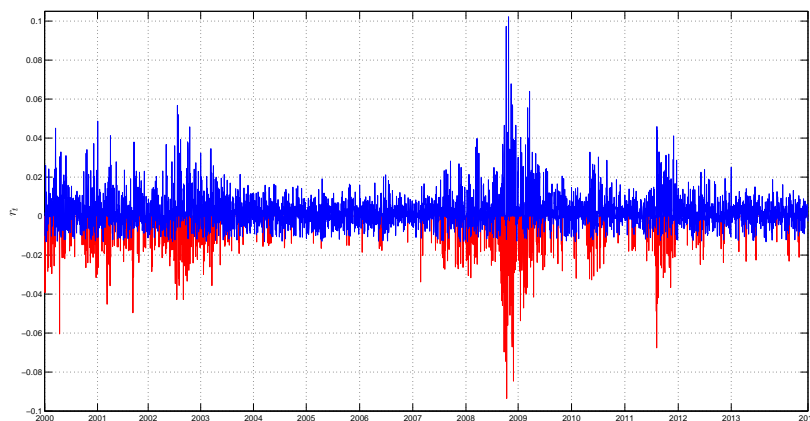
Reported are in-sample estimation results of the linear HAR model and non-linear TAR(2) model. The in-sample covers the period from February 2000 to June 2014 (3582 observations). $p_{bootstrap}$ is computed based on 500 replications using the heteroscedastic bootstrap method. We set the maximum amount of lags equal to 10 in the TAR estimation.

ate a significant surge in volatility, high volatility persistence can lead to pronounced spikes in the future volatility. Indeed, Figure 1.5 shows that the frequency of returns that are lower than the threshold (red line) increased dramatically during recent financial crises. By contrast, returns that exceed the threshold (blue line) completely dominated “very negative returns” during the period of low volatility in 2003–2007.

Table 1.5 shows that parameters β^d , β^w and β^m are very different in high- and low-volatility regimes. In particular, β_1^w is twice as large as the corresponding estimate in the low-volatility regime for $\sqrt{\mathbf{RV}_t}$ specification. Although some estimates have negative signs, they are not statistically significant at 10% for both realized volatility and variance models. By contrast, intercepts in both regimes are statistically negative for logarithmic specifications. Overall, corresponding estimates differ substantially in different regimes, which highlights the importance of using the regime switching model. Next, Figure 1.6 shows that the 95% confidence interval for the threshold parameter is quite narrow ($\tau_{opt} \in [-0.014, -0.012]$), although it includes two disjoint sets.

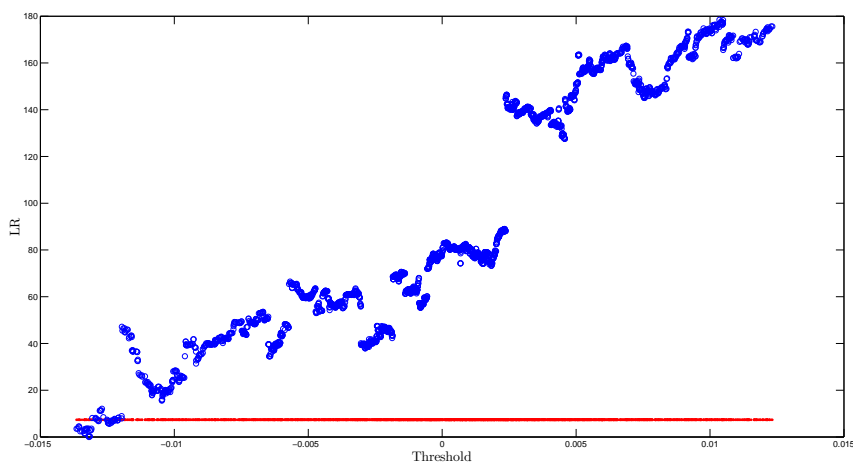
Finally, we compare the in-sample performance of the HAR and TAR(2) models for different indices, including both developing and developed countries: Bovespa (Brazil),

Figure 1.5: The dynamics of returns in two regimes



Daily returns in high (red line) and low (blue line) volatility regimes. The high (low) volatility regime occurs when the return is lower (higher) than the threshold. The sample period goes from February 2000 till June 2014 (3603 observations).

Figure 1.6: The confidence interval of threshold parameter



Ninety five percent confidence interval for the threshold parameter of the TAR(2) model with $\sqrt{RV_t}$ specification. The red line corresponds to $c(0.05) \approx 7$, while the blue points represent LR .

Table 1.5: TAR(2) estimation

	\mathbf{RV}_t		$\sqrt{\mathbf{RV}_t}$		$\mathbf{log} \sqrt{\mathbf{RV}_t}$	
	Estimate	SE	Estimate	SE	Estimate	SE
c_1	-2.6E-06	4.1E-05	-9.3E-05	9.5E-04	-0.321 ***	0.138
β_1^d	0.331 *	0.189	0.332 ***	0.085	0.347***	0.029
β_1^w	1.091 ***	0.372	0.811 ***	0.191	0.475 ***	0.045
β_1^m	-0.138	0.275	-0.018	0.128	0.133 ***	0.037
c_2	2.1E-05 ***	5.6E-06	0.001 ***	1.8E-04	-0.515***	0.150
β_2^d	0.182	0.156	0.340 ***	0.067	0.220 ***	0.038
β_2^w	0.260 *	0.139	0.317 ***	0.067	0.498***	0.050
β_2^m	0.268 ***	0.097	0.204 ***	0.045	0.243 ***	0.041
τ	-0.013		-0.013		0.001	
l	0		0		0	
R^2	58.0%		74.9%		74.7%	

Reported are in-sample estimation results of the non-linear TAR(2) model and corresponding standard errors computed based on the HAC variance-covariance matrix. The in-sample covers the period from February 2000 to June 2014 (3582 observations). The first four rows correspond to the high-volatility, while the last four rows correspond the low-volatility regime, respectively. Here, *** and * mean that the corresponding p -values are lower than 0.01 and 0.1, respectively.

DAX (Germany) and IPC Mexico (Mexico). The main findings remain robust to the different sets of indices: the non-linear model with an exogenous trigger is preferred over the corresponding specification with the endogenous variable.

1.4 Forecast

In this section, we discuss one- and multiple-step-ahead forecasts of realized volatility based on the TAR(2) model and several competing benchmarks. We assess their forecasting performance using low- and high-volatility periods.

Table 1.6: One-day-ahead out-of-sample forecast

	January 2008 to January 2009				July 2011 to December 2011				January 2008 to June 2014			
	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>
<i>RMSE</i>	7.0	0.96	0.78	0.85	4.9	0.96	0.72	0.73	3.8	0.98	0.77	0.82
<i>MAE</i>	4.1	0.97	0.67	0.76	3.6	0.95	0.63	0.66	2.3	0.99	0.67	0.71
R^2	0.70	0.68	0.56	0.64	0.42	0.38	0.24	0.39	0.75	0.74	0.66	0.70
<i>p_{GW}</i>	NA	0.54	0.00	0.00	NA	0.12	0.00	0.00	NA	0.71	0.00	0.00

The first four columns correspond to the period of recent financial crises in the U.S. from January 2008 to January 2009 (247 observations). The next four columns correspond to Eurozone crises from July 2011 to December 2011 (123 observations). The last four columns correspond to the period from January 2008 to June 2014 (1614 observations). The performance metrics are root mean square error (RMSE), mean absolute error (MAE), the R^2 of the Mincer–Zarnowitz regression and the p -value of the Giacomini and White test based on the MAE metric. Two forecasts are identical in population under the null hypothesis, while TAR beats its competitors under the alternative. We compare TAR against all other models, while NA corresponds to the TAR vs. TAR case. The TAR column represents the actual value of RMSE and MAE errors, while the HAR, GARCH and GJR columns, corresponding to the RMSE and MAE rows, equal the ratio of the TAR model to the following benchmark. Thus, a number below one indicates the improvement of the TAR model over its competitor. Observations for RMSE and MAE of the TAR model are standardized by 1000.

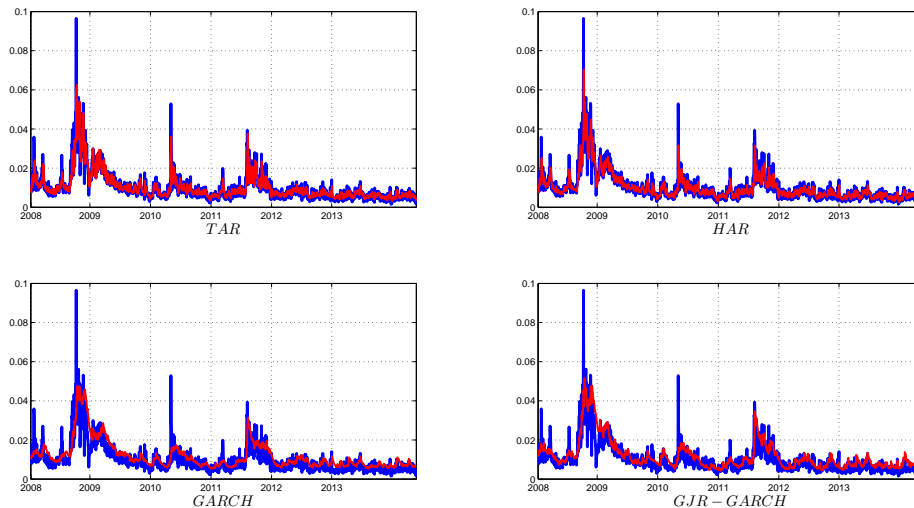
1.4.1 One-Day-Ahead Forecast

We start with the one-day-ahead forecast of the realized volatility, which is measured as the square root of the realized kernel. The in-sample period covers 1968 days from January 2000 to January 2008. In addition to the HAR model, we choose several GARCH specifications as benchmarks, including symmetric GARCH(1,1) and asymmetric GJR-GARCH(1,1). Hansen and Lunde (2005) show that it is extremely hard to outperform a simple GARCH (1,1) model in terms of forecasting ability. Meanwhile, TAR(2) is a non-linear model; therefore, we need to add asymmetric GARCH specification to guarantee a “fair” model comparison. Figure 1.7 and Table 1.6 assess the forecasting performance of high- and low-frequency models.²

Next, we investigate whether the TAR forecast remains superior in population or not

²Although realized volatility ignores overnight returns, the superior performance of the high-frequency models is unlikely to be affected.

Figure 1.7: One-step-ahead forecast in 2008-2014



Comparison of actual and one-day-ahead forecasts based on the TAR(2), HAR, GARCH(1,1) and GJR-GARCH(1,1) models from January 2008 to June 2014 (1614 observations). The red line indicates the one-step forecast, while the blue line the actual data.

using the Giacomini and White test. Recall that the GW test is designed for the situation where in-sample size is fixed, while out-of-sample size is growing. Thus, we assess the forecasting performance of different models using the GW test only for the period from January 2008 to June 2014 and not for U.S. and Eurozone financial crises. In the latter cases, the GW test is likely to perform poorly, since we have a relatively short period of sample periods: 247 and 123 observations, correspondingly.

The main results of this comparison are the following. First, high-frequency models significantly outperform lower frequency symmetric (GARCH) or asymmetric (GJR-GARCH) daily models. This result highlights the importance of more accurate volatility measuring based on the intra-daily data. Second, non-linear TAR(2) specification dominates the linear HAR model thanks to an additional flexibility to capture changes in regimes

according to the first three metrics. Surprisingly, TAR(2) does not outperform the HAR model according to the GW test.

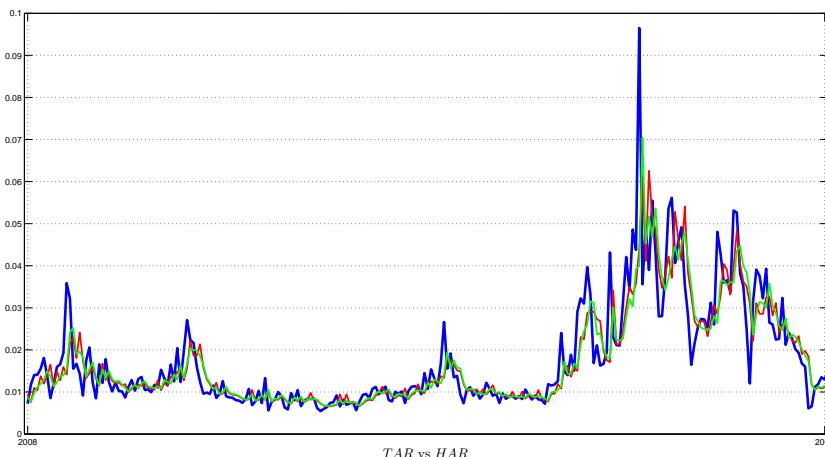
Finally, we assess the performance of volatility forecasts during times of financial turmoil: the U.S. financial crises in 2008 and the Eurozone crises in 2011. Although high-frequency models continue to dominate GARCH specifications, the benefits of using the non-linear TAR(2) model become substantial compared to linear specification: the latter's MAE is higher by 3% (U.S. crises) and 6% (Eurozone crises). By contrast, the MAE of the HAR model is only 1% higher during the whole out-of-sample period. Figure 1.8 shows that TAR(2) better captures spikes in volatility than linear specification during the recent U.S. financial crises. Finally, both RMSE and MAE are lower for Eurozone crises and whole out-of-sample periods compared with recent U.S. financial crises, which reflects the learning process of the model, where recent volatility spikes help to improve the models' performance.

To sum up, the benefits of using the non-linear TAR(2) model are most evident during periods of elevated volatility. In addition, the model is able to predict spikes in volatility, even when we use a relatively calm period for in-sample estimation, since changes in regimes are driven by moderately low returns. As a result, we do not rely on extreme market events to forecast volatility.

1.4.2 Multiple-Step-Ahead Forecast

This section describes multiple-step-ahead forecasts for aggregate volatility. Specifically, the object of interests is the h -step forecast of aggregate realized volatility $\sum_{i=1}^h Y_{t+i|t}$. Table 1.7 compares TAR(2) and other benchmark models during recent U.S. financial crises, Eurozone crisis and the out-of-sample period in 2008–2014. Figure 8 plots five-step-ahead

Figure 1.8: One-step-ahead forecast in 2008-2009



Comparison of actual and one-day-ahead forecasts based on the TAR(2) and HAR models during U.S. financial crises from January 2008 to January 2009 (247 observations). Red and green lines indicate one-step forecasts based on the TAR(2) and HAR models, correspondingly, while the blue line the actual data.

forecasts for all models.

The main findings remain similar to the one-step-ahead forecasts. First, high-frequency models continue to dominate daily models at the five and 10 days' ahead forecast. Second, TAR(2) performs better than the linear HAR model according to RMSE, MAE and R^2 . More importantly, the non-linear model outperforms linear specification, not only in a particular sample, but also in population: we reject the null hypothesis of the GW test that two forecast are identical at the 5% significance level. We based our conclusion on the results of the GW test for the 2008–2014 years to take into account the growing size of the out-of-sample dataset, as discussed in the Section 1.2.3. The GW test is based on the MAE metric. In addition, the U.S. financial crises have substantially higher RMSE and MAE compared with other periods, since periods of elevated volatility allow one to produce more accurate forecasts.

Table 1.7: Multiple-days-ahead out-of-sample forecast

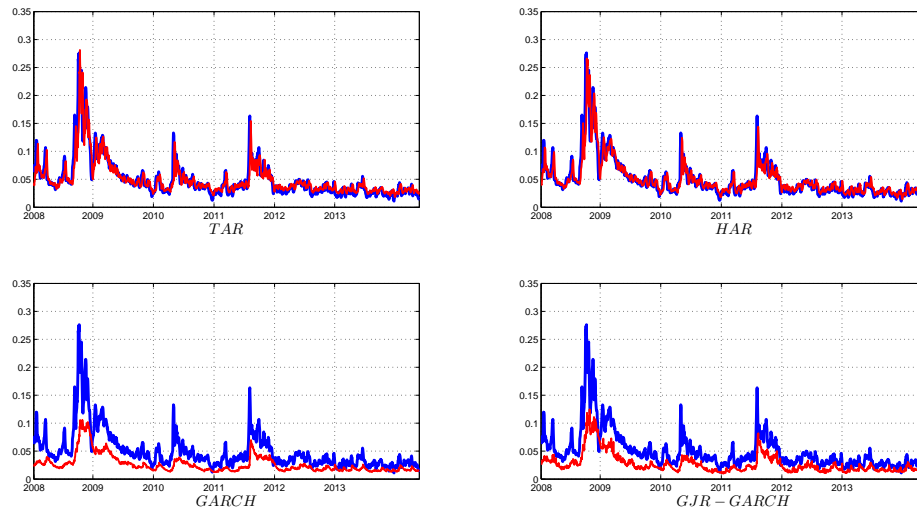
	January 2008 to January 2009				July 2011 to December 2011				January 2008 to June 2014			
	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>	<i>TAR</i>	<i>HAR</i>	<i>GARCH</i>	<i>GJR</i>
5-days-ahead forecast												
<i>RMSE</i>	33.5	0.98	0.56	0.55	23.1	0.99	0.60	0.59	17.3	0.99	0.53	0.55
<i>MAE</i>	22.2	0.98	0.49	0.47	14.5	0.96	0.46	0.45	10.1	0.98	0.43	0.44
R^2	0.67	0.67	0.61	0.64	0.27	0.26	0.16	0.20	0.76	0.75	0.71	0.75
p_{GW}	NA	0.13	0.00	0.00	NA	0.06	0.00	0.00	NA	0.03	0.00	0.00
10-days-ahead forecast												
<i>RMSE</i>	69.2	0.98	0.48	0.48	47.6	0.98	0.50	0.50	35.2	0.98	0.44	0.45
<i>MAE</i>	47.0	0.97	0.41	0.40	30.6	0.96	0.36	0.35	20.6	0.97	0.33	0.33
R^2	0.63	0.63	0.61	0.61	0.15	0.15	0.16	0.14	0.73	0.73	0.71	0.74
p_{GW}	NA	0.21	0.00	0.00	NA	0.31	0.00	0.00	NA	0.01	0.00	0.00

The first four columns correspond to the period of recent financial crises in the U.S. from January 2008 to January 2009 (247 observations). The next four columns correspond to Eurozone crises from July 2011 to December 2011 (123 observations). The last four columns correspond to the period from January 2008 to June 2014 (1604 observations). The performance metrics are the root mean square error (RMSE), mean absolute error (MAE), the R^2 of the Mincer–Zarnowitz regression and the p -value of the Giacomini and White test based on the MAE metric. Two forecasts are identical in population under the null hypothesis, while TAR beats its competitors under the alternative. The TAR column represents the actual value of RMSE and MAE errors, while the HAR, GARCH and GJR columns, corresponding to the RMSE and MAE rows, equal the ratio of TAR model to the following benchmark. Thus, a number below one indicates the improvement of the TAR model over its competitor. Observations for RMSE and MAE of the TAR model are standardized by 1000. Finally, the first four rows correspond to the 5-step-ahead, while the next four to the 10-step-ahead forecast, respectively. Observations from RMSE and MAE are standardized by 1000.

Finally, we compare TAR(2) and its competitors during recent financial crisis. The improvement in the MAE and RMSE metrics is comparable for crisis and longer out-of-sample periods and equal to approximately 2%. Although the GW test indicates that the TAR(2) and HAR model have the same forecasting errors, this can be explained by the relatively short size of the out-of-sample for both U.S. and Eurozone crises.

To sum up, our non-linear model outperforms its competitors thanks to its ability to capture different regimes in volatility and to measure volatility much more accurately than daily models. In addition, our model achieves approximately the same rate of improvement over the HAR model as much more complicated non-linear models, but with lower computational costs, since the TAR(2) model has only two regimes. For example, Scharth and Medeiros (2009) modelled realized volatility with five regimes and achieved an im-

Figure 1.9: Multiple-step-ahead forecast in 2008-2014



Comparison of aggregate volatility over five days and corresponding forecasts based on the TAR(2), HAR, GARCH(1,1) and GJR-GARCH(1,1) models from January 2008 to June 2014 (1604 observations). The red line indicates the aggregate five-step forecast, while the blue line the actual data.

provement in forecasting performance over the HAR model of around 3%. This feature is essential for practical applications.

1.5 Conclusions

This chapter develops a non-linear threshold model for RV (realized volatility), allowing us to obtain a more accurate volatility forecast, especially during periods of financial crisis. The changes in volatility regimes are driven by negative past returns, where the threshold equals approximately -1% . This finding remains robust to different functional forms of volatility and different set of indices from both developing and developed countries. The additional flexibility of the model allows one to produce a more accurate one-day-ahead

forecast compared to the linear HAR specification and GARCH family models. More importantly, the superior multiple-step-ahead forecasting performance of TAR is achieved not only in particular samples, but also in population according to the GW test for the out-of-sample period from 2008 to 2014. Finally, we derive an approximated closed form solution for multiple-step-ahead forecast, which is based on the NIG conditional distribution of returns. The non-linear threshold model primarily outperforms its competitors during periods of financial crisis.

The superior forecasting performance of TAR over other high-frequency models, as well as inter-daily GARCH specifications might warrant further examination. First, while the option pricing literature primarily relies on GARCH-type models, very few works exploit the availability of high-frequency data (e.g., see Stentoft (2008), Corsi et al. (2013), Christoffersen et al. (2014)). Thus, it might be useful to incorporate the TAR model into the option pricing framework, especially during periods of elevated volatility. We conjecture that a more accurate volatility model should result in lower hedging costs and, therefore, produce economic gains.

Second, the extension of the univariate to the multivariate models remains an important area of research given significant demand from practitioners. However, non-synchronicity is the key problem of estimating the covariance matrix despite the abundance of high-frequency data. Alternatively, a copula-based approach allows one to avoid this problem and to estimate the joint distribution of many assets. It would be interesting to incorporate the non-linear TAR model into the copula framework, since the current literature focuses either on the GARCH or HAR models (Patton and Salvatierra (2015) and Oh and Patton (2016)).

Chapter 2

Uncovering a Data Generating Process of Returns with Application to the Derivatives' Pricing, Risk Management and Strategy Development

2.1 Introduction

The importance of volatility for risk management and pricing of financial derivatives is well established in the academic literature and among practitioners. The option pricing literature is primarily dominated by continuous time processes with stochastic volatility after the seminal works of Black and Scholes (1973) and Merton (1973). Although the continuous time models often provide tractable semi-closed form expressions of the security prices in a quite general framework — see for example Heston (1993), Duffie et al. (2000),

their estimation is not straightforward.³ Meanwhile, time-series models with time-varying volatility were primarily developed in the GARCH discrete-time framework motivated by Engle (1982) work. Estimation of these models relies only on historical returns data and thus avoids the filtering problem, where volatility is treated as unobserved state variable.

The availability of high frequency data has generated a lot of research aiming at more accurate measuring, modeling and forecasting latent volatility. As volatility becomes essentially observable, researchers have applied time-series techniques for forecasting purposes. These have been shown to outperform daily models such as different GARCH specifications — Andersen et al. (2007), McAleer and Medeiros (2008). Despite their superior performance, the option literature primarily relies either on stochastic volatility or GARCH specifications — Heston and Nandi (2000), Christoffersen et al. (2006), Christoffersen et al. (2013a). Surprisingly, there are very few papers that exploit model-free measures of volatility, exceptions are Stentoft (2008), Majewski et al. (2015), Christoffersen et al. (2014) and Christoffersen et al. (2015a). Motivated by the superior performance of volatility models based on high frequency data we incorporate it into our option pricing framework.

The goal of this chapter is to build a model of the equity index return, which is (1) consistent under physical and risk-neutral measures, (2) can be easily estimated using daily returns, realized volatility, options data and is convenient for the option valuation, (3) has an economic intuition and (4) document both statistical and economic gains. To achieve these goals, we propose a novel discrete time stochastic volatility model based on available high frequency data, which uses mixture of normal (MN) distribution for the realized

³The literature proposes several approaches to estimate continuous time models using non(semi)parametric approaches (Jiang and Knight (1997), Conley et al. (1997)); empirical characteristic functions (Jiang and Knight (2002), Singleton (2001)); Markov Chain Monte Carlo (Eraker (2001), Eraker (2004)); simulated method of moments (Duffie and Singleton (1993)); and efficient method of moments (Andersen et al. (2002)).

volatility. First, our model addresses Bates (1996) critique, that an accurate model should be able to fit both returns under two measures, thus avoiding overfitting option data and simultaneously providing poor fit of returns data. Second, this flexible distribution is able to produce non-Gaussian dynamics of index returns, while it remains tractable enough for option valuation purposes. Third, we provide economic intuition for our model and relate the mixing structure of our distribution with several theories that explain volatility formation process. Finally, we document both statistical gains under objective and risk-neutral measures compared with nested benchmarks, and translate them into economic gains measured by superior performance of a simple quantitative trading strategy.

Our model essentially corresponds to a multifactor volatility framework, where constant, daily, weekly and monthly volatility, respectively, serve as factors. Thanks to a mixture of normal distribution it has semiparametric flavour and can approximate unknown distribution of realized volatility, while avoiding tight parameteric specification proposed in a number of affine models (Christoffersen et al. (2015a), and Majewski et al. (2015)). In addition, mixture of normal distribution has an appealing economic intuition and can be generated by either heterogeneity in information arrival (Cont (2007)) or evolutionary models (LeBaron et al. (1999)). We also incorporate volatility in volatility effect in our framework, which generates time-varying dynamics of high order moments. We document that this property has important applications to capture time-varying dynamics of both returns and realized volatility and assess statistical gains associated with the volatility of volatility effect.

Having established the key drivers of the data generating process under the objective measure we continue to study its benefits under the risk-neutral measure. For this purposes, we derive the risk-neutral dynamics of our model based on the approach of Christoffersen

et al. (2010) (henceforth CEFJ), which relies on the no-arbitrage principle and the equivalent martingale measure. We document the importance of the multivariate structure in the realized volatility's distribution by assessing gains of MN compared with nested Gaussian distribution. In particular, former achieves twice lower pricing errors for out of the money Put options, including short, medium and long term contracts. The lower option pricing errors of the short out of the money options highlights the ability of MN to generate "jumps" in returns or/and volatility processes.

Finally, we show the importance of market expectation embedded in the option prices for more accurate risk assessment and quantitative strategy development. We assess former by forecasting realized volatility's conditional mean and computing Value-at-Risk (VaR) and compare with models based on the historical data. We document a significant improvement in the VaR prediction based on the MN and Gaussian risk-neutral models compared with physical competitors. This result highlights the importance of forward-looking information embedded in the option prices, especially during periods of financial crises. We also translate statistical gains into economic benefits by building an algorithmic trading strategy, which is market-neutral, has the same volatility as a S&P 500 returns, while significantly outperforms benchmark index (76% compared with 2%) during 2008-2011 years.

The remainder of this chapter is organized as follows. Section 2.2 defines the model under the physical measures, discusses estimation procedure and provides risk neutralization. Section 2.3 presents estimation results and diagnostic checks of the mixture of normal model. The option valuation is described in Section 2.4, while Section 2.5 presents application of our model for predicting returns distribution and conditional mean of realized volatility. Section 2.6 describes our algorithmic strategy, while Section 2.7 concludes.

2.2 Model

In this section we specify our new model for return and conditional volatility under the physical measure. Next, we discuss estimation of the mixture of normal model using the maximum likelihood approach and expectation maximization algorithm. We derive the risk neutral distribution of volatility and innovations given a specific choice of a Radon-Nikodym derivative. Finally, we discuss several identification strategies of a Radon-Nikodym derivative using option prices.

2.2.1 Model of DGP for returns under the objective measure P

We assume that the return is driven by the daily volatility $\sqrt{RV_t^d}$ and i.i.d. innovations z_t under the objective measure P :

$$R_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \mu_t + \sqrt{RV_t^d} z_t, \quad (2.1)$$

where R_t is log return at period t , μ_t is its conditional mean, $\sqrt{RV_t^d}$ is the realized volatility, and z_t is an innovation at period t . This modelling framework is common in the discrete-time literature (see Christoffersen et al. (2012) for overview) and reflects the fact that returns are uncorrelated, but not independent. Indeed, the second moment of returns exhibits persistent behavior, which we capture through the dynamics of realized volatility.

Although there are numerous ways of modeling the conditional mean of realized volatility (Maheu and McCurdy (2011), Andersen et al. (2003)), we choose a heterogeneous au-

toregressive model (HAR) introduced by Corsi (2009) with augmented leverage effect:

$$\sqrt{RV_t^d} = \beta_0 + \beta_1 \sqrt{RV_{t-1}^d} + \beta_2 \sqrt{RV_{t-1}^w} + \beta_3 \sqrt{RV_{t-1}^m} + c \cdot \mathbb{1}[R_{t-1} < 0] \sqrt{RV_{t-1}^d} + \epsilon_t. \quad (2.2)$$

Here $\sqrt{RV_{t-1}^d}$, $\sqrt{RV_{t-1}^w}$ and $\sqrt{RV_{t-1}^m}$ are daily, weekly and monthly realized volatilities, and c captures leverage effect.⁴ The error term of realized volatility ϵ_t determines the conditional distribution at time t . We define lower frequency volatility component, for example weekly realized volatility as:

$$\sqrt{RV_t^w} = \frac{\sqrt{RV_t^d} + \dots + \sqrt{RV_{t-4}^d}}{5}, \quad (2.3)$$

The monthly realized volatility is computed as average of daily volatilities over 22 days. The HAR model and its extensions have proved to be quite successful in modeling and forecasting volatility in a parsimonious way using only few volatility components — Corsi et al. (2008) and Andersen et al. (2007). In particular, the HAR structure is able to replicate the main empirical properties of realized volatility, including volatility clustering and long memory. At the same time, our modelling framework can incorporate other processes for conditional mean of realized volatility, for example regime-switching models (McAleer and Medeiros (2008), Pypko (2015)).

Next, we specify the conditional distribution of both return and realized volatility. Our model has two sources of uncertainty at period $t - 1$, which affect daily returns — innovations z_t and error term in realized volatility ϵ_t , which follow bivariate mixture of normal

⁴In other words, positive value of c implies that negative return at period $t - 1$ leads to a higher realized volatility at period t compared with positive return of the same magnitude.

(MN) distributions:

$$\begin{aligned}
 pdf(\epsilon_t, z_t | I_{t-1}) &= \sum_{k=0}^{K-1} \pi_k N \left((0, 0)', \begin{pmatrix} \sigma_{kt}^2 & 0 \\ 0 & 1 \end{pmatrix} \right) \\
 \sum_{k=0}^{K-1} \pi_k &= 1, \quad K = 4 \\
 \sigma_{kt}^2 &= \sigma_k^2 RV_{t-1}^{(k)} : \quad RV_{t-1}^{(0)} = 1, \quad RV_{t-1}^{(1)} = RV_{t-1}^d, \quad RV_{t-1}^{(2)} = RV_{t-1}^w, \quad RV_{t-1}^{(3)} = RV_{t-1}^m,
 \end{aligned} \tag{2.4}$$

where K is a number of mixtures, π_k is a probability of each mixing distribution and I_{t-1} is and information set available at period $t-1$. Although, volatility is observable at $t-1$, it remains stochastic next period and follows mixture of normal (MN) distributions. In essence, volatility becomes "observable" at period $t-1$ given availability of the high frequency data to construct its non-parametric estimator. Specifically, we choose realized kernel developed by Barndorff-Nielsen et al. (2008) as proxy for the unobservable volatility since it is robust to the market microstructure noise and it performs better than other realized measures in terms of the Value-at-Risk forecasting (Brownlees and Gallo (2009)). In contrast, daily discrete time GARCH models assume that innovations are the only source of randomness, while latent volatility is filtered out from daily returns. Meanwhile, innovations z_t follow the standard normal distribution, since we choose variances within each mixtures equal to one.

We assume that our MN distribution has four components and they are related to the same volatility component as in HAR model: constant, daily, weekly and monthly volatility, respectively. This assumption allows to generate time-varying volatility of volatility σ_{kt}^2 , which is proportional to the current volatility level. This stylized fact of realized

volatility is called "volatility in volatility effect" and is important for both statistical analysis under objective (Section 2.3) and risk-neutral measures (Sections 2.4 and 2.6). In other words, volatility of volatility tends to be high (low) during financial crises (expansions) when current level of volatility is high (low). We deliberately do not employ more elaborate processes for volatility in volatility as proposed by Bollerslev et al. (2009) due to the following reasons. First, our parsimonious specification is able to match quite accurately unconditional distribution of both returns and volatility, which we discuss this in more details in Section 2.3. Second, the more sophisticated dynamics of volatility in volatility process makes the estimation procedure and risk-neutralization much more complicated.

We choose four component in mixing distribution to convey economic intuition about the heterogeneous structure of volatility's distribution in the financial markets.⁵ Specifically, its shape is driven by the interaction of heterogeneous agents with different investment horizons and risk appetites — daily (e.g. high frequency traders), weekly (e.g. banks) and monthly components (e.g. pension funds), correspondingly. Meanwhile, the probabilities of mixing distribution reveal the contribution of each agent's type to the overall volatility of volatility. This mechanism is similar to two types of models proposed in literature, which explain the formation and evolution of volatility: heterogeneous arrival rates of information and evolutionary models. The heterogeneous news approach models volatility clustering and its long memory by assuming heterogeneity in investors time's horizons (Andersen and Bollerslev (1997), Cont (2007) and LeBaron (2001)). Meanwhile, evolutionary models replicate volatility dynamics by simulating artificial financial market populated by different agents, which have specific set of rules (LeBaron et al. (1999) and Arthur et al. (1997)).

⁵It is possible to consider model with more mixtures and either estimate their optimal number as an additional parameter or use statistical criteria to test whether additional component improves a data fit. However, once we add more than four mixtures we lose an appealing economic interpretation of our model.

In addition to economic intuition, our MN model has several appealing statistical properties. First, it is flexible enough to approximate conditional distribution of realized volatility. Although, the unconditional distribution of realized volatility or variance can be closely approximated by Inverse Gaussian distribution (Forsberg and Bollerslev (2002) and Stentoft (2008)), the conditional distribution can be very different from unconditional. In contrast, MN distribution is immune to this problem and able to match conditional distribution of latent volatility given the substantial amount of mixtures. This is an important distinction between this chapter and the affine models literature, which makes a specific distributional assumption to generate an affine structure. In contrast, our framework assumes more flexible distribution of both returns and realized volatility. Not surprisingly, discrete time non-affine models specified at daily frequency outperforms affine models in terms of more accurate option valuation (Christoffersen et al. (2012)). We conjecture that our non-affine model based on the flexible time-varying distribution of returns and realized volatility will continue to outperform high frequency affine models, but leave their systematic comparison beyond the scope of this chapter. Second, multiple factor stochastic volatility models perform reasonably well for option valuation purposes (Christoffersen et al. (2009), which motivates us to include several factors. As a result, we use constant, daily, weekly and monthly volatilities as the corresponding factors in the MN(4) model. Third, basis functions contain only one parameter to estimate within each Gaussian component (σ_k), which allows to include many mixing distributions. An opposite approach is to use more sophisticated model within regime, like NGARCH in Rombouts and Stentoft (2015).

To sum-up, our novel discrete-time model generates flexible distribution of returns and realized volatility, while preserving an appealing economic intuition regarding the volatility formation. Our approach does not require a volatility filtering as in GARCH type models,

while discrete time framework allows its straightforward estimation, which we discuss in the next subsection. Finally, volatility of volatility effect combined with MN distribution allows to generate high order time-varying moments, which are important for both option valuation (Section 2.4) and strategy development (Section 2.6).

2.2.2 Estimation of MN

We estimate the vector of parameters $\theta = (\beta_0, \dots, \beta_{K-1}, c, \sigma_0, \dots, \sigma_{K-1}, \pi_0, \dots, \pi_{K-2}, \lambda)'$ by specifying the conditional mean in the following way:

$$\mu_t = r_t^f + \lambda \sqrt{RV_t^d}, \quad (2.5)$$

where r_t^f is risk free rate and λ is a price of volatility risk. By definition, the conditional log-likelihood for bivariate mixtures of normal distributions is defined as:

$$l(\theta) = \sum_{t=22}^T \ln \left(\sum_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp \left(-\frac{1}{2} (X_t - M_{kt})' \Sigma^{-1} (X_t - M_{kt}) \right) \right), \quad (2.6)$$

where $X_t = (\epsilon_t, z_t)'$, $M_{kt} = (0, 0)'$, $\Sigma = \begin{pmatrix} \sigma_{kt}^2 & 0 \\ 0 & 1 \end{pmatrix}$, while innovations z_t are filtered out from model (2.1):

$$z_t = \frac{R_t - r - \lambda \sqrt{RV_t^d}}{\sqrt{RV_t^d}}. \quad (2.7)$$

After simplification the log-likelihood becomes:

$$l(\theta) = \sum_{t=22}^T \ln \left(\sum_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp \left(-\frac{1}{2\sigma_{kt}^2} [\epsilon_t^2 + z_t^2 \sigma_{kt}^2] \right) \right). \quad (2.8)$$

The main obstacle with estimation of MN models is that direct maximization is challenging due to the sum inside the logarithm. As a result, the mixture log-likelihood function is not well behaved, which might produce local instead of global maximum and unstable estimates (Wirjanto and Xu (2009)). These problems are especially acute when we increase the number of mixtures. Therefore, we use an alternative approach, which is robust to the above mentioned issues — expectation maximization algorithm (EM) proposed by Dempster et al. (1977).

The idea behind the EM algorithm is to swap logarithms and sums and therefore maximizing lower bound of log-likelihood function. For this purposes, we introduce a distribution of hidden variable, which specifies the evolution of each normal distribution. The EM algorithm consists of two part: expectation (E) and maximization (M), correspondingly. At the first step (E), we compute the probability that random variable is drawn from specific normal distribution given some initial guess for θ . Next at the M step, we maximize the expected value of log-likelihood given these probabilities with respect to the hidden variable and find a new set of parameters. Note that maximization of the expected value of log-likelihood is much easier than original likelihood since we interchange logarithm and sum. Finally, we iterate until estimates at the previous step converges to estimates at the next step. Despite the advantages of using EM algorithm there are several major drawbacks, including slow speed of convergence and absence of standard errors. The former problem does result in significant computation time for our model, while we alleviate latter's by computing asymptotic standard errors for MLE estimates. Appendix B.1 provides more details about EM algorithm.

Having discussed the main benefits of Expectation Maximization approach we proceed with estimation of model (2.1) in two steps. First, we estimate parameters $(\beta_0, \dots, \beta_{K-1}, c)$,

which are constant for all mixtures by ordinary least squares. Next, we derive EM estimators for the remaining parameters $(\sigma_0^2, \dots, \sigma_{K-1}^2, \pi_0, \dots, \pi_{K-2}, \lambda)$ in Appendix B.1 as follows:

$$\begin{aligned}
\hat{\sigma}_k^2 &= \frac{\sum_{t=22}^T q_K(k) \frac{\epsilon_t^2}{RV_t^{(k)}}}{\sum_{t=22}^T q_K(k)} \\
\hat{\pi}_k &= \frac{\sum_{k=0}^{K-1} q_K(k)}{T} \\
\hat{\lambda} &= \frac{\sum_{t=22}^T \frac{R_t - r_t^f}{\sqrt{RV_t^d}}}{T - 22} \\
q_K(k) &= \frac{\frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(-\frac{1}{2\sigma_{kt}^2} \cdot [\hat{\epsilon}_t^2 + \hat{z}_t^2 \sigma_{kt}^2]\right)}{\sum_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(-\frac{1}{2\sigma_{kt}^2} \cdot [\hat{\epsilon}_t^2 + \hat{z}_t^2 \sigma_{kt}^2]\right)},
\end{aligned} \tag{2.9}$$

where $q_K(k)$ is a probability that error term ϵ_t is drawn from the k th Normal distribution. Meanwhile, vectors X'_{-kt} and β_{-k} include all volatility components except of those corresponding to the k th frequency. Finally, we iterate parameters for EM algorithm till convergence given our initial guess and estimates $(\hat{\beta}_0, \dots, \hat{\beta}_{K-1}, \hat{c})$ obtained from the least-squares.⁶

2.2.3 Risk neutralization

In this section we derive the joint distribution of errors ϵ_t and innovations z_t under the risk neutral measure Q . We follow the approach developed by CEFJ, which uses the no-arbitrage principle and the equivalent martingale measure (EMM). Specifically, we define a Radon-Nikodym derivative and derive conditions under which the risk neutral measure Q becomes an EMM. This framework does not make any explicit assumptions about the

⁶The estimates for λ is identical to the OLS estimates obtained from (2.5), where z_t is an error term.

underlying economy or investors preferences, while assuming only the existence of the conditional joint moment generating function (MGF).

We define a Radon-Nikodym derivative as:

$$\frac{dQ}{dP} | I_t = \exp \left(\left[- \sum_{i=1}^t (v_{1,i} \epsilon_i + v_{2,i} z_i + \Psi_i(v_{1,i}, v_{2,i})) \right] \right), \quad (2.10)$$

where $\Psi_i(v_{1,i}, v_{2,i})$ is the logarithm of the joint conditional MGF of the vector $X_t = (\epsilon_t, z_t)'$, $\{v_{1,t}\}$ and $\{v_{2,t}\}$ are some deterministic sequences at period $t - 1$. The MGF of bivariate MN vector X_t is given by:

$$\Psi_t(t_1, t_2) = \log \left[\sum_{k=0}^{K-1} \pi_k \exp \left(\frac{t_1^2 \sigma_{kt}^2}{2} + \frac{t_2^2}{2} \right) \right]. \quad (2.11)$$

Theorem 2.2.1 derives the conditional MGF under the risk neutral distribution Q .

Theorem 2.2.1 *Let returns follow (2.1), while conditional volatility and innovations are distributed according to bivariate mixture of normal distributions. Given Radon-Nikodym derivative (2.8) the log of conditional MGF under risk neutral measure are obtained as follows:*

$$\Psi_t^Q(t_1, t_2) = \log \left[\sum \tilde{\pi}_{kt} \exp \left(-t_1 \tilde{\mu}_{kt} - t_2 \tilde{M}_t + \frac{1}{2} (t_1^2 \sigma_{kt}^2 + t_2^2) \right) \right], \quad (2.12)$$

where:

$$\begin{aligned}
 \widetilde{\mu}_{kt} &= -v_{1t}\sigma_{kt}^2, \\
 \widetilde{M}_t &= -v_{2t}, \\
 \widetilde{\pi}_{kt} &= \frac{\pi_k \exp\left(\frac{v_{1t}^2\sigma_{kt}^2 + v_{2t}^2}{2}\right)}{\sum_{k=0}^{K-1} \pi_k \exp\left(\frac{v_{1t}^2\sigma_{kt}^2 + v_{2t}^2}{2}\right)},
 \end{aligned} \tag{2.13}$$

Proof See Appendix B.2.

Note that change of measure from the objective to the risk-neutral affects both conditional mean and probabilities, while the variance of volatility remains unchanged within each Gaussian component. This result is in line with the literature — Rombouts and Stentoft (2015) derive similar result for univariate MN distribution of innovations. Meanwhile, the joint distribution under physical and risk neutral measure belong to the same family of distributions, but with different parameters. Given the results of Theorem 2.2.1 we are ready to derive risk-neutral probability measure for the model (2.1).

Theorem 2.2.2 *Let returns follow (2.1), while conditional volatility and innovations are distributed as bivariate mixture of normal distributions. Given Radon-Nikodym derivative*

(2.8) the risk neutral probability measure becomes:

$$\begin{aligned}
R_t &= \ln\left(\frac{S_t}{S_{t-1}}\right) = r_t^f - \gamma_t + \sqrt{RV_t^d} z_t^* \\
\gamma_t &= -v_{2t}M_t + 0.5M_t^2 + \log\left(\frac{\sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \exp\left(\frac{(M_t - v_{2t})^2 \sigma_{kt}^2 + 2(M_t - v_{2t})\widetilde{\mu}_{kt} + \widetilde{\mu}_{kt}^2}{2(1 - \sigma_{kt}^2)}\right)}{\sqrt{1 - \sigma_{kt}^2}}\right) \\
\sqrt{RV_t^d} &= \beta_0 + \beta_1 \sqrt{RV_{t-1}^d} + \beta_2 \sqrt{RV_{t-1}^w} + \beta_3 \sqrt{RV_{t-1}^m} + c \cdot \mathbb{1}[R_{t-1} < 0] \sqrt{RV_{t-1}^d} + \epsilon_t^* \\
pdf(\epsilon_t^*, z_t^* | I_{t-1}) &= \sum_{k=0}^{K-1} \widetilde{\pi}_{kt} N\left(\widetilde{\mu}_{kt}, \widetilde{M}_t\right), \begin{pmatrix} \sigma_{kt}^2 & 0 \\ 0 & 1 \end{pmatrix} \\
\widetilde{\mu}_{kt} &= -v_{1t} \sigma_{kt}^2, \\
\widetilde{M}_t &= -v_{2t}, \\
\widetilde{\pi}_{kt} &= \frac{\pi_k \exp\left(\frac{v_{1t}^2 \sigma_{kt}^2 + v_{2t}^2}{2}\right)}{\sum_{k=0}^{K-1} \pi_k \exp\left(\frac{v_{1t}^2 \sigma_{kt}^2 + v_{2t}^2}{2}\right)}, \\
\sum_{k=0}^{K-1} \widetilde{\pi}_{kt} &= 1 \\
\sigma_{kt}^2 &= \sigma_k^2 RV_{t-1}^{(k)}, \quad RV_{t-1}^{(0)} = 1, \quad RV_{t-1}^{(1)} = RV_{t-1}^d, \quad RV_{t-1}^{(2)} = RV_{t-1}^w, \quad RV_{t-1}^{(3)} = RV_{t-1}^m.
\end{aligned} \tag{2.14}$$

Proof See Appendix B.3.

Recall that γ_t is an adjustment term to make expected value of return equals to the risk free rate under the risk neutral measure. Our next step is to specify deterministic sequences $\{v_{1t}\}$ and $\{v_{2t}\}$ using conditions when risk neutral measure Q becomes EMM:

Theorem 2.2.3 Given Radon-Nikodym derivative (2.11) the risk neutral probability mea-

sure becomes EMM if and only if:

$$\sum_{k=0}^{K-1} \pi_k \exp\left(\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} - \lambda M_t\right) = \sum_{k=0}^{K-1} \frac{\pi_k}{\sqrt{1 - \sigma_{kt}^2}} \exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t b_t \sigma_{kt}^2}{2(1 - \sigma_{kt}^2)}\right),$$

$$M_t = \beta_0 + \beta_1 \sqrt{RV_{t-1}^d} + \beta_2 \sqrt{RV_{t-1}^w} + \beta_3 \sqrt{RV_{t-1}^m} + c \cdot \mathbb{1}[R_{t-1} < 0] \sqrt{RV_{t-1}^d}, \quad (2.15)$$

$$b_t = M_t - v_{2t},$$

$$c_t = \lambda - v_{1t}.$$

The approximate solution for $\{v_{2,t}\}$ is defined as:

$$v_{2t} \approx \frac{M_t + 2\lambda}{2}. \quad (2.16)$$

Proof See Appendix B.4.

Theorem 2.2.3 does not provide unique solutions for the deterministic sequences $\{v_{1,t}\}$ and $\{v_{2,t}\}$ since market is incomplete. In other words, we have only one condition for EMM and two unknowns variables each period, therefore we can not identify both sequences using returns and realized volatility data only.⁷ Thus, we have to identify the remaining parameter v_{1t} from both physical (returns and realized volatility) and risk neutral measure (option prices), correspondingly.

2.2.4 Identification of v_{1t} sequence

We propose to identify the unknown sequence v_{1t} by calibrating our model to the price of the longest maturity at the money (ATM) option for each week. In addition, we consider

⁷ $E^Q \left[\frac{S_t}{B_t} | I_{t-1} \right] = \frac{S_{t-1}}{B_{t-1}}$, where B_t is a price of zero-coupon risk free bond at period t .

alternative identification strategy to assess robustness of our results and minimize relative RMSE for each day and assume that $\nu_{1t} \equiv \nu_1$:

$$\%RMS E_t = \sqrt{\frac{1}{N_O} \sum_{i,t} \left(\frac{C_{i,t}(\theta, \nu_t) - C_{i,t}}{C_{i,t}} \right)^2}, \quad (2.17)$$

where N_O is a number of options, $C_{i,t}$ and $C_{i,t}(\theta, \nu_t)$ are actual and simulated option prices for a day t , ν_t is a vector of sequences ν_{1t} and ν_{2t} . We choose a calibration approach since it does not require the choice of minimization metric and computationally faster since we use only one option instead of panel options for each week. Meanwhile, both approaches lead to qualitatively similar results.

Our approach continues to work even for new market such that market prices of options do not exist. However, in this case our original identification procedure does not work since we can not identify pricing kernel. We solve this problem by assuming that Radon-Nikodym derivative depends only on the Gaussian shock by assuming that $\nu_{1,i} \equiv 0 \quad \forall i$:

$$\frac{dQ}{dP} |_{I_t} = \exp \left(\left[- \sum_{i=1}^t (\nu_{2,i} z_i + \Psi_i(0, \nu_{2,i})) \right] \right). \quad (2.18)$$

We compare option valuation performance of the model with original pricing kernel (2.8) and its nested specification (2.16) and assess gains of including the additional source of risk.

Having established identification strategy, we are ready to price options. We proceed as follows. First, we use rolling window estimation of length 1988 days (which corresponds to 2000-2007 period) to estimate all physical parameters. Second, we identify parameter ν_{1t} from option data for a given week. Third, we risk-neutralize our model, compute

sequence v_{2t} and obtain option prices using Monte-Carlo simulations. Finally, we assess option performance during 2008-2011 years.

2.3 Empirical analysis

In this section we first introduce high frequency data for realized volatility and daily data for returns, which is used in the empirical analysis. Then we describe our option data set. Next, we provide estimation results of parameters under physical measure obtained through the EM-OLS algorithm and maximum likelihood procedure. Finally, we discuss several diagnostic checks of our model by using simulations.

2.3.1 Data

Returns and realized volatility data

We use high frequency data for S&P 500 index for the empirical analysis. We collect this data from the Realized Library of Oxford-Man Institute of Quantitative Finance (Library version 0.2) designed by Heber et al. (2009), which is freely available.

The authors clean raw data obtained through Reuters Data Scope Tick History and compute high frequency estimators from cleaned data. The sample covers period from January 3 of 2000 to June 12 of 2014, overall 3603 trading days.⁸ We use the realized kernel as a proxy for the integrated variance.

⁸We exclude all days from the sample when market was closed.

Option data

We evaluate performance of our models using data for European Put options on the S&P 500 index for the period from 2008 to 2011. We obtain data set from the Option Metrics, which includes 21,759 European Put contracts. We apply standard filtering techniques adopted in the literature:⁹

1. We consider only options traded on Wednesdays to eliminate weekend effect.
2. We consider only options with time to expiration (tenor) more than 6 and less than 252 days, correspondingly.
3. We exclude options with traded volume less than 100 contracts.
4. We exclude all options with ask prices below 0.5\$ U.S. dollars.
5. We consider only options whose implied volatility can be computed.

Table 2.1 shows descriptive statistics of our option data set, divided into several maturity and moneyness categories, correspondingly. We define moneyness category as $M = \frac{S}{K \exp(-r_f T)}$, where S is an index level, K is a strike, r_f is a risk free interest rate, T is a tenor or time to expiration of the option. We observe the implied volatility smile in our data, that is volatility of out-of-the money (OTM) and in-the-money (ITM) options exceeds corresponding volatility of at-the-money (ATM) options, which is in line with the previous studies — Christoffersen et al. (2012) and Rombouts and Stentoft (2015). Overall, we conclude that our sample contains a large number of contracts and corresponds to the period of both financial stability and financial crises. The latter feature is important for our analysis

⁹See Rombouts and Stentoft (2015) and Heston and Nandi (2000) for details.

Table 2.1: Summary of the S&P 500 index options data

	Maturity				
	ST	MT	LT	VLT	All
<i>IVol %</i>	36.37	40.01	37.55	37.12	38.04
<i>Price</i>	12.58	16.53	28.85	46.49	23.34
<i>Number</i>	5156	7360	6151	3092	21579
	Moneyness				
	OTM	ATM	ITM	All	
<i>IVol %</i>	40.68	28.22	33.30	38.04	
<i>Price</i>	12.87	41.64	94.75	23.34	
<i>Number</i>	16560	3727	1472	21759	

Reported are implied volatility (*Vol*), option price (*Price*) computed as an average between bid and ask, and the number of contracts (*Number*). The maturity category is divided into short term (*ST*) with $T < 21$, medium term (*MT*) with $21 \leq T < 42$, long term (*LT*) with $42 \leq T < 84$ and very long term (*VLT*) with $T \geq 84$. The moneyness category is divided into out of the money (*OTM*) with $M < 0.98$, at the money (*ATM*) with $0.98 \leq M < 1.02$ and in the money (*ITM*) with $M \geq 1.02$. The moneyness is defined as $M = \frac{S}{K \exp(-r_f T)}$, where S is index level, K is strike, r_f is risk free interest rate, T is a tenor or time to expiration.

of option valuation, when periods of financial turmoil present the biggest challenges for academics and practitioners.

2.3.2 Estimation results

Estimation of MN models

Table 2.2 provides estimates of the mixture of normal model introduced in Section 2.2. First column presents results of the combined EM-OLS algorithm. Second column shows MLE estimates, where EM-OLS estimates were chosen as starting points. We also compute MLE asymptotic standard errors, since EM algorithm is an iterative procedure, which does not produce standard errors. Finally, last two columns show MLE estimates and their

Table 2.2: Estimations results

	EM	MLE(4)	S.E.	MLE(1)	S.E.
β_0	0.000	0.000***	(0.000)	0.001***	(0.000)
β_1	0.285	0.210***	(0.020)	0.206***	(0.023)
β_2	0.424	0.512***	(0.028)	0.502***	(0.033)
β_3	0.169	0.150***	(0.021)	0.170***	(0.024)
c	0.136	0.117***	(0.009)	0.119***	(0.010)
σ_0	0.024	0.026*	(0.013)		
σ_1	0.227	0.226***	(0.024)		
σ_2	0.237	0.221***	(0.009)	0.292***	(0.001)
σ_3	0.640	0.602***	(0.039)		
π_0	0.001	0.001	(0.001)		
π_1	0.360	0.245***	(0.063)		
π_2	0.577	0.671***	(0.064)		
π_3	0.062	0.083			
λ	0.077	0.083***	(0.016)	0.078***	(0.016)
<i>Loglik</i>	11424	11458		11205	
<i>BIC</i>	-22751	-22818		-22344	

Reported are in-sample estimation results and corresponding standard errors of the mixture of normal models. The first column represents the estimation result obtained from EM algorithm (7) and OLS estimates of the remaining parameters. The second column provides MLE estimates of the whole model (6). The third column shows asymptotic standard errors of the MLE estimates. Here ***, * and *, means that corresponding p -values are lower than 0.01, 0.05 and 0.1, respectively. The standard errors are MLE asymptotic standard errors. The in-sample covers period from January 2000 to June 2014 (3581 observations). The optimal value of log-likelihood is labeled as *Loglik*. The Bayesian Information criterion is computed as $BIC = -2 \cdot \text{Loglik} + p \cdot \log(n)$, where p is a number of parameters of the model and n is a number of observations.

asymptotic standard errors of Gaussian model with weekly volatility component only.¹⁰

The main conclusions from Table 2.2 are the following. First, all MLE estimates are statistically different from 0 at 1% significance level. Second, daily (36%) and weekly (58%) volatility components have the biggest impact on the variance of error term ϵ_t . Put in other words, the volatility in financial markets is primarily driven by high and medium

¹⁰We choose volatility component with the biggest probability weight and the highest value of likelihood among all Gaussian models with one component.

frequency agents, e.g. high frequency traders and commercial or investment banks. In contrast, agents with longer investment horizons, like pension funds or university endowments, have substantially lower impact on a market's volatility. Third, we can interpret the constant term as a very rare jump with big magnitude in realized volatility, which might happen with 0.1% chance. To sum-up, MN model contains very different volatility components, which produce fat-tailed and asymmetric distribution of stochastic volatility as depicted in Figure 2.3.

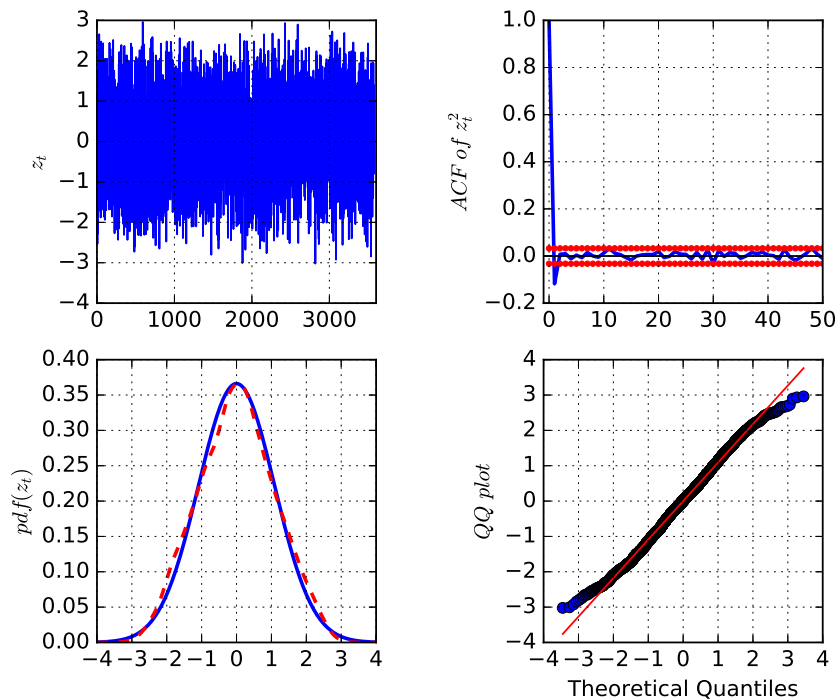
Fourth, the risk premium parameter λ is significantly different from 0, which implies that investors demand premium for holding risky financial assets. In this case, the volatility risk is priced, therefore justifying change of physical measure P , discussed in subsection 2.2.3, for the option valuation purposes. Finally, the direct comparison of our combined EM-OLS approach and MLE shows very little latter's improvement — only 0.3% increase in the value of Log-likelihood. In addition, the majority of estimates are very close in both methods, which motivates us to use combined EM-OLS approach for the option valuation purposes.¹¹ In contrast, Mixture of Normal model does not only improve overall likelihood compared with nested Gaussian specification, but also achieves lower BIC criterion.

Analysis of innovations

Next, we inspect the time series dynamics and distribution of innovations z_t filtered from (2.5). First graph of Figure 2.1 shows that these innovations do not cluster over time, while second graph illustrates that they are uncorrelated according to the sample auto correlation function. Meanwhile, third graph and fourth graphs compare non-parametric distributions of innovations with Gaussian distribution. Both graphs — unconditional distribution and

¹¹The difference between EM-OLS and MLE estimates are statistically insignificant for 10 out of 13 estimates at 5% significance level.

Figure 2.1: Analysis of innovations



Analysis of innovations $z_t = \frac{R_t - r - \lambda \sqrt{RV_t}}{\sqrt{RV_t}}$. First graph plots time series dynamics of innovations from February 2000 till June 2014 (3581 observations). Second graph plots sample auto correlation function of squared innovations (solid blue line) and 95% confidence band. Third graph compares normal distribution of innovations (solid blue line) with non-parametric distribution (dash red line), while fourth graph plots corresponding QQ plot.

QQ (Quantile-Quantile) plot — establish the close match between these distributions and highlight the importance of using Gaussian distribution as an approximation.

Overall, we conclude that distribution of z_t can be approximated by i.i.d. Normal distribution, though their second moments are dependent.

Table 2.3: Comparison of actual and simulated moments for returns and realized volatility

	MEAN	STD	SKEW	KURT
R_t	8.0E-05	1.2E-02	-0.2	10.2
MN(4) w VOL	7.2E-04	1.0E-02	0.2	9.4
Gaussian w VOL	6.8E-04	9.4E-03	0.2	8.8
MN(4)	7.6E-04	1.1E-02	0.1	6.4
Gaussian	7.5E-04	1.0E-02	0.1	6.0
$\sqrt{RV_t}$	9.3E-03	6.1E-03	3.3	24.6
MN(4) w VOL	8.7E-03	4.8E-03	1.8	9.6
Gaussian w VOL	8.3E-03	4.5E-03	1.8	8.9
MN(4)	9.6E-03	4.9E-03	0.9	8.0
Gaussian	9.0E-03	5.1E-03	0.4	2.9

Reported are comparison of actual and simulated under P measure sample moments for returns and realized volatility. The former is computed as an average based on 100 simulations. Each simulation contains 40,000 observations, but we keep only the last 3603 observations to match the size of the actual data. The four columns represent mean, standard deviation, skewness and kurtosis, respectively. First two columns compares moments of simulated and actual returns, while last two columns — of realized volatility.

MN(4) model simulations

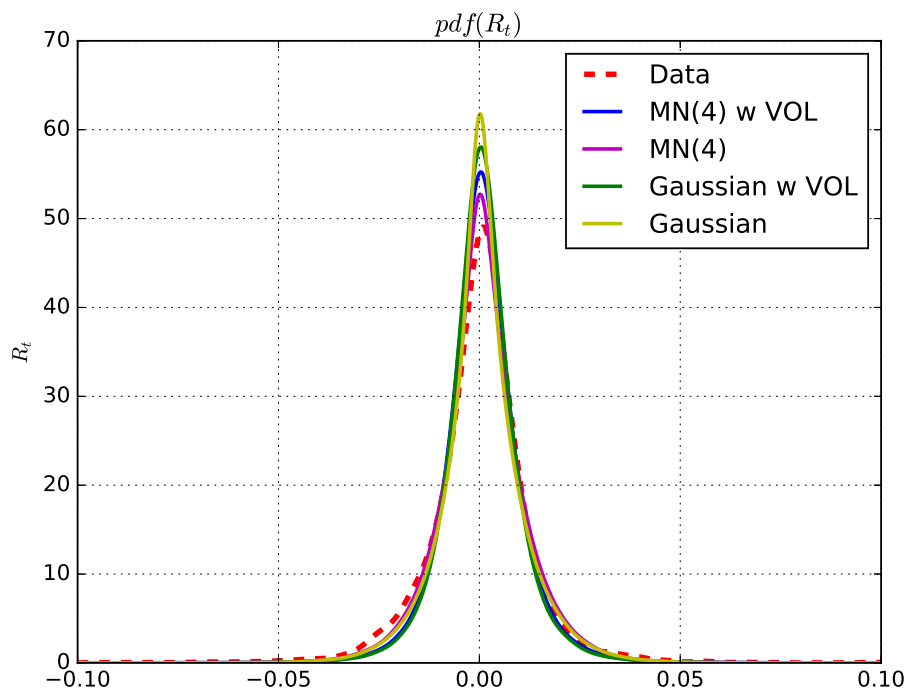
In the previous sections we have discussed a trade-off between flexibility of mixture of normal model and its econometric tractability. Next, we inspect model's ability to replicate stylized facts by using simulations. If the model does a good job in describing actual data then model-implied moments and probability density functions should be similar to the actual values. For this purposes, we simulate 40,000 observations for daily returns and realized volatility based on model (2.1) and (2.3). Then we keep only the last 3603 observations to match the size of our data set, while the previous observations serve as burn-in sample. Finally, we repeat our simulations 100 times and compute averages of all simulated moments presented in Table 2.3.

Overall, Table 2.3 highlights the importance of volatility in volatility effect in the underlying data generating process. Although, first two moments of returns and realized volatility are very similar across four models, the benefits of this effect become evident once we compare higher order moments. In particular, MN with volatility in volatility effect model generates unconditional kurtosis, which is very close to actual return's kurtosis. This result is impressive, since we do not add jumps in either returns or/and volatility process, but model is still able to generate fat tailed distribution. Similarly MN model produces significantly higher skewness and kurtosis in realized volatility, while models without volatility effect unable to match fat tails and positive skewness. Meanwhile, all models generate slightly positive skewness in returns, with actual returns are negatively skewed.

Figures 2.2 and 2.3 illustrate the benefits of volatility in volatility effect by plotting probability distribution functions of actual Returns and realized volatility compared with model implied distributions. As we discussed before, the mixture of normal (solid blue line) and Gaussian model (solid magenta line) with volatility in volatility effect produce much closer fit of unconditional distribution (dash red) compared with corresponding nested models — MN (solid green line) and Gaussian specification (solid yellow line), respectively.

To sum-up, MN and Gaussian specifications with volatility in volatility effect outperform nested models based on a number of metrics, including comparison of moments, probability density functions and autocorrelation functions. At the same time, the difference between MN and Gaussian model is relatively small, which suggests that we can not rank them based on assessment of their unconditional distributions. However, the benefits of MN distribution become more evident once we move to the option valuation, which requires accurate fit of conditional (rather than unconditional) distribution of returns. Thus, we consider option valuation of only those models, which incorporate volatility in volatility

Figure 2.2: Comparison of actual and simulated returns



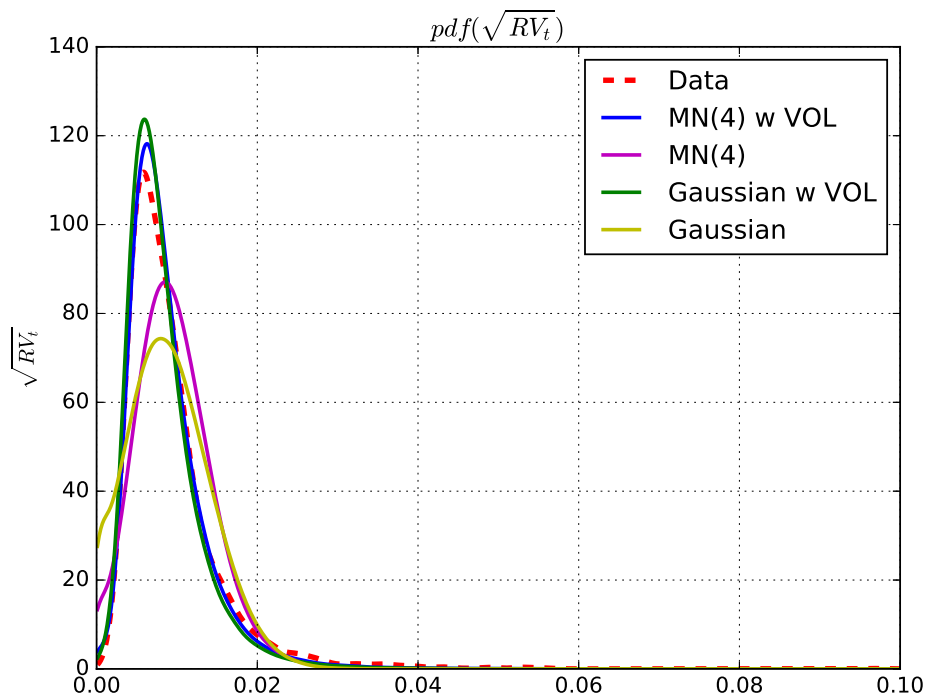
Comparison of actual (top graph) and simulated (bottom graph) daily returns. We simulated 40,000 observations and then keep only last 3603 observations to match the size of actual data. We repeat these simulations 100 times and compute average of corresponding parameters. This graph compares pdf of model-simulated and actual data using non-parametric kernel estimator. We consider the following models: MN(4) with volatility effect (solid blue), MN(4) (solid magenta), Gaussian model with volatility effect (solid green), Gaussian (solid yellow) and actual (dash red) data.

effect in the next section.

2.4 Option valuation

In this section, we discuss valuation of the European Put options using Monte Carlo simulations. First, we consider markets with available market prices and therefore we are able to identify unknown sequence v_{1t} . Next, we consider valuation of new derivatives such

Figure 2.3: Comparison of actual and simulated realized volatility



Comparison of actual (top graph) and simulated (bottom graph) realized volatility. We simulated 40,000 observations and then keep only last 3603 observations to match the size of actual data. We repeat these simulations 100 times and compute average of corresponding parameters. This graph compares pdf of model-simulated and actual data using non-parametric kernel estimator. We consider the following models: MN(4) with volatility effect (solid blue), MN(4) (solid magenta), Gaussian model with volatility effect (solid green), Gaussian (solid yellow) and actual (dash red) data.

that market prices do not exist, which impedes our identification procedure. We solve this problem by considering nested pricing kernel, which includes only Gaussian type of risk. Next, we discuss an implications of our model to analyze time-varying uncertainty in the financial markets during 2008-2011 years using model-implied variance, skewness and kurtosis. Finally, we assess performance of MN and nested Gaussian models across different moneyness and maturity categories defined in the Section 2.3.

2.4.1 Option pricing in the existent markets

The flexible mixture of normal distributions do not allow to obtain closed form expression for the option prices. In contrast, assuming specific parametric assumption researchers derived "semi-closed" solution for the option price — examples are Heston and Nandi (2000), Christoffersen et al. (2006) and Corsi et al. (2013). These studies obtained expression for the conditional MGF of returns and then used Fourier method to compute price of European call option. Meanwhile, we use Monte Carlo simulations to price securities while simulating returns and realized volatility paths. While semi-closed form expression seems to be more appealing for option valuation, Barone-Adesi et al. (2008) argue that computational time required by either semi-closed approach or simulations is approximately the same.¹² In addition, we avoid computationally extensive filtering techniques exploited in continuous time models (Bakshi et al. (1997)), since our model relies only on the returns and realized volatility data. Finally, simulation methods work for both exotic and plain vanilla options, while "semi-closed" form expression is available primarily for the European calls and puts options. The price of European call is computed as follows:

$$C(t, K) = e^{-r_f(T-t)} \frac{1}{MC} \sum_{m=1}^{MC} \max(S_T^m - K, 0), \quad (2.19)$$

where r_f is a risk free rate, T is an option's tenor, MC is a number of simulated paths ($MC = 20,000$), S_T^m is a value of index given m th simulated path. We set risk-free interest rate at a fixed level for each day for all simulated paths.¹³

We use implied volatility root mean square error (*IVRMSE*) to assess option pricing

¹²Both methods are based on running for-loop either for simulating paths (Monte Carlo) or for computing coefficients used for Fourier transformation ("semi-closed" solution).

¹³In fact, Bakshi et al. (1997) show that stochastic interest rate is not a first-order importance for the security pricing.

performance as advocated by Renault (1997):

$$IVRMS E = \sqrt{\frac{1}{N_O} \sum_{i,t}^{N_O} (IV_{i,t}(\theta, \nu_t) - \sigma_{i,t}^{BS})^2}, \quad (2.20)$$

where N_O is a number of options, $IV_{i,t}(\theta, \nu_t)$ and $\sigma_{i,t}^{BS}$ are implied volatilities of simulated and actual prices, $C_{i,t}$ and $C_{i,t}(\theta, \nu_t)$ are actual and simulated option prices for a day t , ν_t is a vector of sequences ν_{1t} and ν_{2t} .

While we use "crude" Monte Carlo method to price options, several approaches can be applied to reduce variance and improve efficiency — see Boyle et al. (1997) for an overview of such techniques. In addition, Duan and Simonato (1998) propose a modification of Monte Carlo, which does not only ensure Martingale property for the discounted simulated prices under Q measure, but also leads to lower price errors. Recall that all aforementioned techniques are important to improve security pricing of the specific option. In contrast, we assess performance of our model based on the averaging of aggregate errors across time and cross section of options, which in essence leads to reduction of pricing errors. Thus, we select the simplest simulation technique to price options.

Table 2.4 shows option valuation performance for European calls of our MN model and compares it with Gaussian model, where all model incorporate volatility in volatility effect.

Panel A displays implied volatility pricing errors across different moneyness and maturity dimensions, respectively. Panel B shows relative improvement of MN model compared with Gaussian model. The benefits of MN model are especially evident for OTM options, and equal to 43%, 56%, 58% and 65% for short, medium, long term and very long maturities, respectively. The superior performance in terms of pricing short term OTM options reveals an important benefit of MN specification to generate "jumps" in returns or/and

Table 2.4: Option valuation I

Panel A: *IVRS ME* of MN(4) w VOL model

	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
OTM	7.93	7.94	6.71	4.14
ATM	4.75	4.47	3.64	1.21
ITM	5.19	4.77	4.92	2.99

Panel B: Relative performance of MN compared with Gaussian model

	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
OTM	0.57	0.44	0.42	0.35
ATM	1.06	0.92	0.77	0.56
ITM	1.06	0.87	0.83	0.97

Panel C: Number of options

	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
OTM	3723	5816	4624	2397
ATM	1055	1098	1121	453
ITM	378	446	406	242

Panel A reports implied volatility root mean square errors (*IVRMS E*) of MN model measured in percentage points sorted by moneyness and maturity. The maturity category is divided into short term (*ST*) with $T < 21$, medium term (*MT*) with $21 \leq T < 42$, long term (*LT*) with $42 \leq T < 84$ and very long term (*VLT*) with $T \geq 84$. The moneyness category is divided into out of the money (*OTM*) with $M < 0.98$, at the money (*ATM*) with $0.98 \leq M < 1.02$ and in the money (*ITM*) with $M \geq 1.02$. The moneyness is defined as $M = \frac{S}{K \exp(-r_f T)}$, where S is index level, K is strike, r_f is risk free interest rate, T is a tenor or time to expiration. Panel B displays ratio of *IVRMS* of MN and Gaussian model, correspondingly. The number below unity implies the lower pricing errors of MN model. Panel C shows number of options in each moneyness and maturity category. The out-of-sample option data set includes European Calls recorded every Wednesdays from 2008 to 2011 years.

volatility. This is especially important since pricing of short OTM puts remains challenging for all models proposed in the literature (Christoffersen et al. (2013b), Christoffersen et al. (2015a)) and reflects precautionary motives of investors to buy short term insurance policy.

2.4.2 Analysis of model-implied higher order moments

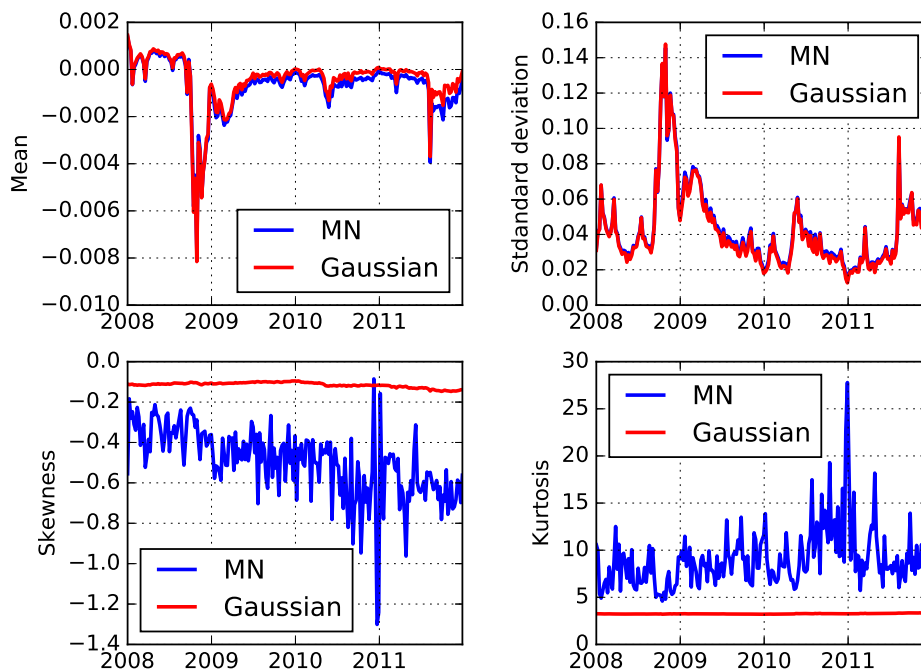
Having established superior performance of MN compared with Gaussian model, we study the reason behind this improvement. Recall, that MN model is able to generate time-varying higher-order moments, while Gaussian distribution has constant skewness and kurtosis.¹⁴ Figure 2.4 illustrates this point by plotting aggregated dynamics (over 10 days) of the first four moments obtained from MN and Gaussian specifications, correspondingly. Although, conditional mean and standard deviations are almost identical for both models, the evolution of higher order moments is very different. In other words, the ability of MN model to generate time-varying skewness and kurtosis allows to capture more accurately the evolution of tails of the returns distribution, and therefore translated into lower option pricing errors.¹⁵ In addition, MN model is able to generate negative skewed returns under risk-neutral measure, which is consistent with findings in the previous studies (Christoffersen et al. (2015b)).

Once we document the importance of the time-varying risk-neutral moments for option valuation, we study next their implications for the assessment of financial risk. Several conclusions emerge from the analysis of Figure 4 during turbulent 2008-2011 years. First, market participants did not anticipate incoming US housing crises as both skewness and kurtosis did not fluctuate substantially in the first half of 2008 year. In contrast, we observe a dramatic shift in the fear of both negative (skewness) and fat-tails returns (kurtosis) in the aftermath of 2008-2009 US housing crises. This change was primarily driven by European debt crises and reflected both positive and negative news for the investors. In particular,

¹⁴We can derive both closed-form expression for higher moments or alternatively use simulations.

¹⁵Gaussian model also generates time-varying skewness and kurtosis due to aggregation of time property even though one day ahead moments are zero and three, correspondingly. However, these moments are substantially less volatile than MN moments.

Figure 2.4: Comparison of MN(4) and Gaussian model-implied moments of returns under the risk-neutral measure



This figure plots mean, standard deviation, skewness and kurtosis of aggregated returns over 10 days under the risk-neutral measure obtained from MN (blue line) and Gaussian model (red line), respectively. The time span covers periods from 2008 till 2011, overall 208 weekly observations.

a surge in volatility of the high order moments started after appearance of news regarding revision of Greek budget and country's potential bankruptcy (November of 2009), followed by uncertainty and approval of first Greek (April of 2010), Irish (October of 2010), Portuguese (April of 2011) bailouts, respectively. After that investors expected that US debt ceiling turmoil in August of 2011 would constitute a significant threat to the stability of financial system – as kurtosis surged almost threefold. Similarly, overall uncertainty related to the shut-down of government for several weeks measured by conditional volatility surged in the beginning of August.

The above-mentioned analysis emphasizes the difference between two types of risk: level of current uncertainty (volatility) and future fear of negative and large returns (skewness and kurtosis), respectively. In contrast to rapid swings in higher order moments during European debt turmoil, a conditional volatility achieved its peak in the fall 2008 amidst Lehman collapse, AIG bailout and liquidity crunch. In other words, time-varying volatility has limited potential to detect early warning signals of financial vulnerability, but it rather measures a current level of uncertainty. Not surprisingly, a model-free measure of risk neutral volatility or VIX is called a "fear index" and widely used by academics, practitioners and policy makers. Therefore, the analysis of higher order moments has a special interest for the central banks, with an application to the development of the early warning system of financial vulnerability.

Although, we provide a qualitative analysis of the time-varying moments and relate it to the recent development in the international financial markets, the natural question arises regarding quantitative assessment of the model-implied moments. Indeed, we do not observe them in the data and rely instead on the model to compute mean, volatility, skewness and kurtosis. One approach is to compare our time-varying moments with others methods proposed in the literature (Audrino et al. (2015)). This chapter recovers both physical and risk-neutral moments in a model-free way from the derivatives prices over longer time-horizon, but report similar latter's dynamics in 2008-2011 years. Finally, we show how to exploit the information content of risk-neutral moments in terms of development quantitative trading strategy, which we discuss in the Section 2.6.

Table 2.5: Option valuation II

		<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
OTM	MN(4)	7.93	7.94	6.71	4.14	0.50	0.38	0.34	0.21
	Gaussian	13.97	18.07	16.11	11.90	0.88	0.87	0.80	0.60
	MN(4) with one risk	14.40	18.87	17.76	18.06	0.91	0.91	0.89	0.90
	Gaussian with one risk	15.91	20.83	20.02	19.98	1.00	1.00	1.00	1.00
		<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
ATM	MN(4)	4.75	4.47	3.64	1.21	0.82	0.60	0.39	0.11
	Gaussian	4.47	4.85	4.73	2.17	0.77	0.66	0.50	0.19
	MN(4) with one risk	5.67	7.21	9.04	10.51	0.98	0.97	0.96	0.93
	Gaussian with one risk	5.80	7.40	9.38	11.26	1.00	1.00	1.00	1.00
		<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>	<i>ST</i>	<i>MT</i>	<i>LT</i>	<i>VLT</i>
ITM	MN(4)	5.19	4.77	4.92	2.99	0.85	0.67	0.25	0.28
	Gaussian	4.92	5.49	5.95	3.09	0.80	0.78	0.30	0.29
	MN (4) with one risk	6.14	12.70	9.97	10.48	1.00	1.79	0.50	0.99
	Gaussian with one risk	6.13	7.08	19.76	10.63	1.00	1.00	1.00	1.00

This table reports implied volatility root mean square errors (*IVRMSE*) and corresponding ratios of *IVRMSE* sorted by moneyness and maturity. *IVRMSEs* are presented in columns 3-6th, while their ratios computed with respect to Gaussian model with one source of risk are displayed in columns 7-10th. The number below unity implies the lower pricing errors of any other model compared to Gaussian specification with one source of risk. The out-of-sample option data set includes European Calls recorded every Wednesdays from 2008 to 2011 years.

2.4.3 Option pricing in new markets

Our approach continues to work in new market where option prices do not exist, once we shut down a fat-tail risk and focus only on the Gaussian source of risk. Recall, that in this case a Radon-Nikodym derivative is identified under historical measure alone, thus we can assess gains of MN model compared with its nested Gaussian specifications. Table 2.5 compares option valuations of these two models and shows *IVRMSE* (3-6th columns) and relative performance MN specification measured by its ratio to Gaussian model's loss metrics (7-10th columns).

Table 2.5 show that MN continues to dominate Gaussian model in terms of lower pricing errors for the most liquid OTM options, where they are lower by 10% across all maturity categories. The improvement decreases once we move to the pricing of ATM option and occasionally Gaussian model provides more accurate valuation of ITM options (medium term contract), which is explained by their lower liquidity and volume. At the same time, we document the importance of an additional source of risk, as Gaussian model with full pricing kernel improves option valuation compared with nested MN specification incorporating one Gaussian risk only. Although, they have comparable performance for short and medium term OTM options, an additional source of risk starts to play more prominent role once we move to the pricing of long term options. Similarly, the performance of model with two sources of risk is substantially better for ATM and ITM options, but it is again driven by lower liquidity and volume. This result is not surprising since we force model to price ATM options accurately by calibrating sequence $v_{1,t}$ on a weekly basis.

2.5 Forecasting volatility and returns distributions

In this section, we highlight the importance of forward-looking information embedded in the option prices and its application for risk management purposes. Specifically, we document an improvement in forecasting of left tails of returns distribution by computing Value at Risk using both risk-neutral and physical models, correspondingly. Finally, we show how to improve forecasting of realized volatility's conditional mean using risk-neutral models compared with specifications based on the historical data only.

2.5.1 Forecasting Value at Risk

Having built a joint model of S&P 500 options and its underlying index we investigate next the information content of option prices in more details. In particular, we show how to improve forecast of the left tail of returns distribution by combining market expectations embedded in option prices with our MN model. For this purposes, we compute Value at Risk (VaR) using both risk-neutral and physical models and conduct corresponding back-testing procedure. This is especially important for risk management since our sample includes three recent periods of financial calamities: US financial crisis in 2008, Euro debt turmoil in 2010 and US debt ceiling crisis in 2011.

The literature proposed several approaches to model time-varying distribution of returns by specifying different dynamics of underlying volatility: GARCH (Engle (1982), Bollerslev (1986)), stochastic volatility (Hull and White (1987), Melino and Turnbull (1990)) and high frequency models (Andersen et al. (2003), Corsi (2009)). These specifications successfully capture a number of stylized facts, including volatility clustering, long memory and asymmetric responses to the past shocks (leverage effect). However, all of them rely on historical data and can not predict structural breaks or regime switching associated with financial crises, which dramatically affect VaR computations. In contrast, we propose a novel approach to improve modelling of returns distribution by exploiting forward-looking expectations embedded in the option prices. These prices reflect options traders' future beliefs about market outcome and might be useful for the forecasting purposes. Indeed, Roll et al. (2010) and Pan and Poteshman (2006) show that option traders' activities contain important information about future market outcomes and can predict future returns.

At the same time, a risk-neutral distribution is different from physical distribution due to presence of the risk premia. As a result, risk-neutral forecast of left tail differ from

desired prediction under the objective measure. Ideally, we want to separately identify both pricing kernel and objective distribution using derivative prices, which was recently done under specific assumptions in discrete time framework (Ross (2015)) and continuous time framework (Carr and Yu (2012)), respectively. However, Borovicka et al. (2016)) argue that this recovery might be misleading and it is still impossible to uniquely identify physical distribution of returns from the derivatives prices under more general set of assumptions. Having discussed the distortions of the objective distribution caused by the risk premia we continue computation of the VaR using risk-neutral model.

The Value at Risk is defined as α th quantile of returns distribution over l time period:

$$VaR_t(\alpha) = \inf\{x | F_R(x) \geq \alpha\}, \quad (2.21)$$

where $VaR_t(\alpha)$ is a α th quantile of the returns distribution, $F_R(x)$ is a CDF of aggregate returns from period t to $t + l$. Despite several critics of this measure, including the inability to predict returns distribution behind α th quantile¹⁶ and the fact that VaR is not a coherent measure,¹⁷ it is still widely used among practitioners.¹⁸

To backtest our models first we compute hit sequences defined as:

$$I_{t+l}(\alpha) = \begin{cases} 1, & \text{if } R_{t:t+l} < VaR_t(\alpha) \\ 0, & \text{if } R_{t:t+l} \geq VaR_t(\alpha) \end{cases}, \quad (2.22)$$

where $R_{t:t+l}$ are actual aggregate returns over period l , while the VaR values are obtained

¹⁶For these purposes, reseachers use expected shortfall measure defined as an expected returns in the worst α th cases or $ES = \frac{1}{\alpha} \int_0^\alpha VaR_t(x) dx$.

¹⁷The coherent measure should satisfy three preoprties: monotonicity, sub-addivity, homgeneity and translation invariance.

¹⁸Basel III regulatory framework is based on VaR measure (on Banking Supervision (2011)).

from either physical or risk neutral models using Monte Carlo simulations. Next, we follow Christoffersen (1998) approach and verify two properties of hit functions:

1. Unconditional coverage property: The probability of returns that are lower than $VaR_t(\alpha)$ should equal to $\alpha \cdot 100\%$ or $Pr(I_{t+l}(\alpha) = 1) = \alpha$. If probability is less than α then model is over conservative and overestimates risk, while opposite leads to the underestimation of the risk level.
2. Independence Property: any two elements of $I_{t+l}(\alpha)$ sequence should be independent. Put in other words, clustering of hit sequences implies that model can not properly adjust to capture time-varying risk.

Table 2.6 presents results for the unconditional coverage property and highlights more accurate risk assessment captured by the risk-neutral models. Gaussian and MN risk-neutral models substantially improve VaR assessment judging by their closer match of probabilities α . In addition, these models do not have systematic bias towards under or overestimation of volatility, while all physical models underestimate risk. This finding has important application for the risk management during period of financial turmoils, when investors are especially interested in accurate assessments of riskiness of their positions. Meanwhile, the difference between MN and Gaussian model is much less substantial, however former still matches probability α more accurately.

Figure 2.5 sheds more light on comparison between two types of models, by plotting VaR of physical and risk neutral MN models. The key difference between risk-neutral and physical models is the former's ability to react faster to the changing markets conditions thanks to ability to incorporate future market beliefs. Put differently, the wedge between physical and risk-neutral models is not constant and reflects changes in financial markets,

Table 2.6: Unconditional coverage property

$$\alpha = 1\%$$

	1 day	5 days	10 days	20 days
MN(4) w VOL under Q	3.8	1.0	1.0	1.0
Gaussian w VOL under Q	3.8	1.0	1.4	1.9
MN(4) w VOL	3.8	1.0	2.9	3.8
Gaussian w VOL	3.8	1.4	2.9	4.3
MN(4)	3.4	1.0	2.9	3.8
Gaussian	3.4	1.0	2.9	4.3

$$\alpha = 5\%$$

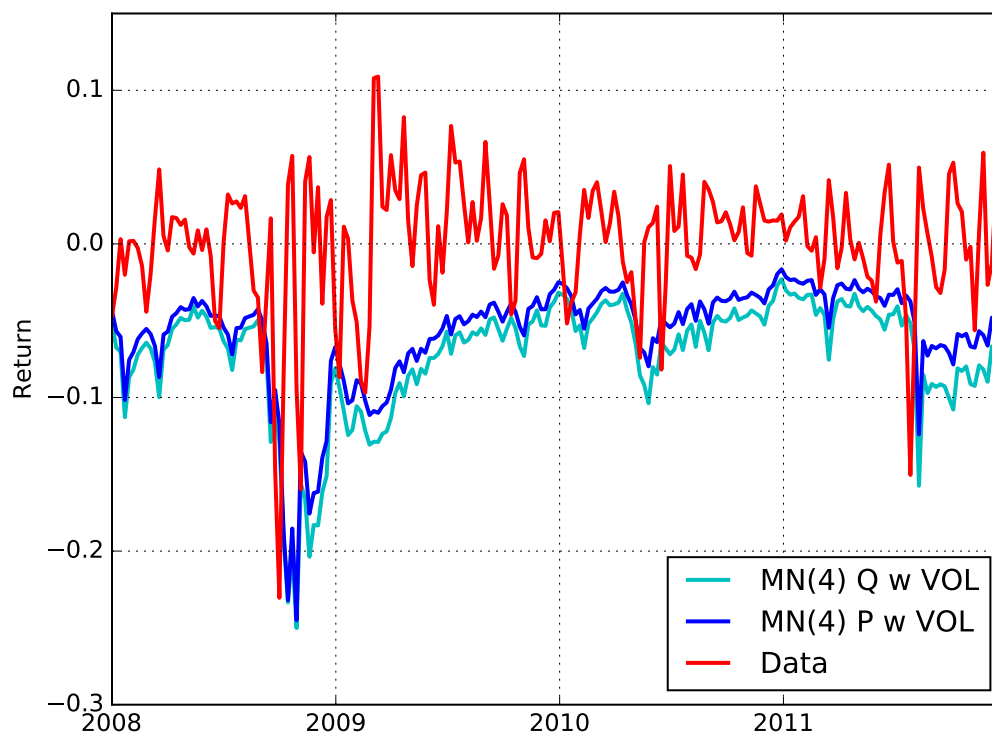
	1 day	5 days	10 days	20 days
MN(4) w VOL under Q	11.5	6.3	3.8	5.8
Gaussian w VOL under Q	11.1	6.3	3.8	6.7
MN(4) w VOL	11.5	9.6	8.7	8.2
Gaussian w VOL	11.5	9.6	8.7	8.2
MN(4)	11.5	8.7	9.1	8.2
Gaussian	11.5	8.7	8.7	8.7

$$\alpha = 10\%$$

	1 day	5 days	10 days	20 days
MN(4) w VOL under Q	14.9	12.5	9.6	9.6
Gaussian w VOL under Q	13.9	11.5	9.6	9.6
MN(4) w VOL	15.4	15.4	13.9	15.4
Gaussian w VOL	15.9	15.4	13.5	13.9
MN(4)	15.4	14.9	13.0	14.9
Gaussian	15.9	14.9	13.0	14.9

Reported are comparison of values of hit functions for Mixture of Normal model with volatility effect under risk neutral measure, Gaussian model with volatility effect under risk neutral measure, Mixture of Normal model with volatility effect under objective measure, Gaussian model with volatility effect under objective measure, Mixture of Normal model under objective measure and Gaussian model under objective measure. The first, second and third panels display values of hit functions for 1%, 5% and 10% levels, correspondingly.

Figure 2.5: Value at Risk for 10 days at 5%

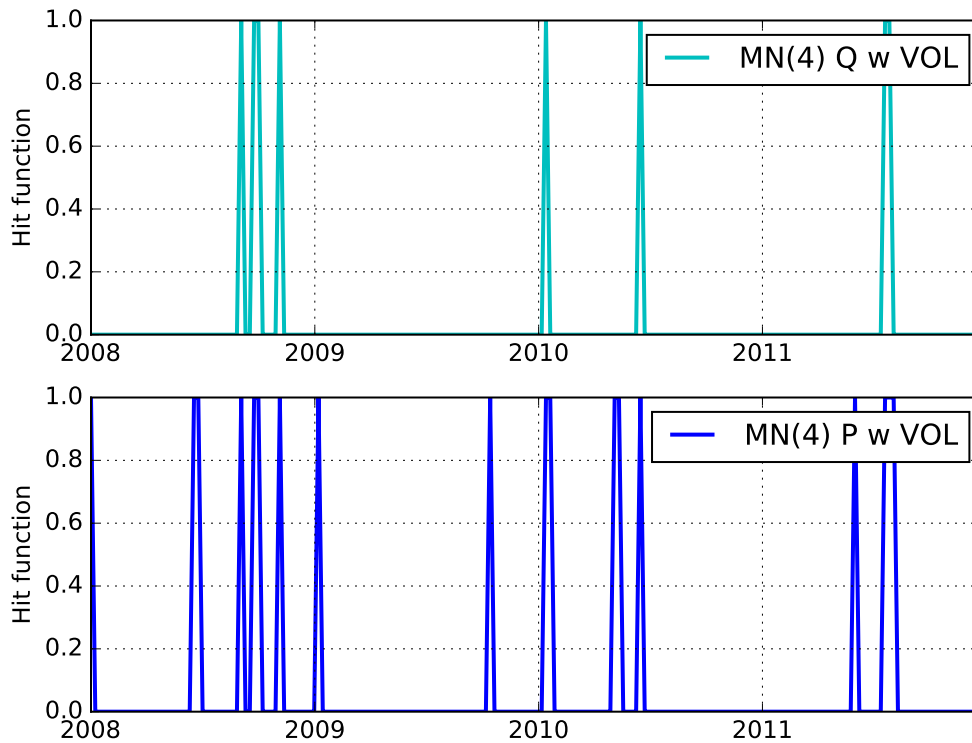


Analysis of Value at Risk measures generated by MN models. We plot aggregate returns over 10 days (red line), Value at Risk at 5% obtained from MN models with volatility effect under objective (blue line) and risk neutral (cyan line) measures, correspondingly. The time span covers periods from 2008 until 2011, overall 208 weekly observations.

while physical model react with lag to surging volatility.

Finally, we check the independence property by plotting hit functions of two models in Figure 2.6. Both models allow clustering of hit functions in the late 2008, which suggests models misspecification. At the same time, this result is not surprising since all our models are linear and do not allow for regime switching which is apparent during financial crises. More importantly, the timing of structural break is extremely hard to identify in out-of-sample exercise.

Figure 2.6: Hit functions over 10 days at 5%



Analysis of the hit functions generated by the risk-neutral and physical MN models. We plot hit functions obtained from MN models with volatility effect under risk neutral (cyan line) at the top graph, and under physical measures (blue line) on the bottom graph, correspondingly. The time span covers periods from 2008 till 2011, overall 208 weekly observations.

To sum-up, our risk-neutral models outperform physical models in terms of unconditional coverage property due to ability to incorporate future beliefs, while the benefits of MN model are less evident compared with Gaussian model. Meanwhile, all models can not successfully reproduce independence property, which might be resolved with regime-switching model. We leave this opportunity for the future research.

2.5.2 Forecasting conditional mean of realized volatility

Next, we assess forecasting performance of risk-neutral and physical models in terms of predicting conditional mean of realized volatility. As we discussed before, our risk-neutral prediction is impeded by the presence of risk premia, which leads to the biased estimate of realized volatility's conditional mean. Thus, we have to make a bias-variance tradeoff: risk-neutral model produces lower variance (due to potential ability to predict structural breaks in volatility), but also leads to the bias caused by risk premia. As a result, our choice of the most accurate model depends on the forecasting horizon as illustrated in Table 2.7. It compares aggregate forecast of realized volatility based on physical MN and Gaussian models with and without volatility effect, together with risk-neutral MN and Gaussian specifications. We use simulations to compute multiple-steps ahead forecast since the closed form expression becomes cumbersome. We assess forecasting performance of different models by computing the root mean square error (RMSE) and mean absolute error (MAE):

$$\begin{aligned}
 e_{t+\tau|t} &= Y_{t+\tau} - Y_{t+\tau|t} \\
 RMSE &= \sqrt{\frac{\sum_{j=t+1}^{t+N} e_{j+\tau|j}^2}{N}} \\
 MAE &= \frac{\sum_{j=t+1}^{t+N} |e_{j+\tau|j}|}{N},
 \end{aligned} \tag{2.23}$$

where $Y_{t+\tau|t}$ is the τ day-ahead conditional forecast of the aggregate realized volatility computed based on either physical or risk-neutral model, and $Y_{t+\tau}$ is the value of aggregate realized volatility over period τ .

Table 2.7: Comparison of volatility forecasts

		MN(4) w VOL Q	Gaussian w VOL Q	MN(4) w VOL P	Gaussian w VOL P	MN P	Gaussian P
RMSE	1 day	1.00	1.00	1.00	1.01	1.00	1.00
RMSE	5 days	0.97	0.98	1.00	0.99	1.00	1.00
RMSE	10 days	0.97	0.98	1.00	0.98	1.00	1.00
RMSE	20 days	0.98	0.99	1.00	0.97	0.99	1.00
MAE	1 day	0.99	1.01	1.00	1.01	1.00	1.00
MAE	5 days	0.99	1.01	1.00	0.99	1.00	1.00
MAE	10 days	1.03	1.04	1.00	1.00	1.00	1.00
MAE	20 days	1.10	1.08	1.00	0.98	1.01	1.00

Reported are comparison of root mean square error (first and second rows) and mean absolute error (third and fourth rows). The time span covers periods from 2008 till 2011, overall 208 weekly observations. Q model column represents the actual value of loss function, while P model column shows ratio between P and Q model forecasts. Thus, the number greater than unity implies an improvement in forecast any model over Gaussian model without volatility effect under physical measure.

Table 2.7 illustrates some benefits of using risk-neutral model in terms of aggregate volatility forecast for short time forecasting horizons – one and five days (one week) ahead. Although, risk-neutral MN model is the best specification for one, five and ten days ahead according to RMSE metric, it is not translated into superior performance judging by MAE metric. Nevertheless, this result reflects the benefits of using forward-looking beliefs and support previous findings, which document that physical HAR model is a tough benchmark to "beat" even with non-linear models— see Scharth and Medeiros (2009), McAleer and Medeiros (2008) for details. In contrast to these studies, we assume the same functional form of realized volatility (linear) model, but incorporate market expectation of the option traders. Thus, we consider our approach as a complement to the standard non-linear specifications, and conjecture that combining them together can achieve more accurate forecast. Meanwhile, risk-neutral model become less accurate (according to MAE metric) compared with physical models as forecasting time horizons rises to ten and twenty days due to stronger affect of risk premia. Interestingly, Gaussian model with volatility effect dominates MN model in terms of forecasting accuracy and remains the best model for long term horizon prediction (for 20 days ahead).

2.6 Simple quantitative trading strategy

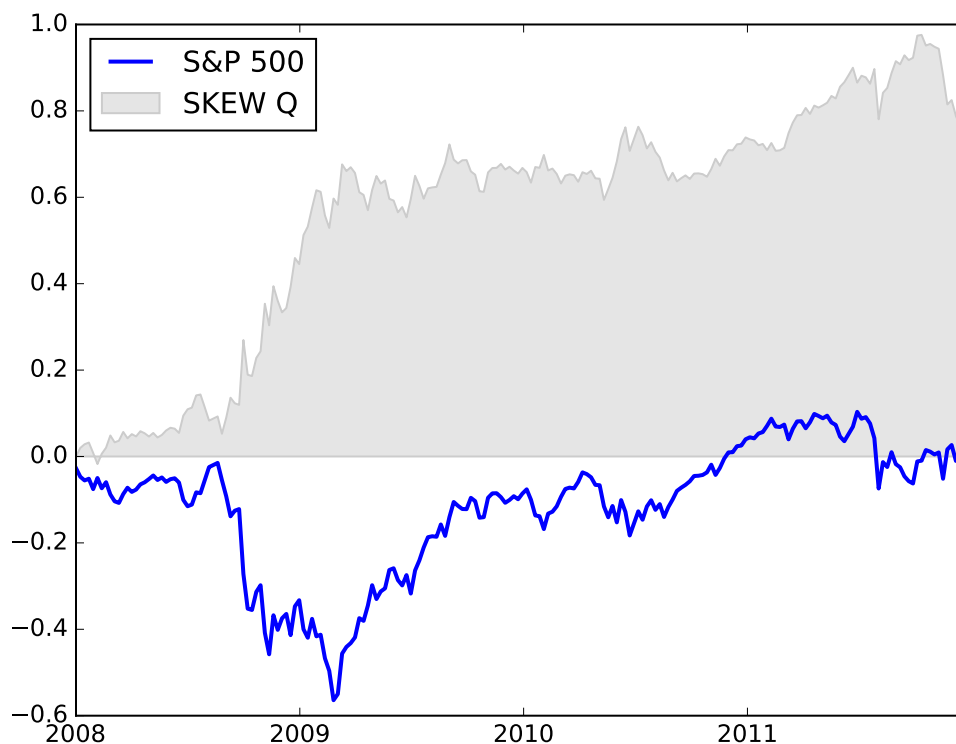
In this section we propose an algorithmic strategy that exploits the evolution of the risk-neutral moments discussed in subsection 2.4.2. We follow a similar idea proposed by Audrino et al. (2015) and build three strategies related to the risk-neutral volatility, skewness and kurtosis, respectively. We go long if the absolute value of either volatility, skewness or kurtosis is larger than last week, and go short if it is lower. The economic intuition be-

hind these strategies reflects the preference of investors. In particular, the investor expects to get higher expected return if financial risk measured by either volatility, skewness or kurtosis increases, which is in line with economic theory (Kimball (1993)). We compare risk-neutral moments and therefore make a portfolio allocation decision (long or short) on a weekly basis, which leads to 135 transactions or 2.7% trading costs by assuming 2 basis-points for transactions costs. Figure 2.7 plots the performance of our strategy (grey shaded area) and compare it with benchmark cumulative returns of the S&P 500 index (blue line) over 2008-2011 years.

Figure 2.7 highlights several appealing features of our strategy. First, it is market neutral as correlation with S&P return is less than 1%. Second, our strategy has the same volatility over 2008-2011 years measured by sample standard deviation as S&P 500 returns. Third, it delivers 76% growth over four years from 2008 to 2011 years, while index was essentially flat (2%). In contrast, Audrino et al. (2015) recovered aggregated moments over fixed interval (30 days) given the structure of the option data set. Although our approach is model-implied we can generate any term structure of implied moments, while model-free method is restricted by options availability, especially for the short time horizon.

Fourth, our strategy generated positive returns during challenging times in the end of 2008 and early 2009, while S&P 500 returns plunged. Meanwhile, it continued to work during market rally started in 2009, while skewness strategy performed poorly in late 2010 and 2011. Fifth, we consider strategy based on the risk-neutral skewness only, because volatility and kurtosis strategies deliver much lower and more unstable returns. The latter result is consistent with findings of Audrino et al. (2015) and might reflect the imprecise estimates of the fourth moment due to the presence of outliers. Finally, we might consider several approaches to improve the performance of our strategy by adding higher leverage in

Figure 2.7: Trading strategy based on a risk neutral skewness



This graph compares cumulative returns of our trading strategy based on risk neutral skewness (shaded area) and buy and hold S&P 500 index strategy (blue line), correspondingly. The risk-neutral skewness is computed over 10 days (2 weeks). The time span covers periods from 2008 till 2011, overall 208 weekly observations.

the position once trading signal becomes stronger, e.g. measured by large increase/decrease in the skewness dynamics. Alternatively, we might combine skewness strategy with other moments or S&P 500 index to avoid negative performance during second half of 2010 and early 2011 years.

To sum-up, we propose simple algorithmic strategies, which delivers substantial returns over the most turbulent years in the recent history. Our strategy is consistent with economic theory and demonstrates the importance of forward looking information embedded in the

option prices.

2.7 Conclusion

This chapter develops a novel discrete-time model for the asset return based on the high-frequency data and mixture of normal distributions of the latent volatility. We use expectation maximization algorithm to avoid problems with direct maximization of mixing likelihood, which allows us to obtain robust estimates. Our model accurately replicates distributions of both returns and realized volatility under physical measure as we demonstrated in our simulation exercise. To compute option prices, we specify a Radon-Nikodym derivative, which includes both Gaussian and non-normal innovations, correspondingly. We identify one parameter of the pricing kernel from EMM condition, while obtain another parameter by fitting option prices for each week. Crucially, our approach avoids calibration of all model's parameters. We price European Put options using Monte Carlo simulations and assess pricing performance of MN and nested Gaussian models during turbulent financial markets in 2008-2011 years.

Our MN model does not only substantially reduce option pricing errors (by around 50% for out of the money Put options) compared with Gaussian model, but provides an appealing econometric framework to assess evolution of investors' risk. Specifically, we distinguish between level of the current uncertainty (volatility) and fear of future financial crash (skewness and kurtosis) and show how dynamics of these measures differ in several financial turmoil episodes. Next, we show a novel approach for predicting returns distribution measured by VaR and forecast of realized volatility's conditional mean by exploiting informational content of option prices and MN model. Finally, we build a simple quanti-

tative strategy, which substantially outperforms returns of S&P 500 index (76% compared with 2%) during turbulent 2008-2011 years, while remaining market-neutral and has the same volatility as a benchmark returns.

The promising performance of the mixture of normal model might warrant further examination. First, we might consider non-linear Radon-Nikodym derivative suggested by Byun et al. (2015) or non-linear pricing kernel proposed by Christoffersen et al. (2013a), and Babaoglu et al. (2014). These studies showed that option pricing models based on daily GARCH specifications outperform corresponding models with linear pricing kernels. Thus, it would be interesting to incorporate non-linear Radon-Nikodym derivative in our framework and study its effect on the option pricing errors.

Second, mixture of exponential power distributions allows to improve volatility modeling compared with mixture of normal distributions (Rombouts and Bouaddi (2009)). In addition, the former requires less components, which is important for econometric tractability. The interesting question is how to translate superior statistical volatility forecast into economic gains measured either by lower option pricing errors or higher returns of the quantitative strategy.

Finally, the application of our model in predicting realized volatility should deserve further investigation. One extension is to produce density forecast of realized volatility and compare it with corresponding forecast based on the physical model. We conjecture that combining high frequency data and flexible mixture distribution with market expectations should lead to more accurate density forecast. This is especially important for predicting structural breaks in realized volatility associated with financial crises. We leave this question for future research.

Chapter 3

Global Factors and Common Idiosyncratic Variance in Exchange Rates Volatility

3.1 Introduction

Verdelhan (2015) documents that two factors, dollar and carry, can summarize contemporaneous variations across a wide range of foreign exchange rates. The dollar and carry factors have meaningful economic interpretations (see Lustig et al. (2011, 2014)). The dollar factor is a return on the equally-weighted portfolio of all currencies relative to the U.S. dollar. The carry factor is a return on a portfolio of high interest rate currencies minus return on a portfolio of low interest rate currencies. This decomposition implies that the volatility of exchange rate returns should be driven by the volatility of the dollar and carry factor returns or by the volatility of country-specific (idiosyncratic) returns.

The first part of the chapter establishes stylized facts about the volatility of exchange rate returns. We find that two distinct but related components summarize 77% of the variations of exchange rate volatility. The first common component corresponds to the volatility of the dollar factor. The second component drives common variations in the volatility of country-specific returns. Country-specific returns are largely idiosyncratic, but nothing prevents closer linkage between their volatility. Indeed, we find that the cross-sectional average of the volatility of idiosyncratic returns plays a significant role. This is closely related to results by Herskovic et al. (2016) for equities. We confirm that the same factor structure is found in the exchange rate volatility implied from option prices.

The second part of the chapter introduces a new multivariate factor model of exchange rate returns that is consistent with the evidence. First, exchange-rate returns and their volatility are driven by common factors. Hence, we can include a large number of exchange rates, avoiding the curse of dimensionality. This factor representation can be derived from a parsimonious construction of the international stochastic pricing kernel. Second, the innovations to expected returns and volatility processes are derived from observable exchange rate returns—including dollar and carry. Hence, we bypass the need to filter latent stochastic volatility factors (Diebold and Nerlov, 1989, Mahieu and Schotman, 1994).

The final section of the chapter brings our model to the data. We estimate our model in two ways. First, we estimate the model under physical measure based on the exchange rate returns of 22 countries between November 2005 and November 2014. We compare filtered variances from our multivariate model and univariate Heston and Nandi (2000) GARCH models for each country. Results are very close, except for four countries: Columbia, Chile, Israel and Switzerland. For these countries, we show that exchange rate returns volatility is mostly uncorrelated with global shocks.

A joint factor model of exchange rate returns and option-implied variances has several important applications. First, the model allows us to decompose the risk premium contribution to exchange rate returns. In addition, we can link the risk premium with the variance premium generated by volatility risk in the exchange rate markets.

The rest of the chapter is organized as follows. Section 3.2 establishes systematic variations in exchange rate volatility and its linkages with the volatility of common factors driving exchange rate returns. Section 3.3 introduces a multivariate dynamic no-arbitrage model of exchange rate returns and variances. Section 3.4 discusses estimation methodology, while Section 3.5 presents estimation results under objective measure. Finally Section 3.5.3 concludes.

3.2 Global and idiosyncratic variance factors

This section documents how global and idiosyncratic components in FX returns generate common volatility across currency and generate low-dimensional risk-returns tradeoffs. We start from the well-known linear factor model for FX returns proposed in Verdelhan (2015):

$$\Delta s_{i,t+1} = \alpha_i + \beta_i^\top f_{t+1} + \epsilon_{i,t+1}, \quad (3.1)$$

where $\Delta s_{i,t+1}$ is a log change in the nominal exchange rate of currency i at time $t + 1$, f_{t+1} includes global dollar and carry factors, β_i are factor loadings, and $\epsilon_{i,t+1}$ is the idiosyncratic or country-specific shock. The conditional variance of $\Delta s_{i,t+1}$ follows:

$$\text{var}_t[\Delta s_{i,t+1}] = \beta_i^\top \text{var}_t[f_{t+1}] \beta_i + \text{var}_t[\epsilon_{i,t+1}], \quad (3.2)$$

where var_t is the conditional variance given the information available at time t , and where we assume $\text{cov}_t[f_{t+1}, \epsilon_{i,t} + 1] = 0$ for simplicity. This equation shows that the conditional variance of dollar and carry $\text{var}_t[f_{t+1}]$ are global factors driving the variances of FX returns. The last term $\text{var}_t[\epsilon_{i,t+1}]$ suggests that the conditional variance can be largely idiosyncratic across countries. By contrast, we show that $\text{var}_t[\epsilon_{i,t+1}]$ is highly correlated in the cross-section of currencies. The fact that the country-specific shock $\epsilon_{i,t+1}$ are uncorrelated has no bearing on this. Put it differently, these shocks are only uncorrelated but dependent through their second moments.

3.2.1 Data

We use daily financial data obtained from Bloomberg include data from 22 currencies: Australia, Brazil, Canada, Chile, Columbia, Czech Republic, Denmark, Euro, Hungary, Israel, Japan, Mexico, New Zealand, Norway, Poland, Singapore, South Africa, South Korea, Sweden, Switzerland, Turkey, United Kingdom. We compile the following data for these currencies:

1. bilateral exchange rates with respect to the U.S. dollar (Jul. 2005–Nov. 2014),
2. 3-month LIBOR rates (Jul. 2005–Nov. 2014),
3. 3-month forward exchange rates (Nov. 2005–Nov. 2014),
4. 3-month at-the-money implied volatility (Jul. 2005–Nov. 2014) and
5. 3-month 25 delta risk-reversals (Jul. 2005–Nov. 2014).

3.2.2 Common variance of country-specific Innovations

Next, we document that the variances of country-specific innovations ($\text{var}_t[\epsilon_{i,t+1}]$) share a common factor structure. For this purpose, we compute the innovations $\epsilon_{i,t} + 1$ from Equation (3.1), estimated with OLS with daily data. These innovations are essentially uncorrelated. The first three components (PCs) of $\epsilon_{i,t+1}$ across 22 currencies explain only 19%, 13%, and 8% of the variations, respectively (first column of Table 3.1).

We estimate the conditional variances $\text{var}_t[\epsilon_{i,t+1}]$ of country-specific shocks using an EGARCH(1,1) specification combined with an AR(1) process:

$$\begin{aligned}\epsilon_{i,t+1} &= a + b\epsilon_{i,t} + \nu_{i,t+1} \\ \nu_{i,t+1} &= \sqrt{\text{var}_t[\epsilon_{i,t+1}]}z_{i,t+1} \\ \log(\text{var}_t[\epsilon_{i,t+1}]) &= \omega_i + b_i \log(\text{var}_{t-1}[\epsilon_{i,t}]) + a_i \left[\frac{|\nu_{i,t}|}{\text{var}_{t-1}[\epsilon_{i,t}]} - \sqrt{\frac{\pi}{2}} \right] + \zeta_i \frac{\nu_{i,t}}{\text{var}_{t-1}[\epsilon_{i,t}]}\end{aligned}$$

Principal component analysis reveals a strong factor structure (Table 3.1). The first PC summarizes 72% of total variations. This PC is essentially the average of $\text{var}_t[\epsilon_{i,t+1}]$ across countries: its loading are spread more or less uniformly across exchange rates. The correlation between the first PC and the cross-sectional average is 0.98.

3.2.3 Dollar and carry conditional variance

Next, we ask whether the variance of global factors $\text{var}_t[f_i]$ explains a large proportion of variations in $\text{var}_t[\Delta s_{i,t+1}]$. Again, we use a simple EGARCH(1,1) model combined with an AR(1) process to estimate the variance $\text{var}_t[f_{t+1}]$ and $\text{var}_t[\Delta s_{i,t+1}]$ for each currency.¹⁹

¹⁹We computed these variances in the same way as variances of the country-specific shocks.

Table 3.1: First component explains most of the total and country-specific variances

	$\epsilon_{i,t+1}$	$\text{var}_t[\epsilon_{i,t+1}]$	$\text{var}_t[\Delta s_{i,t+1}]$
PC1	19%	72%	77%
PC2	13%	6%	9%
PC3	9%	5%	4%

Shares of variations explained by the first three principal components of country-specific shocks $\epsilon_{i,t+1}$, country-specific variances $\text{var}_t[\epsilon_{i,t+1}]$ and total variances $\text{var}_t[\Delta s_{i,t+1}]$. Daily sample between November 2001 and November 2014 and includes 22 countries: Australia, Brazil, Canada, Chile, Columbia, Czech Republic, Denmark, Euro, Hungary, Israel, Japan, Mexico, New Zealand, Norway, Poland, Singapore, South Africa, South Korea, Sweden, Switzerland, Turkey, United Kingdom.

Principal component analysis reveals a strong factor structure across currencies: the first two components explain 77% and 9% of the variations, respectively (See third column of Table 3.1). In turn, these components are closely related to the variance of global FX variance factors. Together, the variances of the dollar factor and the average variance of country-specific shocks capture 93% the variations of the first principal component of $\text{var}_t[\Delta s_{i,t+1}]$ and capture 77% of the total variations in $\text{var}_t[\Delta s_{i,t+1}]$.²⁰

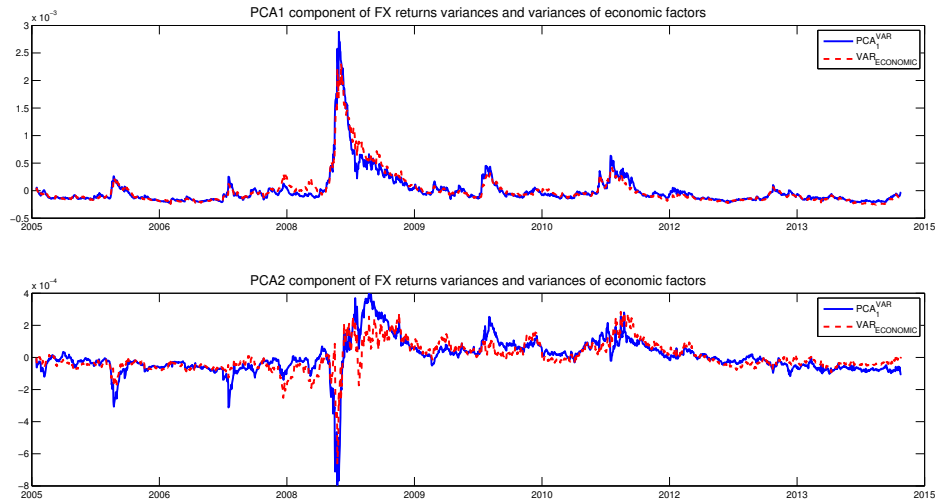
Figure 3.1 compares the first (top) and second (bottom) principal components with our two economic factors. The principal components are plotted in solid blue lines and the regression fits of the principal components on dollar factor are plotted in dashed red lines.²¹ The comparison of the two lines shows a strong co-movement between statistical and economic factors, especially for the first principal component.

We find a very similar factor structure in 3-month at-the-money implied variances of 27 currencies. Principal component analysis reveals that the first two components explain 87% and 5% of the variations, respectively. Moreover, Figure 3 shows that the first component

²⁰We compute contribution as a sum of regression's R^2 weighted by the percentage of the total variance explained by each principal component.

²¹In other words, dash line is a fitted value obtained from regression principal component on variance of dollar factor and constant.

Figure 3.1: PCA1 of exchange rate returns variances, and variances of global and domestic factors



The time-series dynamics of first principal component of FX returns variances (top graph) and first principal component of variances of domestic factors (bottom graph). The time period starts from November 2001 to November 2014.

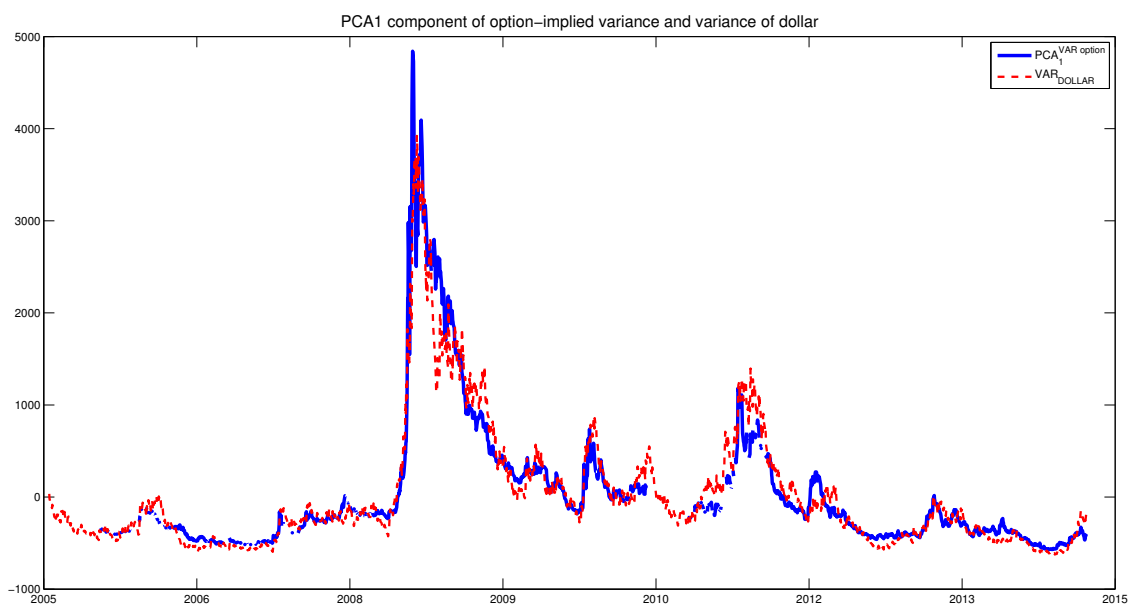
is highly correlated with the variance of the dollar factor as shown in Figure 3.2.

Common factors in other option-implied measures of risk

We also document the presence of similar factor structure in volatility skew and volatility term structure across different currencies.²² Note that volatility skew and volatility term structure are not the main focus of our study. As a proxy for volatility skew, we use 3-month 25-delta risk reversal which is the difference between implied volatility of call and put options with deltas of 0.25 and -0.25, respectively. As a proxy for the volatility term structure, we use the difference between at-the-money implied volatilities with 9-month

²²Christoffersen et al. (2013) find a factor structure in the level of option-implied volatility, volatility skew, and volatility term structure across U.S. equities.

Figure 3.2: PCA1 of implied variances of exchange rate returns, and variances of global and domestic factors



Time series dynamics of first principal component of implied variance (solid blue line) and variance of dollar (red dash). The variance of dollar is available from November 2001 to November 2014, while implied variances are plotted from July 2005 to November 2014.

and 3-month terms.

Table 3.2 reports result from the principal component analysis. The results show that the first principal components explain 71% and 84% of the variations in the volatility skew (risk reversals) and in the volatility term structure, respectively. Although PCA analyses reveal strong factor structures in both volatility skew and volatility term structure of FX options, their formal link is beyond the scope of this chapter.

Table 3.2: Factor Structure in Equity and Exchange Rate Markets

	Equity	FX
Implied volatility	77%	87%
Skewness	77%	71%
Term structure	60%	84%
Idiosyncratic Variance		72%

Share of variations explained by the first principal component of implied volatility, skewness, slope and idiosyncratic variance for equity options and exchange rate options, respectively. The shares of variations for equities are obtained from Christoffersen et al. (2015). The term structure is the difference between implied volatility of 9-months ATM Call options and 3-months ATM Call options.

Robustness

Our main findings remain qualitatively the same after several robustness checks. In particular, the majority of the variations in physical and risk-neutral FX return variances are driven by the variance of the dollar factor and the average variance of country-specific shocks. Specifically, we perform the following checks.

- We consider a sub-sample of only developed countries: Australia, Canada, Denmark, Euro Zone, Israel, Japan, New Zealand, Norway, Sweden, Switzerland, United Kingdom.
- We use squared regression residuals as a model-free proxy of conditional variance of country-specific shocks. PCA documents weaker factor structure in the variances of country-specific shocks, which can be attributed to noisy nature of squared residuals.
- We add the first principal component of the country-specific shocks to the model (1) to capture possible omitted factors. This additional factor is meant to absorb all of the co-movement in country-specific news $\epsilon_{i,t+1}$.

3.3 No-arbitrage dynamic exchange rate model

In this section, we specify a dynamic no-arbitrage model of FX returns and their variance. Our approach combines well-known components. First, FX returns share systematic variations due to their exposure to carry and dollar factors. Second, carry and dollar volatility evolves according to Heston and Nandi (2000) model. Third, idiosyncratic variance across countries share a strong factor structure, that also evolves according to a Heston-Nandi specification.

3.3.1 Individual exchange rate dynamics

We model the FX return of currency i based on the factor structure discussed in the previous section but with time-varying volatility:

$$\Delta s_{i,t+1} = \alpha_i + f_{t+1}^\top \beta_i + \sqrt{\gamma_i (\sigma_\epsilon + h_t^\epsilon)} z_{i,t+1}^{\mathbb{P}} \quad (3.3)$$

$$h_{t+1}^\epsilon = \omega_\epsilon + b_\epsilon h_t^\epsilon + c_\epsilon \left(a_\epsilon^\top z_{t+1}^{\mathbb{P}} - \sqrt{\sigma_\epsilon + h_t^\epsilon} \right)^2, \quad (3.4)$$

where $\Delta s_{i,t+1}$ is the log-return at time $t + 1$, f_{t+1} is a $N \times 1$ vector of FX risk factors, β_i are the factor loadings, $z_{i,t}^{\mathbb{P}} \sim N(0, 1)$ and denoting $z_t^{\mathbb{P}} = [z_{1,t}, \dots, z_{J,t}]$. The country-specific components of FX returns share the same conditional variance factor h_{t+1}^ϵ but scaled by the loading γ_i . The conditional variance has standard dynamics but based on the enlarged information set $a_\epsilon^\top z_t^{\mathbb{P}}$. The $J \times 1$ vector a_ϵ aggregates information from all country-specific innovations to update the conditional variance from t to $t + 1$.²³

²³From a technical standpoint, this generalization does not introduce any difficulty relative to the Heston-Nandi base-case, since the linear combination $a_\epsilon^\top z_{t+1}^{\mathbb{P}}$ is conditional Gaussian.

3.3.2 Exchange rate factor dynamics

The dynamics of the factors f_{t+1} are given in terms of the $N \times 1$ conditional mean m_t and conditional variance h_t :

$$f_{t+1} = m_t + \sqrt{(\sigma + h_t) \Sigma} u_{t+1}^{\mathbb{P}}, \quad (3.5)$$

with $u_{t+1}^{\mathbb{P}} \sim N(0, I_N)$ and the following dynamics:

$$m_t = \phi_0 + \phi_1 m_{t-1} + \Psi(f_t - m_{t-1}) \quad (3.6)$$

$$h_{t+1} = \omega + b h_t + c \left(a^\top u_{t+1}^{\mathbb{P}} - \sqrt{\sigma + h_t} \right)^2. \quad (3.7)$$

The conditional mean m_t has auto-regressive dynamics with innovations $(f_t - m_{t-1})$.

3.4 Estimation

3.4.1 Benchmark univariate model

We define a well-known univariate time-varying volatility model as a benchmark to assess the performance of our encompassing specification. We consider the case where each currency has its own independent Heston-Nandi volatility dynamics. Of course, this greatly expands the number of latent volatility process and provides substantial flexibility in fitting the particular volatility process of each currency. Nonetheless, the results show that a parsimonious model based on common risk factors provides a similar fit. The benchmark

dynamics are given by

$$\Delta s_{i,t+1} = \alpha_i + f_{t+1}^\top \beta_i + \sqrt{h_{i,t}} z_{i,t+1}^{\mathbb{P}} \quad (3.8)$$

$$h_{i,t+1} = \omega_i + b_i h_{i,t} + c_i \left(a_i z_{i,t+1}^{\mathbb{P}} - \sqrt{h_{i,t}} \right)^2, \quad (3.9)$$

which is estimated based on standard routines for each currency individually.

3.4.2 Exchange rate likelihood

Our sample has 22 currencies ($J = 22$) and we use the dollar and carry factors as observable risk factors ($N = 2$). We estimate parameters of exchange rate dynamics based on the joint conditional likelihood of the data:

$$\begin{aligned} \ln L^{\mathbb{P}} &= \sum_t \ln \left[l_t(\Delta s_{t+1}, f_{t+1} | I_t) \right] \\ l(\Delta s_{t+1}, f_{t+1} | I_t) &= l(\Delta s_{t+1} | f_{t+1}, I_t) \times l(f_{t+1} | I_t), \end{aligned} \quad (3.10)$$

where Δs_{t+1} stacks the J exchange rate returns and where the information set I_t contains the history of Δs_t and f_t until time- t . The first term in the conditional likelihood is given by:

$$\begin{aligned} l(\Delta s_{t+1} | f_{t+1}, I_t) &= (2\pi)^{-\frac{J}{2}} \left| \Sigma_t^{\Delta s} \right|^{-\frac{1}{2}} \\ &\times \exp \left[-\frac{1}{2} (\Delta s_{t+1} - E[\Delta s_{t+1} | f_{t+1}])^\top (\Sigma_t^{\Delta s})^{-1} (\Delta s_{t+1} - E[\Delta s_{t+1} | f_{t+1}]) \right], \end{aligned} \quad (3.11)$$

with $E[\Delta s_{t+1}|f_{t+1}] = A + Bf_{t+1}$ and $\Sigma_t^{\Delta s_{t+1}} = (\sigma_\epsilon + h_t^\epsilon)\Gamma_1$ where B and Γ stacks the corresponding loadings β_i and γ_i . The second conditional likelihood is given by:

$$f(f_{t+1}|I_t) = (2\pi)^{-\frac{N}{2}} |\Sigma_t|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(f_{t+1} - m_t)' (\Sigma_t)^{-1} (f_{t+1} - m_t)\right], \quad (3.12)$$

with $\Sigma_t = (\sigma_F + h_t^F)\Sigma_1$.

3.4.3 Targeting unconditional moments

The parameters to be estimated for each currency are $\alpha_i, \beta_i, \gamma_i$. The additional common parameters are $\sigma_\epsilon, \omega_\epsilon, b_\epsilon, c_\epsilon$ and a_ϵ . Parameters of the factor dynamics are $\sigma, \Sigma, \phi_0, \phi_1, \phi_f, \phi_\epsilon, \Psi$. Parameters of the factor variance are ω, b, c and a . For each currency, starting value of α_i, β_i can be obtained from OLS since carry and dollar factors are observable. Starting values of country-specific variance parameters can also be obtained based on the OLS residuals $\epsilon_{i,t}$. Similarly, starting values for parameters of the risk factors, since carry and dollar returns are observable. Convergence of the joint likelihood appears robust to varying these starting values. Nonetheless, we target some of the parameters to the following unconditional moments of the data.

Factors mean

From the factor dynamics, and taking unconditional expectation, we obtain

$$\begin{aligned} E[m_{t+1}] &= \phi_0 + \phi_1 E[m_t] \\ \phi_0 &= (I_N - \phi_1)E[f_{t+1}], \end{aligned} \quad (3.13)$$

where the unconditional mean of h_t^ϵ and h_t are given by

$$\begin{aligned} E[h_t^\epsilon] &= \frac{\omega_\epsilon + c_\epsilon(a_\epsilon^\top a_\epsilon + \sigma_\epsilon)}{1 - (b_\epsilon + c_\epsilon)} \\ E[h_t] &= \frac{\omega + c(a^\top a + \sigma)}{1 - (b + c)}, \end{aligned} \quad (3.14)$$

and where an estimate of $E[f_{t+1}]$ is given by the sample mean of f_{t+1} .

Country-specific variance

From the volatility of country-specific innovations, and taking unconditional expectations, we have

$$\gamma_i = \frac{\text{var}[\epsilon_{i,t+1}]}{\sigma_\epsilon + E[h_t^\epsilon]}, \quad (3.15)$$

and substituting for the unconditional mean of h_t^ϵ we obtain:

$$\gamma_i = \frac{\text{var}[\epsilon_{i,t+1}]}{\sigma_\epsilon + \frac{\omega_\epsilon + c_\epsilon(a_\epsilon^\top a_\epsilon + \sigma_\epsilon)}{1 - (b_\epsilon + c_\epsilon)}}, \quad (3.16)$$

where an estimate of $\text{var}[\epsilon_{i,t+1}]$ is easily obtained from the sample variance of the residuals in a regression of $\Delta s_{i,t+1}$ on the factors f_{t+1} .

Factors variance

Applying law of total variance we have

$$\begin{aligned}\text{var}[f_{t+1}] &= \text{var}[E_t(f_{t+1})] + E[\text{var}_t(f_{t+1})] \\ &\approx E[\text{var}_t(f_{t+1})] = (\sigma_F + E[h_t^F])\Sigma_1^P,\end{aligned}\tag{3.17}$$

where $\text{var}[E_t(f_{t+1})]$ is of second-order importance in practice. We obtain:

$$\Sigma_1^P = \frac{\text{var}[f_{t+1}]}{\sigma_F + \frac{\omega+c(a\tau a+\sigma)}{1-(b+c)}},\tag{3.18}$$

where an estimate of $\text{var}[f_{t+1}]$ is easily obtained from the sample variance of the factors f_{t+1} , which guarantee that Σ_1^P is positive definite.

3.5 Results

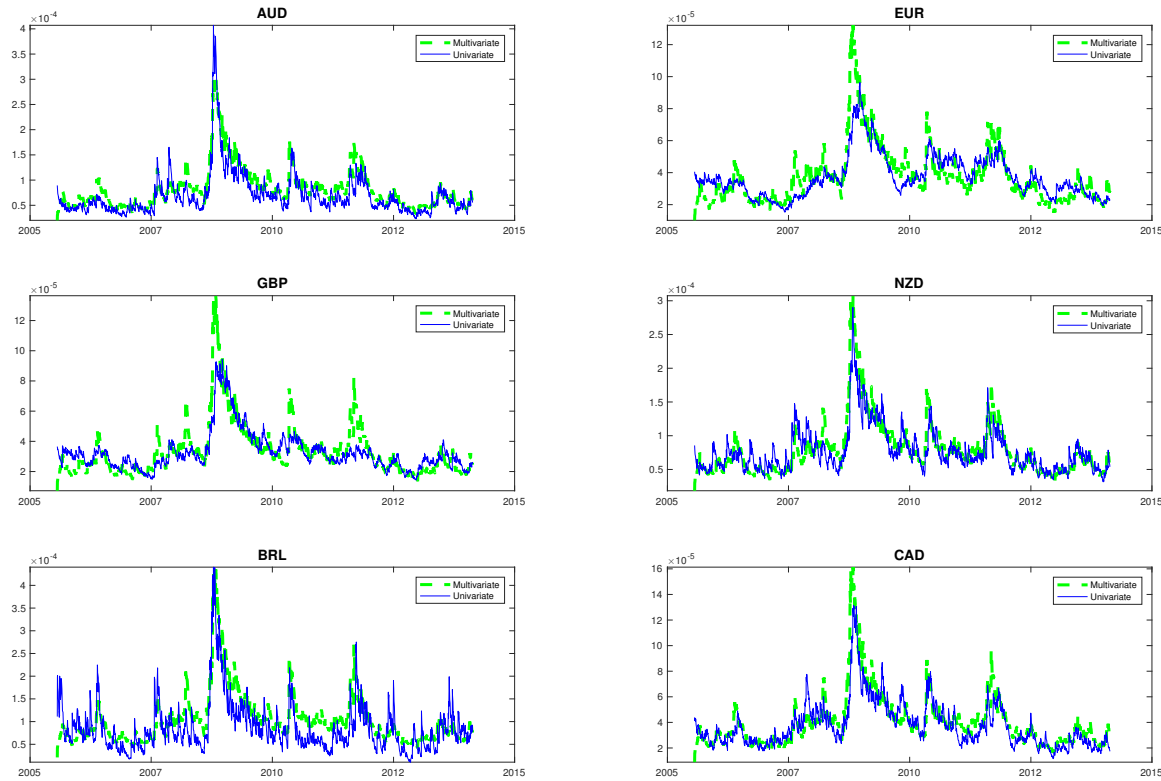
Overall, the performance of our multivariate model is comparable to that of the benchmark. In particular, a factor structure allows us to closely replicate filtered univariate variances with the exception of few emerging countries.

3.5.1 Volatility dynamics

We compare the volatility dynamics between the benchmark model and our specification. We filter the conditional volatility for each currency separately based on the benchmark model. Next, we compute the conditional variance implied from our factor model based on the filtered volatility factors h_t and h_t^ϵ . Figures 3.3-3.6 show the results. Overall, our

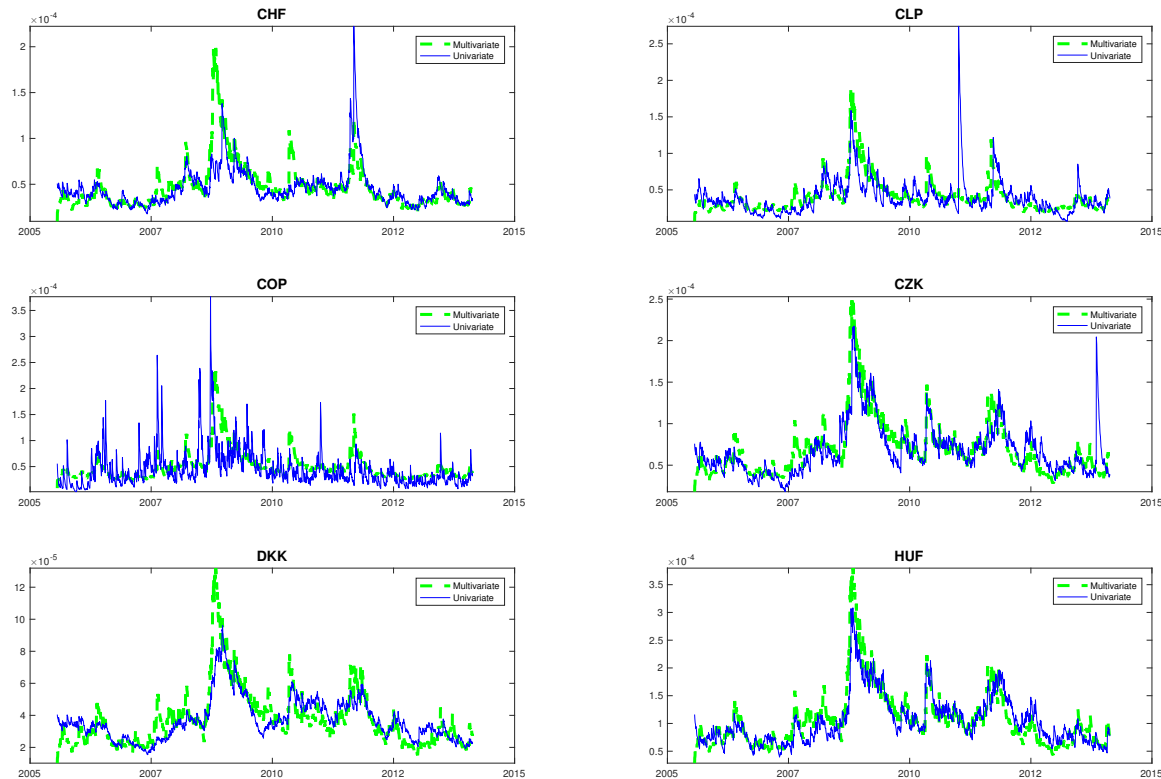
specification does an excellent job in terms of matching the variances. A few stand out: Columbia, Israel, Chile and Switzerland. This result is consistent with the stylized facts documented before, including low persistence of variance process for Columbia (see Table 3.3) and low share of individual country-specific variances attributed to the common factor (see Figure 3.7).

Figure 3.3: Comparison of multivariate and univariate variances. Part I



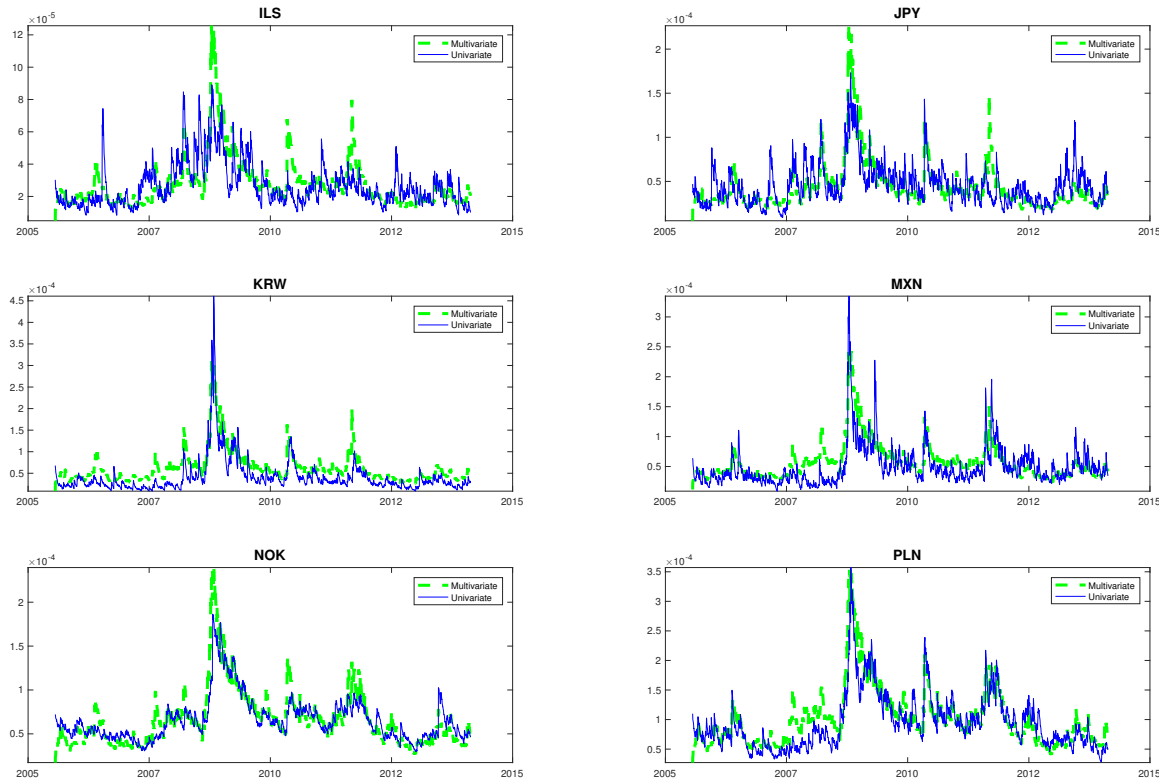
Comparison of univariate variance (solid blue line) and variance implied from multivariate model (dash green) The former is obtained from HN-GARCH model fitted separately for each FX return, while the latter is computed from our factor model. The time period starts from July 2005 to November 2014, overall 2356 observations. The following countries are included: Australia, Euro Zone, United Kingdom, New Zealand, Brazil, Canada.

Figure 3.4: Comparison of multivariate and univariate variances. Part II



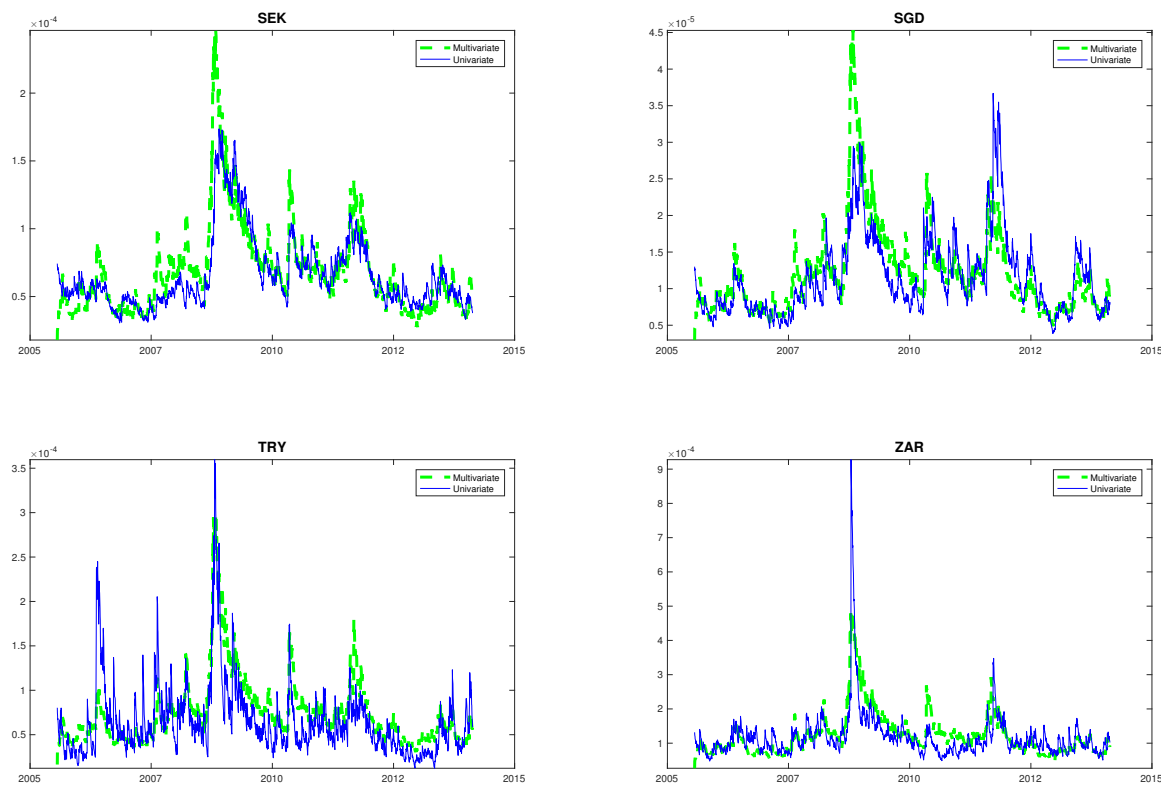
Comparison of univariate variance (solid blue line) and variance implied from multivariate model (dash green) The former is obtained from HN-GARCH model fitted separately for each FX return, while the latter is computed from our factor model. The time period starts from July 2005 to November 2014, overall 2356 observations. The following countries are included: Switzerland, Chile, Columbia, Czech Republic, Denmark and Hungary.

Figure 3.5: Comparison of multivariate and univariate variances. Part III



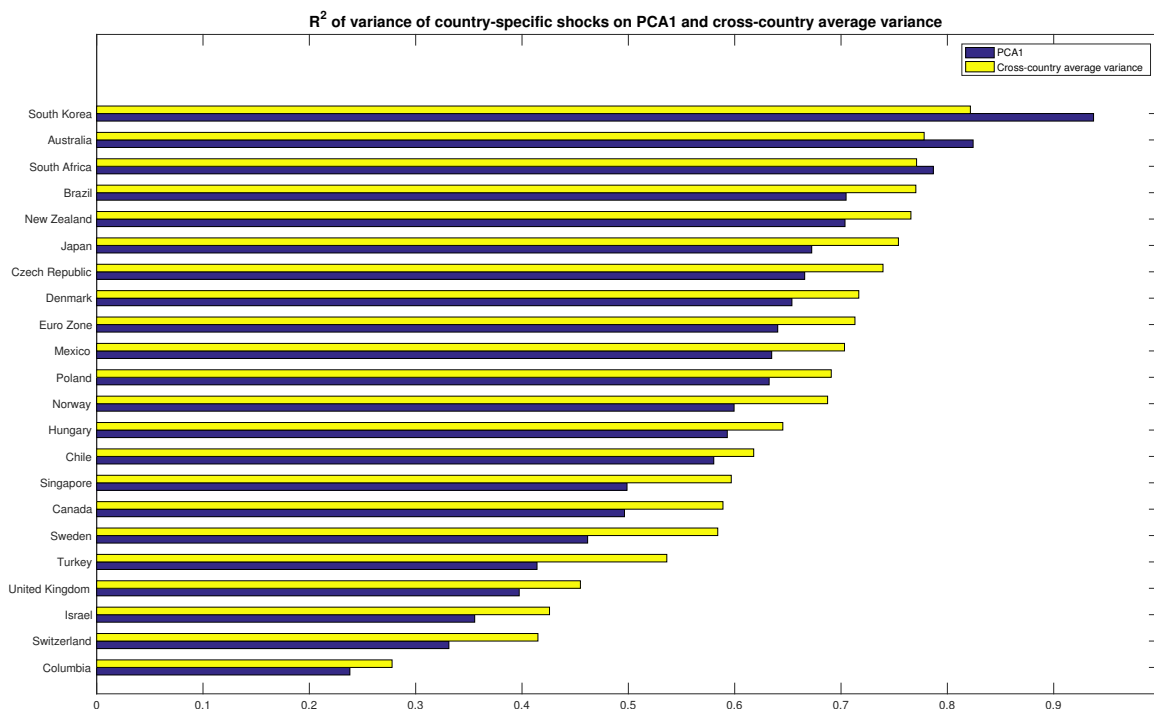
Comparison of univariate variance (solid blue line) and variance implied from multivariate model (dash green). The former is obtained from HN-GARCH model fitted separately for each FX return, while the latter is computed from our factor model. The time period starts from July 2005 to November 2014, overall 2356 observations. The following countries are included: Israel, Japan, South Korea, Mexico, Norway and Poland.

Figure 3.6: Comparison of multivariate and univariate variances. Part IV



Comparison of univariate variance (solid blue line) and variance implied from multivariate model (dash green). The former is obtained from HN-GARCH model fitted separately for each FX return, while the latter is computed from our factor model. The time period starts from July 2005 to November 2014, overall 2356 observations. The following countries are included: Sweden, Singapore, Turkey and South Africa.

Figure 3.7: Share of individual country-specific variance attributed to common factor

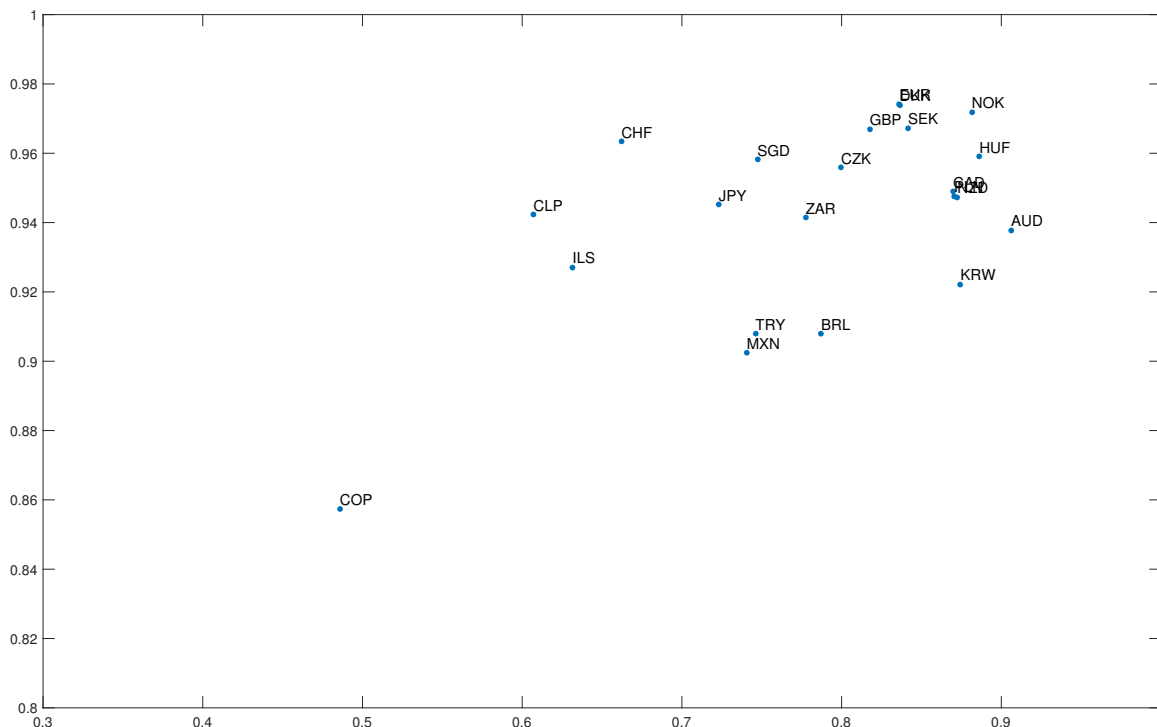


R^2 from regressions of individual country-specific variance on the first principal component (blue bar) or on the cross-country average (yellow bar) of all country-specific variances. The time period starts from November 2001 to November 2014 and includes 22 countries: Australia, Brazil, Canada, Chile, Columbia, Czech Republic, Denmark, Euro, Hungary, Israel, Japan, Mexico, New Zealand, Norway, Poland, Singapore, South Africa, South Korea, Sweden, Switzerland, Turkey, United Kingdom.

3.5.2 Persistence

A small number of emerging countries stand out and do not seem to be captured well. To better understand this challenge, Table 3.3 reports the correlation between volatility estimates. The low-correlation countries (correlation below 70%) are the expected culprits given the results above: Columbia, Chile, Israel and Switzerland. In case of Columbia, the low correlation appears caused by the much lower persistence in the benchmark volatility (second column of Table 3.3). This poses a challenge to our specification where assume

Figure 3.8: The link between univariate model's persistence and correlation between model-implied and univariate variances



This scatter plot compares correlation between model-implied and univariate variances (horizontal axis) and persistence of univariate model (vertical axis). The time period starts from November 2001 to November 2014 and includes 22 countries: Australia, Brazil, Canada, Chile, Columbia, Czech Republic, Denmark, Euro, Hungary, Israel, Japan, Mexico, New Zealand, Norway, Poland, Singapore, South Africa, South Korea, Sweden, Switzerland, Turkey, United Kingdom.

that the volatility persistence variance is the same across currencies (see Figure 3.8 for details). The Colombian Peso appears to truly stand out due to the large spikes in volatility during 2007-2008 years driven by peso depreciation.

3.5.3 Diagnostic checks

Finally, we check whether our multivariate and each univariate models are able to get rid off any remaining conditional volatilities in the standardized returns. For this purpose, we test whether second moments of innovations have any remaining autocorrelations. Table

Table 3.3: Comparison of multivariate and univariate variances

	Correlation	Persistence	Sensitivity to factor's variance (γ_i)
Australia	0.91	0.94	3.82
Euro Zone	0.84	0.97	1.20
United Kingdom	0.82	0.97	2.11
New Zealand	0.87	0.95	4.21
Brazil	0.79	0.91	7.30
Canada	0.87	0.95	2.41
Switzerland	0.66	0.96	3.34
Chile	0.61	0.94	3.61
Columbia	0.49	0.86	4.64
Czech Republic	0.80	0.96	2.67
Denmark	0.84	0.97	1.21
Hungary	0.89	0.96	3.68
Israel	0.63	0.93	2.37
Japan	0.72	0.95	5.00
South Korea	0.87	0.92	6.56
Mexico	0.74	0.90	4.11
Norway	0.88	0.97	2.51
Poland	0.87	0.95	3.33
Sweden	0.84	0.97	2.53
Singapore	0.75	0.96	0.55
Turkey	0.75	0.91	4.55
South Africa	0.78	0.94	7.36

The second column of Table 3.3 displays correlation between univariate variance and variance implied from factor model, correspondingly. Third column shows persistence of each univariate model computed as $b + c$. Fourth column shows sensitivity of each FX variance to the common idiosyncratic variance and scaled by 10^{-6} . The time period starts from July 2005 to November 2014, overall 2356 observations. Countries with correlation less or equal than 70% between univariate and multivariate models are bolded.

3.4 shows p-values of Ljung-Box test for squared returns (columns 1 and 2), squared returns standardized by either univariate (columns 3 and 4) or multivariate (columns 5 and 6) variances. Both univariate models and our joint specification successfully remove conditional variations in the standardized returns approximately for half of all currencies. By contrast, all exchange rate returns exhibit conditional variations in their second moments.

Table 3.4: Ljung-Box test for remaining autocorrelation in squared exchange rate returns and standardized returns

	$\Delta s_{i,t}^2$		$\left(\frac{\Delta s_{i,t}}{h_{univ}}\right)^2$		$\left(\frac{\Delta s_{i,t}}{h_{mult}}\right)^2$	
	1 lag	5 lags	1 lag	5 lags	1 lag	5 lags
Australia	0.00	0.00	0.00	0.00	0.00	0.00
Euro Zone	0.00	0.00	0.66	0.21	0.94	0.61
United Kingdom	0.00	0.00	0.14	0.00	0.71	0.02
New Zealand	0.00	0.00	0.00	0.00	0.00	0.00
Brazil	0.00	0.00	0.01	0.00	0.00	0.00
Canada	0.00	0.00	0.06	0.00	0.00	0.00
Switzerland	0.01	0.00	0.35	0.28	0.43	0.09
Chile	0.00	0.00	0.41	0.45	0.01	0.00
Columbia	0.00	0.00	0.00	0.00	0.00	0.00
Czech Republic	0.00	0.00	0.04	0.00	0.07	0.00
Denmark	0.00	0.00	0.63	0.26	0.87	0.66
Hungary	0.00	0.00	0.73	0.00	0.50	0.00
Israel	0.00	0.00	0.01	0.00	0.00	0.00
Japan	0.00	0.00	0.90	0.91	0.07	0.01
South Korea	0.00	0.00	0.00	0.00	0.00	0.00
Mexico	0.00	0.00	0.06	0.00	0.00	0.00
Norway	0.00	0.00	0.00	0.00	0.03	0.27
Poland	0.00	0.00	0.00	0.00	0.00	0.00
Sweden	0.00	0.00	0.49	0.00	0.30	0.00
Singapore	0.00	0.00	0.39	0.00	0.42	0.00
Turkey	0.00	0.00	0.00	0.00	0.00	0.00
South Africa	0.00	0.00	0.71	0.02	0.13	0.00

This table shows p-values of Ljung-Box test for squared exchange rate returns (second and third columns), squared exchange rate returns standardized by univariate (fourth and fifth columns) and multivariate (sixth and seventh columns) volatilities.

Conclusion

We propose a new multivariate factor model of exchange rate returns and their option-implied variances. This model starts from a simple linear factor model for exchange rate returns whose factors have economic underpinnings such as Verdelhan (2015), then es-

establish a factor structure in their variances. We show that the common factors driving variances of exchange rate returns include the variances of global factors (e.g. variance of dollar factor in Verdelhan (2015) model) and the common factors driving variances of country-specific shocks.

Our model has several appealing features. First, we can include a large number of exchange rates, avoiding the curse of dimensionality, which is crucial for a portfolio construction. Second, the innovations to expected returns and volatility processes are derived from observable exchange rate returns without relying on filtering of latent stochastic volatility factors. We find that our joint model produces very similar dynamics of objective variance comparing with individual univariate models, with exception of few countries.

The promising performance of our multivariate model might warrant further examination. First, it allows us to decompose the risk premium contribution to the exchange rate returns. In other words, by providing the model's specification under both physical and risk-neutral measures, we open the door for examining risk premia in the currency spot and option markets. However, detailed theoretical and empirical investigation of this issue is left for future research. Second, our model can be used to devise a better portfolio construction or hedging strategy for a portfolio containing both currencies and currency options. This has an important application for currency traders, who lost around 30% on their carry trades due to the common depreciation of the majority of currencies against US dollar in the late 2008. By contrast, our model provides an interaction among different currencies, which should be helpful to avoid such losses. We conjecture that it will outperform competing univariate benchmarks during periods of financial crises. We leave its empirical examination of our model's performance for the future research.

Appendix A

Chapter 2 Appendix

A.1 m -step-ahead forecast

First, the GARCH(1,1) model is defined as:

$$\begin{cases} r_t = \mu + \epsilon_t \\ \epsilon_t = z_t \sqrt{h_t}, \quad z_t \sim i.i.d. \quad N(0,1) \\ h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \end{cases} \quad (0.1)$$

The m -step-ahead forecast is computed according to:

$$\begin{aligned} \hat{h}_{t+m|t} &= \omega + \alpha_1 \hat{\epsilon}_{t+m-1|t}^2 + \beta_1 \hat{h}_{t+m-1|t} \\ \hat{\epsilon}_{t+m|t}^2 &= \hat{h}_{t+m|t} \quad \text{if } m > 0 \\ \hat{\epsilon}_{t+m|t}^2 &= \epsilon_{t+m}^2, \quad \hat{h}_{t+m|t} = h_{t+m}, \quad \text{if } m \leq 0. \end{aligned} \quad (A.1)$$

Second, the GJR-GARCH(1,1) model is defined as:

$$\begin{cases} r_t = \mu + \epsilon_t \\ \epsilon_t = z_t \sqrt{h_t}, \quad z_t \sim i.i.d. \quad N(0,1) \\ h_t = \omega + \alpha_1 \epsilon_{t-1}^2 \cdot [1 - \mathbb{1}(\epsilon_{t-1} > 0)] + \gamma_1 \epsilon_{t-1}^2 \cdot \mathbb{1}(\epsilon_{t-1} > 0) + \beta_1 h_{t-1}. \end{cases}$$

The recursive formula for the multiple-step-ahead forecast of the GJR-GARCH(1,1) model is calculated as:

$$\hat{h}_{t+m|t} = \omega + \left(\frac{\alpha_1 + \gamma_1}{2} + \beta_1 \right) \hat{h}_{t+m-1|t}. \quad (\text{A.2})$$

A.2 Proof of Theorem 1.2.1

Recall that process Y_{t+1} is described as:

$$\begin{aligned} Y_{t+1} &= F(Y_t) + \epsilon_{t+1} \\ F(Y_t) &= \mathbb{1}(r_t < \tau) X_t' \theta_1 + \mathbb{1}(r_t \geq \tau) X_t' \theta_2, \end{aligned} \quad (\text{A.3})$$

where $X_t' = \left[1, Y_t, \frac{Y_t + \dots + Y_{t-4}}{5}, \frac{Y_t + \dots + Y_{t-21}}{22} \right]$ and $\theta = (\theta_1', \theta_2')'$. The one-step-ahead forecast is obtained as:

$$\hat{Y}_t(1) = E[Y_{t+1}|I_t] = F(Y_t). \quad (\text{A.4})$$

Next, consider the the-step-ahead forecast from Equation (1.19):

$$\begin{aligned}
\hat{Y}_t(2) &= E[Y_{t+2}|I_t] = \\
&= E[F(Y_{t+1}) + \epsilon_{t+2}|I_t] = E[F(Y_{t+1})|I_t] = \\
&= E[\mathbb{1}(r_{t+1} < \tau)X'_{t+1}\theta_1 + \mathbb{1}(r_{t+1} \geq \tau)X'_{t+1}\theta_2|I_t] = \\
&= E[\mathbb{1}(r_{t+1} < \tau)X'_{t+1}\theta_1|I_t] + E[\mathbb{1}(r_{t+1} \geq \tau)X'_{t+1}\theta_2|I_t] = \\
&= S_1 + S_2.
\end{aligned} \tag{A.5}$$

Simplifying the first summand S_1 , we obtain:

$$\begin{aligned}
S_1 &= E\left[\mathbb{1}(r_{t+1} < \tau) \cdot X'_{t+1}\theta_1|I_t\right] = \\
&= E\left[\mathbb{1}(r_{t+1} < \tau) \cdot \left(1, Y_{t+1}, \frac{Y_{t+1} + \dots + Y_{t-3}}{5}, \frac{Y_{t+1} + \dots + Y_{t-20}}{22}\right)' \theta_1|I_t\right] = \\
&= E\left[\mathbb{1}(r_{t+1} < \tau) \cdot \left(1, F(Y_t) + \epsilon_{t+1}, \frac{F(Y_t) + \epsilon_{t+1} + \dots + Y_{t-3}}{5}, \frac{F(Y_t) + \epsilon_{t+1} + \dots + Y_{t-20}}{22}\right)' \theta_1|I_t\right] = \\
&= [\theta_1 = (c_1, \beta_1^d, \beta_1^w, \beta_1^m)'] = \\
&\approx E[(\mathbb{1}(r_{t+1} < \tau)|I_t)] \cdot E[(c_1 + \beta_1^d(F(Y_t) + \epsilon_{t+1}) + \beta_1^w \left(\frac{F(Y_t) + \epsilon_{t+1} + \dots + Y_{t-3}}{5}\right) + \\
&+ \beta_1^m \left(\frac{F(Y_t) + \epsilon_{t+1} + \dots + Y_{t-20}}{22}\right)|I_t] = \\
&= E\left[(\mathbb{1}(r_{t+1} < \tau)|I_t) \cdot (c_1 + \beta_1^d F(Y_t|I_t) + \beta_1^w \left(\frac{F(Y_t|I_t) + \dots + Y_{t-3}}{5}\right) + \beta_1^m \left(\frac{F(Y_t|I_t) + \dots + Y_{t-20}}{22}\right))\right].
\end{aligned} \tag{A.6}$$

Note that we obtain an approximated expression for the expected value S_1 since r_{t+1} and ϵ_{t+1} are only uncorrelated,²⁴ but not independent. Thus, Expression (A.6) can be simplified

²⁴Recall from (1.20) that $r_{t+1} = z_{t+1} \sqrt{RV_{t+1}}$, where z_{t+1} is i.i.d. $N(0, 1)$ random variable, while z_{t+1} and $\sqrt{RV_{t+1}}$ are independent. Then correlation between r_{t+1} and ϵ_{t+1} is close to zero since μ_n is small, see table 1.1.

as follows:

$$\begin{aligned}
& E \left[(\mathbb{1}(r_{t+1} < \tau) | I_t) \cdot (c_1 + \beta_1^d F(Y_t | I_t) + \beta_1^w \left(\frac{F(Y_t | I_t) + \dots + Y_{t-3}}{5} \right) + \beta_1^m \left(\frac{F(Y_t | I_t) + \dots + Y_{t-20}}{22} \right)) \right] = \\
& = Pr(r_{t+1} < \tau | I_t) \cdot \left(c_1 + \beta_1^d F(Y_t | I_t) + \beta_1^w \left(\frac{F(Y_t | I_t) + \dots + Y_{t-3}}{5} \right) + \beta_1^m \left(\frac{F(Y_t | I_t) + \dots + Y_{t-20}}{22} \right) \right).
\end{aligned} \tag{A.7}$$

Using $\pi_t = Pr(r_{t+1} < \tau | I_t)$, Expression (A.7) becomes:

$$\begin{aligned}
& Pr(r_t < \tau | I_t) \cdot \left(c_1 + \beta_1^d F(Y_t | I_t) + \beta_1^w \left(\frac{F(Y_t | I_t) + \dots + Y_{t-3}}{5} \right) + \beta_1^m \left(\frac{F(Y_t | I_t) + \dots + Y_{t-20}}{22} \right) \right) + \\
& + (1 - Pr(r_t < \tau | I_t)) \cdot \left(c_2 + \beta_2^d F(Y_t | I_t) + \beta_2^w \left(\frac{F(Y_t | I_t) + \dots + Y_{t-3}}{5} \right) + \beta_2^m \left(\frac{F(Y_t | I_t) + \dots + Y_{t-20}}{22} \right) \right) = \\
& = c_1 \pi_t + c_2 (1 - \pi_t) + (\beta_1^d \pi_t + \beta_2^d (1 - \pi_t)) \hat{Y}_t(1) + (\beta_1^w \pi_t + \beta_2^w (1 - \pi_t)) \left[\frac{\hat{Y}_t(1) + \dots + Y_{t-3}}{5} \right] + \\
& + (\beta_1^m \pi_t + \beta_2^m (1 - \pi_t)) \left[\frac{\hat{Y}_t(1) + \dots + Y_{t-20}}{22} \right]
\end{aligned} \tag{A.8}$$

where $\hat{Y}_t(s) = Y_{t+s}$ is $s < 0$. Finally, the formula for the multiple-step-ahead forecast $\hat{Y}_t(h)$ with $h > 2$ is extended recursively from Result (A.8).

Q.E.D.

A.3 Comparison of HAR and SETAR(2) models.

Table A.1: Comparison of the TAR(1) (or HAR) and SETAR(2) models

	\mathbf{RV}_t	$\sqrt{RV_t}$	$\log(\mathbf{RV}_t)$
R^2 of TAR(1)	50.4%	72.6%	73.2%
R^2 of SETAR (2)	51.1%	72.9%	73.5%
τ_{opt}	0.016	0.014	0.007
l	9	1	1
F_{12}	53.8	43.9	44.0

Reported are in-sample estimation results of the linear HAR model and non-linear SETAR(2) model. The in-sample covers the period from February 2000 to June 2014 (3582 observations). We set the maximum amount of lags equal to 10 in the TAR estimation.

Appendix B

Chapter 3 Appendix

B.1 Expectation Maximization algorithm

The Expectation Maximization algorithm consists on two parts: Expectation and Maximization. The idea is to introduce a hidden variable K_t , which governs the choice of particular normal distribution. For instance, if $K_t = 1$ then mixtures of normal distribution becomes $N(0, \sigma_1^2 RV^{(d)})$ at day t , which corresponds to the daily factor.

At the first step (Expectation), we have to compute the posterior distribution for the hidden variable K_t given some initial guess of parameters. The posterior probability that $K_t = 1$ is computed from Bayes' Theorem:

$$Pr(K_t = 1 | X_t = x_t; \theta_0) = \frac{\frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(\frac{-\epsilon_t^2 - z_t^2 \sigma_{kt}^2}{2\sigma_{kt}^2}\right)}{Pr(X_t = x_t; \theta_0)} = q_K(1) \quad (\text{B.1})$$

$$Pr(X_t = x_t; \theta_0) = \sum_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(\frac{-\epsilon_t^2 - z_t^2 \sigma_{kt}^2}{2\sigma_{kt}^2}\right),$$

where vector X_t contains R_t, RV_t, r_t^f , while θ_0 is an initial guess for parameter θ . At the

second step (Maximization), we optimize parameters given the choice of the posterior distribution defined in the Expectation step. Since direct maximization of the log-likelihood is infeasible, we introduce the new distribution of the hidden variable K :

$$\begin{aligned}
 \log (Pr(x; \theta)) &= \log \left(\sum_{k=0}^{K-1} Pr_{X,K}(x, k; \theta) \right) = \\
 &= \log \left(\sum_{k=0}^{K-1} q_K(k) \frac{Pr_{X,K}(x, k; \theta)}{q_K(k)} \right) = \\
 &= \log \left(E_{q_K} \left[\frac{Pr_{X,K}(x, k; \theta)}{q_K(k)} \right] \right) \geq \\
 &\geq E_{q_K} \left[\log \left(\frac{Pr_{X,K}(x, k; \theta)}{q_K(k)} \right) \right].
 \end{aligned} \tag{B.2}$$

The last result follows from Jensen's inequality, which allows us to swap logarithm and sum. As a result, the maximization of the last expression in (B.2) can be done in the closed form. In essence, we maximize lower bound of the log-likelihood and therefore there is no guarantee that EM estimators will be the same or converge to the MLE estimator. Despite the lack of evidences regarding asymptotic efficiency and consistency, several studies investigate performance of EM algorithm via simulations. Nityasuddhi and Bohning (2003) compute bias and RMSE of EM algorithm and find that resulting estimates are quite close to the true parameters. In addition, Arcidiacono and Jones (2003) show that their sequential EM algorithm is asymptotically well-behaved, but it does not converge to the MLE estimator.

To sum-up, EM algorithm is a very pragmatic way to estimate MN model without substantial computational costs. However, the asymptotic efficiency and consistency can not be established. Having discussed the intuition behind EM algorithm we can proceed and derive corresponding estimators.

First, we split the last expression from (B.2) in two parts:

$$E_{q_K} \left[\log \left(\frac{Pr_{X,K}(x, k; \theta)}{q_K(k)} \right) \right] = E_{q_K} [\log(Pr_{X,K}(x, k; \theta))] - E_{q_K} (\log[q_K(k)]). \quad (\text{B.3})$$

Note that second term does not contain vector of parameters θ , and thus can be ignored.

As a result, EM estimator $\hat{\theta}$ is given by:

$$\hat{\theta} = \arg \max_{\theta} E_{q_K} [\log(Pr_{X,K}(x, k; \theta))]. \quad (\text{B.4})$$

Now, we derive EM estimator for our mixtures of normal distributions:

$$\begin{aligned} E_{q_K} [\log(Pr_{X,K}(x, k; \theta))] &= E_{q_K} \left[\log \left(\prod_{t=22}^T \prod_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp \left(\frac{-\epsilon_t^2 - z_t^2 \sigma_{kt}^2}{2\sigma_{kt}^2} \right)^{\mathbb{1}(K_k=k)} \right) \right] = \\ &= E_{q_C} \left[\sum_{t=22}^T \sum_{k=0}^{K-1} \mathbb{1}(K_k = k) \left(\log(\pi_k) + \log \left(\frac{1}{2\pi\sigma_{kt}} \right) - \frac{(\epsilon_t^2 + z_t^2 \sigma_{kt}^2)}{2\sigma_{kt}^2} \right) \right] = \\ &= \sum_{t=22}^T \sum_{k=0}^{K-1} E_{q_C} [\mathbb{1}(K_k = k)] \left(\log(\pi_k) + \log \left(\frac{1}{2\pi\sigma_{kt}} \right) - \frac{(\epsilon_t^2 + z_t^2 \sigma_{kt}^2)}{2\sigma_{kt}^2} \right). \end{aligned} \quad (\text{B.5})$$

Recall that probability of occurring regime k equals to the expected value of the corresponding indicator $E_{q_K}[\mathbb{1}(K_k = k)]$. As a result, we are ready to derive closed form

expression of EM estimates:

$$\begin{aligned}
\tilde{L} &= \sum_{t=22}^T \sum_{k=0}^{K-1} q_K(k) \left(\log(\pi_k) + \log\left(\frac{1}{2\pi\sigma_{kt}}\right) - \frac{(\epsilon_t^2 + z_t^2 \sigma_{kt}^2)}{2\sigma_{kt}^2} \right) \\
\frac{d\tilde{L}}{d\sigma_{kt}^2} = 0 &\quad \implies \hat{\sigma}_k^2 = \frac{\sum_{t=22}^T q_K(k) \frac{\epsilon_t^2}{RV_t^{(k)}}}{\sum_{t=22}^T q_K(k)} \\
\frac{d\tilde{L}}{d\pi_k} = 0 &\quad \implies \hat{\pi}_k = \frac{\sum_{k=0}^{K-1} q_K(k)}{T} \\
q_K(k) &= \frac{\frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(-\frac{1}{2\sigma_{kt}^2} \cdot [\hat{\epsilon}_t^2 + \hat{z}_t^2 \sigma_{kt}^2]\right)}{\sum_{k=0}^{K-1} \frac{\pi_k}{2\pi\sigma_{kt}} \exp\left(-\frac{1}{2\sigma_{kt}^2} \cdot [\hat{\epsilon}_t^2 + \hat{z}_t^2 \sigma_{kt}^2]\right)},
\end{aligned} \tag{B.6}$$

where $q_K(k)$ is a probability of occurring regime k .

B.2 Proof of Theorem 2.2.1

First, we compute the risk-neutral MGF using Radon-Nikodym derivative (2.10):

$$\begin{aligned}
E^Q[\exp(-t_1\epsilon_t - t_2z_t)|I_{t-1}] &= E^P \left[\frac{\frac{dQ}{dP}|I_t}{\frac{dQ}{dP}|I_{t-1}} \exp(-t_1\epsilon_t - t_2z_t)|I_{t-1} \right] = \\
&= E^P [\exp(-v_{1t}\epsilon_t - v_{2t}z_t - \Psi_t(v_{1t}, v_{2t})) \cdot \exp(-t_1\epsilon_t - t_2z_t)|I_{t-1}] = \\
&= \exp[\Psi_t(v_{1t} + t_1, v_{2t} + t_2) - \Psi_t(v_{1t}, v_{2t})].
\end{aligned} \tag{B.7}$$

Recall that conditional MGF is defined as:

$$E^Q[\exp(-t_1\epsilon_t - t_2z_t)|I_{t-1}] = \exp(\Psi^Q(t_1, t_2)). \tag{B.8}$$

By combining (B.7) and (B.8), we obtain the link between risk-neutral and physical log-MGF:

$$\Psi^Q(t_1, t_2) = \Psi(v_{1t} + t_1, v_{2t} + t_2) - \Psi(v_{1t}, v_{2t}). \quad (\text{B.9})$$

Next, the logarithm of conditional MGF under risk-neutral measure is given by:

$$\begin{aligned} \Psi^Q(t_1, t_2) &= \Psi(v_{1t} + t_1, v_{2t} + t_2) - \Psi(v_{1t}, v_{2t}) = \\ &= \ln \left(\sum_{k=0}^{K-1} \pi_k \exp \left[\frac{(v_{1t} + t_1)^2 \sigma_{kt}^2}{2} + \frac{(v_{2t} + t_2)^2}{2} \right] \right) - \ln \left(\sum_{k=0}^{K-1} \pi_k \exp \left[\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} \right] \right) = \\ &= \ln \left(\sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \frac{\exp \left[\frac{(v_{1t} + t_1)^2 \sigma_{kt}^2}{2} + \frac{(v_{2t} + t_2)^2}{2} \right]}{\exp \left[\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} \right]} \right) = \\ &= \ln \left(\sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \exp \left[-(-v_{1t} \sigma_{kt}^2) t_1 - (-v_{2t}) t_2 + \frac{1}{2} (t_1^2 \sigma_{kt}^2 + t_2^2) \right] \right), \end{aligned} \quad (\text{B.10})$$

where time-varying probability $\widetilde{\pi}_{kt}$, mean of innovation z_t^* and error term ϵ_t^* under risk-neutral measure are computed as follows:

$$\begin{aligned} \widetilde{\pi}_{kt} &= \frac{\pi_k \exp \left(\frac{v_{1t}^2 \sigma_{kt}^2 + v_{2t}^2}{2} \right)}{\sum_{k=0}^{K-1} \pi_k \exp \left(\frac{v_{1t}^2 \sigma_{kt}^2 + v_{2t}^2}{2} \right)}, \\ \widetilde{\mu}_{kt} &= -v_{1t} \sigma_{kt}^2, \\ \widetilde{M}_t &= -v_{2t}. \end{aligned} \quad (\text{B.11})$$

Q.E.D.

B.3 Proof of Theorem 2.2.2

From the Theorem 2.2.1 we know that physical and risk-neutral distributions belong to the same family of distributions — mixture of normal, albeit with different parameters. In addition, we have already computed all risk-neutral probabilities except of the correction term γ_t . Under risk-neutral measure the conditional expected returns equals to risk free rate:

$$\begin{aligned} E^Q \left[\frac{S_t}{S_{t-1}} | I_{t-1} \right] &= e^{r_t^f} = E^Q \left[e^{r_t^f - \gamma_t + \sqrt{RV_t^d} z_t^*} | I_{t-1} \right] \\ e^{\gamma_t} &= E^Q \left[e^{\sqrt{RV_t^d} z_t^*} | I_{t-1} \right]. \end{aligned} \tag{B.12}$$

We calculate expression (B.12) using MGF of the non-central chi square random variable. Indeed, since $\sqrt{RV_t^d}$ and z_t^* follow mixtures of normal distributions their product is distributed as mixtures of non-central chi-squared random variables with the following MGF defined for fixed mixture k :

$$MGF(t) = \frac{\exp\left(\frac{\lambda_k t}{1-2t}\right)}{\sqrt{1-2t}}, \tag{B.13}$$

where λ_k is non-centrality parameter. Now, we can derive expression for γ_t :

$$\begin{aligned}
E^Q \left[e^{\sqrt{RV_t^d} z_t^*} | I_{t-1} \right] &= E^Q \left[e^{(M_t + \epsilon_t^*) z_t^*} | I_{t-1} \right] = \\
&= E_{\epsilon_t^*}^Q \left[E_{z_t^*}^Q \left[e^{(M_t + \epsilon_t^*) z_t^*} | I_{t-1} \right] \right] = \\
&= E_{\epsilon_t^*}^Q \left[\exp \left(-v_{2t} (M_t + \epsilon_t^*) + \frac{1}{2} (M_t + \epsilon_t^*)^2 \right) | I_{t-1} \right] = \\
&= \exp \left(-v_{2t} M_t + \frac{1}{2} M_t^2 \right) E_{\epsilon_t^*}^Q \left[\exp \left(\epsilon_t^* b_t + \frac{1}{2} (\epsilon_t^*)^2 \right) | I_{t-1} \right] = \\
&= \exp \left(-v_{2t} M_t + \frac{1}{2} M_t^2 - \frac{b_t^2}{2} \right) E_{\epsilon_t^*}^Q \left[\exp \left(\frac{1}{2} [\epsilon_t^* + b_t]^2 \right) | I_{t-1} \right],
\end{aligned} \tag{B.14}$$

where $b_t = M_t - v_{2t}$. Note that $[\epsilon_t^* + b_t]^2$ is a mixture of non-central chi squared random variables with non-centrality parameters equal to:

$$\begin{aligned}
\lambda_k &= \left(\frac{b_t + \widetilde{\mu}_{kt}}{\sigma_{kt}} \right)^2 \\
\widetilde{\mu}_{kt} &= -v_{1t} \sigma_{kt}^2.
\end{aligned} \tag{B.15}$$

To finish derivation we need to compute MGF of non-central chi-squared random variable defined in (B.15). After rearranging terms we obtain:

$$\begin{aligned}
E^Q \left[e^{\sqrt{RV_t^d} z_t^*} | I_{t-1} \right] &= \exp \left(-v_{2t} M_t + \frac{1}{2} M_t^2 - \frac{b_t^2}{2} \right) \sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \frac{\exp \left(\frac{\lambda_k \cdot \frac{\sigma_{kt}^2}{2}}{1 - 2 \frac{\sigma_{kt}^2}{2}} \right)}{\sqrt{1 - 2 \frac{\sigma_{kt}^2}{2}}} = \\
&= \exp \left(-v_{2t} M_t + \frac{1}{2} M_t^2 \right) \sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \frac{\exp \left(\frac{b_t^2 \sigma_{kt}^2 + 2b_t \widetilde{\mu}_{kt} + \widetilde{\mu}_{kt}^2}{2(1 - \sigma_{kt}^2)} \right)}{\sqrt{1 - \sigma_{kt}^2}}.
\end{aligned} \tag{B.16}$$

We finish proof with the explicit expression for the correction term γ_t :

$$\gamma_t = -v_{2t}M_t + 0.5M_t^2 + \log \left(\frac{\sum_{k=0}^{K-1} \widetilde{\pi}_{kt} \exp \left(\frac{(M_t - v_{2t})^2 \sigma_{kt}^2 + 2(M_t - v_{2t}) \widetilde{\mu}_{kt} + \widetilde{\mu}_{kt}^2}{2(1 - \sigma_{kt}^2)} \right)}{\sqrt{1 - \sigma_{kt}^2}} \right). \quad (\text{B.17})$$

Q.E.D.

B.4 Proof of Theorem 2.2.3

We need to obtain EMM condition $E^Q \left[\frac{S_t}{B_t} | I_{t-1} \right] = \frac{S_{t-1}}{B_{t-1}}$ or $E^Q \left[\frac{S_t/S_{t-1}}{B_t/B_{t-1}} | I_{t-1} \right] = 1$:

$$\begin{aligned} E^Q \left[\frac{S_t/S_{t-1}}{B_t/B_{t-1}} | I_{t-1} \right] &= E^P \left[\frac{\frac{dQ}{dP} | I_t}{\frac{dQ}{dP} | I_{t-1}} \frac{S_t}{S_{t-1}} / \frac{B_t}{B_{t-1}} | I_{t-1} \right] = \\ &= E^P \left[\exp(-v_{1t}\epsilon_t - v_{2t}z_t - \Psi(v_{1t}, v_{2t}) - r_t^f) \frac{S_t}{S_{t-1}} | I_{t-1} \right] = \\ &= \exp(-\Psi(v_{1t}, v_{2t})) \cdot E^P \left[\exp \left(-v_{1t}\epsilon_t - v_{2t}z_t + \lambda \sqrt{RV_t^d} + \sqrt{RV_t^d} z_t \right) | I_{t-1} \right] = \\ &= \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \cdot E_{z_t}^P \left[E_{\epsilon_t}^P \exp(\epsilon_t(-v_{1t} + \lambda + z_t)) \exp(z_t b) | I_{t-1} \right] = \\ &= \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \cdot \sum_{k=0}^{K-1} \pi_k E_{z_t}^P \left[\exp \left(\frac{1}{2} (\lambda - v_{1t} + z_t)^2 \sigma_{kt}^2 + z_t b \right) | I_{t-1} \right]. \end{aligned} \quad (\text{B.18})$$

Note that expression within the expectation is a function of z_t^2 . Thus, we obtain expectation with respect non-central chi-squared random variable, which is its MGF defined in

(B.18). Rearranging terms we obtain:

$$\begin{aligned}
E^Q \left[\frac{S_t/S_{t-1}}{B_t/B_{t-1}} / I_{t-1} \right] &= \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \cdot \sum_{k=0}^{K-1} \pi_k E_{z_t}^P \left[\exp \left(\frac{1}{2} (\lambda - v_{1t} + z_t)^2 \sigma_{kt}^2 + z_t b_t \right) / I_{t-1} \right] = \\
&= \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \times \\
&\times \sum_{k=0}^{K-1} \pi_k E_{z_t}^P \left[\exp \left(\frac{\sigma_{kt}^2}{2} \left(z_t + \frac{c_t \sigma_{kt}^2 + b_t}{\sigma_{kt}^2} \right)^2 + \left(\frac{-b_t^2 - 2c_t \sigma_{kt}^2 b_t}{2\sigma_{kt}^2} \right) \right) \right] = \\
&= \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \cdot \sum_{k=0}^{K-1} \pi_k \left[\frac{\exp \left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)} \right)}{\sqrt{1 - \sigma_{kt}^2}} \right].
\end{aligned} \tag{B.19}$$

Recall that conditional log-MGF under physical measure is:

$$\Psi_t(t_1, t_2) = \log \left[\sum_{k=0}^{K-1} \pi_k \exp \left(\frac{t_1^2 \sigma_{kt}^2}{2} + \frac{t_2}{2} \right) \right]. \tag{B.20}$$

Next, we take logarithm of the above expression and rearrange terms using (B.14) we

obtain:

$$\begin{aligned}
& \exp(\lambda M_t - \Psi(v_{1t}, v_{2t})) \cdot \sum_{k=0}^{K-1} \pi_k \left[\frac{\exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)}\right)}{\sqrt{1-\sigma_{kt}^2}} \right] = 1 \\
& \lambda M_t - \Psi(v_{1t}, v_{2t}) + \log \left[\sum_{k=0}^{K-1} \pi_k \left[\frac{\exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)}\right)}{\sqrt{1-\sigma_{kt}^2}} \right] \right] = 0 \\
& \log \left[\sum_{k=0}^{K-1} \pi_k \left[\exp\left(\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} - \lambda M_t\right) \right] \right] = \log \left[\sum_{k=0}^{K-1} \pi_k \left[\frac{\exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)}\right)}{\sqrt{1-\sigma_{kt}^2}} \right] \right] \\
& \sum_{k=0}^{K-1} \pi_k \left[\exp\left(\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} - \lambda M_t\right) \right] = \sum_{k=0}^{K-1} \pi_k \left[\frac{\exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)}\right)}{\sqrt{1-\sigma_{kt}^2}} \right].
\end{aligned} \tag{B.21}$$

The last part of equation (B.21) is a non-linear equation of v_{2t} . However, we can obtain approximate solution using the fact that $\sigma_{kt}^2 \approx 0$:

$$\sum_{k=0}^{K-1} \pi_k \left[\exp\left(\frac{v_{1t}^2 \sigma_{kt}^2}{2} + \frac{v_{2t}^2}{2} - \lambda M_t\right) \right] \approx \sum_{k=0}^{K-1} \pi_k \left[\frac{\exp\left(\frac{c_t^2 \sigma_{kt}^2 + b_t^2 + 2c_t \sigma_{kt}^2 b_t}{2(1-\sigma_{kt}^2)}\right)}{\sqrt{1-\sigma_{kt}^2}} \right] \tag{B.22}$$

$$\begin{aligned}
& \sum_{k=0}^{K-1} \pi_k \left[\exp\left(\frac{v_{2t}^2}{2} - \lambda M_t\right) \right] \approx \sum_{k=0}^{K-1} \pi_k \exp\left[\frac{b_t^2}{2}\right] \\
& v_{2t} \approx \frac{M_t + 2\lambda}{2}.
\end{aligned} \tag{B.23}$$

Q.E.D.

Bibliography

- Andersen, T., L. Benzoni, and J. Lund (2002) “An Empirical Investigation of Continuous-Time Equity Return Models,” *Journal of Finance*, Vol. 57, pp. 1239–1284.
- Andersen, T. and T. Bollerslev (1997) “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns,” *Journal of Finance*, Vol. 52, pp. 975–1005.
- Andersen, T., T. Bollerslev, and F. Diebold (2007) “Roughing it Up: Disentangling Continuous and Jump Components in Measuring, Modeling and Forecasting Asset Return Volatility,” *Review of Economics and Statistics*, Vol. 89, pp. 701–720.
- Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001) “The Distribution of Realized Stock Return Volatility,” *Journal of Financial Economics*, Vol. 61, pp. 43–76.
- Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003) “Modeling and Forecasting Realized Volatility,” *Econometrica*, Vol. 71, pp. 579–625.
- Andersen, T., T. Bollerslev, P. Frederiksen, and M. Nielsen (2010) “Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns,” *Journal of Applied Econometrics*, Vol. 25, pp. 233–261.
- Andrews, D. (1993) “Tests for parameter instability and structural change with unknown change point,” *Econometrica*, Vol. 61, pp. 821–856.
- Arcidiacono, P. and J. Jones (2003) “Finite Mixture Distributions, Sequential Likelihood and the EM algorithm,” *Econometrica*, Vol. 71, pp. 933–946.

- Arthur, W., J. Holland, B. LeBaron, J. Palmer, and P. Tayler (1997) "Asset pricing under heterogeneous expectations in an artificial stock market," *The Economy as an Evolving Complex System*, Vol. Perseus Books, Reading MA,, pp. 15–44.
- Audrino, F., R. Huitema, and M. Ludwig (2015) "An Empirical Analysis of the Ross Recovery Theorem," *Working Paper University of Zurich*, pp. 1–39.
- Babaoglu, K., P. Christoffersen, S. Heston, and K. Jacobs (2014) "Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels," *Working paper, University of Maryland*.
- Bakshi, G., C. Cao, and Z. Chen (1997) "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, Vol. 52, pp. 2003–2049.
- on Banking Supervision, Basel Committee (2011) "Basel III: A global regulatory framework for more resilient banks and banking systems," pp. 1–77.
- Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008) "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, Vol. 76, pp. 1481–1536.
- Barndorff-Nielsen, O. and N. Shephard (2002a) "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, Vol. 64, pp. 253–280.
- (2002b) "Estimating quadratic variation using realized variance," *Journal of Applied Econometrics*, Vol. 17, pp. 457–477.

- Barone-Adesi, G., R. Engle, and L. Mancini (2008) “A GARCH Option Pricing Model with Filtered Historical Simulation,” *Review of Financial Studies*, Vol. 21, pp. 1223–1258.
- Bates, D. (1996) “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options,” *Review of Financial Studies*, Vol. 9, pp. 69–107.
- Black, F. and M. Scholes (1973) “The Pricing of Options and Corporate Liabilities,” *The Journal of Political Economy*, Vol. 81, pp. 637–654.
- Bollerslev, T. (1986) “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, Vol. 31, pp. 307–327.
- Bollerslev, T., R. Engle, and D. Nelson (1994) “ARCH Models,” in R.F. Engle and D. McFadden eds. *Handbook of Econometrics, Vol.IV*: Elsevier Science B.V.: Amsterdam, The Netherlands, Chap. 49, pp. 2959-3038.
- Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009) “A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects,” *Journal of Econometrics*, Vol. 150, pp. 151–166.
- Borovicka, J., L. Hansen, and J. Scheinkman (2016) “Misspecified Recovery,” *Forthcoming in Journal of Finance*, pp. 1–70.
- Boyle, P., M. Broadie, and P. Glasserman (1997) “Monte Carlo Methods for Security Pricing,” *Journal of Economic Dynamics and Control*, Vol. 21, pp. 1267–1321.
- Brownlees, C. and G. Gallo (2009) “Comparison of volatility measures: A risk management perspective,” *Journal of Financial Economics*, Vol. 8, pp. 29–56.

- Byun, S., B. Jeon, B. Min, and S-J. Yoon (2015) "The role of the variance premium in Jump-GARCH option pricing models," *Journal of Banking & Finance*, Vol. 59, pp. 38–56.
- Carr, P. and J. Yu (2012) "Risk, Return, and Ross Recovery," *The Journal of Derivatives*, Vol. 20, pp. 38–59.
- Chan, K. (1993) "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, Vol. 21, pp. 520–533.
- Chan, K., J. Petrucelli, and H. Tong and S. Woolford (1985) "A multiple threshold AR(1) model," *Journal of Applied Probability*, Vol. 22, pp. 267–279.
- Choi, K., W. Yu, and E. Zivot (2010) "Long Memory versus Structural Breaks in Modeling and Forecasting Realized Volatility," *Journal of International Money and Finance*, Vol. 29, pp. 857–875.
- Christoffersen, P. (1998) "Evaluating Interval Forecasts," *International Economic Review*, Vol. 39, pp. 841–862.
- Christoffersen, P., R. Elkamhi, B. Feunou, and K. Jacobs (2010) "Option Valuation with Conditional Heteroskedasticity and Nonnormality," *Review of Financial Studies*, Vol. 23, pp. 2139–2183.
- Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi (2014) "The Economic Value of Realized Volatility: Using High-Frequency Returns for Option Valuation," *The Economic Value of Realized Volatility: Using High-Frequency Returns for Option Valuation*, Vol. 49, pp. 663–697.

Christoffersen, P., B. Feunou, and Y. Jeon (2015a) “Option Valuation with Observable Volatility and Jump Dynamics,” *Journal of Banking & Finance*, Vol. 61, pp. 101–120.

Christoffersen, P., M. Fournier, and K. Jacobs (2013a) “The Factor Structure in Equity Options,” *Working paper, University of Toronto*, pp. 1–58.

——— (2015b) “The Factor Structure in Equity Options,” *Rotman School of Management Working Paper No. 2224270*, pp. 1–75.

Christoffersen, P., S. Heston, and K. Jacobs (2006) “Option Valuation with Conditional Skewness,” *Journal of Econometrics*, Vol. 131, pp. 253–284.

——— (2009) “The Shape and Term Structure of the Index Option Smirk: Why Multi-factor Stochastic Volatility Models Work So Well,” *Management Science*, Vol. 55, pp. 1914–1932.

Christoffersen, P., S. Heston, and K. Jacobs (2013b) “Capturing Option Anomalies with a Variance-Dependent Pricing Kernel,” *Review of Financial Studies*, Vol. 26, pp. 1962–2006.

Christoffersen, P., K. Jacobs, and C. Ornathanalai (2012) “Dynamic Jump Intensities and Risk Premiums: Evidence from S&P500 Returns and Options,” *Journal of Financial Economics*, Vol. 106, pp. 447–472.

——— (2013c) “GARCH Option Valuation: Theory and Evidence.” *The Journal of Derivatives*, Vol. 21, pp. 8–41.

Conley, T., L. Hansen, E. Luttmer, and J. Scheinkman (1997) “Short-term interest rates as subordinated diffusions,” *Review of Financial Studies*, Vol. 10, pp. 525–577.

- Cont, R. (2007) “Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models,” *Long Memory in Economics*, Vol. Springer Berlin Heidelberg, pp. 289–309.
- Corsi, F. (2009) “A Simple Approximate Long-Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, Vol. 7, pp. 174–196.
- Corsi, F., N. Fusari, and D. Vecchia (2013) “Realizing smiles: Pricing options with realized volatility,” *Journal of Financial Economics*, Vol. 107, pp. 284–304.
- Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008) “The volatility of realized volatility,” *Econometric Reviews*, Vol. 27, pp. 46–78.
- Dacco, R. and S. Satchell (1999) “Why do Regime-switching Models Forecast so Badly?” *Journal of Forecasting*, Vol. 18, pp. 1–16.
- Davies, R. (1977) “Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative,” *Biometrika*, Vol. 64, pp. 247–254.
- (1987) “Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative,” *Biometrika*, Vol. 74, pp. 33–43.
- Dempster, A., N. Laird, and D. Rubin (1977) “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, pp. 1–38.
- Diebold, F. and C. Chen (1996) “Testing structural stability with endogenous breakpoint. A size comparison of analytic and bootstrap procedures,” *Journal of Econometrics*, Vol. 70, pp. 221–241.

- Diebold, F. and M. Nerlov (1989) "The dynamics of exchange rate volatility: A multivariate latent factor ARCH model," *Journal of Applied Econometrics*, Vol. 4, pp. 1–21.
- Duan, J. and J. Simonato (1998) "Empirical Martingale Simulation for Asset Prices," *Management Science*, Vol. 44, pp. 1218–1233.
- Duffie, D., J. Pan, and K. Singleton (2000) "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, Vol. 68, pp. 1343–1376.
- Duffie, D. and K. Singleton (1993) "Simulated moments estimation of Markov models of asset prices," *Econometrica*, Vol. 61, pp. 929–952.
- Engle, R. (1982) "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of UK Inflation," *Econometrica*, Vol. 50, pp. 987–1008.
- Eraker, B. (2001) "MCMC analysis of diffusion models with application to finance," *Journal of Business and Economic Statistics*, Vol. 19, pp. 177–191.
- (2004) "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, Vol. 59, pp. 1367–1404.
- Forsberg, L. and T. Bollerslev (2002) "Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the Euro): The GARCH-NIG Model," *Journal of Applied Econometrics*, Vol. 17, pp. 535–548.
- Franses, P. and D. Dijk (2000) *Non-Linear Time Series Models in Empirical Finance*, Cambridge University Press: Cambridge, United Kingdom.
- Giacomini, R. and H. White (2006) "Tests of Conditional Predictive Ability," *Econometrica*, Vol. 74, pp. 1545–1578.

- Glosten, L., R. Jagannathan, and D. Runkle (1993) "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, Vol. 48, pp. 1779–1801.
- Granger, C. and A. Ding (1996) "Varieties of Long Memory Models," *Journal of Econometrics*, Vol. 73, pp. 61–77.
- Hansen, B. (1997) "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, Vol. 2, pp. 1–14.
- (1999) "Testing for Linearity," *Journal of Economic Surveys*, Vol. 13, pp. 551–576.
- (2000) "Sample splitting and threshold estimation," *Econometrica*, Vol. 68, pp. 575–603.
- Hansen, P. and A. Lunde (2005) "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?" *Journal of Applied Econometrics*, Vol. 20, pp. 873–889.
- Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009) "Oxford-Man Institute's realized library, Library version: 0.2," *Oxford, United Kingdom*.
- Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016) "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, Vol. 119, pp. 249–283.
- Heston, S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, Vol. 6, pp. 327–343.

- Heston, S. and S. Nandi (2000) "A Closed-Form GARCH Option Valuation Model," *The Review of Financial Studies*, Vol. 13, pp. 585–625.
- Hull, J. and A. White (1987) "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, Vol. 42, pp. 281–300.
- Jiang, G. and J. Knight (1997) "A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model," *Econometric Theory*, Vol. 13, pp. 615–645.
- (2002) "Efficient estimation of the continuous time stochastic volatility model via the empirical characteristic function," *Journal of Business and Economic Statistics*, Vol. 20, pp. 198–212.
- Kimball, M. (1993) "Standard Risk Aversion," *Econometrica*, Vol. 61, pp. 589–611.
- Knight, J. and S. Satchell (2011) "Some New Results for Threshold AR(1) Models," *Journal of Time Series Econometrics*, Vol. 3, pp. 1–42.
- LeBaron, B. (2001) "Evolution and time horizons in an agent-based stock market," *Macroeconomic Dynamics*, Vol. 5, pp. 225–254.
- LeBaron, B., B. Arthur, and R. Palmer (1999) "Time series properties of an artificial stock market," *Journal of Economic Dynamics and Control*, Vol. 23, pp. 1487–1516.
- Lustig, H., N. Roussanov, and A. Verdelhan (2011) "Common Risk Factors in Currency Markets," *The Review of Financial Studies*, Vol. 24, pp. 3731–3777.
- (2014) "Countercyclical Currency Risk Premia," *Journal of Financial Economics*, Vol. 111, pp. 527–553.

- Maheu, J. and T. McCurdy (2011) “Do High-frequency Measures of Volatility Improve Forecasts of Return Distributions?” *Journal of Econometrics*, Vol. 160, pp. 69–76.
- Mahieu, R. and P. Schotman (1994) “Neglected common factors in exchange rate volatility,” *Journal of Empirical Finance*, Vol. 1, pp. 279–311.
- Majewski, A., G. Bormettia, and F. Corsi (2015) “Smile from the Past: A general option pricing framework with multiple volatility and leverage components,” *Journal of Econometrics*, Vol. 187, pp. 521–531.
- McAleer, M. and M. Medeiros (2008) “A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries,” *Journal of Econometrics*, Vol. 147, pp. 104–119.
- Melino, A. and S. Turnbull (1990) “Pricing Foreign Currency Options with Stochastic Volatility,” *Journal of Econometrics*, Vol. 45, pp. 239–265.
- Merton, R. (1973) “Theory of Rational Option Pricing,” *The Bell Journal of Economics and Management Science*, Vol. 4, pp. 141–183.
- Muller, U., M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. Ward (14-15 October, 1993) “Fractals and Intrinsic Time – A Challenge to Econometricians,” in *Proceedings of the XXXIXth International AEA Conference on Real Time Econometrics*: Luxembourg.
- Nityasuddhi, D. and D. Bohning (2003) “Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances,” *Computational Statistics & Data Analysis*, Vol. 41, pp. 591–601.

- Oh, D. and A. Patton (2016) “High Dimension Copula-Based Distributions with Mixed Frequency Data,” *Journal of Econometrics*, Vol. 193, pp. 349–366.
- Ornthanalai, C. (2014) “Levy jump risk: Evidence from options and returns,” *Journal of Financial Economics*, Vol. 112, pp. 69–90.
- Pan, J. and A. Poteshman (2006) “The information in options volume for future stock prices,” *Review of Financial Studies*, Vol. 19, pp. 871–908.
- Patton, A. and I. Salvatierra (2015) “Dynamic Copula Models and High Frequency Data,” *Journal of Empirical Finance*, Vol. 30, pp. 120–135.
- Perron, P. (1989) “The Great Crash, the Oil Price Shock and the Unit Root Hypothesis,” *Econometrica*, Vol. 57, pp. 1361–1401.
- Pypko, S. (2015) “Volatility Forecast in Crises and Expansions,” *Journal of Risk and Financial Management*, Vol. 8, pp. 311–326.
- Renault, E. (1997) “Econometric models of option pricing errors,” *Econometric Society Monograph*, Vol. 28, pp. 223–278.
- Roll, R., E. Schwartz, and A. Subrahmanyam (2010) “O/S: The relative trading activity in options and stock,” *Journal of Financial Economics*, Vol. 96, pp. 1–17.
- Rombouts, J. and M. Bouaddi (2009) “Mixed Exponential Power Asymmetric Conditional Heteroskedasticity,” *Studies in Nonlinear Dynamics & Econometrics*, Vol. 13, pp. 1–30.
- Rombouts, J. and L. Stentoft (2015) “Option pricing with asymmetric heteroskedastic normal mixture models,” *International Journal of Forecasting*, Vol. 31, pp. 635–650.

- Ross, S. (2015) "The Recovery Theorem," *The Journal of Finance*, Vol. 70, pp. 615–648.
- Scharth, M. and M. Medeiros (2009) "Asymmetric effects and long memory in the volatility of Dow Jones stocks," *International Journal of Forecasting*, Vol. 25, pp. 304–327.
- Singleton, K. (2001) "Estimation of affine asset pricing models using the empirical characteristic function," *Journal of Econometrics*, Vol. 102, pp. 111–141.
- Stentoft, L. (2008) "Option Pricing Using Realized Volatility," *CREATES Research Paper; Aarhus, Denmark*.
- Tong, H. (1978) "On a Threshold Model.," in C.H. Chen ed. *Pattern Recognition and Signal Processing*: Sijthoff and Noordhoff: Amsterdam, the Netherlands, pp. 101–141.
- Tong, H. and K. Lim (1980) "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 42, pp. 245–292.
- Verdelhan, A. (2015) "The Share of Systematic Variation in Bilateral Exchange Rates," *Working paper, MIT*, pp. 1–71.
- Wirjanto, T. and D. Xu (2009) "The Applications of Mixtures of Normal Distributions in Empirical Finance: A Selected Survey," *Working paper, University of Waterloo*, pp. 1–35.
- Zivot, E. and D. Andrews (1992) "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, Vol. 10, pp. 251–270.

Curriculum Vitae

Name: Sergii Pypko

Post-Secondary Education and Degrees: Odessa National Mechnikov University, Odessa, Ukraine
2003-2007, B.S. in Mathematics (Cum Laude)

Kyiv School of Economics, Kyiv, Ukraine
2007-2009, M.A. in Economics (Magna Cum Laude)

The University of Houston, Houston, USA
2009-2009, M.A. in Economics

The University of Western Ontario, London, Canada
2012-2016, Ph.D. in Economics

Honours and Awards: Gervan Fearon PhD Award in Economics

Related Work Experience: Quantitative Researcher, National Bank
2016-present

Senior Analyst, Bank of Canada
2015-2015

Teaching and Research Assistant, The University of Western Ontario
2012-2016

Lecturer, The University of Western Ontario
2014-12/2016

Quantitative Analyst, Elyland LLC
2011-2012

Economist, SigmaBleyzer
2010-2011

Publication: "Volatility Forecast in Crises and Expansions",
Journal of Risk and Financial Management, Vol. 8, pp. 311-336, 2015