ESSAYS ON STOCHASTIC VOLATILITY AND JUMPS

by

DIEP NGOC DUONG

A Dissertation submitted to the

Graduate School - New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Norman Rasmus Swanson

and approved by

____

____

____

____

New Brunswick, New Jersey
May 2013

ABSTRACT OF THE DISSERTATION

Essays on Stochastic Volatility and Jumps

By DIEP NGOC DUONG

Dissertation Director:

Norman Rasmus Swanson

This dissertation comprises three essays on financial economics and econometrics. The first essay outlines and expands upon further testing results from Bhardwaj, Corradi and Swanson (BCS: 2008) and Corradi and Swanson (2011). In particular, specification tests in the spirit of the conditional Kolmogorov test of Andrews (1997) that rely on block bootstrap resampling methods are first discussed. We then broaden our discussion from single process specification testing to multiple process model selection by discussing how to construct predictive densities and how to compare the accuracy of predictive densities derived from alternative (possibly misspecified) diffusion models. In particular, we generalize simulation steps outlined in Cai and Swanson (2011) to multifactor models where the number of latent variables is larger than three. In the second essay, we begin by discussing important developments in volatility modeling, with a focus on time varying and stochastic volatility as well as the "model free" estimation of volatility via the use of so-called realized volatility, and variants thereof called realized measures. In an empirical investigation, we use realized measures to investigate the role of “small” and “large” jumps in the realized variation of stock price returns and show that jumps do

matter in the relative contribution to the total variation of the process, when examining individual stock returns, as well as market indices. The third essay examines the predictive content of a variety of realized measures of jump power variations, all formed on the basis of power transformations of instantaneous returns. Our prediction involves estimating members of the linear and nonlinear extended Heterogeneous Autoregressive of the Realized Volatility (HAR-RV) class of models, using S&P 500 futures data as well as stocks in the Dow 30, for the period 1993-2009. Our findings suggest that past "large" jump power variations helpless in the prediction of future realized volatility, than past "small" jump power variations. Our empirical findings also suggest that past realized signed jump power variations, which have not previously been examined in this literature, are strongly correlated with future volatility.

Acknowledgments

I am deeply indebted to my dissertation advisor, Professor Norman Swanson, for his invaluable research guidance and support during my entire graduate study at Rutgers University. He gave me many valuable lessons and expanded my perspective in the research world as well as provided the optimal advices whenever I needed them. For me, he sets a role model as a researcher and an advisor.

I am extremely grateful to Professor Roger Klein and Professor Richard Mclean for their advices, encouragement and support. Professor Roger Klein was so patient to teach me various statistical topics and provided me the opportunities to do research in semi-parametric econometrics. The discussion with him has always been motivational and inspiring. Professor Mclean showed so much enthusiasm in helping me improve my second year and job market papers which are parts of my dissertation. The unsurpassed insight and profound knowledge that I learnt from Professor Klein and Professor Mclean will stay with me as an invaluable asset.

I would like to express my sincere thanks to Nii Ayi Armah, Professor John Landon-Lane, Professor Bruce Mizrach, Professor Carolyn Moehling, Professor Hiroki Tsurumi for helpful comments as well as insightful discussions at various stages of my dissertation. Professor Mizrach introduced me to the financial market microstructure topics and I greatly appreciate the valuable discussions with him.

I owe special thanks to Anh Le for his encouragement and support over the years. He has been patient in explaining things to me and sharing his unique insights in the area of fixed income, which has helped me shape the research agenda for the years to come.

I am also grateful to Professor Joel Rogers for the mathematical knowledge and advices that he taught me at Polytechnic University.

Many thanks are owed to Dorothy Rinaldi and Mayra Lopez for their support. I also appreciate the financial support from Economics Department and Graduate School at Rutgers University.

I would also like to thank my friends, in particular Huyn Hak Kim, Minh Le, Tan Le, Hien Ngo, Giang Nguyen, Yoichi Otsubo, Hoa Tran for the support and time we spent together.

Finally, I would like to thank my parents and my wife for their immense sacrifices.

Dedication

To my parent, my wife, Son Duong and Hali Duong.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation considers the specification tests of diffusion processes, the measurement and forecast of volatility in the presence of stochastic volatility and jumps. The technique of using densities and conditional distributions to carry out consistent specification testing and model selection amongst multiple diffusion processes have received considerable attention from both financial theoreticians and empirical econometricians over the last two decades. One reason for this interest is that correct specification of diffusion models describing dynamics of financial assets is crucial for many areas in finance including equity and option pricing, term structure modeling, and risk management, for example. In the second chapter, we discuss advances to this literature introduced by Corradi and Swanson (2005), who compare the cumulative distribution (marginal or joint) implied by a hypothesized null model with corresponding empirical distributions of observed data. We also outline and expand upon further testing results from Bhardwaj, Corradi and Swanson (BCS: 2008) and Corradi and Swanson (2011). In particular, parametric specification tests in the spirit of the conditional Kolmogorov test of Andrews (1997) that rely on block bootstrap resampling methods in order to construct test critical values are first discussed. Thereafter, extensions due to BCS (2008) for cases where the functional form of the conditional density is unknown are introduced, and related continuous time simulation methods are introduced. Finally, we broaden our discussion from single process specification testing to multiple process model selection by discussing how to construct predictive densities and how to compare the accuracy of predictive densities derived from alternative (possibly misspecified) diffusion models. In particular, we generalize simulation Steps outlined in Cai and Swanson (2011) to multifactor models where the number of latent variables is larger than three. These final tests can be thought of as continuous time generalizations of the discrete time "reality check" test statistics of White (2000), which are widely used in empirical finance (see e.g. Sullivan, Timmermann and White (1999, 2001)). We finish the chapter with an empirical illustration of model selection amongst alternative short term interest rate models.

In the third chapter, we begin by surveying models of volatility, both discrete and continuous, and then we summarize some selected empirical findings from the literature. In particular, in the first sections of this chapter, we discuss important developments in volatility models, with focus on time varying and stochastic volatility as well as nonparametric volatility estimation. The models discussed share the common feature that volatilities are unobserved, and belong to the class of missing variables. We then provide empirical evidence on "small" and "large" jumps from the perspective of their contribution to overall realized variation, using high frequency price return data on 25 stocks in the DOW 30. Our "small" and "large" jump variations are constructed at three truncation levels, using extant methodology of Barndorff-Nielsen and Shephard (2006), Andersen, Bollerslev and Diebold (2007) and Aït-Sahalia and Jacod (2009a,b). Evidence of jumps is found in around 22.8% of the days during the 1993-2000 period, much higher than the corresponding figure of 9.4% during the 2001-2008 period. While the overall role of jumps is lessening, the role of large jumps has not decreased, and indeed, the relative role of large jumps, as a proportion of overall jumps has actually increased in the 2000s.

Volatility predictability is important in numerous areas of financial econometrics ranging from the pricing of volatility-based derivative products to asset management. In light of this, a number of recent papers have addressed volatility predictability, some from the perspective of the usefulness of jumps in forecasting volatility. Key papers in this area of research include Andersen, Bollerslev, Diebold and Labys (2003), Andersen, Bollerslev and Diebold (2007), Barndoff, Kinnebrock, and Shephard (2010), Corsi (2004), Corsi, Pirino and Reno (2008), Patton and Shephard (2011), and the references cited therein. In the fourth chapter, we examine the predictive content of a variety of realized measures of jump power variations, all formed on the basis of power transformations of instantaneous returns (i.e., $|r_t|^q$), as first discussed in Ding, Granger and Engle (1993) and Ding and Granger (1996). More specifically, we consider jump power variations with $0 \leq q \leq 6$, and construct a variety of estimators of jump risk, including upside and downside risk, jump asymmetry (i.e., realized signed jump power variation), and truncated jump measures. Our prediction experiments use high frequency price returns constructed using S&P 500 futures data as well as stocks in the Dow 30, for the period 1993-2009 period; and our empirical implementation involves estimating members of the linear and nonlinear extended Heterogeneous Autoregressive of the Realized Volatility (HAR-RV) class of models. Our findings suggest that past "large" jump power variations help less in the prediction of future realized volatility, than past "small" jump power variations. This in turn suggests the "larger" jumps might help less in the prediction of future realized volatility than "smaller" jumps. Our empirical findings also suggest that past realized signed jump power variations, which have not previously been examined in this literature, are strongly correlated with future volatility, and that past downside jump variations matter in prediction. Finally, incorporation of downside and upside jump power variations does improve predictability, albeit to a limited extent. Overall, our findings are consistent with ABD (2007) in the concluding that continuous components dominate, when predicting volatility.

Chapter 2

Density and Conditional Distribution Based Specification Analysis

2.1 Introduction

The last three decades have provided a unique opportunity to observe numerous interesting developments in finance, financial econometrics and statistics. For example, although starting as a narrow sub-field, financial econometrics has recently transformed itself into an important discipline, equipping financial economic researchers and industry practitioners with immensely helpful tools for estimation, testing and forecasting. One of these developments has involved the development of "state of the art" consistent specification tests for continuous time models, including not only the geometric Brownian motion process used to describe the dynamics of asset returns (Merton (1973)), but also a myriad of other diffusion models used in finance, such as the Ornstein-Uhlenbeck process introduced by Vacisek (1977), the constant elastic volatility process applied by Beckers (1980), the square root process due to Cox, Ingersoll and Ross (CIR: 1985), the so called CKLS model by Chan, Karolyi, Longstaff and Sanders (CKLS: 1992), various three factor models proposed Chen (1996), stochastic volatility processes such as generalized CIR of Andersen and Lund (1997), and the generic class of affine jump diffusion processes discussed in Duffie, Pan and

Singleton (2000).[1]

The plethora of available diffusion models allow decision makers to be flexible when choosing a specification to be subsequently used in contexts ranging from equity and option pricing, to term structure modeling and risk management. Moreover, the use of high frequency data when estimating such models, in continuous time contexts, allows investors to continuously update their dynamic trading strategies in real-time.[2] However, for statisticians and econometricians, the vast number of available models has important implications for formalizing model selection and specification testing methods. This has led to several key papers that have recently been published in the area of parametric and non-parametric specification testing. Most of the papers focus on the ongoing "search" for correct Markov and stationary models that "fit" historical data and associated dynamics. In this literature, it is important to note that correct specification of a joint distribution is not the same as that of a conditional distribution, and hence the recent focus on conditional distributions, given that most models have an interpretation as conditional models. In summary, the key issue in the construction of model selection and specification tests of conditional distributions is the fact that knowledge of the transition density (or conditional distribution) in general cannot be inferred from knowledge of the drift and variance terms of a diffusion model. If the functional form of the density is available parametrically, though, one can test the hypothesis of correct specification of a diffusion via the probability integral transform approach of Diebold, Gunther, and Tay (1998); the cross-spectrum approach of Hong (2001), Hong and Li (2005) and Hong, Li, and Zhao (2007); the martingalization-type Kolmogorov test of Bai (2003); or via the normality transformation approaches of Bontemps and Meddahi (2005) and Duan (2003). Furthermore, if the transition density is unknown, one can construct a non-parametric test by comparing a kernel density estimator of the actual and simulated data, for example, as in Altissimo and

[1] For complete details, see Section 2.2.

[2] For further discussion, see Duong and Swanson (2010, 2011).

Mele (2009) and Thompson (2008); or by comparing the conditional distribution of the simulated and the historical data, as in Bhardwaj, Corradi, and Swanson (BCS: 2008). One can also use the methods of Aït-Sahalia (2002) and Aït-Sahalia, Fan, and Peng (2009), in which they compare closed form approximations of conditional densities under the null, using data-driven kernel density estimates.

For clarity and ease of presentation, we categorize the above literature into two areas. The first area, initiated by the seminal work of Aït-Sahalia (1996) and later followed by Pritsker (1998) and Jiang (1998), breaks new ground in the continuous time specification testing literature by comparing marginal densities implied by hypothesized null models with nonparametric estimates thereof. These sorts of tests examine one factor specifications. The second area of testing, as initiated in Corradi and Swanson (CS: 2005) does not look at densities. Instead, they compare cumulative distributions (marginal, joint, or conditional) implied by a hypothesized null model with corresponding empirical distributions. A natural extension of these sorts of tests involves model selection amongst alternative predictive densities associated with competing models. While CS (2005) focus on cases where the functional form of the conditional density is known, BCS (2008) use simulation methods to examine testing in cases where the functional form of the conditional density is unknown. Corradi and Swanson (CS: 2011) and Cai and Swanson (2011) take the analysis of BCS (2008) on Step further, and focus on the comparison of out of sample predictive accuracy of possibly misspecified diffusion models, when the conditional distribution is not known in closed form (i.e., they "choose" amongst competing models based on predictive density model performance). The "best" model is selected by constructing tests that compare both predictive densities and/or predictive conditional confidence intervals associated with alternative models

In this chapter, we primarily focus our attention on the second area of the model selection and testing literature.[3] One feature of all of the tests that we shall discuss is that, given that they are based on the comparison of CDFs, they obtain parametric rates. Moreover, the tests can be used to evaluate single and multiple factor and dimensional models, regardless of whether or not the functional form of the conditional distribution is known.

In addition to discussing simple diffusion process specification tests of CS (2005), we discuss tests discussed in BCS (2008) and CS (2011), and provide some generalizations and additional results. In particular, parametric specification tests in the spirit of the conditional Kolmogorov test of Andrews (1997) that rely on block bootstrap resampling methods in order to construct test critical values are first discussed. Thereafter, extensions due to BCS (2008) for cases where the functional form of the conditional density is unknown are introduced, and related continuous time simulation methods are introduced. Finally, we broaden our discussion from single dimensional specification testing to multiple dimensional selection by discussing how to construct predictive densities and how to compare the accuracy of predictive densities derived from alternative (possibly misspecified) diffusion models as in CS (2011). In addition, we generalize simulation and testing procedures introduced in Cai and Swanson (2011) to more complicated multi-factor and multi-dimensional models where the number of latent variables larger than three. These final tests can be thought of as continuous time generalizations of the discrete time "reality check" test statistics of White (2000), which are widely used in empirical finance (see e.g. Sullivan, Timmermann and White (1999, 2001)). We finish the chapter with an empirical illustration of model selection amongst alternative short term interest rate models, drawing on BCS (2008), CS (2011) and Cai and Swanson (2011).

Of the final note is that the test statistics discussed here are implemented via use of simple bootstrap methods for critical value simulation. We use the bootstrap because

[3] For a recent survey on results in the first area of this literature, see Aït-Sahalia (2007).

the covariance kernels of the (Gaussian) asymptotic limiting distributions of the test statistics are shown to contain terms deriving from both the contribution of recursive parameter estimation error (PEE) and the time dependence of data. Asymptotic critical value thus cannot be tabulated in a usual way. Several methods can easily be implemented in this context. First one can use block bootstrapping procedures, as discussed below. Second one can use the conditional p-value approach of Corradi and Swanson (2002) which extends the work of Hansen (1996) and Inoue (2001) to the case of non vanishing parameter estimation error. Third is the subsampling method of Politis, Romano and Wolf (1999), which has clear efficiency "costs", but is easy implement. Use of the latter two methods yields simulated (or subsample based) critical values that diverge at rate equivalent to the blocksize length under the alternative. This is the main drawback to their use in our context. We therefore focus on use of a block bootstrap that mimics the contribution of parameter estimation error in a recursive setting and in the context of time series data. In general, use of the block bootstrap approach is made feasible by establishing consistency and asymptotic normality of both simulated generalized method of moments (SGMM) and nonparametric simulated quasi maximum likelihood (NPSQML) estimators of (possibly misspecified) diffusion models, in a recursive setting, and by establishing the first-order validity of their bootstrap analogs.

The rest of the chapter is organized as follows. In Section 2.2, we present our setup, and discuss various diffusion models used in finance and financial econometrics. Section 2.3 outlines the specification testing hypotheses, presents the cumulative distribution based test statistics for one factor and multiple factor models, discusses relevant procedures for simulation and estimation, and outlines bootstrap techniques that can be used for critical value tabulation. In Section 2.4, we present a small empirical illustration. Section 2.5 summarizes and concludes.

2.2 Setup

2.2.1 Diffusion Models in Finance and Financial Econometrics

For the past two decades, continuous time models have taken center stage in the field of financial econometrics, particularly in the context of structural modeling, option pricing, risk management, and volatility forecasting. One key advantage of continuous time models is that they allow financial econometricians to use the full information set that is available. With the availability of high frequency data and current computation capability, one can update information, model estimates, and predictions in milliseconds. In this Section we will summarize some of the standard models that have been used in asset pricing as well as term structure modelling. Generally, assume that financial asset returns follow Ito-semimartingale processes with jumps, which are the solution to the following stochastic differential equation system.

$$X(t_-) = \int_0^t b(X(s_-), \theta_0) ds - \lambda_0 t \int_Y y\phi(y) dy + \int_0^t \sigma(X(s_-), \theta_0) dW(s) + \sum_{j=1}^{J_t} y_j, \quad (2.1)$$

where $X(t_-)$ is a cadlag process (right continuous with left limit) for $t \in \Re^+$, and is an N dimensional vector of variables, $W(t)$ is an $N-$dimensional Brownian motion, $b(\cdot)$ is $N-$dimensional function of $X(t_-)$, and $\sigma(\cdot)$ is an NxN matrix-valued function of $X(t_-)$, where θ_0 is an unknown true parameter. J_t is a Poisson process with intensity parameter λ_0, λ_0 finite, and the $N-$dimensional jump size, y_j, is *iid* with marginal distribution given by ϕ. Both J_t and y_j are assumed to be independent of the driving Brownian motion, $W(t)$.[4] Also, note that $\int_Y y\phi(y)dy$ denotes the mean jump size, hereafter denoted by μ_0. Over a unit time interval, there are on average λ_0 jumps;

[4] Hereafter, $X(t_-)$ denotes the cadlag, while X_t denotes discrete skeleton for $t = 1, 2, \ldots$.

so that over the time span $[0,t]$, there are on average $\lambda_0 t$ jumps. The dynamics of $X(t_-)$ is then given by:

$$dX(t) = \left(b(X(t_-),\theta_0) - \lambda_0\mu_{y,0}\right)dt + \sigma(X(t_-),\theta_0)dW(t) + \int_Y yp(dy,dt), \qquad (2.2)$$

where $p(dy,dt)$ is a random Poisson measure giving point mass at y if a jump occurs in the interval dt, and $b(\cdot),\sigma(\cdot)$ are the "drift" and "volatility" functions defining the parametric specification of the model. Hereafter, the same (or similar) notation is used throughout when models are specified.

Though not an exhaustive list, we review some popular models. Models are presented with the "true" parameters.

Diffusion Models Without Jumps:

Geometric Brownian Motion (log normal model). In this set-up, $b(X(t_-),\theta_0) = b_0 X(t)$ and $\sigma(X(t_-),\theta_0) = \sigma_0 X(t)$

$$dX(t) = b_0 X(t)dt + \sigma_0 X(t)dW(t),$$

where b_0 and σ_0 are constants and and $W(t)$ is a one dimensional standard Brownian motion. (Below, other constants such as α_0 , β_0, λ_0, γ_0, δ_0, η_0, κ_0, and Ω_0 are also used in model specifications.)

This model is popular in the asset pricing literature. For example, one can model equity prices according to this process, especially in the Black-Scholes option set-up or in structured corporate finance.[5] The main drawback of this model is that the return process (log(price)) has constant volatility, and is not time varying. However, it is widely used as a convenient "first" econometric model.

Vasicek (1977) and Ornstein-Uhlenbeck process. The process is used to model

[5] See Black and Scholes (1973) for details.

asset prices, specifically in term structure modelling, and the specification is:

$$dX(t) = (\alpha_0 + \beta_0 X(t))dt + \sigma_0 dW(t)$$

where $W(t)$ is a standard Brownian motion, and α_0, β_0 and σ_0 are constants. β_0 is negative to ensure the mean reversion of $X(t)$.

Cox, Ingersoll and Ross (1995) use the following square root process to model the term structure of interest rates:

$$dX(t) = \kappa(\alpha_0 - X(t))dt + \sigma_0 \sqrt{X(t)} dW(t)$$

where $W(t)$ is a standard Brownian motion, α_0 is the long-run mean of $X(t)$, κ measures the speed of mean-reversion, and σ_0 is a standard deviation parameter and is assumed to be fixed. Also, non-negativity of the process is imposed, as $2\kappa\beta_0 > \sigma_0^2$.

Wong (1964) points out that in the CIR model, $X(t)$ with the dynamics evolving according to:

$$dX(t) = ((\alpha_0 - \lambda_0) - X(t))dt + \sqrt{\alpha_0 X(t)} dW(t), \ \alpha_0 > 0 \text{ and } \alpha_0 - \lambda_0 > 0 \quad (2.3)$$

belongs to the linear exponential (or Pearson) family with a closed form cumulative distribution. α_0 and λ_0 are fixed parameters of the model.

The Constant Elasticity of Variance, or CEV model is specified as follows:

$$dX(t) = \alpha_0 X(t)dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

where $W(t)$ is a standard Brownian motion and α_0, σ_0 and β_0 are fixed constants.

Of note is that the interpretation of this model depends on β_0, i.e. in the case of stock prices, if $\beta_0 = 2$, then the price process $X(t)$ follows a lognormal diffusion; if

$\beta_0 < 2$, then the model captures exactly the leverage effect as price and volatility are inversely correlated.

Among other authors, *Beckers (1980)* uses this CEV model for stocks, *Marsha and Rosenfeld (1983)* apply a CEV parametrization to interest rates and *Emanuel and Macbeth (1982)* utilize this set-up for option pricing.

The *Generalized constant elasticity of variance model is defined as follows:*

$$dX(t) = (\alpha_0 X(t)^{-(1-\beta_0)} + \lambda_0 X(t))dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

where the notation follows the CEV case. λ_0 is another parameter of the model. This process nests log diffusion when $\beta_0 = 2$, and nests square root diffusion when $\beta_0 = 1$.

Brennan and Schwartz (1979) and Courtadon (1982) analyze the model:

$$dX(t) = (\alpha_0 + \beta_0 X(t))dt + \sigma_0 X(t)^2 dW(t)$$

where $\alpha_0, \beta_0, \sigma_0$ are fixed constants and $W(t)$ is a standard Brownian motion.

Duffie and Kan (1993) study the specification:

$$dX(t) = (\alpha_0 - X(t))dt + \sqrt{\beta_0 + \gamma_0 X(t)} dW(t)$$

where $W(t)$ is a standard Brownian motion and α_0, β_0 and γ_0 are fixed parameters.

Aït-Sahalia (1996) looks at a general case with general drift and CEV diffusion:

$$dX(t) = (\alpha_0 + \beta_0 X(t) + \gamma_0 X(t)^2 + \eta_0 / X(t))dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

In the above expression, $\alpha_0, \beta_0, \gamma_0, \eta_0, \sigma_0$ and β_0 are fixed constants and $W(t)$ is again a standard Brownian motion.

Diffusion Models with Jumps*:*

For term structure modeling in empirical finance, the most widely studied class of models is the family of affine processes, including diffusion processes that incorporate jumps.

Affine Jump Diffusion Model: $X(t_{-})$ is defined to follow an affine jump diffusion if

$$dX(t) = \kappa_0(\alpha_0 - X(t))dt + \Omega_0\sqrt{D(t)}dW(t) + dJ(t)$$

where $X(t_{-})$ is an $N-$dimensional vector of variables of interest and is a cadlag process, $W(t)$ is an $N-$dimensional independent standard Brownian motion, κ_0 and Ω_0 are square $N \times N$ matrices, α_0 is a fixed long-run mean, $D(t)$ is a diagonal matrix with ith diagonal element given by

$$d_{ii}(t) = \theta_{0i} + \delta'_{0i}X(t)$$

In the above expressions, θ_{0i} and δ'_{0i} are constants. The jump intensity is assumed to be a positive, affine function of $X(t)$ and the jump size distribution is assumed to be determined by it's conditional characteristic function. The attractive feature of this class of affine jump diffusions is that, as shown in Duffie, Pan and Singleton (2000), it has an exponential affine structure that can be derived in closed form, i.e.

$$\Phi(X(t)) = \exp(a(t) + b(t)'X(t))$$

where the functions $a(t)$ and $b(t)$ can be derived from Riccati equations.[6] Given a known characteristic function, one can use either GMM to estimate the parameters of this jump diffusion, or one can use quasi-maximum likelihood (QML), once the first two moments are obtained. In the univariate case without jumps, as a special case, this corresponds to the above general CIR model with jumps.

[6]For details, see Singleton (2006), page 102.

Multifactor and Stochastic Volatility Model: Multifactor models have been widely used in the literature; particularly in option pricing, term structure, and asset pricing. One general set-up has $(X(t), V(t))' = \left(X(t), V^1(t), ..., V^d(t)\right)'$ where only the first element, the diffusion process X_t, is observed while $V(t) = (V^1(t), ..., V^d(t))'_{dx1}$ is latent. In addition, $X(t)$ can be dependent on $V(t)$. For instance, in empirical finance, the most well-known class of the multifactor models is the stochastic volatility model expressed as:

$$\begin{pmatrix} dX(t) \\ dV(t) \end{pmatrix} = \begin{pmatrix} b_1(X(t), \theta_0) \\ b_2(V(t), \theta_0) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(V(t), \theta_0) \\ 0 \end{pmatrix} dW_1(t) \tag{2.4}$$

$$+ \begin{pmatrix} \sigma_{12}(V(t), \theta_0) \\ \sigma_{22}(V(t), \theta_0) \end{pmatrix} dW_2(t), \tag{2.5}$$

where $W_1(t)_{1x1}$ and $W_2(t)_{1x1}$ are independent standard Brownian motions and $V(t)$ is latent volatility process. $b_1(\cdot)$ is a function of $X(t)$ and $b_2(\cdot), \sigma_{11}(\cdot), \sigma_{22}(\cdot)$ and $\sigma_{22}(\cdot)$ are general functions of $V(t)$, such that system of equations (2.4) is well-defined. Popular specifications are the square-root model of Heston (1993), the GARCH diffusion model of Nelson (1990), lognormal model of Hull and White (1987) and the eigenfunction models of Meddahi (2001). Note that in this stochastic volatility case, the dimension of volatility is $d = 1$. More general set-up can involve d driving Brownian motions in $V(t)$ equation.

As an example, *Andersen and Lund (1997)* study the generalized CIR model with stochastic volatility, specifically

$$dX(t) = \kappa_{x0}(\overline{x}_0 - X(t))dt + \sqrt{V(t)}dW_1(t)$$

$$dV(t) = \kappa_{v0}(\overline{v}_0 - V(t))dt + \sigma_{v0}\sqrt{V(t)}dW_2(t)$$

where $X(t)$ and $V(t)$ are price and volatility processes, respectively, $\kappa_{x0}, \kappa_{v0} > 0$ to ensure stationarity, $\overline{x}_0$ is the long-run mean of (log) price process, and $\overline{v}_0$ and σ_{v0} are constants. $W_1(t)$ and $W_2(t)$ are scalar Brownian motions. However, $W_1(t)$ and $W_2(t)$ are correlated such that $dW_1(t)dW_2(t) = \rho dt$ where the correlation ρ is some constant $\rho \in [-1, 1]$. Finally, note that volatility is a square-root diffusion process, which requires that $\kappa_{v0}\overline{v}_0 > \sigma_{v0}^2$.

Stochastic Volatility Model with Jumps (SVJ): A standard specification is:

$$
\begin{aligned}
dX(t) &= \kappa_{x0}\left(\overline{x}_0 - X(t)\right)dt + \sqrt{V(t)}dW_1(t) + J_u dq_u - J_d dq_d, \\
dV(t) &= \kappa_{v0}\left(\overline{v}_0 - V(t)\right)dt + \sigma_{v0}\sqrt{V(t)}dW_2(t),
\end{aligned}
$$

where q_u and q_d are Poisson processes with jump intensity parameters λ_u and λ_d respectively, and are independent of the Brownian motions $W_1(t)$ and $W_2(t)$. In particular, λ_u is the probability of a jump up, $\Pr(dq_u(t) = 1) = \lambda_u$ and λ_d is the probability of a jump down, $\Pr(dq_d(t) = 1) = \lambda_d$. J_u and J_d are jump up and jump down sizes and have exponential distributions: $f(J_u) = \frac{1}{\zeta_u}\exp\left(-\frac{J_u}{\zeta_u}\right)$ and $f(J_d) = \frac{1}{\zeta_d}\exp\left(-\frac{J_d}{\zeta_d}\right)$, where ζ_u, $\zeta_d > 0$ are the jump magnitudes, which are the means of the jumps, J_u and J_d.

Three Factor Model (CHEN): The three factor model combines various features of the above models, by considering a version of the oft examined 3-factor model due to Chan, Karolyi, Longstaff and Sanders (1992), which is discussed in detail in Dai and Singleton (2000). In particular,

$$
\begin{aligned}
dX(t) &= \kappa_{x0}\left(\theta(t) - X(t)\right)dt + \sqrt{V(t)}dW_1(t), \\
dV(t) &= \kappa_{v0}\left(\overline{v} - V(t)\right)dt + \sigma_{v0}\sqrt{V(t)}dW_2(t), \qquad (2.6) \\
d\theta(t) &= \kappa_{\theta 0}\left(\overline{\theta}_0 - \theta(t)\right)dt + \sigma_{\theta 0}\sqrt{\theta(t)}dW_3(t),
\end{aligned}
$$

where $W_1(t)$, $W_2(t)$ and $W_3(t)$ are independent Brownian motions, and V and θ are the stochastic volatility and stochastic mean of $X(t)$, respectively. $\kappa_{x0}, \kappa_{v0}, \kappa_{\theta 0}, \overline{v_0}, \overline{\theta_0}, \sigma_{v0}$, $\sigma_{\theta 0}$ are constants. As discussed above, non-negativity for $V(t)$ and $\theta(t)$ requires that $2\kappa_{v0}\overline{v}_0 > \sigma_{v0}^2$ and $2\kappa_{\theta 0}\overline{\theta_0} > \sigma_{\theta 0}^2$.

Three Factor Jump Diffusion Model (CHENJ): Andersen, Benzoni and Lund (2004) extend the three factor Chen (1996) model by incorporating jumps in the short rate process, hence improving the ability of the model to capture the effect of outliers, and to address the finding by Piazzesi (2004, 2005) that violent discontinuous movements in underlying measures may arise from monetary policy regime changes. The model is defined as follows:

$$\begin{aligned} dX(t) &= \kappa_{x0}(\theta(t) - X(t))\,dt + \sqrt{V(t)}dW_1(t) + J_u dq_u - J_d dq_d, \qquad (2.7) \\ dV(t) &= \kappa_{v0}(\overline{v_0} - V(t))\,dt + \sigma_{v0}\sqrt{V(t)}dW_2(t), \\ d\theta(t) &= \kappa_{\theta 0}\left(\overline{\theta_0} - \theta(t)\right)dt + \sigma_{\theta 0}\sqrt{\theta(t)}dW_3(t) \qquad (2.8) \end{aligned}$$

where all parameters are similar as in (2.6), $W_1(t)$, $W_2(t)$ and $W_3(t)$ are independent Brownian motions, q_u and q_d are Poisson processes with jump intensities λ_{u0} and λ_{d0}, respectively, and are independent of the Brownian motions $W_r(t)$, $W_v(t)$ and $W_\theta(t)$. In particular, λ_{u0} is the probability of a jump up, $\Pr(dq_u(t) = 1) = \lambda_{u0}$ and λ_{d0} is the probability of a jump down, $\Pr(dq_d(t) = 1) = \lambda_{d0}$. J_u and J_d are jump up and jump down sizes and have exponential distributions $f(J_u) = \frac{1}{\zeta_{u0}} \exp\left(-\frac{J_u}{\zeta_{u0}}\right)$ and $f(J_d) = \frac{1}{\zeta_{d0}} \exp\left(-\frac{J_d}{\zeta_{d0}}\right)$, where ζ_{u0}, $\zeta_{d0} > 0$ are the jump magnitudes, which are the means of the jumps J_u and J_d.

2.2.2 Overview on Specification Tests and Model Selection

The focus in this chapter is specification testing and model selection. The "tools" used in this literature have been long established. Several key classical contributions include the Kolmogorov-Smirnov test (see e.g. Kolmogorov (1933) and Smirnov (1939)), various results on empirical processes (see e.g. Andrews (1993) and the discussion in chapter 19 of van der Vaart (1998) on the contributions of Glivenko, Cantelli, Doob, Donsker and others), the probability integral transform (see e.g. Rosenblatt (1952)), and the Kullback-Leibler Information Criterion (see e.g. White (1982) and Vuong (1989)). For illustration, the empirical distribution mentioned above is crucial in our discussion of predictive densities because it is useful in estimation, testing, and model evaluation. Let Y_t is a variable of interest with distribution F and parameter θ_0. The theory of empirical distributions provides a result that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{Y_t \leq u\} - F(u|\theta_0))$$

satisfies a central limit theorem (with a parametric rate) if T is large (i.e., asymptotically). In the above expression, $1\{Y_t \leq u\}$ is the indicator function which takes value 1 if $Y_t \leq u$ and 0 otherwise. In the case where there is parameter estimation error, we can use more general results in chapter 19 of van der Vaart (1998). Define

$$P_T(f) = \frac{1}{T}\sum_{i=1}^{T} f(Y_i) \text{ and } P(f) = \int f dP$$

where P is a probability measure associated with F. Here, $P_n(f)$ converges to $P(f)$ almost surely for all the measurable functions f for which $P(f)$ is defined . Suppose one wants to test the null hypothesis that P belongs to a certain family $\{P_{\theta_0} : \theta_0 \in \Theta\}$, where θ_0 is unknown; it is natural to use a measure of the discrepancy between P_n and $P_{\widehat{\theta}}$ for a reasonable estimator $\widehat{\theta}_t$ of θ_0. In particular, if $\widehat{\theta}_t$ converges to θ_0 at a root-T rate, $\frac{1}{\sqrt{T}}(P_T - P_{\widehat{\theta}_t})$ has been shown to satisfy a central limit theorem.[7]

With regard to dynamic misspecification and parameter estimation error, the approach discussed for the class of tests in this chapter allows for the construction of

[7] See Theorem 19.23 in van der Vaart (1998) for details.

statistics that admit for dynamic misspecification under both hypotheses. This differs from other classes of tests such as the framework used by Diebold, Gunther and Tay (DGT: 1998), Hong (2001), and Bai (2003) in which correction dynamic specification under the null hypothesis is assumed. In particular, DGT use the probability integral transform to show that $F_t(Y_t|\Im_{t-1},\theta_0) = \int_{-\infty}^{Y_t} f_t(y|\Im_{t-1},\theta_0)dy$ is identically and independently distributed as a uniform random variable on $[0;1]$, where $F_t(\cdot)$ and $f_t\ (\cdot)$ are a parametric distribution and density with underlying parameter θ_0, Y_t is again our random variable of interest, and $\Im_t$ is the information set containing all "relevant" past information. They thus suggest using the difference between the empirical distribution of $F_t(Y_t|\Im_{t-1},\widehat{\theta}_t)$ and the $45°$ - degree line as a measure of "goodness of fit", where $\widehat{\theta}_t$ is some estimator of θ_0. This approach has been shown to be very useful for financial risk management (see e.g. Diebold, Hahnand, Tay (1999)), as well as for macroeconomic forecasting (see e.g. Diebold, Tay and Wallis (1998) and Clements and Smith (2000,2002)). Similarly, Bai (2003) develops a Kolmogorov type test of $F_t(Y_t|\Im_{t-1},\theta_0)$ on the basis of the discrepancy between $F_t(Y_t|\Im_{t-1},\widehat{\theta}_t)$ and the CDF of a uniform on $[0;1]$. As the test involves estimator $\widehat{\theta}_t$, the limiting distribution reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) proposes a novel approach based on a martingalization argument to construct a modified Kolmogorov test which has a nuisance parameter free limiting distribution. This test has power against violations of uniformity but not against violations of independence. Hong (2001) proposes another related interesting test, based on the generalized spectrum, which has power against both uniformity and independence violations, for the case in which the contribution of parameter estimation error vanishes asymptotically. If the null is rejected, Hong (2001) also proposes a test for uniformity robust to non independence, which is based on the comparison between a kernel density estimator and the uniform density. Two features differentiate the tests surveyed in this chapter from the tests outlined in the other papers mentioned above. First, the tests discussed here assume strict stationarity. Second, they allow for dynamic misspecification under the null hypothesis. The second feature allows us to obtain asymptotically valid critical values even when the conditioning information set does not contain all of the relevant past history. More precisely, assume that we are interested in testing for correct specification, given a particular information set which may or may not contain all of the relevant past information. This is important when a Kolmogorov test is constructed, as one is generally faced with the problem of defining $\Im_{t-1}$. If enough history is not included, then there may be dynamic misspecification. Additionally, finding out how much information (e.g. how many lags) to include may involve pre-testing, hence leading to a form of sequential test bias. By allowing for dynamic misspecification, such pre-testing is not required. Also note that critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Consider the following example. Assume that we are interested in testing whether the conditional distribution of $Y_t|Y_{t-1}$ follows normal distribution $N(\alpha_1 Y_{t-1}, \sigma_1)$. Suppose also that in actual fact the "relevant" information set has $\Im_{t-1}$ including both $Y_{t-1}$and Y_{t-2}, so that the true conditional model is $Y_t|\Im_{t-1} = Y_t|Y_{t-1},\, Y_{t-2} = N(\alpha_1 Y_{t-1} + \alpha_2 Y_{t-2}, \sigma_2)$. In this case, correct specification holds with respect to the information contained in X_{t-1}; but there is dynamic misspecification with respect to $Y_{t-1}$and Y_{t-2}. Even without taking account of parameter estimation error, the critical values obtained assuming correct dynamic specification are invalid, thus leading to invalid inference. Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under the null) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is not false. In summary, if one is interested in the particular problem of testing for correct specification for a given information set, then the approach of tests in this chapter is appropriate.

2.3 Consistent Distribution-Based Specification Tests and Predictive Density Type Model Selection for Diffusion Processes

2.3.1 One Factor Models

In this Section we outline the set-up for the general class of one factor jump diffusion specifications. All analysis carry through to the more complicated case of multi-factor stochastic volatility models which we will elaborate upon in the next Subsection. In the presentation of the tests, we follow a view that all candidate models, either single or multiple dimensional ones, are approximations of reality, and can thus be misspecified. The issue of correct specification (or misspecification) of a single model and the model selection test for choosing amongst multiple competing models allow for this feature.

To begin, fix the time interval $[0, T]$, consider a given single one factor candidate model the same as (2.1), with the true parameters θ_0, λ_0, μ_0 to be replaced by it's the pseudo true analogs $\theta^{\dagger}, \lambda, \mu$, respectively and $0 \leq t \leq T$:

$$X(t_{-}) = \int_0^t b(X(s_{-}), \theta^{\dagger})ds - \lambda t \int_Y y\phi(y)dy + \int_0^t \sigma(X(s_{-}), \theta^{\dagger})dW(s) + \sum_{j=1}^{J_t} y_j,$$

or

$$dX(t-) = \big(b(X(t-), \theta^{\dagger}) - \lambda\mu\big)\, dt + \sigma(X(t-), \theta^{\dagger})dW(t) + \int_Y yp(dy, dt), \qquad (2.9)$$

where variables are defined the same as in (2.1) and (2.2). Note that as the above model is the one factor version of (2.1) and (2.2) where the dimension of $X(t_{-})$ is 1x1, $W(t)$ is a one-dimensional standard Brownian motion and jump size, and y_j is one dimensional variable for all j. Also note that both J_t and y_j are assumed to be independent of the driving Brownian motion.

If the single model is correctly specified, then ,

$$
\begin{aligned}
b(X(t-),\theta^{\dagger}) &= b_0(X(t-),\theta_0) \\
\sigma(X(t-),\theta^{\dagger}) &= \sigma_0(X(t-),\theta_0) \\
\lambda &= \lambda_0, \mu = \mu_0 \ , \phi = \phi_0
\end{aligned}
$$

where $b_0(X(t-),\theta_0), \sigma_0(X(t-),\theta_0), \lambda_0, \mu_0 \ , \phi_0$ are unknown and belong to the true specification.

Now consider a different case (not a single model) where m candidate models are involved. For model k with $1 \leq k \leq m$, denote it's corresponding specification to be $(b_k(X(t_-),\theta_k^{\dagger})$, $\sigma_k(X(t_-),\theta_k^{\dagger})$, λ_k, μ_k, $\phi_k)$. Two scenarios immediate arise. Firstly, if the model k is correctly specified, then $b_k(X(t_-),\theta_k^{\dagger}) = b_0(X(t_-),\theta_0)$, $\sigma_k(X(t_-),\theta_k^{\dagger}) = \sigma_0(X(t_-),\theta_0)$, $\lambda_k = \lambda_0$, $\mu_k = \mu_0$ and $\phi_k = \phi_0$ which are similar to the case of a single model. In the second scenario, all the models are likely to be misspecified and modelers are faced with the choice of selecting the "best" one. This type of problem is well-fitted into the class of accuracy assessment tests initiated earlier by Diebold and Mariano (1995) or White (2000).

The tests discussed hereafter are Kolomogorov type tests based on the construction of cumulative distribution functions (CDFs). In a few cases, the CDF is known in closed form. For instance, for the simplified version of the CIR model as in (2.3), $X(t)$ belongs to the linear exponential (or Pearson) family with the gamma CDF of the form:[8]

$$
F(u,\alpha,\lambda) = \frac{\int_0^u (\frac{\lambda}{2})^{-2(1-\alpha/\lambda)-1} \exp(-x/(\frac{\lambda}{2}))dx}{\Gamma(2(1-\alpha/\lambda))}, \text{ where } \Gamma(x) = \int_0^{\infty} t^x \exp(-t)dt, \tag{2.10}
$$

and α, λ are constants.

[8] See Wong (1964) for details.

Furthermore, if we look at the pure diffusion process without jumps:

$$dX(t) = b(X(t), \theta^\dagger)dt + \sigma(X(t), \theta^\dagger)dW(t), \qquad (2.11)$$

where $b(\cdot)$ and $\sigma = \sigma(\cdot)$ are drift and volatility functions, it is known that the stationary density, say $f(x, \theta^\dagger)$, associated with the invariant probability measure can be expressed explicitly as:[9]

$$f(x, \theta^\dagger) = \frac{c(\theta^\dagger)}{\sigma^2(x, \theta^\dagger)} \exp\left(\int^x \frac{2b(u, \theta^\dagger)}{\sigma^2(u, \theta^\dagger)} du \right)$$

where $c(\theta^\dagger)$ is a constant ensuring that f integrates to one. The CDF, say $F(u, \theta^\dagger) = \int^u f(x, \theta^\dagger)dx$, can then be obtained using available numerical integration procedures.

However, in most cases, it is impossible to derive the CDFs in closed form. To obtain a CDF in such cases, a more general approach is to use simulation. Instead of estimating the CDF directly, simulation techniques estimates the CDF indirectly utilizing it's generated sample paths and the theory of empirical distributions. The specification of a specific diffusion process will dictate the sample paths and thereby corresponding test outcomes.

Note that in the historical context, many early papers in this literature are probability density-based. For example, in a seminal paper, Ait-Sahalia (1996) compares the marginal densities implied by hypothesized null models with nonparametric estimates thereof. Following the same framework of correct specification tests, CS(2005) and BCS (2008), however, do not look at densities. Instead, they compare the cumulative distribution (marginal or joint) implied by a hypothesized null model with the corresponding empirical distribution. While CS (2005) focus on the known unconditional distribution, BCS (2008) look at the conditional simulated distributions. CS (2011) make extensions to multiple models in the context of out of sample accuracy

[9] See Karlin and Taylor (1981) for details.

assessment tests. This approach is somewhat novel to this continuous time model testing literature.

Now suppose we observe a discrete sample path $X_{1,}X_2, ..., X_T$ (also referred as skeletons).[10] The corresponding hypotheses can be set up as follows:

Hypothesis 1: Unconditional Distribution Specification Test of a Single Model

$H_0 : F(u, \theta^\dagger) = F_0(u, \theta_0)$, for all u, a.s.

H_A : $\Pr\left(F(u, \theta^\dagger) - F_0(u, \theta_0) \neq 0\right) > 0$, for some $u \in U$, with non-zero Lebesgue measure.

where $F_0(u, \theta_0)$ is the true cumulative distribution implied by the above density, i.e. $F_0(u, \theta_0) = \Pr(X_t \leq u)$. $F(u, \theta^\dagger) = \Pr\left(X_t^{\theta^\dagger} \leq u\right)$ is the cumulative distribution of the proposed model. $X_t^{\theta^\dagger}$ is a skeleton implied by model (2.9).

Hypothesis 2: Conditional Distribution Specification Test of A Single Model

$H_0 : F_\tau(u|X_t, \theta^\dagger) = F_{0,\tau}(u|X_t, \theta_0)$, for all u, a.s.

H_A : $\Pr\left(F_\tau(u|X_t, \theta^\dagger) - F_{0,\tau}(u|X_t, \theta_0) \neq 0\right) > 0$, for some $u \in U$, with non-zero Lebesgue measure.

where

$$F_\tau(u|X_t, \theta^\dagger) = \Pr\left(X_{t+\tau}^{\theta^\dagger} \leq u|X_t^{\theta^\dagger} = X_t\right)$$

is τ-Step ahead conditional distributions and $t = 1, ..., T - \tau$. $F_{0,\tau}(u|X_t, \theta_0)$ is τ-Step ahead true conditional distributions.

Hypothesis 3: Predictive Density Test for Choosing Amongst Multiple Competing Models

The null hypothesis is that no model can outperform model 1 which is the benchmark model.[11]

[10] As mentioned earlier, we follow CS (2005) by using notation $X(\cdot)$ when defining continuous time processes and X_t for a skeleton.

[11] See White (2000) for a discussion of a discrete time series analog to this case, whereby point

$$H_0 : \max_{k=2,\ldots,m} E_X\left(\left(F_{X^{\theta_1^\dagger}_{1,t+\tau}(X_t)}(u_2) - F_{X^{\theta_1^\dagger}_{1,t+\tau}(X_t)}(u_1)\right) - (F_0(u_2|X_t) - F_0(u_1|X_t))\right)^2$$
$$-E_X\left(\left(F_{X^{\theta_k^\dagger}_{k,t+\tau}(X_t)}(u_2) - F_{X^{\theta_k^\dagger}_{k,t+\tau}(X_t)}(u_1)\right) - (F_0(u_2|X_t) - F_0(u_1|X_t))\right)^2 \leq 0$$

H_A : negation of H_0'

where $F_{X^{\theta_k^\dagger}_{k,t+\tau}(X_t)}(u) = F_k^\tau(u|X_t, \theta_k^\dagger) = P^\tau_{\theta_k^\dagger}\left(X^{\theta_k^\dagger}_{t+\tau} \leq u | X_t^{\theta_k^\dagger} = X_t\right)$, which is the conditional distribution of $X_{t+\tau}$, given X_t, and evaluated at u under the probability law generated by model k. $X^{\theta_k^\dagger}_{k,t+\tau}(X_t)$ with $1 \leq \tau \leq T-t$ is the skeleton implied by model k, parameter $\theta_k^\dagger$ and initial value X_t. Analogously, define $F_0^\tau(u|X_t, \theta_0) = P^\tau_{\theta_0}(X_{t+\tau} \leq u|X_t)$ to be the "true" conditional distribution.

Note that the three hypotheses expressed above apply exactly the same to the case of multifactor diffusions. Now, before moving to the statistics description section, we briefly explain the intuitions in facilitating construction of the tests:

In the first case (Hypothesis 1), CS (2005) construct a Kolomogorov type test based on comparison of the empirical distribution and the unconditional CDF implied by the specification of the drift, variance and jumps. Specifically, one can look at the scaled difference between

$$F(u, \theta^\dagger) = \Pr\left(X_t^{\theta^\dagger} \leq u\right) = \int^u f(x, \theta^\dagger)dx$$

and estimator of the true $F_0(u|X_t, \theta_0)$, the empirical distribution of X_t defined as:

$$\frac{1}{T}\sum_{t=1}^{T} 1\left\{X_t \leq u\right\}$$

where $1\{Y_t \leq u\}$ is indicator function which takes value 1 if $Y_t \leq u$ and 0 otherwise.

Similarly for the second case of conditional distribution (Hypothesis 2), the test

rather than density-based loss is considered; Corradi and Swanson (2007b) for an extension of White (2000) that allows for parameter estimation error; and Corradi and Swanson (2006a) for an extension of Corradi and Swanson (2007b) that allows for the comparison of conditional distributions and densities in a discrete time series context.

statistic V_T can be a measure of the distance between the $\tau - Step$ ahead conditional distribution of $X^{\theta^\dagger}_{t+\tau}$, given $X^{\theta^\dagger}_t = X_t$, as:

$$F_\tau(u|X_t, \theta^\dagger) = \Pr\left(X^{\theta^\dagger}_{t+\tau} \leq u|X^{\theta^\dagger}_t = X_t\right),$$

to an estimator of the true $F_{0,\tau}(u|X_t, \theta_0)$, the conditional empirical distribution of $X_{t+\tau}$ conditional on the initial value X_t defined as:

$$\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} 1\left\{X_{t+\tau} \leq u\right\},$$

In the third case (Hypothesis 3), model accuracy is measured in terms of a distributional analog of mean square error. As is commonplace in the out-of-sample evaluation literature, the sample of T observations is divided into two subsamples, such that $T = R + P$, where only the last P observations are used for predictive evaluation. A $\tau-$Step ahead prediction error under model k is $1\{u_1 \leq X_{t+\tau} \leq u_2\} - \left(F^\tau_k(u_2|X_t, \theta^\dagger_k) - F^\tau_k(u_1|X_t, \theta^\dagger_k)\right)$ where $2 \leq k \leq m$ and similarly for model 1 by replacing index k with index 1. Suppose we can simulate $P - \tau$ paths of $\tau-$Step ahead skeleton[12] using X_t as starting values where $t = R, ..., R + P - \tau$, from which we can construct a sample of $P - \tau$ prediction errors. Then, these prediction errors can be used to construct a test statistic for model comparison. In particular, model 1 is defined to be more accurate than model k if:

$$\begin{aligned} & E\left(\left((F^\tau_1(u_2|X_t, \theta^\dagger_1) - F^\tau_1(u_1|X_t, \theta^\dagger_1)) - (F^\tau_0(u_2|X_t, \theta_0) - F^\tau_0(u_1|X_t, \theta_0))\right)^2\right) \\ < \; & E\left(\left((F^\tau_k(u^\tau_2|X_t, \theta^\dagger_k) - F^\tau_k(u^\tau_1|X_t, \theta^\dagger_k)) - (F^\tau_0(u_2|X_t, \theta_0) - F^\tau_0(u_1|X_t, \theta_0))\right)^2\right). \end{aligned}$$

where $E(\cdot)$ is an expectation operator and $E\left(1\{u_1 \leq X_{t+\tau} \leq u_2\}|X_t\right) = F^\tau_0(u_2|X_t, \theta_0) - F^\tau_0(u_1|X_t, \theta_0)$. Concretely, model k is worse than model 1 if on average $\tau-$Step ahead

[12] See Section 2.3.3.1 for model simulation details.

prediction errors under model k is larger than that of model 1.

Finally, it is important to point out some main features characterized by all the three test statistics. Processes $X(t)$ hereafter is required to satisfy the regular conditions, i.e. assumptions A1-A8 in CS (2011). Regarding model estimation (in Section 2.3.3), $\theta^\dagger$ and $\theta_k^\dagger$ are unobserved and need to be estimated. While CS (2005), BCS (2008) utilize (recursive) Simulated General Method of Moments (SGMM), CS (2011) make extension to (recursive) Nonparametric Simulated Quasi Maximum Likelihood (NPSQML). For the unknown distribution and conditional distribution, it will be pointed out in Section 2.3.3.2 that $F(u, \theta^\dagger)$, $F_\tau(u|X_t, \theta^\dagger)$ and $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u)$ can be replaced by their simulated counterparts using the (recursive) SGMM and NPSQML parameter estimators. In addition, test statistics converge to functional of Gaussian processes with covariance kernels that reflect time dependence of the data and the contribution of parameter estimation error (PEE). Limiting distributions are not nuisance parameter free and critical values thereby cannot be tabulated by the standard approach. All the tests discussed in this chapter rely on the bootstrap procedures for obtaining the asymptotically valid critical values, which we will describe in Section 2.3.4.

2.3.1.1 Unconditional Distribution Tests

For one-factor diffusions, we outline the construction of unconditional test statistics in the context where CDF is known in closed form. In order to test the **Hypothesis 1**, consider the following statistic:

$$V_{T,N,h}^2 = \int_U V_{T,N,h}^2(u)\pi(u)du,$$

where

$$V_{T,N,h} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(1\{X_t \leq u\} - F(u, \widehat{\theta}_{T,N,h})\Big)$$

In the above expression, U is a compact interval and $\int_U \pi(u)du = 1$, $1\{X_t \leq u\}$ is again the indicator function which returns value 1 if $X_t \leq u$ and 0 otherwise. Further, as defined in Section 2.3.3, $\widehat{\theta}_{T,N,h}$ hereafter is a simulated estimator where T is sample size and h is the discretization interval used in simulation. In addition, with the abuse of notation, N is a generic notation throughout this chapter, i.e. $N = L$, the length of each simulation path for (recursive) SGMM and $N = M$, the number of random draws (simulated paths) for (recursive) NPQML estimator.[13] Also note in our notation that as the above test is in sample specification test, the estimator and the statistics are constructed using the entire sample, i.e. $\widehat{\theta}_{T,N,h}$.

It has been shown in CS (2005) that under regular conditions and if the estimator is estimated by SGMM, the above statistics converges to a functional of Gaussian process.[14] In particular, pick the choice $T, N \to \infty, h \to 0, T/N \to 0$ and $Th^2 \to 0$

Under the null,

$$V^2_{T,N,h} \to \int_U Z^2(u)\pi(u)$$

where Z is a Gaussian process with covariance kernel. Hence, the limiting distribution of $V^2_{T,N,h}$ is a functional of a Gaussian process with a covariance kernel that reflects both PEE and the time series nature of the data. As $\widehat{\theta}_{T,N,h}$ is root-T consistent, PEE does not disappear in the asymptotic covariance kernel.

Under H_A, there exists an $\varepsilon > 0$ such that

$$\lim_{T\to\infty} \Pr(\frac{1}{T} V^2_{T,N,h} > \varepsilon) = 1$$

For the asymptotic critical value tabulation, we use the bootstrap procedure. In order to establish validity of the block bootstrap under SGMM with the presence of PEE, the simulated sample size should be chosen to grow at a faster rate than the

[13] M is often chosen to coincide with S, the number of simulated paths used when simulating distributions.

[14] For details and the proof, see Theorem 1 in CS (2005).

historical sample, i.e. $T/N \to 0$.

Thus, we can follow Steps in appropriate bootstrap procedure in Section 2.3.4. For instance, if the SGMM estimator is used, the bootstrap statistic is

$$V^{2*}_{T,N,h} = \int_U V^{2*}_{T,N,h}(u)\pi(u)du,$$

where

$$V^{*}_{T,N,h} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big((1\{X^*_t \le u\} - 1\{X_t \le u\}) - (F(u, \widehat{\theta}^*_{T,N,h}) - F(u, \widehat{\theta}_{T,N,h})) \Big) .$$

In the above expression, $\widehat{\theta}^*_{T,N,h}$ is the bootstrap analog of $\widehat{\theta}_{T,N,h}$ and is estimated by the bootstrap sample $X^*_1, ..., X^*_T$ (see Section 2.3.4). With appropriate conditions, CS (2005) show that under the null, $V^{2*}_{T,N,h}$ has a well defined limiting distribution which coincides with that of $V^2_{T,N,h}$. We then can straightforwardly derive the bootstrap critical value by following Step 1-5 Section 2.3.4. In particular, in Step 5, the idea is to perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrap statistics. Reject H_0 if $V^2_{T,N,h}$ is greater than the $(1-\alpha)th-$percentile of this empirical distribution. Otherwise, do not reject H_0.

2.3.1.2 Conditional Distribution Tests

Hypothesis 2 tests correct specification of the conditional distribution, implied by a proposed diffusion model. In practice, the difficulty arises from the fact that the functional form of neither τ-Step ahead conditional distributions $F_\tau(u|X_t, \theta^\dagger)$ nor $F_{0,\tau}(u|X_t, \theta_0)$ is unknown in most cases. Therefore, BCS (2008) develop bootstrap specification test on the basis of simulated distribution using the SGMM estimator.[15]

[15] In this chapter, we assume that $X(\cdot)$ satisfies the regularity conditions stated in CS (2011), i.e. assuptions A1-A8. Those conditions also reflect requirements A1-A2 in BCS (2008). Note that, the SGMM estimator used in BCS (2008) satisfies the root-N consistency condition that CS (2011) impose on their parameter estimator (See Assumption 4).

With the important inputs leading to the test such as simulated estimator, distribution simulation and bootstrap procedures to be presented in the next Section[16], the test statistic is defined as:

$$Z_T = \sup_{u \times v \in U \times V} |Z_T(u, v)|$$

where

$$Z_T(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1 \{X_t \leq v\},$$

with U and V compact sets on the real line. $\widehat{\theta}_{T,N,h}$ is the simulated estimator using entire sample $X_{1,...,}X_T$ and S is the number of simulated replications used in the estimation of conditional distributions as described in Section 2.3.3. If SGMM estimator is used (similar to unconditional distribution case and the same as in BCS (2008)), then $N = L$, where L is the simulation length used in parameter estimation.

The above statistic is a simulation-based version of the conditional Kolmogorov test of Andrews (1997), which compare the joint empirical distribution

$$\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} 1\{X_{t+\tau} \leq u\} 1 \{X_t \leq v\}$$

with its semi-empirical/semi-parametric analog given by the product of

$$\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} F_{0,\tau}(u|X_t, \theta_0) 1 \{X_t \leq v\}.$$

Intuitively, if the null is not rejected, the metric distance between the two should asymptotically disappear. In the simulation context with parameter estimation error, the asymptotic limit of Z_T however is a nontrivial one. BCS (2008) show that with the

[16] See Sections 2.3.3 and 2.3.4 for further details.

proper choice of T, N, S, h, i.e. $T, N, S, T^2/S \to \infty$ and $h, T/N, T/S, Nh, h^2 T \to 0$, then

$$Z_T \xrightarrow{d} \sup_{u \times v \in U \times V} |Z(u, v)|,$$

where $Z(u, v)$ is a Gaussian process with a covariance kernel that characterizes: 1) long-run variance we would have if we knew $F_{0,\tau}(u|X_1, \theta_0)$; 2) the contribution of parameter estimation error; 3) The correlation between the first two.

Furthermore, under H_A, there exists some $\varepsilon > 0$ such that:

$$\lim_{P \to \infty} \Pr\left(\frac{1}{\sqrt{T}} Z_T > \varepsilon\right) = 1.$$

As $T/S \to 0$, the contribution of simulation error is asymptotically negligible. The limiting distribution is not nuisance parameter free and hence critical values cannot be tabulated directly from it. The appropriate bootstrap statistic in this context is:

$$Z_T^* = \sup_{u \times v \in U \times V} |Z_T^*(u, v)|,$$

where

$$\begin{aligned} Z_T^*(u, v) &= \frac{1}{\sqrt{T - \tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}^*} \le u \right\} - 1\{X_{t+\tau}^* \le u\} \right) 1\{X_t^* \le v\} \\ & \quad - \frac{1}{\sqrt{T - \tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \le u \right\} - 1\{X_{t+\tau} \le u\} \right) 1\{X_t \le v\} \end{aligned}$$

In the above expression, $\widehat{\theta}_{T,N,h}^*$ is the bootstrap parameter estimated using the resampled data X_t^* for $t = 1, ..., T - \tau$. $X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}^*}$, $s = 1, ..., S$ and $t = 1, ..., T - \tau$ is the simulated data under $\widehat{\theta}_{T,N,h}^*$ and X_t^*, $t = 1, ..., T - \tau$ is a resampled series constructed using standard block-bootstrap methods as described in 2.3.4. Note that in the original paper, BCS (2008) propose bootstrap SGMM estimator for conditional distribution of diffusion processes. CS (2011) extend the test to the case of simulated recursive NPSQML estimator. Regarding the generation of the empirical distribution of Z_T^* (asthmatically the same as Z_T), follow Step 1-5 in the bootstrap procedure in Section 2.3.4. This yields B bootstrap replications (B large) of Z_T^*. One can then compare Z_T with the percentiles of the empirical distribution of Z_T^*, and reject H_0 if Z_T is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject H_0. Tests carried out in this manner are correctly asymptotically sized, and have unit asymptotic power.

2.3.1.3 Predictive Density Tests for Multiple Competing Models

In many circumstances, one might want to compare one (benchmark) model (model 1) against multiple competing models (models k, $2 \leq k \leq m$). In this case, recall in the null in **Hypothesis 3** is that no model can outperform the benchmark model. In testing the null, we first choose a particular interval i.e., $(u_1, u_2) \in U\text{x}U$ where U is a compact set so that the objective is evaluation of predictive densities for a given range of values. In addition, in the recursive setting (not full sample is used to estimate parameters), if we use the recursive NPSQML estimator, say $\widehat{\theta}_{1,t,N,h}$ and $\widehat{\theta}_{k,t,N,h}$, for models 1 and k, respectively, then the test statistic is defined as

$$D_{k,P,S}^{Max}(u_1, u_2) = \max_{k=2,\ldots,m} D_{k,P,S}(u_1, u_2).$$

where

$$\begin{aligned} & D_{k,P,S}(u_1, u_2) \\ = \ & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\left[\frac{1}{S} \sum_{i=1}^{S} 1 \left\{ u_1 \leq X_{1,i,t+\tau}^{\widehat{\theta}_{1,t,N,h}}(X_t) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau} \leq u_2\} \right]^2 \right. \\ & \left. - \left[\frac{1}{S} \sum_{i=1}^{S} 1 \left\{ u_1 \leq X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau} \leq u_2\} \right]^2 \right). \end{aligned}$$

All notation is consistent with previous Sections where S is the number of simulated replications used in the estimation of conditional distributions. $X_{1,i,t+\tau}^{\widehat{\theta}_{1,t,N,h}}(X_t)$ and $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}$, $i = 1, ..., S$, $t = 1, ..., T - \tau$, are the ith simulated path under $\widehat{\theta}_{1,t,N,h}$ and $\widehat{\theta}_{k,t,N,h}$. If models 1 and k are nonnested for at least one $k = 2, ..., m$. Under regular conditions and if P, R, S, h are chosen such as $P, R, N \to \infty$ and $h, P/N, h^2P \to 0$, $P/R \to \pi$ where π is finite then

$$\max_{k=2,..,m} \left(D_{k,P,N}(u_1, u_2) - \mu_k(u_1, u_2)\right) \xrightarrow{d} \max_{k=2,...,m} Z_k(u_1, u_2),$$

where, with an abuse of notation, $\mu_k(u_1, u_2) = \mu_1(u_1, u_2) - \mu_k(u_1, u_2)$, and

$$\mu_j(u_1, u_2) = E\left(\left(\left(F_{X_{j,t+\tau}^{\theta_j^\dagger}(X_t)}(u_2) - F_{X_{j,t+\tau}^{\theta_j^\dagger}(X_t)}(u_1)\right) - \left(F_0(u_2|X_t) - F_0(u_1|X_t)\right)\right)^2\right),$$

for $j = 1, ..., m$, and where $(Z_1(u_1, u_2), ..., Z_m(u_1, u_2))$ is an $m-$dimensional Gaussian random variable the covariance kernels that involves error in parameter estimation. Bootstrap statistics are therefore required to reflect this parameter estimation error issue.[17]

In the implementation, we can obtain the asymptotic critical value using a recursive version of the block bootstrap. The idea is that when forming block bootstrap samples in the recursive setting, observations at the beginning of the sample are used more frequently than observations at the end of the sample. We can replicate the Step 1-5 in bootstrap procedure in Section 2.3.4. It should be stressed the re-sampling in the Step 1 is the recursive one. Specifically, begin by resampling b blocks of length l from the full sample, with $lb = T$. For any given τ, it is necessary to jointly resample $X_t, X_{t+1}, ..., X_{t+\tau}$. More precisely, let $Z^{t,\tau} = (X_t, X_{t+1}, ..., X_{t+\tau})$, $t = 1, ..., T - \tau$. Now, resample b overlapping blocks of length l from $Z^{t,\tau}$. This yields $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T - \tau$. Use these data to construct bootstrap

[17] See CS (2011) for further discussion.

estimator $\widehat{\theta}^*_{k,t,N,h}$. Recall that N is chosen in CS (2011) as the number of simulated series used to estimate the parameters ($N = M = S$) and such as $N/R, N/P \to \infty$. Under this condition, simulation error vanishes and there is no need to resample the simulated series.

CS (2011) show that

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}^*_{k,t,N,h} - \widehat{\theta}_{k,t,N,h}\right)$$

has the same limiting distribution as

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}_{k,t,N,h} - \theta^\dagger_k\right),$$

conditional on all samples except a set with probability measure approaching zero. Given this, the appropriate bootstrap statistic is:

$$\begin{aligned}
& D^*_{k,P,S}(u_1, u_2) \\
= \; & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\Bigg\{\Bigg(\Bigg[\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X^{\widehat{\theta}^*_{1,t,N,h}}_{1,i,t+\tau}(X^*_t) \leq u_2\right\} - 1\{u_1 \leq X^*_{t+\tau} \leq u_2\}\Bigg]^2 \\
& - \left(\frac{1}{T}\sum_{j=1}^{T}\left[\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X^{\widehat{\theta}_{1,t,N,h}}_{1,i,t+\tau}(X_j) \leq u_2\right\} - 1\{u_1 \leq X_{j+\tau} \leq u_2\}\right]^2\right)\Bigg) \\
& - \Bigg(\Bigg[\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X^{\widehat{\theta}^*_{k,t,N,h}}_{k,i,t+\tau}(X^*_t) \leq u_2\right\} - 1\{u_1 \leq X^*_{t+\tau} \leq u_2\}\Bigg]^2 \\
& - \left(\frac{1}{S}\sum_{j=1}^{S}\left[\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X^{\widehat{\theta}_{k,t,N,h}}_{k,i,t+\tau}(X_j) \leq u_2\right\} - 1\{u_1 \leq X_{j+\tau} \leq u_2\}\right]^2\right)\Bigg)\Bigg\}.
\end{aligned}$$

As the bootstrap statistic is calculated from the last P resampled observations, it is necessary to have each bootstrap term recentered around the (full) sample mean. This is true even in the case there is no need to mimic PEE, i.e. the choice of P, R is such that $P/R \to 0$. In such a case, above statistic can be formed using $\widehat{\theta}_{k,t,N,h}$ rather than $\widehat{\theta}^*_{k,t,N,h}$.

For any bootstrap replication, repeat B times (B large)) bootstrap replications which yield B bootstrap statistics $D^*_{k,P,S}$. Reject H_0 if $D_{k,P,S}$ is greater than the $(1-\alpha)th$-percentile of the bootstrap empirical distribution. For numerical implementation, it is of importance to note that in the case where $P/R \to 0, P, T, R \to \infty$, there is no need to re-estimate $\widehat{\theta}^*_{1,t,N,h}$ $(\widehat{\theta}^*_{k,t,N,h})$. Namely, $\widehat{\theta}_{1,t,N,h}(\widehat{\theta}_{k,t,N,h})$ can be used in all bootstrap experiments.

Of course, the above framework can also be applied using entire simulated distributions rather than predictive densities, by simply estimating parameters once, using the entire sample, as opposed to using recursive estimation techniques, say, as when forming predictions and associated predictive densities.

2.3.2 Multifactor Models

Now, let us turn our attention to multifactor diffusion models of the form

$$\left(X(t), V^1(t), ..., V^d(t)\right)',$$

where only the first element, the diffusion process X_t, is observed while $V(t) = (V^1(t), ..., V^d(t))'$ is latent. The most popular class of the multifactor models is stochastic volatility model expressed as below:

$$\begin{pmatrix} dX(t) \\ dV(t) \end{pmatrix} = \begin{pmatrix} b_1(X(t), \theta^\dagger) \\ b_2(V(t), \theta^\dagger) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(V(t), \theta^\dagger) \\ 0 \end{pmatrix} dW_1(t) \quad (2.12)$$

$$+ \begin{pmatrix} \sigma_{12}(V(t), \theta^\dagger) \\ \sigma_{22}(V(t), \theta^\dagger) \end{pmatrix} dW_2(t), \quad (2.13)$$

where $W_1(t)_{1\text{x}1}$ and $W_2(t)_{1\text{x}1}$ are independent Brownian Motions.[18] For instance, many term structure models require the multifactor specification of the above form (see Dai and Singleton (2000)). In a more complicated case, the drift function can also be specified to be a stochastic process which poses even more challenges to testing. As mentioned earlier, the hypotheses (**Hypothesis 1,2,3**) and the test construction strategy for multifactor models are the same as for one factor model. All theory essentially applies immediately to multifactor cases. In implementation, the key difference is in the simulated approximation scheme facilitating parameter and CDF estimation. $X(t)$ cannot simply be expressed as a function of $d+1$ driving Brownian motions but instead involves a function of $(W_{jt}, \int_0^t W_{js} dW_{is}\ )$, $i,j = 1, ..., d+1$ (see e.g. Pardoux and Talay (1985) p.30-32 and CS(2005)).

For illustration, we hereafter focus on the analysis of a stochastic volatility model (2.12) where drift and diffusion coefficients can be written as

$$b = \begin{pmatrix} b_1(X(t), \theta^\dagger)) \\ b_2(V(t), \theta^\dagger)) \end{pmatrix}, \ \sigma = \begin{pmatrix} \sigma_{11}(V(t), \theta^\dagger) & \sigma_{12}(V(t), \theta^\dagger) \\ 0 & \sigma_{22}(V(t), \theta^\dagger) \end{pmatrix}$$

We also examine a three factor model (i.e., the Chen Model as in (2.6)) and a three factor model with jumps, (i.e., CHENJ as in (2.7)). By presenting two and three factor models as an extension of our above discussion, we make it clear that specification tests of multiple factor diffusions with $d \geq 3$ can be easily constructed in similar manner.

In distribution estimation, the important challenge for multifactor models lies in the missing variable issue. In particular, for simulation of X_t, one needs initial values of the latent processes $V_{1,...}, V_d$, which are unobserved. To overcome this problem, it suffices to simulate the process using different random initial values for the volatility

[18]Note that the dimension of $X(\cdot)$ can be higher and we can add jumps to the above specification such that it satisfies the regularity conditions outlined in the one factor case. In addition, CS (2005), provide a detailed discussion of approximation schemes in the context of stochastic volatility models.

process, then construct the simulated distribution using those initial values and average them out. This allows one to integrate out the effect of a particular choice of volatility initial value.

For clarity of exposition, we sketch out a simulation strategy for a general model of d latent variables in Section 2.3.3. This generalizes the simulation scheme of three factor models in the Cai and Swanson (2011). As a final remark before moving to the statistic presentation, note that the class of multifactor diffusion processes considered in this chapter is required to match the regular conditions as in previous Section (assumption from A1-A8 in CS (2011) with A4 being replaced by A4').

2.3.2.1 Unconditional Distribution Tests

Following the above discussion on test construction, we specialize to the case of two-factor stochastic volatility models. Extension to general multidimensional and multifactor models follows similarly. As the CDF is rarely known in closed form for stochastic volatility models, we rely on simulation technique. With the simulation scheme, estimators, simulated distributed and bootstrap procedures to be presented in the next sections (see Section 2.3.3 and 2.3.4), the test statistics for **Hypothesis 1** turns out to be:

$$SV_{T,S,h} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(1\{X_t \leq u\} - \frac{1}{S} \sum_{t=1}^{S} 1(X_{t,h}^{\widehat{\theta}_{T,N,L,h}} \leq u) \right)$$

In the above expression, recall that S is the number of simulation paths used in distribution simulation, $\widehat{\theta}_{T,N,L,h}$ is a simulated estimator (see Section 2.3.3). N is a generic notation throughout this chapter, i.e. $N = L$, the length of each simulation path for SGMM and $N = M$, the number of random draws (simulated paths) for NPQML estimator. h is the discretization interval used in simulation. Note that $\widehat{\theta}_{T,N,L,h}$ is chosen in CS (2005) to be SGMM estimator using full sample and therefore

$N = L = S$.[19] To put it simply, one can write $\widehat{\theta}_{T,S,h} = \widehat{\theta}_{T,N,L,h}$.

Under the null, choose T, S to satisfy $T, S \to \infty, Sh \to 0, T/S \to 0$ then:

$$SV_{T,S,h}^2 \to \int_U SV^2(u)\pi(u)$$

where Z is a Gaussian process with covariance kernel that reflects both PEE and the time dependent nature of the data. The relevant bootstrap statistic is:

$$SV_{T,S,h}^{2*} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\begin{array}{c} (1\{X_t^* \leq u\} - 1\{X_t \leq u\}) \\ -\frac{1}{S}(\sum_{t=1}^{S} 1(X_{t,h}^{\widehat{\theta}^*_{T,N,L,h}} \leq u) - 1(X_{t,h}^{\widehat{\theta}_{T,N,L,h}} \leq u)) \end{array} \right)$$

where $\widehat{\theta}^*_{T,S,h}$ is the bootstrap analogue of $\widehat{\theta}_{T,S,h}$. Repeat the Step 1-5 in the bootstrap procedure in Section 2.3.4 to obtain critical value which are the percentiles of the empirical distribution of Z_T^*. Compare $SV_{T,S,h}$ with the percentiles of the empirical distribution of the bootstrap statistic and reject H_0 if $SV_{T,S,h}$ is greater than the $(1-\alpha)th$-percentile thereof. Otherwise, do not reject H_0.

2.3.2.2 Conditional Distribution Tests

To test **Hypothesis 2** for the multifactor models, first we present the test statistic for the case of the stochastic volatility model (X_t, V_t) in (2.12), (i.e., for two factor diffusion), and then we discuss testing with the three factor model (X_t, V_t^1, V_t^2) as in (2.6). Other multiple factor models can be tested analogously. Note that for illustration, we again assume use of the SGMM estimator $\widehat{\theta}_{T,N,L,h}$, as in the original work of BCS (2008) (namely, $\widehat{\theta}_{T,N,L,h}$ is the simulated estimator described in Section 2.3.3). Specifically, N is chosen as the length of sample path L used in parameter

[19] As seen in assumption A4' in CS (2011) and Section 2.3.3 of this chapter, $\widehat{\theta}_{T,N,L,h}$ can be other estimators such as the NPSQML estimator. Importantly, $\widehat{\theta}_{T,N,L,h}$ satisfies condition A4' in CS (2011).

estimation. The associated test statistic is:

$$SZ_T = \sup_{u \times v \in U \times V} |SZ_T(u,v)|$$

$$SZ_T(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{i=1}^{S} 1\left\{ X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,,L,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\left\{X_t \leq v\right\}$$

where $X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,,L,h}}$ is is τ - Step ahead simulated skeleton obtained by simulation procedure for multi-factor model in Subsection 2.3.4.1.

In a similar manner, the bootstrap statistic analogous to SZ_T is

$$SZ_T^* = \sup_{u \times v \in U \times V} |SZ_T^*(u,v)|,$$

$$\begin{aligned}
& SZ_T^*(u,v) \\
= & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{i=1}^{S} 1\left\{ X_{j,i,t+\tau}^{\widehat{\theta}^*_{T,N,L,h}} \leq u \right\} - 1\{X^*_{t+\tau} \leq u\} \right) 1\left\{X^*_t \leq v\right\} \\
& - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{i=1}^{S} 1\left\{ X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,L,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\left\{X_t \leq v\right\}.
\end{aligned}$$

where $\widehat{\theta}^*_{T,N,L,h}$ is the bootstrap estimator described in Section 2.3.4. For the three factor model, the test statistic is defined as

$$MZ_T = \sup_{u \times v \in U \times V} |MZ_T(u,v)|,$$

$$\begin{aligned}
& MZ_T(u,v) \\
= & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{L^2 S} \sum_{j=1}^{L} \sum_{k=1}^{L} \sum_{i=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\left\{X_t \leq v\right\}
\end{aligned}$$

and bootstrap statistics is:

$$
\begin{aligned}
& MZ_T^*(u,v) \\
= \; & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{L^2 S} \sum_{j=1}^{L} \sum_{k=1}^{L} \sum_{i=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}} \leq u \right\} - 1\{X^*_{t+\tau} \leq u\} \right) 1\left\{X_t^* \leq v\right\} \\
& - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{L^2 S} \sum_{j=1}^{L} \sum_{k=1}^{L} \sum_{i=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\left\{X_t \leq v\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}} &= X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}}(X_t, V_j^{1,\widehat{\theta}_{T,N,L,h}}, V_k^{2,\widehat{\theta}_{T,N,L,h}}) \\
X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}} &= X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}}(X_t, V_j^{1,\widehat{\theta}^*_{t,N,L,h}}, V_k^{2,\widehat{\theta}^*_{t,N,L,h}}).
\end{aligned}
$$

The first order asymptotic validity of inference carried out using bootstrap statistics formed as outlined above follows immediately from BCS (2008). For testing decisions, one compares the test statistics $SZ_{T,S,h}$ and $MZ_{T,S,h}$ with the percentiles or the empirical distributions of SZ_T^* and $MZ_{T,S,h}^*$, respectively. Then, reject H_0 if the actual statistic is greater than the $(1-\alpha)th$-percentile of the empirical distribution of the bootstrap statistic, as in Section 2.3.4. Otherwise, do not reject H_0.

2.3.2.3 Predictive Density Tests for Multiple Competing Models

For illustration, we present the test for the stochastic volatility model (two factor model). Again, note that extension to other multi-factor models follows immediately. In particular, all steps in the construction of the test in the one factor model case carry through immediately to the stochastic volatility case with the statistic defined as:

$$DV_{P,L,S} = \max_{k=2,\ldots,m} DV_{k,P,L,S}(u_1, u_2)$$

where

$$
\begin{aligned}
& DV_{k,P,L,S}(u_1,u_2) \\
= \; & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\left(\begin{array}{c} \frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,L,h}}(X_t, V_{1,j}^{\widehat{\theta}_{1,t,N,L,h}}) \leq u_2 \right\} \\ -1\{u_1 \leq X_{t+\tau} \leq u_2\} \end{array} \right)^2 \right.
\end{aligned}
$$

$$
\left. - \left(\frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,L,h}}(X_t, V_{k,j}^{\widehat{\theta}_{k,t,N,L,h}}) \leq u_2 \right\} -1\{u_1 \leq X_{t+\tau} \leq u_2\} \right)^2 \right),
$$

Critical values for these tests can be obtained using a recursive version of the block bootstrap. The corresponding bootstrap test statistic is:

$$
DV_{P,L,S}^{*} = \max_{k=2,\ldots,m} DV_{k,P,L,S}^{*}(u_1,u_2)
$$

where

$$
\begin{aligned}
& DV_{k,P,L,S}^{*}(u_1,u_2) \\
= \; & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left\{ \left(\left(\left[\begin{array}{c} \frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{1,t+\tau,i,j}^{\widehat{\theta}^{*}_{1,t,N,L,h}}(X_t^{*}, V_{1,j}^{\widehat{\theta}^{*}_{1,t,N,L,h}}) \leq u_2 \right\} \\ -1\{u_1 \leq X_{t+\tau}^{*} \leq u_2\} \end{array} \right]^2 \right. \right. \right. \\
& \left. \left. - \left(\frac{1}{T}\sum_{l=1}^{T} \left[\begin{array}{c} \frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,L,h}}(X_l, V_{1,j}^{\widehat{\theta}_{1,t,N,L,h}}) \leq u_2 \right\} \\ -1\{u_1 \leq X_{l+\tau} \leq u_2\} \end{array} \right] \right)^2 \right) \right) \\
& - \left(\left[\frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}^{*}_{k,t,N,L,h}}(X_t^{*}, V_{k,j}^{\widehat{\theta}^{*}_{k,t,N,L,h}}) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau}^{*} \leq u_2\} \right]^2 \right. \\
& \left. \left. - \left(\frac{1}{T}\sum_{l=1}^{T} \left[\begin{array}{c} \frac{1}{SL}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{ u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,L,h}}(X_l, V_{k,j}^{\widehat{\theta}_{k,t,N,L,h}}) \leq u_2 \right\} \\ -1\{u_1 \leq X_{l+\tau} \leq u_2\} \end{array} \right] \right)^2 \right) \right\}.
\end{aligned}
$$

Of note is that we follow CS (2011) by adopting the recursive NPSQML estimator $\widehat{\theta}_{1,t,N,L,h}$ and $\widehat{\theta}_{k,t,N,L,h}$ for model 1 and k, respectively, as introduced in Section 2.3.3.4

with the choice $N = M = S$. $\widehat{\theta}^*_{1,t,N,L,h}$ and $\widehat{\theta}^*_{k,t,N,L,h}$ are bootstrap analogs of $\widehat{\theta}_{1,t,N,L,h}$ and $\widehat{\theta}_{k,t,N,L,h}$ respectively (see Section 2.3.4). In addition, we do not need to resample the volatility process, although volatility is simulated under both $\widehat{\theta}_{k,t,N,L,h}$ and $\widehat{\theta}^*_{k,t,N,L,h}$, $k = 1, ..., m$.

Repeat Steps 1-5 in the bootstrap procedure in Section 2.3.4 to obtain critical values. Compare $DV_{P,L,S}$ with the percentiles of the empirical distribution of $DV^*_{P,L,S}$, and reject H_0 if $DV_{P,L,S}$ is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject H_0. Again, in implementation, there is no need to re-estimate $\widehat{\theta}^*_{k,t,N,L,h}$ for each bootstrap replications if $P/R \to 0, P, T, R \to \infty$, ass parameter estimation error vanishes asymptotically in this case.

2.3.3 Model Simulation and Estimation

2.3.3.1 Simulating Data

Approximation schemes are used to obtain simulated distributions and simulated parameter estimators, which are needed in order to construct the tests statistics outlined in previous sections. We therefore devote the first part of this section to a discussion of the Milstein approximation schemes that have been used in CS (2005), BCS (2008) and CS (2011). Let L be the length of each simulation path and h be the discretization interval, $L = Qh$ and θ be a generic parameter in simulation expression. We consider three cases:

The pure diffusion process as in (2.11)*:*

$$\begin{aligned} X^{\theta}_{qh} - X^{\theta}_{(q-1)h} &= b(X^{\theta}_{(q-1)h}, \theta)h + \sigma(X^{\theta}_{(q-1)h}, \theta)\epsilon_{qh} \\ &\quad -\frac{1}{2}\sigma(X^{\theta}_{(q-1)h}, \theta)'\sigma(X^{\theta}_{(q-1)h}, \theta)h \\ &\quad +\frac{1}{2}\sigma(X^{\theta}_{(q-1)h}, \theta)'\sigma(X^{\theta}_{(q-1)h}, \theta)\epsilon^2_{qh}, \end{aligned}$$

where

$$\left(W_{qh} - W_{(q-1)h}\right) = \epsilon_{qh} \overset{iid}{\sim} N(0,h),$$

$q = 1, \ldots, Q$, with $\epsilon_{qh} \overset{iid}{\sim} N(0,h)$; and where σ' is the derivative of $\sigma(\cdot)$ with respect to its first argument. Hereafter, X^{θ}_{qh} denotes the values of the diffusion at time qh, simulated under generic θ, and with a discrete interval equal to h, and so is a fine grain analog of $X^{\theta}_{t,h}$.

The pure jump diffusion process without stochastic volatility as in (2.9)*:*

$$\begin{aligned} & X^{\theta}_{(q+1)h} - X^{\theta}_{qh} \\ = \; & b(X^{\theta}_{qh},\theta)h + \sigma(X^{\theta}_{qh},\theta)\epsilon_{(q+1)h} - \frac{1}{2}\sigma(X^{\theta}_{qh},\theta)'\sigma(X^{\theta}_{qh},\theta)h \end{aligned}$$

$$+\frac{1}{2}\sigma(X^{\theta}_{qh},\theta)'\sigma(X^{\theta}_{qh},\theta)\epsilon^{2}_{(q+1)h} - \lambda\mu_y h + \sum_{j=1}^{\mathcal{J}} y_j 1\left\{qh \leq \mathcal{U}_j \leq (q+1)h\right\}, \qquad (2.14)$$

The only difference between this approximation and that used for the pure diffusion is the jump part. Note that the last term on the right-hand-side (RHS) of (2.14) is nonzero whenever we have one (or more) jump realization(s) in the interval $[(q-1)h, qh]$. Moreover, as neither the intensity nor the jump size is state dependent, the jump component can be simulated without any discretization error, as follows. Begin by making a draw from a Poisson distribution with intensity parameter $\widehat{\lambda}\tau$, say $\mathcal{J}$. This gives a realization for the number of jumps over the simulation time span. Then, draw $\mathcal{J}$ uniform random variables over $[0, L]$, and sort them in ascending order so that $\mathcal{U}_1 \leq \mathcal{U}_2 \leq ... \leq \mathcal{U}_{\mathcal{J}}$. These provide realizations for the $\mathcal{J}$ jump times. Then, make $\mathcal{J}$ independent draws from ϕ, say $y_1, ..., y_{\mathcal{J}}$.

SV models without jumps as in (2.4) *(*using a generalized Milstein scheme):

$$\begin{aligned} X^{\theta}_{(q+1)h} &= X^{\theta}_{qh} + \widetilde{b}_1(X^{\theta}_{qh},\theta)h + \sigma_{11}(V^{\theta}_{qh},\theta)\epsilon_{1,(q+1)h} \\ &+\sigma_{12}(V^{\theta}_{qh},\theta)\epsilon_{2,(q+1)h} + \frac{1}{2}\sigma_{22}(V^{\theta}_{qh},\theta)\frac{\partial\sigma_{12,k}(V^{\theta}_{qh},\theta)}{\partial V}\epsilon^2_{2,(q+1)h} \\ &+\sigma_{22}(V^{\theta}_{qh},\theta)\frac{\partial\sigma_{11}(V^{\theta}_{qh},\theta)}{\partial V}\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right)dW_{2,s} \end{aligned} \tag{2.15}$$

$$\begin{aligned} V^{\theta}_{(q+1)h} &= V^{\theta}_{qh} + \widetilde{b}_2(V^{\theta}_{qh},\theta)h + \sigma_{22}(V^{\theta}_{qh},\theta)\epsilon_{2,(q+1)h} \\ &+\frac{1}{2}\sigma_{22}(V^{\theta}_{qh},\theta)\frac{\partial\sigma_{22}(V^{\theta}_{qh},\theta)}{\partial V}\epsilon^2_{2,(q+1)h} \end{aligned} \tag{2.16}$$

where $h^{-1/2}\epsilon_{i,qh} \sim N(0,1)$, $i = 1,2$, $E(\epsilon_{1,qh}\epsilon_{2,q'h}) = 0$ for all $q \neq q'$, and

$$\widetilde{b}(V,\theta) = \begin{pmatrix} \widetilde{b}_1(V,\theta) \\ \widetilde{b}_2(V,\theta) \end{pmatrix} = \begin{pmatrix} b_1(V,\theta) - \frac{1}{2}\sigma_{22}(V,\theta)\frac{\partial\sigma_{12}(V,\theta)}{\partial V} \\ b_2(V,\theta) - \frac{1}{2}\sigma_{22}(V,\theta)\frac{\partial\sigma_{22}(V,\theta)}{\partial V} \end{pmatrix}.$$

The last terms on the RHS of (2.15) involve stochastic integrals and cannot be explicitly computed. However, they can be approximated, up to an error of order $o(h)$ by (see, for example, equation (3.7), pp. 347 in Kloeden and Platen (1999)):

$$\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right)dW_{2,s} \approx h\left(\frac{1}{2}\xi_1\xi_2 + \sqrt{\rho_p}\left(\mu_{1,p}\xi_2 - \mu_{2,p}\xi_1\right)\right)$$

$$+\frac{h}{2\pi}\sum_{r=1}^{p}\frac{1}{r}\left(\varsigma_{1,r}\left(\sqrt{2}\xi_2 + \eta_{2,r}\right) - \varsigma_{2,r}\left(\sqrt{2}\xi_1 + \eta_{1,r}\right)\right),$$

where for $j = 1,2$, $\xi_j, \mu_{j,p}, \varsigma_{j,r}, \eta_{j,r}$ are *iid* $N(0,1)$ random variables, $\rho_p = \frac{1}{12} - \frac{1}{2\pi^2}\sum_{r=1}^{p}\frac{1}{r^2}$, and p is such that as $h \to 0$, $p \to \infty$.

Stochastic Volatility with Jumps

Simulation of sample paths of diffusion processes with stochastic volatility and jumps follows straightforwardly from the previous two cases. Whenever both intensity and jump size are not state dependent, a jump component can be simulated and added to either $X(t)$ and/or the $V(t)$ in the same manner as above. Extension to general multidimensional and multifactor models both with and without jumps also follows directly.

2.3.3.2 Simulating Distributions

In this section we sketch out methods used to construct $\tau-$step ahead simulated conditional distributions using simulated data. In applications, simulation techniques are needed when the functional form conditional distribution is unknown. We first illustrate the technique for one factor models and then discuss multifactor models.

One factor models:

Consider the one factor model as in (2.9). To estimate the simulated CDFs,

Step 1: Obtain $\widehat{\theta}_{T,N,h}$ (using the entire sample) or $\widehat{\theta}_{t,N,h}$ (recursive estimator) where $\widehat{\theta}_{T,N,h}$ and $\widehat{\theta}_{t,N,h}$ are estimators as discussed in Section 2.3.3.3 and 2.3.3.4.

Step 2: Under $\widehat{\theta}_{T,N,h}$ or $\widehat{\theta}_{t,N,h}$[20], simulate S paths of length τ, all having the same starting value, X_t. In particular, for each path $i = 1, ...S$ of length τ, generate $X_{i,t+\tau}^{\widehat{\theta}_{T,N,h}}(X_t)$ according to a Milstein schemes detailed in previous section, with $\theta = \widehat{\theta}_{T,N,h}$ or $\widehat{\theta}_{t,N,h}$. The errors used in simulation are $\epsilon_{qh} \overset{iid}{\sim} N(0,h)$, and $Qh = \tau$. ϵ_{qh} is assumed to be independent across simulations, so that $E(\epsilon_{i,qh}\epsilon_{j,qh}) = 0$,for all $i \neq j$ and $E(\epsilon_{i,qh}\epsilon_{i,qh}) = h$, for any i, j. In addition, as the simulated diffusion is ergodic, the effect of the starting value approaches zero at an exponential rate, as $\tau \to \infty$.

Step 3: If $\widehat{\theta}_{T,N,h}$ ($\widehat{\theta}_{t,N,h}$) is used, an estimate for the distribution, at time $t+\tau$, conditional on X_t, with estimator $\widehat{\theta}_{T,N,h}(\widehat{\theta}_{t,N,h})$, is defined as:

$$= \frac{1}{S}\sum_{i=1}^{S} 1\left\{X_{i,t+\tau}^{\widehat{\theta}_{T,N,h}}(X_t) \leq u\right\}$$

[20] Note that $N = L$ for the SGMM estimator while $N = M = S$ for NSQML estimator.

BCS (2008) show that if the model is correctly specified, then $\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h})$ provides a consistent estimate of the conditional distribution $F_\tau(u|X_t, \theta^\dagger)$.

Specifically, assume that $T, N, S \to \infty$. Then, for the case of SGMM estimator, if $h \to 0$, $T/N \to 0$, and $h^2T \to 0$, $T^2/S \to \infty$, the following result holds for any X_t, $t \geq 1$, uniformly in u

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h}) - F_\tau(u|X_t, \theta^\dagger) \xrightarrow{pr} 0,$$

In addition, if the model is correctly specified (i.e. if $\mu(\cdot,\cdot) = \mu_0(\cdot,\cdot)$ and $\sigma(\cdot,\cdot) = \sigma_0(\cdot,\cdot)$) then:

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h}) - F_{0,\tau}(u|X_t, \theta_0) \xrightarrow{pr} 0,$$

Step 4: Repeat Steps 1-3 for $t = 1, ..., T-\tau$. This yields $T-\tau$ conditional distributions that are $\tau-$Steps ahead which will be used in the construction of the specification tests.

The CDF simulation in the case selection test of multiple models with recursive estimator is similar. For model k, let $\widehat{\theta}_{k,t,N,h}$ be the recursive estimator of "pseudo true" $\theta_k^\dagger$ computed using all observations up to varying time t. Then, $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$ is generated according to a Milstein schemes as in Section 2.3.3.1, with $\theta = \widehat{\theta}_{k,t,N,h}$ and the initial value X_t, $Qh = \tau$. The corresponding empirical distribution of the simulated series $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$ can then be constructed. Under some regularity conditions,

$$\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t) \leq u_2\right\} \xrightarrow{pr} F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_1),\ t = R, ..., T-\tau,$$

where $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u)$ is the marginal distribution of $X_{t+\tau}^{\theta_k^\dagger}(X_t)$ implied by k model (i.e., by the model used to simulate the series), conditional on the (simulation) starting value X_t. Furthermore, the marginal distribution of $X_{t+\tau}^{\theta^\dagger}(X_t)$ is the distribution of $X_{t+\tau}$ conditional on the values observed at time t. Thus, $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u) = F_k^\tau(u|X_t, \theta_k^\dagger)$.

Of important note is that in the simulation of $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$, $i=1,...,S$, for each t, $t=R,...,T-\tau$, we must use the same set of randomly drawn errors and similarly the same draws for numbers of jumps, jump times and jump sizes. Thus, we only allow for the starting value to change. In particular, for each $i=1,...,S$, we generate $X_{k,i,R+\tau}^{\widehat{\theta}_{k,R,N,h}}(X_R),...,X_{k,i,T}^{\widehat{\theta}_{k,T-\tau,N,h}}(X_{T-\tau})$. This yields an SxP matrix of simulated values, where $P=T-R-\tau+1$ refers to the length of the out-of-sample period. $X_{k,i,R+j+\tau}^{\widehat{\theta}_{k,R+j,N,h}}(X_{R+j})$ (at time $R+j+\tau$) can be seen as τ periods ahead value "predicted" by model k using all available information up to time $R+j_{R+j}$, $j=1,...,P$ (the initial value X_{R+j} and $\widehat{\theta}_{k,R+j,N,h}$ estimated using $X_1,...,X_{R+j}$). The key feature of this setup is that it enables us to compare "predicted " τ periods ahead values (i.e. $X_{k,i,R+j+\tau}^{\widehat{\theta}_{k,R+j,N,h}}(X_{R+j})$) with actual values that are τ periods ahead (i.e., $X_{R+j+\tau}$), for $j=1,...,P$. In this manner, simulation based tests under *ex-ante* predictive density comparison framework can be constructed.

Multifactor model:

Consider the multi-factor model with a skeleton $\left(X_t, V_t^1,...,V_t^d\right)'$ (e.g. stochastic mean, stochastic volatility models, stochastic volatility of volatility, etc.) where only the first element X_t is observed. For simulation of the CDF, the difficulty arises as we do not know the initial values of latent variables $(V_t^1,...,V_t^d)'$ at each point in time. We generalize the simulation plan of BCS (2008) and Cai and Swanson (2011) to the case of d dimensions. Specifically, to overcome the initial value difficulty, a natural strategy is to simulate a long path of length L for each latent variable $V_t^1,...,V_t^d$, use them to construct $X_{t+\tau}$ and the corresponding simulated CDF of $X_{t+\tau}$; and finally, we average out the volatility values. Note that there are L^d combinations of the initial values $V_t^1,...,V_t^d$. For illustration, consider the case of stochastic volatility $(d=1)$ and the Chen three factor model as in (2.6) $(d=2)$, using recursive estimators.

For the case of stochastic volatility $(d=1)$, i.e. $(X_t,V_t)'$, the steps are as follows:

Step 1: Estimate $\widehat{\theta}_{t,N,L,h}$ using recursive SGMM or NSQML estimation methods.

Step 2. Using the scheme in (2.16) with $\theta = \widehat{\theta}_{t,N,L,h}$, generate the path $V_{qh}^{\widehat{\theta}_{t,N,L,h}}$ for $q = 1/h, ..., Qh$ with $Qh = L$ and hence obtain $V_{j}^{\widehat{\theta}_{t,N,L,h}}$ $j = 1, ...L$.

Step 3: Using schemes in (2.15), (2.16), simulate LxS paths of length τ, setting the initial value for the observable state variable to be X_t. For the initial values of unobserved volatility, use $V_{j,qh}^{\widehat{\theta}_{t,N,L,h}}$, $j = 1, ...L$ as retrieved in Step 2. Also, keep the simulated random innovations (i.e.$\epsilon_{1,qh}$,.$\epsilon_{1,qh}$, $\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right) dW_{2,s}$) to be constant across each j and t. Hence, for each replication i,using initial values X_t and $V_{j,qh}^{\widehat{\theta}_{T,N,h}}$, we obtain $X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t)$ which is a τ - step ahead simulated value.

Step 4: Now the estimator of $F_\tau(u|X_t, \theta^\dagger)$ is defined as:

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{t,N,L,h}) = \frac{1}{LS}\sum_{j=1}^{L}\sum_{i=1}^{S} 1\left\{X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,h}}(X_t) \leq u\right\}$$

Note that, by averaging over the initial value of the volatility process, we have integrated out it's effect. In other words, $\frac{1}{S}\sum_{i=1}^{S} 1\left\{X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,h}}(X_t) \leq u\right\}$ is an estimate of $F_\tau(u|X_t, V_{j,h}^{\widehat{\theta}_{t,N,h}}, \theta^\dagger)$.

Step 5: Repeat the Steps 1-4 for $t = 1, ..., T - \tau$. This yields $T - \tau$ conditional distributions that are $\tau-$steps ahead which will be used in the construction of the specification tests.

For three factor model ($d = 2$), i.e., (X_t, V_t^1, V_t^2), consider model (2.6), where $W_t = (W_t^1, W_t^2, W_t^3)$ are mutually independent standard Brownian motions.

Step 1: Estimate $\widehat{\theta}_{t,N,L,h}$ using SGMM or NSQML estimation methods.

Step 2*:* Given the estimated parameter $\widehat{\theta}_{t,N,L,h}$, generate the path $V_{qh}^{1,\widehat{\theta}_{t,N,L,h}}$ and $V_{ph}^{2,\widehat{\theta}_{t,N,L,h}}$ for $q, p = 1/h, ..., Qh$ with $Qh = L$ and hence obtain $V_{j}^{1,\widehat{\theta}_{T,N,L,h}}$,$V_{k}^{2,\widehat{\theta}_{T,N,L,h}}$ $j, k = 1, ..., L$.

Step 3: Given the observable X_t and the $L \times L$ simulated latent paths ($V_{j}^{1,\widehat{\theta}_{t,N,L,h}}$ and $V_{k}^{2,\widehat{\theta}_{t,N,L,h}}$ $j, k = 1, ..., L$) as the start values , we simulate τ-Step ahead

$X_{t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t, V_{j}^{1,\widehat{\theta}_{t,N,L,h}}, V_{k}^{2,\widehat{\theta}_{t,N,L,h}})$. Since the start values for the two latent variables are $L \times L$ length, so for each X_t we have N^2 path. Now to integrate out the initial effect of latent variables, form the estimate of conditional distribution as

$$\widehat{F}_{\tau,s}(u|X_t,\widehat{\theta}) = \frac{1}{L^2}\sum_{j=1}^{L}\sum_{k=1}^{L} 1\left\{X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t, V_j^{1,\widehat{\theta}_{t,N,L,h}}, V_k^{2,\widehat{\theta}_{t,N,L,h}}) \leq u\right\},$$

where s denotes the sth simulation.

Step 4*:* Simulate $X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}}$ S times, that is, repeat Step 3 S times i.e. $s = 1, ..., S$. The estimate of $F_\tau(u|X_t, \theta^\dagger)$ is

$$\widehat{F}_\tau(u|X_t,\widehat{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\widehat{F}_{\tau,s}(u|X_t,\widehat{\theta}_{T,N,h})$$

Step 5: Repeat the Steps 1-4 for $t = 1, ..., T-\tau$. This yields $T-\tau$ conditional distributions that are $\tau-$steps ahead which will be used in the construction of the specification tests.

As a final remark, for the case of multiple competing models, we can proceed similarly. In addition, in the next two subsections, we present the exactly identified simulated (recursive) general method of moments and recursive nonparametric simulated quasi-maximum likelihood estimators that can be used in simulating distributions as well as constructing test statistics described in Section 2.3.2. The bootstrap analogs of those estimators will be discussed in Section 2.3.4.

2.3.3.3 Estimation: (Recursive) Simulated General Method of Moments (SGMM) Estimators

Suppose that we observe a discrete sample of T observations, say $(X_1, X_2, ..., X_T)'$, from the underlying diffusion in (2.9). The (recursive) SGMM estimator $\widehat{\theta}_{t,L,h}$ with

$1 \leq t \leq T$ is specified as:

$$
\begin{aligned}
& \widehat{\theta}_{t,L,h} \\
= \; & \arg\min_{\theta \in \Theta} \left(\frac{1}{t} \sum_{j=1}^{t} g(X_j) - \frac{1}{L} \sum_{j=1}^{L} g(X_{j,h}^{\theta}) \right)' W_t^{-1} \left(\frac{1}{t} \sum_{j=1}^{t} g(X_j) - \frac{1}{L} \sum_{j=1}^{L} g(X_{j,h}^{\theta}) \right) \quad (2.17) \\
= \; & \arg\min_{\theta \in \Theta} G_{t,L,h}(\theta)' W_t G_{t,L,h}(\theta), \quad (2.18)
\end{aligned}
$$

where g is a vector of p moment conditions, $\Theta \subset \Re^p$ (so that we have as many moment conditions as parameters), and $X_{j,h}^{\theta} = X_{[Qjh/L]}^{\theta}$, with $L = Qh$ is the simulated path under generic parameter θ and with discrete interval h. $X_{j,h}^{\theta}$ is simulated using the Milstein schemes.

Note that in the above expression, in the context of the specification test $\widehat{\theta}_{t,L,h}$ is estimated using the whole sample, i.e. $t = T$. In the out of sample context, the recursive SGMM estimator $\widehat{\theta}_{t,L,h}$ is estimated recursively using the using sample from 1 up to t.

Typically, the p moment conditions are based on the difference between sample moments of historical and simulated data or, between sample moments and model implied moments, whenever the latter are known in closed form. Finally, W_t is the heteroskedasticity and autocorrelation (HAC) robust covariance matrix estimator, defined as

$$
W_t^{-1} = \frac{1}{t} \sum_{\nu=-l_t}^{l_t} w_{\nu} \sum_{j=\nu+1+l_t}^{t-l_t} \left(g(X_j) - \frac{1}{t} \sum_{j=1}^{t} g(X_j) \right) \left(g(X_{j-\nu}) - \frac{1}{t} \sum_{j=1}^{t} g(X_j) \right)', \quad (2.19)
$$

where $w_v = 1 - v/(l_T + 1)$. Further, the pseudo true value, $\theta^{\dagger}$, is defined to be:

$$
\theta^{\dagger} = \arg\min_{\theta \in \Theta} G_{\infty}(\theta)' W_0 G_{\infty}(\theta),
$$

where

$$G_\infty(\theta)' W_0 G_\infty(\theta) = p \lim_{L,T\to\infty,h\to 0} G_{T,L,h}(\theta)' W_T G_{T,L,h}(\theta);$$

and where $\theta^\dagger = \theta_0$, if the model is correctly specified.

In the above set up, the exactly identified case is considered rather than the overidentified (S)GMM. This choice guarantees that $G_\infty(\theta^\dagger) = 0$ even under misspecification, in the sense that the model differs from the underlying DGP. As pointed out by Hall and Inoue (2003), the root-N consistency does not hold for overidentified (S)GMM estimators of misspecified models. In addition,

$$\nabla_\theta G_\infty(\theta^\dagger)' W^\dagger G_\infty(\theta^\dagger) = 0.$$

However, in the case for which the number of parameters and the number of moment conditions is the same, $\nabla_\theta G_\infty(\theta^\dagger)' W^\dagger$ is invertible, and so the first order conditions also imply that $G_\infty(\theta^\dagger) = 0$.

Also note that other available estimation methods using moments include the efficient method of moments (EMM) estimator as proposed by Gallant and Tauchen (1996, 1997), which calculates moment functions by simulating the expected value of the score implied by an auxiliary model. In their setup, parameters are then computed by minimizing a chi-square criterion function.

2.3.3.4 Estimation: Recursive Nonparametric Simulated Quasi Maximum Likelihood Estimators

In this section we outline a recursive version of the NPSQML estimator of Fermanian and Salani´e (2004), proposed by CS (2011). The bootstrap counterpart of the recursive NPSQML estimator will be presented in the next section.

One factor models:

Hereafter, let $f\left(X_t | X_{t-1}, \theta^\dagger\right)$ be the conditional density associated with the above jump diffusion. If f is known in closed form, we can just estimate $\theta^{\dagger}$ recursively, using standard QML as:[21]

$$\widehat{\theta}_t = \arg\max_{\theta\in\Theta} \frac{1}{t}\sum_{j=2}^{t} \ln f\left(X_j|X_{j-1},\theta\right),\ t = R, ..., R+P-1.. \tag{2.20}$$

Note that, similarly to the case of SGMM, the pseudo true value $\theta^{\dagger}$ is optimal in the sense:

$$\theta^{\dagger} = \arg\max_{\theta\in\Theta} E\left(\ln f\left(X_t|X_{t-1},\theta\right)\right). \tag{2.21}$$

for the case f is not known in closed form, we can follow Kristensen and Shin (2008) and CS (2011) to construct the simulated analog $\widehat{f}$ of f and then use it to estimate $\theta^{\dagger}$. $\widehat{f}$ is estimated as function of the simulated sample paths $X^{\theta}_{t,i}(X_{t-1})$, for $t = 2, ..., T-1$, $i = 1, ..., M$. First, generate $T-1$ paths of length one for each simulation replication, using X_{t-1} with $t = 1, ...T$ as starting values. Hence, at time t and simulation replication i we obtain skeletons $X^{\theta}_{t,i}(X_{t-1})$, for $t = 2, ..., T-1$, $i = 1, ..., M$ where M is the number of simulation paths (number of random draws or $X^{\theta}_{t,j}(X_{t-1})$ and $X^{\theta}_{t,l}(X_{t-1})$ are i.i.d.) for each simulation replication. M is fixed across all initial values. Then the recursive NPSQML estimator is defined as follows:

$$\widehat{\theta}_{t,M,h} = \arg\max_{\theta\in\Theta} \frac{1}{t}\sum_{i=2}^{t} \ln \widehat{f}_{M,h}\left(X_i|X_{i-1},\theta\right)\tau_M\left(\widehat{f}_{M,h}\left(X_i|X_{i-1},\theta\right)\right),\ t \geq R,$$

where

$$\widehat{f}_{M,h}\left(X_t|X_{t-1},\theta\right) = \frac{1}{M\xi_M}\sum_{i=1}^{M} K\left(\frac{X^{\theta}_{t,i,h}(X_{t-1}) - X_t}{\xi_M}\right).$$

Note that with abuse of notation, we define $\widehat{\theta}_{t,L,h}$ for SGMM and $\widehat{\theta}_{t,M,h}$ for NPSQML estimators where L and M have different interpretations (L is the length

[21] Note that as model k is, in general, misspecified, $\sum_{t=1}^{T-1} f_k\left(X_t|X_{t-1},\theta_k\right)$ is a quasi-likelihood and $f_k\left(X_t|X_{t-1},\theta^{\dagger}_k\right)$ is not necessarily a martingale difference sequence.

of each simulation path and M is number of random draws).

The function $\tau_M\left(\widehat{f}_{M,h}\left(X_t|X_{t-1},\theta\right)\right)$ is a trimming function. It has some characteristics such as positive and increasing,

$$\tau_M\left(\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)\right)=0, \text{if } \widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)<\xi_M^\delta,$$

and

$$\tau_M\left(\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)\right)=1, \text{if } \widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)>2\xi_M^\delta,$$

for some $\delta>0$.[22] Note that when the log density is close to zero, the derivative tends to infinity and thus even very tiny simulation errors can have a large impact on the likelihood. The introduction of the trimming parameter into the optimization function ensures the impact of this case to be minimal asymptotically.

Multifactor Models:

Since volatility is not observable, we cannot proceed as in the single factor case when estimating the SV model using NPSQML estimator. Instead, let V_j^θ be generated according to (2.16), setting $qh=j$, and $j=1,...,L$. The idea is to simulate L different starting values for unobservable volatility, construct the simulated likelihood functions accordingly and then average them out. For each simulation replication at time t, we simulate L different values of $X_t\ (X_{t-1},V_j^\theta)$ by generating L paths of length one, using fixed observable X_{t-1} and unobservable V_j^θ, $j=1,...,L$ as starting values. Repeat this procedure for any $t=1,...,T-1$, and for any set j, $j=1,...,L$ of random errors $\epsilon_{1,t+(q+1)h,j}$ and $\epsilon_{2,t+(q+1)h,j}$, $q=1,...,1/h$. Note that it is important to use the same set of random errors $\epsilon_{1,t+(q+1)h,j}$ and $\epsilon_{2,t+(q+1)h,j}$ across different initial values for volatility. Denote the simulated value at time t, simulation replication i, under

[22] Fermanian and Salanie (2004) suggest using the following triming function:

$$\tau_N(x)=\frac{4(x-a_N)^3}{a_N^3}-\frac{3(x-a_N)4}{a_N^4},$$

for $a_N\leq x\leq 2a_N$.

generic parameter θ, using X_{t-1}, V_j^θ as starting values as $X_{t,i,h}^\theta(X_{t-1}, V_j^\theta)$. Then:

$$\widehat{f}_{M,L,h}\left(X_t|X_{t-1},\theta\right) = \frac{1}{L}\sum_{j=1}^{L}\frac{1}{M\xi_M}\sum_{i=1}^{M}K\left(\frac{X_{t,i,h}^\theta(X_{t-1},V_j^\theta) - X_t}{\xi_M}\right),$$

and note that by averaging over the initial values for the unobservable volatility, its effect is integrated out. Finally, define:[23]

$$\widehat{\theta}_{t,M,L,h} = \arg\min_{\theta\in\Theta}\frac{1}{t}\sum_{s=2}^{t}\ln\widehat{f}_{M,L,h}\left(X_s|X_{s-1},\theta\right)\tau_M\left(\widehat{f}_{M,L,h}\left(X_s|X_{s-1},\theta\right)\right),\ t \geq R.$$

Note that in this case, X_t is no longer Markov (i.e., X_t and V_t are jointly Markovian, but X_t is not). Therefore, even in the case of true data generating process, the joint likelihood cannot be expressed as the product of the conditional and marginal distributions. Thus, $\widehat{\theta}_{t,M,L,h}$ is necessarily a QML estimator. Furthermore, note that $\nabla_\theta f(X_t|X_{t-1},\theta^\dagger)$ is no longer a martingale difference sequence; therefore, we need to use HAC robust covariance matrix estimators, regardless of whether the model is the "correct" model or not.

2.3.4 Bootstrap Critical Value Procedures

The test statistics presented in Section 2.3.1 and 2.3.2 are implemented using critical values constructed via the bootstrap. As mentioned earlier, motivation for using the bootstrap is clear. The covariance kernel of the statistics limiting distributions contain both parameter estimation error and the data related time dependence components. Asymptotic critical value cannot thus be tabulated in a usual way. Several methods have been proposed to tackle this issue. One is the block bootstrap procedures which we discuss. Others have been mentioned above.

With regarding to the validity of the bootstrap, note that, in the case of dependent

[23]For discussion of asymptotic properties of $\widehat{\theta}_{k,t,M,L,h}$, as well as of regularity conditions, see CS(2011).

observations without PEE, we can tabulate valid critical value using a simple empirical version of the Künsch (1989) block bootstrap. Now, the difficulty in our context lies in accounting for parameter estimation error. Goncalves and White (2002) establish the first order validity of the block bootstrap for QMLE (or m-estimator) for dependent and heterogeneous data. This is an important result for the class of SGMM and NSQML estimators surveyed in this chapter, and allows Corradi and Swanson in CS (2011) and elsewhere to develop asymptotically valid version of the bootstrap that can be applied under generic model misspecification, as assumed throughout this chapter.

For the SGMM estimator, as shown in CS (2005) the first order validity of the block bootstrap is valid in the exact identification case, and when $T/S \to 0$. In this case, SGMM is asymptotically equivalent to GMM, and consequently there is no need to bootstrap the simulated series. In addition, in the exact identification case, GMM estimators can be treated the same way that QMLE estimators are treated. For the NSQML estimator, CS (2011) point out that the NPSQML estimator is asymptotically equivalent to the QML estimator. Thus, we do not need to resample the simulated observations as the negligible contribution of simulation errors.

Also note that critical values for these tests can be obtained using a recursive version of the block bootstrap. When forming block bootstrap samples in the recursive case, observations at the beginning of the sample are used more frequently than observations at the end of the sample. This introduces a location bias to the usual block bootstrap, as under standard resampling with replacement, all blocks from the original sample have the same probability of being selected. Also, the bias term varies across samples and can be either positive or negative, depending on the specific sample. A first-order valid bootstrap procedure for non simulation based $m-$estimators constructed using a recursive estimation scheme is outlined in Corradi and Swanson (2007a). Here we extend the results of Corradi and Swanson (2007a) by establishing asymptotic results for cases in which simulation-based estimators are bootstrapped in a recursive setting.

Now the details of bootstrap procedure for critical value tabulation can be outlined in 5 steps as follows:

Step 1: Let $T = bl$, where b denotes the number of blocks and l denotes the length of each block. We first draw a discrete uniform random variable, I_1, that can take values $0, 1, ..., T - l$ with probability $1/(T - l + 1)$. The first block is given by $X_{I_{1+1}}, ..., X_{I_1+l}$. We then draw another discrete uniform random variable, say I_2, and a second block of length l is formed, say $X_{I_2+1}, ..., X_{I_2+l}$. Continue in the same manner, until you draw the last discrete uniform say I_b, and so the last block is $X_{I_b+1}, ..., X_{I_b+l}$. Let's call the X_t^* the resampled series, and note that $X_1^*, X_2^*, ..., X_T^*$ corresponds to $X_{I_1+1}, X_{I_1+2}, ..., X_{I_b+l}$. Thus, conditional on the sample, the only random element is the beginning of each block. In particular

$$X_1^*, ..., X_l^*, X_{l+1}^*, ..., X_{2l}^*, X_{T-l+1}^*, ..., X_T^*,$$

conditional on the sample, can be treated as b *iid* blocks of discrete uniform random variables. For a simple illustration the link between the bootstrap sample and the original sample. Note that it can be shown that except a set of probability measure approaching zero,

$$E^* \left(\frac{1}{T} \sum_{t=1}^{T} X_t^* \right) = \frac{1}{T} \sum_{t=1}^{T} X_t + O_P^*(l/T) \tag{2.22}$$

$$\begin{aligned} Var^* \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} X_t^* \right) &= \frac{1}{T} \sum_{t=l}^{T-l} \sum_{i=-l}^{l} (X_t - \frac{1}{T} \sum_{t=1}^{T} X_t)(X_{t+i} - \frac{1}{T} \sum_{t=1}^{T} X_t) \\ &\quad + O_{P^*}(l^2/T), \end{aligned} \tag{2.23}$$

where E^* and Var^* denotes the expectation and the variance operators with respect to P^* (the probability law governing the resampled series or the probability law governing the *iid* uniform random variables, conditional on the sample), and where $O_{P^*}(l/T)$ $(O_{P^*}(l^2/T))$ denotes a term converging in probability P^* to zero, as $l/T \to 0$ $(l^2/T \to 0)$.

In the case of recursive estimators, we proceed the bootstrap similarly as follows. Begin by resampling b blocks of length l from the full sample, with $lb = T$. For any given τ, it is necessary to jointly resample $X_t, X_{t+1}, ..., X_{t+\tau}$. More precisely, let $Z^{t,\tau} = (X_t, X_{t+1}, ..., X_{t+\tau})$, $t = 1, ..., T-\tau$. Now, resample b overlapping blocks of length l from $Z^{t,\tau}$. This yields $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T-\tau$.

Step 2: Re-estimate $\widehat{\theta}^*_{t,N,h}(\widehat{\theta}^*_{T,N,L,h})$ using the bootstrap sample,

$$Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*), t = 1, ..., T-\tau$$

(or full sample $X_1^*, X_2^*, ..., X_T^*$). Recall that if we use the entire sample for the estimation, as the specification test in CS(2005) and BCS(2008), then $\widehat{\theta}^*_{t,N,h}$ is denoted as $\widehat{\theta}^*_{T,N,h}$. The bootstrap estimators for SGMM and NPSQML are presented below:

Bootstrap (recursive) SGMM Estimators

If the full sample is used in the specification test as in CS (2005) and BCS(2008), the bootstrap estimator is constructed straightforward as

$$\begin{aligned}
&\widehat{\theta}^*_{T,L,h} \\
= \quad &\arg\min_{\theta\in\Theta} \left(\frac{1}{T}\sum_{j=1}^{T} g(X_j^*) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta})\right)' W_T^{*-1} \left(\frac{1}{T}\sum_{j=1}^{T} g(X_j^*) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta})\right),
\end{aligned}$$

where W_T^{-1} and $g(.)$ are defined in (2.19) and L is the length of each simulation path.

Note that it is convenient not to resample the simulated series as the simulation error vanishes asymptotically. In implementation, we do not have mimic its contribution to the covariate kernel.

In the case of predictive density type model selection where recursive estimators are needed, define the bootstrap analog as

$$
\begin{aligned}
& \widehat{\theta}^*_{t,L,h} \\
= \; & \arg\min_{\theta\in\Theta} \left(\frac{1}{t}\sum_{j=1}^{t}\left(\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right) - \left(\frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta}) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right)\right)' \\
& \Omega_t^{*-1}\left(\frac{1}{t}\sum_{j=1}^{t}\left(\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right) - \left(\frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta}) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right)\right) \\
= \; & \arg\min_{\theta\in\Theta} G^*_{t,L,h}(\theta)'\Omega_t^{*-1}G^*_{t,L,h}(\theta),
\end{aligned}
$$

where

$$
\Omega_t^{*-1} = \frac{1}{t}\sum_{\nu=-l_t}^{l_t} w_{\nu,t}\sum_{j=\nu+1+l_t}^{t-l_t}\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right)\left(g(X_{j-\nu}^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right)
$$

Note that each bootstrap term is recentered around the (full) sample mean. The intuition behind the particular recentering in bootstrap recursive SGMM estimator is that it ensures that the mean of the bootstrap moment conditions, evaluated at $\widehat{\theta}_{t,L,h}$ is zero, up to a negligible term. Specifically, we have

$$
\begin{aligned}
& E^*\left(\frac{1}{t}\sum_{j=1}^{t}\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right) - \left(\frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}^*_{t,L,h}}) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right) \\
= \; & E^*(g(X_j^*)) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'}) = O(l/T), \text{ with } l = o(T^{1/2}),
\end{aligned}
$$

where the $O(l/T)$ term is due to the end block effect (see Corradi and Swanson (2007b) for further discussion).

Bootstrap Recursive NPSQML Estimators

Let $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T-\tau$. For each simulation replication, generate $T-1$ paths of length one, using $X_1^*, ..., X_{T-1}^*$ as starting values, and so obtaining $X^{\theta}_{t,j}(X^*_{t-1})$, for $t = 2, ..., T-1$, $i = 1, ..., M$. Further, let:

$$\widehat{f}^*_{M,h}\left(X^*_t | X^*_{t-1}, \theta\right) = \frac{1}{M\xi_M} \sum_{i=1}^{M} K\left(\frac{X^{\theta}_{t,i,h}(X^*_{t-1}) - X^*_t}{\xi_M}\right),$$

Now, for $t = R, ..., R+P-1$, define:

$$\begin{aligned}
& \widehat{\theta}^*_{t,M,h} \\
= \; & \arg\max_{\theta\in\Theta} \frac{1}{t} \sum_{l=2}^{t} \left(\ln \widehat{f}_{M,h}\left(X^*_l | X^*_{l-1}, \theta\right) \tau_M\left(\widehat{f}_{M,h}\left(X^*_l | X^*_{l-1}, \theta\right)\right) \right. \\
& -\theta' \left(\frac{1}{T} \sum_{l'=2}^{T} \frac{\nabla_\theta \widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \theta\right)}{\widehat{f}_{M,h}\left(X^*_{l'} | X^*_{t-l'}, \theta\right)} \Big|_{\theta=\widehat{\theta}_{t,M,h}} \tau_M\left(\widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \widehat{\theta}_{t,M,h}\right)\right) \right. \\
& \left.\left. +\tau'_M\left(\widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \widehat{\theta}_{t,M,h}\right)\right) \nabla_\theta \widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \theta\right) \Big|_{\widehat{\theta}_{t,M,h}} \ln \widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \widehat{\theta}_{t,M,h}\right)\right)\right)
\end{aligned}$$

where $\tau'_M(\cdot)$ denotes the derivative of $\tau_M(\cdot)$ with respect to its argument. Note that each term in the simulated likelihood is recentered around the (full) sample mean of the score, evaluated at $\widehat{\theta}_{t,M,h}$. This ensures that the bootstrap score has mean zero, conditional on the sample. The recentering term requires computation of $\nabla_\theta \widehat{f}_{M,h}\left(X_{l'} | X_{l'-1}, \widehat{\theta}_{t,M,h}\right)$, which is not known in closed form. Nevertheless, it can be computed numerically, by simply taking the numerical derivative of the simulated likelihood.

Bootstrap Estimators for Multifactor Model

The SGMM and the bootstrap SGMM estimators in the case of multifactor model are similar as in one factor model. The difference is that the simulation scheme (2.15) and (2.16) are used instead of (2.14).

For recursive NPSQML estimators, to construct the bootstrap counterpart $\widehat{\theta}^*_{t,M,L,h}$ of $\widehat{\theta}_{t,M,L,h}$, since $M/T \to \infty$ and $L/T \to \infty$, the contribution of simulation error is asymptotically negligible. Hence, there is no need to resample the simulated observations or the simulated initial values for volatility. Define:

$$\widehat{f}_{M,L,h}\left(X_t^*|X_{t-1}^*,\theta\right)=\frac{1}{L}\sum_{j=1}^{L}\frac{1}{M\xi_M}\sum_{i=1}^{M}K\left(\frac{X_{t,i,h}^{\theta}(X_{t-1}^*,V_j^{\theta})-X_t^*}{\xi_M}\right).$$

Now, for $t=R,...,R+P-1$, define:

$$\begin{aligned}
&\widehat{\theta}^*_{t,M,L,h}\\
=\;& \arg\max_{\theta\in\Theta}\frac{1}{t}\sum_{l=2}^{t}\Bigg(\log\widehat{f}_{t,M,L,h}\left(X_l^*|X_{l-1}^*,\theta\right)\tau_M\left(\widehat{f}_{t,M,L,h}\left(X_l^*|X_{l-1}^*,\theta\right)\right)\\
&-\theta'\Bigg(\frac{1}{T}\sum_{l'=2}^{T}\frac{\nabla_\theta\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\theta\right)}{\widehat{f}_{t,M,L,h}\left(X_{l'}^*|X_{t-l'}^*,\theta\right)}\Big|_{\theta_{t,M,L,h}}\tau_M\left(\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\right)\\
&\qquad+\tau_M'\left(\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\right)\\
&\times\nabla_\theta\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\theta\right)\Big|_{\widehat{\theta}_{t,M,L,h}}\ln\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\Bigg)\Bigg),
\end{aligned}$$

where $\tau_M'(\cdot)$ denotes the derivative with respect to its argument.

Of note is that each bootstrap term is recentered around the (full) sample mean. This is necessary because the bootstrap statistic is constructed using the last P resampled observations, which in turn have been resampled from the full sample. In particular, this is necessary regardless of the ratio, P/R. In addition, in the case $P/R\to 0$, so that there is no need to mimic parameter estimation error, the bootstrap statistics can be constructed using $\widehat{\theta}_{t,M,L,h}$ instead of $\widehat{\theta}^*_{t,M,L,h}$.

Step 3: Using the same set of random variables used in the construction of the actual statistics, construct $X_{i,t+\tau,*}^{\widehat{\theta}^*_{t,N,h}}$ or $X_{k,i,t+\tau,*}^{\widehat{\theta}^*_{t,N,h}}$, $i=1,....,S$ and $t=1,...,T-\tau$. Note that we do not need resample the simulated series (as $L/T\to\infty$,simulation error is asymptotically negligible). Instead, simulate the series using bootstrap estimators and using bootstrapped values as starting values.

Step 4: Corresponding bootstrap statistics $V_{T,N,h}^{2*}$ (or $Z_{T,N,h}^*$, $D_{k,P,S}^*$, $SV_{T,N,h}^{2*}$, $SZ_{T,N,h}^*$, $SD_{k,P,S}^*$ depending on the types of tests) which are built on $\widehat{\theta}^*_{t,N,h}$ ($\widehat{\theta}^*_{t,N,L,h}$)

then are followed correspondingly. For the numerical implementation, again, of importance note is that in the case where we pick the choice $P/R \to 0, P, T, R \to \infty$, there is no need to re-estimate $\widehat{\theta}^*_{t,N,h}(\widehat{\theta}^*_{t,N,L,h})$. $\widehat{\theta}_{t,N,h}$ $(\widehat{\theta}^*_{t,N,L,h})$ can be used in all the bootstrap replications.

Step 5: Repeat the bootstrap Steps 1-4 B times, and generate the empirical distribution of the B bootstrap statistics.

2.4 Summary of Empirical Applications of the Tests

In this section, we briefly review some empirical applications of the methods discussed above. We start with unconditional distribution test, as in CS (2005), then give a specific empirical example using the conditional distribution test from BCS (2008). Finally, we briefly discuss on conditional distribution specification test applied to multiple competing models. The list of the diffusion models considered are provided in Table 2.1.

Table 2.1: Specification Test Hypotheses of Short Rate Processes[24]

Model	Specification	Reference, Hypothesis
Wong[25]	$dr(t) = (\alpha - \lambda - r(t))dt + \sqrt{\alpha r(t)}dW_r(t)$	CS (2005),H1 BCS (2008),H2
CIR	$dr(t) = k_r\left(\overline{r} - r(t)\right)dt + \sqrt{V(t)}dW_r(t),$	Cai& Swanson (2011) H2&H3
CEV	$dr(t) = k_r\left(\overline{r} - r(t)\right)dt + \sigma_r r(t)^{\rho}dW_r(t)$	Cai & Swanson (2011) H2&H3
SVJ[26]	$dr(t) = k_r\left(\overline{r} - r(t)\right)dt + \sqrt{V(t)}dW_r(t)$ $+J_u dq_u - J_d dq_d,$ $dV(t) = k_v\left(\overline{v} - V(t)\right)dt + \sigma_v\sqrt{V(t)}dW_v(t),$	BCS (2008), H2 Cai & Swanson (2011)
CHEN	$dr(t) = \kappa_r\left(\theta(t) - r(t)\right)dt + \sqrt{V(t)}dW_r,$ $dV(t) = \kappa_v\left(\overline{v} - V(t)\right)dt + \sigma_v\sqrt{V(t)}dW_v(t),$ $d\theta(t) = \kappa_\theta\left(\overline{\theta} - \theta(t)\right)dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t),$	Cai & Swanson (2011) H2&H3
CHENJ	$dr(t) = \kappa_r\left(\theta(t) - r(t)\right)dt + \sqrt{V(t)}dW_r(t)$ $+J_u dq_u - J_d dq_d,$ $dV(t) = \kappa_v\left(\overline{v} - V(t)\right)dt + \sigma_v\sqrt{V(t)}dW_v(t),$ $d\theta(t) = \kappa_\theta\left(\overline{\theta} - \theta(t)\right)dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t),$	Cai & Swanson (2011) H2&H3

Note that specification testing of the first model - a simplified version of the CIR model (we refer to this model as Wong) is carried out using the unconditional distribution test. With the cumulative distribution function known in closed form as in (2.10), the test statistic can be straightforwardly calculated. It is also convenient

[24] In the 3rd column, H1, H2 and H3 denote Hypothesis 1, Hypothesis 2 and Hypothesis 3, respectively. The hypotheses are presented corresponding to the references in the third column. For example, for CIR model, H2 corresponds to BCS (2008) and H2, H3 correspond to Cai and Swanson (2011).

[25] This model is a simplified version of the CIR model. For convenience, we refer to this model as Wong.

[26] Data used for Stochatic Volatility and Jump (SVJ) model is the same as in CIR Model.

to use GMM estimation in this case as the first two moments are known in closed form, i.e. $\alpha - \lambda$ and $\alpha/2(\alpha - \beta)$, respectively. CS (2005) examine Hypothesis 1 using simulated data. Their Monte Carlo experiments suggest that the test is useful, even for samples as small as 400 observations.

Hypothesis 2 is tested in BCS (2008) and Cai and Swanson (2011). For illustration, we focus on the results in BCS (2008) where CIR, SV and SVJ models are empirically tested using the one-month Eurodollar deposit rate (as a proxy for short rate) for the sample period January 6, 1971 - September 30, 2005, which yields 1,813 weekly observations. Note that one might apply these tests to other datasets including the monthly federal funds rate, the weekly 3-month T-bill rate, the weekly US dollar swap rate, the monthly yield on zero-coupon bonds with different maturities, and the 6-month LIBOR. Some of these variables have been examined elsewhere, for example in Ait-Sahalia (1999), Andersen, Benzoni and Lund (2004), Dai and Singleton (2000), Diebold and Li (2006, 2007), and Piazzesi (2001).

The statistic needed to apply the test discussed in Section 2.3.1.2 is:

$$Z_T = \sup_{v \in V} |Z_T(v)|,$$

where

$$Z_T(v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1\left\{X_t \leq v\right\};$$

and

$$Z_T^* = \sup_{v \in V} |Z_T^*(v)|,$$

where

$$
\begin{aligned}
& Z_T^*(v) \\
= \; & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ \underline{u} \leq X_{s,t+\tau,*}^{\widehat{\theta}^*_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau}^* \leq \overline{u}\} \right) 1 \left\{ X_t^* \leq v \right\} \\
& - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ \underline{u} \leq X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1 \left\{ X_t \leq v \right\}.
\end{aligned}
$$

For the case of stochastic volatility models, similarly we have:

$$
SZ_T = \sup_{v \in V} \left| SZ_T(v) \right|,
$$

where

$$
\begin{aligned}
& SZ_T(v) \\
= \; & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1 \left\{ \underline{u} \leq X_{j,s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1 \left\{ X_t \leq v \right\};
\end{aligned}
$$

and its bootstrap analog

$$
SZ_T^* = \sup_{v \in V} \left| SZ_T^*(v) \right|,
$$

where

$$
\begin{aligned}
& SZ_T^*(v) \\
= \; & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1 \left\{ \underline{u} \leq X_{j,s,t+\tau,*}^{\widehat{\theta}^*_{i,T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau}^* \leq \overline{u}\} \right) 1 \left\{ X_t^* \leq v \right\} \\
& - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1 \left\{ \underline{u} \leq X_{j,s,t+\tau}^{\widehat{\theta}_{i,T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1 \left\{ X_t \leq v \right\}.
\end{aligned}
$$

BCS (2008) carry out these tests using $\tau - Step$ ahead confidence intervals. They set $\tau = \{1, 2, 4, 12\}$ which corresponds to one week, two week, one month, and one quarter ahead intervals and set $(\underline{u}, \overline{u}) = (\overline{X} \pm 0.5\sigma_X, \, \overline{X} \pm \sigma_X)$, covering 46.3% and

72.4% coverage, respectively. $\overline{X}$ and σ_X are the mean and variance of an initial sample of data. In addition, $S = \{10T, 20T\}$ and $l = \{5, 10, 20, 50\}$.

For illustrative purposes, we report one case from BCS (2008). The test is implemented by setting $S = 10T$ and $l = 25$ for the calculation of both Z_T and SZ_T. In the Table 2.2, single, double, and triple starred entries represent rejection using 20%, 10%, and 5% size tests, respectively. Not surprisingly, the findings are consistent with some other papers in the specification test literature such as such as Aït-Sahalia (1996) and Bandi (2002). Namely, the CIR model is rejected using 5% size tests in almost all cases. When considering SV and SVJ models, smaller confidence intervals appear to lead to more model rejections. Moreover, results are somewhat mixed when evaluating the SVJ model, with a slightly higher frequency of rejection than in the case of SV models.

Table 2.2: Empirical Illustration of Specification Testing - CIR, SV, SVJ Models

	$(\underline{u}, \overline{u})$		CIR			SV			SVJ	
		Z_T	5% CV	10% CV	SZ_T	5% CV	10% CV	SZ_T	5% CV	10% CV
					$l = 25$					
1	$\overline{X} \pm 0.5\sigma_X$	0.5274***	0.2906	0.3545	0.9841***	0.8729	0.9031	1.1319	1.8468	2.1957
	$\overline{X} \pm \sigma_X$	0.4289***	0.2658	0.3178	0.6870	0.6954	0.7254	1.2272*	1.1203	1.3031
2	$\overline{X} \pm 0.5\sigma_X$	0.6824***	0.4291	0.4911	0.4113	1.3751	1.4900	0.9615*	0.8146	1.1334
	$\overline{X} \pm \sigma_X$	0.4897*	0.4264	0.5182	0.3682	1.1933	1.2243	1.2571	1.3316	1.4096
4	$\overline{X} \pm 0.5\sigma_X$	0.8662**	0.7111	0.8491	1.2840	2.3297	2.6109	1.5012*	1.1188	1.6856
	$\overline{X} \pm \sigma_X$	0.8539*	0.7521	0.9389	1.0472	2.2549	2.2745	0.9901*	0.9793	1.0507
12	$\overline{X} \pm 0.5\sigma_X$	1.1631*	1.0087	1.3009	1.7687	4.9298	5.2832	2.4237*	2.0818	3.0640
	$\overline{X} \pm \sigma_X$	1.0429	1.4767	2.0222	1.7017	5.2601	5.6522	1.4522	1.7400	2.1684

$^{(*)}$ Notes: Tabulated entries are test statistics and 5%, 10% and 20% level critical values. Test intervals are given in the second column of the table, for τ =1, 2, 4, 12. All tests are carried out using historical one-month Eurodollar deposit rate data for the period January 1971 - September 2005, measured at a weekly frequency. Single, double, and triple starred entries denote rejection at the 20%,10%, and 5% levels, respectively. Additionally, $\overline{X}$ and σ_X are the mean and standard deviation of the historical data. See above for complete details.

Finally, turning to **Hypothesis 3,** Cai and Swanson (2011) use an extended version of the above dataset, i.e. the one-month Eurodollar deposit rate from January 1971 to April 2008 (1,996 weekly observations). Specifically, they examine whether the CHEN model is the "best" model amongst multiple alternative models including those outlined in Table 2.1. The answer is "yes". In this example, the test was implemented using $D_{k,p,N}(u1, u2)$, as described in Sections 2.3.1 and 2.3.2, where $P = T/2$ and predictions are constructed using recursively estimated models and the simulation sample length used to address latent variable initial values is set at $L = 10T$. The choice of other inputs to the test such as τ and interval $(\underline{u}, \overline{u})$ are the same as in BCS (2008). The number of replications S, the block length l and number of bootstrap replications are $S = 10T$, $l = 20$ and $B = 100$.

Cai and Swanson (2011) also compare the Chen model with the so called Smooth Transition Autoregression Model (STAR) model defined as follows:

$$r_t = (\theta_1 + \beta_1 r_{t-1})G(\gamma, z_t, c) + (\theta_1 + \beta_2 r_{t-1})(1 - G(\gamma, z_t, c)) + u_t$$

where u_t is a disturbance term, θ_1, β_1, γ , β_2, and c are constants, $G(\cdot)$ is the logistic CDF (i.e. $G(\gamma, z_t, c) = \frac{1}{1+\exp(\gamma(z_t - c))}$), and the number of lags, p is selected via the use of Schwarz information criterion. Test statistics and predictive density type "mean square forecast errors" (MSFEs) values are again calculated as in Section 2.3.1 and 2.3.2.[27] Their results indicate that at a 90% level of confidence, one cannot reject the null hypothesis that the CHEN model generates predictive densities at least as accurate as the STAR model, regardless of forecast horizon and confidence interval width. Moreover, in almost all cases, the CHEN model has lower MSFE, and the magnitude of the MSFE differential between the CHEN model and STAR model rises as the forecast horizon increases. This confirms their in-sample findings that the CHEN model also wins when carrying out in-sample tests.

[27] See Table 6 in Cai and Swanson (2011) for complete details.

2.5 Concluding Remarks

This chapter reviews a class of specification and model selection type tests developed by CS (2005), BCS (2008) and CS (2011) for continuous time models. We begin with outlining the setup used to specify the types of diffusion models considered in this chapter. Thereafter, diffusion models in finance are discussed, and testing procedures are outlined. Related testing procedures are also discussed, both in contexts where models are assumed to be either correctly specified under the null hypothesis or generically misspecified under both the null and alternative test hypotheses. In addition to discussing tests of correct specification and test for selecting amongst alternative competing models, using both in-sample methods and via comparison of predictive accuracy, methodology is outlined allowing for parameter estimation, model and data simulation, and bootstrap critical value construction.

Several extensions that are left to future research are as follows. First, it remains to construct specification tests that do not integrate out the effects of latent factors. Additionally, it remains to examine the finite sample properties of the estimators and bootstrap methods discussed in this chapter.

Chapter 3

Volatility in Discrete and Continuous Time Models: A Survey with New Evidence on Large and Small Jumps

3.1 Introduction

This chapter explains how a mixture of discrete jump processes and continuous diffusion processes came to be used to model volatility in modern empirical finance. In the first part of the chapter, we present a selective survey of discrete-time and continuous-time volatility models and include a quick summary of the empirical findings. After noting that volatility modeling can be viewed as a missing data problem, we quickly discuss the traditional discrete-time models, some newer parametric continuous-time models, and some new nonparametric continuous-time models. Our focus is how the earlier literature link to the newer literature that mixes jump processes and diffusion processes. Finally, we use methods presented in Duong and Swanson (2010), to look for jumps in tick-by-tick data on 25 stocks. Modeling and forecasting financial asset return volatility constitute two cornerstones in modern econometric research. Volatility is a key factor in portfolio allocation and in pricing options and more advanced financial instruments known as derivatives.[28] So, volatility lies at the heart of many

[28] See Carr and Lee (2003, 2009) for more disussion on volatility-based derivative products.

modern financial calculations. Unfortunately, the true process generating volatility is unobserved, despite the ironic fact that different levels of "volatility" are reported on the evening news.

The literature on estimating volatility is vast, which is to be expected given it's practical and theoretical importance to modern finance. This chapter tells the story of how a mixture of discrete jump processes and continuous diffusion process came to be used to model volatility, it discusses some methods for detecting or estimating jumps, and it presents some more empirical evidence that jumps can be detected in high-frequency financial data. The availability of high-frequency financial data is an essential precursor to the continuous-time and jump-process literature, and enables financial econometricians to estimate continuous-time models.

We begin by looking at the key ARCH class of models, followed by a discussion of the class of continuous time processes frequently used in finance and the link between discrete time and continuous time models. We then discuss the construction of implied volatility in the Black-Scholes framework, and generalizations thereof. and it's extensions. Finally, we discuss recent research in the area of "model free" estimation of integrated volatility via use of so-called realized volatility, and variants thereof called realized measures. In our empirical investigation, we use realized measures to investigate the role of jumps in the realized variation of stock price returns.

The importance of realized volatility to econometric modelling is now obvious. For example, future realized volatilities are often used in the so-called variance swap, an important product in the volatility derivative market. Other products that use realized volatility such as caps on the variance swap, corridor variance swaps, and options on realized volatility have also been introduced into the class of volatility related financial instruments traded in financial markets. The key here is that investors worry about future volatility risk, and hence often choose to opt for this type of contract in order to hedge against it.[29] Realized volatility is also needed for calculation of the variance risk premium, a new financial variable that has interesting implications in asset pricing. For example, Bollerslev, Tauchen and Zhou (2009) find that the variance risk premium is able to explain time-series variation in post-1990 aggregate stock market returns with high (low) premia predicting high (low) future returns. Finally, note that in the context of realized volatility, jumps have a significant impact on modeling and forecasting volatility and it's realized measures. For example, when jumps are present, realized volatility is a biased estimator of integrated volatility. Thus, practitioners who are interested in modeling risks associated with continuous components of return processes, or integrated volatility, should use carefully designed realized measures that take jumps effects into account.[30] Careful analysis of jumps and realized measures in the presence of jumps are crucial elements to any reasonable quantification of risk. Moreover, several authors[31] have found that separation of continuous components from jump components can improve forecasts of future realized volatility. This finding should be of great interest to practitioners, especially when their objective is hedging. In summary, risk, or volatility plays an important role in many areas of econometrics used in the finance industry. However, as volatility is generally unmeasured, it poses a standard sort of "missing variable" problem.

Turning back to our discussion of jumps, note that evidence of jumps in financial markets is plentiful. In an important paper, Huang and Tauchen (2005) find evidence of jumps for S&P cash and future (log) returns from 1997 to 2002, in approximately 7% of the trading days. Their test for jumps requires the jump component to be a com-

[29] Volatility and variance swaps are newer hedging instruments, adding to the traditional volatility "Vega", which is derived from options data. See Hull (1997, pp. 328) for a definition of Vega. For example, as noted in Carr and Lee (2009), the UBS book was short many millions of vega in 1993, and they were the first to use variance swaps and options on realized volatility to hedge against volatility risk.

[30] See Corradi, Distaso and Swanson (2009, 2011) for discussion of prediction of integrated volatility.

[31] For instance, see Andersen, Bollerslev and Diebold (2007).

pound Poisson process. Several authors, including Cont and Mancini (2007), Tauchen and Todorov (2010) and Aït-Sahalia and Jacod (2009b) have taken the analysis of jumps one step further by developing tests to ascertain whether the process describing an asset contains "infinite activity jumps" - those jumps that are tiny and look similar to continuous movements, but whose contribution to the jump risk of the process is not negligible. Cont and Mancini (2007) implement their method of testing for the existence of infinite activity jumps using foreign exchange rate data, and find no evidence infinite activity jumps. Aït-Sahalia and Jacod (2009b) estimate that the degree of activity of jumps in Intel and Microsoft log returns is approximately 1.6, which implies evidence of infinite activity jumps for these, and possibly many other stocks. Andersen, Bollerslev and Diebold (ABD: 2007) find that separating out the volatility jump component results in improved out-of-sample volatility forecasting, and find that jumps are closely related to macroeconomic announcements. In summary, it is now generally accepted that many return processes contain jumps.

In the part of this chapter that present empirical findings relevant to the topic discussed in previous sections, we examine high frequency data for 25 stocks in the DOW 30, using 5 minute interval observations, and for the sample period from 1993 to 2008. Some of the stocks in our data set, (e.g. Microsoft and Intel) have been found to be characterized by infinite activity jumps by Aït-Sahalia and Jacod (2009b), and therefore do not belong to the class of finite activity jump processes that Barndorff-Nielsen and Shephard (BNS: 2006) has often been applied to. This fact underscores the importance of the recent papers by Jacod (2008), Tauchen and Todorov (2010) and Aït-Sahalia and Jacod (2009a,b), where new limit theory applicable to infinite activity is implemented and developed; and underscores why the results of these papers are used in our empirical investigation. In summary, we find evidence of jumps in around 22.8% of the days in the 1993-2000 period, and 9.4% in the 2001-2008 period. This degree of jump activity implies more (jump induced) turbulence in financial markets in the previous decade than the current decade. However, and as expected, the prevalence of "large" jumps varies across these periods. (Note that we examine large jumps by picking 3 different fixed γ levels, corresponding to $50th, 75th$ and $90th$ percentiles of samples of the monthly maximum return increments, i.e. our monthly “abnormal event" samples.) In particular, large jump activity increases markedly during the 2001-2008 period, with respect to its contribution to the realized variation of jumps and with respect to the contribution of large jumps to the total variation of the (log) price process. This suggests that while the overall role of jumps is lessening, the role of large jumps has not decreased, and indeed, the relative role of large jumps, as a proportion of overall jumps has actually increased in the 2000s. Note that this result holds on average across all 25 stocks examined. In summary, it appears that frequent “small" jumps of the 1990s have been replaced with relatively infrequent "large" jumps in recent years. Interestingly, this result holds for all of the stocks that we examine, supporting the notion that their is strong co-movement across jump components for a wide variety of stocks, as discussed in Bollerslev, Law and Tauchen (2008).

The rest of the chapter is organized as follows: Section 3.2 discusses volatility models in discrete time. Section 3.3 discusses volatility models in continuous time, and outlines various measures used to estimate (and forecast) volatility. Section 3.4 summarizes results from the extant testing and prediction literatures that are often used in the study of volatility, and in particular that are used in our empirical investigation that is presented in Section 3.5. Concluding remarks are contained Section 3.6.

3.2 Volatility Models in Discrete Time

A number of important stylized facts have emerged from empirical studies of discrete-time models. After discussing these stylized facts, we review discrete-time models and then discuss Nelson (1990) and Corradi (2000), which link the discrete-time and continuous-time literatures. This chapter is not a survey of the huge literature on discrete-time models that began with the seminal papers by Engle (1982), Bollerslev (1986) and Nelson (1991). See Bollerslev, Chou and Kronner (1992) and Bollerslev, Engle and Nelson (1994) for more complete introductions.

3.2.1 Stylized Facts in Financial Market - Directions for Volatility Models

It is well-known in empirical finance that asset returns share various regularities, all of which guide financial economists and econometricians in their choice of models. These stylized facts have been discussed by many authors. Here, we highlight some of them that pertain to stock returns.

Leptorkurtosis: Asset returns have been noted by Mandelbrot (1963) and Fama (1965) to have fat tails and one therefore should use non-normal distributions to model their dynamics. Fama (1965) shows evidence of excess kurtosis in the distribution of stock returns. According to Clark (1973), a stochastic process is fat tailed if it is conditionally normal with a randomly changing conditional variance. Engle and González-Rivera (1991) introduce a semi-parametric volatility model, which allows for generic return distributions.

Volatility Clustering and Persistence: By observing cotton prices, Mandelbrot (1963) stressed that ".... large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes...". The persistence of shocks to the conditional variance of stock returns seems to be clear. The interpretation of this persistence as well as how long the shocks persist are crucial in specifying the "correct" dynamics. Porterba and Summers (1986) note that volatility shocks may affect the entire term structure, associated risk premia, and investment in long-lived capital goods.

Volatility persistence is an important feature that pertains to models with time varying and codependent variance. Black and Scholes (1973) wrote that "...there is evidence of non-stationary in the variance. More work must be done to predict variances using the information available.". Since their chapter, numerous autoregressive conditional heteroskedasticity, volatility and stochastic volatility models have been developed.

Leverage Effects: Black (1976) observes that changes in stock prices seem to be negatively correlated with changes in stock volatility. Volatility seems to increase after bad news and decrease after good news. Schwert (1989, 1990) presents empirical evidence that stock volatility is higher during recessions and financial crises. Christie (1982) discusses economic mechanisms that explain this effect. Specifically, reductions in equity value raise the riskiness of firms, as implied by debt to equity ratios, and therefore lead to increases in future volatility. For modeling, Nelson (1991) suggests a new model that captures the asymmetric relation between returns and changes in volatility.

Co-movement in Volatilities. This is also first commented on by Black (1976). He points out the commonality in volatility changes across stocks. When stock volatilities change, they all tend to change in the same direction. This suggests that (few) common (unobserved or missing) factors might be specified when modelling individual asset return volatility.

3.2.2 ARCH and GARCH Models

In this section, we discuss the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and the Generalized ARCH (GARCH) model of Bollerslev (1986), as well as related models. These models are very well known, but we cover them in moderate detail in order to trace the evolution to the modern continuous-time and jump-process literatures.

Uncovering the correct conditional volatility specification is important in finance because misspecified models can leave modellers with erroneous information about the risk return trade-off of investments. For investment decision-making, risk-averse investors take into account not only expected returns, but also the level of risk. Investors demand risk premia for risk and the risk premium may include premia due to changes in volatility. In risk neutral pricing, a measure of volatility is needed for the derivation of the market price of risk.

The difficulties in specifying an acceptable trade-off between flexibility and parsimony, as well as settling on a model capable of picking up the key stylized facts caused the discrete-time literature to grow very quickly. From our perspective, we note that similar difficulties have had a similar effect on the continuous-time literature. In this survey, we focus on a few basic discrete-time models and how they relate to continuous-time models.

Turning to our discussion of ARCH type models, let X_t be a financial asset return, say, and F_{t-1} denotes a filtration of all information through time $t-1$. The prototypical autoregressive conditional heteroskedasticity (ARCH) model has:

$$X_t = \varepsilon_t \sigma_t$$

$$\varepsilon_t \sim i.i.d \text{ with}$$

$$E(\varepsilon_t) = 0 \text{ and } Var(\varepsilon_t) = 1$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$$

and in the more general case:

$$X_t | F_{t-1} \sim N(Z_t \beta, \sigma_t^2)$$

$$\sigma_t^2 = h(\varepsilon_{t-1}, \varepsilon_{t-2,\ldots,}\varepsilon_{t-p}, \alpha)$$

$$\varepsilon_t = X_t - Z_t \beta$$

where Z_t may contain lags of X_t. If the function h contains current and lagged X's, then

$$\sigma_t^2 = h(\varepsilon_{t-1}, \varepsilon_{t-2,\ldots,}\varepsilon_{t-p}, x_t, x_{t-1,\ldots,}x_{t-p}, \alpha)$$

In this class of models, ARCH(p) is the most popular where

$$X_t | F_{t-1} \sim N(Z_t \beta, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \ldots \alpha_p \varepsilon_{t-p}^2$$

Engle (1982) proposes a convenient estimation and testing methodology for the model using maximum likelihood. He shows that α and β could be estimated separately under some regularity conditions.[32] To capture the trade-off between risk and expected return, Engle, Lilien and Robins (1987) introduce ARCH in mean, or ARCH-M models. Let

$$X_t = g(Z_{t-1}, \sigma_t^2; b) + \varepsilon_t$$

The appealing feature of this model is that the conditional mean, μ_t, is a function of the variance, i.e. $\mu_t = g(Z_{t-1}, \sigma_t^2; b)$. This helps us to model directly the risk-return relationship, and has important implications for predicting the conditional

[32] For details, see Sections 4 and 5 in Engle (1982).

mean function, since the conditional volatility enters therein. The parametric choice for the function g depends on the modeler. In practice, many papers set g to be a linear or logarithmic function.

An important improvement to these models is made by Bollerslev (1986), where the ARCH model is generalized to the Generalized ARCH (GARCH) model. As noted in Bollerslev (1986), the extension from ARCH to GARCH is similar to the extension in time series modelling of an AR to an ARMA model. Specifically, as in the case of the ARCH model, let ε_t be the innovation in a linear regression

$$\varepsilon_t = X_t - Z_t'\beta$$

Then the GARCH (p,q) specification is given by

$$\varepsilon_t | F_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$
$$\varepsilon_t = X_t - Z_t \beta$$

where p and q denote lag orders, and

$$p \geq 0, q > 0$$
$$\alpha_0 > 0,\ \ \alpha_i \geq 0,\ \ i = 1, ..., q$$
$$\beta_i \geq 0,\ \ i = 1, ..., p$$

It is clear that the difference between the above set up and the ARCH model is the linear lagged conditional variances. Conditional volatility today not only depends on the lagged innovations but also on lagged conditional volatilities.

Bollerslev (1986) presents a complete set of results on the conditions under which the model is appropriate, as well as maximum likelihood and testing procedures for implementing the general GARCH (p,q) model. The most successful model, empirically, is the GARCH(1,1) model. Engle and Bollerslev (1986) discuss the so-called Integrated GARCH or IGARCH model. Under this specification, $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$, and this leads to a unit root in the volatility equation.

In other key papers, Nelson (1990, 1991) discusses the use of EARCH (i.e., exponential ARCH) to approximate continuous time processes. Nelson (1991) points out that the GARCH model has several limitations in empirical applications to financial markets. For instance, in the GARCH model, volatility responds symmetrically to positive and negative residuals and therefore does not explain the stylized leverage effect. In lieu of this, Nelson (1991) proposes the Exponential ARCH, or EARCH in which the volatility function is constructed as follows:

$$X_t = \sigma_t \varepsilon_t$$

$$\varepsilon_t \sim i.i.d \text{ with}$$

$$E(\varepsilon_t) = 0 \text{ and } Var(\varepsilon_t) = 1$$

and

$$\ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(\varepsilon_{t-k}), \qquad \beta_1 \equiv 1$$

where $\{\alpha_t\}_{t=-\infty,\infty}$ and $\{\beta_k\}_{k=1,\infty}$ are parameters.

The choice for the functional form of g is

$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma(|\varepsilon_t| - E|\varepsilon_t|)$$

This set-up allows the conditional variance process to respond asymmetrically to rises and falls in stock prices. It is straightforward to verify this as when ε_t is positive $g(\varepsilon_t) = (\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|)$ and when ε_t is negative $g(\varepsilon_t) = (\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|)$. In each case, the $g(\varepsilon_t)$ is linear function with a different slope.

In addition, Nelson (1991) points out, while in GARCH, it is difficult to verify the persistence the shocks to the variance, in the EARCH model, the stationarity and ergodicity of the logarithm of the variance process are easily checked. He states conditions for the ergodicity and strict stationarity of $\{exp(-\alpha_t \sigma_t^2)\}$ and $\{exp(-\alpha_t/2X_t)\}$.[33] Other modifications of the GARCH (1,1) model include the GJR model proposed by Glosten, Jaganathan and Runkle (1993). This model imposes structure that induces asymmetry in shocks to returns in a different way. Namely, they define

$$\sigma_t^2 = \omega + \alpha\varepsilon_t^2 + \gamma\varepsilon_t^2 1_{\{\varepsilon_t \geq 0\}} + \beta\sigma_{t-1}^2$$

Note that when $\gamma < 0$, positive return shocks increase volatility less than negative shocks.

The above discussion summarizes a very few of the important models in class of discrete ARCH models.[34] In addition to these models, there have been many modifications and improvements. For a complete list and discussion, see Bollerslev (2008), where he provides a Glossary to ARCH. For models with multivariate specifications (see Bollerslev, Engle and Wooldbridge (1988)). In the next section, we will highlight some links between discrete time models and continuous time models in the framework of modeling volatility.

3.2.3 From ARCH and GARCH to Continuous Time Models

An interesting aspect of the volatility literature is the connection between discrete time and continuous time models. In the case of constant volatility, the classical

[33] See Theorem 2.1 in Nelson (1991).

[34] We also present the work of Heson and Nandi (2000) and Barone-Adesi, Engle, Mancini (2008) which use modifications of ARCH models for option pricing in Section 3.2.2.

result by Cox and Ross (1976) shows that the limiting form of the jump process

$$dX_t = \mu X_t dt + cX_t dN_t(\lambda)$$

as $\lambda \to 0$ is the diffusion process

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

where σ is a function of c. $N_t(\lambda)$ is a continuous time Poisson process with intensity λ, (i.e., dN_t is the number of jumps of X_t during dt and is Poisson-distributed with parameter λdt). cX_t is the jump amplitude and W_t is a standard Brownian motion. In another important research, Nelson (1990) bridges the gap between discrete and continuous time stochastic volatility models by using AR(1) Exponential ARCH and GARCH (1,1) models as approximations for continuous time processes. It should be noted that in this approximation framework, only the discrete models with one lag are relevant due to the characteristics of continuous time models. GARCH models with two more lags as explanatory variables are not relevant. Under certain assumptions,[35] Nelson (1990) at first looks at the GARCH (1,1) - M process of Engle and Bollerslev (1986)

$$\begin{aligned} X_t &= X_{t-1} + c\sigma_t^2 + \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega + \sigma_t^2[\beta + \alpha \varepsilon_t^2], \end{aligned}$$

If time is partitioned more finely, one can write the above difference equation as

$$\begin{aligned} X_{kh} &= X_{(k-1)h} + hc\sigma_{kh}^2 + \sigma_{kh}\varepsilon_{kh} \\ \sigma_{(k+1)h}^2 &= \omega_h + \sigma_{kh}^2[\beta_h + h^{-1}\alpha_h \varepsilon_{kh}^2], \end{aligned}$$

[35] For details, see Section 2 in Nelson (1990).

where h is the time increment and $\varepsilon_{kh} \sim i.i.d\ N(0,h)$. He shows that if h goes to 0 in the limit, this system converges weakly in distribution[36] to a diffusion process of the form

$$dX_t = c\sigma_t^2 dt + \sigma_t dW_{1t}$$

$$d\sigma_t^2 = (\omega - \theta\sigma_t^2)dt + \alpha\sigma_t^2 dW_{2t}$$

where W_{1t} and W_{2t} are linearly independent standard Brownian motions, independent of the initial values (X_0, σ_0^2).

In another important paper in this line of research, Corradi (2000) considers the limit when $h \to 0$, of the GARCH (1,1) process

$$X_{kh} - X_{(k-1)h} = \sigma_{(k-1)h}\varepsilon_{kh},$$

$$\sigma_{kh}^2 - \sigma_{(k-1)h}^2 = \omega_h + (\omega_{1h} - 1)\sigma_{(k-1)h}^2 + h^{-1}\omega_{2h}\sigma_{(k-1)h}^2\varepsilon_{kh}$$

She shows that in the limit, this system converges to either one of the following continuous time processes, depending on the parameters ω_{1h} and w_{2h}[37]

$$dY_t = \sigma_t dW_t$$

$$d\sigma_t^2 = (\omega_0 + \theta\sigma_t^2)dt$$

or

$$dY_t = \sigma_t dW_{1t}$$

$$d\sigma_t^2 = (\omega_0 + \theta\sigma_t^2)dt + \alpha\sigma_t^2 dW_{2t}$$

where (W_{1t}, W_{2t}) are two standard independent Brownian motions.

[36] For a definition of weak convergence for stochastic processes, see Billingsley (1978).
[37] For details, see Proposition 2.1 in the Corradi (2000).

Nelson (1990) also introduces a class of ARCH models which can approximate a wide range of stochastic differential equations. He investigates the approximations of the system of stochastic differential equations defined as follows:

$$dX_t = f dt + g dW_{1t}$$

$$dY_t = F dt + G dW_{2t}$$

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} [dW_{1,t} \ \ {}_{2,t}] = \Omega dt$$

where the first equation is univariate, Y_t is vector of latent state variables, W_{1t} can be correlated with elements in $W_{2t.}$, and F, G, f, g are functions of X_t, Y_t, and t. He shows that the above system is the limit of the following ARCH type discrete time system of difference equations

$$X_{(k+1)h} = X_{kh} + fh + g\varepsilon_{kh}$$

$$Y_{(k+1)h} = Y_{kh} + Fh + G\varepsilon^*_{kh}$$

where ε^*_{kh} corresponds to W_{2t} and is constructed on the basis of $\varepsilon_{kh} \sim N(0, h)$ and Ω.[38]

3.3 Volatility in Continuous Time

3.3.1 Continuous Time Models

While many financial economists have long preferred the continuous-time framework,[39] the availability of high-frequency data made feasible the econometric analysis of continuous-time models. As with discrete time models, the difficult choices involved

[38]For details on the construction of ε^*_{kh}, see Section 3.2 in Nelson (1990).
[39]For example, see Duffie (2001) for asset pricing in continuous time.

in specifying the volatility function have given rise to a plethora of models.

Over the last 20 years, continuous-time models have taken on a central role in option pricing, risk management and volatility forecasting. Continuous-time models allow decision makers to model and react to the nearly continuous flow of data. In principle, financial managers can update their trading strategies every second. For econometricians, the use of high-frequency data has interesting implications for both estimation and prediction.

The list of models discussed in this section is not meant to be exhaustive. Rather, the list is designed to lead to the jump-processes and the realized volatilities models in which we are interested.

Diffusion Processes:

Brownian Motion with Drift:

$$dX_t = \mu dt + \sigma dW_t$$

This specification has been used a lot in early work in economics and finance due to its simplicity. It is obvious that X_t is normally distributed with mean μt and variance $\sigma^2 t$.

Geometric Brownian Motion (Log Normal Model):

$$dX_t = \mu X_t dt + \sigma_t X_t dW_t$$

This model has been very popular for asset prices. It has been extensively used in the Black and Scholes (1973) option pricing framework and in structured corporate finance. The main drawback of this model is that the (log) return process has constant volatility. To see this, apply Itô' Lemma[40] to the function $f(x) = \log(x)$, yielding

$$d\log(X_t) = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t.$$

Ornstein-Uhlenbeck Process (sometimes referred to as the Vasicek (1977) model, and often used to model interest rates):

$$dX_t = (\alpha + \beta X_t)dt + \sigma dW_t.$$

Cox, Ingersoll and Ross (1985) analyze the following square root process, also known as the CIR model, again for modelling the term structure of interest rates, although this model, unlike the Vasicek model, ensures positivity of rates.

$$dX_t = (\alpha + \beta X_t)dt + \sigma\sqrt{X_t}dW_t.$$

Brennan and Schwartz (1979) and Courtadon (1982) analyze the process:

$$dX_t = (\alpha + \beta X_t)dt + \sigma X_t^2 dW_t.$$

Constant Elasticity of Variance (CEV):

$$dX_t = \mu X_t dt + \sigma X_t^{\beta/2} dW_t$$

[40] See Protter (1990) and Duffie (2001) for details on Itô' Lemma.

Note that the interpretation of this model depends on β. In particular, in the case using this process to model stock price, if $\beta = 2$, the price process X_t follows geometric Brownian motion and therefore the volatility of the (log) stock return process is constant. If $\beta < 2$, this model captures the leverage effects discussed above. Among other authors, Beckers (1980) uses this CEV model for stocks. Marsh and Rosenfeld (1983) apply the CEV model to interest rates and Emanuel and Macbeth (1982) use this set-up for option pricing.

Generalized Constant Elasticity of Variance:

$$dX_t = (\alpha X_t^{-(1-\omega)} + \beta X_t)dt + \sigma X_t^{\beta/2} dW_t$$

This process nests the log diffusion when $\beta = 2$, and square root process when $\beta = 1$.

Chan, Karolyi, Longstaff, and Sanders (1992) study the case of linear drift and CEV diffusion with $\beta \geq 2$

$$dX_t = (\alpha + \beta X_t)dt + \sigma X_t^{\beta/2} dW_t.$$

Duffie and Kan (1996) specify a mean reversion and square root structure in volatility for modelling of interest rates. In the univariate case:

$$dX_t = (\alpha - X_t)dt + \sqrt{\sigma_0 + \sigma_1 X_t} dW_t.$$

Aït-Sahalia (1996) looks at more the general case of diffusions with general drift and CEV:

$$dX_t = (\alpha + \beta X_t + \gamma X_t^2 + \eta / X_t)dt + \sigma X_t^{\beta/2} dW_t.$$

Needless to say, model selection is important issue when specifying diffusion models. Note that in the general setting, the diffusion process is written as

$$dX_t = \mu(X_t, t, \theta)dt + \sigma(X_t, t, \theta)dW_t$$

Also, it is known that the drift and diffusion terms $\mu(X_t, t)$ and $\sigma(X_t, t)$ respectively uniquely determine the stationary density, say $f(x, \theta_0)$, associated with the invariant probability measure of the above diffusion process.[41] In particular,

$$f(x, \theta_0) = \frac{c(\theta_0)}{\sigma^2(x, \theta_0)} \exp\left(\frac{2\mu(u, \theta_0)}{\sigma^2(u, \theta_0)} du\right)$$

In a seminal paper, Aït-Sahalia (1996) constructs a nonparametric test for interest rate models on the basis of the comparison of such stationary densities. In his empirical application to spot interest rates, he finds that the misspecification of the models in the literature on spot interest rates is mainly due to the linearity of the drift function in such models. In addition, his proposed model (general drift and CEV) could not be rejected. In this same line of research, Corradi and Swanson (2005, 2011) develop bootstrap specification tests for univariate and multifactor diffusion process that do not require knowledge of the transition density. Instead of comparing of densities, their method is based on a comparison of cumulative distribution functions. They also extend their methods to diffusion process with jumps and stochastic volatility.

Jump Diffusions:

[41] See Karlin and Taylor (1981), pp. 241.

Many high-frequency financial time series appear to be a mixture of sudden relatively large changes and smooth small changes. This image suggests modeling a high-frequency financial time series as a mixture of a discrete jump process and a continuous diffusion process. A jump-process J_t is a discrete process specified by a distribution, ν, for the magnitudes of the jumps and a distribution, $\lambda(X_t)$, for the intensity with which jumps occur. A jump-diffusion process is the sum of a continuous diffusion process and a jump process,

$$dX_t = \mu(X_t, t, \theta)dt + \sigma(X_t, t, \theta)dW + dJ_t$$

One pioneering work which incorporates jumps into continuous time processes is Merton (1976), where he models the continuous component of the log price process to be Gaussian as in the case of geometric Brownian motion. The magnitude of jumps also follows a Gaussian distribution, and jumps follow Poison distribution in his paper. Newer developments in this area do not "append" a "discrete" jump process onto the diffusion, but instead specify the jumps using other means, such as via the use of Levy processes.

Affine Jump Diffusion Model: This class of models is widely studied in the empirical finance literature, especially in term structure modelling. The family of affine processes X_t including jumps is parametrized as follow

$$dX_t = \kappa(\alpha - X_t)dt + \Omega\sqrt{D_t}dW + dJ_t$$

where W_t is an $N-$dimensional independent standard Brownian motion, κ and Ω are square $N \times N$ matrices. D_t is a diagonal matrix with ith diagonal element given by

$$d_{ii,t} = \theta_i + \delta_i' X_t$$

The jump intensity is assumed to be a positive, affine function of X_t and the jump size distribution is assumed to be determined by its conditional characteristic function. As shown by transform analysis in Duffie, Pan and Singleton (2000), the attractive feature of this class of affine jump diffusions is that the exponential affine structure characteristics function is known in closed form. Namely

$$\Phi_t(X_t) = \exp(a_t + b_t' X_t)$$

where functions a_t and b_t can be derived from Riccati equations.[42] With known characteristics function, one can use either GMM to estimate the parameters of this system of this jump diffusion, and can use quasi-maximum likelihood (QML), once the first two moments are obtained. In the univariate case without jumps, as a special case, this corresponds to the above general CIR model with jumps.

Stochastic Volatility Models:

Stochastic volatility models are popular, particularly for modelling asset prices and interest rates. They are first introduced by Harvey, Ruiz and Shephard (1994) in discrete time. Stochastic volatility implies that unobserved volatility follows another stochastic process. For example, one specification could be

$$dX_t = (\alpha + \beta X_t)dt + \sigma dW_{1t},$$

and the volatility process follows:

$$d\sigma_t^2 = \kappa(\vartheta - \sigma_t^2)dt + \delta\sigma_t dW_{2t},$$

where $Cov(dW_{1t}, dW_{2t}) = \rho dt$.

Andersen and Lund (1997) estimate the generalized CIR model with stochastic

[42] For details, see Singleton (2006), pp. 102.

volatility:

$$dX_t = \kappa_1(\alpha - X_t)dt + \sigma_t X_t^{\beta} dW_{1t},$$

$$d\log\sigma_t^2 = \kappa_1(\alpha - \log\sigma_t^2)dt + \delta dW_{2t}.$$

Mixed Stochastic Volatility, Jump Diffusion Models. An example of these models comes from the application of spectral GMM in Chacko and Viceira (2003) where the return process is specified as:

$$dX_t = (\mu - \frac{\sigma_t^2}{2})dt + \sigma_t dW_{1t} + [\exp(J_u) - 1]dN_u(\lambda_u) + [\exp(-J_d) - 1]dN_d(\lambda_d)$$
$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \delta\sigma_t dW_{2t},$$

where λ_u, λ_d are jump intensity parameters and are constant, and where. J_u and J_d >0 are stochastic jump magnitudes that follow an exponential distribution, i.e.

$$f(J_u) = \frac{1}{\eta_u}\exp(\frac{-J_u}{\eta_u}),$$

$$f(J_u) = \frac{1}{\eta_d}\exp(\frac{-J_d}{\eta_d}).$$

In addition, in the option pricing literature, many models are nested in the following data generating process which allows for jumps in both equations

$$dX_t = \mu_t dt + \sigma_t X_t dW_{1t} + dJ_{1t}$$
$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \delta\sigma_t(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}) + dJ_{2t}$$

where W_{1t} and W_{2t} are two independent Brownian motions process, and J_{1t} and J_{2t} are two jump processes. Popular models that are nested in this class include Heston (1993) with no jumps in either price or volatility, Bates (2000), Chernov and Ghysels

(2000) and Pan (2002).[43]

3.3.2 Implied Volatility from Option Pricing

3.3.2.1 Black-Scholes Framework as an Illustration

Implied volatility is considered to be the market prediction of future volatility, used in the context of option pricing. In this section, we firstly provide a standard method to show how econometricians can construct implied volatility from Black-Scholes option prices. Though stochastic volatility better captures the dynamics of asset returns, Black-Sholes model is still considered to be an important element of option pricing theory and practice. However, that said, after the current discussion we turn to a discussion of "model free" measures of implied volatility, and the so-called VIX volatility index.

Within the Black-Scholes framework, we restate the derivation of the European call price. In this model, stock prices are assumed to be log-normally distributed. A nice feature of this assumption is that option prices can then be derived in closed form, i.e. using the Black-Scholes (BS) formula. Once an option pricing function is known, implied volatility can in turn be backed out using stock, option and interest rate data. Note that one can also derive the price of any derivative whose payoff is a function of stock prices. Also, even under more complicated assumptions about the stock price process, the risk neutral pricing methodology presented below can still be applied.

Let's assume that we are interested in an asset market which has 3 assets: a riskless bond, a stock, and a derivative whose payoff is a function of the stock price. Also, say that under the physical probability measure, P, the price of a non-dividend

[43] For a more detailed discussion, see Singleton (2006), chapter 15.

paying stock, X_t, follows a geometric Brownian motion, i.e.

$$dX_s = \mu X_s ds + \sigma X_s dW_s$$

and the interest rate, r,associated with riskless bond (i.e. the short rate) is assumed to be constant. For simplicity, we analyze the dynamics of the process $\overline{X_s} = \frac{X_s}{ers}$. Using Itô' Lemma, $\overline{X_s}$ follows:

$$d\overline{X_s} = (\mu - r)\overline{X_s} ds + \sigma \overline{X_s} dW_s$$

Say that the call option on the stock has strike price K and maturity date T. It's payoff at time T is

$$C_T = (X_T - K)^+ = \max(X_T - K, 0)$$

Option pricing means we look for the price of the derivative c_t whose payoff at maturity is C_T. With the no arbitrage assumption, there must exist a risk neutral measure Q. Under this risk neutral measure, the discount process $\overline{X_s}$, $\overline{C_s} = \frac{C_s}{esr}$ has no drift term, or is a martingale:

$$\begin{aligned} d\overline{X_s} &= \sigma \overline{X_s} d\overline{W_s} \\ \overline{c_t} &= E^Q[\frac{C_T}{e^{rt}}|X_t] = e^{-rT} E^Q[((X_T - K)^+)|X_t] \end{aligned}$$

where $\overline{W_s}$ is a Brownian motion under measure Q. The expectation above is taken under the probability measure Q. Recall that probability measures P and Q can be transformed back and forth though Girsanov's Theorem.[44] In particular $\overline{X_T}/\overline{X_t}$ is log-normally distributed with mean $-(\sigma^2/2)(T-t)$ and variance $\sigma^2(T-t)$. Or, given that the asset $\overline{C_s}$ has no drift term, one can easily verify that the c_t process is the

[44] See Duffie (2001) for details on Girsanov's Theorem.

solution to the following partial differential equation

$$-rf + f_t + rSf_X + 1/2\sigma^2 f_{XX} = 0$$

with the boundary condition at the expiration date $f(S,T) =$ Payoff at time $T = \max(X_T - K, 0)$. Note that one can use this derivation approach for any asset, with a change in payoff function. In more complicated settings with more complicated price process dynamics, the same PDE approach can be applied.

However, in such cases, to solve the PDE, a numerical algorithm is needed as there may not be a closed form solution for the option price. Turning again to our closed form BS solution, note that it takes the form

$$c_t = \Phi(z) - e^{-r(T-t)}\Phi(z - \sigma\sqrt{T-t})$$

where Φ is the cumulative normal cdf and,

$$z = \frac{\log(X_t/K) + (r + \sigma^2/2(T-t))}{\sigma\sqrt{T-t}}$$

From the BS formula, one can invert the unobservable σ. In particular, σ is function of current time t variables, $X_t, r, \tau = T - t, K$ and c_t which are all observed and available in the data. This is useful in the framework of no-arbitrage pricing, and this option pricing tool is the key to backing out implied volatility. Generally, one can write

$$c_t = F(X_t, r, K, \tau, \sigma_\tau)$$

and the volatility process can be inferred once the nonlinear function F is known. If it is not known in closed form, we still can back out implied volatility via numerical analysis.

3.3.2.2 Deviation from Black-Scholes

The main drawback of Black-Scholes is that it is not consistent with empirical evidence that implied volatility varies across different maturities and strike price. Dumas, Flemning and Wahley (1998) propose an *ad hoc* Black-Scholes model in which volatility is not constant. This measure is now widely used by practitioners. *Ad hoc* BS allows for different implied volatilities to price options differently. In particular, implied volatilities are modeled as follows:

$$\sigma(\tau, K) = \alpha_0 + \alpha_1 K + \alpha_2 K^2 + \beta_0 \tau + \beta_1 \tau^2 + \beta_2 \tau K$$

where $\sigma(\tau, K)$ is the Black-Scholes implied volatility for and option with strike K and time to maturity τ. This method has been shown by Dumas et al. (1998) to be better than the constant volatility approach. Option pricing with varying volatility, such as a square root diffusion model, are also proposed by many researchers (for example, see the work of Bates (1996, 2000), Bakshi, Cao and Chen (1997) and Scott (1997)). Heston (1993) provides a parametric stochastic volatility model, i.e. square root process for volatility, and solves for closed-form prices. Chernov and Ghysels (2000) and Pan (2002) introduces a more general pricing framework, i.e. stochastic volatility with jumps. For a survey of this option pricing literature, see Bates (2003), and for a discussion of the econometrics of option pricing see Garcia, Ghysels and Renault (2010). In an interesting discrete-time set-up, Heston and Nandi (2000) proposes a closed form GARCH option price model in which they present an option formula for a stochastic volatility model with GARCH. This discrete-time set-up is close to the Heston (1993) continuous time stochastic volatility model, but is easier to implement. They specify the following dynamics for asset returns

$$\log(X(_t) = \log(X_{t-h}) + r + \lambda\sigma_t^2 + \sigma_t\varepsilon_t$$
$$\sigma_t^2 = \omega + \beta_1\sigma_{t-h}^2 + \alpha_1(\varepsilon_{t-h} - \gamma_1\sigma_{t-h})^2$$

where α_1 determines the kurtosis of the distribution and γ_1 allows for asymmetric effects of shocks. Specifically

$$Cov_{t-h}(\sigma^2(t+h), \log(X_t)) = -2\alpha_1\gamma_1\sigma^2(t).$$

In testing the empirical implications of this GARCH option pricing model, they find that the model produces smaller valuation errors compared to the *ad hoc* BS model mentioned above. In different work, Barone-Adesi, Engle and Mancini (2008) propose a new option pricing method with filtered historical innovations. The new feature in their methodology is that it fits in an incomplete markets framework and is not based on the specification of the change of measure, i.e. from physical measure to risk neutral measure and state price density.[45] Instead, they estimate separate GARCH parameters in the risk neutral world. They show that their pricing outperforms other discrete GARCH models.

In recent papers in the semi-parametric literature, as opposed to Black-Scholes, Carr and Madan (1998), Demeterfi et al. (1999), Britten-Jones and Neuberger (2000), Lynch and Panigirtzoglou (2003), Jiang and Tian (2005), Car and Wu (2009) develop variants of so called "model-free" implied volatility. These estimators are referred to as semiparametric measures, as volatility is implied from option prices via risk neutral pricing without many of the usual parametric assumptions on the dynamics of asset returns. In addition, these volatility measures provide ex ante risk neutral expec-

[45] For further discussion on risk neutral pricing and change of measures, see Duffie (2001), chapter 6.

tations of future volatilities, which is an important input to calculate variance risk premia. Variance risk premia are defined as the difference between implied volatility and realized volatility. When (log) stock prices follow a continuous process, the implied volatility between time t and $t+h$ can be derived by the formula.

$$IMV_{t,t+h} = 2\int_0^{\infty} \frac{C(t+h,K)-C(t,K)}{K^2}dK = E_t^Q(\sigma_{t+h}^2)$$

where $C(t,K)$ is the price of European call option written on strike price K and maturing at time t. Here, $E_t^Q(\sigma_{t+h}^2)$ is the expectation of the variation of the log price process, or of the realized volatility. An advantage to using the above model free implied volatility in equity markets is that one now can rely on a published volatility index (usually the VIX) as a standard measure of implied volatility on the S&P 500 index. VIX is considered a key measure of market expectations of near-term volatility implied from S&P 500 stock index option prices. VIX was first introduced by the Chicago Board of Exchange (CBOE) in 1993, and often referred to by many as a "fear" index. In 2003, CBOE updated the calculation of VIX and the general formula is as follows:

$$IMV = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2}e^{r\tau}Q(K_i) - \frac{1}{\tau}\left[\frac{F}{K_0}-1\right]^2$$

where $VIX = IMV * 100, \tau$ is time to expiration, F is the forward index level derived from index option prices, K_0 is the first strike price below F, K_i is the strike price of ith out of the money option, r is risk free rate, $\Delta K_i = (K_{i+1} - K_{i-1})/2$, $Q(K_i)$ is the mid point bid-ask spread for each option with strike price K_i.[46]

[46] For details, see "CBOE Volatility Index - VIX" at the link http://www.cboe.com/micro/vix/vixwhite.pdf and Demeterfi et al. (1999)

3.3.3 Realized Volatility - Nonparametric Measures

The latest developments in the volatility literature largely center on the use of so called realized volatility (RV) as a "model-free" estimator of latent variance of stock returns or other financial variables. Daily realized volatility is simply the sum squared returns from high-frequency data over short time interval within any given day. As noted in the key early works of Andersen and Bollerslev (1998), Banrdorff-Nielsen and Shephard (2002) and Medahhi (2002), RV and it's variants yield much more accurate ex post observations of volatility than the traditional sample variances which used daily or lower frequency data. Many papers have been written on this topic since these first papers. In practice, RV has been an important variable in the volatility derivatives market. For instance, trading of forward contracts on future realized volatility was sporadically seen in financial market as early as 1993. This type of product is now common, and is often referred to as the variance swap. One important feature of this product is that it's payoff is a linear function of RV and therefore is simple to use as a hedging tool, compared to traditional vega hedging. Moreover, there is much market demand for this product as practitioners prefer implied volatility to variance and they need additional instruments to hedge against future volatility risk. Other products that used realized volatility such as caps on variance swaps, corridor variance swaps, and options on realized volatility are also available in financial markets.[47] In research, several authors have developed the concept of variance risk premia which directly depend on RV and they argue that this variable is useful in asset pricing. Variance risk premium (VRP) is defined as

$$VRP_t = IMV_t - RV_t$$

[47] For a discussion of volatility and variance swaps, see for instance, Carr and Lee (2009).

where IMV_t is implied volatility defined under the risk neutral measure Q as

$$IMV_t = E^Q[\text{Return variation between } t-1 \text{ and } t]$$

Bollerslev, Tauchen and Zhou (2009) use this premium to predict stock market returns and they find that the premium is able to explain time-series variation in post-1990 aggregate stock market returns with high (low) premia predicting high (low) future returns.

RV non-parametrically measures the variation of return processes, and the dynamics of RV can be driven by components other than those directly involved with returns. When the return process is continuous, it's variation is due to the continuous component, and is known as the integrated volatility (IV). Realized volatility is a proxy for IV.[48] Several authors (for example, Huang and Tauchen (2005) and Aït-Sahalia and Jacod (2009a,b) find important evidence of active jumps in equity markets. If jumps occur, variation of the return process is greater than integrated volatility as it contains a jump variation component. Realized volatility therefore is not an estimator of integrated volatility. ABD (2007) construct a simple measure of the variation of this jump component and then show that incorporation of jumps can affect estimation and volatility prediction. (Our empirical application discussed below fits within this strand of the literature. In particular, we provide evidence of jumps and large jumps as well as providing a measure of large jump variation in equity markets.)

Following the general set-up of Aït-Sahalia and Jacod (2009a), consider the filtered probability space $(\Omega, F, (F_t)_{t\geq 0}, P)$, in which $(F_t)_{t\geq 0}$ is denoted as a filtration (i.e., a family of sub-sigma algebras F_t of F, being increasing $t : F_s \subset F_t$ if $s \leq t$). The log price process, $X_t = log(P_t)$, is assumed to be an Itô semimartingale process that can

[48] For instance, under the assumption that the return process is continuous, Kristensen (2010) develops a kernel based approach to estimate integrated and spot volatilty using realized volatility.

be written as:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s + \sum_{s \leq t} \Delta X_s, \tag{3.24}$$

where $X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s$ is the continuous semimartingale component of the process, which is the sum of a local martingale plus an adapted process with finite variation component. Additionally, ΔX_s is a jump at time s, defined as:

$$\Delta X_s = X_s - \lim_{\tau < s, \tau \rightarrow s} X_\tau.$$

Given this definition, the jump part of X_t in the time interval $[0, t]$ is defined to be $\sum_{s \leq t} \Delta X_s$. Note that when the jump is a Compound Poisson Process (CPP) - i.e. a finite activity jump process - then it can be expressed as:

$$J_t = \sum_{s \leq t} \Delta X_s = \sum_{i=1}^{N_t} Y_i,$$

where N_t is number of jumps in $[0, t]$, N_t follows a Poisson process, and the Y_i's are i.i.d. and are the sizes of the jumps. The CCP assumption has been widely used in the literature on modeling, forecasting, and testing for jumps. However, as discussed above, recent evidence suggests that processes may contain infinite activity jumps - i.e. infinite tiny jumps that look similar to continuous movements. In such cases, the CCP assumption is clearly violated, and hence we draw in such cases on the theory of Jacod (2008) and Aït-Sahalia and Jacod (2009a) when applying standard BNS (2006) type jump tests in the sequel. The empirical evidence discussed in this chapter involves examining the structure of the jump component of the log return process, X_t, using one historically observed price sample path $\{X_0, X_{\Delta_n}, X_{2\Delta_n} ... X_{n\Delta_n}\}$, where Δ_n is deterministic. The increment of the process at time $i\Delta_n$ is denoted by:

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

For convenience, we consider the case $t = n\Delta_n$ in the sequel.[49] In general, integrated volatility and quadratic variation are formally defined as:

$$IV_t = \int_0^t \sigma_s^2 ds = [\text{variation due to continuous component}]_t$$

$$QV_t = [X, X]_t = \int_0^t \sigma_s^2 ds + \sum_{t-1 \leq s < t} \Delta X_s^2 = IV_t + [\text{variation due to jump component}]_t$$

Volatility processes QV_t and IV_t are not observable. However, one could exploit high or ultra high frequency data in financial markets to estimate these variables. If the process is continuous, $IV_t = QV_t$ and their noisy estimators, hereby referred to as realized measures (RM) could be written as

$$RM_{t,n} = IV_t + N_{t,n}$$

where $N_{t,n}$ denotes the measurement error associated with the realized measure $RM_{t,n}$. There are two sources of measurement errors. One is due to the so-called microstructure noise effect of high frequency data, and the second is due to standard noise. There are a few realized measures and methods proposed to alleviate the effect of the contaminated high frequency noise. For example, Corradi, Distaso and Swanson (2009, 2011) derive consistent estimators of predictive conditional densities of integrated volatility using these noisy realized measures. They show that by choosing an appropriate realized measure, one can achieve consistent estimation, even in the presence of jumps and microstructure noise in prices. They thereby construct conditional predictive densities and confidence intervals for integrated volatility using realized measures, which may be of interest to volatility derivatives traders. Note that as microstructure noise is not the focus of our chapter, we will focus mainly on three key realized measures that are commonly used, i.e. realized volatility, bipower, and

[49] See Jacod (2008) for further details.

tripower variation (many papers in the extant literature now look also at multipower variation). For a list of other realized measures that are robust to microstructure noise, see for instance Corradi, Distaso and Swanson (2009, 2010). The realized volatility of equity return process X is defined as follows:

$$RV_{t,n} = \sum_{i=1}^{n} (\Delta_i^n X)^2$$

It has been shown that when n is large, realized volatility converges almost surely to the quadratic variation of the process.[50] If we measure volatility within a day then,

$$RV_{t,n} = \sum_{i=1}^{n} (\Delta_i^n X)^2 \simeq \int_{t-1}^{t} \sigma_s^2 ds + [\text{jump variation between } day \text{ day } t-1 \text{ and } t]_t$$

where n here is used to denote the number of incremental returns within a day or any other fixed time period. Multipower variations are constructed on the basis of

$$V_{r_1,r_2...,r_j} = \sum_{i=j+1}^{n} |\Delta_i^n X|^{r_1} |\Delta_{i-1}^n X|^{r_2} .... |\Delta_{i-j}^n X|^{r_j}.$$

where $r_{1,}r_{2,}..., r_j$ are positive, such that $\sum_1^j r_i = k$. Bipower (BV) and tripower variation (TP) are special cases of multipower variations. Specifically, Bipower variation is given by

$$BV_{t,n} = (\mu_1)^{-1} \sum_{i=2}^{n} V_{1,1} \simeq \int_{t-1}^{t} \sigma_s^2 ds$$

where $\mu_1 = E|Z| = 2^{1/2}\Gamma(1)/\Gamma(1/2)$ and Z is a standard normal random variable; and tripower variation is given by

$$TV_{t,n} = V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}} \mu_{\frac{2}{3}}^{-3} \simeq \int_{t-1}^{t} \sigma_s^2 ds$$

where $\mu_r = E(|Z|^r)$and Z is a $N(0,1)$ random variable. Finally, to illustrate the nuts

[50] See Barndorff-Nielsen and Shephard (2002).

and bolts of microstructure robust realized measures, we include the formula for a commonly used subsample based realized volatility measure, $\widehat{RV}_{t,l,M}$, suggested by Zhang, Mykland and Ait-Sahalia (2005), and defined as

$$\widehat{RV}_{t,l,n} = RV^{avg}_{t,l,n} - 2l\widehat{v}_{t,n},$$

where

$$\widehat{v}_{t,n} = \frac{RV_{t,l,n}}{2n} = \frac{1}{2n}\sum_{j=1}^{n}(\Delta_i^n X)^2$$

and

$$RV^{avg}_{t,l,n} = \frac{1}{B}\sum_{b=1}^{B} RV^b_{t,l} = \frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{t-1}(X_{t+\frac{jB+b}{n}} - X_{t+\frac{(j-1)B+b}{n}})^2$$

Here, $Bl \cong n, l = O(n^{1/3})$

Forecasting Realized Volatility With the availability of high frequency data in recent years, much effort has been dedicated to building good models to forecast realized volatility, a "model free" estimator of ex post variance. To see the link of this type of forecast to the volatility forecasting in discrete time models, note that the daily volatility implied by a discrete time model, let's say GARCH, is equivalent to $\sqrt{QV_t}$ or $\sqrt{IV_t}$ (without jumps) in continuous time,[51] which are proxied by $\sqrt{RV_{t,n}}$ or $\sqrt{BV_{t,n}}$ or $\sqrt{TV_{t,n}}$. The link justifies the rationale of this type of forecast in the literature. We highlight several important papers in this area of research. The key paper is Andersen, Bollerslev, Diebold, Laby (2003), who show empirically that a long memory Gaussian VAR can capture the dynamics of volatility. They apply a simple trivariate VAR (VAR-RV) to model the dynamics of volatilities of logarithmic exchange rates, i.e. DM/\$, Yen/\$ and Yen/DM. In particular, denote y_t as the vector

[51] To see this, in the case of continuous process, $Var(X_t - X_{t-1}) = E(\int_{t-1}^{t}\sigma_s dB_s)^2 = \int_{t-1}^{t}\sigma_s^2 ds$.

In addition, for convenience in notation, in this section, $\sqrt{QV_t}$ or $\sqrt{IV_t}$ are refered to as volatilities in a day.

.

of the the exchange rates, and define the forecasting equation as

$$\Phi(L)(1-L)^d(y_t-\mu)=\varepsilon_t$$

where ε_t is a vector white noise process. The authors point out that this simple framework performs better than many alternative models that have been used in this literature. In the same regression model, Corsi (2003) proposes the so called HAR - RV model in which realized volatility is specified as a linear sum of the lagged realized volatilities over different horizons, i.e.

$$RV_{t,t+h}=\beta_0+\beta_d RV_t+\beta_w RV_{t,t-5}+\beta_m RV_{t,t-22}+\epsilon_{t+h}$$

where h is the forecasting horizon, i.e. $h=1,5,22$. and $RV_{t,t+h}=h^{-1}[RV_{t+1}+RV_{t+2}+...+RV_{t+h}]$. ABD (2007) generalizes HAR - RV to linear and nonlinear HAR-RV, HAR-RV-J and HAR-RV-CJ. In particular, the new feature of the model that they propose involves incorporating the variation of jump components of the log price process into their forecasts. The advantage of these models is that they are rather simple to implement via least squares estimation and they take advantage of recent developments in the construction of robust jump tests. In the next section, we summarize robust jump testing and discuss the quantification of large jumps and small jumps, a departure that can potentially help improve the model's forecasting performance.

3.4 Volatility and Jumps

Thus far, we have summarized important developments in volatility models, with focus on time varying and stochastic volatility as well as nonparametric volatility estimators. All of our models share the same feature that volatilities are unobserved, or belong to class of missing variables. We now turn to a discussion of jumps, testing for jumps, and disentangling the effects of jumps from measures of volatility.

This section also contains the results of our empirical analysis of jumps and volatility. See Duong and Swanson (2010) for derivations of the tests and measures discussed and used in this section.

3.4.1 Testing for Jumps

In this section, we review some theoretical results relating to testing for jumps, namely testing whether $J_t = \sum_{s\leq t} \Delta X_s \neq 0$. In pioneering work, BNS (2006) proposes a robust and simple test for a class of Brownian Itô Semimartingales plus Compound Poisson jumps. In recent work, Aït-Sahalia and Jacod (2009a) among others develop a different test which applies to a large class of Itô-semimartingales, and allows the log price process to contain infinite activity jumps - small jumps with infinite concentrations around 0. In this chapter, we follow the jump test methodology of Huang and Tauchen (2005) as well as Barndorff-Nielsen and Shephard (2006), which looks at the difference between the continuous component and total quadratic variation in order to test for jumps. However, we make use of the limit theorems developed and used by Jacod (2008) and Aït-Sahalia and Jacod (2009a) in order to implement the Barndorff -Nielsen and Shephard (2006) type test under the presence of both infinite activity and finite activity jumps (see Section 3.4 for further discussion). A simplified version of the results of the above authors applied to (4.31) for the one-dimensional case is as follows. If the process X is continuous, let $f(x) = x^n$ (exponential growth), let ρ_{σ_s} be the law $N(0, \sigma_s^2)$, and let $\rho_{\sigma_s}(f)$ be the integral of f with respect to this law. Then:

$$\sqrt{\frac{1}{\Delta_n}}\left(\Delta_n \sum_{i=1}^{n} f(\frac{\Delta_i^n X}{\sqrt{\Delta_n}})^2 - \int_0^t \rho_{\sigma_s}(f)ds\right) \xrightarrow{L-S} \int_0^t \sqrt{\rho_{\sigma_s}(f^2) - \rho_{\sigma_s}^2(f)}dB_s \qquad (3.25)$$

Here, $L-S$ denotes stable convergence in law, which also implies convergence in distribution. For $n=2$, the above result is the same as BNS (2006). More generally:

$$\sqrt{\frac{1}{\Delta_n}}\left(\sum_{i=1}^{n}(\Delta_i^n X)^2-\int_0^t \sigma_s^2 ds\right)\xrightarrow{D} N(0,\int_0^t \vartheta\sigma_s^4 ds) \tag{3.26}$$

or

$$\frac{\sqrt{\frac{1}{\Delta_n}}\left(\sum_{i=1}^{n}(\Delta_i^n X)^2-\int_0^t \sigma_s^2 ds\right)}{\sqrt{\int_0^t \vartheta\sigma_s^4 ds}}\xrightarrow{D} N(0,1), \tag{3.27}$$

where ϑ is constant and where $\int_0^t \sigma_s^2 dt$ is known as the integrated volatility or the variation of the continuous component of the model and $\int_0^t \sigma_s^2 dt$ is integrated quarticity. From the above result, notice that if the process does not have jumps, then $\sum_{i=1}^{n}(\Delta_i^n X)^2$,which is an approximation of quadratic variation of the process, should be "close" to the integrated volatility. This is the key idea underlying the BNS (2006) jump test. For appropriate central limit theorems, in tests with both finite and infinite activity jumps, refer to Barndorff-Nielsen, Graverson, Jacod, Podolskij, and Shephard (2006), in the case of continuous semimartingales and Barndorff-Nielsen, Shephard, and Winkel (2006) for discontinuous process with Lévy jumps. A final crucial issue in this jump test is the estimation of $\int_0^t \sigma_s^2 dt$ and $\int_0^t \sigma_s^4 dt$ in the presence of both finite and infinite activity jumps. As remarked in BNS (2006), in order to ensure that tests have power under the alternative, integrated volatility and integrated quarticity estimators should be consistent under the presence of jumps. The authors note that robustness to jumps is straightforward so long as there are a finite number of jumps, or in cases where the jump component model is a Lévy or non-Gaussian OU model (Barndorff-Nielsen, Shephard, and Winkel (2006)). Moreover, under infinite activity jumps, note that as pointed out in Jacod (2007), there are available limit results for volatility and quarticity estimators for the case of semimartingales with jumps.

Turning again to our discussion of volatility and quarticity, note that in a continuation of work initiated by Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen, Graverson, Jacod, Podolskij, and Shephard (2006) and Jacod (2007) develop general so-called multipower variation estimators of $\int_0^t \sigma_s^k ds$, in the case of continuous semimartingales and semimartingales with jumps, respectively, which are based on

$$V_{r_1,r_2...,r_j} = \sum_{i=j+1}^{n} |\Delta_i^n X|^{r_1} |\Delta_{i-1}^n X|^{r_2} .... |\Delta_{i-j}^n X|^{r_j}.$$

where $r_{1,}r_{2,}..., r_j$ are positive, such that $\sum_1^j r_i = k$. For cases where $k = 2$ and $k = 4$, BNS (2006) use $V_{1,1}$ (bipower variation) and $V_{1,1,1,1}$. In our jump test implementation, we use $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}$(tripower variation) and $V_{\frac{4}{5},\frac{4}{5},\frac{4}{5}}$. The reason we use tripower variation, $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}$, instead of bipower variation, $V_{1,1}$, is that it is more robust to clustered jumps. Denote the estimators of $\int_0^t \sigma_s^2 ds$ and $\int_0^t \sigma_s^4 ds$ to be $\widehat{IV}$ and $\widehat{IQ}$, and note that:

$$\widehat{IV} = V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}} \mu_{\frac{2}{3}}^{-3} \simeq \int_0^t \sigma_s^2 ds \tag{3.28}$$

and

$$\widehat{IQ} = \Delta_n^{-1} V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}} \mu_{\frac{4}{5}}^{-5} \simeq \int_0^t \sigma_s^4 ds, \tag{3.29}$$

where $\mu_r = E(|Z|^r)$and Z is a $N(0,1)$ random variable.

Regardless of the estimator that is used, the appropriate test hypotheses are:

H_0 : X_t is a continuous process

H_1 : the negation of H_0 (there are jumps)

If we use multi-power variation, under the null hypothesis the test statistic which directly follows from the CLT mentioned above is:

$$LS_{jump} = \frac{\sqrt{\frac{1}{\Delta_n}}\left(\sum_{i=1}^{n}(\Delta_i^n X)^2 - \widehat{IV}\right)}{\sqrt{\vartheta \widehat{IQ}}} \xrightarrow{D} N(0,1)$$

and the so-called jump ratio test statistic is:

$$RS_{jump} = \frac{\sqrt{\frac{1}{\Delta_n}}}{\sqrt{\vartheta \widehat{IQ}/(\widehat{IV})^2}}\left(1 - \frac{\widehat{IV}}{\sum_{i=1}^{n}(\Delta_i^n X)^2}\right) \xrightarrow{D} N(0,1).$$

Of note is that an adjusted jump ratio statistic has been shown by extensive Monte Carlo experimentation in Huang and Tauchen (2005), in the case of CCP jumps, to perform better than the two above statistics, being more robust to jump over-detection. This adjusted jump ratio statistic is:

$$AJ_{jump} = \frac{\sqrt{\frac{1}{\Delta_n}}}{\sqrt{\vartheta max(t^{-1}, \widehat{IQ}/(\widehat{IV})^2)}}\left(1 - \frac{\widehat{IV}}{\sum_{i=1}^{n}(\Delta_i^n X)^2}\right) \xrightarrow{L} N(0,1) \qquad (3.30)$$

In general if we denote the daily test statistics to be $Z_{t,n}(\alpha)$,where n is the number of observations per day and α is the test significance level,[52] then we reject the null hypothesis if $Z_{t,n}(\alpha)$ is in excess of the critical value Φ_α, leading to a conclusion that there are jumps. The converse holds if $Z_{t,n}(\alpha)$ is less than Φ_α. In our empirical application, $Z_{t,n}(\alpha)$ is the adjusted jump ratio statistic, and we calculate the percentage of days that have jumps, for the period from 1993 to 2008. We now turn to a discussion of large jumps and constructing measures of the daily variation due to continuous and jump components.

[52] i.e., $\Delta_n = 1/n$

3.4.2 Large Jumps and Small Jumps

There is now clear evidence that jumps are prevalent in equity market. For example, Huang and Tauchen (2005) construct the above jump test statistics, and find that jumps contribute about 7% to the total variation of daily stock returns. Aït-Sahalia and Jacod (2009b) not only find jumps but given the existence of jumps, they look more deeply into the structure of the jumps, and for Intel and Microsoft returns they find evidence of the existence of infinite activity jumps.

An important focus in our chapter is to the decomposition of jumps into "large" and "small" components so that we may assess their contributions to the overall variation of the price process. In particular, for some fixed level γ, define large and small jump components as follows, respectively:

$$LJ_t(\gamma) = \sum_{s \leq t} \Delta X_s I_{|\Delta X_s| \geq \gamma}.$$

and

$$SJ_t(\gamma) = \sum_{s \leq t} \Delta X_s I_{|\Delta X_s| < \gamma}.$$

The choice of γ may be data driven, but in this chapter we are more concerned with scenarios where there is some prior knowledge concerning the magnitude of γ. For example, under various regulatory settings, capital reserving and allocation decisions may be based to a large extent on the probability of jumps or shocks occurring that are of a magnitude greater than some known value, γ. In such cases, planners may be interested not only in knowledge of jumps of magnitude greater than γ, but also in characterizing the nature of the variation associated with such large jumps. The procedure discussed in this section can readily be applied to uncover this sort of information.

3.4.3 Realized measures of daily jump variation

The partitioning of variation due to continuous and jump components can be done, for example, using truncation based estimators which have been developed by Mancini (2001,2004,2009) and Jacod (2008). One can also simply split quadratic variation into continuous and jump components by combining various measures of integrated volatility, such as bipower or tripower variation and realized volatility. Andersen, Bollerslev, and Diebold (2007) do this, and construct measures of the variation of the daily jump component as well as the continuous component. In this chapter we use their method, but apply it to both small and large jumps. In particular, once jumps are detected, the following risk measures introduced by Andersen et al. (2007) are constructed:

$VJ_t =$ *Variation of the jump component* $= max\{0, RV_t - \widehat{IV_t}\} * I_{jump,t}$

$VC_t =$ *Variation of continuous component* $= RV_t - VJ_t$,

where $RV_t = \sum_{i=1}^{n}(\Delta_i^n X)^2$ is the daily realized volatility (i.e. a measure of the variation of the entire (log) stock return process), I_{jump} is an indicator taking the value 0 if there are no jumps and 1 otherwise, and n is the number of intra-daily observations. One can then calculate daily jump risk. Note that in these formulae, the variation of the continuous component has been adjusted (i.e. the variation of the continuous component equals realized volatility if there are no jumps and equals $\widehat{IV_t}$ if there are jumps). In addition, note that $\sum_{i=1}^{n}(\Delta_i^n X)^2 I_{|\Delta_i^n X| \geq \gamma}$ converges uniformly in probability to $\sum_{s \leq t}(\Delta X_s)^2 I_{|\Delta X_s| \geq \gamma}$, as n goes to infinity.[53] Thus, the contribution of the variation of jumps with magnitude larger than γ and smaller than γ are denoted and calculated as follows:

Realized measure of large jump variation: $VLJ_{t,\gamma}$=min$\{VJ_t, (\sum_{i=1}^{n}(\Delta_i^n X)^2 I_{|\Delta_i^n X| \geq \gamma} * I_{jump,t})\}$,

Realized measure of small jump variation: $VSJ_t, \gamma = VJ_t - VLJ_{t,\gamma}$,

[53] See Jacod (2008), Aït-Sahalia and Jacod (2011) for further details.

where I_{jump} is defined above and $I_{jump,\gamma}$ is an indicator taking the value 1 if there are large jumps and 0 otherwise. This condition simply implies that large jump risk is positive if the process has jumps and has jumps with magnitude greater than γ.

Now we can write the relative contribution of the variation of the different jump components to total variation in a variety of ways:

Relative contribution of continuous component $= \frac{VC_t}{RV_t}$

Relative contribution of jump component $= \frac{VJ_t}{RV_t}$

Relative contribution of large jump component $= \frac{VLJ_{t,\gamma}}{RV_t}$

Relative contribution of small jump component $= \frac{VLS_{t,\gamma}}{RV_t}$

Relative contribution of large jumps to jump variation $= \frac{VLJ_{t,\gamma}}{VJ_t}$

Relative contribution of small jumps to jump variation $= \frac{VLS_{t,\gamma}}{VJ_t}$

We are now ready to discuss some empirical findings based on the application of the tools discussed in this section.

3.5 Empirical Findings

3.5.1 Data description

We use a large tick by tick data set of 25 DOW 30 stocks available for the period 1993-2008. The data source is the TAQ database. We use only 25 stocks because we purge our data set of those stocks that not frequently traded or are not available across the entire sample period. For the market index, we follow several other papers and look at S&P futures. We also follow the common practice in the literature of eliminating from the sample those days with infrequent trades (less than 60 transactions at our 5 minute frequency).

One problem in data handling involves determining the method to filter out an evenly-spaced sample. In the literature, two methods are often applied - *previous tick* filtering and *interpolation* (Dacorogna, Gencay, Müller, Olsen, and Pictet (2001)).

As shown in Hansen and Lund (2006), in applications using quadratic variation, the *interpolation* method should not be used, as it leads to realized volatility with value 0 (see *Lemma 3* in their paper). Therefore, we use the *previous tick* method (i.e. choosing the last price observed during any interval). We restrict our dataset to regular time (i.e. 9:30am to 4:00pm) and ignore ad hoc transactions outside of this time interval. To reduce microstructure effects, the suggested sampling frequency in the literature is from 5 minutes to 30 minutes[54]. As mentioned above, we choose the 5 minute frequency, yielding a maximum of 78 observations per day.

3.5.2 Jump and Large Jump Results

We implement our analysis in two stages. In the first stage we test for jumps and in second stage we examine large jump properties, in cases where evidence of jumps is found. The list of the companies for which we examine asset returns is given in Table 3.1, along with a summary of our jump test findings. The rest of the tables and figures summarize the results of our empirical investigation. Before discussing our findings, however, we briefly provide some details about the calculations that we have carried out.

All daily statistics are calculated using the formulae in Section 3.4 with:

$$\Delta_n = \frac{1}{n} = \frac{1}{\text{\# of 5 minute transactions / day}}$$

Therefore, $\Delta_n = 1/78$ for most of the stocks in the sample, except during various shortened and otherwise nonstandard days, and except for some infrequently traded stocks. This also implies the choice of time to be the interval $[0, 1]$, where the time from $[0, 1]$ represents the standardizing time with beginning (9 am) set to 0 and end (4.30 pm) set to 1. In our calculations of estimates of integrated volatility and integrated quarticity, we use multipower variation, as given in (4.45) and (4.46).

[54] See Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005)

Recall also that $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ is simply the incremental return of $X_{i\Delta_n}$. For any trading day, X_0 and X_1 correspond to the first and the last observations of the day. Denote T as the number of days in the sample. We construct the time series $\{Z_{t,n}(\alpha)\}_{t=0}^{T}$ and $\left\{\frac{VC_t}{RV_t}, \frac{VJ_t}{RV_t}, \frac{VLJ_{t,\gamma}}{RV_t}, \frac{VSJ_{t,\gamma}}{RV_t}\right\}_{t=0}^{T}$. The number of days and proportion of days identified as containing jumps can easily be calculated as:

Number of days identified as jumps $= \sum_{i=0}^{T} I(Z_{i,n}(\alpha) > \Phi_\alpha)$.

Proportion of days identified as jumps $= \frac{\sum_{i=0}^{T} I(Z_{t,n}(\alpha)>\Phi_\alpha)}{T}$.

In addition, we construct the following monthly time series

Proportion of days identified as jumps in a month$= \frac{\sum_{i=m}^{m+h} I(Z_{t,n}(\alpha)>\Phi_\alpha)}{h}$

Monthly average relative contribution of jump component $= \frac{\sum_{i=m}^{m+h} \frac{VJi}{RVi}}{h}$

Monthly average relative contribution of large jump component truncated at level $\gamma = \frac{\sum_{i=m}^{m+h} \frac{VLJ_{i,\gamma}}{RV_t}}{h}$,

where m is the starting date and h is the number of days in each month. On average, there are 22 business days per month. Note that there are 12 statistics each year for each time series.

Here, $I(\cdot)$ denotes the indicator function. The average relative contribution of continuous, jump, and large jump components to the variation of the process is reported using the mean of the sample (i.e. we report the means of $\frac{VC_t}{RV_t}$, $\frac{VJ_t}{RV_t}$, $\frac{VLJ_{t,\gamma}}{RV_t}$, and$\frac{VSJ_{t,\gamma}}{RV_t}$).

In addition to reporting findings based on examination of the entire sample period, we also split the sample into two periods. The first period is from 1993 to 2000 and the second period is from 2001 to 2008. The reason for doing this is that we would like to see whether jump activity changes over time. Moreover, these subsamples correspond roughly to break dates for financial data found in Cai and Swanson (2011).

In the sequel, we provide figures for representative individual stocks in our sample (i.e. Walmart, IBM, Bank of America and Citigroup). These stocks are chosen on the basis of their market systematic risk beta. Namely, Walmart has low beta of around 0.3, IBM has a beta close to 0.7, and Bank of America and Citigroup are more risky stocks with betas of around 2.6 and 2.8.

Turning now to our results, a first sense of the prevalence of jumps can be formulated by inspecting Panels A,B, C and D of Figure 3.1, where statistics higher than 3.09 (i.e. the 0.001 significance level critical value) are presented for the entire sample from 1993 to 2008. It is obvious that jumps are prevalent. Additionally, it should be noted that there is a marked difference in jump frequency between 1993-2000 and 2001-2008, where the first period is much more densely populated with jumps than the latter period. The highest statistic values are around 11, for Walmart in 1997, 11 for IBM in 1994, 10 for Bank of America in 1996 and 7 for Citigroup from 1996 to 1998. Post 2000, the highest statistics are consistently located in 2002 and 2006-2008. Moreover, a simple visual check of the statistic magnitudes in this figure suggests that jumps are more prevalent in the earlier sample period, with respect to both frequency and significance level (more will be said on this later).

Regarding our choice of the large jumps, an important step is to choose truncation levels, γ. If we choose arbitrarily large truncation levels, then clearly we will not find evidence of large jumps. Also one may easily proceed by just picking the truncation level based on the percentiles of the entire historical sample of the 5 minute log return. However, results could then turn out to be difficult to interpret, as in one case the usual choice of $90th$ or $75th$ percentiles leads to virtually no large jumps while the choice $25th$ or $10th$ percentiles leads to a very large number of large jumps. In addition, "large" jumps are often thought of as abnormal events that arise at a frequency of one in several months or even years. Therefore, a reasonable way to proceed is to pick the truncation level on the basis of the sample of the monthly maximum increments - monthly based abnormal events. Specifically, we set three levels $\gamma = 1, 2, 3$ to be the $50th, 75th$, and $90th$ percentiles of the entire sample from 1993 to 2008. Panels A,B,C, and D of Figure 3.2 depict the monthly largest absolute increments and the jump truncation levels used in our calculations of the variation of large and small jump components. Again, it is quite obvious that the monthly maximum increments are dominant in the previous decade. The larger monthly increments in current decades are mostly located in 2006-2008 and 2002-2004. As a result, the fixed truncation levels which are chosen across the entire sample result in more "hits" in previous decade than in the current one. The truncation level of Citigroup is the largest of the four stocks depicted (for example at $\gamma = 3$ the level is approximately 0.04 for Citigroup and 0.025 for IBM).

Notice that the graphs in Figures 3.3A and 3.3B depict magnitudes of the variation of continuous, jump, and truncated jump components of returns for our 4 sample stocks. Namely, the plots are of daily realized volatility, and realized variance of continuous, jump and large jump components at different truncation levels. As might be expected, inspection of the graphs suggests a close linkage between the greater number of jumps in the first decade of the sample and the and large jump risk over the same period. For example, in the case of IBM, the variation of the jump components is clearly dominant in the earlier decade. The highest daily jump risk occurs in late 1998, and is above 0.018. Indeed, at jump truncation level 3, we only see large jump risk for the years 1994, 1996, 1998, 2000 and 2008. Combined with the results of Figure 3.1, this again strongly suggests that there was much more turbulence in the earlier decade.

Turning now to our tabulated results, first recall that Table 3.1 reports the proportion of days identified as having jumps, at 6 different significant levels, $\alpha = \{0.1, 0.05, 0.01, 0.005, 0.001, 0.0001\}$. Again, there is clear evidence of jumps in both periods. However, the jump frequency in the 1993-2000 sample is significantly higher than that in the 2001-2008 sample, across all stocks and test significance levels. For example, at the $\alpha = 0.005$ and 0.001 levels, the average daily jump frequencies are 46.9% and 22.8% during the 1993-2000 period, as compared with 16.8% and 9.4% during the 2001-2008 period, respectively. When considering individual stocks, the story is much the same. As illustrated in Figures 3.1, and tabulated in Table 3.1, the proportion of "jump-days" for IBM and for the Bank of America are 5.9% and 8.8% during the 2000s, which is much smaller than the value of 19.2% and 21.3% for the two stocks during the 1990s, based on tests implemented using a significance level of $\alpha = 0.001$.

Of course, when calculating jump frequencies, we ignore the magnitudes of the jumps. Table 3.2 addresses this issue by summarizing another measure of jumps - namely the average percentage contribution of jumps to daily realized variance. Details of the measures reported are given above and in Section 3.4. In support of our earlier findings, it turns out that jumps account for about 15.6% and 8.1% of total variation at significance levels $\alpha = 0.005$ and 0.001, respectively, when considering the entire sample period from 1993-2008. Moreover, analogous statistics for the period 1993-2000 are 25.1% and 12.7%, while those for the 2001-2008 period are 7% and 5%. The statistics for IBM and Bank of America are 25.3% and 10.7% for the period 1993-2000 and 3.5% and 2.3% for the period 2001-2008 while those for the entire samples are 7.9% and 6.6%. This result is consistent with our earlier findings through figure analysis.

In summary, without examining the impact of large jumps, we already have evidence that: (i) There is clear evidence that jumps characterize the structure of the returns of all of the stocks that we examine. (ii) The 1990s are characterized by the occurrence of more jumps than the 2000s. (iii) The contribution of jumps to daily realized variance is substantively higher during the 1990s than the 2000s. (iv) Our results are consistent across all stocks, suggesting the importance of jump risk comovement during turbulence periods.

In our empirical analysis of large jumps, we carry out the same steps as those employed above when examining overall jump activity. Results are reported in Tables 3.3A-C are for truncation levels $\gamma = 1, 2, 3$ at 6 different significant levels, $\alpha =$

$\{0.1, 0.05, 0.01, 0.005, 0.001, 0.0001\}$. As mentioned earlier, Figures 3.1 and 3.3 contain plots of jump test statistics and realized variation not only for overall jump activity, as discussed above, but also for large jumps. Examination of these tables suggest a number of conclusions.

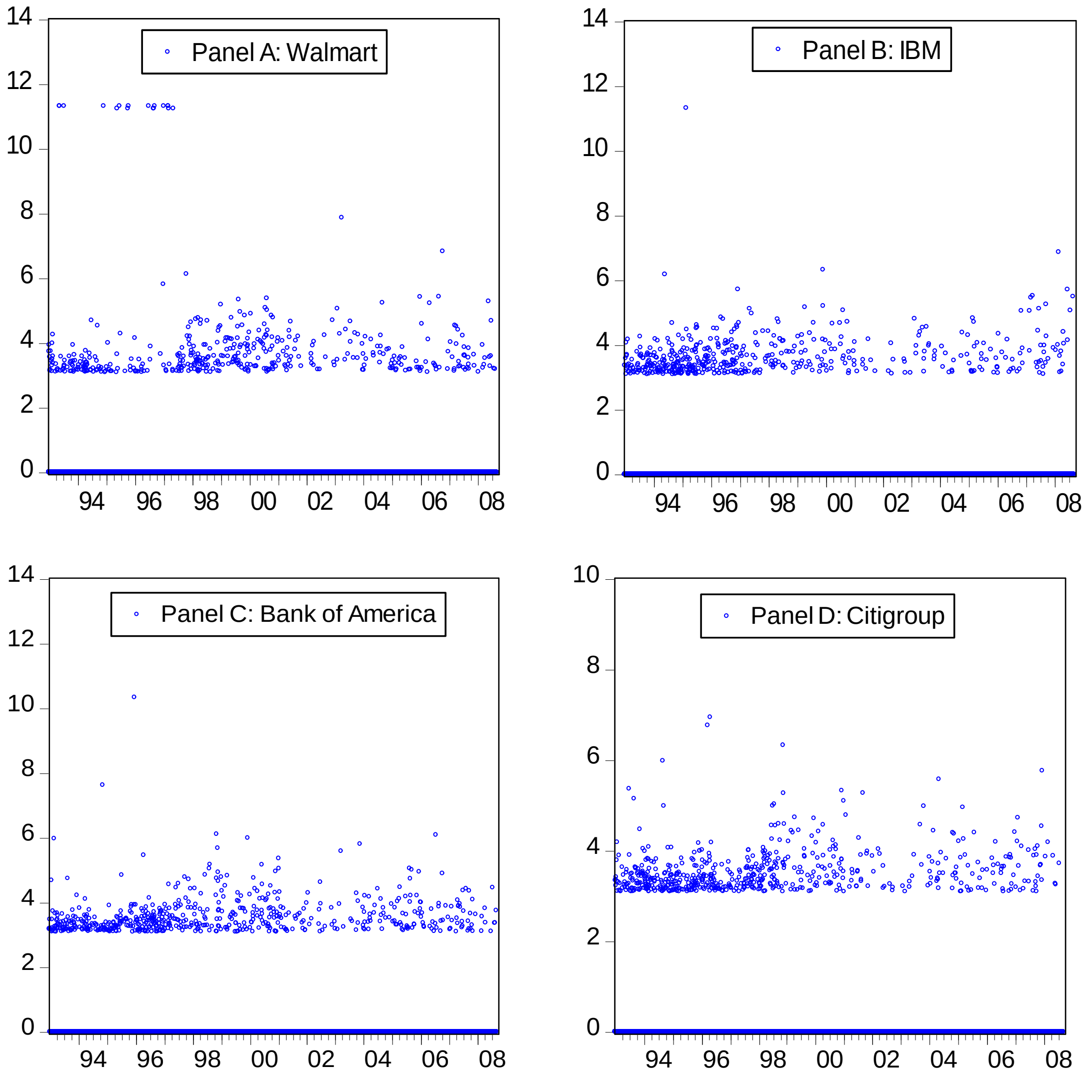
Across the entire sample, there is evidence of large jumps at all levels by measure of variation. Table 3.3A reports the proportion of days identified as having large jumps for truncation level $\gamma = 1$. It can be seen that the proportion of variation due to large jumps at truncation level $\gamma = 1$ accounts for about 0.9% and 0.6% of total variation (regardless of stock), at significance levels $\alpha = 0.005$ and 0.001, respectively. Values at significance level 0.001 for the periods 1993-2000 and 2001-2008 are around 0.8% and 0.4%, respectively. For $\gamma = 2$, values are 0.4% and 0.3% at significant levels $\alpha = 0.005$ and 0.001, respectively, when considering the entire sample. Values at significance level 0.001 for the periods 1993-2000 and 2001-2008 are around 0.4% and 0.2% for period 1993-2000 and 2001-2008, respectively. A similar result obtains for $\gamma = 3$, suggesting that large jump variability is around twice as big (as a proportion of total variability) for the latter sub-sample, regardless of truncation level. As previously, these results are surprisingly stable across stocks. Although not included here, our analysis of the market index data discussed above yielded a similar result. Further examination of the statistics in the Tables 3.3A-C also yields another interesting finding. In particular, though proportions of jumps and large jumps at truncation level $\gamma = 1, 2, 3$ are all larger in the previous decade, the difference is smaller and increasingly narrows as higher truncation levels are considered, when examining large jumps. This result, which is true for many of our stocks, suggests an increased role of large jumps in explaining daily realized variance during the latter sub-sample. To illustrate this point, which is apparent upon inspection of average statistics constructed for all 25 stocks, we investigate the case of of ExxonMobil, where we look at all statistics at significance level $\alpha = 0.001$. The proportion of variation of jumps to total variation is 17% for the period 1993-2000 (as shown in Table 3.2), almost 3 times as much as the corresponding value of 6.2% in 2001-2008. However for large jumps at truncation level $\gamma = 1$, the analogous value is 0.6% for 1993-2008, which is just 1.5 times as much as the 0.4% value during 2001-2008. Similarly at truncation level $\gamma = 2$, the value is 0.4% for 1993-2008 and 0.2% for 2001-2008. Interestingly, at truncation level $\gamma = 3$, the proportion of variation of jumps is 0 for period 1993-2000 while it is 0.1% for period 2001-2008. Therefore, with respect to large jump we find that: (i) Large jumps incidence and magnitudes are consistent with our earlier finding that the 1990s are much more turbulent than the 2000s. (ii) However, for higher truncation levels, the contribution of jump risk during the two periods becomes much closer, and indeed the contribution during the latter period can actually become marginally greater. This suggests that while the overall role of jumps is lessening, the role of large jumps has not decreased, and indeed, the relative role of large jumps has actually increased in the 2000s.

3.6 Concluding Remarks

In this chapter we review some of the recent literature on volatility modelling and jumps, with emphasis on the notion that these variables are unobserved latent variables, and thus can be viewed in some sense as "missing data". Many estimators of volatility, both continuous and discrete, as well as both parametric and nonparameteric are also reviewed.

In an empirical investigation, we provide new evidence of jumps in individual log price processes, and note that there are clearly comovements during turbulent times, for all stocks. More noticeably, jump incidence is greater during the 1990s than during the 2000s, although the incidence of "very large" jumps is similar across both decades, and the relative importance of large jumps has increased.

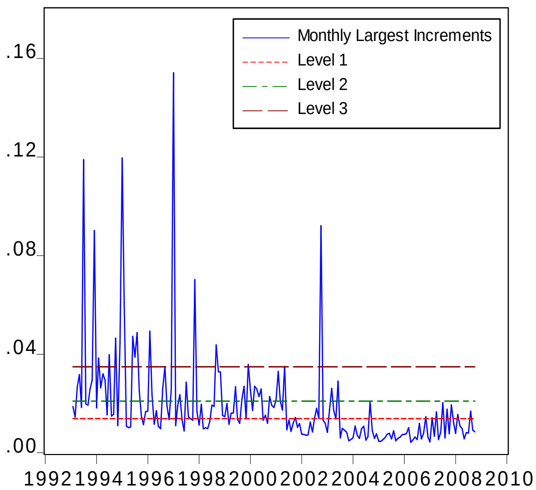
Figure 3.1: Jump Test Statistics of Days Identified as Having Jumps of (Log) Stock Prices: Sample Period 1993-2008 *



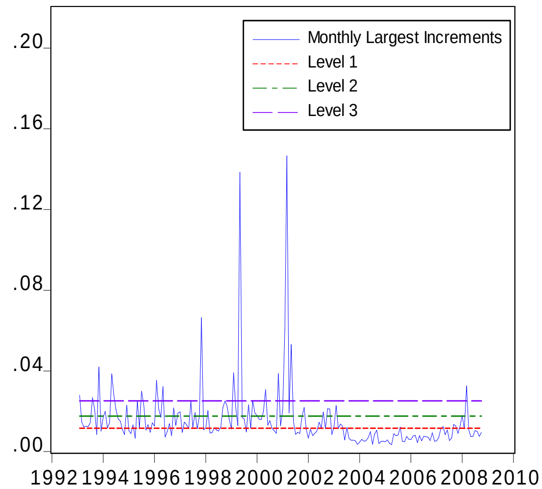
* Panel A, B, C, D depict daily test statistics of days identified as having jumps for Walmart, IBM, Bank of America, Citigroup (Log) Stock Price using 0.001 significant level. Specifically, all statistics in the figure are larger than 3.09. See section 3.5 for further details.

Figure 3.2: Monthly Largest Increments and Truncation Levels $\gamma = 1, 2, 3$: Sample Period 1993-2008 *

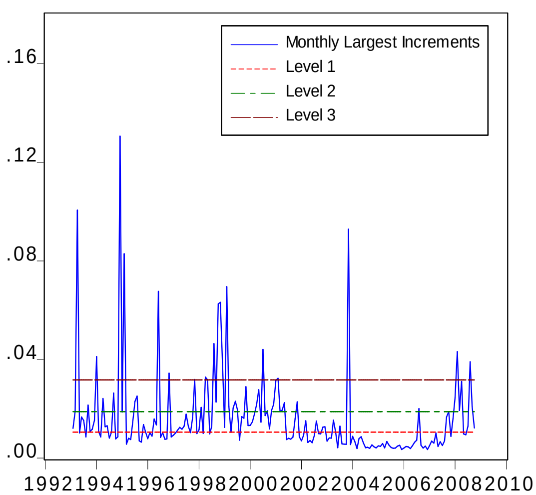
Panel A: Walmart



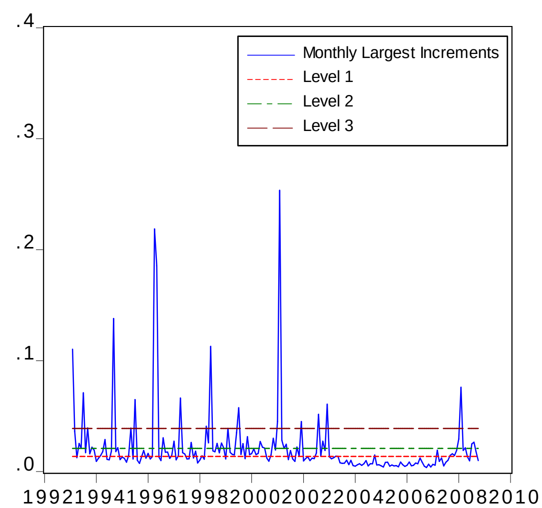
Panel B: IBM



Panel C: Bank of America



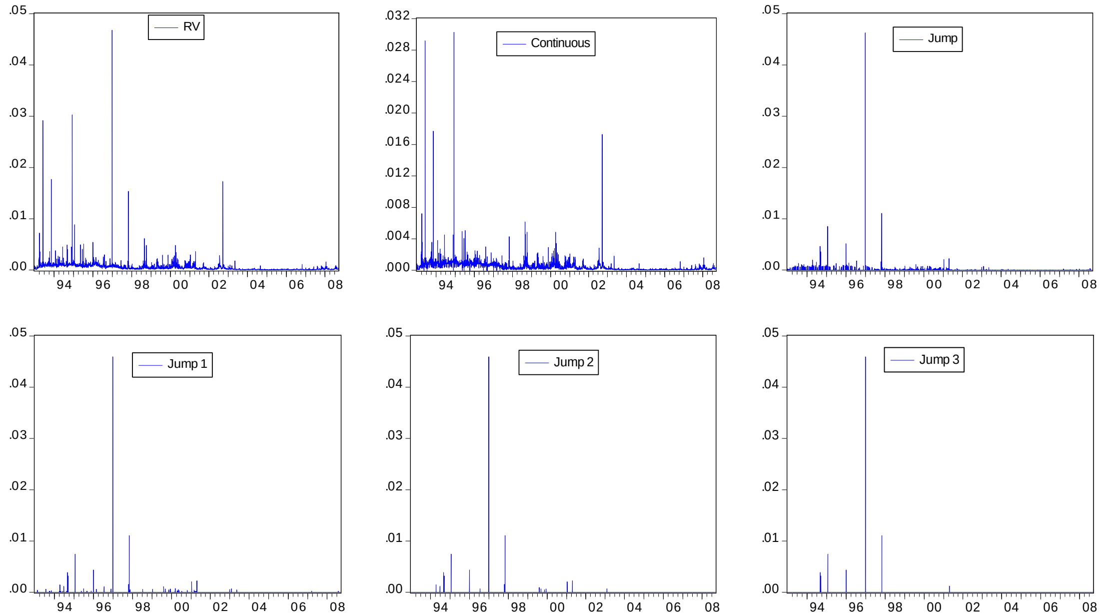
Panel D: Citigroup



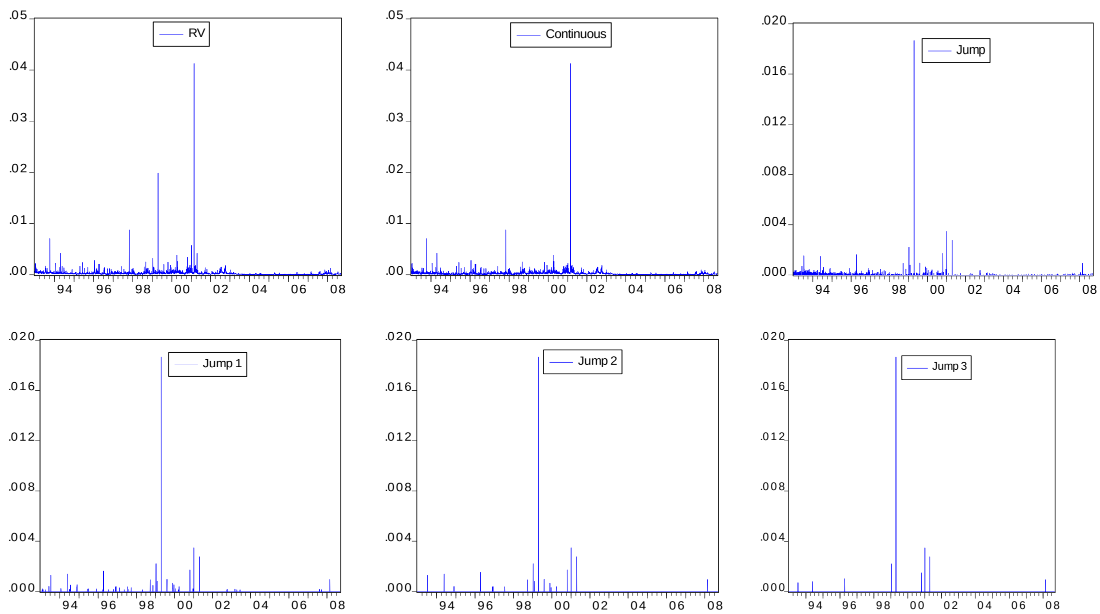
* Panel A,B,C,D depict the monthly largest absolute increments and the jump truncation levels used as thresholds in our calculations of the variations of large and small jump components, where level γ=1 corresponds to the median of monthly maximum increments , level γ=2 corresponds to 75th percentile of monthly maximum increments, and level γ=3 corresponds to 90th percentile monthly maximum increments of(log) stock prices of Walmart, IBM, Bank of America and Citigroup for the sample period is from 1993 to 2008.

Figure 3.3A: Daily Realized Volatility (RV) and Realized Variation of Continuous, Jump and Truncated Jump Components (Log) Stock Prices for Truncation Levels $\gamma = 1, 2, 3$ *

Panel A: Walmart



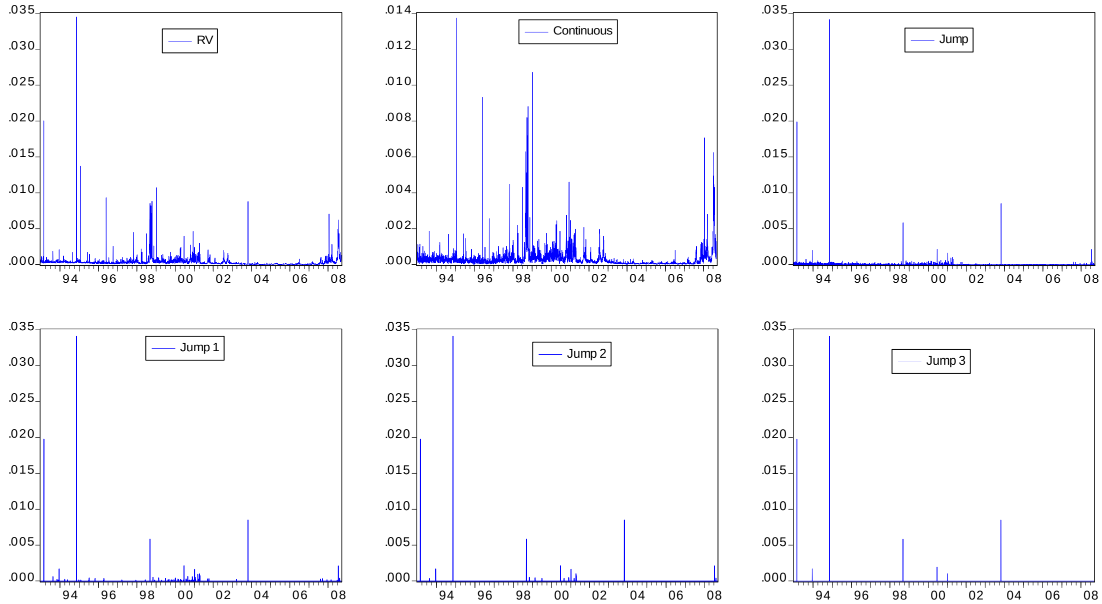
Panel B: IBM



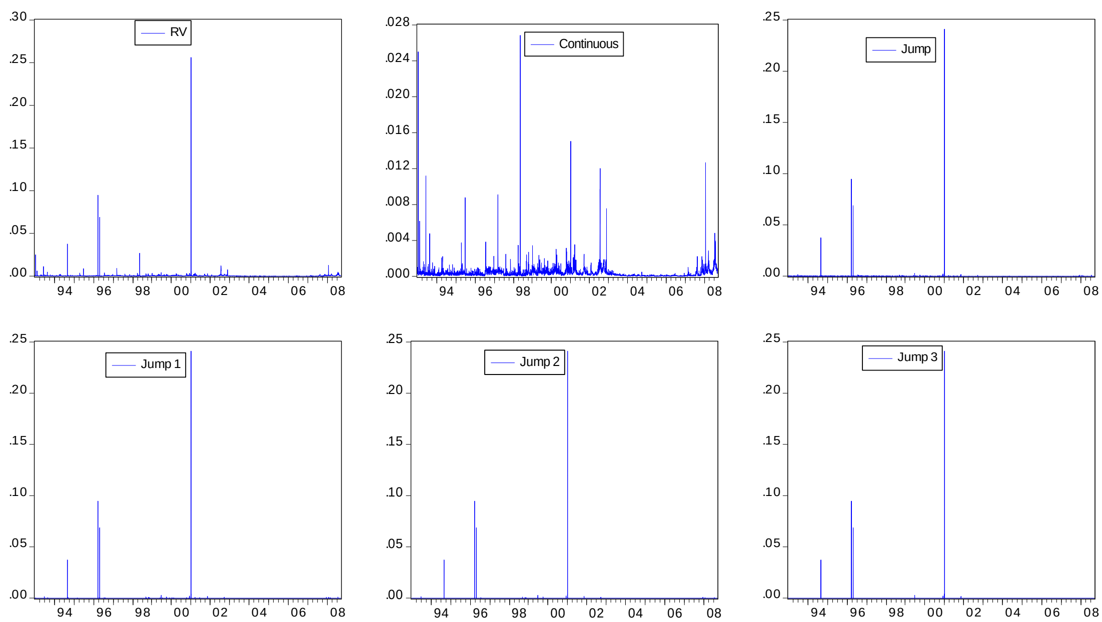
* See Figure 3.2 for details about the jump truncation levels. The above panels plot daily realized volatility, realized measures of the variation of continuous, jump and large jump components at truncation levels $\gamma = 1, 2, 3$, which are shortly referred to as jump 1, jump 2 and jump 3 for the period 1993-2008. The realized measures of variations are calculated as discussed in Section 3.4 and 3.5.

Figure 3.3B: Daily Realized Volatility (RV) and Realized Variation of Continuous, Jump and Truncated Jump Components of (Log) Stock Prices for Truncation Levels $\gamma = 1, 2, 3$ *

Panel C: Bank of America



Panel D: Citigroup



* See notes in Figure 3.3A.

Table 3.1: Percentage of Days Identified as Having Jumps Using Daily Statistics *

Stock Name	*Panel A: Sample Period 1993-2000 ($T \simeq 2000$)*						*Panel B: Sample Period 2001-2008 ($T \simeq 1900$)*					
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	88	80.9	62.1	52	26.3	8.7	49.2	39.3	21.4	16.5	9.7	4.3
American Express	82.9	76.2	56.4	46.6	19.3	4.6	47.4	36.8	20.2	14.7	8.0	4.0
Bank of America	81.8	75.6	56.7	46.4	21.3	5.0	45.1	34.1	20.1	15.7	8.8	3.1
Citigroup	86.3	80.5	63.3	51.9	23.3	4.9	43.6	32.9	17.9	14.6	7.1	2.6
Caterpillar	87.2	81.5	61.8	51	25.9	7.3	46	35.3	19.9	16.3	9.5	4.3
Dupont	83.8	76.5	57.2	48.3	24.2	5.6	49.5	38.8	21.8	17.1	9.5	3.9
Walt Disney	89.3	83.9	65.9	56.0	27.3	5.3	55.6	43.9	23.9	17.6	10.1	3.9
General Electric	79.6	73.5	54.5	45.5	22.3	4.5	49.2	39.3	21.8	16.2	9.4	3.9
GM	88.1	83.1	65.4	54	25.4	6.2	51.8	40.4	22.8	17.8	10.5	4.7
Home Depot	87.7	81	62.1	51.4	24.6	5.1	49.5	38.5	22.1	16.8	10	4.3
IBM	73.8	65	47.3	39.6	19.2	5.9	39.9	30.1	15.1	11.7	5.9	2.8
Intel	69.2	58.9	39.5	33.0	18.0	6.3	51.7	41.4	23.6	18.7	11.3	4.7
Johnson & Johnson	86.7	81.2	62.8	52.5	25.2	5.7	47.5	37.7	22.1	18.0	10.9	4.6
JPM	79.5	73.2	55.7	47.6	21.4	5.0	47.9	35.9	20.8	16.1	9.0	3.3
Coca Cola	86.4	80.8	63.3	54.2	23.9	4.8	52.5	41.9	23.3	18.5	10.2	4.6
McDonald's	90.5	85	66.1	55.9	25.8	4.9	51.3	40.8	24.6	19.8	11.5	4.8
3M	85.7	78.8	59.2	49.9	25.6	6.9	43.1	33.1	18.8	14.2	7.9	3.6
Microsoft	68.5	58.7	38.6	30.5	16.4	7.0	56.3	44.8	25.7	21.5	11.1	4.4
Pfizer	82.6	75.4	56.6	49.1	26.3	6.5	50	40	23.5	17.7	9.4	4.1
Procter & Gamble	80.1	72.4	55.6	46.4	25.5	6.4	46.9	35.6	18.5	14.4	7.2	2.8
AT & T	89.3	83.3	65.8	54.7	23.1	4.4	58.8	48.4	29	22.8	13.8	6.1
United Tech.Corp.	84.2	77.1	54.3	43.9	22.8	8.2	46.3	36.3	20.5	16.0	9.1	3.6
Verizon	81.5	67.7	46	39.5	24.2	8.1	51.4	40.9	24.5	19.4	11.2	5.0
Walmart	86.7	81.5	59.8	46.9	15.5	5.1	44.7	34.3	18.7	14.0	7.4	2.6
ExxonMobil	61.3	49.8	32.8	26.2	17	5.2	44.2	33.6	17.5	12.9	6.2	2.9
Average	82.4	75.3	56.4	46.9	22.8	5.9	48.8	38.2	21.5	16.8	9.4	4.0

* See notes to Figure 3.1. Entries denote the percentage of days identified as having jumps based on the calculation of daily statistics. Statistics are the adjusted ratio jump statistics of Barndorff-Nielsen and Shephard (2006) and Huang and Tauchen (2005), as discussed in Section 3.4. Test results are summarized in Panel A for the sample period from 1993-2000 and for the sample period 2001-2008 in Panel B. These sample periods have approximately 2000 and 1900 daily statistics, respectively. Statistics are reported for six different significance levels, $\alpha = 0.1, 0.05, 0.01, 0.005, 0.001, 0.0001$.

Table 3.2: Daily Realized Variation: Ratio of Jump to Total Variation *

Stock Name	*Panel A: Sample Period 1993-2000* ($T \simeq 2000$)						*Panel B: Sample Period 2001-2008* ($T \simeq 1900$)					
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	41.2	39.5	32.6	28.4	14.8	5.2	13.6	11.8	7.7	6.4	4.1	2.1
American Express	39.5	38.1	30.5	26.2	10.8	2.6	12.7	10.9	7.1	5.7	3.5	1.9
Bank of America	37.2	36.0	29.7	25.3	12.1	2.9	11.6	9.8	6.8	5.6	3.5	1.5
Citigroup	41.5	40.3	34.1	29.2	13.7	3.0	11.0	9.3	6.0	5.1	2.8	1.2
Caterpillar	40.0	38.6	32.0	27.5	14.7	4.5	12.3	10.5	7.1	6.2	4.0	2.0
Dupont	36.8	35.4	29.1	25.8	13.3	3.2	13.3	11.5	7.6	6.4	4.0	1.8
Walt Disney	42.0	40.8	34.9	30.7	15.8	3.4	15.3	13.2	8.6	6.9	4.2	1.8
General Electric	34.0	32.8	27.1	23.8	12.3	2.5	13.2	11.6	7.7	6.1	3.8	1.8
GM	42.5	41.4	35.2	30.2	14.7	3.9	14.6	12.5	8.3	6.9	4.4	2.2
Home Depot	40.0	38.6	32.3	27.9	14.1	3.1	13.3	11.5	7.8	6.3	4.1	2.0
IBM	30.1	28.3	23.1	20.4	10.7	3.5	9.8	8.2	5.1	4.2	2.3	1.3
Intel	24.0	22.0	16.8	14.7	8.9	3.5	13.9	12.2	8.3	6.9	4.6	2.2
Johnson &Johnson	39.2	38.1	32.3	28.3	14.2	3.4	13.0	11.3	7.8	6.7	4.4	2.1
JPM	36.0	34.7	29.2	25.6	12.0	2.8	12.6	10.6	7.2	5.9	3.7	1.6
Coca Cola	39.9	38.8	33.1	29.4	13.7	2.8	14.0	12.3	8.2	6.8	4.2	2.1
McDonald's	43.9	42.8	36.0	31.5	15.1	3.1	14.9	13.1	9.2	7.8	5.0	2.3
3M	39.0	37.5	30.9	27.2	14.4	4.1	11.0	9.4	6.2	5.1	3.2	1.6
Microsoft	23.2	21.4	16.1	13.5	8.0	3.7	15.1	13.2	9.0	7.8	4.5	2.0
Pfizer	35.5	34.1	28.2	25.4	14.7	3.9	13.8	12.1	8.3	6.7	3.9	1.9
Procter &Gamble	33.9	32.4	27.8	24.5	14.2	3.7	11.9	10.1	6.3	5.2	3.0	1.4
AT &T	43.8	42.6	36.0	31.1	13.5	2.8	17.7	15.9	11.1	9.3	6.1	3.0
United Tech.Corp.	37.1	35.5	27.6	23.0	12.4	4.8	12.2	10.5	7.0	5.8	3.6	1.6
Verizon	29.4	26.8	21.0	18.8	12.0	4.4	14.4	12.6	8.7	7.4	4.7	2.4
Walmart	44.4	43.2	33.8	27.4	9.1	3.4	11.5	9.8	6.3	5.1	3.0	1.2
ExxonMobil	19.3	17.1	12.7	10.9	7.6	2.6	10.8	9.1	5.7	4.6	2.5	1.3
Average	36.5	35.1	28.9	25.1	12.7	3.5	13.1	11.3	7.6	6.3	3.9	1.9

Panel C: Sample Period 1993-2008 ($T \simeq 3900$)

Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	26.7	25.0	19.5	16.9	9.2	3.6
American Express	26.4	24.8	19.1	16.2	7.2	2.3
Bank of America	24.7	23.2	18.5	15.7	7.9	2.2
Citigroup	26.6	25.1	20.3	17.4	8.3	2.1
Caterpillar	26.2	24.7	19.6	16.9	9.4	3.3
Dupont	25.4	23.7	18.6	16.3	8.8	2.5
Walt Disney	29.0	27.3	22.0	19.1	10.2	2.6
General Electric	23.8	22.5	17.6	15.1	8.2	2.2
GM	28.9	27.3	22.1	18.8	9.7	3.1
Home Depot	27.0	25.3	20.3	17.4	9.2	2.5
IBM	20.2	18.5	14.3	12.5	6.6	2.4
Intel	19.0	17.2	12.6	10.9	6.8	2.9
Johnson &Johnson	26.4	25.0	20.3	17.7	9.4	2.8
JPM	24.6	23.0	18.4	16.0	7.9	2.2
Coca Cola	27.3	25.9	20.9	18.4	9.0	2.5
McDonald's	30.4	29.0	23.5	20.5	10.4	2.7
3M	26.0	24.4	19.4	16.9	9.2	2.9
Microsoft	19.4	17.6	12.8	10.8	6.3	2.9
Pfizer	24.9	23.3	18.5	16.3	9.4	2.9
Procter &Gamble	23.2	21.5	17.3	15.1	8.7	2.6
AT &T	31.2	29.6	23.9	20.5	9.9	2.9
United Tech.Corp.	23.7	22.1	16.5	13.8	7.7	3.1
Verizon	15.3	13.5	9.5	8.1	5.1	2.5
Walmart	28.3	26.9	20.4	16.5	6.1	2.3
ExxonMobil	11.9	10.1	6.6	5.4	3.1	1.5
Average	24.7	23.1	18.1	15.6	8.1	2.6

* See notes to Figure 3.2. The entries in the table denote the average percentage of daily variation of the jump component relative to daily realized variance for the sample periods 1993-2000, 2001-2008 and 1993-2008. The realized measure of variation of the jump component is calculated as discussed in Section 3.4. In addition to frequency of jumps, realized measures of variations also take the magintude of jumps into account. Entries are caculcated accross 6 different significant levels, $\alpha = 0.1, 0.05, 0.01, 0.005, 0.001, 0.0001$.

Table 3.3A: Daily Realized Variation: Ratio of Large Jump to Total Variation, Jump Truncation Level $\gamma = 1$ *

Stock Name	*Panel A: Sample Period 1993-2000 ($T \simeq 2000$)*						*Panel B: Sample Period 2001-2008 ($T \simeq 1900$)*					
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	1.8	1.7	1.2	1.1	0.9	0.6	1.3	1.2	0.9	0.8	0.5	0.4
American Express	1.5	1.3	1.0	0.8	0.6	0.4	1.2	1.1	0.8	0.6	0.5	0.3
Bank of America	3.0	2.7	1.8	1.5	1.0	0.7	1.0	0.9	0.8	0.6	0.4	0.2
Citigroup	2.1	1.9	1.2	1.0	0.6	0.4	0.8	0.7	0.5	0.4	0.3	0.1
Caterpillar	2.3	2.2	1.6	1.5	1.0	0.6	0.7	0.6	0.5	0.5	0.3	0.2
Dupont	2.1	1.9	1.2	1.0	0.7	0.3	0.9	0.9	0.6	0.5	0.4	0.2
Walt Disney	2.1	1.8	1.1	0.9	0.6	0.4	1.6	1.4	1.0	0.9	0.6	0.3
General Electric	1.5	1.4	0.8	0.7	0.4	0.2	1.3	1.2	0.9	0.6	0.3	0.2
GM	1.5	1.4	1.0	0.8	0.6	0.4	1.3	1.2	0.8	0.7	0.5	0.2
Home Depot	1.9	1.7	1.3	1.1	0.6	0.3	0.7	0.6	0.5	0.3	0.2	0.1
IBM	2.3	2.1	1.7	1.6	1.0	0.7	0.5	0.5	0.4	0.4	0.2	0.1
Intel	2.3	2.1	1.6	1.2	0.8	0.5	0.7	0.7	0.4	0.4	0.3	0.2
Johnson &Johnson	2.2	2.0	1.5	1.3	0.9	0.6	0.6	0.6	0.4	0.3	0.2	0.1
JPM	1.3	1.1	0.7	0.6	0.3	0.2	1.7	1.5	1.0	0.8	0.6	0.3
Coca Cola	2.3	2.1	1.3	1.2	0.8	0.5	0.8	0.8	0.6	0.5	0.4	0.3
McDonald's	1.8	1.6	1.2	0.9	0.7	0.4	1.0	1.0	0.7	0.6	0.4	0.2
3M	2.1	2.0	1.3	1.1	0.7	0.5	0.6	0.6	0.4	0.4	0.4	0.3
Microsoft	2.9	2.7	2.0	1.7	1.0	0.5	0.6	0.6	0.5	0.4	0.2	0.1
Pfizer	2.0	1.9	1.3	1.1	0.8	0.5	0.7	0.6	0.5	0.4	0.3	0.2
Procter &Gamble	2.4	2.2	1.7	1.4	0.9	0.5	0.8	0.7	0.5	0.4	0.3	0.2
AT &T	2.3	2.2	1.6	1.3	0.9	0.7	1.7	1.5	1.2	1.1	0.7	0.4
United Tech.Corp.	3.2	2.9	2.1	1.9	1.2	0.6	1.3	1.1	0.8	0.7	0.5	0.2
Verizon	6.9	6.3	5.0	4.4	2.7	1.1	1.5	1.3	1.0	0.9	0.6	0.4
Walmart	2.7	2.4	1.4	1.2	0.9	0.6	0.6	0.6	0.4	0.4	0.3	0.1
ExxonMobil	2.1	1.8	1.3	1.0	0.6	0.4	1.1	0.9	0.7	0.6	0.4	0.2
Average	2.3	2.1	1.5	1.3	0.8	0.5	1.0	0.9	0.7	0.6	0.4	0.2

Panel C: Sample Period 1993-2008 ($T \simeq 3900$)						
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	1.6	1.4	1.0	0.9	0.7	0.5
American Express	1.3	1.2	0.9	0.7	0.5	0.4
Bank of America	2.0	1.9	1.3	1.1	0.7	0.5
Citigroup	1.4	1.3	0.8	0.7	0.5	0.3
Caterpillar	1.5	1.4	1.1	1.0	0.7	0.4
Dupont	1.5	1.4	0.9	0.8	0.5	0.2
Walt Disney	1.8	1.6	1.1	0.9	0.6	0.4
General Electric	1.4	1.3	0.8	0.7	0.4	0.2
GM	1.4	1.3	0.9	0.8	0.5	0.3
Home Depot	1.3	1.2	0.9	0.7	0.4	0.2
IBM	1.4	1.3	1.0	1.0	0.6	0.4
Intel	1.5	1.4	1.0	0.8	0.6	0.3
Johnson &Johnson	1.4	1.3	1.0	0.8	0.6	0.4
JPM	1.5	1.3	0.9	0.7	0.5	0.3
Coca Cola	1.6	1.5	1.0	0.8	0.6	0.4
McDonald's	1.4	1.3	1.0	0.8	0.5	0.3
3M	1.4	1.3	0.9	0.8	0.6	0.4
Microsoft	1.8	1.7	1.3	1.1	0.6	0.3
Pfizer	1.3	1.2	0.9	0.8	0.6	0.4
Procter &Gamble	1.6	1.5	1.1	0.9	0.6	0.3
AT &T	2.0	1.9	1.4	1.2	0.8	0.5
United Tech.Corp.	2.1	2.0	1.4	1.3	0.8	0.4
Verizon	1.8	1.6	1.2	1.1	0.7	0.4
Walmart	1.7	1.5	0.9	0.8	0.6	0.3
ExxonMobil	1.2	1.1	0.7	0.6	0.4	0.2
Average	1.6	1.4	1.0	0.9	0.6	0.3

* See notes to Figure 3.2. Entries in the table denote the average percentage of daily variation due to jumps constructed using truncation level $\gamma = 1$, relative to the daily realized variance, for the sample periods 1993-2000, 2001-2008 and 1993-2008. The realized measure of variation of the jump component is calculated as discussed in Section 3.4. Entries are caculcated accross 6 different significance levels (α =$0.1, 0.05, 0.01, 0.005, 0.001, 0.0001$).

Table 3.3B: Daily Realized Variation: Ratio of Large Jump to Total Variation, Jump Truncation Level $\gamma = 2$ *

Stock Name	*Panel A: Sample Period 1993-2000* ($T \simeq 2000$)						*Panel B: Sample Period 2001-2008* ($T \simeq 1900$)					
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	0.8	0.8	0.6	0.5	0.5	0.3	0.5	0.5	0.4	0.4	0.2	0.1
American Express	0.8	0.7	0.6	0.5	0.3	0.3	0.5	0.4	0.3	0.2	0.2	0.1
Bank of America	0.9	0.9	0.5	0.4	0.3	0.2	0.4	0.4	0.3	0.3	0.2	0.1
Citigroup	1.0	1.0	0.7	0.5	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.1
Caterpillar	1.0	1.0	0.8	0.8	0.5	0.3	0.3	0.3	0.3	0.3	0.2	0.2
Dupont	0.9	0.8	0.4	0.3	0.2	0.1	0.4	0.4	0.2	0.2	0.1	0.0
Walt Disney	1.0	0.9	0.4	0.3	0.3	0.2	0.5	0.5	0.3	0.3	0.2	0.1
General Electric	0.7	0.7	0.4	0.4	0.2	0.1	0.6	0.6	0.5	0.3	0.1	0.1
GM	0.7	0.7	0.4	0.3	0.2	0.2	0.7	0.6	0.4	0.4	0.3	0.1
Home Depot	1.0	0.9	0.7	0.6	0.3	0.2	0.2	0.2	0.2	0.1	0.0	0.0
IBM	1.0	0.9	0.8	0.7	0.5	0.3	0.2	0.2	0.2	0.2	0.1	0.1
Intel	0.9	0.9	0.7	0.6	0.4	0.2	0.2	0.2	0.1	0.1	0.1	0.0
Johnson &Johnson	0.9	0.8	0.6	0.5	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.1
JPM	0.4	0.4	0.2	0.2	0.1	0.0	0.7	0.7	0.5	0.4	0.3	0.2
Coca Cola	0.9	0.9	0.4	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.2	0.2
McDonald's	0.8	0.7	0.5	0.4	0.3	0.3	0.5	0.5	0.4	0.2	0.2	0.0
3M	1.0	0.9	0.5	0.4	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.1
Microsoft	1.1	1.0	0.8	0.6	0.3	0.1	0.2	0.2	0.2	0.2	0.1	0.0
Pfizer	0.8	0.8	0.6	0.6	0.4	0.3	0.3	0.3	0.2	0.2	0.1	0.1
Procter &Gamble	1.1	1.0	0.8	0.6	0.4	0.3	0.3	0.2	0.2	0.2	0.1	0.1
AT &T	1.1	1.0	0.7	0.6	0.4	0.4	0.7	0.6	0.6	0.5	0.3	0.1
United Tech.Corp.	1.1	1.0	0.7	0.6	0.5	0.3	0.4	0.4	0.3	0.3	0.2	0.1
Verizon	2.8	2.6	2.2	1.9	0.7	0.2	0.5	0.5	0.4	0.4	0.3	0.2
Walmart	1.2	1.1	0.6	0.5	0.5	0.4	0.2	0.2	0.2	0.2	0.1	0.0
ExxonMobil	0.8	0.6	0.5	0.4	0.4	0.2	0.5	0.5	0.4	0.3	0.2	0.1
Average	1.0	0.9	0.6	0.5	0.4	0.2	0.4	0.4	0.3	0.3	0.2	0.1

Panel C: Sample Period 1993-2008 ($T \simeq 3900$)

Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	0.6	0.6	0.5	0.4	0.3	0.2
American Express	0.7	0.6	0.4	0.3	0.3	0.2
Bank of America	0.7	0.6	0.4	0.4	0.3	0.2
Citigroup	0.7	0.7	0.5	0.4	0.3	0.2
Caterpillar	0.7	0.6	0.5	0.5	0.4	0.3
Dupont	0.7	0.6	0.3	0.3	0.2	0.1
Walt Disney	0.8	0.7	0.4	0.3	0.2	0.2
General Electric	0.7	0.6	0.5	0.3	0.1	0.1
GM	0.7	0.7	0.4	0.3	0.2	0.2
Home Depot	0.6	0.6	0.4	0.3	0.2	0.1
IBM	0.6	0.6	0.5	0.5	0.3	0.2
Intel	0.6	0.5	0.4	0.3	0.2	0.1
Johnson &Johnson	0.6	0.6	0.4	0.4	0.3	0.2
JPM	0.6	0.5	0.4	0.3	0.2	0.1
Coca Cola	0.6	0.6	0.3	0.3	0.2	0.2
McDonald's	0.7	0.6	0.5	0.3	0.3	0.2
3M	0.6	0.6	0.3	0.3	0.2	0.2
Microsoft	0.6	0.6	0.5	0.4	0.2	0.0
Pfizer	0.6	0.5	0.4	0.4	0.3	0.2
Procter &Gamble	0.7	0.6	0.5	0.4	0.3	0.2
AT &T	0.9	0.8	0.6	0.6	0.4	0.3
United Tech.Corp.	0.7	0.7	0.5	0.4	0.3	0.2
Verizon	0.7	0.6	0.5	0.4	0.3	0.2
Walmart	0.7	0.6	0.4	0.3	0.3	0.2
ExxonMobil	0.6	0.5	0.4	0.3	0.2	0.1
Average	0.7	0.6	0.4	0.4	0.3	0.2

* See notes to Table 3.3A.

Table 3.3C: Daily Realized Variation: Ratio of Large Jump to Total Variation, Jump Truncation Level $\gamma = 3$ *

Stock Name	*Panel A: Sample Period 1993-2000* ($T \simeq 2000$)						*Panel B: Sample Period 2001-2008* ($T \simeq 1900$)					
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	0.3	0.3	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.0
American Express	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.0
Bank of America	0.5	0.5	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
Citigroup	0.5	0.5	0.4	0.3	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.0
Caterpillar	0.5	0.5	0.4	0.4	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1
Dupont	0.4	0.4	0.2	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Walt Disney	0.5	0.5	0.2	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.1	0.0
General Electric	0.3	0.3	0.2	0.1	0.0	0.0	0.3	0.3	0.3	0.2	0.0	0.0
GM	0.3	0.3	0.2	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.0
Home Depot	0.4	0.4	0.3	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
IBM	0.3	0.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Intel	0.4	0.4	0.3	0.2	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Johnson &Johnson	0.4	0.4	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.0
JPM	0.2	0.2	0.1	0.1	0.0	0.0	0.2	0.2	0.2	0.2	0.1	0.1
Coca Cola	0.5	0.5	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1
McDonald's	0.4	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.0
3M	0.3	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
Microsoft	0.4	0.4	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0
Pfizer	0.3	0.3	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.0	0.0
Procter &Gamble	0.3	0.3	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.1	0.1
AT &T	0.6	0.5	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.0
United Tech.Corp.	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.1	0.0
Verizon	0.7	0.7	0.7	0.7	0.0	0.0	0.2	0.1	0.1	0.1	0.1	0.0
Walmart	0.5	0.5	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.0	0.0
ExxonMobil	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.1	0.0
Average	0.4	0.4	0.3	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.0

Panel C: Sample Period 1993-2008 ($T \simeq 3900$)						
Significant Level α	0.1	0.05	0.01	0.005	0.001	0.0001
Alcoa	0.3	0.3	0.2	0.2	0.1	0.1
American Express	0.3	0.2	0.2	0.1	0.1	0.1
Bank of America	0.3	0.3	0.2	0.2	0.1	0.1
Citigroup	0.3	0.3	0.2	0.2	0.2	0.1
Caterpillar	0.3	0.3	0.3	0.2	0.2	0.1
Dupont	0.3	0.3	0.1	0.0	0.0	0.0
Walt Disney	0.4	0.4	0.2	0.2	0.1	0.1
General Electric	0.3	0.3	0.2	0.2	0.0	0.0
GM	0.3	0.3	0.1	0.1	0.1	0.1
Home Depot	0.2	0.2	0.2	0.1	0.0	0.0
IBM	0.2	0.2	0.2	0.2	0.1	0.1
Intel	0.2	0.2	0.2	0.1	0.1	0.0
Johnson &Johnson	0.3	0.3	0.2	0.2	0.2	0.1
JPM	0.2	0.2	0.1	0.1	0.0	0.0
Coca Cola	0.3	0.3	0.1	0.1	0.1	0.1
McDonald's	0.3	0.3	0.2	0.1	0.1	0.1
3M	0.2	0.2	0.1	0.1	0.1	0.1
Microsoft	0.2	0.2	0.2	0.1	0.1	0.0
Pfizer	0.2	0.2	0.2	0.1	0.1	0.1
Procter &Gamble	0.2	0.2	0.2	0.2	0.1	0.1
AT &T	0.4	0.4	0.2	0.2	0.2	0.1
United Tech.Corp.	0.2	0.2	0.2	0.2	0.1	0.1
Verizon	0.2	0.2	0.1	0.1	0.1	0.0
Walmart	0.3	0.3	0.2	0.2	0.1	0.1
ExxonMobil	0.2	0.2	0.2	0.2	0.1	0.0
Average	0.3	0.3	0.2	0.1	0.1	0.1

* See notes to Table 3.3A.

Chapter 4

Volatility Predictability and Jump Asymmetry

4.1 Introduction

Many recent modelling advances in asset pricing and management are predicated on the importance of jumps, or discontinuous movements in asset returns. In an important paper, Huang and Tauchen (2005) find evidence of discrete large jumps in S&P cash and future (log) returns from 1997 to 2002, in approximately 7% of the trading days. Aït-Sahalia and Jacod (2009b) develop methods to ascertain whether the process describing an asset contains "infinite activity jumps" - those jumps that are tiny and look similar to continuous movements, but whose contribution to the jump risk of the process is not negligible. In an empirical analysis of Intel and Microsoft returns, they find evidence of the presence of infinite active jumps in historical data. In summary, it is now generally accepted that many return processes contain jumps.[55] Once jumps are found, the economic implications of including them in dynamic asset pricing exercises are substantial. For example, the incorporation of jumps lead to break-downs in the typical market completeness condition needed for portfolio replication strategy in derivatives valuation. Additionally, jumps complicate the implementation of the "state of the art" change of risk measure in risk neutral pricing. As a result, asset allocation and risk management, which heavily depend on risk measures and underlying asset return dynamics, are affected. In volatility measurement, it is necessary to separate out the volatility due to jumps or construct variables that appropriately summarize information generated by jumps.

In volatility forecasting, once jumps are detected, understanding the role of variables that capture jump information is potentially important for applied practition-

[55] For other examples of work in this area, see Aït-Sahalia (2002), Carr et al., (2002), Carr and Wu (2003), Barndorff-Nielsen and Shephard (BNS: 2006), Woerner (2006), Jacod (2008), Jiang and Oomen (2008), Lee and Mykland (2008), Tauchen and Todorov (2009), Aït-Sahalia and Jacod (2009a,b) and the references cited therein.

ers, especially in the construction of hedging strategies.[56] In general, volatility predictability is important in numerous areas ranging from the pricing of volatility-based derivative products to asset management. In light of this, a number of recent papers have addressed volatility predictability, some from the perspective of the usefulness of jumps in forecasting volatility. However, although there is strong evidence of the importance of jumps in pricing, investment and risk management, there is mixed evidence concerning whether information extracted from jumps is useful for volatility forecasting. In a seminal work, Andersen, Bollerslev and Diebold (ABD: 2007) show that almost all of the predictability in daily, weekly, and monthly return volatilities comes from the non-jump component for DM/$ exchange rate, the S&P500 market index, and the 30-year U.S. Treasury bond yield. Corsi, Pirino and Reno (2008) find that jumps are positively correlated with, and have a significant impact on future volatility of the S&P500 index, various individual stocks and US bond yields. Patton and Shephard (2011) point out that the impact of a jump on future volatility critically depends on the sign of the jump, for both the S&P 500 index, as well as 105 individual stocks. In this chapter we add to the empirical literature on this topic by providing results on volatility forecasting using a variety of "new" variables that capture information generated by jumps.

When undertaking empirical research using volatility, a key issue involves the choice of the volatility estimator. One approach involves "backing out" volatility from parametric from ARCH, GARCH, Stochastic Volatility, or Option pricing models. The approach that we adopt involves using recently developed "model free" estimators (see the influential work of Andersen, Bollerslev, Diebold and Laby (2001)), including realized volatility (RV), and variants thereof such as bipower variation, tripower variation, multipower variation, semivariance, and various others.[57] One key reason for the use of these "model free" realized measures (RMs), is that they allow us to treat volatility as if it is observed, when we fit regressions in order to assess jump predictability. Modeling and forecasting RMs are important not only because RMs are a natural proxy for volatility, but also because of the many practical appli-

[56] See Andersen, Bollerslev and Diebold (2007) and Aït-Sahalia and Jacod (2011) for further discussion.

[57] See e.g., Barndorff-Nielsen and Shephard (2004), Aït-Sahalia, Mykland and Zhang, (2005), Zhang (2006), Barndorf-Nielsen, Hansen, Lunde, and Shephard (2006,2008), Jacod (2008), Barndoff, Kinnebrock, Shephard (2010), and the references cited therein.

cations and uses of RMs in constructing synthetic measures of risk in the financial markets. For example, since shortly after the inception in 1993 of the VIX (index of implied volatility), a variety of volatility-based derivative products have been engineered using RV as an input. These include variance swaps, caps on variance swaps, corridor variance swaps, covariance swaps, options on RV overshooters, and up and downcrossers. The key here is that investors worry about future volatility risk, and hence often opt for this type of contract in order to hedge against risk, adding to the traditional volatility "Vega".[58] In light of the above uses of RV, several authors have advocated forecasting RV (and more generally RMs) using extensions of ARMA models (see e.g., Andersen, Bollerslev, Diebold and Labys (2003), Corsi (2004), and ABD (2007)). In related work, Corradi, Distaso and Swanson (2012) develop model -free conditional predictive density estimators and confidence intervals for integrated volatility.

Given the availability of volatility estimators, as discussed above, it remains to choose variables that capture information generated by jumps. In this chapter, we examine four realized measures of jump power variations, all formed on the basis of power transformation of the instantaneous return (i.e., $|r_t|^q$). The analysis of power transformations of returns is not new. Ding, Granger and Engle (1993) and Ding and Granger (1996) develop long memory Asymmetric Power ARCH models based on power transformations of daily absolute returns. They find that the autocorrelations of power transformations of S&P 500 returns are the strongest for $q < 1$. In the context of high frequency data, Liu and Maheu (LH: 2005) and Ghysels and Sohn (GS: 2009) study the predictability of future realized volatility using past absolute power variations and multipower variations. GS (2009) find that the optimal value of q is approximately unity. However, their empirical evidence considers the continuous class of models, and does not account for jumps. Andersen, Bollerslev and Diebold (ABD: 2007), on the other hand, develop an interesting framework for separating jump and continuous components of RV, and carry out predictability experiments indicating that jumps play a small but notable role in forecasting volatility. In related recent

[58] Volatility and variance swaps are newer hedging instruments, adding to the traditional volatility "Vega", which is derived from options data. See Hull and White (1997, pp. 328) for a definition of Vega. For example, as noted in Carr and Lee (2008), the UBS book was short many millions of vega in 1993, and they were the first to use variance swaps and options on realized volatility to hedge against volatility risk. See Duong and Swanson (2011) for further discussion.

work, Barndoff, Kinnebrock, and Shephard (BKS: 2010) construct new estimators of downside (and upside) risk (i.e., so-called realized semivariances), using square transformations of positive and negative intra-daily return, and find that downside risk measures are important when attempting to model and understand risk. They note, as quoted from Granger (2008), that: *'It was understood that risk relates to an unfortunate event occurring, so for an investment this corresponds to a low, or even negative, return. Thus getting returns in the lower tail of the return distribution constitutes this "downside risk." However, it is not easy to get a simple measure of this risk.'* This point is noteworthy, since it is argued in the literature (see e.g., Ang, Chen and Xing (2006)), that investors treat downside losses differently than upside gains. As a result, agents who put higher weight on downside risk demand additional compensation for holding stocks with high sensitivity to downside market movements. Most authors in this literature pay attention to co-skewness as a measure of downside risk, and use daily data for estimation thereof. Patton and Shephard (2011) build on these ideas and use semivariance estimators to forecast volatility.

Building on the work of above authors, and in particular BKS (2010), we contribute to the volatility prediction literature by examining recently proposed realized measures of (downside) jump power variations. The measures are constructed using power transformations of absolute intra-daily returns, based on recent limit theory advances due to Jacod (2008) and BKS (2010). Theoretically, the measures do not require the use of a jump test in order to "pre-test" for jumps. Although construction of the measures is closely related to the work of Ghysels and Sohn (2009), our approach differs in that we focus on jump power variations with $q > 2$. Furthermore, the limit theory that we adopt allows us to construct estimators of downside and upside jump power variations using intra-daily positive and negative returns. These estimators are suggested by BKS (2010) as alternatives to the semivariances implemented in Patton and Shephard (2011). We also examine jump asymmetry (i.e., realized signed jump power variation) in realized volatility prediction experiments. Of note is that the role of the size of jumps that are most useful for forecasting can be inferred through examination of the order of q. For this reason, we consider jump power variations with $0 \leq q \leq 6$. While previous authors have focused on $q \leq 2$,allowing for a wider range of values for q is sensible, given that convergence to jump power variation occurs only when $q > 2$ (see e.g. Todorov and Tauchen (2010) and BKS (2010)). [59]We also use an approach recommended in Duong and Swanson (2010) for constructing truncated jump measures, in order to assess whether jumps of a particular range of magnitudes are more useful than measures based upon the use of all jumps, or of signed jumps. Our dataset includes high frequency price returns constructed using S&P futures index data as well as stocks in the Dow 30, for the period 1993-2009; and our empirical implementation involves estimating members of the linear and non-linear extended Heterogeneous Autoregressive of the Realized Volatility (HAR-RV) class of models. Our findings suggest that past "large" jump power variations help less in the prediction of future realized volatility, than past "small" jump power variations. This in turn suggests the "larger" jumps might help less in the prediction of future realized volatility than "smaller" jumps. Our empirical findings also suggest that past realized signed jump power variations, which have not previously been examined in this literature, are strongly correlated with future volatility, and that past downside jump variations matter in prediction. Moreover, our results include various experimental setups in which the (forecast) best values of q are larger than 2 for S&P 500 futures. Interestingly, whether or nor jump tests are implemented prior to the construction of jump power variations also affects the choice of q, in a variety of in-sample and out-of-sample forecasting contexts. Finally, incorporation of downside and upside jump power variations does improve predictability, albeit to a limited extent. Overall, our findings are consistent with ABD (2007) in the concluding that continuous components dominate, when predicting volatility.

The rest of the chapter is organized as follows: Section 4.2 discusses volatility and price jump variation, and Section 4.3 discusses the various realized measures of price jump variation that we examine. Section 4.4 outlines our experimental setup, and Section 4.5 gathers our empirical findings. Concluding remarks are contained Section 4.6.

---

[59]In our implementation, for $q > 6$, the prediction results are almost the same as the case $q = 6$ and therefore are not presented.

4.2 Volatility and Price Jump Variations

We adopt a general semi-parametric specification for asset prices. Following Todorov and Tauchen (2010), the log-price of asset, $p_t = \log(P_t)$, is assumed to be an Itô semimartingale process,

$$p_t = p_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s + J_t, \tag{4.31}$$

where $p_0 + \int_0^t \mu_s ds + \int_0^t \sigma_{s-} dB_s$ is a Brownian semi-martingale and J_t is a pure jump process which is the sum of all "discontinuous" price movements up to time t,

$$J_t = \sum_{s \leq t} \Delta p_s.$$

J_t is assumed to be finite[60] and a jump at time s is defined as $\Delta p_s = p_s - p_{s-}$.

When the jump component is a Compound Poisson Process (CPP) - i.e. a finite activity jump process - then,

$$J_t = \sum_{i=1}^{N_t} Y_i, \tag{4.32}$$

where N_t is number of jumps in $[0, t]$. N_t follows a Poisson process, and the $Y_i's$ are i.i.d. and are the sizes of the jumps. The CCP assumption has been widely used in the literature on modeling, forecasting, and testing for jumps. However, jumps could have more general specifications, which contain so called - infinite activity jumps as in Todorov and Tauchen (2010).

The empirical evidence discussed in this chapter involves examining the variation of the log-price jump component using an equally spaced path of a historically observed price sample, i.e. $\{p_0, p_{1\Delta_n}, p_{2\Delta_n} ..., p_{n\Delta_n}\}$, where the sampling frequency $\Delta_n = \frac{t}{n}$ is deterministic[61]. The intra-daily return or increment of p_t is

$$r_{i,n} = p_{i\Delta_n} - p_{(i-1)\Delta_n}.$$

Returns are observed at various frequencies. However, volatility of log-price is often treated as an unobserved variable. The "true" value of variance of price (risk)

[60] See, for example Jacod (2008) or Todorov and Tauchen (2009) for the conditions for the finiteness of jump.

[61] For instance, if we use 5 minute sampling frequency to calculate daily measure in our application. Then $t = 1$ and $n = 78$ and $\Delta_n = \frac{1}{78}$.

is defined in the literature by so-called quadratic variation of the process p_t, i.e.,

$$V_t = [p,p]_t = \int_0^t \sigma_s^2 ds + QJ_t,$$

where the variation of continuous component is $\int_0^t \sigma_s^2 ds$ (integrated volatility) and the variation of jump component is $QJ = \sum_{s\leq t}(\Delta p_s)^2$.

The realized volatility (RV), constructed by simply summing up all successive intra-daily squared returns, converges to the quadratic variation of the process as sampling frequency $n \to \infty$. Andersen, Bollerslev, Diebold and Labys (2001) use realized volatility as an estimator of variation or volatility of the process,

$$RV_t = \sum_{i=1}^{n} r_{i,n}^2 \xrightarrow{ucp} V_t, \tag{4.33}$$

where ucp denotes uniform convergence in probability. RV is useful, in particular in volatility modeling and forecasting.

As jumps are often linked to abnormal or tailed behaviors of returns, the assessment of different RMs of jump variations is important. One way is to decompose price jumps Δp_s as in Duong and Swanson (DS: 2010) and Ait-Sahalia and Jacod (2011) using pre-fixed truncation level γ, $\gamma \geq 0$,

$$JT_{t,\gamma} = \sum_{0<s\leq t} (\Delta p_s)^2 I_{\Delta p_s > \gamma} + \sum_{0<s\leq t} (\Delta p_s)^2 I_{\Delta p_s < -\gamma}, \tag{4.34}$$

where I is an indicator taking 1 if jump size is larger than γ (upside truncated jumps) or less than $-\gamma$ (downside truncated jumps). Intuitively, $JT_{t,\gamma}$ keeps all jumps with absolute magnitude larger than γ.

In this chapter, we assess jump variations using different measures, jump power variations formulated by power transformation of absolute log-price jumps ($|\Delta p_s|^q$),

$$JP_{q,t} = \sum_{0<s\leq t} |\Delta p_s|^q, \tag{4.35}$$

and "upside" jump power variation measure, defined as

$$JPV_{q,t}^{+} = \sum_{0<s\leq t} |\Delta p_s|^q I_{\Delta p_s > 0}. \tag{4.36}$$

$JPV_{q,t}^{+}$ retains the "upside" jump movements. Similarly, we could consider the "downside" jump power variation which keeps all the "downside" jump movements, i.e.,

$$JPV_{q,t}^{-} = \sum_{0<s\leq t} |\Delta p_s|^q I_{\Delta p_s<0}, \tag{4.37}$$

Finally, jump asymmetry could be measured by the so-called signed jump power variation, defined as

$$JA_{q,t} = \sum_{0<s\leq t} |\Delta p_s|^q I_{\Delta p_s>0} - \sum_{0<s\leq t} |\Delta p_s|^q I_{\Delta p_s<0}. \tag{4.38}$$

In the above expression, we are particularly interested in the case where q is larger or equal to 2. Note that for a large value of q, $JP_{q,t}$, $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$, $JA_{q,t}$ are dominated by large jumps. For $q < 2$, the jump variations are not always guaranteed to be finite. The natural estimators for the above jump variations are based on power transformation of intra-daily return, $|r_{i,n}|^q$, which we will discuss in the next section.

4.3 Realized Measures of Price Jump Variations

Our interest in this chapter is to construct and examine the realized measures (RMs) of jump power variations such as $JP_{q,t}$, $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$, $JA_{q,t}$, for a wide range of values of q, and then use them for various prediction experiments. In this line of research, note that for the case $q = 2$, BKS (2010) develop the so-called realized semivariances which are the estimators of $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$. PS (2011) build on these results and make use of realized semivariances to forecast volatility. For the variations with $q \neq 2$, GS (2009) study the predictability of future RV using realized power variations. Realized power variations are formed on the basis of the power transformation of absolute return. They look for the optimal predictors of this type in the forecast. In their set-up, the log-price process is a continuous semimartingale. In the following sections, we briefly review the estimators used in GS (2009), BKS (2010) and PS (2011) and then present the RMs of jump power variations $JP_{q,t}$, $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$, $JA_{q,t}$ used in our chapter.

4.3.1 Semivariances and Realized Power Variations

We start by reviewing the estimators used in BKS (2010) and PS (2011). BKS (2010) construct realized semivariances on the basis of square transformation of intra-daily return, $r_{i,n}^2$, defined as follows:

$$RS^- = \sum_{i=1}^{n} (r_{i,n})^2 \, I_{\{r_{i,n}<0\}},$$

and

$$RS^+ = \sum_{i=1}^{n} (r_{i,n})^2 \, I_{\{r_{i,n}>0\}}.$$

RS^- (RS^+) retains only negative (positive) intra-daily returns and could serve as a measure of downside (upside) risks as pointed out in BKS (2010). They show that RS^+ and RS^- converge uniformly in probability to semi-variances,

$$RS^+ \to \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum (\Delta p_s)^2 I_{\Delta p_s>0}, \tag{4.39}$$

and

$$RS^- \to \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum (\Delta p_s)^2 I_{\Delta p_s<0}.$$

With the above limit results, realized measure of "downside" jump variation is obtained by replacing $\int_0^t \sigma_s^2 ds$ with it's estimator $\widehat{IV}$,

$$\sum_{i=1}^{n} r_{i,n}^2 I_{\{r_{i,n}<0\}} - \frac{1}{2}\widehat{IV} \to \sum (\Delta p_s)^2 I_{\Delta p_s \le 0}. \tag{4.40}$$

In volatility forecasting experiments, PS (2011) use bipower variation for $\widehat{IV}$ [62]. In addition, they construct the so-called "signed" jump variation variable, $\Delta RJ = RS^+ - RS^-$ that captures jump variation asymmetry,

$$\Delta RJ \to \sum (\Delta p_s)^2 I_{\Delta p_s>0} - \sum (\Delta p_s)^2 I_{\Delta p_s<0}.$$

When jumps are not present, ΔRJ converges to 0 and there is no asymmetry in volatility of (log) price process. When the process has jumps, ΔRJ could be a proxy for jump variation asymmetry.

Turning to the discussion of variations with $q \neq 2$, to our knowledge, very few papers empirically examined realized power variations for forecasting. GS (2009) examine the optimal realized power variation, $n^{-1+q/2}\sum_{i=1}^{n} |r_{i,n}|^q$ (optimal q) in forecasting future RV. They build their estimators on the assumption that the price

[62] See BNS (2004) for the dicussion on bipower variation and integrated volatility.

process follows Brownian Semi-martingale. Their implications are therefore restricted to the higher order variation of log-price continuous component, $\int_0^t \sigma_s^q ds$, involving no jumps. In such case, Ait-Sahalia and Jacod (2011) point out that for all $q > 0$,

$$n^{-1+q/2} \sum_{i=1}^{n} |r_{i,n}|^q \rightarrow \mu_q \int_0^t \sigma_s^q ds, \tag{4.41}$$

where $\mu_q = E(|u|^q)$ and u is a standard normal random variable.

4.3.2 Realized Downside and Signed Jump Power Variation

Understanding the role of variables that capture jump information is potentially important for applied practitioners. In this section, we first study recently proposed realized measures of jump power variations $JP_{q,t}$. The measures are constructed using power transformations of absolute intra-daily returns, based on recent limit theory advances due to Jacod (2008) and BKS (2010). Furthermore, the limit theory that we adopt then allows us to construct estimators of downside and upside jump power variations, $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$ for $q > 2$, using intra-daily positive and negative returns. These estimators are suggested by BKS (2010) as alternatives to the semivariances implemented in PS (2011). Finally. making use of the RMs of $JPV_{q,t}^{+}$, $JPV_{q,t}^{-}$, we develop a novel proxy for jump asymmetry (i.e., realized signed jump power variation). The RMs of jump power variations are defined as:

$$RPV_{q,t} = \sum_{i=1}^{n} |r_{i,n}|^q,$$

for $q > 0$.

The realized downside and upside power variations are defined as:

$$RJ_{q,t}^{+} = \sum_{i=1}^{n} |r_{i,n}^{+}|^q,$$

and

$$RJ_{q,t}^{-} = \sum_{i=1}^{n} |r_{i,n}^{-}|^q,$$

for $q > 2$.

For a brief discussion of the above realized measures, the convergences of the above RMs to jump power variations occur when $q > 2$. Therefore, in the prediction experiments, different from previous work, we are particularly interested a range of q from 2 to 6 and allow for price process to contain jumps. Regarding $RPV_{q,t}$, we also look at at the range of q from 0 to 2 by applying a jump robust limit result of Jacod (2008).

Regarding the limiting behavior of $RPV_{q,t}$, Todorov and Tauchen (2010) summarize selected results from Barndorff-Nielsen et. al. (2005) and Jacod (2008). In their set-up, the log-price process contains continuous martingale, jump and drift components. The value of q directly affects the limiting behavior of $RPV_{q,t}$. For instance, for $q < 2$, the limit of $RPV_{q,t}$ is determined by the continuous martingale. For $q > 2$, the limit is driven by jump component. When $q = 2$, both continuous and jump components contribute to the limit of $RPV_{q,t}$. The results are as follows:

$$\begin{cases} \Delta_n^{1-q/2} RPV_{q,t} \xrightarrow{ucp} \mu_q \int_0^t \sigma_s^q ds \text{ , if } 0 < q < 2, \\ RPV_{q,t} \xrightarrow{ucp} V \text{ if } q = 2, \\ RPV_{q,t} \xrightarrow{ucp} JP_{q,t} \text{ if } q > 2. \end{cases} \tag{4.42}$$

BKS (2010) point out that we can go one step further to decompose jump power variations into upside movements and downside movements, i.e.

$$\begin{cases} RJ^+_{q,t} \xrightarrow{ucp} JPV^+_{q,t} \\ RJ^-_{q,t} \xrightarrow{ucp} JPV^-_{q,t} \end{cases} \text{if } q > 2 \tag{4.43}$$

As earlier mentioned, for $q < 2$, the scaled $RPV_{q,t}$ converges to power variations of the continuous component, involving no jumps. Intuitively, with $q > 2$, the scaled $RPV_{q,t}, RJ^+_{q,t}, RJ^-_{q,t}$ eliminate all variations due to the continuous component and keep all the large jumps. In addition, the realized measures are more dominated by large jumps for the high value of q. Conversely, for the case $q < 2$, all jumps are eliminated asymptotically.

Building on (4.43), we could construct the novel RMs of jump power variation asymmetry, so-called "signed" jump power variation. It is straightforward to verify that:

$$RJA_{q,t} = RJ^+_{q,t} - RJ^-_{q,t} \xrightarrow{ucp} JA_{q,t}$$

Note that this variable has not been studied in volatility forecasting literature. PS (2011) study the predictability of the similar estimator, ΔRJ, constructed on the basis of realized semivariance. In our forecasting experiments, we examine the usefulness of this new jump asymmetry variable, $RJA_{q,t}$ with a wide range of values of $q > 2$, in future volatility forecasting.

As the last remark in our discussion of RMs of variations, in the predictive comparison of variables that capture information generated by jumps and the continuous component, we need to select variables that measure the variation of the continuous movements of price process. In this chapter we use multi-power variations, which are estimators of $\int_0^t \sigma_s^q ds$. Those estimators are robust to the existence of jumps. We also utilize these estimators for the jump test implementation highlighted in the next section. The multipower variations discard the impact of jumps by multiplying power transformations of successive absolute intra-daily returns, i.e.,

$$V_{m_1,m_2...,m_j} = \sum_{i=2}^{n} |r_{i,n}|^{m_1} |r_{i-1,n}|^{m_2} .... |r_{i-j,n}|^{m_j},$$

where $m_1, m_2, ..., m_j$ are positive, such that $\sum_1^j m_i = q$.

4.3.3 Testing for Jumps

As discussed in the previous section, realized measures of jump power variations $RPV_{q,t}$, $RJ^+_{q,t}$, $RJ^-_{q,t}$, $RJA_{q,t}$ converge asymptotically to jump power variations $JP_{q,t}$, $JPV^+_{q,t}$, $JPV^-_{q,t}$, $JA_{q,t}$ of the price process. Theoretically, this result also holds for price process without jumps, yielding the limits of zeros. However, in finite sample, it might be useful to implement a pre-testing step to determine whether the log-price process has jumps. The pre-testing approach is developed by ABD (2007) and is empirically examined in DS (2010) for the construction of RMs of truncated jump quadratic variation. We follow this approach in our construction of variables that capture information generated by jumps, in particular we use the jump test statistics developed by BNS (2006) and Huang and Tauchen (2005).

Firstly, we review some theoretical results relating to testing for jumps, namely testing whether $J_t \neq 0$.

In pioneering work, BNS (2006) propose a robust and simple test for a class of Brownian Itô Semimartingales plus Compound Poisson jumps[63]. In recent work, Aït-

[63] A simplified version of the results of the above authors applied to (4.31) for the one-dimensional case is as follows. If the process X is continuous, let $f(x) = x^n$ (exponential growth), let ρ_{σ_s} be the

Sahalia and Jacod (2009a) among others develop a different test which applies to a large class of Itô-semimartingales, and allows the log price process to contain infinite activity jumps - small jumps with infinite concentrations around 0.

Regardless of the estimator that is used, the appropriate test hypotheses are:

$$H_0 : p_t \text{ is a continuous process in the interval } [0,t]$$

$$H_1 : \text{the negation of } H_0 \text{ (there are jumps)}$$

If we use multipower variation, under the null hypothesis the test statistic which directly follows from the CLT mentioned above is:

$$LS_{jump} = \frac{\sqrt{\frac{t}{n}}\left(\sum_{i=1}^{n}(r_{i,n})^2 - \widehat{IV}\right)}{\sqrt{\vartheta \widehat{IQ}}} \xrightarrow{D} N(0,1)$$

and the so-called jump ratio test statistic is:

$$RS_{jump} = \frac{\sqrt{\frac{t}{n}}}{\sqrt{\vartheta \widehat{IQ}/(\widehat{IV})^2}}\left(1 - \frac{\widehat{IV}}{\sum_{i=1}^{n}(r_{i,n})^2}\right) \xrightarrow{D} N(0,1).$$

where $\widehat{IV}$ and $\widehat{IQ}$ are multipower variation estimators of integrated volatility $\int_0^t \sigma_s^2 ds$ and $\int_0^t \sigma_s^4 ds$. BNS (2006) use $V_{1,1}$ (bipower variation) and $V_{1,1,1,1}$. In jump test implementation with multipower estimators, ABD (2007) suggest the use $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}$ (tripower variation) and $V_{\frac{4}{5},\frac{4}{5},\frac{4}{5}}$. The reason we use tripower variation, $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}$, instead of bipower variation, $V_{1,1}$, is that it is more robust to clustered jumps and note that:

$$\widehat{IV} = V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}} \mu_{\frac{2}{3}}^{-3} \tag{4.45}$$

and

$$\widehat{IQ} = \frac{n}{t} V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}} \mu_{\frac{4}{5}}^{-5} \tag{4.46}$$

where $\mu_r = E(|Z|^r)$ and Z is a $N(0,1)$ random variable.

---

law $N(0, \sigma_s^2)$, and let $\rho_{\sigma_s}(f)$ be the integral of f with respect to this law. Then:

$$\sqrt{\frac{1}{\Delta_n}}\left(\Delta_n \sum_{i=1}^{n} f\left(\frac{\Delta_i^n X}{\sqrt{\Delta_n}}\right)^2 - \int_0^t \rho_{\sigma_s}(f) ds\right) \longrightarrow \int_0^t \sqrt{\rho_{\sigma_s}(f^2) - \rho_{\sigma_s}^2(f)} dB_s \tag{4.44}$$

Andersen, Dobrev, Schaumburg (2008) suggest a different estimator that could handle the case of consecutive jumps. This estimator is also more robust to occurrence of zero-return. This robust jump measure is as follows:

$$\widehat{IV} = MedRV_n = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{n}{n-2} \right) \sum_{i=2}^{n-1} med \left(|r_{i-1,n}| \, |r_{i-2,n}| \, |r_{i-3,n}| \right)^2$$

Of note is that an adjusted jump ratio statistic has been shown by extensive Monte Carlo experimentation in Huang and Tauchen (2005), in the case of CCP jumps, to perform better than the two above statistics, being more robust to jump over-detection. This adjusted jump ratio statistic is:

$$AJ_{jump} = \frac{\sqrt{\frac{n}{t}}}{\sqrt{\vartheta max(t^{-1}, \widehat{IQ}/(\widehat{IV})^2)}} \left(1 - \frac{\widehat{IV}}{\sum_{i=1}^{n} (r_{i,n})^2} \right) \xrightarrow{D} N(0,1)$$

In general if we denote the daily test statistics to be $Z_{t,n}(\alpha)$, where n is the number of observations per day and α is the test significance level [64], then we reject the null hypothesis if $Z_{t,n}(\alpha)$ is in excess of the critical value Φ_α, leading to a conclusion that there are jumps. The converse holds if $Z_{t,n}(\alpha)$ is less than Φ_α. In our empirical application, $Z_{t,n}(\alpha)$ is the adjusted jump ratio statistic.

4.3.4 Realized Measures of Daily Variations

With the availability of the RMs such as $RPV_{q,t}, RJ^+_{q,t}, RJ^-_{q,t}$, $RJA_{q,t}$ and the jump tests, in this section, we elaborate further on how to construct daily time series of variables that captures information generated by the variations of log-price process for forecast experiments.

For each day, we calculate the realized measures of jump power variations using a high frequency price path. To mitigate the effect of microstructure noises,[65] we sample data at five-minute frequency as suggested in Aït-Sahalia, Mykland and Zhang (2005). The first group of predictors is constructed without jump tests. The second group of predictors utilizes jump test adjusted technique by ABD (2007). We set $n = 78$, the number of five-minute observations within a day, and consider the range of q from 0.1

[64] i.e., $\Delta_n = 1/n$

[65] The main drawback of realized measures constructed on the basis of high frequency data is that they are contaminated by mictrostructure noises. See Aït-Sahalia, Mykland and Zhang (2005) for further dicussion.

to 6, i.e. $q = \{0.1, 0.2, ..., 5.9, 6\}$.

4.3.4.1 Predictors with No Jump Test

The daily times series of realized measures of jump power variations are formed at a particular day t as follows:

$RPV_{q,t}$ = Realized qth order power variation at day $t = \sum_{i=1}^{78} |r_{i,78}|^q$ with $q > 0$,

$RJ^+_{q,t}$ = Realized Measure of qth order upside jump power variation at day $t = \sum_{i=1}^{78} \left(|r^+_{i,78}|^q\right)$, $q > 2$,

$RJ^-_{q,t}$ = Realized Measure of qth order downside jump power variation at day $t = \sum_{i=1}^{n} \left(|r^-_{i,78}|^q\right)$, $q > 2$,

$RJA_{q,t}$ = Realized Measure of qth order signed jumps power variation at day $t = RJ^+_{q,t} - RJ^-_{q,t}$, $q > 2$,

As noted before, realized qth order power variation with $q < 2$ does not involve jumps.

4.3.4.2 Predictors with Jump Test

First, the predictors are calculated as in section 4.3.4.1. Jump tests are then implemented on daily basis and the predictors are adjusted accordingly. Specifically, RMs of jump power variations at day t are positive if jumps are detected and 0 otherwise. This simple approach is first studied by ABD (2007) in the construction of time series of RMs of quadratic variations of jump component. Let $I_{jump,t}$ be the indicator of jumps, i.e. $I_{jump,t} = 1$ if jumps occur at day t and $I_{jump,t} = 0$ otherwise. Then the adjusted realized measure of jump power variations are expressed as,

$RPV_{q,t}$ = Realized $q - th$ order power variation $= I_{jump,t} * \{\sum_{i=1}^{78} |r_{i,78}|^q\}$,

$RJ^+_{q,t}$ = Realized $q-th$ order upside jump power variation $= I_{jump,t} * \{\sum_{i=1}^{78} \left(|r^+_{i,78}|^q\right)\}$,

$RJ^-_{q,t}$ = Realized $q-th$ order downside jump power variation $= I_{jump,t} * \{\sum_{i=1}^{n} \left(|r^-_{i,78}|^q\right)\}$,

$RJA_{q,t}$ = Realized $q - th$ order signed jumps power variation $= I_{jump,t} * \{RJ^+_{q,t} - RJ^-_{q,t}\}$.

4.3.4.3 Benchmark Realized Variations of the Continuous and Jump Components

The RMs of quadratic variation (RV) and variation of continuous component are formalized as in ABD (2007),

$RVJ_t =$ *Variation of the jump component* $= max\{0, RV_t - \widehat{IV_t}\} * I_{jump,t}$,

$RVC_t =$ *Variation of continuous component* $= RV_t - RVJ_t$,

where $\widehat{IV_t}$ is an estimator of variation of continuous component $\int_0^t \sigma_s^2 ds$. One could use Tripower Variation or Truncated Power Variation. In the chapter, we use Tripower Variation:

$RVJ_t =$ *Variation of the jump component* $= max\{0, RV_t - \widehat{IV_t}\} * I_{jump,t}$,

$RVC_t =$ *Variation of continuous component* $= RV_t - RVJ_t$.

As the above measures in section 4.3.4.1 and 4.3.4.2 depend on q, we take into account the fact that larger q magnifies larger jump in the sum. $RPV_{q,t}, RJ^+_{q,t}$, $RJ^-_{q,t}$ and $RJA_{q,t}$ are assessed for a wide range of values of q, from 2.1 to 6, i.e. $q = \{2.1, 2.2, ..., 5.9, 60\}$. We end up with 40 sub-models (predictors) for each case. For the realized power variation RPV, we set q from 0.1 to 6.

4.4 Models and Forecast Evaluations

4.4.1 Model Specifications

In a classical paper, Ding, Granger and Engle (DGE:1993) find that the auto-correlation of power transformation of daily return of S&P 500 is strongest when $q = 1$, as opposed to the value $q = 2$ widely used in the literature. This leads them to generalize ARCH type model to the class of so-called Asymmetric Power ARCH (APARCH) model. The APARCH specification allows for the flexibility of q in the power qth transformation of absolute returns. GS (2009) point out that this class of models ends up working with volatility that is not measured by squared returns, which researchers and practitioners care the most. Using the five-minute intra-daily returns of the Down Jones Composite over the period 1993-2000, GS (2009) make a thorough empirical correlation analysis of daily RV and realized power variations, with the forecasting horizon from one to four weeks. They conclude that realized power variation with $q = 1$ and future RV display the strongest cross-correlation over the first 10 lags. Beyond this first 10 lags, the cross-correlation holds for $q = 0.5$. This suggests that the prediction of RVs using variables such as realized power variation might be a better approach compared to the lag of RVs. GS (2009) use the Mixed Data Sampling Regression (MIDAS) models to investigate the predictive power of realized power variation for $q < 2$.

We add to the empirical research on this topic by providing results on volatility forecasting using a variety of "new" variables that capture information generated by jumps. In particular, we utilize RMs of jump power variations discussed in the previous section. We estimate an extended Heterogeneous Autoregressive of the Realized Volatility (HAR-RV) class of models. The HAR-RV model, initially developed in Corsi (2009), has been implemented with success in a number of recent contributions. These models are formulated on the basis of the so-called Heterogeneous ARCH, or HARCH, a class of models analyzed by Müller et al. (1997), in which the conditional variance of the discretely sampled returns is parameterized as a linear function of the lagged squared returns over the identical return horizon together with the squared returns over shorter return horizons. Intuitively, different groups of investors have different investment horizons, and consequently behave differently. The genuine HAR-RV model is formally a constrained AR(22) model and is convenient in application as volatility is treated as if it is observed, when we fit regressions in order to assess predictability. In the following, we describe the set-up of HAR-RV and present the specifications that extends HAR-RV to incorporate our new jump variables.

Define the multi-period normalized realized measures for jump and continuous components as the average of the corresponding one-period measures. Namely for daily time series Y_t, we construct $Y_{t,t+h}$ such that

$$Y_{t,t+h} = h^{-1}[Y_{t+1} + Y_{t+2} + ... + Y_{t+h}], \tag{4.47}$$

where h is an integer. $Y_{t,t+h}$ aggregates information between time $t+1$ and $t+h$. The daily time series Y_t could be the RMs such as RV_t, RVJ_t, RVC_t, $RPV_{q,t}$, $RJ^{+}_{q,t}$, $RJ^{-}_{q,t}$, $RJA_{q,t}$ and $q = \{0.1 + 0.1k\}_{k=0}^{k=59}$.

In standard linear and nonlinear HAR-RV models, future RV depends on the past of RV,

$$\phi(RV_{t,t+h}) = \beta_0 + \beta_d \phi(RV_t) + \beta_w \phi(RV_{t-5,t}) + \beta_m \phi(RV_{t-22,t}) + \epsilon_{t+h}, \tag{4.48}$$

where ϕ is a linear, square root or log function.

The incorporations of RMs of jump variations, RVJ_t could be done as in ABD

(2007), using the HAR-RV-J,

$$\phi(RV_{t,t+h}) = \beta_0 + \beta_d \phi(RV_t) + \beta_w \phi(RV_{t-5,t}) + \beta_m \phi(RV_{t-22,t}) + \beta_j \phi(RVJ_t) + \epsilon_{t+h},$$

or HAR-RV-CJ,

$$\phi(RV_{t+h}) = \beta_0 + \beta_d \phi(RVC_t) + \beta_w \phi(RVC_{t-5,t}) + \beta_m \phi(RVC_{t-22,t}) + \beta_{jd} \phi(RVJ_t),$$
$$+ \beta_{jw} \phi(RVJ_{t-5,t}) + \beta_{jm} \phi(RVJ_{t-22,t}) + \epsilon_{t+h}.$$

ABD (2007) find that the class of log HAR-RV, log HAR-RV-J and log HAR-RV-CJ models performs the best for several market indexes. DS (2010) revisit this class of models but focus on the predictive performance of the models applied to Dow 30 individual stock returns. PS (2011) extend this class of models to assess different predictors, the realized semivariance and realized signed jump measure. Their extended HAR-RV model is,

$$\phi(RV_{t,t+h}) = \beta_0 + \beta_1^+ \phi(RS_t^+) + \beta_1^- \phi(RS_t^-) + \beta_5^+ \phi(RS_{t-5,t}^+) + \beta_5^- \phi(RS_{t-5,t}^-)$$
$$+ \beta_{22}^+ \phi(RS_{t-22,t}^+) + \beta_{22}^- \phi(RS_{t-22,t}^-) + \varepsilon_{t+h}.$$

Building on Corsi (2004), ABD (2007) and PS (2011), we extend the HAR- RV to incorporate time series of RMs of jump power variations. In addition, we modify the forecast set-up by examining the forecast of RV_{t+h}, rather than $RV_{t,t+h}$. The specifications are presented as follows:

Specification 1: Class of standard HAR-RV-C Model (Benchmark Model),

$$\phi(RV_{t+h}) = \beta_0 + \beta_{cd} \phi(RVC_t) + \beta_{cw} \phi(RVC_{t-5,t}) + \beta_{cm} \phi(RVC_{t-22,t}) + \epsilon_{t+h}. \quad (4.49)$$

In this benchmark case, future RVs depend on lags of the variation of the continuous component of the process.

Specification 2: Class of HAR-RV-C-PV(q) Model,

$$\phi(RV_{t+h}) = \beta_0 + \beta_{cd} \phi(RVC_t) + \beta_{cw} \phi(RVC_{t-5,t}) + \beta_{cm} \phi(RVC_{t-22,t})$$
$$+ \beta_{jd} \phi(RPV_{q,t}) + \beta_{jw} \phi(RPV_{q,t-5,t}) + \beta_{jm} \phi(RPV_{q,t-22,t}) + \epsilon_{t+h}, \quad (4.50)$$

where $RPV_{q,t}$ is qth order variation of the jump component. $RPV_{q,t-5,t}$ and

$RPV_{q,t-22,t}$ are calculated using (4.47). As discussed in the previous section, we allow for a wide range of values of q from 0.1 to 6. Note that when $q > 2$, the implication of this variable is jump power variations. With $q < 2$, actually the limit is robust to jumps as discussed in Section 2.

Specification 3: HAR-RV-C-UJ(q) Model (upside jump) is defined as,

$$\begin{aligned}\phi(RV_{t+h}) &= \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) \\ &+ \beta^+_{jd}\phi(RJ^+_{q,t}) + \beta^+_{jw}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jm}\phi(RJ^+_{q,t-22,t}) + \epsilon_{t+h}. \qquad (4.51)\end{aligned}$$

This specification incorporates the RMs of qth order upside jump power variations as explanatory variables to forecast future RV. Specifically, $RJ^+_{q,t}, RJ^+_{q,t-5,t}, RJ^+_{q,t-22,t}$ measure the qth order power variation of positive jumps of today, previous week, and previous month, respectively. $RJ^+_{q,t-5,t}, RJ^+_{q,t-22,t}$ are calculated using (4.47). The range of q varies from 2.1 to 6.

Specification 4: HAR-RV-C-DJ(q) Model (downside jump),

$$\begin{aligned}\phi(RV_{t+h}) &= \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) \\ &+ \beta^-_{jd}\phi(RJ^-_{q,t}) + \beta^-_{jw}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jm}\phi(RJ^-_{q,t-22,t}) + \epsilon_{t+h}. \qquad (4.52)\end{aligned}$$

This specification incorporates the RMs of qth order downside jump variations as explanatory variables. Specifically, $RJ^-_{q,t}, RJ^-_{q,t-5,t}, RJ^-_{q,t-22,t}$ are the RMs of the qth order power variations of negative jumps of today, previous week, and previous month, respectively. $RJ^-_{q,t-5,t}, RJ^-_{q,t-22,t}$ are calculated using (4.47). The range of q varies from 2.1 to 6.

Specification 5: HAR-RV-C-UDJ(q) Model (Full decomposition),

$$\begin{aligned}\phi(RV_{t+h}) &= \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) \\ &+ \beta^+_{jd}\phi(RJ^+_{q,t}) + \beta^+_{jw}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jm}\phi(RJ^+_{q,t-22,t}) \\ &+ \beta^-_{jd}\phi(RJ^-_{q,t}) + \beta^-_{jw}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jm}\phi(RJ^-_{q,t-22,t}) + \epsilon_{t+h}. \qquad (4.53)\end{aligned}$$

This specification fully decomposes realized measure of the qth order jump power variation into upside and downside components. The predictors therefore contain both upside and downside jump power variations, i.e. $RJ^+_{q,t}, RJ^+_{q,t-5,t}, RJ^+_{q,t-22,t}$ and

$RJ^{-}_{q,t}, RJ^{-}_{q,t-5,t}, RJ^{-}_{q,t-22,t}$. We set the range of q to vary from from 2.1 to 6 for this specification.

Specification 6: HAR-RV-C-APJ(q) Model,

$$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t})$$
$$+ \beta_{jd}\phi(RJA_{q,t}) + \beta_{jw}\phi(RJA_{q,t-5,t}) + \beta_{jm}\phi(RJA_{q,t-22,t}) + \epsilon_{t+h}. \quad (4.54)$$

This class of models uses RMs of signed jump power variations, measures of jump asymmetry, as explanatory variables for RV forecast. Specifically, predictors are $RJA_{q,t}$, $RJA_{q,t-5,t}$ and $RJA_{q,t-22,t}$, calculated using (4.47).

The estimation of the above models is simply done by OLS regression. We report the parameters and measures of fit. Across all the specifications, there is a potential issue of the serial correlation due to the long forecast horizons ($h = 5, 22$). Though serial correlation does not affect the consistency of the estimated parameters, robust estimates of covariance matrix need be addressed. In our empirical experiments, we apply both standard and robust heteroskedasticity-and- autocorrelation-consistent (HAC) estimators of covariance matrix [66] .

In a different forecast experiment, we construct realized measures of the truncated jump power variations following the similar approach as in DS (2010). Specifically, we define the RMs of jump power variations truncated at a fixed level γ to be:

$$RJ_q(\gamma) = \sum_{i=1}^{n} |r_{i,n}|^q I_{|r_{i,n}|<\gamma}.$$

Similarly for the RMs of downside jump power variations truncated at fixed level γ,

$$RJ^{-}_q(\gamma) = \sum_{i=1}^{n} |r_{i,n}|^q I_{-\gamma<r_{i,n}<0},$$

and for realized measure of upside jump power variations truncated at fixed level γ,

$$RJ^{+}_q(\gamma) = \sum_{i=1}^{n} |r_{i,n}|^q I_{0<r_{i,n}<\gamma}.$$

[66] For Hac estimator, we use Newey-West estimator.

Then, if one is interested in jumps with magnitude less than γ in the forecast of future RV, the time series of $RJ_{q,t}, RJ^{-}_{q,t}, RJ^{+}_{q,t}$ on the right hand side of forecasting equation(4.50) (4.51) (4.52) (4.53) (4.54) could be replaced by $RJ_{q,t}(\gamma), RJ^{-}_{q,t}(\gamma), RJ^{+}_{q,t}(\gamma)$. In this context, we assume that the modeler has a predetermined knowledge of γ. In our empirical implementation, we choose γ on the basis of the sample of maximum of monthly increments which represents monthly abnormal events. A question we would want to see is whether the choice of larger γ has an impact on the volatility prediction.

Note that we could obtain the optimal value of q for prediction of volatility under a certain measure of fit criteria such as the minimum mean square error. However, this is not the aim of our chapter. By using a wide rage of values of q, we are more interested in capturing the pattern of the predictive powers of the RMs of jump power variations. The pattern also helps in approximating the optimal value of q in the prediction.

Regarding the predictive regression of the above models, note that for specification 2, we need to estimate 60 linear regression equations, depending on q from 0.1 to 6. For each specification 3,4,5 and 6, we need to estimate 40 linear regression equations, depending on q from 2.1 to 6. A straightforward way to assess the usefulness of the RMs is to compare the predictive accuracy, measured by mean square errors or R^2 across all values of q. In the next section, we discuss the forecast evaluation methods that are being used in empirical implementation.

4.4.2 In-Sample and Out-of-Sample Forecast Evaluation

For each specification, we fit the above forecasting equations by ordinary least square. The forecast horizons that we examine in this chapter are set to be $h = 1, 5, 22$ which are the one day ahead, one week ahead and one month ahead horizon, respectively. Our model specifications extend the standard HAR-RV models, as presented in previous section. For each specification, we have 40 sub-models, corresponding to 40 different values of q. Once a measure of fit is obtained, we present it as a function of q and the relationship between q and the measure of fit could be plotted. Regarding the measure of fit, a straightforward way is using in-sample adjusted R^2. The other favored measure of fit is the out-of-sample R^2, calculated from projection of the predictive RV sample on the sample of forecasted RV implied by the model. For the pairwise model comparisons, we use the Diebold-Mariano (DM:1995) test and quadratic loss function.

Specifically, the entire sample of T observations is divided into two samples, the estimation sample containing R observations and the prediction sample containing $P = T - R$ observations. The traditional in-sample adjusted R^2 is calculated using entire sample T.

For the out of sample forecast, we calculate the R^2 using recursive, rolling or fixed estimation schemes. If the forecast horizon $h = 1$ and the recursive estimation are used, the model is to be fitted by P regressions using data chunks from 1 to $R, 1$ to $R + 1, ...1$ to $T - 1$. Alternatively, we could use the rolling scheme where the P regressions are implemented using data chunks 1 to R, 2 to $R + 1, ..., T - P$ to $T - 1$. The fixed scheme requires the estimation using the entire sample. After this step, we could calculate P predicted values implied by the models. Next, the out-of-sample R^2s are obtained by simply regressing the prediction sample on the forecasts implied by the models. Note that the above procedure is presented with forecasting horizon $h = 1$. For the general forecast horizon h and the recursive scheme, the models are fitted P times using data chunk from 1 to $R - h + 1, 1$ to $R + 1, ...1$ to $T - h$. For the rolling scheme, P model-implied forecasts are obtained by the estimation using data chunk from 1 to $R - h + 1, 2$ to $R + 1, ..., T - P - h + 1$ to $T - h$.

Now turn to predictive equality accuracy test of Diebold and Mariano (DM: 1995), we could formally make a pairwise comparisons of any two models by applying this test. Suppose we are interested in the comparison of two models $i = 1, 2$ using the times series y_t, $t = 1, 2, ..., T$. The mean square forecast error (MSFE) is defined as

$$MSFE = \sum_{\tau=R-h+2}^{T-h+1} \left(y_{t+h} - \widehat{y}_{i,t+h}\right)^2,$$

where $\widehat{y}_{i,t+h}$ is the forecast for horizon h for model i. Denote $\varepsilon_{i,t+h}$ to be model's prediction error of model i. The hypothesis could be set up as. The null H_0 : $E(\varepsilon_{1,t+h}^2) - E\left(\varepsilon_{2,t+h}^2\right) = 0$ and alternative H_1 : not H_0. The actual statistics is constructed as: $DM = P^{-1} \sum_{k=1}^{P} (d_t/\widehat{\sigma})$ where $d_t = \widehat{\varepsilon}_{1,t+h}^2 - \widehat{\varepsilon}_{2,t+h}^2$, a the comparative measure of fit between the two models. $\widehat{\sigma}$ is the the estimator of standard deviation of $(\sum_{k=1}^{P} d_t)/P$. The choice of this estimator could be set as a heteroskedasticity and auto-correlation robust estimator (HAC). In addition, to the acceptance and rejection outcome of the test on the basis of the test statistics, we could also infer that the negative statistics implies that model 2 is preferred to model 1 as it's statistics measure of fit over the out of sample forecast is superior.

4.4.3 Alternative Models to HAR-RV and Other Issues

Given the main focus of our chapter is to assess the predictability of the new group of variables that capture information generated by jumps, we use the simple predictive regression models, i.e. the extended HAR-RV in this chapter. For an alternative to the extended HAR-RV class of models, the GARCH-based model as in BKS (2010) could be considered. The other approach is using stochastic volatility models. Both approaches require us to treat true volatility as an unobserved variable. In this context, RVs are additional variables that capture rich information generated by high frequency data sets. Stochastic Volatility (SV) model in discrete time is discussed in depth by Shephard (2005). We could use models that inputs RV variable into volatility equation. One way is to estimate the bivariate return - stochastic volatility system building on Liesfield and Richard (2003) filtering framework. In the context of mixed data sampling, one could also implement non-linear regression MIDAS as used in Ghysels and Sohn (2009) as an alternative to HAR-RV model.

In addition, with the choice of volatility estimator, variables that capture information generated by jumps (RMs of Jump Power Variations), jump test statistics, and predictive models as discussed in previous sections, before moving to the discussion on empirical findings, it is useful make an comparative overview on empirical strategies implemented in our chapter and other related papers. The following list summarizes the selected papers that examine jumps and higher order power transformation of absolute returns to predict future volatility.

Summary of Related Work using RMs of Power Variations for Volatility Prediction[67]

Paper	HFD	Jumps	Dow/Upside	Power q	Jump Test	Truncation
DGE (1993)	No	No	No	0-5	No	No
LM (2005)	Yes	Yes	No	0-2	No	No
ABD (2007)	Yes	Yes	No	2	Yes	No
GS (2009)	Yes	No	No	0-2	No	No
PS (2011)	Yes	Yes	Yes	2	No	No
Duong (2012)	Yes	Yes	Yes	0-6	Yes	Yes

In the above list, note that our work makes a thorough examination of jumps variations by using a wide range of values of q compared to other papers and we also consider jump test adjusted RMs in predictions. In the next section, we present empirical findings of our chapter.

4.5 Empirical Findings

4.5.1 Data Description

For empirical implementation in this chapter, we implement the forecasting experiments on S&P 500 futures for the period 1993-2009. We also look at Dow 30 components in the period 1993-2008 as in DS (2010). The data source for stocks is the TAQ database. In the data processing, we follow the common practice in the literature by eliminating from the sample those days with infrequent trades (less than 60 transactions at our 5 minute frequency).

One problem in data handling involves determining the method to filter out an evenly-spaced sample. In the literature, two methods are often applied - *previous tick* filtering and *interpolation* (Dacorogna, Gencay, Müller, Olsen, and Pictet (2001)).

[67] The table summarizes the selected papers that examine jumps and higher order power transformation of absolute returns to predict future volatility. The first column is the list of papers under consideration. The second, third and fourth column provide information whether the paper in the list utilizes high frequency data (HFD), jumps, downside/upside jumps, respectively. The fifth column provides the range of order q used in each paper. The sixth and seventh column provide information whether the paper implements jump test adjustment technique as in ABD (2007) and whether the paper looks at truncated jumps (truncation) in the construction of jump variables for volatility prediction, respectively.

As shown in Hansen and Lund (2006), in applications using quadratic variation, the *interpolation* method should not be used, as it leads to realized volatility with value 0 (see *Lemma 3* in their paper). Therefore, we use the *previous tick* method (i.e. choosing the last price observed during any interval). We restrict our data-set to regular time (i.e. 9:30am to 4:00pm) and ignore ad hoc transactions outside of this time interval. To reduce microstructure effects, the suggested sampling frequency in the literature is from 5 minutes to 30 minutes[68]. As mentioned above, we choose the 5 minute frequency, yielding a maximum of 78 observations per day.

4.5.2 Prediction without Jump Test

First, we calculate all daily RMs as discussed in section 4.3.4 for S&P 500 futures. For each realized measure, we end up with a time-series of size $T = 4123$. In the out-of-sample forecasting experiments, we choose prediction sample size, $P = 410$, and estimation sample size, $R = 3713$, respectively.[69]

Models considered in our empirical application are discussed in Section 4.3. We present all the specifications in Table 4.2. For a quick summary, the specification 1 (benchmark model), HAR-RV-C incorporates only RMs of continuous component variations as predictors. The specification 2, the HAR-RV-C-PV ($q > 0$) uses RMs of continuous component variations and RMs of qth (jump) power variation components as predictors[70]. The specification 3, the HAR-RV-C-UJ ($q > 2$) uses RMs of continuous component variations and RMs of the qth order "upside" jump power variation components as predictors. The specification 4, HAR-RV-C-DJ ($q > 2$) utilize continuous component variations and the qth order "downside" jump power variation components as predictors. The specification 5, HAR-RV-C-UDJ ($q > 2$) builds directly on specification 3 and 4, and uses both RMs of the qth order upside and downside jump power variation components in predictions. The specification 6, HAR-RV-C-APJ examines variables that capture jump asymmetry by incorporating RMs of the signed qth order jump power variations in the prediction. The formulation

[68] See Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005)

[69] We also implement other choices for P =210, 310, 510, 610,710 and results show the same pattern, which are available upon request.

[70] For q>2,the realized power variation converges to the jump power variation. For q<2, the standardized realized power variation converges to power variation of continuous component as discussed in Section 3.

of the time series, $RPV_{q,t}$, $RJ^+_{q,t}$, $RJ^-_{q,t}$ and $RJA_{q,t}$ is shown in details Section 4.3.4.

The empirical analyses of exchange rates, equity index returns, and bond yields in ABD (2007) suggest that the volatility jump component is both highly important and distinctly less persistent than the continuous component, and that separating the "rough" jump movements from the smooth continuous movements results in significant in-sample volatility forecast improvements (i.e. the linear and nonlinear HAR-RV-CJ models perform better than the other two classes of models).

We first provide a brief discussion on the performance of the models for S&P futures. In our chapter, the predictive performance of a model is measured by both its in-sample and out-of sample R^2, which is similar to approach taken in ABD (2007). We also carry out the Diebold-Mariano (1995) predictive equivalence tests to determine whether the choice of order q matters for the qth order jump power variations in forecasting future RV.

Turning to our regression results, Table 4.1 reports the regression estimates, in-sample and out-of-sample R^2 values for the linear, square root and log HAR-RV-C models at daily ($h = 1$), weekly ($h = 5$) and monthly ($h = 22$) prediction horizons. The entries in bracket are t-statistics calculated using the Newey-West estimator with auto-correlation up to 44 lags[71]. Regarding in-sample and out-of-sample R^2s, as shown in the table, the square root models and log models perform much better than their linear counterpart regardless of the prediction horizon. For instance, at the forecasting horizon $h = 1$, the in-sample and out-of-sample R^2 of square root models are 0.45 and 0.34 while those of the linear counterpart are 0.35 and 0.24, respectively. In addition, the estimates of β_{cd}, β_{cw}, β_{cm} and t-statistics confirm the long memory persistent feature of volatility. For the linear model with $h = 1$, the t-statistics of the monthly forecast parameter is 7.81, implying that the continuous component of the previous month could help the one-day-ahead prediction of volatility. This statistical pattern holds for square root and log models across all forecast horizons. In addition, at prediction horizon $h = 22$, while the in-sample R^2s are large, the out-of sample results show an opposite direction.

In the formulation of RMs of jump power variations, $RPV_{q,t}$, $RJ^+_{q,t}$, $RJ^-_{q,t}$, and $RJA_{q,t}$, order q is gridded by 0.1 from 2.1 and 6, i.e $q = \{2.1, 2.2, ..., 5.8, 5.9, 6\}$.

[71]HAC estimator is known to be robust to both heteroscedasticity and auto-correlation.

The choice of maximal $q = 6$ is sufficient to determine the effect of large jumps and their predictive power[72]. With $q > 2$, the realized (jump) power variation, $RPV_{q,t}$, converges asymptotically to jump power variations of log-price process. In addition, larger q effectively eliminates the effect of continuous component and smaller jumps and magnify the impacts of large jumps. In the presentation of results, we choose $q = 2.5$ and $q = 5$, the two representative cases for small and large jump power variations. Table 4.3A, 4.3B, 4.3C, 4.3D report predictive regression estimates of the two cases. Each table involves linear, square root or log model. All the numbers in the brackets are t-statistics. For the in-sample forecast results, jump coefficients are not statistically significant for $q = 5$ (large jumps). The results hold across all model specifications. For $q = 2.5$, the t-statistics are significant for β_{jm} in HAR-RV-C-PV linear and square root models. Similarly, the t-statistics are 2.366 and 2.1 for forecasting horizon $h = 1$ in HAR-RV-C-PDJ linear and square root models. Regarding the full "decomposition" HAR-RV-C-PDUJ model, we find that the downside jumps have an impact on future RV at one-day-ahead forecast horizon ($h = 1$). In particular, for linear model, Table 4.3C shows that the t-statistics for β^{-}_{jd} is 2.138. All the upward jump variations have small impacts on the prediction. More interestingly, correlation between the past $RAJ(q)$ and and future RV is strong across all forecast horizons (daily, weekly and monthly) for all models under consideration (linear, square root, log). Table 4.4 depicts the findings in the group of log models, showing the t-statistics for β_{jd} of 10.05 (daily), 9.15 (weekly), 10.01(monthly) for case $q = 2.5$ and 10.76 (daily), 9.91 (weekly), 11.08 (monthly) for case $q = 5$. The finding strongly suggests that jump asymmetry matters for modeling future RV, at least at the shorter horizon. In addition, the asymmetry holds for both large and small jumps.

Turn to the analysis of the predictive comparison, our prediction experiments show improvements for both in sample and out of sample once RMs of jump power variations are used as additional predictors in volatility forecasting. For example, at the forecast horizons $h = 1$ and $h = 5$, the out-of sample R^2s of the HAR-RV-C square root models are 0.341 for $h = 1$ and 0.244 for $h = 5$ compared to that numbers of 0.368 and 0.262 of HAR-RV-C-PV models. This is equivalent to 8% and 7.5% increases in R^2 if we switch from HAR-RV-C to HAR-RV-C-PV models. However, as shown in

[72] In our implementation, for $q > 6$, the prediction results are almost the same as the case $q = 6$ and therefore are not presented.

the table, the continuous component, RVC, dominates in all prediction experiments, which is consistent with the previous findings in the literature on volatility forecasting using high frequency data. There are little improvements in R^2 for HAR-RV-C-PDUJ in the prediction. Interestingly, the table suggests in-sample and out-of sample R^2 be smaller for the larger q when we examine prediction experiments for the case $q = 2.5$ and $q = 5$. Once we consider a wider range of values of q, this pattern is clear as shown next.

Table 4.4 shows the Diebold and Mariano (DM) test statistics for fixed, recursive and rolling schemes. In the construction of the statistics, denoted in the table as DM Stat, we make a restriction for q to be larger or equal to 2.5, i.e. ($q = \{2.5+k*0.1\}_{k=0}^{k=35}$) and then search for the values of q that yield the maximal and minimal mean square errors. More specifically, q_b denotes the the value of q that yields biggest R^2 and q_s denotes the value of q that yields the smallest R^2. The table shows that q_b is smaller than q_s. In addition, for most of the models, the value of q_b is 2.5 and the value of q_s is 6. We test whether the predictions of future volatility using RMs of (jump) power variations as predictors differ if q_b and q_s are used. The results of DM tests show that most of the t-statistics are significant, regardless of which forecast scheme is used. In particular, the results are stronger for downside jump measures.

Finally, the pattern involving R^2s suggested in the above discussion is confirmed by our figures shown in the appendix. In Figure 4.1, we plot the in-sample adjusted R^2s of all linear and nonlinear models across horizon $h = 1, 5$ and 22. The vertical axis ranges from 0 to 1 for the value of the R^2. The horizontal axis ranges from 0.1 to 6, representing the 60 grid points of value of q, i.e. $q = \{0 + 0.1 * k\}_{k=1}^{60}$. In those plots, except for HAR-RV-PV models, we focus on the part of the curves on the right side of 2 as convergence to jump power variation occurs only when $q > 2$ (see e.g. Todorov and Tauchen (2010) and BKS (2010)). The purple curve represents the R^2s of HAR-RV-C-PDUJ model (full decomposition). The light blue curve represents the R^2s of HAR-RV-C-AJP model. The dark purple curve represents HAR-RV-C-PUJ model. The light green curve represents HAR-RV-C-PDJ. The dark blue curve and orange curve represent the R^2s of HAR-RV-C-PV and the benchmark model HAR-RV-C model, respectively. Notably, for $q > 2$ the R^2 is monotonically decreasing in q. These results are consistent with what we found in Table 4.3. With the monotonic pattern, all the curves look very close to one another, except for HAR-RV-C-PDUJ and HAR-RV-C-AJP model. HAR-RV-C-PDUJ model is slightly better for daily and weekly horizon while HAR-RV-C-AJP model is slightly superior for monthly horizon. It is also clear that the R^2s of all models are higher than those of the benchmark model. This suggests that higher order jumps be helpful in prediction of future RV. The predictive power of large jumps power variations dies out as q becomes bigger. As the higher order jump power variations are dominated by large jumps, the observation suggests that large jumps play less important role in the prediction and dies out when the value of q is larger than 6.

Finally, we also plot the results for the mean square errors across power q, as shown in Figure 4.4. The shapes of plots are in opposite direction to R^2, supporting our earlier findings. In addition to S&P futures, we also implement the volatility prediction applied to individual stocks in the Down 30. We get the similar pattern. However, the optimal values of q in the prediction for those stocks are smaller than 2 and mostly stick around 1, which is consistent with the findings in GS (2009). We show the results for several individual stocks of the Dow 30 components in Figure 4.5[73]. The patterns are obviously similar to S&P futures.

4.5.3 Prediction with Jump Test and Truncated Jump

In the previous section, we present a set of results which are purely based on the RMs of jump power variations which are not adjusted for the jump tests. Theoretically, the realized measures should converge to jump power variations. In finite sample with the sampling choice of 5 minutes ($n = 78$ per day), ABD (2007) develop a straightforward procedure to separate the variation of log-price process due to jumps. We follow this approach to adjust the realized measures of jumps for any day that jump does not occur. In particular day, we first test for jumps using the simple jump test procedure and set the realized measures of jumps to be 0 once the jump statistics is significant. With the new time series of $RPV_{q,t}, RJ^+_{q,t}$, $RJ^-_{q,t}$ and $RJA_{q,t}$, we then carry out the similar forecast experiments as in previous section.

The results show similar pattern as earlier findings. Figure 4.2 plots the in-sample R^2s for all specification and horizons while Figure 4.3 plot the out of sample R^2s for

[73] We present the result for 4 stocks. Results for other stocks in Dow 30 are avaialbe upon request.

linear, square root at the forecast horizon $h = 1$ and $h = 5$. Across all plots, HAR-RV-C-PDUJ is slightly better for and daily and weekly horizon while HAR-RV-C-AJP is slightly better for monthly horizon. Regarding all specifications from 2 to 6 of linear square root and log models, the R^2s are higher than those of the benchmark model. This result is consistent with the earlier findings that higher order jumps help in prediction of future. In addition, similar to the above discussion, large jumps play less important role in the prediction and dies out. In comparison of out-of sample R^2s between jump test and no jump test cases, we see a marginal improvement in the jump test case. Though the increase is very small, this would suggest jump test might be helpful in the forecast experiments using jump variations.

As an additional remark, as shown in Figure 4.3A and 4.3B for S&P futures, the out-of-sample results for the no-jump-test case point out the scenarios where the optimal predictive values of q are larger than 2, as opposed to the the results found in earlier in the literature where q is around 1. Interestingly, the curves change when jump tests are implemented. For linear models at horizon $h = 1$ and $h = 5$, the optimal q is larger than 2 in Figure 4.3A (no jump tests) and is less than 2 in Figure 4.3B (with jump tests). Conversely, for the linear model, the optimal q is larger than 2 when jump tests are implemented. This illustration therefore also suggests that the implementation of jump tests could affect the results of the prediction.

Now turn to the truncated jumps variables, as discussed in section 4.4.1, we truncate large jumps on the basis of percentiles of the time series of monthly largest increments, as implemented in DS (2010). For the experiments, we pick $\gamma = 5\text{th}, 10\text{th}$ and 25th percentile of the sample spanning period 1993-2009, i.e. we discard all the jumps larger than this threshold and construct the new time series $RPV_{q,t}(\gamma), RJ^+_{q,t}(\gamma)$, $RJ^-_{q,t}(\gamma)$ and $RJA_{q,t}(\gamma)$. We then implement forecast using specifications as in 4.1. Interestingly, the results are almost the same as in the above discussion, implying that the larger jumps matter little in the prediction of future volatility.

In summary, our analysis demonstrates that: (i) Continuous component dominates in the predictions. (ii) There is a strong correlation between our jump power variation based jump asymmetry variable and future realized volatility and downside jumps matters more than upward jumps in the prediction. (iii) Incorporation of downside and upside jump power variations might help in prediction but to a limited extent in term of both in-sample and out-of-sample prediction. (iv) There is a strong pattern that higher order jump power variations help less in the prediction of realized volatility, regardless of model specifications that we consider.(v) We find the evidence that the optimal value of q could be larger than 2, depending on the set-up and jump implementation. (vi) Finally, the implementation of jump tests might change the results in the predictions.

4.6 Concluding Remarks

In this chapter, we build on the recent theoretical results of Jacod (2008) and Barndorff-Nielsen and Shephard (2004, 2006) and BKS (2010) to assess large jump power variations, downside (upside) jump power variations, and asymmetry jump power variations. In particular, we look at the role of those variables in the prediction of future realized volatility. We do so by extending the class of approximate long memory model, HAR-RV. Our results are consistent with the earlier findings in the literature, such as ABD (2007) that continuous component dominates in the prediction of future realized volatility. The separation of continuous and jump components could help in increasing in-sample and out-of- sample R^2. In addition, we find a pattern of predictability in which past "large" jump power variations help less in the prediction of future realized volatility, than past "small" jump power variations. This suggests the "larger" jumps might help less in the prediction of future realized volatility than "smaller" jumps. Regarding jump asymmetry, there is an evidence that the signed jump power variation has a strong correlation with future RV. Our results also show that downside jump power variation might matter for modeling future RV. Moreover, in various experimental setups, the (forecast) best values of q are larger than 2 for S&P futures. Finally, incorporation of downside and upside jump power variations does improve predictability, albeit to a limited extent.

Table 4.1: HAR-RV-C Predictive Regression for S&P 500 futures (Benchmark)*

Linear Model				Square Root Model				Log Model			
β_0	β_d	β_w	β_m	β_0	β_d	β_w	β_m	β_0	β_d	β_w	β_m
h=1 (Daily Forecast)											
0.000	0.089	0.060	1.654	-0.002	0.067	0.117	1.001	-0.200	0.167	0.099	0.716
(-0.67)	(1.93)	(0.38)	(7.81)	(-1.569)	(2.70)	(1.70)	(11.91)	(-1.15)	(7.04)	(1.66)	(11.71)
$R^2_{in}(R^2_{out})$ = 0.35(0.244)				$R^2_{in}(R^2_{out})$ = 0.45(0.341)				$R^2_{in}(R^2_{out})$ = 0.45(0.388)			
h=5 (Weekly Forecast)											
0.000	0.058	-0.075	1.832	-0.001	0.055	0.037	1.077	-0.346	0.134	0.142	0.688
(-0.71)	(0.51)	(-0.43)	(10.31)	(-1.00)	(0.94)	(0.40)	(12.62)	(-1.84)	(5.80)	(2.40)	(10.85)
$R^2_{in}(R^2_{out})$ = 0.35(0.169)				$R^2_{in}(R^2_{out})$ = 0.44(0.244)				$R^2_{in}(R^2_{out})$ = 0.43(0.296)			
h=22 (Monthly Foreast)											
0.000	-0.034	0.387	1.375	0.000	0.011	0.137	0.978	-0.772	0.076	-0.014	0.847
(0.32)	(-0.89)	(3.47)	(11.86)	(0.18)	(0.41)	(1.92)	(15.04)	(-3.08)	(3.24)	(-0.19)	(12.39)
$R^2_{in}(R^2_{out})$ = 0.33(0.026)				$R^2_{in}(R^2_{out})$ = 0.41 (0.0357)				$R^2_{in}(R^2_{out})$ = 0.38 (0.033)			

Table 4.2: Summary of Model Specifications for RV Forecasting

Specification 1	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) + \epsilon_{t+h}$
Specification 2 (HAR-RV-C-PV(q))	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t,t-5}) + \beta_{cm}\phi(RVC_{t,t-22}) + \beta_{jd}\phi(RPV_{q,t}) + \beta_{jw}\phi(RPV_{q,t,t-5}) + \beta_{jm}\phi(RPV_{q,t,t-22}) + \epsilon_{t+h}$
Specification 3 (HAR-RV-C-UJ(q))	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t,t-5}) + \beta_{cm}\phi(RVC_{t,t-22}) + \beta^+_{jd}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jw}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jm}\phi(RJ^+_{q,t-22,t}) + \epsilon_{t+h}$
Specification 4 (HAR-RV-C-DJ(q))	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t,t-5}) + \beta_{cm}\phi(RVC_{t,t-22}) + \beta^-_{jd}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jw}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jm}\phi(RJ^-_{q,t-22,t}) + \epsilon_{t+h}$
Specification 5 (HAR-RV-C-UDJ(q))	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) + \beta^+_{jd}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jw}\phi(RJ^+_{q,t-5,t}) + \beta^+_{jm}\phi(RJ^+_{q,t-22,t}) + \beta^-_{jd}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jw}\phi(RJ^-_{q,t-5,t}) + \beta^-_{jm}\phi(RJ^-_{q,t-22,t}) + \epsilon_{t+h}$
Specification 6 (HAR-RV-C-APJ(q))	$\phi(RV_{t+h}) = \beta_0 + \beta_{cd}\phi(RVC_t) + \beta_{cw}\phi(RVC_{t-5,t}) + \beta_{cm}\phi(RVC_{t-22,t}) + \beta_{jd}\phi(RJA_{q,t}) + \beta_{jw}\phi(RJA_{q,t-5,t}) + \beta_{jm}\phi(RJA_{q,t-22,t}) + \epsilon_{t+h}$

* The table 4.1 summarizes the estimation of HAR-RV-C model at daily (h=1), weekly (h=5) and montly (h=22) horizon. For each horizon, the first row entries are the parameter estimates, the second row entries in bracket are t-statistics. The third row reports R_I and R_O, the in and out-of- sample R-square of the predictive regressions, respectively.

Table 4.3A: Predictive Regression for q=2.5 and q=5 for S&P 500 Futures *

		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
β_0	q=2.5	0.001	0	0.001	0.006	0.005	0.008	-0.519	-0.702	-1.228
		-3.035	-1.737	-2.107	-3.444	-2.885	-3.286	(-1.902)	(-2.369)	(-3.290)
	q=5	0.001	0	0.001	0.003	0.004	0.006	-0.38	-0.55	-1.024
		-3.171	-2.345	-2.468	-2.486	-2.455	-3.136	(-1.620)	(-2.173)	(-3.203)
β_{cd}	q=2.5	0.066	0.004	-0.072	0.02	0.024	0.031	0.174	0.136	0.075
		-1.448	-0.031	(-1.090)	-0.599	(0.347	-0.674	-6.808	-5.691	-2.978
	q=5	0.073	0.019	-0.078	0.056	0.056	0.029	0.169	0.133	0.071
		-1.771	-0.17	(-1.319)	-1.786	(0.961	-0.71	-6.893	-5.73	-2.944
β_{cw}	q=2.5	-0.122	-0.251	0.421	0.015	-0.116	0.111	0.087	0.128	-0.017
		(-0.812)	(-1.419)	-2.79	-0.177	(-1.058)	-1.247	-1.346	-2.102	(-0.217)
	q=5	-0.067	-0.197	0.429	0.061	-0.065	0.088	0.098	0.137	-0.012
		(-0.443)	(-1.151)	-2.944	-0.777	(-0.616)	-1.015	-1.594	-2.289	(-0.164)
β_{cm}	q=2.5	0.701	1.23	0.686	0.52	0.695	0.37	0.681	0.653	0.793
		-2.472	-4.925	-2.215	-3.628	-4.685	-1.933	-10.055	-9.149	-10.088
	q=5	1.25	1.569	1.029	0.854	0.973	0.784	0.691	0.666	0.818
		-6.84	-10.869	-5.276	-9.239	-10.025	-7.699	-10.764	-9.918	-11.084

Table 4.3A: Predictive Regression for q=2.5 and q=5 for S&P 500 Futures (Cont.)

β_{jd}	q=2.5	0.072	0.214	0.165	0.108	0.067	-0.042	-16.351	-9.511	-2.187
		-0.434	1.317)	-0.7	-1.942	-0.879	(-0.426)	(-1.441)	(-0.787)	(-0.143)
	q=5	18.272	56.886	71.9	0.625	-0.05	-0.84	-2597	-1047	3180
		-0.351	1.171)	-1.002	-0.563	(-0.035)	(-0.451)	(-0.877)	(-0.302)	-0.889
β_{jw}	q=2.5	0.788	0.793	-0.186	0.321	0.471	0.08	27.949	30.637	10.902
		-1.628	1.775)	(-0.543)	-1.916	-2.573	-0.544	-0.872	-0.97	-0.362
	q=5	194.637	195.392	-76.421	3.984	7.134	3.415	2032	6163	127
		-1.158	-1.469	(-0.808)	-1.064	-1.884	-1.223	-0.215	-0.68	-0.016
β_{jm}	q=2.5	1.18	0.46	1.236	0.387	0.195	0.746	32.416	26.803	51.586
		-2.238	-0.893	-1.976	-1.969	-0.793	-2.742	-0.984	-0.726	-1.389
	q=5	114.426	2.21	211.491	0.879	-1.852	2.98	10776	6132	10073
		-0.714	-0.017	-1.346	-0.245	(-0.454)	-0.987	-1.066	-0.59	-1.056
R^2_{in}	q=2.5	0.376	0.372	0.333	0.463	0.452	0.418	0.452	0.434	0.383
	q=5	0.368	0.368	0.333	0.455	0.446	0.414	0.451	0.434	0.383
R^2_{out}	q=2.5	0.315	0.199	0.033	0.368	0.262	0.04	0.39	0.297	0.032
	q=5	0.244	0.167	0.027	0.344	0.24	0.037	0.389	0.296	0.032

* The table 4.3A summarizes the regression parameter estimates for HAR-RV-C model at daily (h=1), weekly (h=5) and montly (h=22) horizon. For each parameters corresponding to q=2.5 or q=5, the first row entries are the parameter estimates. The entries in bracket in the second row are t-statistics. The four rows at the bottom report R_{in} and R_{out}, the in-sample and out-of- sample R-squares of the predictive regressions for the case q=2.5 and q=5, respectively.

Table 4.3B: Predictive Regression for q=2.5 and q=5 for S&P 500 Futures *

		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
β_0	q=2.5	0.001	-0.702	-1.228	0.006	0.005	0.008	-0.525	-0.710	-1.236
		(3.024)	(-2.369)	(-3.290)	(3.462)	(2.886)	(3.281)	(-1.917)	(-2.385)	(-3.300)
	q=5	0.001	-0.550	-1.024	0.004	2.489	0.006	-0.384	-0.554	-1.027
		(3.175)	(-2.173)	(-3.203)	(2.533)	(0.004)	(3.145)	(-1.632)	(-2.184)	(-3.203)
β_{cd}	q=2.5	0.069	0.136	0.075	0.023	0.025	0.030	0.174	0.136	0.074
		(1.465)	(5.691)	(2.978)	(0.675)	(0.359)	(0.653)	(6.812)	(5.653)	(2.955)
	q=5	0.078	0.133	0.071	0.058	0.968	0.027	0.169	0.133	0.071
		(1.779)	(5.730)	(2.944)	(1.813)	(0.058)	(0.673)	(6.894)	(5.715)	(2.927)
β_{cw}	q=2.5	-0.122	0.128	-0.017	0.018	-0.113	0.109	0.087	0.130	-0.016
		(-0.800)	(2.102)	(-0.217)	(0.222)	(-1.037)	(1.226)	(1.345)	(2.125)	(-0.203)
	q=5	-0.070	0.137	-0.012	0.063	-0.633	0.085	0.099	0.136	-0.012
		(-0.456)	(2.289)	(-0.164)	(0.804)	(-0.061)	(0.976)	(1.601)	(2.286)	(-0.156)
β_{cm}	q=2.5	0.656	0.653	0.793	0.503	0.682	0.371	0.679	0.651	0.792
		(2.211)	(9.149)	(10.088)	(3.442)	(4.541)	(1.966)	(10.050)	(9.121)	(10.043)
	q=5	1.228	0.666	0.818	0.845	10.021	0.788	0.690	0.666	0.817
		(6.486)	(9.918)	(11.084)	(8.960)	(0.982)	(7.668)	(10.738)	(9.897)	(11.055)

Table 4.3B: Predictive Regression for q=2.5 and q=5 for S&P 500 Futures (Cont.)

β_{jd}	q=2.5	0.115	-9.511	-2.187	0.143	0.091	-0.057	-33.777	-16.717	-2.734
		(0.347)	(-0.787)	(-0.143)	(1.798)	(0.859)	(-0.411)	(-1.493)	(-0.731)	(-0.091)
	q=5	21.144	-1047	3180	0.720	-0.083	-1.051	-5414	-2223	7005
		(0.204)	(-0.302)	(0.889)	(0.449)	(-0.424)	(-0.411)	(-0.915)	(-0.341)	(0.994)
β_{jw}	q=2.5	1.548	30.637	10.902	0.441	0.654	0.126	56.900	57.260	17.944
		(1.638)	(0.970)	(0.362)	(1.917)	(2.516)	(0.623)	(0.893)	(0.901)	(0.304)
	q=5	390	6163	127	5.415	1.885	5.192	3453	12940	-688
		(1.184)	(0.680)	(0.016)	(1.058)	(17.888)	(1.385)	(0.182)	(0.710)	(-0.043)
β_{jm}	q=2.5	2.528	26.803	51.586	0.582	0.300	1.038	65.742	56.053	106.019
		(2.366)	(0.726)	(1.389)	(2.104)	(0.863)	(2.748)	(0.999)	(0.755)	(1.416)
	q=5	263	6132	10073	1.719	-0.446	3.727	22752	12109	20505
		(0.817)	(0.590)	(1.056)	(0.348)	(-5.091)	(0.858)	(1.104)	(0.576)	(1.051)
R^2_{in}	q=2.5	0.376	0.372	0.333	0.463	0.452	0.418	0.452	0.434	0.383
	q=5	0.365	0.368	0.333	0.455	0.446	0.414	0.451	0.434	0.383
R^2_{out}	q=2.5	0.318	0.201	0.033	0.364	0.260	0.039	0.390	0.297	0.032
	q=5	0.244	0.167	0.027	0.345	0.241	0.037	0.390	0.296	0.032

* See notes in Table 4.3A.

Table 4.3C: Predictive Regression for q=2.5 and q= 5 for S&P 500 Futures*

		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
β_0	q=2.5	0.001	0.000	0.001	0.006	0.006	0.008	-0.569	-0.748	-1.279
		(2.950)	(1.648)	(2.055)	(3.550)	(2.901)	(3.284)	(-2.038)	(-2.468)	(-3.357)
	q=5	0.001	0.000	0.001	0.004	0.004	0.006	-0.415	-0.582	-1.026
		(3.221)	(2.326)	(2.607)	(2.727)	(2.561)	(3.128)	(-1.746)	(-2.262)	(-3.156)
β_{cd}	q=2.5	0.067	0.004	-0.072	0.019	0.023	0.031	0.173	0.136	0.074
		(1.477)	(0.030)	(-1.135)	(0.583)	(0.335)	(0.671)	(6.758)	(5.646)	(2.952)
	q=5	0.072	0.020	-0.074	0.056	0.054	0.028	0.168	0.133	0.071
		(1.845)	(0.182)	(-1.300)	(1.778)	(0.940)	(0.712)	(6.859)	(5.733)	(2.958)
β_{cw}	q=2.5	-0.102	-0.239	0.407	0.018	-0.110	0.113	0.094	0.133	-0.011
		(-0.633)	(-1.439)	(2.755)	(0.217)	(-1.034)	(1.291)	(1.462)	(2.172)	(-0.141)
	q=5	-0.040	-0.185	0.407	0.067	-0.056	0.084	0.104	0.140	-0.013
		(-0.235)	(-1.132)	(2.757)	(0.834)	(-0.557)	(0.980)	(1.673)	(2.333)	(-0.174)
β_{cm}	q=2.5	0.578	1.153	0.713	0.509	0.680	0.364	0.668	0.643	0.782
		(1.713)	(3.750)	(2.468)	(3.549)	(4.638)	(1.943)	(9.919)	(8.889)	(9.762)
	q=5	1.160	1.504	1.062	0.831	0.951	0.793	0.682	0.659	0.818
		(5.209)	(8.904)	(6.022)	(8.645)	(10.120)	(8.485)	(10.528)	(9.598)	(10.980)
β_{jd}^{-}	q=2.5	-1.116	0.419	1.836	-0.303	-0.114	0.082	-144	89	75
		(-0.659)	(0.395)	(2.138)	(-1.002)	(-0.317)	(0.263)	(-1.037)	(0.784)	(0.638)
	q=5	-526	8	532	-5.807	-4.136	3.523	-16818	-7675	39133
		(-0.785)	(0.026)	(1.961)	(-0.831)	(-0.700)	(0.667)	(-0.426)	(-0.312)	(1.644)

Table 4.3C: Predictive Regression for q=2.5 and q= 5 for S&P 500 Futures (Cont.)*

β^-_{jw}	q=2.5	0.523	-0.579	0.404	0.002	0.588	0.608	214	-130	-173
		(0.111)	(-0.241)	(0.082)	(0.002)	(0.756)	(0.699)	(0.449)	(-0.340)	(-0.412)
	q=5	400	245	328	-2.498	16.943	11.885	-22222	58161	-39700
		(0.292)	(0.347)	(0.215)	(-0.144)	(1.479)	(0.954)	(-0.199)	(0.747)	(-0.433)
β^-_{jm}	q=2.5	14.004	9.028	-3.694	2.772	1.644	0.097	946	989	1155
		(1.682)	(0.798)	(-0.495)	(1.435)	(0.742)	(0.048)	(1.217)	(1.063)	(1.262)
	q=5	3597	2303	-1813	47.313	18.591	-22.755	277860	154197	28836
		(1.692)	(0.825)	(-0.916)	(1.689)	(0.572)	(-0.717)	(1.530)	(0.735)	(0.122)
β^+_{jd}	q=2.5	1.272	0.001	-1.536	0.460	0.212	-0.141	113.345	-109.981	-81.229
		(0.755)	(0.001)	(-1.547)	(1.595)	(0.574)	(-0.481)	(0.817)	(-0.884)	(-0.690)
	q=5	567.860	103.841	-397.293	6.768	4.195	-4.735	11699	5426	-33271
		(0.833)	(0.331)	(-1.310)	(1.018)	(0.658)	(-0.897)	(0.293)	(0.190)	(-1.373)
β^+_{jw}	q=2.5	0.893	2.094	-0.688	0.445	0.062	-0.498	-172.658	185.572	187.449
		(0.184)	(0.792)	(-0.131)	(0.373)	(0.089)	(-0.551)	(-0.362)	(0.493)	(0.434)
	q=5	-79.637	108.397	-441.164	7.847	-7.332	-6.750	22772	-48575	40677
		(-0.056)	(0.143)	(-0.274)	(0.440)	(-0.785)	(-0.507)	(0.208)	(-0.656)	(0.428)
β^+_{jm}	q=2.5	-11.313	-7.911	6.100	-2.226	-1.355	0.968	-875	-934	-1049
		(-1.438)	(-0.726)	(0.784)	(-1.167)	(-0.623)	(0.462)	(-1.131)	(-1.013)	(-1.163)
	q=5	-3255.071	-2218.650	2196.218	-45.618	-20.656	26.622	-252245	-138328	-8874
		(-1.634)	(-0.795)	(1.066)	(-1.648)	(-0.628)	(0.813)	(-1.469)	(-0.678)	(-0.038)
R^2_{in}	q=2.5	0.378	0.373	0.335	0.464	0.452	0.418	0.480	0.434	0.384
	q=5	0.372	0.369	0.335	(0.456	0.447	.0415	0.452	0.434	0.383
R^2_{out}	q=2.5	0.342	0.212	0.03	0.373	0.263	0.038	0.391	0.298	0.033
	q=5	0.249	0.169	0.027	(0.352	0.246	0.035	0.390	0.297	0.032

*See notes in Table 4.3A.

Table 4.3D: Predictive Regression for q=2.5 and q = 5 for S&P 500 Futures*

		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
β_{cd}	q=2.5	0.001	0.000	0.001	0.006	0.005	0.008	-0.519	-0.702	-1.228
		(3.035)	(1.737)	(2.107)	(3.449)	(2.885)	(3.285)	(-1.902)	(-2.369)	(-3.290)
	q=5	0.001	0.000	0.001	0.003	0.004	0.006	-0.380	-0.550	-1.024
		(3.171)	(2.345)	(2.468)	(2.484)	(2.451)	(3.135)	(-1.620)	(-2.173)	(-3.203)
β_{cw}	q=2.5	0.066	0.004	-0.072	0.019	0.023	0.032	0.174	0.136	0.075
		(1.448)	(0.031)	(-1.090)	(0.575)	(0.338)	(0.680)	(6.807)	(5.688)	(2.977)
	q=5	0.073	0.019	-0.078	0.056	0.055	0.028	0.169	0.133	0.071
		(1.771)	(0.170)	(-1.319)	(1.780)	(0.947)	(0.710)	(6.893)	(5.730)	(2.944)
β_{cm}	q=2.5	-0.122	-0.251	0.421	0.013	-0.115	0.111	0.087	0.129	-0.017
		(-0.812)	(-1.419)	(2.790)	(0.163)	(-1.053)	(1.242)	(1.348)	(2.104)	(-0.216)
	q=5	-0.067	-0.197	0.429	0.058	-0.063	0.088	0.098	0.137	-0.012
		(-0.443)	(-1.151)	(2.944)	(0.741)	(-0.602)	(1.005)	(1.594)	(2.289)	(-0.164)

Table 4.3D: Predictive Regression for q=2.5 and q = 5 for S&P 500 Futures (Cont.)

	q=2.5	0.701	1.230	0.686	0.522	0.695	0.370	0.681	0.653	0.793
β_{jd}		(2.472)	(4.925)	(2.215)	(3.649)	(4.681)	(1.934)	(10.058)	(9.152)	(10.092)
	q=5	1.250	1.569	1.029	0.858	0.973	0.784	0.691	0.666	0.818
		(6.840)	(10.869)	(5.276)	(9.293)	(9.994)	(7.693)	(10.764)	(9.918)	(11.084)
	q=2.5	0.072	0.214	0.165	0.078	0.048	-0.030	-16.138	-9.318	-2.037
β_{jw}		(0.434)	(1.317)	(0.700)	(1.970)	(0.898)	(-0.431)	(-1.431)	(-0.777)	(-0.134)
	q=5	18.273	56.886	71.902	0.450	-0.002	-0.591	-2596.867	-1046.806	3180.150
		(0.352)	(1.171)	(1.002)	(0.571)	(-0.002)	(-0.449)	(-0.876)	(-0.302)	(0.889)
	q=2.5	0.788	0.793	-0.186	0.229	0.330	0.057	27.695	30.419	10.641
		(1.628)	(1.775)	(-0.543)	(1.930)	(2.588)	(0.549)	(0.867)	(0.967)	(0.355)
	q=5	194.636	195.391	-76.421	2.935	4.928	2.420	2031.944	6162.489	127.111
		(1.158)	(1.469)	(-0.808)	(1.108)	(1.897)	(1.230)	(0.215)	(0.680)	(0.016)
	q=2.5	1.180	0.460	1.236	0.271	0.139	0.528	32.365	26.705	51.504
β_{jm}		(2.238)	(0.893)	(1.976)	(1.947)	(0.805)	(2.740)	(0.985)	(0.726)	(1.391)
	q=5	114.426	2.211	211.490	0.499	-1.222	2.100	10776.175	6131.779	10073.197
		(0.714)	(0.017)	(1.346)	(0.197)	(-0.435)	(0.991)	(1.066)	(0.590)	(1.056)
R^2_{in}	q=2.5	0.376	0.372	0.335	0.463	0.451	0.418	0.452	0.434	0.384
	q=5	0.368	0.368	0.335	0.455	0.445	0.415	0.451	0.434	0.383
R^2_{out}	q=2.5	0.315	0.199	0.033	0.368	0.262	0.040	0.39	0.297	0.032
	q=5	0.244	0.167	0.027	0.343	0.241	0.037	0.389	0.296	0.032

*See notes in Table 4.3A.

Table 4.4: Diebold - Mariano Predictive Tests for S&P 500 futures*

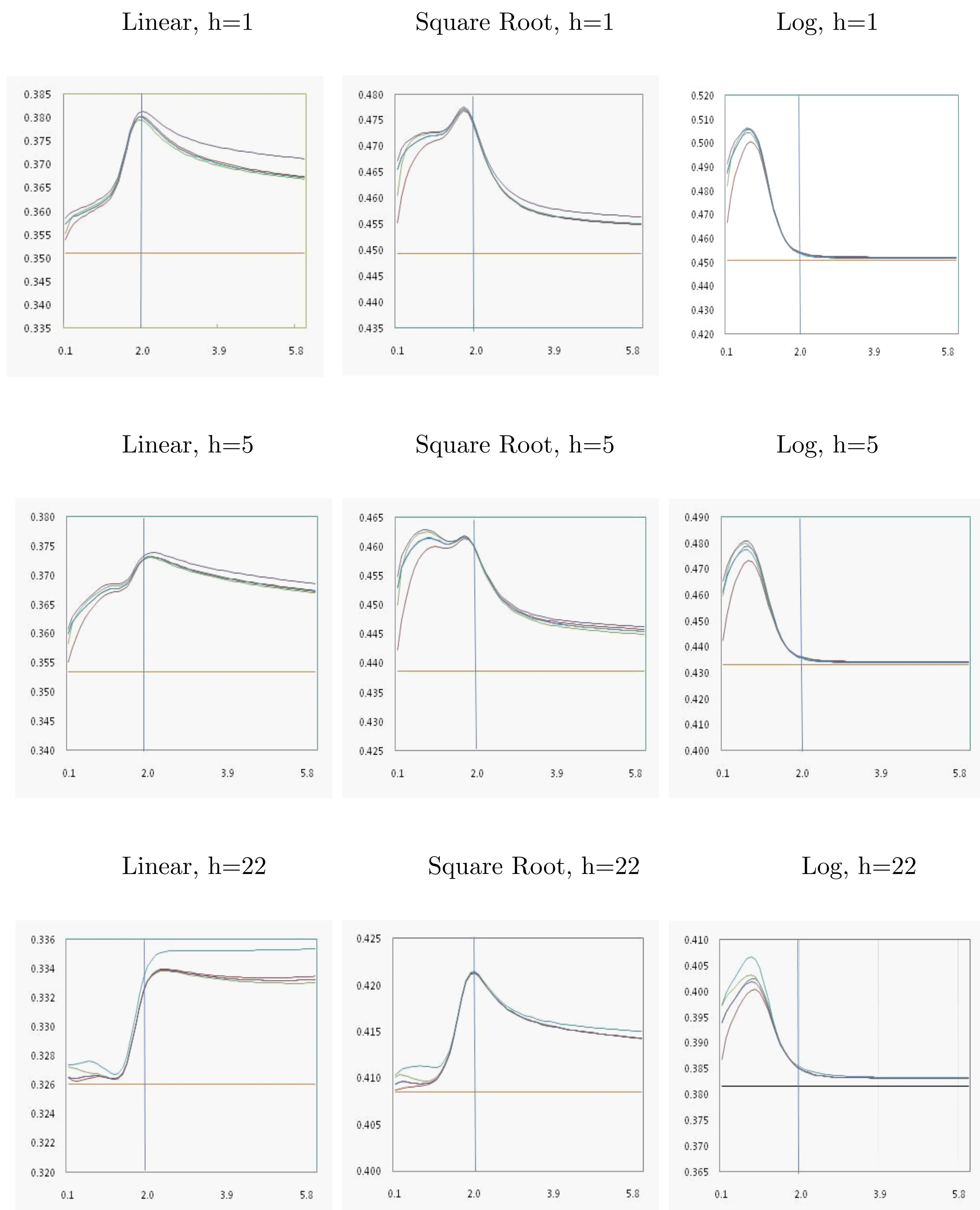
		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
Panel A: Recursive Scheme										
HAR-C-PV	DM Stat	5.30	2.75	-3.04	3.42	2.60	3.25	2.08	2.84	2.29
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.20	2.50	2.50
	$.q_s$	4.40	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
HAR-C-PDJ	DM Stat	4.05	1.61	-2.40	3.20	2.90	3.41	2.51	3.20	2.64
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.20	2.50	2.50
	$.q_s$	4.30	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
HAR-C-PUJ	DM Stat	5.80	2.36	-1.69	3.20	2.05	3.30	2.20	1.49	1.31
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.10	2.50	2.50
	$.q_s$	4.40	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
HAR-C-PDUJ	DM Stat	6.16	2.89	-1.75	3.51	2.31	3.16	2.19	2.28	1.63
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.10	2.50	2.50
	$.q_s$	4.40	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
Panel B: Rolling Scheme										
HAR-C-PV	DM Stat	6.17	3.19	-3.41	3.29	2.55	3.19	0.99	2.87	2.49
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.40	2.50	2.50
	$.q_s$	4.20	4.70	5.40	6.00	6.00	6.00	6.00	6.00	6.00
HAR-C-PDJ	DM Stat	4.28	2.15	-2.67	3.09	2.84	3.32	-3.18	3.30	2.83
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.40	2.50	2.50
	$.q_s$	4.20	4.70	5.40	6.00	6.00	6.00	2.50	6.00	6.00
HAR-C-PUJ	DM Stat	6.56	2.86	-1.89	3.04	1.99	3.20	0.92	1.46	1.53
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.30	2.50	2.50
	$.q_s$	4.20	4.70	5.40	6.00	6.00	6.00	6.00	6.00	6.00
HAR-C-PDUJ	DM Stat	7.13	3.40	-1.75	3.35	2.26	3.11	2.11	2.24	1.88
	q_b	2.50	2.50	2.50	2.50	2.50	2.50	3.10	2.50	2.50
	$.q_s$	4.20	4.70	5.30	6.00	6.00	6.00	6.00	6.00	6.00

Panel C: Fixed Scheme

		Linear Models			Square Root Models			Log Models		
		h=1	h=5	h=22	h=1	h=5	h=22	h=1	h=5	h=22
HAR-C-PV	DM Stat	6.17	3.11	-3.23	4.27	3.15	3.41	-4.82	3.32	2.35
	q_b	2.5	2.5	2.5	2.5	2.5	2.5	3.5	2.5	2.5
	.q_s	4.3	4.8	5.8	6.0	6.0	6.0	2.5	6.0	6.0
HAR-C-PDJ	DM Stat	4.23	1.98	-2.46	3.67	3.25	3.52	-3.75	3.46	2.66
	q_b	2.5	2.5	2.5	2.5	2.5	2.5	3.5	3.5	3.5
	.q_s	4.3	4.3	4.3	6.0	6.0	6.0	6.0	6.0	6.0
HAR-C-PUJ	DM Stat	6.82	2.76	-1.84	3.86	2.43	3.37	0.21	1.80	1.23
	q_b	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	.q_s	4.3	5.8	5.8	6.0	6.0	6.0	6.0	6.0	6.0
HAR-C-PDUJ	DM Stat	7.13	3.05	-1.80	4.37	2.87	3.34	1.73	3.19	1.85
	q_b	2.5	2.5	2.5	2.5	2.5	2.5	3.2	2.5	2.5
	.q_s	4.3	4.9	3.4	6.0	6.0	6.0	6.0	6.0	6.0

* The table reports Diebold-Mariano (1995) test statistics, calculated using Hac estimators with auto-correlated lags up to 44 as discussed in section 4.4.2, for linear, square root and log specifications HAR- C-PV, HAR-C-PDJ, HAR-C-DUJ, HAR-C-AJP at forecast horizon h=1,5,22 respectively. For each specification, the entries in the first rorws, DM Stat are statistics. The entries in the second row, $q_b(2.5 <= q_b <= 6)$ is the value of q that yields the highest R-square and $q_s(2.5 <= q_b <= 6)$ is the value of q that yields the smallest R-square.

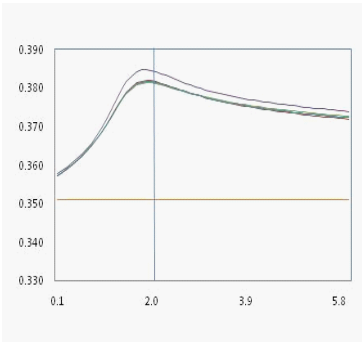
Figure 4.1: In-sample R^2 for S&P 500 futures, No Jump Test*



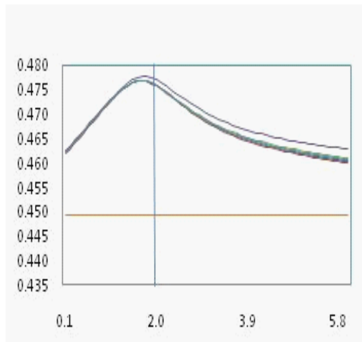
* The figure depicts 9 plots in-sample R-square of 6 specifications summarized in Table 4.2 for all linear, square root and log models across forecast horizon h=1, 5 and 22. For each plot, the vertical axis represents R-square, with range from 0 to 1. The horizontal represents the order q with range from 0 to 6 , i.e, q=0.1,0.2,...,5.9,6. The orange line plots R-square of HAR-RV-C. The purple curve plots R-square of HAR-RV-C-PDUJ model. The dark purple represents HAR-RV-C-PUJ and the light green represents HAR-RV-C-PDJ and dark blue is for HAR-RV-C-PV. The light blue plots R-square of HAR-RV-C-AJP.

Figure 4.2: In-sample R^2 for S&P 500 futures, with Jump Test*

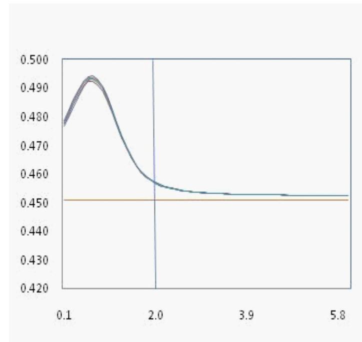
Linear, h=1



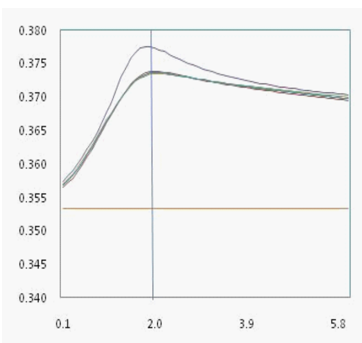
Square Root, h=1



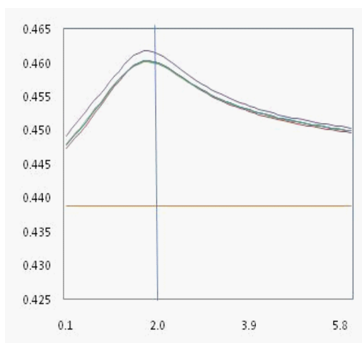
Log, h=1



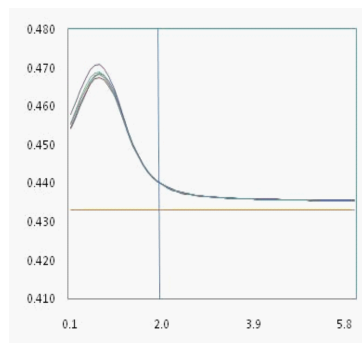
Linear, h=5



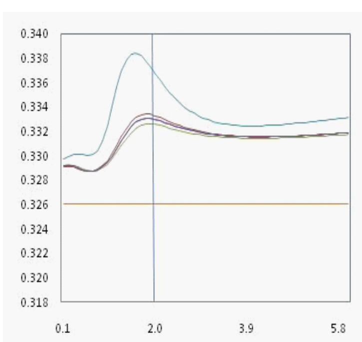
Square Root, h=5



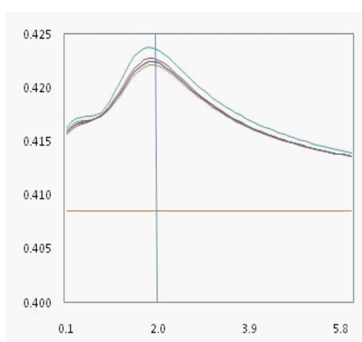
Log, h=5



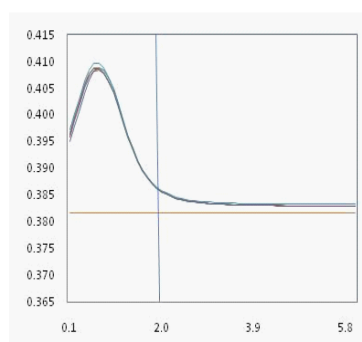
Linear, h=22



Square Root, h=22



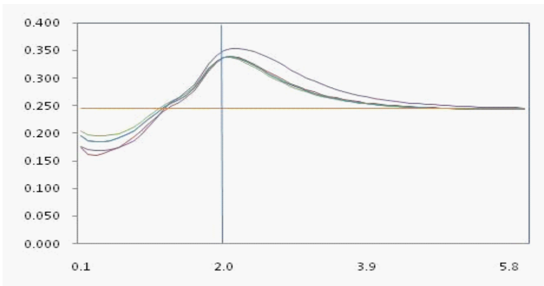
Log, h=22



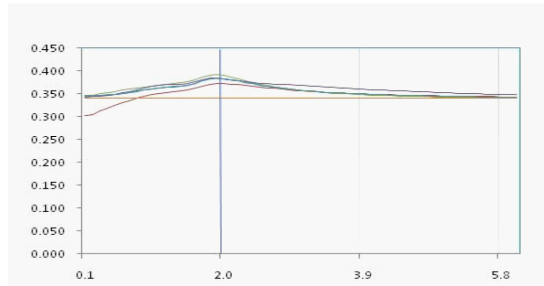
* The Figure depicts 9 plots in-sample Rš where jump tests are taken into account. See footnote in Figure 4.1.

Figure 4.3A: Out of sample R^2 for S&P 500 futures, No Jump Test*

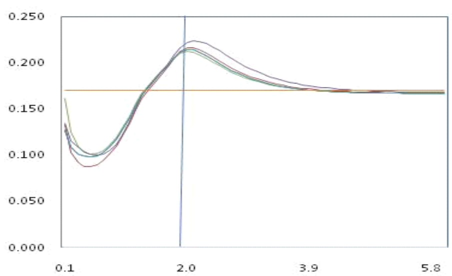
Linear, h=1



Square Root, h=1



Linear, h=5



Square Root, h=5

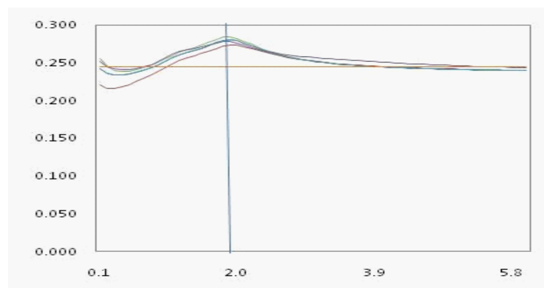
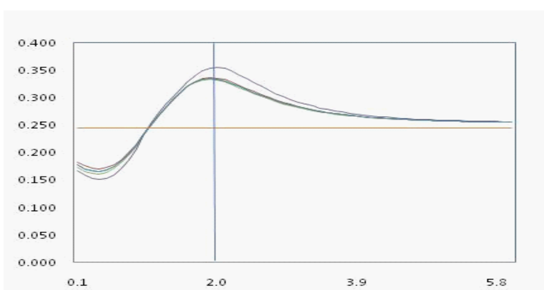
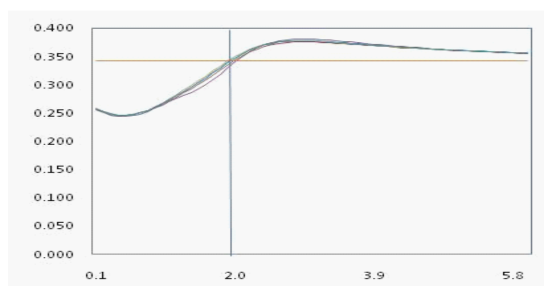


Figure 4.3B: Out of sample R^2 for S&P 500 Futures, with Jump Test*

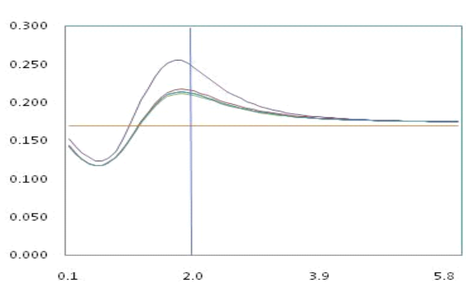
Linear, h=1



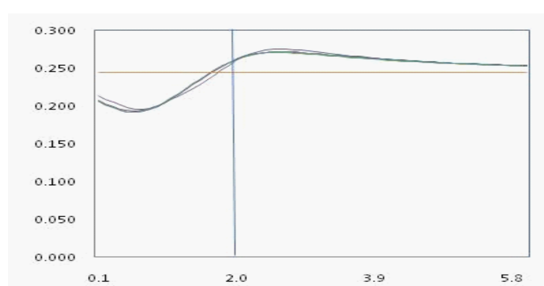
Square Root, h=1



Linear, h=5



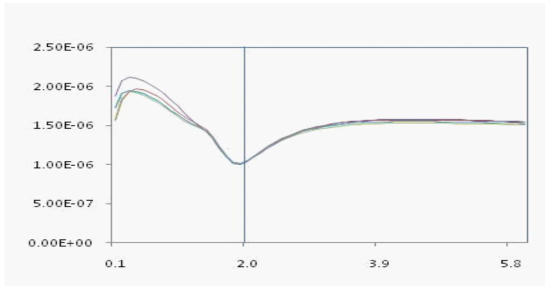
Square Root, h=5



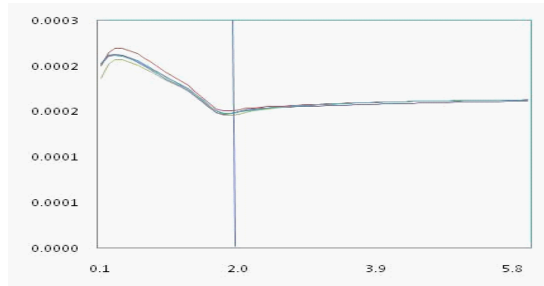
* The Figure 4.3A depicts 4 plots in-sample R-squares of 6 specifications summarized in Table 4.2 for linear, square root across forecast horizon h=1, 5 See footnote in Figure 4.1 for further details. Figure 4.3B takes jump tests into account.

Figure 4.4A: Mean Square Errors for S&P 500 futures, No Jump Test*

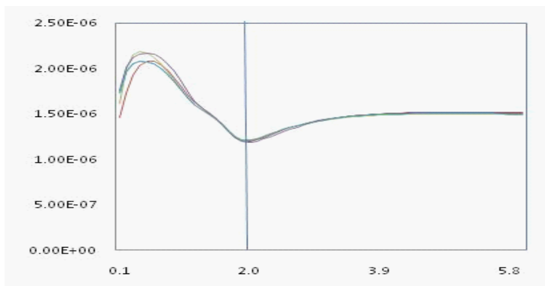
Linear, h=1



Square Root, h=1



Linear, h=5



Square Root, h=5

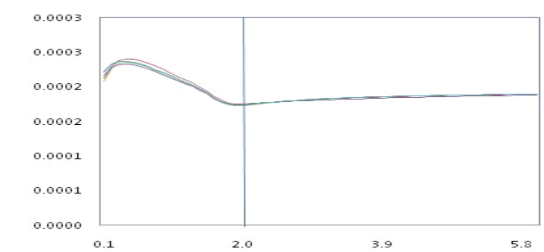
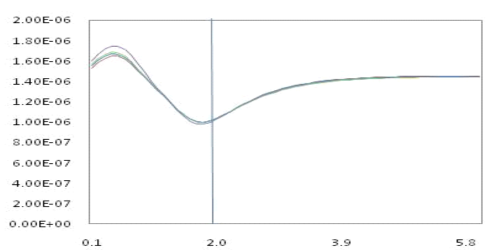
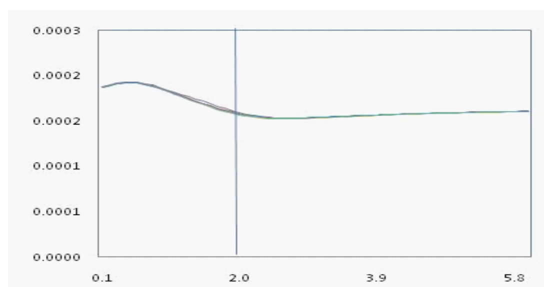


Figure 4.4B:Mean Square Errors for S&P 500 futures, Jump Test*

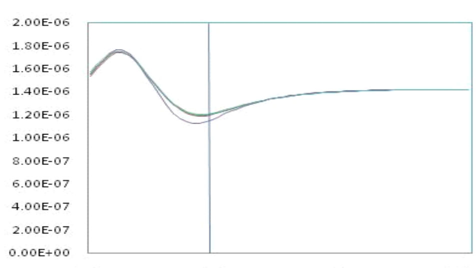
Linear, h=1



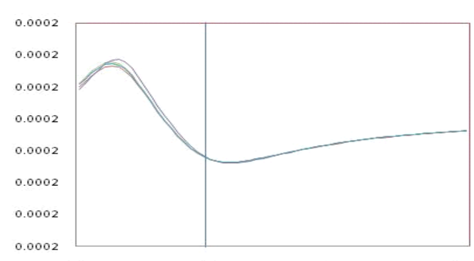
Square Root, h=1



Linear, h=5



Square Root, h=5

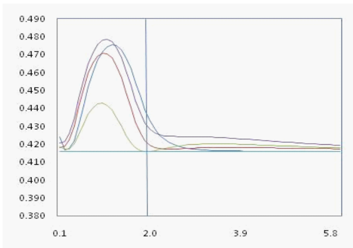


* The figure 4.4A depicts 4 plots mean square errors of specifications from 2 to 6 summarized in Table 4.2 for linear, square root across forecast horizon h=1, 5. Figure 4.4B takes jump tests into account. See footnote in Figure 4.1 for further details.

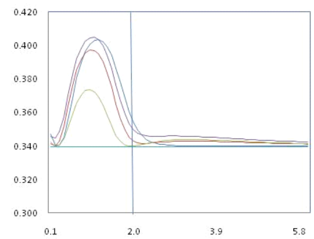
Figure 4.5: R^2 for Dow 30 components for Square Root Models, No Jump Test*

Panel A: Intel

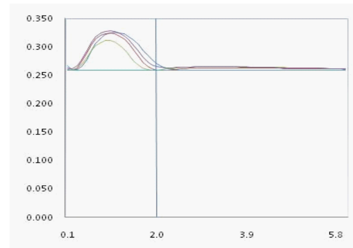
Square Root, h=1



Square Root, h=5

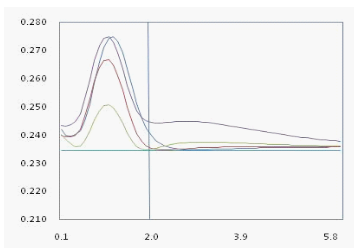


Square Root, h=22

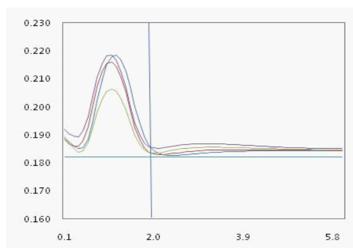


Panel B: City

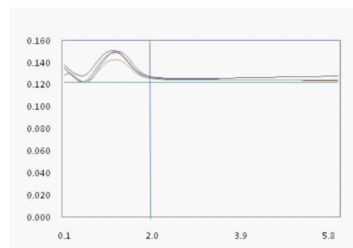
Square Root, h=1



Square Root, h=5

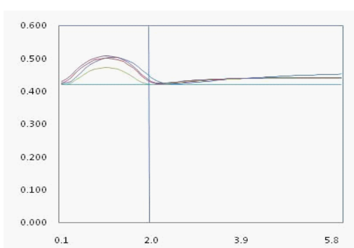


Square Root, h=22

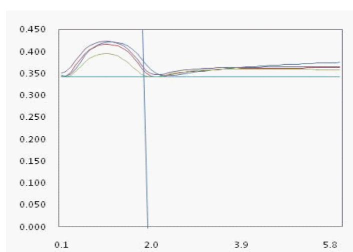


Panel C: Microsoft

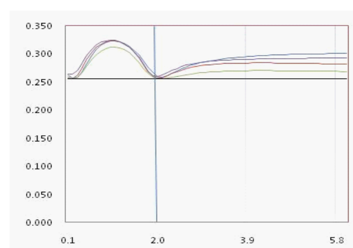
Square Root, h=1



Square Root, h=5

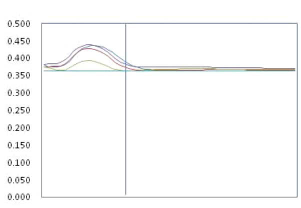


Square Root, h=22

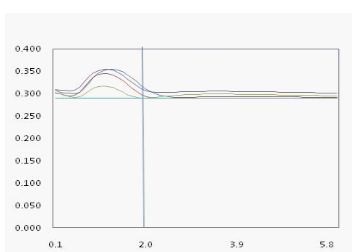


Panel D: Home Depot

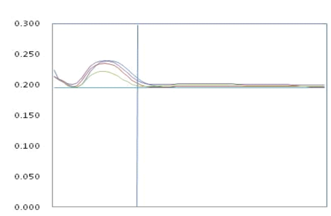
Square Root, h=1



Square Root, h=5



Square Root, h=22



* The Panel A,B,C,D in Figure 4.5 depict the in-sample R-squares for the 4 representative stocks in Down 30 components. The four stocks are Intel, Citi, MSFT and Home Depot, respectively and the models are square root at daily, weekly and monthly forecast horizon. See footnotes in Figure 4.1A for further details about the plot.

References

Aït-Sahalia, Y., (1996), Testing Continuous Time Models of the Spot Interest Rate, *Review of Financial Studies*, 9, 385-426.

Aït-Sahalia, Y. (2002), Telling from discrete data whether the underlying continuous-time model is a diffusion. *Journal of Finance* 57, 2075–2112.

Aït-Sahalia, Y., (2002), Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed Form Approximation Approach, *Econometrica*, 70, 223-262.

Aït-Sahalia, Y., (2007), Estimating Continuous-Time Models Using Discretely Sampled Data, Econometric Society World Congress Invited Lecture, in Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, edited by Richard Blundell, Persson Torsten and Whitney K. Newey, Econometric Society Monographs, Cambridge University Press.

Aït-Sahalia, Y., J. Fan and H. Peng, (2009), Nonparametric Transition-Based Tests for Diffusions, *Journal of the American Statistical Association*, 104, 1102-1116.

Aït-Sahalia, Y. and J. Jacod (2009a), Testing for Jumps in a Discretely Observed Process. *Annals of Statistics* 37, 184-222.

Aït-Sahalia, Y. and J. Jacod (2009b), Estimating the Degree of Activity of Jumps in High Frequency Data. *Annals of Statistics* 37, 2202-2244.

Aït-Sahalia, Y. and Jacod, J. (2011), Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data. *Journal of Economic Literature*, forthcoming.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005), How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* 18, 351–416.

Altissimo, F. and A. Mele, (2009), Simulated Nonparametric Estimation of Dynamic Models with Application in Finance, *Review of Economic Studies*, 76, 413-450.

Andersen, T.G., L. Benzoni and J. Lund, (2004), Stochastic Volatility, Mean Drift, and Jumps in the Short-Term Interest Rate, Working Paper, Northwestern University.

Andersen, T.G. and T. Bollerslev (1998), Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, *Journal of Finance* 53, 219-265.

Andersen, T.G., T. Bollerslev, and F.X. Diebold (2007), Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, *Review of Economics and Statistics* 89, 701-720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001), The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55. Correction published in 2003, volume 98, page 501.

Andersen, T.G., T. Bollerslev, F.X. Diebold, P. Labys (2003), Modeling and Forecasting Realized Volatility, *Econometrica* 71, 579 625.

Andersen, T.G., and J. Lund, (1997), Estimating Continuous Time Stochastic Volatility Models of Short Term Interest Rates, *Journal of Econometrics*, 72, 343-377.

Andersen, T. G., D. P. Dobrev, and E. Schaumburg (2008), Jump robust volatility estimation, Working Paper.

Andrews, D.W.K., (1993), An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Variables, *Econometric Reviews*, 12, 183-216.

Andrews, D.W.K., (1997), A Conditional Kolmogorov Test, *Econometrica,* 65, 1097-1128.

Ang, A., J. Chen, and Y. Xing. (2006), Downside risk, *Review of Financial Studies* 19, 1191–1239.

Bai, J., (2003), Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics,* 85, 531-549.

Bakshi, G., C. Cao, and Z. Chen (1997), Empirical Performance of Alternative Option Pricing Models, *Journal of Finance* 52, 2003–2049.

Bandi, F., (2002), Short-Term Interest Rate Dynamics: A Spatial Approach, *Journal of Financial Economics,* 65, 73-110.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006), A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales, in: **From Stochastic Analysis to Mathematical Finance**, Festschrift for Albert Shiryaev, Springer and Verlag, Y. Kabanovand R. Lipster (eds.): New York: U.S.A.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008), Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica*, 76, 1481-1536.

Barndorff-Nielsen O.E., S. Kinnebrock, and N. Shephard (2010), Measuring Downside Risk – Realised Semivariance, in Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle, ed. by T. Bollerslev, J. Russell, and M. Watson. Oxford University Press.

Barndorff-Nielsen, O.E. and N. Shephard (2002), Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models, *Journal of Royal Statistical Society*, 64, 253–80.

Barndorff-Nielsen, O.E. and N. Shephard (2004), Power and Bipower Variation with Stochastic Volatility and Jumps (with discussion), *Journal of Financial Econometrics* 2, 1-48.

Barndorff-Nielsen, O.E. and N. Shephard (2006), Econometrics of Testing for Jumps in Financial Economics using Bipower Variation, *Journal of Financial Econometrics* 4, 1–30.

Barndorff-Nielsen, O.E., N. Shephard, and M. Winkel (2006), Limit Theorems for Multipower Variation in the Presence of Jumps. *Stochastic Processes and Their Applications* 116, 796-806.

Barone-Adesi, G., R.F. Engle, and L. Mancini (2008), A GARCH Option Pricing Model with Filtered Historical Simulation. *Review of Financial Studies* 21, 1223–1258.

Bates, D.S. (1996), Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies* 9, 69-107.

Bates, D.S. (2000), Post-87 Crash Fears in the S&P 500 Futures Option Market. *Journal of Econometrics* 94, 181-238

Bates, D.S. (2003), Empirical Option Pricing: a Retrospection. *Journal of Econometrics* 116, 387–404.

Beckers, S. (1980), The Constant Elasticity of Variance Model and US Implications for Option Pricing. *Journal of Finance* 35, 661-673.

Bhardwaj, G., V. Corradi and N.R. Swanson, (2008), A Simulation Based Specification Test for Diffusion Processes, *Journal of Business and Economic Statistics*, 26, 176-93.

Billingsley, P. (1968), **Convergence of Probability Measures**. New York: Wiley.

Black, F. (1976), Studies in Stock Prie Volatility Changes. *In Proceedings of the 1976 Meetings of the Business and Economic Statistics Section*, 177-181. American Statistical Association.

Black, F. and M. Scholes (1973), The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 637-654.

Bollerslev, T. (1986), Generalized Autogressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327.

Bollerslev, T., R. Chou, and K. Kroner (1992), ARCH Modeling in Finance: A Review of Theory and Empirical Evidence. *Journal of Econometrics* 52, 5-59.

Bollerslev, T., R.F. Engle, and D.B. Nelson (1994), ARCH Models. In R. Engle and D. McFadden (eds.), **Handbook of Econometrics**, vol. 4. Amsterdam: North-Holland.

Bollerslev, T., R.F. Engle, and J.M. Wooldridge (1988), A Capital Asset Pricing Model with Time Varying Covariances. *Journal of Political Economy* 95, 116–131.

Bollerslev, T., U. Kretschmer, C. Pigorsch and G. Tauchen (2007), A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects, *Journal of Econometrics* 150, 151–166.

Bollerslev, T., T. Law, and G. Tauchen (2008), Risk, Jumps, and Diversification, *Journal of Econometrics* 144, 234–256.

Bollerslev, T., G. Tauchen, and H. Zhou (2009), Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* 22, 4463–4492.

Bontemps, C., and N. Meddahi, (2005), Testing Normality: a GMM Approach, *Journal of Econometrics*, 124, 149-186.

Brennan, M.J., and E.S. Schwartz (1979), A Continuous-Time Approach to the Pricing Bonds, *Journal of Banking and Finance* 3, 133-155.

Britten-Jones, M. and A. Neuberger (2000), Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance* 55, 839-866.

Cai, L. and N.R. Swanson, (2011), An Empirical Assessment of Spot Rate Model Stability. *Journal of Empirical Finance*, 18, 743–764.

Carr, P., H. Geman, D.B. Madan, and M. Yor (2002), The Fine Structure of Asset Returns: an Empirical Investigation. *Journal of Business* 75, 305–332.

Carr, P. and R. Lee (2003), Trading Autocorrelation, Manuscript, New York University.

Carr, P. and R. Lee (2009), Volatility Derivatives, Annual Review of Financial Economics 1, 1-21.

Carr, P. and D. Madan (1998), Towards a Theory of Volatility Trading. Risk Books, 417-427 (Chapter 29).

Carr, P. and L. Wu (2003), What Type of Process Underlies Options? A Simple Robust Test, *Journal of Finance* 58, 2581–2610.

Carr, P. and L. Wu (2009), Variance Risk Premia, *Review of Financial Studies* 22, 1311-1341.

Carr, P., Geman, H., Madan, D.B., Yor, M., (2002), The fine structure of asset returns: an empirical investigation. *Journal of Business* 75, 305–332.

Carr, P. and Wu, L. (2003), What type of process underlies options? A simple robust test, *Journal of Finance* 58, 2581–2610.

Chacko, G. and L.M. Viceira (2003), Spectral GMM Estimation of Continuous-Time Processes, *Journal of Econometrics* 116, 259-292.

Chan, K.C., G. A. Karolyi, F. A. Longstaff and A. B. Sanders, (1992), An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance,* 47, 1209-1227.

Chen, L, (1996), Stochastic Mean and Stochastic Volatility - A Three-Factor Model of Term Structure of Interest Rates and its Application to the Pricing of Interest Rate Derivatives, Blackwell Publishers, Oxford, UK.

Chen, B. and Y. Hong, (2005), Diagnostics of Multivariate Continuous-Time Models with Application to Affine Term Structure Models, Working Paper, Cornell University.

Chernov, M. and E. Ghysels (2002), A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, *Journal of Financial Economics* 56, 407-458.

Christie, A. (1982), The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* 10, 407–432.

Clark, P. K. (1973), A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41, 135-156.

Clements, M.P. and J. Smith, (2000), Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment, *Journal of Forecasting*, 19, 255-276.

Clements, M.P. and J. Smith, (2002), Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches, *International Journal of Forecasting*, 18, 397-407.

Cont, R. and C. Mancini (2007), Nonparametric Tests for Analyzing the Fine Structure of Price Fluctuations, Working Paper, Columbia University Center for Financial Engineering.

Corradi, V. (2000), Reconsidering the Continuous Time Limit of the GARCH(1,1) Process, *Journal of Econometrics* 96, 145–153.

Corradi, V., W. Distaso and N.R. Swanson (2009), Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures, *Journal of Econometrics* 150, 119–138.

Corradi, V., W. Distaso, and N. Swanson (2011), Predictive Infererence for Integrated Volatility. *Journal of the American Statistical Association*, 106(496), 1496-1512.

Corradi, V. and N.R. Swanson, (2005), A Bootstrap Specification Test for Diffusion Processes, *Journal of Econometrics*, 124, 117-148.

Corradi, V. and N.R. Swanson, (2006), Predictive Density and Conditional Confidence Interval AccuracyTests, *Journal of Econometrics*, 135, 187-228.

Corradi, V. and N.R. Swanson, (2007a), Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes, *International Economic Review*, February 2007, 48, 67-109.

Corradi, V. and N.R. Swanson, (2007b), Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes, *International Economic Review*, 48, 67-109.

Corradi, V. and N.R. Swanson, (2011), Predictive Density Construction and Accuracy Testing with Multiple Possibly Misspecified Diffusion Models, *Journal of Econometrics,* 161,304 - 324.

Corsi, F. (2004), A Simple Long Memory Model of Realized Volatility, Working Paper, University of Southern Switzerland.
Corsi, F., Pirino, D., Reno, R, (2008). Volatility Forecasting: The Jumps do matter, Working Paper

Courtadon, G., (1982), The Pricing of Options on Default-Free Bonds, *Journal of Financial and Quantitative Analysis*, 17, 75-100.

Cox, J.C., J.E. Ingersoll and S.A. Ross, (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-407.

Cox, J.C. and S. Ross (1976), The Valuation of Options for Alternative Stochastic Processes, *Journal of Financial Economics* 3, 145-166.

Dacorogna, M.M., R. Gencay, U. Müller, R.B. Olsen, and O.B. Pictet (2001), **An Introduction to High-Frequency Finance**. Academic Press: London.

Dai, Q. and Kenneth J. Singleton, (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance,* 55, 1943-1978.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou. (1999), A Guide to Volatility and Variance Swaps, *Journal of Derivatives* 6, 9–32.

Diebold, F.X., T. Gunther and A.S. Tay, (1998), Evaluating Density Forecasts with Applications to Finance and Management, *International Economic Review*, 39, 863-883.

Diebold, F.X., J. Hahn and A.S. Tay, (1999), Multivariate Density Forecast Evaluation and Calibrationin Financial Risk Management: High Frequency Returns on Foreign Exchange, *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X. and C. Li, (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, 130, 337-364.

Diebold, F.X. and R.S. Mariano, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X., A.S. Tay and K.D. Wallis, (1998), Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters, in Festschrift in Honor of C.W.J. Granger, eds. R.F. Engle and H. White, Oxford University Press, Oxford.

Ding, Z. and C. W. J. Granger (1996), Modeling volatility persistence of speculative returns: A new approach, *Journal of Econometrics* 73, 185–215.

Ding, Z., C.W.J. Granger, and R. F. Engle (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.

Duan, J.C., (2003), A Specification Test for Time Series Models by a Normality Transformation, Working Paper, University of Toronto.

Duffie, D. (2001), **Dynamic Asset Pricing Theory**, Princeton University Press: Princeton: NJ.

Duffie, D. and R. Kan, (1996), A Yield-Factor Model of Interest Rates. *Mathematical Finance,* 6:379-406.

Duffie, D., J. Pan, and K. Singleton, (2000), Transform Analysis and Asset Pricing for Affine Junp-Diffusions. *Econometrica*, 68, 1343–1376.

Duffie, D. and K. Singleton, (1993), Simulated Moment Estimation of Markov Models of Asset Prices, *Econometrica,* 61, 929-952.

Dumas, B., J. Fleming, and R. E. Whaley (1998), Implied Volatility Functions: Empirical Tests, *Journal of Finance* 53, 2059–2106.

Duong, D., and N.R. Swanson, (2010), Empirical Evidence on Jumps and Large Fluctuations in Individual Stocks. Working Paper, Rutgers University.

Duong, D., and N.R. Swanson, (2011), Volatility in Discrete and Continuous Time Models: A Survey with New Evidence on Large and Small Jumps, *Missing Data Methods: Advances in Econometrics,* 27B, 179-233.

Emanuel, D. and J. Mcbeth, (1982), Further Results on the Constant Elasticity of Variance Call Option Pricing Model, *Journal of Financial Quantitative Analysis,* 17, 533-554.

Engle, R.F. (1982), Autogressive Condititional Heteroskedasticity with Esimates of Variance of U.K. Inflation, *Econometrica* 50, 987-1008.

Engle, R.F., D.M. Lilien, and R.P. Robins (1987), Estimating Time-varying Risk Premia inthe Term Structure: The ARCH-M Model, *Econometrica* 55, 391–408.

Fama, F.E. (1965), The Behaviour of Stock Market Prices, *Journal of Business* 38, 34-105.

Fermanian, J.-D. and B. Salanié, (2004), A Nonparametric Simulated Maximum Likelihood Estimation Method, *Econometric Theory*, 20, 701-734.

Gallant, A.R. and G. Tauchen, (1996), Which Moments to Match, *Econometric Theory*, 12, 657-681.

Gallant, A.R., and G. Tauchen (1997), Estimation of Continuous Time Models for Stock Returns and Interest Rates, *Macroeconomic Dynamics,* 1, 135-168.

Garcia, R., E. Ghysels, and E. Renault (2010), The Econometrics of Option Pricing. In Y.Ait-Sahalia, L.Hansen (eds.), Handbook of financial econometrics, North Holland, Oxford and Amsterdam.

Ghysel, E. and B. Sohn (2009), Which Power Variation Predicts Volatility Well ?, *Journal of Empirical Finance*, 16, 686-700.

Glosten, L.R., 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* 48, 1779-1801.

Goncalves, S. and H. White, (2002), The Bootstrap of the Mean for Dependent Heterogeneous Arrays, *Econometric Theory*, 18, 1367-1384.

Granger, C.W. J. (2008), In praise of pragmatic econometrics. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: A Festschrift in honour of David F Hendry*, pp. 105–116. Oxford University Press.

Hall, A.R. and A. Inoue, (2003), The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models, J*ournal of Econometrics*, 361-394.

Hansen, B.E., (1996), Inference when a Nuisance Parameter is Not Identified Under the Null Hypothesis, *Econometrica*, 64, 413-430.

Hansen, P.R. and A. Lunde (2006), Realized Variance and Market Microstructure noise (with comments and rejoinder). *Journal of Business & Economic Statistics* 24, 127-218.

Harvey, A.C., E. Ruiz, and N. Shephard (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies* 61, 247-264.

Heston, S.L., (1993), A Closed Form Solution for Option with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6, 327-344.

Heston, S.L. and S. Nandi. (2000), A Closed-form GARCH Option Valuation Model, *Review of Financial Studies* 13, 585–625.

Hong, Y., (2001), Evaluation of out-of-sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Working Paper, Cornell University.

Hong, Y.M., and H.T. Li, (2005), Nonparametric Specification Testing for Continuous Time Models with Applications to Term Structure Interest Rates, *Review of Financial Studies*, 18, 37-84.

Hong, Y.M., H.T. Li, and F. Zhao, (2007), Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates, *Journal of Econometrics*, 141, 736-776.

Huang, X. and G. Tauchen (2005), The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics* 3, 456-499.

Hull, J. (1997), **Options, Futures, and Other Derivatives,** Englewood Cli s, NJ: Prentice Hall.

Hull, J., and A. White, (1987), The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, 42, 281-300.

Inoue, (2001), Testing for Distributional Change in Time Series, *Econometric Theory*, 17, 156-187.

Jacod, J. (2007), Statistics and High-Frequency Data. SEMSTAT Princeton University, Course Notes, Tech. Rep., Université de Paris - 6.

Jacod, J. (2008), Asymptotic Properties of Realized Power Variations and Related Functionals of Semimartingales, *Stochastic Processes and Their Applications* 118*, 517-559.*

Jiang, G.J., (1998), Nonparametric Modelling of US Term Structure of Interest Rates and Implications on the Prices of Derivative Securities, *Journal of Financial and Quantitative Analysis*, 33, 465-497.

Jiang, G.J. and R.C.A. Oomen (2008), Testing for Jumps when Asset Prices are Observed with Noise - A "Swap Variance" Approach, *Journal of Econometrics* 144, 352-370.

Jiang, G.J. and Y.S. Tian (2005), The Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies* 18, 1305-1342.

Karlin, S. and H.M. Taylor, (1981), **A Second Course in Stochastic Processes**. Academic Press, San Diego.

Kloeden, P.E. and E. Platen, (1999), **Numerical Solution of Stochastic Differential Equations**, Springer and Verlag, New York.

Kolmogorov A.N., (1933), Sulla Determinazione Empirica di una Legge di Distribuzione, Giornale dell'Istitutodegli Attuari, 4, 83-91.

Kristensen, D. (2010), Nonparametric Filtering of the Realized Spot Volatility: A Kernel-based Approach, *Econometric Theory* 26, 60-93.

Kristensen, D. and Y. Shin, (2008), Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood, CREATES Research Paper 2008-58, University of Aarhus and Columbia University.

Künsch H.R., (1989), The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics*, 17, 1217-1241.

Lee, S. and Mykland, P.A. (2008), Jumps in Financial Markets: A New Nonparametric Test and Jump Clustering. *Review of Financial Studies* 21, 2535-2563.

Liesenfeld, R. & J.F. Richard, (2003), Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance* 10, 505–531.

Liu, C. and Maheu, M.J. (2005), Modeling and Forecasting Realized Volatility: The Role of Power Variation, working paper.

Lynch, D. and N. Panigirtzoglou (2003), Option Implied and Realized Measures of Variance, Working Paper. Monetary Instruments and Markets Division, Bank of England.

Mancini, C. (2001), Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion. Giornale dell'Istituto Italiano degli Attuari LXIV 19–47.

Mancini, C. (2004), Estimating the Integrated Volatility in Stochastic Volatility Models with Lévy Type Jumps, Technical report, Univ. Firenze.

Mancini, C. (2009), Non-Parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps, *Scandinavian Journal of Statistics* 36, 270-296.

Mandelbrot, B. (1963), The Variation of Certain Speculative Prices, *Journal of Business* 36, 394-419.

Marsh, T. and E. Rosenfeld (1983), Stochastic Processes for Interest Rates and Equilibrium Bond Prices, *Journal of Finance* 38, 635-646.

Meddahi, N., (2001), An Eigenfunction Approach for Volatility Modeling, Working Paper, University of Montreal

Meddahi, N. (2002), Theoretical Comparison between Integrated and Realized Volatility, *Journal of Applied Econometrics* 17, 479-508.

Mele, A., and F. Fornari (2006), Approximation Volatility Diffusions with CEV-ARCH Models, *Journal of Economic Dynamics and Control* 30, 931-966.

Merton, C.R., (1973), Theory of Rational Option Pricing, *The Bell Journal of Economics and Management Science,* 4, 141-183

Merton, R.C. (1976), Option Pricing when Underlying Stock Returns are Discontinuous, *Journal of Financial Economics 3*, 125-144.

Müller, U.A., M.M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Puctet, and J. von Weizsäcker (1997), Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components, *Journal of Empirical Finance* 4, 213–239.

Nelson, D.B., (1990), ARCH as Diffusion Approximations, *Journal of Econometrics*, 45, 7-38

Nelson, D.B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

Pan, J. (2002), The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-series Study. *Journal of Financial Economics* 63, 3-50

Pardoux, E. and D. Talay, (1985), Discretization and Simulation of Stochastic Differential Equations, *Acta Applicandae Mathematicae,* 3, 23-47.

Patton, A. and K. Shephard. (2011) Good Volatility, Bad Volatility: Signed Jumps and Persistence of Volatility, working paper, Duke University.

Piazzesi, M., (2001), Macroeconomic jump effects and the yield curve, working paper, UCLA.

Piazzesi, M., (2004), Affine Term Structure Models, Manuscript prepared for the Handbook of Financial Econometrics, University of California at Los Angeles.

Piazzesi, M., (2005), Bond Yields and the Federal Reserve, *Journal of Political Economy,* 113, 311-344.

Politis, D.N., J.P. Romano and M. Wolf, (1999), **Subsampling***,* Springer and Verlag, New York.

Pritsker, M., (1998), Nonparametric Density Estimators and Tests of Continuous Time Interest Rate Models, *Review of Financial Studies*, 11, 449-487.

Protter, P. (1990), **Stochastic Integration and Differential Equations: A New Approach**. Springer-Verlag: New York: U.S.
Rosenblatt, M., (1952), Remarks on a Multivariate Transformation, *Annals of Mathematical Statistics*, 23,470-472.

Schwert, G.W. (1989), Why Does Stock Market Volatility Change Over Time? *Journal of Finance* 44, 1115-1153.

Schwert, G.W. (1990), Stock Volatility and the Crash of '87, *Review of Financial Studies* 3, 77-102.

Scott, L. (1997), Pricing Stock Options in a Jump-diffusion Model with Stochastic Volatility and Interest Rates: Application of Fourier Inversion Methods. *Mathematical Finance* 7, 345–358.

Shephard, N. (2005), **Stochastic Volatility: Selected Readings**, Oxford: Oxford University Press.

Singleton, K.J. (2006), **Empirical Dynamic Asset Pricing - Model Specification and Econometric Assessment**, Princeton University Press: Princeton and Oxford.

Smirnov N., (1939), On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples, *Bulletin Mathematique de l'Universite' de Moscou*, 2, fasc. 2.

Sullivan R., A. Timmermann, and H. White, (1999), Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, *The Journal of Finance*, vol. 54, issue 5, 1647-1691.

Sullivan, R., A. Timmermann, and H. White, (2001), Dangers of data-driven inference: The case of calendar effects in stock returns, *Journal of Econometrics*, 249–286.

Tauchen, G. and V. Todorov (2010), Activity Signature Functions with Application for High-Frequency Data Analysis, *Journal of Econometrics* 154, 125-138.

Thompson, S.B., (2008), Identifying Term Structure Volatility from the LIBOR-swap Curve, *Review of Financial Studies*, 21, 819-854.

Todorov, V. (2009), Variance Risk Premia Dynamics: The Role of Jumps, *Review of Financial Studies* 23, 345-283.

Vasicek, O. A., (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics,* 5, 177-188.

Vuong, Q., (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333

White, H., (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25

White, H., (2000), A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.

Woerner, J.H. (2006), Power and Multipower Variation: Inference for High-Frequency Data. In: **Stochastic Finance** (A. Shiryaev, M. do Rosário Grosshino, P. Oliviera and M. Esquivel, eds.) 264–276. Springer, Berlin.

Wong, E., (1960), The Construction of a Class of Stationary Markov Processes, in Sixtenn Symposia in Applied Mathematics: Stochastic Processes in Mathematical Physics and Engineering, ed. by R. Bellman, American Mathematical Society, Providence, R.I.

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

Curriculum Vitae

Diep Duong

EDUCATION

Ph.D. Candidate in Economics, Rutgers University, USA, May 2013

M.A. in Economics, Rutgers University, 2008

M.S. in Financial Engineering, Polytechnic University (NYU Poly), 2006

Bachelor in Economics, Foreign Trade University, Vietnam, 2000

TEACHING EXPERIENCE

Lecturer, Department of Economics, Rutgers University, 2008-2013

Lecturer, Foreign Trade University, Vietnam, 2002-2004

RESEARCH EXPERIENCE

Research Assistant, Department of Economics, Rutgers University, 2008-2013

OTHER EXPERIENCE

Graduate Student Representative, Department of Economics, Rutgers University

LANGUAGES

Vietnamese (Native), English (Fluent)