Some Like it Smooth, and Some Like it Rough:
Disentangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility[*]

Torben G. Andersen[a], Tim Bollerslev[b] and Francis X. Diebold[c]

First Draft: September 2003
This Version: June 2005

Abstract: A rapidly growing literature has documented important improvements in financial return volatility measurement and forecasting performance through the use of realized variation measures constructed from the summation of high-frequency squared returns coupled with relatively simple reduced-form time series modeling procedures. Building on recent theoretical results in Barndorff-Nielsen and Shephard (2004a, 2005) for related bi-power variation measures involving the sum of adjacent absolute high-frequency returns, the present paper provides a practical and robust framework for non-parametrically measuring and assessing the statistical significance of the jump component in asset return volatility. Exploiting these ideas for a decade of high-frequency five-minute returns for the DM/$ exchange rate, the S&P500 market index, and the 30-year U.S. Treasury bond yield, we find the jump component of the price process to be distinctly less persistent than the continuous sample path variation process. Also, the occurrences of many of the most significant jumps appear to be directly associated with specific macroeconomic news announcements. Moreover, including the time series of significant jumps along with the measurements of the corresponding continuous sample path variability in an easy-to-implement reduced form HAR-RV-CJ volatility forecasting model, we find that almost all of the predictability in the daily, weekly and monthly volatilities come from the lagged continuous variation process. Our results thus set the stage for a number of interesting future econometric developments and important financial applications by separately modeling, forecasting and pricing the continuous and jump components of total return variation process.

Keywords: Continuous-time methods; jumps; quadratic variation; realized volatility; bi-power variation; high-frequency data; volatility forecasting; macroeconomic news; HAR-RV model; HAR-RV-CJ model.

JEL Codes: C1, G1

Correspondence: Send all correspondence (by email) to T. Bollerslev at boller@econ.duke.edu

[*] This research was supported by the National Science Foundation, the Guggenheim Foundation, and the Wharton Financial Institutions Center. The word "disentangling" in the title was adapted from Aït-Sahalia (2004). We are grateful to Olsen and Associates for generously supplying their intraday exchange rate data. Xin Huang provided excellent research assistance. We would also like to thank two anonymous referees, Neil Shephard and George Tauchen for many insightful discussions and comments, as well as seminar participants at the September 2003 NBER/NSF Time Series Conference in Chicago, the November 2003 Realized Volatility Conference in Montreal, the December 2003 Conference on Analysis of High-Frequency Financial Data and Market Microstructure at Academia Sinica, Taipei, Taiwan, the May 2004 Journal of Applied Econometrics Conference at Erasmus University, Rotterdam, the May 2004 Aarhus Econometrics Conference at Svinkløv, Denmark, the July 2004 Far Eastern Meetings of the Econometric Society in Seoul, South Korea, the September 2004 Innovations in Financial Econometrics Conference at Stern NYU, as well as Baruch College, Nuffield College, Princeton, Toronto, UCLA, and Uppsala Universities.

[a] Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER,
phone: 847-467-1285, e-mail: t-andersen@kellogg.northwestern.edu
[b] Department of Economics, Duke University, Durham, NC 27708, and NBER,
phone: 919-660-1846, e-mail: boller@econ.duke.edu
[c] Department of Economics, University of Pennsylvania, Philadelphia, PA 19104, and NBER,
phone: 215-898-1507, e-mail: fdiebold@sas.upenn.edu

I. Introduction

Volatility is central to asset pricing, asset allocation and risk management. In contrast to the estimation of expected returns, which generally requires long time spans of data, the results in Merton (1980) and Nelson (1992) suggest that volatility may be estimated arbitrarily well through the use of sufficiently finely sampled high-frequency returns over any fixed time interval. However, the assumption of a continuous sample path diffusion underlying these theoretical results is invariably violated in practice at the highest intradaily sampling frequencies. Thus, despite the increased availability of high-frequency data for a host of different financial instruments, practical complications have hampered the implementation of direct high-frequency volatility modeling and filtering procedures (see, e.g., the discussion in Aït-Sahalia, Mykland and Zhang, 2005; Andersen, Bollerslev and Diebold, 2003; Engle 2000; Russell and Engle, 2005; and Rydberg and Shephard, 2003).

In response to this, Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001) (henceforth ABDL), Barndorff-Nielsen and Shephard (2002a,b), and Meddahi (2002), among others, have recently advocated the use of so-called realized volatility, or variation, measures constructed from the summation of high-frequency intradaily squared returns as a way of conveniently circumventing the data complications, while retaining (most of) the relevant information in the intraday data for measuring, modeling and forecasting volatilities over daily and longer horizons. Indeed, the empirical results in ABDL (2003) suggest that simple reduced form time series models for realized volatility perform as well, if not better, than the most commonly used GARCH and related stochastic volatility models in terms of out-of-sample forecasting.[1]

At the same time, other recent studies have pointed to the importance of explicitly allowing for jumps, or discontinuities, in the estimation of specific parametric stochastic volatility models, and in the pricing of options and other derivatives instruments (e.g., Andersen, Benzoni and Lund, 2002; Bates, 2000; Chan and Maheu, 2002; Chernov, Gallant, Ghysels, and Tauchen, 2003; Drost, Nijman and Werker, 1998; Eraker, 2004; Eraker, Johannes and Polson, 2003; Johannes, 2004; Johannes, Kumar and Polson, 1999; Maheu and McCurdy, 2004; Khalaf, Saphores and Bilodeau, 2003; and Pan, 2002). In particular, it appears that the conditional variance of many assets is best described by a combination of a smooth and very slowly mean-reverting continuous sample path process, along with a much less persistent jump component.[2]

Set against this backdrop, the present paper seeks to further advance the reduced-form volatility

---

[1] These empirical findings are further corroborated by the analytical results for specific stochastic volatility models reported in Andersen, Bollerslev and Meddahi (2004).

[2] Earlier influential work on homoskedastic jump-diffusions include Ball and Torous (1983), Beckers (1981), Jarrow and Rosenfeld (1984) and Merton (1976), while Jorion (1988) and Vlaar and Palm (1993) have previously incorporated jumps in the estimation of discrete-time ARCH and GARCH models; see also the discussion in Das (2002).

forecasting approach advocated in ABDL (2003) through the development of a practical non-parametric procedure for separately measuring the continuous sample path variation and the discontinuous jump part of the quadratic variation process. Our approach builds directly on the new theoretical results in Barndorff-Nielsen and Shephard (2004a, 2005) involving so-called bi-power variation measures constructed from the summation of appropriately scaled cross-products of adjacent high-frequency absolute returns.[3] Implementing these ideas empirically with more than a decade long sample of five-minute high-frequency returns for the DM/$ foreign exchange market, the S&P500 market index, and the 30-year U.S. Treasury yield, we shed new light on the dynamic dependencies and the relative importance of jumps across the different markets. We also demonstrate important gains in terms of volatility forecast accuracy by explicitly differentiating impact of the jump and continuous sample path component. These gains obtain at daily, weekly, and even monthly forecast horizons. Our new HAR-RV-CJ forecasting model incorporating the jumps builds directly on the reduced form heterogenous AR model for the realized volatility, or HAR-RV model, due to Müller et al. (1997) and Corsi (2003), in which the realized volatility is parameterized as a linear function of the lagged realized volatilities over different horizons.

The plan for the rest of the paper is as follows. The next section briefly reviews the relevant bi-power variation theory. Section III details the high-frequency data and highlights the most important qualitative features of the raw jump measurements for each of the three markets. Section IV describes the HAR-RV volatility forecasting model and the resulting gains obtained by explicitly including the raw jump measures as additional explanatory variables. Section V presents a simple statistical procedure for measuring only the most "significant" jumps. Guided by the extensive simulation evidence in Huang and Tauchen (2005), we also discuss how the test statistic may be adapted to guard against empirically realistic market microstructure frictions in the actual high-frequency data. We then illustrate how many of the most significant jumps identified by the robust-to-market-microstructure frictions test statistics may be directly associated with specific macroeconomic news announcements, and further go on to characterize the temporal dependencies in the resulting significant jump time series. Building on this, Section VI shows how separately including the significant jumps and the corresponding continuous sample path variability measures as explanatory variables in a reduced form HAR-RV-CJ forecasting model further enhance the accuracy of the realized volatility forecasts. Section VII concludes with several suggestions for future research.

---

[3] This approach is distinctly different from the recent work of Aït-Sahalia (2002), who relies on direct estimates of the transition density function for identifying jumps.

II. Theoretical Framework

Let $p(t)$ denote the time t logarithmic price of the asset. The continuous-time jump diffusion processes traditionally used in asset pricing finance are then most conveniently expressed in stochastic differential equation form as,

$$dp(t) \;=\; \mu(t)\,dt \;+\; \sigma(t)\,dW(t) + \;\kappa(t)\,dq(t)\,, \qquad 0 \leq t \leq T, \tag{1}$$

where $\mu(t)$ is a continuous and locally bounded variation process, the stochastic volatility process $\sigma(t)$ is strictly positive and continuous, $W(t)$ denotes a standard Brownian motion, and $q(t)$ is a counting process with (possibly) time-varying intensity $\lambda(t)$. That is $P[dq(t)=1] = \lambda(t)dt$, where $\kappa(t) \equiv p(t)-p(t-)$ refers to the size of the corresponding discrete jumps in the logarithmic price process. The quadratic variation (or notional volatility/variance in the terminology of ABD, 2003) for the cumulative return process, $r(t) \equiv p(t) - p(0)$, is then given by

$$[r,r]_t \;=\; \int_0^t \sigma^2(s)\,ds \;+\; \sum_{0<s\leq t} \kappa^2(s)\;, \tag{2}$$

where by definition the summation consists of the $q(t)$ squared jumps that occurred between time 0 and time t. Of course, in the absence of jumps, or $q(t) \equiv 0$, the summation vanishes, and the quadratic variation simply equals the integrated volatility.

Several recent studies concerned with the direct estimation of continuous time stochastic volatility models have highlighted the importance of explicitly incorporating jumps in the price process along the lines of the formulation equation (1) (e.g., Andersen, Benzoni and Lund, 2002; Eraker, Johannes and Polson, 2003; Eraker, 2004; Johannes, 2004; Johannes, Kumar and Polson, 1999). Moreover, the specific parametric model estimates reported in this literature have generally suggested that any dynamic dependencies in the occurrences or sizes of the jumps are much less persistent than the dependencies in the continuous sample path volatility process. However, rather than relying on these more traditional model-driven procedures for estimating each of the two components in equation (2), we will here rely on a new non-parametric and purely high-frequency-data-driven approach for separately measuring the two components.

A. High-Frequency Data, Bi-Power Variation, and Jumps

Let the discretely sampled Δ-period returns be denoted by, $r_{t,\Delta} \equiv p(t) - p(t-\Delta)$. For ease of notation we normalize the daily time interval to unity and label the corresponding discretely sampled daily returns by a single time subscript, $r_{t+1} \equiv r_{t+1,1}$. Also, we define the daily *realized volatility*, or *variation*, by the summation of the corresponding $1/\Delta$ high-frequency intradaily squared returns,[4]

$$RV_{t+1}(\Delta) \quad \equiv \quad \sum_{j=1}^{1/\Delta} r^2_{t+j\cdot\Delta,\Delta}\,, \tag{3}$$

where for notational simplicity and without loss of generality $1/\Delta$ is assumed to be an integer. Then, as emphasized in the series of recent papers by Andersen and Bollerslev (1998), ABDL (2001), Barndorff-Nielsen and Shephard (2002a,b) and Comte and Renault (1998), among others, it follows directly by the theory of quadratic variation that the realized variation converges uniformly in probability to the increment to the quadratic variation process as the sampling frequency of the underlying returns increases. That is,

$$RV_{t+1}(\Delta) \quad \rightarrow \quad \int_t^{t+1} \sigma^2(s)\,ds \quad + \sum_{t<s\leq t+1} \kappa^2(s)\,, \tag{4}$$

for $\Delta \rightarrow 0$. Thus, in the absence of jumps the realized variation is consistent for the integrated volatility that figures prominently in the stochastic volatility option pricing literature. This result, in part, motivates the reduced-form time series modeling and forecasting procedures for realized volatilities advocated in ABDL (2003). It is clear, however, that in general the realized volatility will inherit the dynamic dependencies in both the integrated volatility, and if present, the jump dynamics. Although this does not impinge upon the theoretical justification for directly modeling and forecasting $RV_{t+1}(\Delta)$ through simple reduced-form time series procedures, it does suggest that even better forecasting models may be constructed by separately measuring and modeling the two components in equation (4).

Set against this backdrop, the present paper seeks to further enhance on the predictive gains demonstrated in ABDL (2003) through the use of new and powerful asymptotic results (for $\Delta \rightarrow 0$) in Barndorff-Nielsen and Shephard (2004a, 2005) that allow for separate (non-parametric) identification of the two components of the quadratic variation process. Specifically, define the standardized *realized bi-power variation* measure,

$$BV_{t+1}(\Delta) \quad \equiv \quad \mu_1^{-2} \sum_{j=2}^{1/\Delta} |\, r_{t+j\cdot\Delta,\Delta} \,|\; |\, r_{t+(j-1)\cdot\Delta,\Delta} \,|\,, \tag{5}$$

where $\mu_1 \equiv \sqrt{2/\pi} = E(|Z|)$ denotes the mean of the absolute value of standard normally distributed random

---

[4] We will use the terms realized volatility and realized variation interchangeably in the following.

variable, Z. It is then possible to show that for $\Delta \to 0$,[5]

$$BV_{t+1}(\Delta) \;\; \to \;\; \int_t^{t+1} \sigma^2(s)\, ds \,. \tag{6}$$

Hence, as first noted by Barndorff-Nielsen and Shephard (2004a), combining the results in equations (4) and (6), the contribution to the quadratic variation process due to the discontinuities (jumps) in the underlying price process may be consistently (for $\Delta \to 0$) estimated by

$$RV_{t+1}(\Delta) \; - \; BV_{t+1}(\Delta) \;\; \to \;\; \sum_{t<s\le t+1} \kappa^2(s). \tag{7}$$

This is the central insight on which the theoretical and empirical results in the paper builds. Of course, nothing prevents the estimates of the squared jumps defined by the right hand-side of (7) from becoming negative in a given finite ($\Delta > 0$) sample. Thus, following the suggestion of Barndorff-Nielsen and Shephard (2004a), we simply truncate the actual empirical measurements at zero,

$$J_{t+1}(\Delta) \;\; \equiv \;\; \max[\, RV_{t+1}(\Delta) \, - \, BV_{t+1}(\Delta) \, , \, 0 \,] \;, \tag{8}$$

to ensure that all of the daily estimates are non-negative.

III. Data and Summary Statistics

To highlight the generality of our empirical results related to the improved forecasting performance obtained by separately measuring the contribution to the overall variation coming from the discontinuous price movements, we present the results for three different markets. We begin this section by a brief discussion of the data sources, followed by a summary of the most salient features of the resulting realized volatility and

---

5 Corresponding general asymptotic results for so-called realized power variation measures have recently been established by Barndorff-Nielsen and Shephard (2003, 2004a); see also Barndorff-Nielsen, Graversen and Shephard (2004) for a survey of related results. In particular, it follows that in general for $0<p<2$ and $\Delta \to 0$,

$$RPV_{t+1}(\Delta,p) \;\; \equiv \;\; \mu_p^{-1}\, \Delta^{1-p/2} \sum_{j=1}^{1/\Delta} |\, r_{t+j\cdot\Delta,\Delta} \,|^p \;\; \to \;\; \int_t^{t+1} \sigma^p(s)\, ds,$$

where $\mu_p \equiv 2^{p/2}\Gamma(½(p+1))/\Gamma(½) = E(|Z|^p)$. Hence, the impact of the discontinuous jump process disappears in the limit for the power variation measures with $0<p<2$. In contrast, $RPV_{t+1}(\Delta,p)$ diverges to infinity for $p>2$, while $RPV_{t+1}(\Delta,2) \equiv RV_{t+1}(\Delta)$ converges to the integrated volatility plus the sum of the squared jumps, as in equation (4). Related expressions for the conditional moments of different powers of absolute returns have also been utilized by Aït-Sahalia (2003) in the formulation of a GMM-type estimator for specific parametric homoskedastic jump-diffusion models.

jump series for each of the three markets.

A Data Description

We present the results for three distinct markets: the DM/$ foreign exchange market, the aggregate U.S. stock market as comprised by the S&P500 index, and the fixed income market as represented by the 30-year U.S. Treasury bond yield. The DM/$ volatilities cover the period from December 1986 through June 1999, for a total of 3,045 daily observations. The underlying high-frequency spot quotations were kindly provided by Olsen & Associates in Zurich, Switzerland. This same series has been previously analyzed in the series of papers by ABDL (2001, 2003). The S&P500 volatility measurements are based on tick-by-tick transactions prices from the Chicago Mercantile Exchange (CME) augmented with overnight prices from the GLOBEX automated trade execution system, and cover the period from January 1990 through December 2002. The U.S. T-bond volatilities are similarly constructed from tick-by-tick transactions prices for the 30 year U.S. Treasury Bond futures contract traded on the Chicago Board of Trade (CBOT), and cover the identical January 1990 through December 2002 period. After removing holidays and other inactive trading days, this leaves us with a total of 3,213 observations for each of the two futures markets.[6] A more detailed description of the S&P and T-bond data is available in Andersen, Bollerslev, Diebold and Vega (2005), where the same high-frequency data are analyzed from a very different perspective. All of the volatility measures are based on linearly interpolated logarithmic five-minute returns, as in Müller et al. (1990) and Dacorogna et al. (1993).[7] For the foreign exchange market this results in a total of $1/\Delta = 288$ high-frequency return observations per day, while the two futures contracts are actively traded for $1/\Delta = 97$ five-minute intervals per day. For notational simplicity, we omit the explicit reference to Δ in the following, referring to the five-minute realized volatility and jump measures defined by equations (3) and (8) as RV_t and J_t, respectively.

B. Realized Volatilities and Jumps

The first panels in Figures 1A-C show the resulting three daily realized volatility series in standard

[6] We explicitly exclude all days with sequences of more than twenty consecutive five-minute intervals of no new prices for the S&P500, and forty consecutive five-minute intervals of no new prices for the T-bond market.

[7] In order to mitigate the impact of market microstructure frictions in the construction of unbiased and efficient realized volatility measurements, a number of recent studies have proposed ways of "optimally" choosing Δ (e.g., Aït-Sahalia, Mykland and Zhang, 2005; Bandi and Russell, 2004a,b), sub-sampling schemes (e.g., Zhang, Aït-Sahalia and Mykland, 2005; Zhang, 2004), pre-filtering (e.g., Andreou and Ghysels, 2002; Areal and Taylor, 2002; Bollen and Inder, 2002; Corsi, Zumbach, Müller and Dacorogna, 2001; Oomen 2002, 2004), Fourier methods (Barucci and Reno, 2002; and Malliavin and Mancino, 2002), or other kernel type estimators (e.g., Barndorff-Nielsen, Hansen, Lunde and Shephard, 2004; Hansen and Lunde 2004a,b; and Zhou, 1996). For now we simply follow ABDL (2002, 2001), along with most of the existing empirical literature, in the use of unweighted five-minute returns for each of the three actively traded markets analyzed here. However, we will return to a more detailed discussion of the market microstructure issue and pertinent jump measurements in Section V below.

deviation form, or $RV_t^{1/2}$. Each of the three series clearly exhibits a high degree of own serial correlation. This is confirmed by the Ljung-Box statistics for up to tenth order serial correlation reported in Tables 1A-C equal to *5,714*, *12,184*, and *1,718*, respectively. Similar results obtain for the realized variances and logarithmic transformations reported in the first and third columns in the tables. Comparing the volatility across the three markets, the S&P500 returns are the most volatile, followed by the exchange rate returns. Also, consistent with earlier evidence for the foreign exchange market in ABDL (2001), and related findings for individual stocks in Andersen, Bollerslev, Diebold and Ebens (2002) and the S&P500 in Deo, Hurvich and Lu (2005) and Martens, van Dijk and Pooter (2004), the logarithmic standard deviations are generally much closer to being normally distributed than are the raw realized volatility series. Hence, from a modeling perspective, the logarithmic realized volatilities are more amenable to the use of standard time series procedures.[8]

The second panels in Figures 1A-C display the separate measurements of the jump components (again in standard deviation form) based on the truncated estimator in equation (8).[9] As is evident from the figures, many of the largest realized volatilities are directly associated with jumps in the underlying price process. Some of the largest jumps in the DM/$ market occurred during the earlier 1986-88 part of the sample, while the size of the jumps for the S&P500 has increased significantly over the most recent 2001-02 two-year period. Meanwhile, the size of the jumps in the T-Bond market seem to be much more evenly distributed throughout the sample. Overall, both the size and occurrence of jumps appear to be much more predictable for the S&P500 than for the other two markets.

These visual observations are readily confirmed by the standard Ljung-Box portmanteau statistics for up to tenth order serial correlation in the J_t, $J_t^{1/2}$, and $log(J_t+1)$ series reported in the last three columns in Tables 1A-C. It is noteworthy that although the Ljung-Box statistics for the jumps are generally significant at conventional significance levels (especially for the jumps expressed in standard deviation or logarithmic form), the actual values are markedly lower than the corresponding test statistics for the realized volatility series reported in the first three columns. This indicates decidedly less own dynamic predictable dependencies in the portion of the overall quadratic variation originating from the discontinuous sample path price process compared to the dynamic dependencies in the continuous sample path price movements. The numbers in the table also indicate that the jumps are relatively least important for the DM/$ market, with the mean of the J_t series accounting for *0.072* of the mean of RV_t, while the same ratios for the S&P500 and T-bond markets

---

[8] Modeling and forecasting log volatility also has the virtue of automatically imposing non-negativity of fitted and forecasted volatilities.

[9] The difference between the daily realized variation and bi-power variation measures result in negative estimates for the squared daily jumps on *30.6*, *27.9* and *18.3* percent of the days for each of the three markets, respectively. As discussed further below, in the absence of jumps, the difference should be negative asymptotically ($\Delta \to 0$) for half of the days in the sample.

equal *0.144* and *0.126*, respectively.

Motivated by these observations, we now put the idea of separately measuring the jump component to work in the construction of new and simple-to-implement realized volatility forecasting models. More specifically, we follow ABDL (2003) in directly estimating a set of reduced-form time series models for each of the different realized volatility measures in Tables 1A-C; i.e., RV_t, $RV_t^{1/2}$, and $log(RV_t)$. Then, in order to assess the added value of separately measuring the jump component in forecasting the realized volatilities, we simply include the raw J_t, $J_t^{1/2}$, and $log(J_t + 1)$ jump series as additional explanatory variables in the various forecasting regressions.

IV. Reduced-Form Realized Volatility Modeling and Forecasting

A number of empirical studies have argued for the importance of long-memory dependencies in financial market volatility. Several different parametric ARCH and stochastic volatility formulations have also been proposed in the literature for best capturing this phenomenon (e.g., Andersen and Bollerslev, 1997; Baillie, Bollerslev, and Mikkelsen, 1996; Breidt, Crato and de Lima, 1998; Dacorogna et al., 2001; Ding, Granger and Engle, 1993; Robinson, 1991). These same empirical observations have similarly motivated the estimation of long-memory type ARFIMA models for realized volatilities in ABDL (2003), Areal and Taylor (2002), Deo, Hurvich and Lu (2005), Koopman, Jungbacker and Hol (2005), Martens, van Dijk, and Pooter (2004), Oomen (2002), Pong, Shackleton, Taylor and Xu (2004), Thomakos and Wang (2003), among others.

Instead of these exact, and somewhat complicated-to-estimate, fractionally integrated long-memory formulations, we will here rely on the simple-to-estimate HAR-RV class of volatility models first proposed by Corsi (2003). The HAR-RV formulation is based on a straightforward extension of the so-called Heterogeneous ARCH, or HARCH, class of models analyzed by Müller et al. (1997), in which the conditional variance of the discretely sampled returns is parameterizes as a linear function of the lagged squared returns over the identical return horizon together with the squared returns over longer and/or shorter return horizons.[10] Although the HAR type structure doesn't formally possess long-memory, the mixing of relatively few volatility components is capable of reproducing a remarkably slow decay that is almost indistinguishable from that of a hyperbolic pattern over most empirically relevant forecast horizons.[11]

---

[10] Müller et al. (1997) heuristically motivates the HARCH model through the existence of distinct group of traders with different investment horizons.

[11] Mixtures of low-order ARMA models have similarly been used in approximating and forecasting long-memory type dependencies in the conditional mean by Basak, Chan and Palma (2001), Cox (1991), Hsu and Breidt (2003), Man (2003), O'Connell (1971) and Tiao and Tsay (1994), among others. The component GARCH model in Engle and Lee (1999) and the multi-factor continuous time stochastic volatility model in Gallant, Hsu and Tauchen (1999) are both motivated by similar considerations; see also the discussion of the related multifractal regime switching models in Calvet and Fisher (2001, 2002).

A. The HAR-RV-J Model

To define the HAR-RV model, let the multi-period *normalized* realized variation, defined by the sum of the corresponding one-period measures, be denoted by,

$$RV_{t,t+h} = h^{-1}[\, RV_{t+1} \; + \; RV_{t+2} \; + ... + RV_{t+h} \,] \, , \qquad (9)$$

where $h = 1, 2, ...$. Note that, by definition $RV_{t,t+1} \equiv RV_{t+1}$. Also, provided that the expectations exist, $E(RV_{t,t+h}) \equiv E(RV_{t+1})$ for all h. For ease of reference, we will refer to these normalized measures for $h=5$ and $h=22$ as the weekly and monthly volatilities, respectively. The daily HAR-RV model of Corsi (2003) may then be expressed as,[12]

$$RV_{t+1} \;=\; \beta_0 \;+\; \beta_D RV_t \;+\; \beta_W RV_{t-5,t} \;+\; \beta_M RV_{t-22,t} \;+\; \epsilon_{t+1} \, , \qquad (10)$$

where $t = 1, 2, ..., T$. Of course, realized volatilities over other horizons could easily be included as additional explanatory variables on the right-hand-side of the regression equation, but the daily, weekly and monthly measures employed here afford a natural economic interpretation.[13]

This HAR-RV forecasting model for the one-day volatilities extends straightforwardly to models for the realized volatilities over longer horizons, $RV_{t,t+h}$. Moreover, given the separate non-parametric measurements of the jump component discussed above, the corresponding time series are readily included as an additional explanatory variable over and above the realized volatility components, resulting in the new HAR-RV-J model,

$$RV_{t,t+h} \;=\; \beta_0 \;+\; \beta_D RV_t \;+\; \beta_W RV_{t-5,t} \;+\; \beta_M RV_{t-22,t} \;+\; \beta_J J_t \;+\; \epsilon_{t,t+h} \, . \qquad (11)$$

With observations every period and longer forecast horizons, or $h>1$, the error term will generally be serially correlated up to (at least) order $h-1$. This will not affect the consistency of the regression coefficient estimates, but the corresponding standard errors for the estimates obviously need to be adjusted for this overlapping data

---

[12] The time series of realized volatilities in this and all of the subsequent HAR-RV regressions are implicitly assumed to be stationary. Formal tests for a unit root in $RV_{t,+1}$ easily rejects the null hypothesis of non-stationarity for each of the three markets. Also, the standard log-periodogram estimates for the degree of fractional integration in $RV_{t,+1}$ equal *0.347*, *0.383*, and *0.437*, respectively, with a theoretical asymptotic standard error of *0.087*.

[13] Related mixed data sampling, or MIDAS in the terminology of Ghysels, Santa-Clara and Valkanov (2004), regressions have recently been estimated by Ghysels, Santa-Clara and Valkanov (2005).

problem. In the results discussed below, we rely on the Bartlett/Newey-West heteroskedasticity consistent covariance matrix estimator with *5*, *10*, and *44* lags for the daily ($h=1$), weekly ($h=5$), and monthly ($h=22$) regression estimates, respectively.

Turning to the results reported in the first three columns in Tables 2A-C, the estimates for β_D, β_W, and β_M confirm the existence of highly persistent dependencies in the volatilities. Interestingly, the relative importance of the daily volatility component decreases from the daily to the weekly to the monthly regressions, whereas the monthly volatility component tends to be relatively more important for the longer-run monthly regressions. Importantly, the estimates for the jump component, β_J, are systematically negative across all models and markets, and with few exceptions, overwhelmingly significant.[14] Thus, whereas the realized volatilities are generally highly persistent, the impact of the lagged realized volatility is significantly reduced by the jump component. For instance, for the daily DM/$ realized volatility a unit increase in the daily realized volatility implies an average increase in the volatility on the following day of $0.430 + 0.196/5 + 0.244/22 = 0.480$ for days where $J_t = 0$, whereas for days in which part of the realized volatility comes from the jump component the increase in the volatility on the following day is reduced by *-0.486* times the jump component. In other words, if the realized volatility is entirely attributable to jumps, it carries no predictive power for the following day's realized volatility. Similarly for the other two markets, the combined impact of a jump for forecasting the next day's realized volatility equal $0.341 + 0.485/5 + 0.165/22 - 0.472 = -0.027$ and $0.074 + 0.317/5 + 0.358/22 - 0.152 = -0.002$, respectively.

Comparing the R^2's for the HAR-RV-J models to the R^2's for the "standard" HAR model reported in the last row in which the jump component is absent and the realized volatilities on the right-hand-side but not the left-hand-side of equation (11) are replaced by the corresponding lagged squared daily, weekly, and monthly returns clearly highlights the added value of the high-frequency data. Although the coefficient estimates for the β_D, β_W, and β_M coefficients in the "standard" HAR models (available upon request) generally align fairly closely with those of the HAR-RV-J models reported in the tables, the explained variation is systematically lower.[15] Importantly, the gains afforded by the use of the high-frequency based realized volatilities are not restricted to the daily and weekly horizons. In fact, the longer-run monthly forecasts result

---

[14] Note, that nothing prevents the forecasts for the realized volatilities from the HAR-RV-J model with $\beta_J<0$ from becoming negative. We did not find this to be a problem for any of our in-sample model estimates, however. A more complicated multiplicative error structure, along the lines of Engle (2002) and Engle and Gallo (2005), could be employed to ensure positivity of the conditional expectations.

[15] Note that although the relative magnitude of the R^2's for a given volatility series are directly comparable across the two models, as discussed in Andersen, Bollerslev and Meddahi (2005), the measurement errors in the left-hand-side realized volatility measures invariably result in a systematic downward bias in the reported R^2's vis-a-vis the inherent predictability in the true latent quadratic variation process.

in the largest relative increases in the R^2's, with those for the S&P500 and T-Bonds tripling for the HAR-RV-J models relative to those from the HAR models based on the coarser daily, weekly and monthly squared returns. These large gains in forecast accuracy through the use of realized volatilities are, of course, entirely consistent with the earlier empirical evidence in ABDL (2003), Bollerslev and Wright (2001) and Martens (2002), among others, and further corroborated by the analytical results of Andersen, Bollerslev and Meddahi (2004).

B. Non-Linear HAR-RV-J Models

Practical uses of volatility models and forecasts often involve standard deviations as opposed to variances. The second set of columns in Tables 2A-C thus reports the parameter estimates and R^2's for the corresponding HAR-RV-J model cast in standard deviation form,

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_D RV_t^{1/2} + \beta_W (RV_{t-5,t})^{1/2} + \beta_M (RV_{t-22,t})^{1/2} + \beta_J J_t^{1/2} + \epsilon_{t,t+h}. \qquad (12)$$

The qualitative features and ordering of the different parameter estimates are generally the same as for the variance formulation in equation (11). In particular, the estimates for β_J are systematically negative. Similarly, the R^2's indicate quite dramatic gains for the high-frequency based HAR-RV-J model relative to the standard HAR model. The more robust volatility measurements provided by the standard deviations also result in higher R^2's than for the variance-based models reported in the first three columns.[16]

As noted in Table 1 above, the logarithmic daily realized volatilities are approximately unconditionally normally distributed for each of the three markets. This empirical regularity motivated ABDL (2003) to model the logarithmic realized volatilities, in turn allowing for the use of standard normal distribution theory and related mixture models.[17] Guided by this same idea, we report in the last three columns of Tables 2A-C the estimates for the logarithmic HAR-RV-J model,

$$log(RV_{t,t+h}) = \beta_0 + \beta_D log(RV_t) + \beta_W log(RV_{t,t-5}) + \beta_M log(RV_{t,t-22}) + \beta_J log(J_t+1) + \epsilon_{t,t+h}. \qquad (13)$$

[16] The $R^2 = 0.431$ for the daily HAR-RV-J model for the DM/$ realized volatility series in the fourth column in Table 2A also exceeds the comparable in-sample one-day-ahead $R^2 = 0.355$ for the long-memory VAR model reported in ABDL (2003).

[17] This same transformation has subsequently been used for other markets by Deo, Hurvich and Lu (2005), Koopman, Jungbacker and Hol (2005), Martens, van Dijk and Pooter (2004), and Oomen (2002) among others. Of course, the log-normal distribution isn't closed under temporal aggregation. Thus, if the daily logarithmic realized volatilities are normally distributed, the weekly and monthly volatilities can not also be log-normally distributed. However, as argued by Barndorff-Nielsen and Shephard (2002a) and Forsberg and Bollerslev (2002), the log-normal distributions for the volatility may be closely approximated by Inverse Gaussian distributions, which are formally closed under temporal aggregation.

The estimates are again directly in line with those for the HAR-RV-J models for $RV_{t,t+h}$ and $(RV_{t,t+h})^{1/2}$ discussed earlier. In particular, the β_D coefficients are generally the largest in the daily models, the β_W's are the most important in the weekly models, and the β_M's in the monthly models. At the same time, the negative estimates for the β_J coefficients temper the persistency in the forecasts, suggesting that discontinuities, or jumps, in the price processes tend to be associated with short-lived bursts in volatility.

V. Significant Jumps

The empirical results discussed in the previous two sections rely on the simple non-parametric jump estimates defined by the difference between the realized volatility and the bi-power variation. As discussed in Section II, the theoretical justification for these measurements is based on the notion of increasingly finer sampled returns, or $\Delta \rightarrow 0$. Of course, any practical implementation with a fixed sampling frequency, or $\Delta > 0$, is invariably subject to measurement errors. The non-negativity truncation in equation (8) alleviates part of this finite-sample problem by eliminating theoretically non-sensible negative estimates for the squared jumps. However, the resulting $J_t^{1/2}$ series depicted in Figures 1A-C arguably exhibit an unreasonably large number of non-zero small positive values as well. From a more structural modeling perspective it may be desirable to treat these small jumps as measurement errors, or part of the continuous sample path variation process, only associating abnormally large values of $RV_t(\Delta)-BV_t(\Delta)$ with the jump component. The next sub-section provides a theoretical framework for doing so.

A. Asymptotic Distribution Theory

The distributional results developed in Barndorff-Nielsen and Shephard (2004a, 2005) imply that under ideal conditions and in the absence of jumps,[18]

$$\Delta^{-1/2} \frac{RV_{t+1}(\Delta) - BV_{t+1}(\Delta)}{[\,(\mu_1^{-4} + 2\mu_1^{-2} - 5)\int_t^{t+1} \sigma^4(s)ds\,]^{1/2}} \Rightarrow N(0,1), \tag{14}$$

for $\Delta \rightarrow 0$. Hence, an abnormally large value of this standardized difference between $RV_{t+1}(\Delta)$ and $BV_{t+1}(\Delta)$ is naturally interpreted as evidence in favor of a "significant" jump over the $[t,t+1]$ time interval. Of course, the

[18] Formally, the logarithmic price process belongs to the continuous stochastic volatility semimartingale class of models defined by equation (1) above in which $q(t)\equiv 0$. As discussed by Barndorff-Nielsen and Shephard (2005), all continuous local martingales with absolute continuous quadratic variation may be expressed in this way.

integrated quarticity that appears in the denominator needs to be estimated in order to actually implement this statistic. In parallel to the arguments underlying the robust estimation of the integrated volatility by the realized bi-power variation, it is possible to show that even in the presence of jumps, the integrated quarticity may be consistently estimated by the normalized sum of the product of $n \geq 3$ adjacent absolute returns raised to the power of $4/n$. In particular, on defining the standardized *realized tri-power quarticity* measure,

$$TQ_{t+1}(\Delta) \equiv \Delta^{-1} \mu_{4/3}^{-3} \sum_{j=3}^{1/\Delta} | r_{t+j\cdot\Delta,\Delta} |^{4/3} | r_{t+(j-1)\cdot\Delta,\Delta} |^{4/3} | r_{t+(j-2)\cdot\Delta,\Delta} |^{4/3} , \tag{15}$$

where $\mu_{4/3} \equiv 2^{2/3} \cdot \Gamma(7/6) \cdot \Gamma(½)^{-1} = E(|Z|^{4/3})$, it follows that for $\Delta \to 0$,

$$TQ_{t+1}(\Delta) \Rightarrow \int_t^{t+1} \sigma^4(s)\, ds . \tag{16}$$

Combining the results in equations (14)-(16), the "significant" jumps may therefore be identified by comparing realizations of the feasible test statistics,[19]

$$W_{t+1}(\Delta) \equiv \Delta^{-1/2} \frac{RV_{t+1}(\Delta) - BV_{t+1}(\Delta)}{[\, (\mu_1^{-4} + 2\mu_1^{-2} - 5)\, TQ_{t+1}(\Delta)\,]^{1/2}} , \tag{17}$$

to a standard normal distribution.

Meanwhile, the extensive simulation-based evidence for specific parametric continuous time diffusions reported in Huang and Tauchen (2005), suggests that the $W_{t+1}(\Delta)$ statistic defined in (17) tends to over-reject the null hypothesis of no jumps for large critical values. At the same time, following the approach advocated

---

[19] Similar results were obtained by using the robust *realized quad-power quarticity* measure advocated in Barndorff-Nielsen and Shephard (2004a, 2005),

$$QQ_{t+1}(\Delta) \equiv \Delta^{-1} \mu_1^{-4} \sum_{j=4}^{1/\Delta} | r_{t+j\cdot\Delta,\Delta} | \, | r_{t+(j-1)\cdot\Delta,\Delta} | \, | r_{t+(j-2)\cdot\Delta,\Delta} | \, | r_{t+(j-3)\cdot\Delta,\Delta} | .$$

Note however, that the realized quarticity,

$$RQ_{t+1}(\Delta) \equiv RPV_{t+1}(\Delta,4) \equiv \Delta^{-1} \mu_4^{-1} \sum_{j=1}^{1/\Delta} r_{t+j\cdot\Delta,\Delta}^4 ,$$

used in estimating the integrated quarticity by Barndorff-Nielsen and Shephard (2002a) and Andersen, Bollerslev, and Meddahi (2005) is not consistent in the presence of jumps, which in turn would result in a complete loss of power for the corresponding test statistic obtained by replacing $TQ_{t+1}(\Delta)$ in equation (17) with $RQ_{t+1}(\Delta)$.

by Barndorff-Nielsen and Shephard (2004b), different variance stabilizing transforms for the joint asymptotic distribution of the realized volatility and bi-power variation measures generally give rise to much better performing test statistics. In particular, on applying the delta-rule to the joint bivariate distribution, Huang and Tauchen (2005) find that the ratio-statistic,

$$Z_{t+1}(\Delta) \equiv \Delta^{-1/2} \frac{[\, RV_{t+1}(\Delta) - BV_{t+1}(\Delta)\,]\, RV_{t+1}(\Delta)^{-1}}{[\,(\mu_1^{-4} + 2\mu_1^{-2} - 5)\max\{\,1\,,\, TQ_{t+1}(\Delta)\, BV_{t+1}(\Delta)^{-2}\,\}\,]^{1/2}}\ , \tag{18}$$

where the max adjustment follows by a Jensen's inequality type argument as in Barndorff-Nielsen and Shephard (2004b), is very closely approximated by a standard normal distribution throughout its entire support.[20] Moreover, the ratio-statistic in (18) also has reasonable power against several empirically realistic calibrated stochastic volatility jump diffusion models.

Hence, the "significant" jumps are naturally identified by the realizations of $Z_{t+1}(\Delta)$ in excess of some critical value, say Φ_α,

$$J_{t+1,\alpha}(\Delta) \equiv I\,[\, Z_{t+1}(\Delta) > \Phi_\alpha\,] \cdot [\, RV_{t+1}(\Delta) - BV_{t+1}(\Delta)\,]\,, \tag{19}$$

where $I[\,\cdot\,]$ denotes the indicator function.[21] Moreover, in order to ensure that the measurements of the continuous sample path variation and the jump component add up to the total realized variation, the former component is naturally estimated by the residual relationship,

$$C_{t+1,\alpha}(\Delta) \equiv I\,[\, Z_{t+1}(\Delta) \le \Phi_\alpha\,] \cdot RV_{t+1}(\Delta) + I\,[\, Z_{t+1}(\Delta) > \Phi_\alpha\,] \cdot BV_{t+1}(\Delta)\,. \tag{20}$$

Note that for $\Phi_\alpha > 0$, the definitions in equations (19) and (20) automatically guarantee that both $J_{t+1,\alpha}(\Delta)$ and

[20] In an earlier version of this paper, we relied on the log-based statistic,

$$U_{t+1}(\Delta) \equiv \Delta^{-1/2} \frac{\log(RV_{t+1}(\Delta)) - \log(BV_{t+1}(\Delta))}{[\,(\mu_1^{-4} + 2\mu_1^{-2} - 5)\, TQ_{t+1}(\Delta)\, BV_{t+1}(\Delta)^{-2}\,]^{1/2}}\ .$$

Details of these, qualitatively very similar, results are available upon request.

[21] As noted in personal communication with Neil Shephard, this may alternatively be interpreted as a shrinkage type estimator for the jump component.

$C_{t+1,\alpha}(\Delta)$ are positive.[22] Of course, the non-negativity truncation imposed in equation (8) underlying the empirical jump measurements employed in the preceding two sections corresponds directly to $\alpha = 0.5$, or $J_{t,0.5}$.

B. Market Microstructure Noise

As already discussed in Section III.A, a host of practical market microstrucure frictions, including the use of discrete price grid points and bid-ask spreads, invariably renders the assumption of a continuously observed logarithmic price process adhering to the semimartigale assumption fictitious. Instead, following Aït-Sahalia, Mykland and Zhang (2005), Bandi and Russell (2004b), Zhang, Mykland and Aït-Sahalia (2005), among others, assume that the observed price process is "contaminated" by a market microstructure noise component, say $p(t)=p^*(t)+\nu(t)$, where $p^*(t)$ refers to the true (latent) semimartingale logarithmic price process that would obtain in the absence of any frictions, while $\nu(t)$ denotes an *i.i.d.* white noise component.[23] The discretely sampled Δ-period observed returns,

$$r_{t,\Delta} \;\equiv\; p^*(t) - p^*(t-\Delta) + \nu(t) - \nu(t-\Delta) \;\equiv\; r^*_{t,\Delta} + \eta_{t,\Delta}, \tag{21}$$

then equals the true (latent) returns plus the first-order moving average process, $\eta_{t,\Delta}$. Assuming that the variance of $v(t)$ does not depend upon Δ, the noise term will eventually (for $\Delta \to 0$) dominate the contribution to the overall realized variation in (3) coming from the squared true (latent) high-frequency returns, formally rendering $RV_{t+1}(\Delta)$ inconsistent as a measure for the quadratic variation of $p^*(t)$. In practice, the impact of the market microstructure noise is most easily controlled through the choice of Δ. And, it was exactly this bias-variance tradeoff that motivated our choice of a five-minute sampling frequency for each of the three actively traded markets analyzed here, as the frequency at which the bias in the realized variation measure in equation (3) has largely disappeared.

By analogous arguments, the noise term will generally result in an upward bias in the new bi-power variation measure in equation (5) for Δ "too small," as $E(|r^*_{t,\Delta}|) < E(|r^*_{t,\Delta} + \eta_{t,\Delta}|)$. The first-order serial correlation in $\eta_{t,\Delta}$ further implies that any two adjacent observed returns, say $r_{t+j\cdot\Delta,\Delta}$ and $r_{t+(j-1)\cdot\Delta,\Delta}$, will be

[22] It is possibly, that by specifying $\alpha(\Delta)\to 1$ as an explicit function of $\Delta\to 0$, this approach may formally be shown to result in period-by-period consistent (as $\Delta\to 0$) estimates of the jump component. Of course, data limitations invariably restricts the sampling frequency ($\Delta>0$), so that such a result would arguably be of only limited practical use.

[23] More complicated non *i.i.d.* market microstructure noise components have been analyzed in the realized volatility setting by Bandi and Russell (2004a) and Hansen and Lunde (2004a,b), among others.

serially correlated. In comparison to the realized variation measure based on the sum of the squared high-frequency returns, this spuriously induced first-order serial correlation will therefore result in an additional source of bias in the $BV_{t+1}(\Delta)$ measure. Of course, similar arguments apply to the tri-power quarticity measure in equation (15). As discussed at length in Huang and Tauchen (2005), this in turn implies that in the presence of market microstructure noise, the jump test statistics discussed in the previous section will generally be biased *against* finding significant jumps. In particular, it is possible to show that in the absence of jumps, $\lim_{\Delta \to \infty} [RV_{t+1}(\Delta) - BV_{t+1}(\Delta)] = \kappa < 0$, so that the $W_{t+1}(\Delta)$ test statistic defined in (17) will be negatively biased. Although comparable analytical results are not available for the ratio-statistic in equation (18), the numerical calculations and extensive simulation evidence reported in Huang and Tauchen (2005) confirm that for small values of Δ, the test tends to be under-sized, and this tendency to under-reject further deteriorates with the magnitude of the variance of the $v(t)$ noise component.

Meanwhile, the spuriously induced first order serial correlation in the observed returns defined in equation (21) is readily broken through the use of staggered, or skip-one, returns. Specifically, replacing the sum of the absolute adjacent returns in equation (5) with the corresponding staggered absolute returns, a modified realized bi-power variation measure may be defined by,

$$BV_{1,t+1}(\Delta) \;\equiv\; \mu_1^{-2}\,(1-2\Delta)^{-1} \sum_{j=3}^{1/\Delta} |\, r_{t+j\cdot\Delta,\Delta} \,|\; |\, r_{t+(j-2)\cdot\Delta,\Delta} \,| \,, \tag{22}$$

where the normalization factor in front of the sum reflects the loss of two observations due to the staggering. Of course, higher order serial dependencies could be broken in an analogous fashion by further increasing the lag length. Similarly, the integrated quarticity may alternatively be estimated by the staggered realized tri-power quarticity,

$$TQ_{1,t+1}(\Delta) \;\equiv\; \Delta^{-1}\,\mu_{4/3}^{-3}\,(1-4\Delta)^{-1} \sum_{j=5}^{1/\Delta} |\, r_{t+j\cdot\Delta,\Delta} \,|^{4/3}\, |\, r_{t+(j-2)\cdot\Delta,\Delta} \,|^{4/3}\, |\, r_{t+(j-4)\cdot\Delta,\Delta} \,|^{4/3} \,. \tag{23}$$

Importantly, as shown by Barndorff-Nielsen and Shephard (2004a), in the absence of the noise component, these staggered realized variation measures remain consistent for the corresponding integrated variation measures. Consequently, the asymptotic distribution of the test statistic obtained by replacing $BV_{t+1}(\Delta)$ and $TQ_{t+1}(\Delta)$ in equation (18) with their staggered counterparts, $BV_{1,t+1}(\Delta)$ and $TQ_{1,t+1}(\Delta)$, respectively, say $Z_{1,t+1}(\Delta)$, will also be asymptotically (for $\Delta \to 0$) standard normally distributed. However, following the discussion above, the staggering should help alleviate the confounding influences of the market microstructure noise, resulting in empirically more accurate finite sample approximations.

This conjecture is indeed confirmed by the comprehensive simulation results reported in Huang and Tauchen (2005), which show that the ratio-statistic calculated with the staggered realized bi-power and tri-power variation measures performs admirably for a wide range of market microstructure contaminants. Quoting from the conclusion in Huang and Tauchen (2005): "The Monte Carlo evidence suggests that, under the arguably realistic scenarios considered here, the recently developed tests for jumps perform impressively and are not easily fooled."

Hence, in the empirical results reported on below we will rely on the $J_{t,\alpha}(\Delta)$ and $C_{t,\alpha}(\Delta)$ measures previously defined in equations (19) and (20) calculated on the basis of the staggered $Z_{1,t}(\Delta)$ statistic. To facilitate the notation, we will again omit the explicit reference to the sampling frequency, Δ, simply referring to the "significant" jump and continuous sample path variability measures calculated from the five-minute returns as $J_{t,\alpha}$ and $C_{t,\alpha}$, respectively. The subsequent section summarizes various features of these jump measurements for values of α ranging from *0.5* to *0.9999*, or Φ_α ranging from *0.0* to *3.719*.

C. Significant Jump Measurements and Macroeconomic News

Before summarizing the full sample time series evidence, it is instructive to look at a few specific days to illustrate the working of the jump statistic. To this end, Figure 2 displays the five-minute increments in the logarithmic prices for a highly significant jump day and a day with a large continuous price move for each of the three markets. For ease of comparison, the logarithmic price has been normalized to zero at the beginning of each day, so that a unit increment corresponds to a one-percent return in all of the graphs.

The first panel shows the movements in the DM/$ exchange rate on December 10, 1987. The $Z_{1,t}$ statistics for this day equals *10.315*, thus indicating a highly significant jump. The timing of the jump, as evidenced by the apparent discontinuity at 13:30 GMT, corresponds exactly to the 8:30 EST release of the U.S. trade deficit for the month of October.[24] Quoting from the *Wall Street Journal*: "The trade gap swelled to a record $17.63 billion in October, sending the dollar and bonds plunging." Meanwhile, the second panel in the first row depicts a similar large daily decline in the value of the dollar on September 17, 1992. In fact, this is the day in the entire sample with the highest value of $BV_{1,t} = 4.037$. At the same time, the $Z_{1,t}$ statistic for this day equals *-0.326*, and thus in spite of the overall large daily move, does not signify any jump(s). This

[24] The systematic news announcement analysis in Andersen, Bollerslev, Diebold and Vega (2003, 2005) also point to the U.S. trade balance as one of the most important regularly scheduled macroeconomic news releases for the foreign exchange market.

particular day succeeds the day following the temporary withdrawal of the British Pound from the European Monetary System, and it has previously been highlighted in the study by Barndorff-Nielsen and Shephard (2005). Again, quoting from the *Wall Street Journal*: "The dollar and the pound each sank more than 2% against the mark as nervousness persisted in the currency market." Of course, without the benefit of the intraday high-frequency data, the day-to-day price moves in the first two panels would look almost identical.

Turning to the second row in the figure, the first panel shows the five-minute movements in the S&P500 on June 30, 1999. As suggested by visual inspection of the plot, the $Z_{1,t}= 7.659$ statistic is again highly significant. Moreover, the apparent timing of the jump at 13:15 CST, or 14:15 EST, corresponds exactly to the time of the 1/4% increase in the FED funds rate on that day. However, that rate hike was accompanied by a statement by the FED that it "might not raise rates again in the near term due to conflicting forces in the economy," which apparently was viewed as a positive sign by the market. In contrast, on July 24, 2002, as depicted in the panel on the right, $Z_{1,t}= -0.704$, while $BV_{1,t}$ achieves its maximum value of *29.247*. The abnormally large daily return of *7.157* is also the largest over the whole sample. Yet, this "rough" daily move is made up of the sum of many "smooth" intraday price moves, with no apparent jump(s) in the process. Interestingly, the NYSE also saw a record trading volume of *2.77* billion shares on that day.

The last row in the figure refers to the T-Bond market. The apparent timing of the highly significant jump, $Z_{1,t}= 6.877$, on August 1, 1996, corresponds directly to the release of the National Association of Purchasing Manager's (NAPM) index at 9:00 CST, or 10:00 EST.[25] Meanwhile, the second T-Bond panel for December 7, 2002 again depicts a large daily, but generally "smooth" intraday, price move. Interestingly, most of the movements occurred in the morning following the release of a higher than expected jobless rate. While this didn't result in an immediate jump in the T-Bond price, it reassured most economist's that the FED would cut its rate at the next Board meeting the following business day, which in fact it did. According to *Wall Street Journal*: "Economists said the jobs report removed any lingering doubts that the Federal Reserve will reduce interest rates for the 11th time in the past 12 months when it meets tomorrow."

The direct association of the highly significant jump days in Figure 2 with readily identifiable macroeconomic news affirm earlier case studies for the DM/$ foreign exchange market in Barndorff-Nielsen and Shephard (2005), and is directly in line with the aforementioned evidence in Andersen, Bollerslev, Diebold and Vega (2003, 2005) among others, documenting significant intra-daily price moves in response to a host of macroeconomic news announcements. Similarly, Johannes (2004) readily associates the majority of the

---

[25] The results in Andersen, Bollerslev, Diebold and Vega (2005) again indicate that for the T-Bond market, news about the NAPM index result in the overall highest five-minute return regression R^2 among all of the regularly scheduled macroeconomic announcements.

estimated jumps in a parametric jump-diffusion model for daily interest rates with specific macroeconomic news events. At the same time, informal inspection suggests that not all of the jump days identified by large values of the non-parametric high-frequency $Z_{1,t}$ statistic are as easily linked to specific "news" arrivals. Indeed, it would be interesting, but beyond the scope of the present paper, to attempt a more systematic characterization of the types of events that causes the different markets to jump. Instead, we next turn to a discussion of various summary statistics related to the time series of significant jumps employed in our subsequent volatility forecasting models.

D. Significant Jump Measurements and Dynamic Dependencies

To begin, the first row in Tables 3A-C reports the proportion of days with significant jumps for each of the three markets based on the $Z_{1,t}$ statistic as a function of the significance level, α. Although the use of α's in excess of *0.5* has the intended effect of reducing the number of days with jumps, the procedure still identifies many more significant jumps than would be expected if the underlying price process was continuous. Comparing the jump intensities across the three markets, the foreign exchange and the T-Bond markets generally exhibit the highest proportion of jumps, whereas the stock market has the lowest.[26] For instance, employing a cutoff of $\alpha = 0.999$, or $\Phi_\alpha = 3.090$, results in *417*, *244*, and *424* significant jumps for each of the three markets respectively, all of which far exceed the expected three jumps for a continuous price process (*0.001* times *3,045* and *3,213*, respectively). Indeed, all of the daily jump proportions are much higher than the jump intensities estimated with specific parametric jump diffusion models applied to daily or coarser frequency returns, which typically suggest only a few jumps a year; see, e.g., the estimates for the S&P500 in Andersen, Benzoni and Lund (2002). Intuitively, just as the stock market crash of 1987 and the corresponding large negative daily return on October 17 isn't visible in the time series of annual equity returns, many of the jumps identified by the high-frequency based realized variation measures employed here will invariably be blurred in the coarser daily or lower frequency returns through an aliasing type phenomenon.

Turning to the second and third rows in the table, it is noteworthy that even though the proportions of jumps depend importantly on the particular choice of α, the sample means and standard deviations of the resulting jump time series aren't nearly as sensitive to the significance level. This observation is further corroborated by the time series plots for each of the three markets in the third and fourth panels in Figures 1A-1C, which show the $Z_{1,t}$ statistics and a horizontal line at *3.090*, along with the resulting significant jumps,

[26] This ordering among the three markets is again consistent with the aforementioned evidence in Andersen, Bollerslev, Diebold and Vega (2005), showing that equity markets generally respond the least to macroeconomic news announcements.

or $J_{t,0.999}^{1/2}$. It is evident that the test statistic generally picks out the largest values of $J_t^{1/2}$ as being significant, so that the sample means and standard deviations for the time series depicted in the second and the fourth panels are all fairly close.

The Ljung-Box statistics for up to tenth order serial correlation in the $J_{t,\alpha}$ series for the S&P500 reported in the fourth row in Table 3B are all highly significant, irregardless of the choice of α. This is, of course, at odds with most of the parametric jump-diffusion model estimates reported in the recent literature, which as previously noted suggest very little, or no, predictable variation in the jump process. Still, it is noteworthy that the values of the Ljung-Box statistics for the significant S&P500 jumps are all much less than the corresponding statistics for the realized variation series reported in Table 1B. Meanwhile, the corresponding Ljung-Box tests for the DM/$ and T-Bond jump series are not nearly as large, and generally insignificant for the jumps defined by α's in excess of *0.990*.

This same general picture is further corroborated by the Likelihood Ratio test for the null of *i.i.d.* jump occurrences against the alternative of a first-order Markov chain reported in the fifth row; see Christoffersen (1998) for further details. Under the null of no dependencies this test statistic should be asymptotically chi-square distributed with one degree of freedom. None of the test statistics for α equal to *0.999* or *0.9999* for the DM/$ and T-Bond markets exceed the corresponding *95*-percent critical value of *3.84*, while the tests for the S&P500 are highly significant.

Interestingly, when looking beyond the simply own linear dependencies and the Ljung-Box test for the $J_{t,\alpha}$ series, a somewhat different picture emerges. In particular, decomposing the $J_{t,\alpha}$ series into the times between jumps and the sizes of the corresponding jumps, there appears to be strong evidence for clustering in the occurrences of the significant jumps for both the S&P500 and T-Bond markets, as evidenced by the Ljung-Box test for up to tenth order serial correlation in the durations between the jumps, denoted by LB_{10}, $D_{t,\alpha}$ in Table 3. Similarly, the Ljung-Box tests for serial correlation in the time series of only the significant jumps, denoted by LB_{10}, $J_{t,\alpha}{}^{+}$ in Table 3, strongly suggest that large (small) jumps tend to cluster together in time with other large (small) jumps for both the DM/$ and S&P500 markets. In contrast, for the T-Bond market and α *=0.999*, only the durations but not the sizes of the jumps appear to cluster in time.

These more complex dynamic dependencies in the significant jump time series are further illustrated in Figure 3, which plots the smoothed jump intensities and jump sizes for each of the three markets. The graphs are constructed by exponentially smoothing (with a smoothing parameter of *0.94*) the average monthly jump intensities and sizes for the significant jumps based on $\alpha = 0.999$. The jump sizes are again expressed in standard deviation form, or $J_{t,0.999}^{1/2}$. From the very first panel the DM/$ jump intensities are approximately constant throughout the sample. Similarly, the smoothed jump sizes for the T-Bond market depicted in the last panel vary very little over the sample period. Meanwhile, all of the other four panels suggest the existence of potentially important temporal dependencies in the jump arrival processes and jump sizes.[27] It would be interesting, but beyond the scope of the present paper, to further explore the formulation of parametric jump-diffusion models best designed to capture these non-linear dependencies.[28]

Instead, we next turn to a simple extension of the HAR-RV-J volatility forecasting model introduced in Section IV, in which we incorporate the time series of significant jumps as additional explanatory variables in a straightforward linear fashion.

VI. Reduced Form Realized Volatility Modeling and Forecasting Revisited

The regression estimates for the HAR-RV-J model reported in Section IV show that the inclusion of the simple consistent daily jump measure corresponding to $\alpha=0.5$ as an additional explanatory variable over-and-above the daily realized volatilities result in highly significant and negative parameter estimations for the jump coefficient. These results are, of course, entirely consistent with the summary statistics for the jump measurements discussed above, which indicate markedly less own (linear) serial correlation in the significant jump series in comparison to the realized volatility series. Building on these results, the present section extends the HAR-RV-J model by explicitly decomposing the realized volatilities that appear as explanatory variables on the right-hand-side into the continuous sample path variability and the jump variation utilizing the separate non-parametric measurements based on the $Z_{1,t}$ statistic along with equations (19) and (20), respectively. In so doing, we rely exclusively on $\alpha=0.999$, and the jump series depicted in the bottom panels of Figures 1A-C.[29] To facilitate the exposition, we omit the 0.999 subscript on the $J_{t,0.999}$ and $C_{t,0.999}$ series in what follows.

A. The HAR-RV-CJ Model

---

[27] Of course, exponential smoothing automatically induces some serial correlation, so that the dependencies in the figure should be carefully interpreted. Note also that the common scale enforced on the three jump size panels tend to hide the subtle, but systematic, decline in the sizes of the jumps for the DM/$ market over the sample period, as evidenced by the highly significant LB10, $J_{t,\alpha}^{+}$ in Table 3; see also the plot for the raw jump series in Figure 1A.

[28] For instance, following Bates (2000), Pan (2002) and Eraker (2004), the jump intensity could be specified as a function of the diffusive, or spot volatility. In the notation of equation (1), $\lambda(t)=\lambda_0+\lambda_1\sigma(t)$. Similarly, the sizes of the jumps could be allowed to depend on the volatility and/or lagged past jump sizes, as in, e.g., $\kappa^2(t)=\kappa_0+\kappa_1\sigma(t)+\kappa_2\kappa^2(t_{-1})$. The recent discrete-time parametric model estimates reported in Chan and Maheu (2002) and McCurdy and Maheu (2004) also point to the existence of time-varying jump intensities in U.S. equity index returns.

[29] By "optimally" choosing α, it may be possible to further improve upon the empirical results reported below. However, for simplicity and to guard against obvious data snooping biases, we simply restrict $\alpha=0.999$.

Defining the normalized multi-period jump and continuous sample path variability measures,

$$J_{t,t+h} = h^{-1}[\, J_{t+1} \; + J_{t+2} \; + ... + J_{t+h}\,] \;, \tag{24}$$

and,

$$C_{t,t+h} = h^{-1}[\, C_{t+1} \; + C_{t+2} \; + ... + C_{t+h}\,] \;, \tag{25}$$

respectively, the new HAR-RV-CJ model may be expressed as

$$\begin{aligned} RV_{t,t+h} \;\; = \;\; & \beta_0 \;\; + \;\; \beta_{CD}\, C_t \;\; + \;\; \beta_{CW}\, C_{t\text{-}5,t} \;\; + \;\; \beta_{CM}\, C_{t\text{-}22,t} \;\; + \\ & + \;\; \beta_{JD}\, J_t \;\;\; + \;\; \beta_{JW}\, J_{t\text{-}5,t} \;\;\; + \;\; \beta_{JM}\, J_{t\text{-}22,t} \;\;\; + \;\; \varepsilon_{t,t+h} \,. \end{aligned} \tag{26}$$

The model obviously nests the HAR-RV-J model in (11) for $\beta_D = \beta_{CD} + \beta_{JD}$, $\beta_W = \beta_{CW} + \beta_{JW}$, $\beta_M = \beta_{CM} + \beta_{JM}$, and $\beta_J = \beta_{JD}$, but in general it allows for a more flexible dynamic lag structure.[30]

Turning to the empirical estimates in the first three columns in Tables 4A-C, most of the coefficient estimates for the jump components are insignificant. In other words, the predictability in the HAR-RV realized volatility regressions are almost exclusively due to the continuous sample path components. For the DM/$ and the S&P500 the HAR-RV-CJ models typically result in relatively modest increases in the R^2 of less than *0.01* in absolute value compared to the HAR-RV-J models in Tables 2A-B, whereas for the T-Bond market the improvements are closer to *0.02,* or about *4-5* percent in a relative sense. The test for first, sixth, and twenty-third order serial correlation in the residuals from the estimated daily (h=1), weekly (h=5) and monthly (h=22) regressions, also indicate that the HAR-RV-CJ models have eliminated most of the strong serial correlation in the $RV_{t,t+h}$ series.[31] Still, some statistically significant autocorrelations remain at higher lags for some of the models, suggesting that further refinements might be possible.

These same qualitative results carry over to the non-linear HAR-RV-CJ models cast in standard deviation and logarithmic form; i.e.,

$$(RV_{t,t+h})^{1/2} \;\; = \;\; \beta_0 \;\; + \;\; \beta_{CD}\, C_t^{1/2} \;\; + \;\; \beta_{CW}\,(C_{t\text{-}5,t})^{1/2} \;\; + \;\; \beta_{CM}\,(C_{t\text{-}22,t})^{1/2} \tag{27}$$

[30] For the two models to be nested the (implicit) choice of α employed in the measurements of $J_{t,t+h}$ and $C_{t,t+h}$ should, of course, also be the same across models.

[31] The lag one, six and twenty-three autocorrelations for the residuals from the three DM/$ HAR-RV-CJ models equal *-0.014*, *-0.026* and *0.003*, respectively. For the S&P500 the same residual autocorrelations are *-0.011*, *0.007*, and *-0.081*, while for the T-Bond market the correlations equal *0.003*, *-0.006* and *-0.060*, respectively.

$$+ \ \beta_{JD} J_t^{1/2} \ + \ \beta_{JW} (J_{t\text{-}5,t})^{1/2} \ + \ \beta_{JM} (J_{t\text{-}22,t})^{1/2} \ + \ \varepsilon_{t,t+h} \, ,$$

and,

$$\begin{aligned} log(RV_{t,t+h}) \ &= \ \beta_0 \ + \ \beta_{CD}\, log(C_t) \ + \ \beta_{CW}\, log(C_{t\text{-}5,t}) \ + \ \beta_{CM}\, log(C_{t\text{-}22,t}) \\ &\quad + \ \beta_{JD}\, log(J_t + 1) \ + \ \beta_{JW}\, log(J_{t\text{-}5,t} + 1) \ + \ \beta_{JM}\, log(C_{t\text{-}22,t} + 1) \ + \ \varepsilon_{t,t+h} \, , \end{aligned} \qquad (28)$$

respectively. The coefficient estimates for the jump components, reported in the last six columns in Tables 4A-C, are again insignificant for most of the markets and forecast horizons. In contrast, the estimates of β_{CD}, β_{CW} and β_{CM}, which quantify the impact of the continuous sample path variability on the total future variation, are all generally highly significant.

To further illustrate the predictability afforded by the HAR-RV-CJ model, Figures 4A-C plot the daily, weekly, and monthly realized volatilities (again in standard deviation form) together with the corresponding forecasts from the model in equation (27). The close coherence between the different pairs of realizations and forecasts is immediately evident across all of the markets and forecast horizons. Visual inspection of the graphs also show that the volatility in the U.S. T-Bond market is the least predictable, followed by the DM/$, and then the S&P500. Nonetheless, the forecasts for the T-Bond volatilities still track the overall patterns fairly well, especially for the longer weekly and monthly horizons.

All told, these results further underscore the potential benefit from a volatility forecasting perspective of separately measuring the individual components of the realized volatility. It is possible that even further improvements may be obtained by a more structured approach in which the jump component, J_t, and the continuous sample path component, C_t, are each modeled separately. These individual models for J_t and C_t could then be used in the construction of separate out-of-sample forecasts for each of the components, as well as combined forecasts for the total realized variation process, $RV_{t,t+h} = C_{t,t+h} + J_{t,t+h}$. We leave further work along these lines for future research.

VII. Concluding Remarks

Building on recent theoretical results in Barndorff-Nielsen and Shephard (2004a, 2005) for so-called bi-power variation measures, we provide a simple and easy-to-implement practical framework for measuring "significant" jumps in financial asset prices. Applying the theory to more than a decade long sample of high-frequency prices from the foreign exchange, equity, and fixed income markets, we find that the procedure works well empirically. Consistent with recent parametric model estimates, our non-parametric measurements of the squared jumps are much less persistent (and predictable) than the continuous sample path, or integrated volatility, component of the quadratic variation process. Meanwhile, the high-frequency data underlying our estimates allow us to identify many more jumps than do the parametric models based on daily or courser frequency data hitherto reported in the literature. It also appears that many of the most significant jumps are readily associated with specific macroeconomic news announcements. When separately including the continuous sample path and jump variability measures in a simple linear volatility forecasting model, we find that only the former measure caries any predictive power, in turn resulting in significant gains relative to the simple reduced form realized volatility forecasting models advocated in some of the recent literature.

The ideas and empirical results presented here are suggestive of several interesting extensions. First, it seems natural that jump risk may be priced differently from easier-to-hedge continuous price variability; see, e.g., Santa-Clara and Yan (2004). Hence, separately modeling and forecasting the continuous sample path, or integrated volatility, and jump components of the quadratic variation process, as discussed above, is likely to result in important improvements in derivatives and other pricing decisions. Second, our choice of a five-minute sampling frequency and the new skip-one realized bi-power and tri-power variation measures to mitigate the market microstructure frictions in the high-frequency data were guided by somewhat ad hoc considerations. It would be interesting to further investigate the "optimal" choice of sampling frequency, or the use of sub-sampling schemes in the construction of the bi- and tri-power variation measures. The related results for the realized variation measures in, e.g., Bandi and Russell (2004a,b), Hansen and Lunde (2004a,b) and Zhang, Aït-Sahalia and Mykland (2005), should be helpful. Third, if interest centers exclusively on volatility forecasting, the use of more traditional robust power variation measures defined by the sum of absolute high-frequency returns raised to powers less than two might afford additional gains over and above the improvements provided by the bi-power variation and significant jump measures used here; the recent empirical results in Forsberg and Ghysels (2004) are suggestive. Fourth, casual empirical observations suggest that very large price moves, or jumps, often occur simultaneously across different markets. It would be interesting to extend the present analysis to a multivariate framework explicitly incorporating such commonalities through the use of quadratic covariation and appropriately defined co-power variation measures; the abstract theoretical results in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2005) are intriguing. In addition to allowing for more accurate statistical identification of the most important, or significant, jumps, this should also enhance our understanding of the underlying economic influences that "drive" financial markets and prices.

References

Aït-Sahalia, Y. (2002), "Telling from Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion," *Journal of Finance*, 57, 2075-2121.

Aït-Sahalia, Y. (2004), "Disentangling Volatility from Jumps," *Journal of Financial Economics*, 74, 487-528.

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*, forthcoming.

Andersen, T.G., L. Benzoni and J. Lund (2002), "Estimating Jump-Diffusions for Equity Returns," *Journal of Finance*, 57, 1239-1284.

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

Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2003), "Parametric and Non-Parametric Volatility Measurement," in *Handbook of Financial Econometrics* (L.P Hansen and Y. Aït-Sahalia, eds.). Elsevier Science, New York, forthcoming.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43-76.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2000), "Great Realizations," *Risk*, 13, 105-108.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.

Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38-62.

Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2005), "Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets," Manuscript, Northwestern University, Duke University, University of Pennsylvania, and University of Rochester.

Andersen, T.G., T. Bollerslev and N. Meddahi (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review,* 45, 1079-1110.

Andersen, T.G., T. Bollerslev and N. Meddahi (2005), "Correcting the Errors: Volatility Forecast Evaluation using High-frequency Data and Realized Volatilities," *Econometrica*, 72, 279-296.

Andreou, E. and E. Ghysels (2002), "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results," *Journal of Business and Economic Statistics*, 20, 363-376.

Areal, N.M.P.C. and S.J. Taylor (2002), "The Realized Volatility of FTSE-100 Futures Prices," *Journal of Futures Market*, 22, 627-648.

Bai, X., J.R. Russell and G.C. Tiao (2004), "Effects of Non-Normality and Dependence on the Precision of Variance Estimates using High-Frequency Financial Data," Manuscript, University of Chicago.

Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3-30.

Ball, C.A. and W.N. Torous (1983), "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis*, 18, 53-65.

Bandi, F.M. and J.R. Russell (2004a), "Microstructure Noise, Realized Volatility, and Optimal Sampling," Manuscript, University of Chicago.

Bandi, F.M. and J.R. Russell (2004b), "Separating Microstructure Noise from Volatility" Manuscript, University of Chicago.

Barndorff-Nielsen, O.E., S.E. Graversen and N. Shephard (2004), "Power Variation and Stochastic Volatility: A Review and Some New Results," *Journal of Applied Probability*, 41A, 133-143.

Barndorff-Nielsen, O.E. and N. Shephard (2002a), "Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, 64, 253-280.

Barndorff-Nielsen, O.E. and N. Shephard (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457-478.

Barndorff-Nielsen, O.E. and N. Shephard (2003), "Realised Power Variation and Stochastic Volatility," *Bernoulli*, 9, 243-265.

Barndorff-Nielsen, O.E. and N. Shephard (2004a), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1-37.

Barndorff-Nielsen, O.E. and N. Shephard (2004b), "How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility," in *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg* (D. Andrews, J. Powell, P.A. Ruud, and J. Stock, eds.). Cambridge, UK: Cambridge University Press.

Barndorff-Nielsen, O.E. and N. Shephard (2005), "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, forthcoming.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, M. Podolskij and N. Shephard (2005), "A Central Limit Theorem for Realised Power and Bipower Variation of Continuous Semimartingales," in *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev* (Y. Kabanov and R. Lipster, eds.). Springer Verlag.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2005), "Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise," Manuscript, Stanford University.

Basak, G., N.H. Chan, and W. Palma (2001), "The Approximation of Long-Memory Processes by an ARMA Model," *Journal of Forecasting*, 20, 367-389.

Bates, D.S. (2000), "Post-`87 Crash fears in the S&P500 Futures Option Market," *Journal of Econometrics*, 94, 181-238.

Beckers, S. (1981), "A Note on Estimating the Parameters of the Diffusion-Jump Model of Stock Returns," *Journal of Financial and Quantitative Analysis*, 16, 127-140.

Bollen, B. and B. Inder (2002), "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551-562.

Bollerslev, T. and J.H. Wright (2001), "Volatility Forecasting, High-Frequency Data, and Frequency Domain Inference," *Review of Economics and Statistics*, 83, 596-602.

Breidt, F.J., N. Crato and P. de Lima (1998), "The Detection and Estimation of Long-Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325-348.

Calvet, L. and A. Fisher (2001), "Forecasting Multifractal Volatility," *Journal of Econometrics*, 105, 27-58.

Calvet, L. and A. Fisher (2002), "Multifractality in Asset Returns: Theory and Evidence," *Review of Economics and Statistics*, 84, 381-406.

Chan, W.H. and J.M. Maheu (2002), "Conditional Jump Dynamics in Stock Market Returns," *Journal of Business and Economic Statistics*, 20, 377-389.

Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2003), "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225-257.

Christoffersen, P. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

Comte, F. and E. Renault (1998), "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291-323.

Corsi, F. (2003), "A Simple Long Memory Model of Realized Volatility," Manuscript, University of Southern Switzerland.

Corsi, F., Zumbach, U.A. Müller and M. Dacorogna (2001), "Consistent High-Precision Volatility from High-frequency Data," *Economic Notes*, 30, 183-204.

Cox, D.R. (1991), "Long-Range Dependence, Non-Linearity and Time Irreversibility," *Journal of Time Series Analysis*, 12, 329-335.

Curci, G. and F. Corsi (2003), "A Discrete Sine Transform Approach for Realized Volatility Measurement," Manuscript, University of Pisa.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen and O.V. Pictet (1993), "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance*, 12, 413-438.

Dacorogna, M.M., R. Gencay, U.A. Müller, O.V. Pictet and R.B. Olsen (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

Das, S.R. (2002), "The Surprise Element: Jumps in Interest Rates," *Journal of Econometrics*, 106, 27-65.

Deo, R., C. Hurvich, and Y. Lu (2005), "Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment," *Journal of Econometrics*, forthcoming.

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.

Drost, F.C., T.E. Nijman and B.J.M. Werker (1998), "Estimation and Testing in Models Containing Both Jumps and Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*, 16, 237-243.

Engle, R.F. (2000), "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1-22.

Engle, R.F. (2002), "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425-446.

Engle, R.F. and G.M. Gallo (2005), "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics*, forthcoming.

Engle, R.F. and G.J. Lee (1999), "A Permanent and Transitory Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger* (R.F. Engle and H. White, eds.). Oxford, UK: Oxford University Press.

Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367-1403.

Eraker, B., M.S. Johannes and N.G. Polson (2003), "The Impact of Jumps in Volatility," *Journal of Finance*, 58, 1269-1300.

Forsberg, L. and T. Bollerslev (2002), "Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the Euro): The GARCH-NIG Model," *Journal of Applied Econometrics*, 17, 535-548.

Forsberg, L. and E. Ghysels (2004), "Why Do Absolute Returns Predict Volatility So Well?," Manuscript, University of North Carolina, Chapel Hill..

Gallant, A.R., C.T. Hsu and G. Tauchen (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, 81, 617-631.

Ghysels, E, P. Santa-Clara and R. Valkanov (2004), "The MIDAS Touch: Mixed Data Sampling Regressions," Manuscript, University of North Carolina and UCLA.

Ghysels, E, P. Santa-Clara and R. Valkanov (2005), "Predicting Volatility: How to Get the Most out of Returns Data Sampled at Different Frequencies," *Journal of Econometrics*, forthcoming.

Hansen, P.R. and A. Lunde (2004a), "Realized Variance and Market Microstructure Noise," Manuscript, Stanford University.

Hansen, P.R. and A. Lunde (2004b), "An Unbiased Measure of Realized Variance," Manuscript, Stanford University.

Hsu, N.J. and F.J. Breidt (2003), “Bayesian Analysis of Fractionally Integrated ARMA with Additive Noise,” *Journal of Forecasting*, 22, 491-514.

Huang, X. and G. Tauchen (2005), “The Relative Contribution of Jumps to Total Price Variation,” Manuscript, Duke University.

Jarrow, R.A. and E. Rosenfeld (1984), “Jump Risk and the Intertemporal Capital Asset Pricing Model,” *Journal of Business*, 57, 337-351.

Johannes, M. (2004), “The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models,” *Journal of Finance*, 59, 227-260.

Johannes, M., R. Kumar and N.G. Polson (1999), “State Dependent Jump Models: How do US Equity Markets Jump?” Manuscript, University of Chicago.

Jorion, P. (1988), “On Jump Processes in the Foreign Exchange and Stock Markets,” *Review of Financial Studies*, 1, 427-445.

Khalaf, L., J.-D. Saphores and J.-F. Bilodeau (2003), “Simulation-Based Exact Jump Tests in Models with Conditional Heteroskedasticity,” *Journal of Economic Dynamics and Control*, 28, 531-553.

Koopman, S.J., B. Jungbacker and E. Hol (2005), “Forecasting Daily Variability of the S&P100 Stock Index Using Historical, Realised and Implied Volatility Measures,” *Journal of Empirical Finance*, forthcoming.

Maheu, J.M. and T.H. McCurdy (2004), “News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns,” *Journal of Finance*, 59, 755-793.

Malliavin, P. and M.E. Mancino (2002), “Fourier Series Methods for Measurement of Multivariate Volatilities,” *Finance and Stochastics*, 6, 49-61.

Man, K.S. (2003), “Long Memory Time Series and Short Term Forecasts,” *International Journal of Forecasting*, 19, 477-491.

Mancini, C. (2003), “Statistics of a Poisson-Gaussian Process,” Manuscript, Universita di Firenze.

Mancini, C. (2004), “Estimation of the Parameters of Jump of a General Poisson-Diffusion Model,” *Scandinavian Actuarial Journal*, 104, 42-52.

Martens, M. (2002), “Measuring and Forecasting S&P500 Index Futures Volatility Using High-Frequency Data,” *Journal of Futures Markets*, 22, 497-518.

Martens, M., (2004), “Estimating Unbiased and Precise Realized Covariances,” Manuscript, Erasmus University Rotterdam.

Martens, M., D. van Dijk and M. de Pooter (2004), “Modeling and Forecasting &P500 Volatility: Long-Memory, Structural Breaks and Nonlinearity,” Manuscript, Erasmus University Rotterdam.

Meddahi, N. (2002), “A Theoretical Comparison Between Integrated and Realized Volatility,” *Journal of Applied Econometrics*, 17, 479-508.

Merton, R.C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125-144.

Merton, R.C. (1980), "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323-361.

Müller, U.A., M.M. Dacorogna, R.B. Olsen, O.V. Puctet, M. Schwarz, and C. Morgenegg (1990), "Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis," *Journal of Banking and Finance*, 14, 1189-1208.

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. (1992), "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61-90.

Nielsen, M.Ø. and P.H. Frederiksen (2004), "Finite Sample Accuracy of Integrated Volatility Estimators," Manuscript, Cornell University.

O'Connell, P.E. (1971), "A Simple Stochastic Modelling of Hurst's Law," *Proceedings of International Symposium on Mathematical Models in Hydrology*, 1, 169-187.

Oomen, R.C.A. (2002), "Modelling Realized Variance when Returns are Serially Correlated," Manuscript, University of Warwick.

Oomen, R.C.A. (2004), "Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling," Manuscript, University of Warwick.

Pan, J. (2002), "the Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study," *Journal of Financial Economics*, 63, 3-50.

Pong, S., M.B. Shackleton, S.J. Taylor and X. Xu (2004), "Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models," *Journal of Banking and Finance*, forthcoming.

Robinson, P.M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regressions," *Journal of Econometrics*, 47, 67-84.

Russell, J.R. and R.F. Engle (2005), "A Discrete-State Continuous-Time Model of Financial Transaction Prices and Times: The Autoregressive Conditional Multinomial - Autoregressive Duration Model," *Journal of Business and Economic Statistics*, 23, 166-180.

Rydberg, T. and N. Shephard (2003), "The Dynamics of Trade-by-Trade Price Movements," *Journal of Financial Econometrics*, 1, 2-25.

Santa-Clara, P. and S. Yan (2004), "Jump and Volatility Risk and Risk Premia: A New Model and Lessons from S&P500 Options," Manuscript, UCLA.

Tiao, G.C. and R.S. Tsay (1994), "Some Advances in Non-Linear and Adaptive Modeling in Time Series," *Journal of Forecasting*, 14, 109-131.

Thomakos, D.D. and T. Wang (2003), “Realized Volatility in the Futures Market,” *Journal of Empirical Finance*, 10, 321-353.

Vlaar, P.J.G. and F.C. Palm (1993), “The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity, and Jumps,” *Journal of Business and Economic Statistics*, 11, 351-360.

Zhang, L., Y. Aït-Sahalia and P.A. Mykland (2005), “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, forthcoming.

Zhang, L. (2004), “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” Manuscript, Carnegie Mellon University.

Zhou, B. (1996), “High-Frequency Data and Volatility in Foreign Exchange Rates,” *Journal of Business and Economic Statistics*, 14, 45-52.

Table 1A
Summary Statistics for Daily DM/$ Realized Volatilities and Jumps

	RV_t	$RV_t^{1/2}$	$log(RV_t)$	J_t	$J_t^{1/2}$	$log(J_t+1)$
Mean	0.508	0.670	-0.915	0.037	0.129	0.033
St.Dev.	0.453	0.245	0.657	0.110	0.142	0.072
Skewness	3.925	1.784	0.408	16.52	2.496	7.787
Kurtosis	26.88	8.516	3.475	434.2	18.20	108.5
Min.	0.052	0.227	-2.961	0.000	0.000	0.000
Max.	5.245	2.290	1.657	3.566	1.889	1.519
LB_{10}	3786	5714	7060	16.58	119.4	63.19

Table 1B
Summary Statistics for Daily S&P500 Realized Volatilities and Jumps

	RV_t	$RV_t^{1/2}$	$log(RV_t)$	J_t	$J_t^{1/2}$	$log(J_t+1)$
Mean	1.137	0.927	-0.400	0.164	0.232	0.097
St.Dev.	1.848	0.527	0.965	0.964	0.332	0.237
Skewness	7.672	2.545	0.375	20.68	5.585	6.386
Kurtosis	95.79	14.93	3.125	551.9	59.69	59.27
Min.	0.058	0.240	-2.850	0.000	0.000	0.000
Max.	36.42	6.035	3.595	31.88	5.646	3.493
LB_{10}	5750	12184	15992	558.0	1868	2295

Table 1C
Summary Statistics for Daily U.S. T-Bond Realized Volatilities and Jumps

	RV_t	$RV_t^{1/2}$	$log(RV_t)$	J_t	$J_t^{1/2}$	$log(J_t+1)$
Mean	0.286	0.506	-1.468	0.036	0.146	0.033
St.Dev.	0.222	0.173	0.638	0.069	0.120	0.055
Skewness	3.051	1.352	0.262	8.732	1.667	5.662
Kurtosis	20.05	6.129	3.081	144.6	10.02	57.42
Min.	0.026	0.163	-3.633	0.000	0.000	0.000
Max.	2.968	1.723	1.088	1.714	1.309	0.998
LB_{10}	1022	1718	2238	20.53	34.10	26.95

Key: The first six rows in each of the panels report the sample mean, standard deviation, skewness and kurtosis, along with the sample minimum and maximum. The rows labeled LB_{10} give the Ljung-Box test statistic for up to tenth order serial correlation. The daily realized volatilities and jumps for the DM/$ in Panel A are constructed from five-minute returns spanning the period from December 1986 through June 1999, for a total of 3,045 daily observations. The daily realized volatilities and jumps for the S&P500 and U.S. T-Bonds in Panels B and C are based on five-minute returns from January 1990 through December 2002, for a total of 3,213 observations.

Table 2A

Daily, Weekly, and Monthly DM/$ HAR-RV-J Regressions

$$RV_{t,t+h} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \beta_J J_t + \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_D RV_t^{1/2} + \beta_W (RV_{t-5,t})^{1/2} + \beta_M (RV_{t-22,t})^{1/2} + \beta_J J_t^{1/2} + \epsilon_{t,t+h}$$

$$log(RV_{t,t+h}) = \beta_0 + \beta_D log(RV_t) + \beta_W log(RV_{t-5,t}) + \beta_M log(RV_{t-22,t}) + \beta_J log(J_t+1) + \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.083	0.132	0.231	0.096	0.159	0.293	-0.122	-0.142	-0.269
	(0.015)	(0.018)	(0.025)	(0.015)	(0.021)	(0.034)	(0.021)	(0.030)	(0.048)
β_D	0.430	0.222	0.110	0.392	0.216	0.124	0.408	0.252	0.162
	(0.043)	(0.040)	(0.022)	(0.033)	(0.028)	(0.020)	(0.030)	(0.029)	(0.027)
β_W	0.196	0.216	0.218	0.248	0.264	0.243	0.271	0.265	0.221
	(0.063)	(0.055)	(0.043)	(0.046)	(0.050)	(0.040)	(0.042)	(0.051)	(0.039)
β_M	0.244	0.323	0.225	0.223	0.296	0.219	0.212	0.284	0.226
	(0.061)	(0.068)	(0.062)	(0.041)	(0.056)	(0.063)	(0.036)	(0.054)	(0.065)
β_J	-0.486	-0.297	-0.166	-0.103	-0.047	-0.026	-0.762	-0.590	-0.395
	(0.096)	(0.070)	(0.056)	(0.039)	(0.031)	(0.029)	(0.161)	(0.176)	(0.178)
$R^2_{HAR-RV-J}$	0.364	0.417	0.353	0.431	0.472	0.386	0.476	0.502	0.407
R^2_{HAR}	0.252	0.261	0.215	0.261	0.272	0.225	0.160	0.159	0.138

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-J volatility forecast regressions. The realized volatilities and jumps are constructed from five-minute returns spanning the period from December 1986 through June 1999, for a total of 3,045 daily observations. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively. The last two rows labeled $R^2_{HAR-RV-J}$ and R^2_{HAR} give the coefficients of multiple correlation from the HAR-RV-J model along with a HAR model without any jumps replacing the realized volatilities on the right-hand-side of the regression with the corresponding lagged daily, weekly, and monthly squared returns.

Table 2B
Daily, Weekly, and Monthly S&P500 HAR-RV-J Regressions

$$RV_{t,t+h} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \beta_J J_t + \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_D RV_t^{1/2} + \beta_W (RV_{t-5,t})^{1/2} + \beta_M (RV_{t-22,t})^{1/2} + \beta_J J_t^{1/2} + \epsilon_{t,t+h}$$

$$log(RV_{t,t+h}) = \beta_0 + \beta_D log(RV_t) + \beta_W log(RV_{t-5,t}) + \beta_M log(RV_{t-22,t}) + \beta_J log(J_t+1) + \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.088	0.186	0.387	0.060	0.102	0.205	-0.065	0.001	0.022
	(0.054)	(0.066)	(0.074)	(0.020)	(0.030)	(0.038)	(0.015)	(0.019)	(0.036)
β_D	0.341	0.220	0.109	0.375	0.262	0.169	0.348	0.235	0.164
	(0.094)	(0.064)	(0.040)	(0.041)	(0.038)	(0.029)	(0.028)	(0.025)	(0.024)
β_W	0.485	0.430	0.287	0.354	0.395	0.291	0.329	0.363	0.260
	(0.111)	(0.097)	(0.093)	(0.064)	(0.067)	(0.071)	(0.041)	(0.050)	(0.049)
β_M	0.165	0.220	0.279	0.239	0.270	0.357	0.285	0.334	0.444
	(0.067)	(0.079)	(0.088)	(0.040)	(0.058)	(0.069)	(0.031)	(0.046)	(0.054)
β_J	-0.472	-0.228	-0.075	-0.213	-0.134	-0.054	-0.260	-0.187	-0.105
	(0.102)	(0.078)	(0.067)	(0.051)	(0.041)	(0.046)	(0.062)	(0.064)	(0.080)
$R^2_{HAR-RV-J}$	0.415	0.569	0.474	0.604	0.697	0.634	0.693	0.761	0.727
R^2_{HAR}	0.248	0.239	0.159	0.322	0.320	0.275	0.197	0.219	0.218

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-J volatility forecast regressions. The realized volatilities and jumps are constructed from five-minute returns spanning the period from January 1990 through December 2002, for a total of 3,213 daily observations. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively. The last two rows labeled $R^2_{HAR-RV-J}$ and R^2_{HAR} give the coefficients of multiple correlation from the HAR-RV-J along with a HAR model without jumps replacing the realized volatilities on the right-hand-side of the regression with the corresponding lagged daily, weekly, and monthly squared returns.

Table 2C

Daily, Weekly, and Monthly U.S. T-Bond HAR-RV-J Regressions

$$RV_{t,t+h} \ = \ \beta_0 \ + \ \beta_D RV_t \ + \ \beta_W RV_{t-5,t} \ + \ \beta_M RV_{t-22,t} \ + \ \beta_J J_t \ + \ \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} \ = \ \beta_0 \ + \ \beta_D RV_t^{1/2} \ + \ \beta_W (RV_{t-5,t})^{1/2} \ + \ \beta_M (RV_{t-22,t})^{1/2} \ + \ \beta_J J_t^{1/2} \ + \ \epsilon_{t,t+h}$$

$$log(RV_{t,t+h}) \ = \ \beta_0 \ + \ \beta_D log(RV_t) \ + \ \beta_W log(RV_{t-5,t}) \ + \ \beta_M log(RV_{t-22,t}) \ + \ \beta_J log(J_t+1) \ + \ \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.077	0.088	0.126	0.118	0.152	0.222	-0.353	-0.323	-0.473
	(0.012)	(0.013)	(0.017)	(0.017)	(0.021)	(0.031)	(0.044)	(0.051)	(0.080)
β_D	0.074	0.084	0.044	0.066	0.082	0.044	0.104	0.100	0.063
	(0.031)	(0.016)	(0.011)	(0.026)	(0.014)	(0.009)	(0.024)	(0.015)	(0.012)
β_W	0.317	0.217	0.160	0.319	0.215	0.162	0.324	0.206	0.154
	(0.050)	(0.043)	(0.040)	(0.045)	(0.039)	(0.036)	(0.043)	(0.039)	(0.032)
β_M	0.358	0.416	0.373	0.369	0.429	0.388	0.379	0.440	0.395
	(0.056)	(0.055)	(0.070)	(0.046)	(0.051)	(0.068)	(0.046)	(0.053)	(0.070)
β_J	-0.152	-0.202	-0.140	-0.047	-0.062	-0.042	-0.869	-0.693	-0.545
	(0.099)	(0.042)	(0.034)	(0.032)	(0.018)	(0.015)	(0.240)	(0.129)	(0.123)
$R^2_{HAR\text{-}RV\text{-}J}$	0.130	0.308	0.340	0.171	0.332	0.360	0.200	0.351	0.379
R^2_{HAR}	0.067	0.128	0.105	0.072	0.117	0.093	0.025	0.042	0.035

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-J volatility forecast regressions. The realized volatilities and jumps are constructed from five-minute returns spanning the period from January 1990 through December 2002, for a total of 3,213 daily observations. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively. The last two rows labeled $R^2_{HAR\text{-}RV\text{-}J}$ and R^2_{HAR} give the coefficients of multiple correlation from the HAR-RV-J model along with a HAR model without jumps replacing the realized volatilities on the right-hand-side of the regression with the corresponding lagged daily, weekly, and monthly squared returns.

Table 3A

Summary Statistics for Significant Daily DM/$ Jumps

α	*0.500*	*0.950*	*0.990*	*0.999*	*0.9999*
Prop.	0.859	0.409	0.254	0.137	0.083
Mean.	0.059	0.047	0.037	0.028	0.021
St.Dev.	0.136	0.137	0.135	0.131	0.127
LB_{10}, $J_{t,\alpha}$	65.49	26.30	6.197	3.129	2.414
LR , $I(J_{t,\alpha}>0)$	0.746	2.525	0.224	0.994	0.776
LB_{10} ,$D_{t,\alpha}$	10.78	9.900	7.821	6.230	19.95
LB_{10} ,$J_{t,\alpha}{}^{+}$	73.62	116.4	94.19	87.69	34.57

Table 3B

Summary Statistics for Significant Daily S&P500 Jumps

α	*0.500*	*0.950*	*0.990*	*0.999*	*0.9999*
Prop.	0.737	0.255	0.141	0.076	0.051
Mean.	0.163	0.132	0.111	0.095	0.086
St.Dev.	0.961	0.961	0.958	0.953	0.950
LB_{10}, $J_{t,\alpha}$	300.6	271.9	266.4	260.9	221.6
LR , $I(J_{t,\alpha}>0)$	2.415	1.483	12.83	8.418	7.824
LB_{10}, $D_{t,\alpha}$	50.83	31.47	22.67	36.18	49.25
LB_{10}, $J_{t,\alpha}{}^{+}$	320.8	146.0	77.06	35.11	25.49

Table 3C

Summary Statistics for Significant Daily U.S. T-Bond Jumps

α	*0.500*	*0.950*	*0.990*	*0.999*	*0.9999*
Prop.	0.860	0.418	0.254	0.132	0.076
Mean.	0.048	0.038	0.030	0.021	0.016
St.Dev.	0.094	0.096	0.096	0.090	0.085
LB_{10}, $J_{t,\alpha}$	30.34	30.37	27.85	19.80	18.85
LR , $I(J_{t,\alpha}>0)$	4.746	21.62	13.69	3.743	1.913
LB_{10}, $D_{t,\alpha}$	45.55	100.1	59.86	103.3	81.42
LB_{10}, $J_{t,\alpha}{}^{+}$	21.23	17.18	15.18	9.090	11.98

Key: The significant jumps for each of the three market, $J_{t,\alpha}$, are determined by equation (19) along with the staggered bi-power and tri-power variation measures in equations (23) and (23), respectively. The first row in each of the panels gives the proportion significant jump days for each of the different α's. The next two rows report the corresponding mean and standard deviation of the jump series, while the row labeled LB_{10}, $J_{t,\alpha}$ gives the Ljung-Box tests for up to tenth order serial correlation. LR , $I(J_{t,\alpha}>0)$ denotes the Likelihood Ratio test for *i.i.d.* jump occurrences against a first Markov chain, while LB_{10}, $D_{t,\alpha}$ and LB_{10}, $J_{t,\alpha}{}^{+}$ refer to the Ljung-Box tests for serial correlation in the corresponding durations, or times between jumps, and the sizes of the significant jumps, respectively.

Table 4A
Daily, Weekly, and Monthly DM/$ HAR-RV-CJ Regressions

$$RV_{t,t+h} = \beta_0 + \beta_{CD} C_t + \beta_{CW} C_{t-5,t} + \beta_{CM} C_{t-22,t} + \beta_{JD} J_t + \beta_{JW} J_{t-5,t} + \beta_{JM} J_{t-22,t} + \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_{CD} C_t^{1/2} + \beta_{CW} (C_{t-5,t})^{1/2} + \beta_{CM} (C_{t-22,t})^{1/2} + \beta_{JD} J_t^{1/2} + \beta_{JW} (J_{t-5,t})^{1/2} + \beta_{JM} (J_{t-22,t})^{1/2} + \epsilon_{t,t+h}$$

$$log(RV_{t,t+h}) = \beta_0 + \beta_{CD} log(C_t) + \beta_{CW} log(C_{t-5,t}) + \beta_{CM} log(C_{t-22,t}) + \beta_{JD} log(J_t+1) + \beta_{JW} log(J_{t-5,t}+1) + \beta_{JM} log(C_{t-22,t}+1) + \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.083	0.131	0.231	0.096	0.158	0.292	-0.095	-0.114	-0.249
	(0.015)	(0.018)	(0.025)	(0.015)	(0.021)	(0.034)	(0.024)	(0.036)	(0.057)
β_{CD}	0.407	0.210	0.101	0.397	0.222	0.127	0.369	0.205	0.130
	(0.044)	(0.040)	(0.021)	(0.032)	(0.029)	(0.019)	(0.026)	(0.021)	(0.016)
β_{CW}	0.256	0.271	0.259	0.264	0.289	0.264	0.295	0.318	0.258
	(0.077)	(0.054)	(0.046)	(0.048)	(0.051)	(0.042)	(0.039)	(0.048)	(0.040)
β_{CM}	0.226	0.308	0.217	0.212	0.281	0.205	0.217	0.270	0.213
	(0.072)	(0.078)	(0.074)	(0.044)	(0.060)	(0.068)	(0.036)	(0.055)	(0.071)
β_{JD}	0.096	0.006	-0.002	0.022	0.001	0.003	0.043	0.024	-0.004
	(0.089)	(0.040)	(0.017)	(0.027)	(0.017)	(0.010)	(0.111)	(0.076)	(0.044)
β_{JW}	-0.191	-0.179	-0.073	-0.006	0.001	0.002	-0.076	-0.317	-0.127
	(0.168)	(0.199)	(0.125)	(0.033)	(0.044)	(0.028)	(0.239)	(0.327)	(0.242)
β_{JM}	-0.001	0.055	-0.014	-0.034	-0.011	0.014	-0.690	-0.301	-0.261
	(0.329)	(0.460)	(0.604)	(0.057)	(0.087)	(0.127)	(0.408)	(0.668)	(0.990)
$R^2_{HAR\text{-}RV\text{-}CJ}$	0.368	0.427	0.361	0.443	0.486	0.397	0.485	0.514	0.415

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-CJ volatility forecast regressions. All of the realized volatility measures are constructed from five-minute returns spanning the period from December 1986 through June 1999, for a total of 3,045 daily observations. The weekly and monthly measures are given by the scaled sum of the corresponding daily measures. The significant daily jump and continuous sample path variability measures are based on equations (19) and (20), respectively, along with the staggered power variation measures in equations (22) and (23), using a critical value of $\alpha = 0.999$. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively.

Table 4B

Daily, Weekly, and Monthly S&P500 HAR-RV-CJ Regressions

$$RV_{t,t+h} = \beta_0 + \beta_{CD} C_t + \beta_{CW} C_{t-5,t} + \beta_{CM} C_{t-22,t} + \beta_{JD} J_t + \beta_{JW} J_{t-5,t} + \beta_{JM} J_{t-22,t} + \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_{CD} C_t^{1/2} + \beta_{CW} (C_{t-5,t})^{1/2} + \beta_{CM} (C_{t-22,t})^{1/2} + \beta_{JD} J_t^{1/2} + \beta_{JW} (J_{t-5,t})^{1/2} + \beta_{JM} (J_{t-22,t})^{1/2} + \epsilon_{t,t+h}$$

$$log(RV_{t,t+h}) = \beta_0 + \beta_{CD} log(C_t) + \beta_{CW} log(C_{t-5,t}) + \beta_{CM} log(C_{t-22,t}) +$$
$$\beta_{JD} log(J_t + 1) + \beta_{JW} log(J_{t-5,t} + 1) + \beta_{JM} log(C_{t-22,t} + 1) + \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.143	0.222	0.393	0.062	0.103	0.202	-0.063	0.003	0.026
	(0.040)	(0.057)	(0.075)	(0.018)	(0.028)	(0.037)	(0.013)	(0.019)	(0.036)
β_{CD}	0.356	0.224	0.135	0.381	0.262	0.183	0.320	0.224	0.162
	(0.067)	(0.043)	(0.023)	(0.041)	(0.031)	(0.024)	(0.028)	(0.022)	(0.020)
β_{CW}	0.426	0.413	0.204	0.367	0.413	0.272	0.368	0.383	0.274
	(0.120)	(0.114)	(0.070)	(0.063)	(0.072)	(0.061)	(0.043)	(0.053)	(0.049)
β_{CM}	0.111	0.168	0.319	0.163	0.206	0.322	0.246	0.297	0.403
	(0.063)	(0.076)	(0.070)	(0.042)	(0.062)	(0.065)	(0.032)	(0.049)	(0.056)
β_{JD}	-0.153	-0.016	0.005	-0.043	-0.013	0.005	-0.006	-0.027	0.018
	(0.063)	(0.049)	(0.022)	(0.043)	(0.027)	(0.017)	(0.066)	(0.049)	(0.031)
β_{JW}	0.465	0.362	0.456	0.082	0.096	0.132	0.062	0.163	0.198
	(0.233)	(0.205)	(0.287)	(0.071)	(0.075)	(0.113)	(0.105)	(0.126)	(0.176)
β_{JM}	0.355	0.458	0.215	0.133	0.170	0.190	0.207	0.233	0.246
	(0.304)	(0.448)	(0.202)	(0.054)	(0.084)	(0.105)	(0.085)	(0.136)	(0.201)
$R^2_{HAR\text{-}RV\text{-}CJ}$	0.421	0.574	0.478	0.613	0.700	0.639	0.696	0.763	0.722

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-CJ volatility forecast regressions. All of the realized volatility measures are constructed from five-minute returns spanning the period from January 1990 through December 2002, for a total of 3,213 daily observations. The weekly and monthly measures are given by the scaled sum of the corresponding daily measures. The significant daily jump and continuous sample path variability measures are based on equations (19) and (20), respectively, along with the staggered power variation measures in equations (22) and (23), using a critical value of $\alpha = 0.999$. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively.

Table 4C
Daily, Weekly, and Monthly U.S. T-Bond HAR-RV-CJ Regressions

$$RV_{t,t+h} = \beta_0 + \beta_{CD} C_t + \beta_{CW} C_{t-5,t} + \beta_{CM} C_{t-22,t} + \beta_{JD} J_t + \beta_{JW} J_{t-5,t} + \beta_{JM} J_{t-22,t} + \epsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_{CD} C_t^{1/2} + \beta_{CW} (C_{t-5,t})^{1/2} + \beta_{CM} (C_{t-22,t})^{1/2} + \beta_{JD} J_t^{1/2} + \beta_{JW} (J_{t-5,t})^{1/2} + \beta_{JM} (J_{t-22,t})^{1/2} + \epsilon_{t,t+h}$$

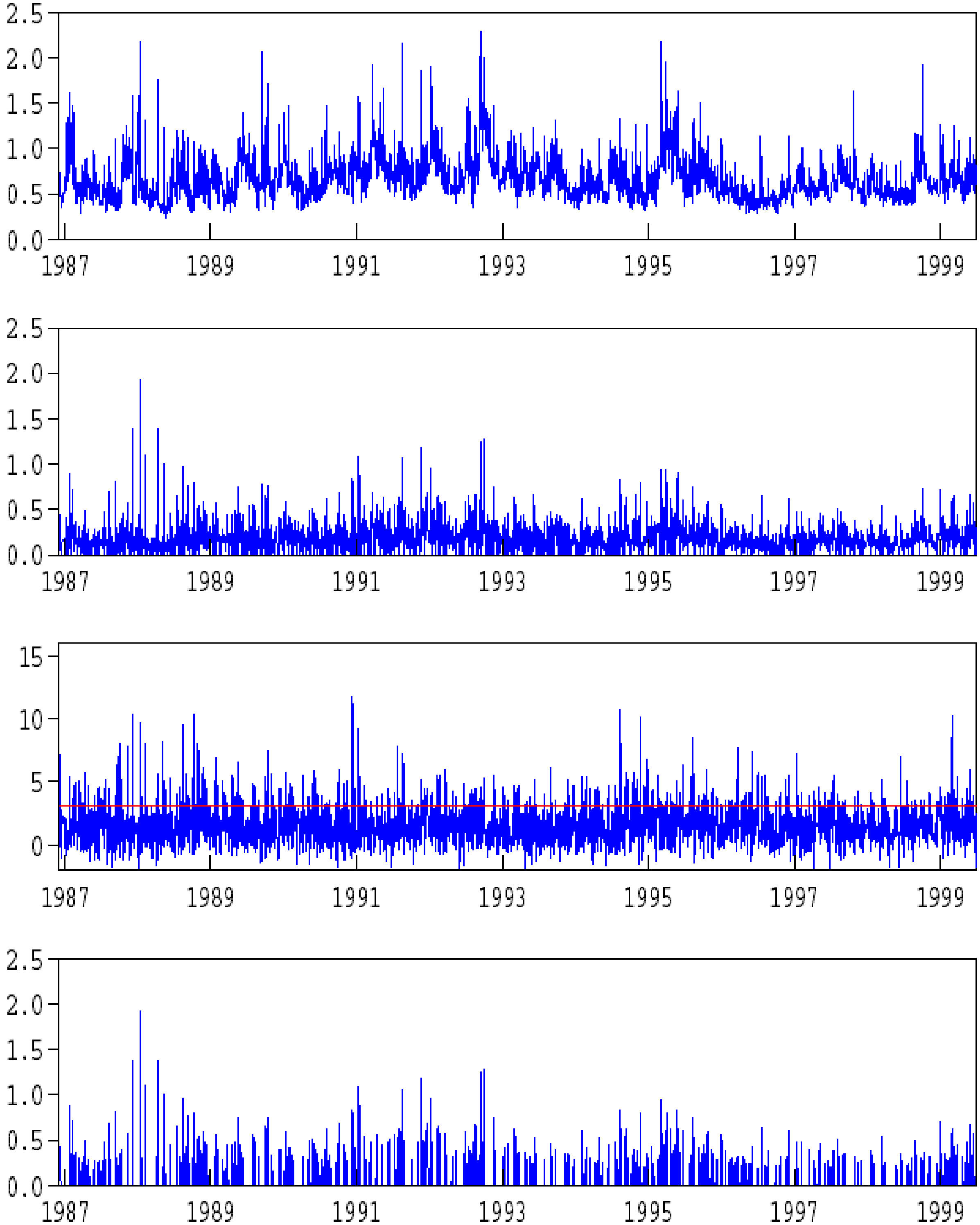
$$log(RV_{t,t+h}) = \beta_0 + \beta_{CD} log(C_t) + \beta_{CW} log(C_{t-5,t}) + \beta_{CM} log(C_{t-22,t}) +$$
$$\beta_{JD} log(J_t + 1) + \beta_{JW} log(J_{t-5,t} + 1) + \beta_{JM} log(C_{t-22,t} + 1) + \epsilon_{t,t+h}$$

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$log(RV_{t,t+h})$		
h	1	5	22	1	5	22	1	5	22
β_0	0.085 (0.011)	0.095 (0.012)	0.133 (0.017)	0.133 (0.016)	0.166 (0.019)	0.236 (0.031)	-0.337 (0.040)	-0.335 (0.052)	-0.473 (0.079)
β_{CD}	0.107 (0.031)	0.064 (0.015)	0.031 (0.006)	0.087 (0.025)	0.069 (0.013)	0.034 (0.006)	0.091 (0.022)	0.068 (0.012)	0.036 (0.007)
β_{CW}	0.299 (0.051)	0.238 (0.047)	0.196 (0.037)	0.306 (0.045)	0.223 (0.042)	0.180 (0.033)	0.297 (0.043)	0.203 (0.042)	0.168 (0.030)
β_{CM}	0.366 (0.062)	0.426 (0.062)	0.369 (0.068)	0.367 (0.048)	0.428 (0.055)	0.380 (0.065)	0.389 (0.046)	0.439 (0.055)	0.382 (0.064)
β_{JD}	-0.136 (0.055)	-0.010 (0.021)	-0.019 (0.008)	-0.080 (0.026)	-0.006 (0.012)	-0.007 (0.006)	-0.769 (0.185)	-0.090 (0.082)	-0.091 (0.041)
β_{JW}	0.230 (0.122)	0.050 (0.081)	-0.075 (0.067)	0.090 (0.033)	0.043 (0.029)	-0.004 (0.025)	0.775 (0.390)	0.227 (0.298)	-0.289 (0.271)
β_{JM}	-0.271 (0.177)	-0.145 (0.216)	-0.116 (0.245)	-0.113 (0.045)	-0.076 (0.058)	-0.057 (0.075)	-1.319 (0.589)	-0.477 (0.773)	-0.034 (0.918)
$R^2_{HAR-RV-CJ}$	0.144	0.325	0.377	0.192	0.353	0.393	0.222	0.365	0.400

Key: The table reports the OLS estimates for daily (h=1) and overlapping weekly (h=5) and monthly (h=22) HAR-RV-CJ volatility forecast regressions. All of the realized volatility measures are constructed from five-minute returns spanning the period from January 1990 through December 2002, for a total of 3,213 daily observations. The weekly and monthly measures are given by the scaled sum of the corresponding daily measures. The significant daily jump and continuous sample path variability measures are based on equations (19) and (20), respectively, along with the staggered power variation measures in equations (22) and (23), using a critical value of $\alpha = 0.999$. The standard errors reported in parentheses are based on a Newey-West/Bartlett correction allowing for serial correlation of up to order 5 (h=1), 10 (h=5) and 44 (h=22), respectively.

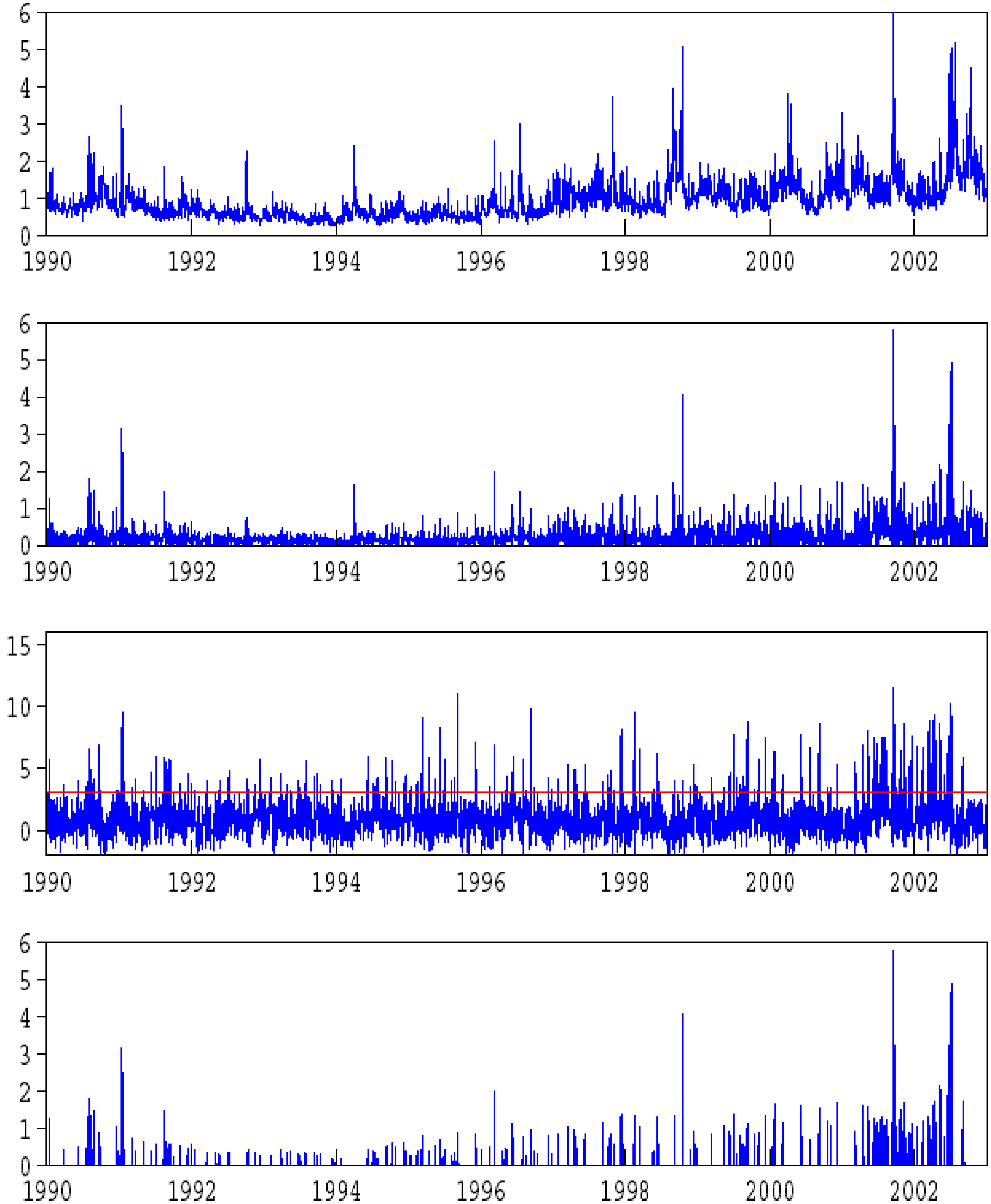
Figure 1A

Daily DM/$ Realized Volatilities and Jumps



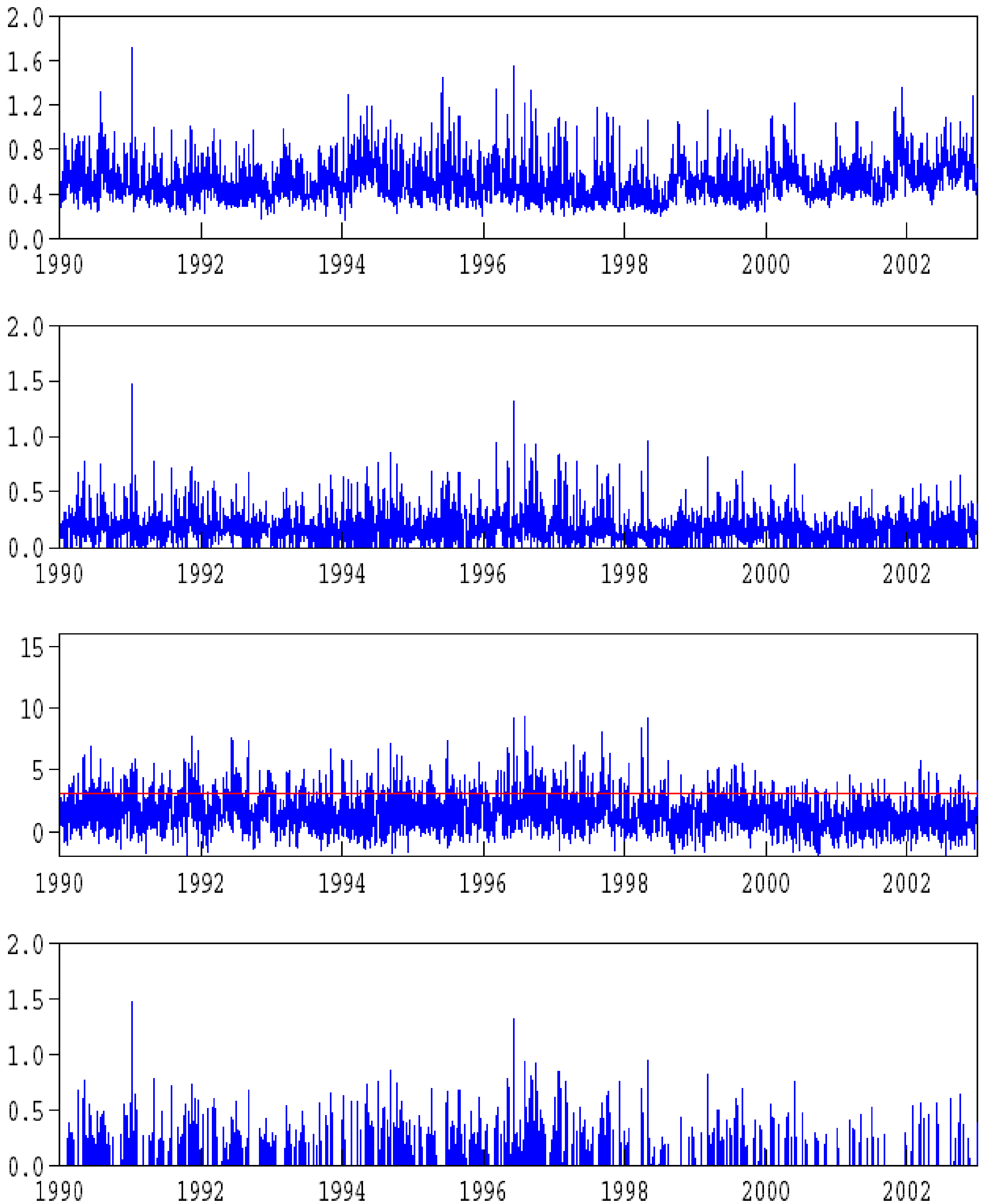
Key: The top panel shows daily realized volatility in standard deviation form, or $RV_t^{1/2}$. The second panel graphs the jump component defined in equation (8), $J_t^{1/2}$. The third panel shows the $Z_{l,t}(\Delta)$ statistic, with the *0.999* significance level indicated by the horizontal line. The bottom panel graphs the significant jumps corresponding to $\alpha = 0.999$, or $J_{t,0.999}^{1/2}$. See the main text for further details.

Figure 1B
Daily S&P500 Realized Volatilities and Jumps



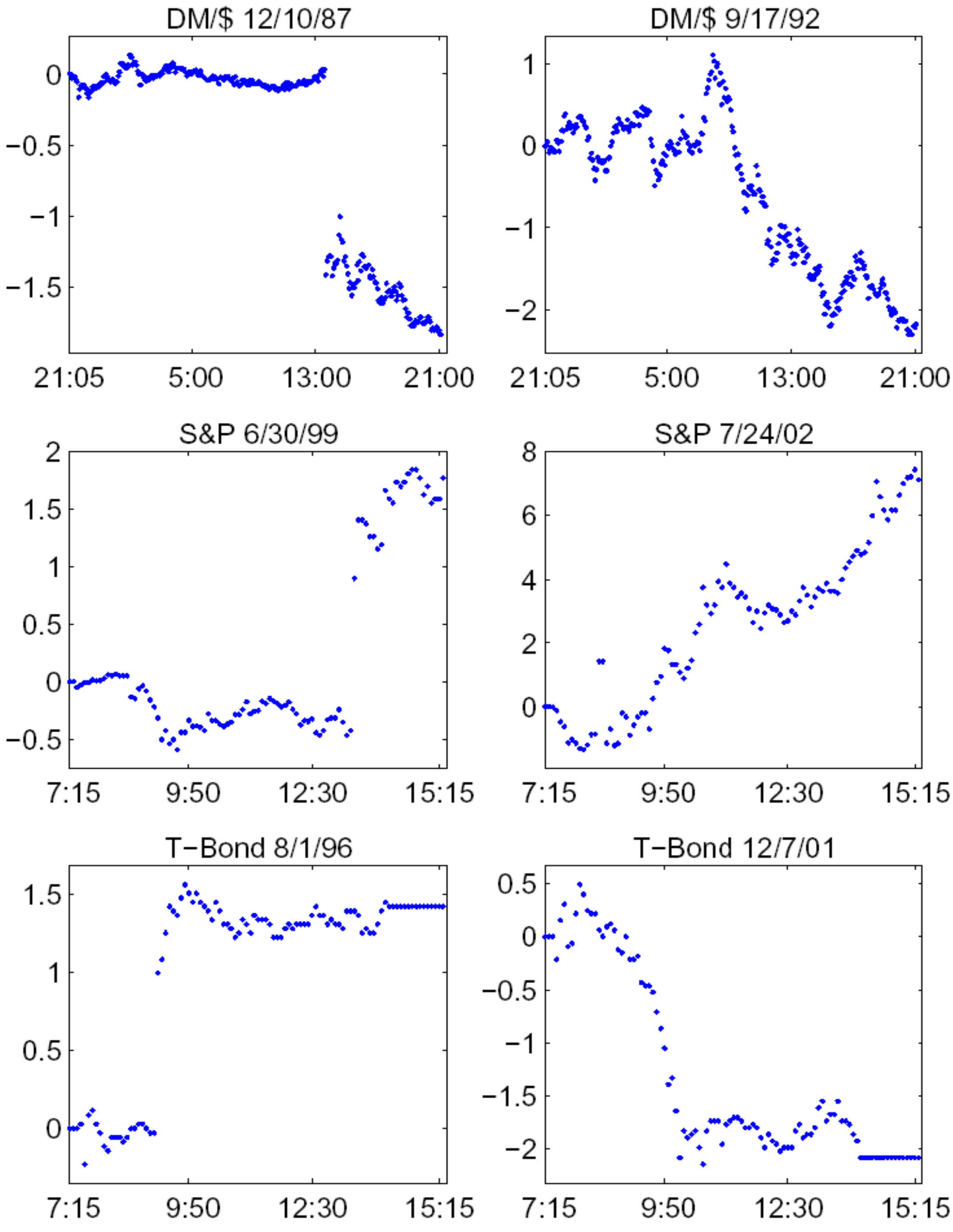
Key: The top panel shows daily realized volatility in standard deviation form, or $RV_t^{1/2}$. The second panel graphs the jump component defined in equation (8), $J_t^{1/2}$. The third panel shows the $Z_{1,t}(\Delta)$ statistic, with the *0.999* significance level indicated by the horizontal line. The bottom panel graphs the significant jumps corresponding to α =*0.999*, or $J_{t,0.999}^{1/2}$. See the main text for further details.

Figure 1C
Daily U.S. T-Bond Realized Volatilities and Jumps



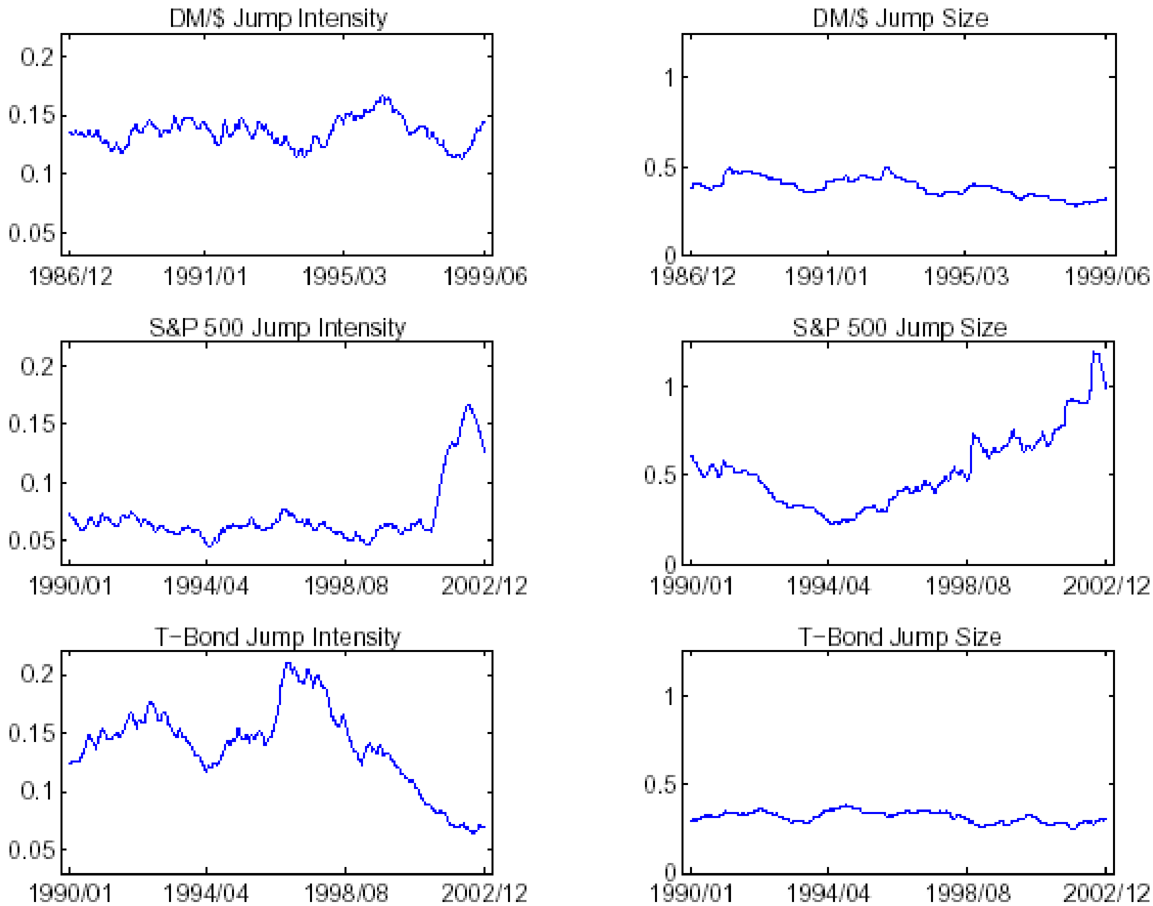
Key: The top panel shows daily realized volatility in standard deviation form, or $RV_t^{1/2}$. The second panel graphs the jump component defined in equation (8), $J_t^{1/2}$. The third panel shows the $Z_{1,t}(\Delta)$ statistic, with the *0.999* significance level indicated by the horizontal line. The bottom panel graphs the significant jumps corresponding to $\alpha = 0.999$, or $J_{t,0.999}^{1/2}$. See the main text for further details.

Figure 2
Intraday Price Movements



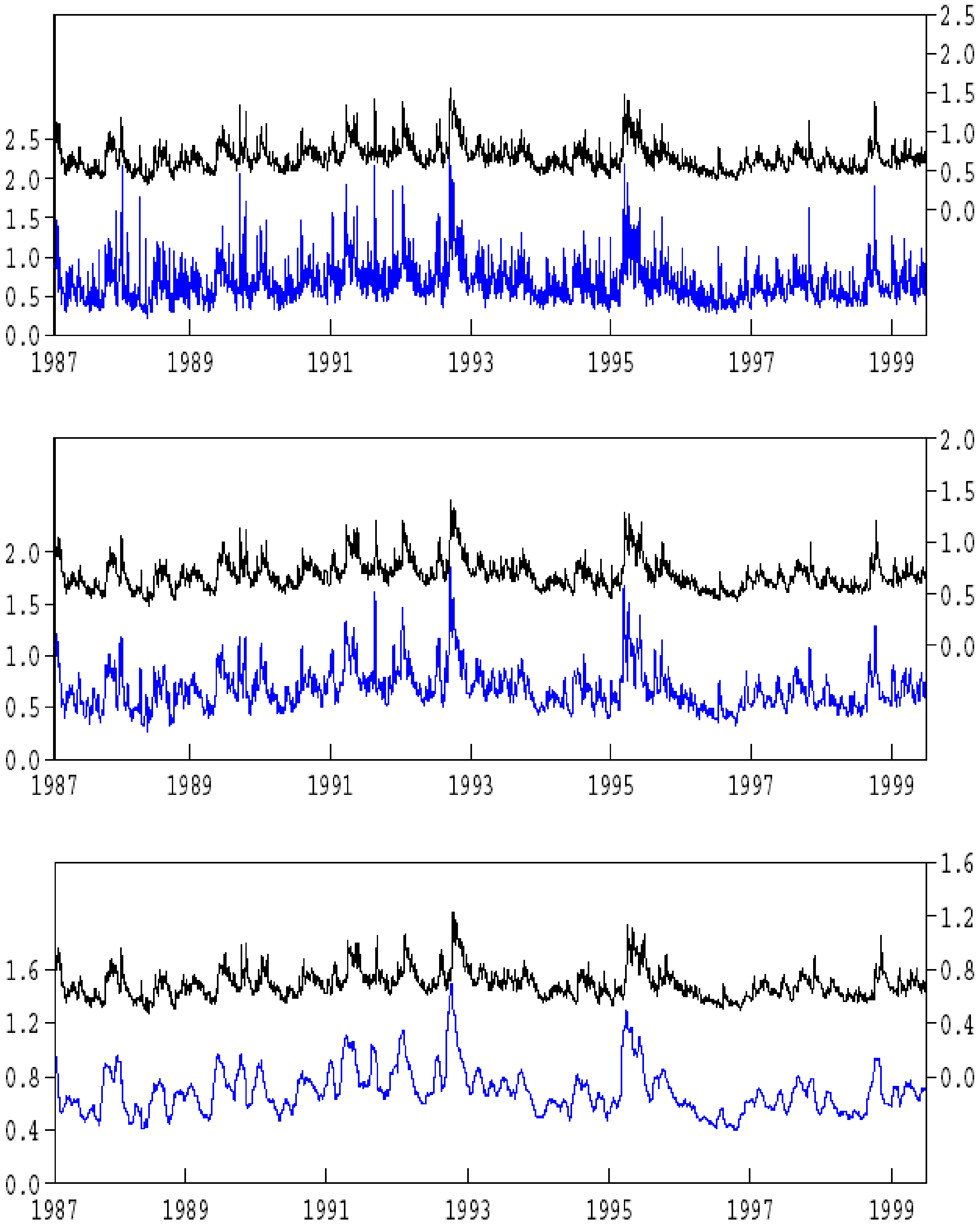
Key: The figure graphs the five-minute intraday price increments for days with large jump statistics $Z_{1,t}(\Delta)$ (left-side panels), and days with large daily price moves but numerically small jump statistics (right-side panels).

Figure 3
Smoothed Jump Intensities and Jump Sizes



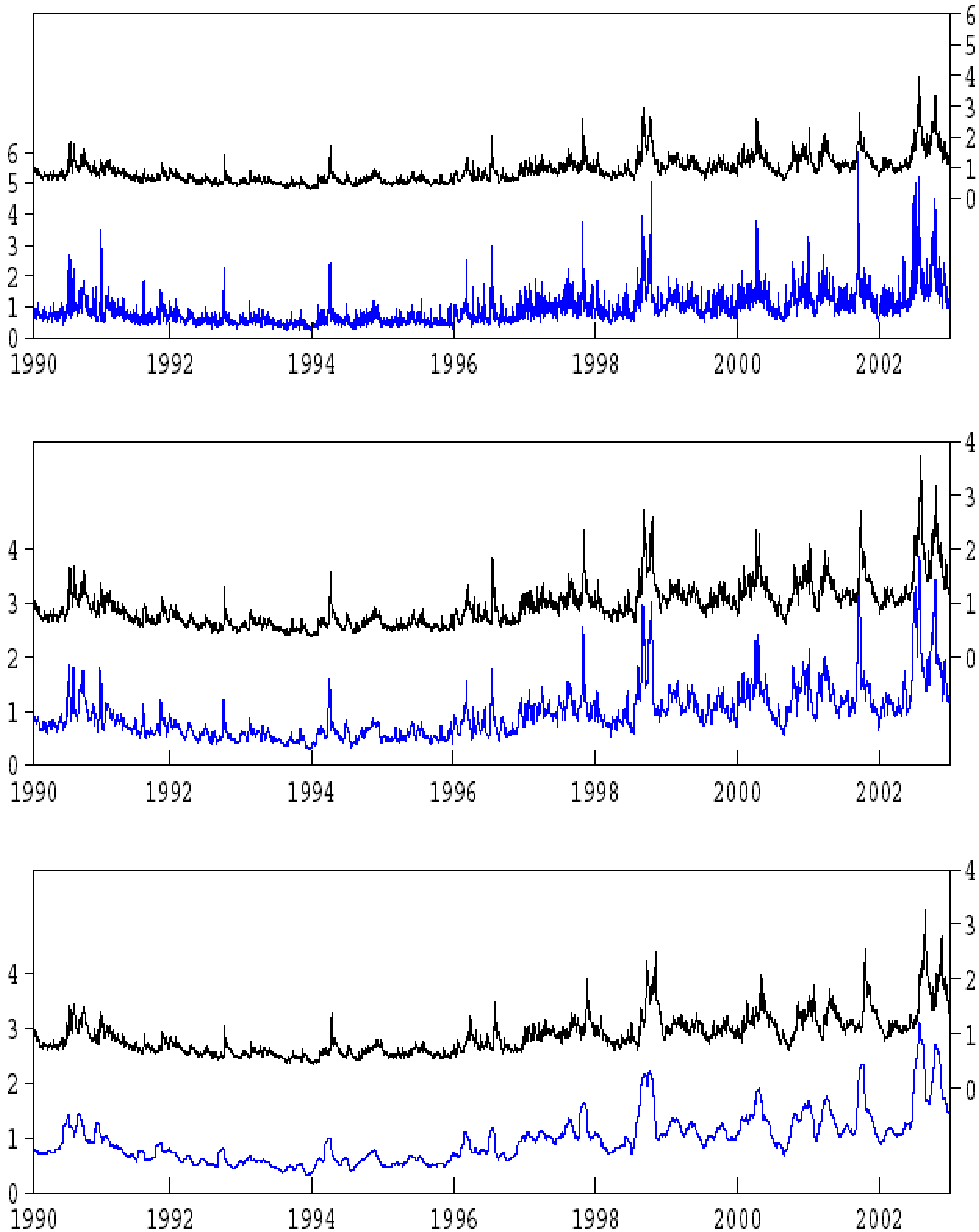
Key: The figure graphs the exponentially smoothed (with a smoothing parameter of *0.94*) average monthly jump intensities and sizes for the significant jumps based on $\alpha = 0.999$. The jump sizes are expressed in standard deviation form, or $J_{t,0.999}^{1/2}$.

Figure 4A
Daily, Weekly and Monthly DM/$ Realized Volatilities and HAR-RV-CJ Forecasts



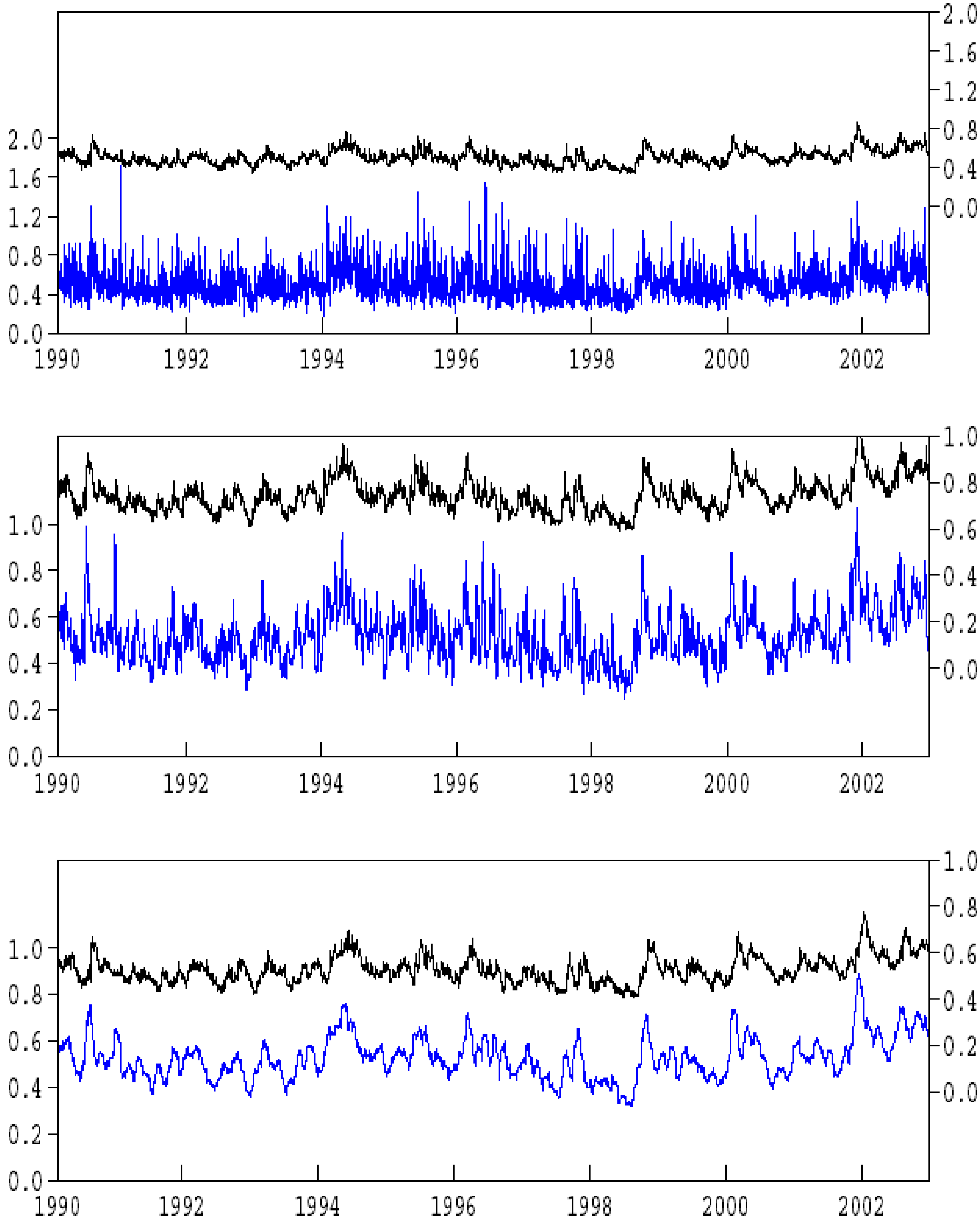
Key: The top, middle and bottom panels show daily (h=1), weekly (h=5) and monthly (h=22) realized volatilities, $RV_{t,t+h}^{1/2}$ (left scale), and the corresponding forecasts from the HAR-RV-CJ model in standard deviation form in equation (27) (right scale). See the main text for further details.

Figure 4B
Daily, Weekly and Monthly S&P500 Realized Volatilities and HAR-RV-CJ Forecasts



Key: The top, middle and bottom panels show daily (h=1), weekly (h=5) and monthly (h=22) realized volatilities, $RV_{t,t+h}^{1/2}$ (left scale), and the corresponding forecasts from the HAR-RV-CJ model in standard deviation form in equation (27) (right scale). See the main text for further details.

Figure 4C
Daily, Weekly and Monthly U.S. T-Bond Realized Volatilities and HAR-RV-CJ Forecasts



Key: The top, middle and bottom panels show daily (h=1), weekly (h=5) and monthly (h=22) realized volatilities, $RV_{t,t+h}^{1/2}$ (left scale), and the corresponding forecasts from the HAR-RV-CJ model in standard deviation form in equation (27) (right scale). See the main text for further details.