Overnight News and Daily Equity Trading Risk Limits

Katja Ahoniemi[a] Ana-Maria Fuertes[b,*]

Jose Olmo[c]

[a] *Imperial College Business School, London, U.K.*
[b] *Faculty of Finance, Cass Business School, City University London, U.K.*
[c] *University of Southampton, U.K.*

Abstract

This paper proposes a new bivariate modeling approach for setting daily equity-trading risk limits using high-frequency data. We construct one-day-ahead Value-at-Risk (VaR) forecasts by taking into account the different dynamics of the overnight and daytime return processes and their covariance. The covariance is motivated by market microstructure effects such as price staleness and news spillover. Among the competitors we include a simpler bivariate model where the overnight return is redefined by moving the open price further into the trading day, and a univariate model based on the close-to-close return and an overnight-adjusted realized volatility. We illustrate the different approaches using data on the S&P 500 and Russell 2000 indices. The evidence in favour of modeling the covariance is more convincing for the latter index due to the lower trading volumes and, relatedly, the less efficient price discovery at market open for small-cap stocks.

Keywords: Overnight; Price discovery; Realized volatility; Risk management; Value-at-Risk.

JEL Classification: C52; C53; G15.

[*] Corresponding address: Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ, UK. E-mail: a.fuertes@city.ac.uk. The first and third authors acknowledge the OP-Pohjola Group Research Foundation and the Spanish Government through project MICINN ECO2011-22650, respectively, for financial support. This paper was partly written while the first author was a visiting researcher at Cass Business School. We are grateful to Torben Andersen (the Editor) and two anonymous referees for their constructive comments. We would like to thank Charlie Cai, Marwan Izzeldin, Petri Jylha, Simone Manganelli, Anthony Neuberger, Matthew Osborne, Andrew Patton, participants at the CEFAGE-UE Conference on Robust Econometric Methods for Modeling Economic and Financial Variables (Lisboa, September 2012), EEA-ESEM Conference (Malaga, August 2012), 6th BMRC-QASS Conference on Macro and Financial Economics (Brunel University, May 2012), the 2nd FEBS Conference on Recent Developments in Financial Markets and Banking (London, June 2012), and seminar participants at Cass Business School, Helsinki Centre for Economic Research, Leeds University Business School and Durham Business School for useful discussions.

1 INTRODUCTION

The world's major stock exchanges are open for a limited number of hours each trading day. Hence, although investors receive news on a continuous basis they are able to trade immediately for only a part of the day in those exchanges. The overnight 'surprise' or close-to-open return reflects local market news accumulated during non-trading hours, such as earnings announcements, and foreign news which are immediately reflected in prices in stock exchanges from other time zones. The information flow is greater during trading (daytime) hours than during non-trading (overnight) hours.[1] However, both the increasing globalization of securities markets and the proliferation of electronic trading systems are likely to emphasize the importance of the overnight information flow, as events from around the globe can trigger investor reactions in all markets.[2] The events that occur during overnight non-trading hours will typically be impounded into prices very rapidly as an exchange opens for a new trading day.

There is ample evidence that the return process exhibits different dynamics during non-trading and trading hours (French and Roll, 1986; Lockwood and Lin, 1990; Hasbrouck, 1991; Masulis and Ng, 1995; George and Hwang, 2001). The contrast has been acknowledged already in various theoretical models of security returns (Oldfield and Rogalski, 1980; Slezak, 1994; Hong and Wang, 2000). Assuming an absence of (or thin) trading during the overnight period, a relevant question is how best to exploit the observed close-to-open price variation for daily equity tail-risk forecasting.

Our main goal is to compare various methods for embedding overnight information into forecasts of equity portfolio tail-risk behavior. Value-at-Risk (VaR) has become a standard risk management tool for setting day-to-day loss limits of trading desks, and it is widely employed by commercial banks. VaR is the αth quantile of a portfolio's value change over a given day, expressed in probability terms as the value $VaR_{t,\alpha}$ for which $P(r_t \leq VaR_{t,\alpha}|\Im_{t-1}) = \alpha$, with r_t the return on day t, and $\Im_{t-1}$ the conditioning information set at the time the forecast is made. Thus for nominal coverage α equal to 0.01 the VaR is, effectively, the lower end of a 99% confidence interval. Appropriate day-ahead VaR forecasts can aid risk managers in ensuring that trading desks stay within predefined risk limits.

Many studies in the high-frequency equity volatility forecasting literature "bundle" the squared overnight return and the daytime realized volatility into an overnight-adjusted realized volatility measure that is unconditionally unbiased for the 24-hour daily price variation. This adjustment has been usually done either by upward-scaling the realized volatility, by summing together the squared overnight return and the realized volatility, or by forming a weighted average of the squared overnight return and the realized volatility. For instance, Hansen and Lunde (2005) suggest to optimally weight, in a minimum mean-squared-error sense, the squared overnight return and the daytime realized vari-

[1] We use the terminology daytime, open-to-close, and trading period interchangeably to refer to the regular trading hours in the U.S. market from 9:30am to 4:00pm EST. Likewise, we use the terminology overnight, close-to-open, and non-trading period interchangeably to refer to the time interval from 4:00pm to 9:30am EST.

[2] In order to predict the opening price of an index in a home market, de Gooijer et al. (2012) exploit information conveyed by high-frequency stock price patterns in foreign markets during non-trading hours in the home market.

ance. We refine this approach by modeling the overnight and daytime processes separately.

This paper contributes to the parametric VaR modeling literature by proposing a novel bivariate approach that not only models the dynamics of the overnight and daytime returns separately, but also accommodates their covariance. Setting trading limits using VaR forecasts has become a routine task for risk managers in banks and other financial institutions. Our approach is motivated by continuous-time price theory where each 24-hour period is explicitly decomposed into an *overnight* and a *daytime* segment. Bivariate VaR modeling has two theoretically appealing features. One is that it acknowledges different return generating processes during trading and non-trading hours (and hence, different degrees of predictability for the overnight and daytime variation), and another is that it accommodates the potentially non-zero covariance between overnight and daytime returns. During the overnight segment, equity markets are closed for trading, although news do not cease. Market microstructure effects such as *price staleness* and *news spillover* can induce a non-zero ex-ante covariance between overnight and daytime returns which may contain useful information for establishing 1-day-ahead equity VaR limits. After the market officially opens, it takes some time before all stocks begin trading. Thus, the stock index quotes available in the first k minutes of each trading day can contain numerous stale prices, namely, individual transaction prices from the previous day.[3] The fact that the first index quotes are stale means that their values do not yet fully reflect all overnight news. Put another way, the impounding of overnight information into the prices of all portfolio constituents can spill into the trading day, inducing a non-zero ex-ante covariance between overnight and daytime returns.

We formulate and test two main hypotheses. Hypothesis I states that the overnight-daytime covariance component of a bivariate VaR model has merit for setting 1-day-ahead equity trading limits. Hypothesis II states that modeling the overnight and daytime segments of the day separately produces VaR forecasts superior to the widespread univariate modeling of an overnight-adjusted realized volatility measure. To the best of our knowledge, Hypothesis I is novel in the literature, and the original aspect of Hypothesis II is its formulation in the context of VaR for setting day-ahead equity trading limits.[4] An important aspect of the paper is that the merit of the bivariate modeling approach is assessed through the lens of equity portfolio risk management rather than statistical predictability. Thus the paper not only contributes to academic research on tail risk prediction but also speaks to the managers of VaR-based trading books at banks and other financial institutions.

The hypotheses are examined through the lens of two distinct VaR backtesting methods. One is the unconditional Equal Predictive Ability test proposed by Giacomini and White (2006) which assesses the significance of differences in out-of-sample forecasting performance. This test has several appealing aspects: it naturally controls for parameter estimation uncertainty, it can be used to confront nested

[3] Stoll and Whaley (1990) document empirically that in the 1980s it took around 5 to 6 minutes for large stocks to open for trading in the NYSE. With the advent of electronic trading, the average time for stocks to commence trading in today's markets is much shorter, but large cap stocks will typically begin trading faster than small cap stocks.

[4] Studies that model both the overnight and daytime processes include Taylor (2007) and Andersen et al. (2011).

or non-nested models, and it can be deployed for any loss function. In our context, the test uses the 'tick' loss function implicit in quantile regression theory, and provides a formal statistical answer to the question of whether the expected forecast error losses associated with the proposed bivariate VaR model are smaller than those from simpler VaR models. The second backtesting method is the Correct Conditional Coverage test of Engle and Manganelli (2004), which we deploy in its original version and as a probit-based test to address the distinct question of whether a VaR model is correctly specified out-of-sample. The null hypothesis is that the actual losses in excess of the predicted VaR are independently distributed and occur at a frequency equal to the nominal coverage rate.

The analysis is based on high-frequency data on the large-cap S&P 500 and the small-cap Russell 2000, over the 14-year period beginning in November 1997 and ending in September 2011. All the modeling approaches under comparison condition the out-of-sample 1-day-ahead VaR forecasts on histories of five-minute returns during the daytime period and the previous-close-to-open return to capture the overnight non-trading period. Seeking to acknowledge time-variation in the degree of downside risk predictability, the comparative evaluation of VaR forecast accuracy is conducted dynamically by deploying the Equal Predictive Ability test and Correct Conditional Coverage tests sequentially over rolling and non-overlapping windows of out-of-sample forecasts. An important landmark of the recent financial crisis, the collapse of Lehman Brothers in September 2008, is chosen as the start of the VaR evaluation period in order to make the forecasting task more challenging.

Our paper produces two main empirical findings which add to current knowledge on tail risk modeling and have practical relevance for market practitioners in search of a good short-term equity VaR forecasting approach. First, the evidence does not refute Hypothesis I, which suggests that modeling the covariance between overnight and daytime returns can prove beneficial for setting accurate daily equity trading risk limits. This finding stems from formal comparisons of the proposed bivariate VaR model and simpler bivariate models that do not account for the overnight-daytime covariance. One such model arises from moving the opening price k minutes further into the trading day (several k values are considered) to avoid the strong impact of market microstructure inefficiencies such as price staleness and news spillover at the beginning of the trading day. This modeling approach is empirically motivated by various studies which acknowledge a slight delay in equity price discovery at the market open (Chan et al., 1991; Stoll and Whaley, 1990; Hecq et al., 2012; Ahoniemi and Lanne, 2013).

Second, the evidence is also in favor of Hypothesis II, which conveys that the separate modeling of overnight and daytime returns has economic value for setting 1-day-ahead equity trading limits relative to the widespread univariate modeling approach. Indeed, it is shown that even simpler bivariate models that neglect the covariance are able to produce superior VaR forecasts than univariate models fitted to an overnight-adjusted realized volatility series. In this sense, our findings add a risk management perspective to extant evidence on the merit of modeling separately the overnight and daytime returns.

The remainder of the paper is organized as follows. Section 2 introduces the proposed bivariate model alongside the competing models to obtain parametric location-scale VaR forecasts. Section 3 outlines the forecast evaluation tests. Section 4 discusses the empirical results. Section 5 concludes.

2 HIGH-FREQUENCY PARAMETRIC VALUE-AT-RISK

Risk managers face the task of setting risk limits for separate business lines or trading desks (e.g., equity, FX, fixed income) which amounts to predicting a specific quantile of the close-to-close return distribution. The 1% VaR quantile of a portfolio is the maximum daily loss that is expected to be exceeded only once every 100 days. VaR has a long history dating back to the 1990s when banks started using it as a one-dimensional snapshot of downside risk and real-time risk monitoring tool. The widespread use of VaR was formally acknowledged by the Basel Accord in 1996 when it introduced a VaR-based capital requirement framework for positions held for trading intent (BCBS, 1996).

VaR measures can be non-parametric (e.g., historical simulation VaR), semi-parametric (e.g., CAViaR) or parametric (e.g., GARCH-based VaR) depending on the assumptions the risk manager is prepared to make. Parametric VaR measures from location-scale models remain widely used; see, e.g. Giot and Laurent (2004), Clements et al. (2008), Brownlees and Gallo (2010), and Fuertes and Olmo (2012). In the parametric modeling strategy, under the assumption that the daily volatility process is independent of the return innovation process ε_t, the quantile process for the 24-hour return r_t is obtained as a simple location-scale transformation of the quantile process of ε_t.

Next we discuss several modeling approaches within the parametric location-scale framework to obtain 1-day-ahead VaR forecasts. The novel bivariate approach that we propose models separately the daytime and overnight returns, and their ex-ante covariance. As reasonable competitors, we consider two simpler bivariate modeling approaches that neglect the covariance, and a univariate modeling approach that relies on the close-to-close return and an overnight-adjusted realized volatility measure.

2.1 Bivariate Modeling Approach

As in the continuous-time finance literature, the diffusion of the log price process is assumed to belong to the class of semimartingales and formalized by the stochastic differential equation

$$dp_t = \mu_t dt + \sigma_t dW_t + \kappa_t dN_t, \quad 0 \leq t \leq T \tag{1}$$

where μ_t denotes the instantaneous deterministic drift term, σ_t the instantaneous or spot volatility process which is stationary and independent of the random Brownian motion W_t, and dN_t a counting process for the jumps of size κ_t such that $dN_t = 1$ if a jump occurs at time t and $dN_t = 0$ otherwise.

To establish notation, let O_t and C_t represent the official market opening and closing times of day t, respectively. Thus, the time period from C_{t-1} to C_t represents the entire 24-hour day t, comprising an *overnight* period during which the market is closed from C_{t-1} to O_t, and an official trading period from O_t to C_t referred to as *daytime*. Likewise, the 24-hour period from C_t to C_{t+1} is called day $t+1$ and so forth. The daily return is $r_t \equiv p_{C_t} - p_{C_{t-1}}$ where prices are measured in logarithms. From the above equation, it follows that the *quadratic variation* of the daily return process is given by

$$QV_t \equiv V(r_t) = IV_t + J_t = \int_{C_{t-1}}^{C_t} \sigma_s^2 ds + \sum_{C_{t-1}<j\leq C_t}^{n_t} \kappa_{t,j}^2 \tag{2}$$

where the integrated variance (IV_t) and the jump component (J_t) represent the contribution of the continuous-time and discrete processes, respectively; n_t is the random number of jumps on day t.

This model can be extended to accommodate the different dynamics of the overnight and daytime components of the continuous-time log price process. The original aspect of the proposed modeling approach for VaR is to accommodate co-variation between the log-price differentials corresponding to the overnight and daytime periods. This is done by assuming the following model for dp_t for $0 \leq t \leq T$,

$$dp_t = \begin{cases} \mu_{o,t}dt + \sigma_{o,t}dW_{o,t} + \kappa_{o,t}dN_{o,t}, & \text{if } C_{t-1} < t \leq O_t, \\ \mu_{d,t}dt + \sigma_{d,t}dW_{d,t} + \kappa_{d,t}dN_{d,t} + \rho_t dW_{o,t-\bar{s}}, & \text{if } O_t < t \leq C_t, \end{cases} \tag{3}$$

where the subscripts o and d denote overnight and daytime, respectively; $t-\bar{s}$ for $t = O_t$ represents the first point in the overnight period from when any news arriving will spill over into the daytime period; and the term $\rho_t dW_{o,t-\bar{s}}$, with ρ_t an instantaneous covariance term, represents the spillover effect from the overnight news onto the daytime prices. This effect vanishes after $t = O_t + \bar{s}$ by construction, and it could vanish earlier if ρ_t is equal to zero within the interval $(O_t, O_t + \bar{s}]$. Hence, the random process $\rho_t dW_{o,t-\bar{s}}$ captures dependence between the overnight and daytime returns. It is important to note that the processes $dW_{o,t}$ in the overnight period and $dW_{o,t-\bar{s}}$ in the daytime period reflect the same random events implying that $E[dW_{o,s}dW_{o,t-\bar{s}}] = ds$, with $C_{t-1} < s \leq O_t$ and $O_t < t \leq C_t$.

It follows from equation (3) that the daily return can be expressed as $r_t = r_{o,t} + r_{d,t}$ where $r_{o,t} \equiv p_{O_t} - p_{C_{t-1}}$ and $r_{d,t} \equiv p_{C_t} - p_{O_t}$ are, respectively, the *overnight* and *daytime* return processes

$$r_{o,t} = \int_{C_{t-1}}^{O_t} \mu_{o,s}ds + \int_{C_{t-1}}^{O_t} \sigma_{o,s}dW_{o,s} + \int_{C_{t-1}}^{O_t} \kappa_{o,s}dN_{o,s}, \tag{4a}$$

and

$$r_{d,t} = \int_{O_t}^{C_t} \mu_{d,s^*}ds^* + \int_{O_t}^{C_t} \sigma_{d,s^*}dW_{d,s^*} + \int_{O_t}^{C_t} \kappa_{d,s^*}dN_{d,s^*} + \int_{O_t}^{O_t+\bar{s}} \rho_{s^*}dW_{o,s^*-\bar{s}}. \tag{4b}$$

For notational simplicity, we define $\mu_{o,t} \equiv \int_{C_{t-1}}^{O_t} \mu_{o,s}ds$ as the integrated deterministic drift driving the overnight return, and $\varepsilon_{o,t} \equiv \int_{C_{t-1}}^{O_t} \sigma_{o,s}dW_{o,s} + \int_{C_{t-1}}^{O_t} \kappa_{o,s}dN_{o,s}$ as its random component; similarly, for the *daytime* return, $\mu_{d,t} \equiv \int_{O_t}^{C_t} \mu_{d,s*}ds^*$ and $\varepsilon_{d,t} \equiv \int_{O_t}^{C_t} \sigma_{d,s^*}dW_{d,s^*} + \int_{O_t}^{C_t} \kappa_{d,s^*}dN_{d,s^*} + \int_{O_t}^{O_t+\bar{s}} \rho_{s^*}dW_{o,s^*-\bar{s}}$.

The latent quadratic variation of r_t can thus be expressed as

$$QV_t = QV_{o,t} + QV_{d,t} + 2Cov(r_{o,t}, r_{d,t}) \tag{5}$$

where $QV_{o,t} = IV_{o,t} + J_{o,t}$ with $IV_{o,t} \equiv \int_{C_{t-1}}^{O_t} \sigma_{o,s}^2 ds$ and $J_{o,t} \equiv \sum_{C_{t-1}<j\leq O_t}^{n_{o,t}} \kappa_{o,j}^2$; $n_{o,t}$ denotes the number of jumps occurring overnight. Similarly, $QV_{d,t} = IV_{d,t} + J_{d,t}$ with $IV_{d,t} \equiv \int_{O_t}^{C_t} \sigma_{d,s^*}^2 ds^* + \int_{O_t}^{O_t+\bar{s}} \rho_{s^*}^2 ds^*$ and $J_{d,t} \equiv \sum_{O_t<j\leq C_t}^{n_{d,t}} \kappa_{d,j}^2$; $n_{d,t}$ denotes the number of daytime jumps. It follows that

$$Cov(r_{o,t}, r_{d,t}) \equiv \int_{O_t-\bar{s}}^{O_t} \rho_{s+\bar{s}} ds.$$

This decomposition of the daily variance is obtained by applying the properties of Brownian motions, $E[dW_{o,s}dW_{d,s^*}] = 0$ and $E[dW_{o,s}dW_{o,s^*}] = 0$ with $s < s^*$, and the property of independence between Brownian motions and jump processes, $E[dW_{o,s}dN_{i,s^*}] = 0$ and $E[dW_{d,s}dN_{i,s^*}] = 0$ with $i = o, d$.

2.1.1 Bivariate Daily Return Level and Volatility Forecasts

In order to formalize the bivariate modeling approach for VaR prediction, let the conditional mean and variance of the daily return process, r_t, be written as

$$\mu_{t|t-1} = \mu_{o,t|t-1} + \mu_{d,t|t-1}, \tag{6}$$

and

$$QV_{t|t-1} = QV_{o,t|t-1} + QV_{d,t|t-1} + 2Cov(r_{o,t}, r_{d,t}| \Im_{t-1}), \tag{7}$$

respectively, where $\Im_{t-1}$ is the sigma-algebra generated by the information available to the forecaster at time $t-1$; the individual components are defined as $\mu_{o,t|t-1} \equiv E[r_{o,t}| \Im_{t-1}]$, $\mu_{d,t|t-1} \equiv E[r_{d,t}| \Im_{t-1}]$, $QV_{o,t|t-1} \equiv V(r_{o,t}| \Im_{t-1})$ and $QV_{d,t|t-1} \equiv V(r_{d,t}| \Im_{t-1})$. Equation (7) is the forecasting counterpart of the daily quadratic variation decomposition formalized in equation (5).

We model both overnight and daytime returns as location-scale processes. Forecasts of the *overnight* return and its variance are obtained, respectively, from the following AR and GARCH equations

$$r_{o,t} = a_0 + a_1 r_{d,t-1} + b_1 r_{o,t-1} + \varepsilon_{o,t} \tag{8a}$$

$$h_{o,t} = \alpha_0 + \alpha_1 \varepsilon_{o,t-1}^2 + \beta_1 h_{o,t-1} + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 I_{t-1}^- \varepsilon_{t-1}^2 \tag{8b}$$

where $\varepsilon_{o,t} \equiv \sqrt{h_{o,t}} \cdot \epsilon_{o,t}$, and $\epsilon_{o,t}$ are i.i.d. standardized skewed Student-$t(0, 1, \xi_o, \eta_o)$ innovations. The parameters ξ_0 and η_o capture, respectively, asymmetry and fat-tailedness in the overnight return distribution. Equation (8a) is an AR(1) parameterization for $r_{o,t}$ augmented with the lagged daytime return. Equation (8b) is a threshold GARCH specification where ε_t are innovations in a standard

AR(1) model for the daily return r_t, and $I_t^- = 1$ if $\varepsilon_t < 0$ and $I_t^- = 0$ otherwise; $\gamma_2 > 0$ implies that bad news increase future volatility more than good news (leverage effect). The parameters of this overnight price variation model can be obtained by Quasi Maximum Likelihood (QML).

Forecasts of the *daytime* return and variance are obtained from the AR-ARIMA model

$$r_{d,t} = \widetilde{a}_0 + \widetilde{a}_1 r_{d,t-1} + \widetilde{b}_1 r_{o,t-1} + \sqrt{{\sigma_d}^2 \cdot RV_t} \cdot \epsilon_{d,t} \tag{9a}$$

$$(1-L)^d (\ln RV_t - \tau_0 - \tau_1 r_{t-1} - \tau_2 I_{t-1}^- r_{t-1}) = (1+\theta L) u_t \tag{9b}$$

where $\epsilon_{d,t}$ is an i.i.d. standardized skewed Student-$t(0, 1, \xi_d, \eta_d)$ innovation, and ${\sigma_d}^2$ is a scaling parameter. This location-scale formulation treats the daytime volatility as observable ex-post by constructing (from high frequency open-to-close data) the daytime *realized variance* measure

$$RV_t = \sum_{j=1}^{M} r_{j,t}^2, \tag{10}$$

where M is the number of equal-length intraday intervals, and $r_{t,j} \equiv p_{t,j} - p_{t,j-1}$ is the jth intraday return on day t; a large literature has established that realized volatilities exhibit long-range dependence and log-normality. Equation (9a) is an AR(1) parameterization for $r_{d,t}$ augmented with lagged overnight returns. Equation (9b) is an ARFIMA$(0, d, 1)$ parameterization for the realized variance where u_t are i.i.d.$(0, \sigma_u^2)$ innovations, and the leverage indicator is defined as $I_t^- = 1$ if $r_t < 0$ and $I_t^- = 0$ otherwise; $\tau_2 < 0$ implies that bad news increase the future daytime volatility more than good news. The parameters can be consistently estimated following Giot and Laurent's (2004) two-step approach. First, the ARFIMA equation is estimated by Maximum Likelihood (ML) under a normality assumption for u_t. Second, the parameter vector $(\widetilde{a}_o, \widetilde{a}_1, \widetilde{b}_1, {\sigma_d}^2, \xi_d, \eta_d)$ of the AR equation is estimated by QML assuming that all the dynamics in the conditional variance is of ARFIMA type.[5]

Finally, we adopt the Dynamic Conditional Correlation modeling framework introduced by Engle (2002; DCC) to accommodate time-varying dependence between $r_{o,t}$ and $r_{d,t}$ which is formalized as[6]

$$q_{o,d,t} = \overline{\rho}_{o,d}(1-\alpha-\beta) + \alpha \epsilon_{o,t-1} \epsilon_{d,t-1} + \beta q_{o,d,t-1}, \tag{11}$$

where $\epsilon_{o,t}$ and $\epsilon_{d,t}$ denote overnight and daytime standardized innovations, respectively, associated to the AR-GARCH and AR-ARFIMA models outlined above; $\overline{\rho}_{o,d} \equiv E[\epsilon_{o,t} \cdot \epsilon_{d,t}]$ denotes unconditional

[5]Explicit long-memory modeling of log realized volatilities via ARFIMA specifications is a well-established approach dating back at least to Andersen et al. (2001, 2003). Alternative easier-to-handle approximations to long memory have been put forward in the recent literature. Examples include the heterogeneous autoregressive (HAR) model of Corsi (2009) and the multiplicative error model (MEM) of Brownlees and Gallo (2010).

[6]Engle (2002) proposes a DCC framework based on individual GARCH processes, but notes that nothing would change if this was generalized. Given the data available over each of the two daily segments, the overnight variance is latent and modeled as GARCH, and the daytime ex-post observed (realized) variance is modeled as ARFIMA.

correlation. The conditional covariance of interest is $Cov(r_{o,t}, r_{d,t}|\, \Im_{t-1}) \equiv q_{o,d,t}\sqrt{h_{o,t} \times RV_{t|t-1}}$ where $h_{o,t}$ and $RV_{t|t-1}$ are the conditional variances derived from equations (8b) and (9b), respectively.

2.1.2 Bivariate VaR Forecasts

Let $F_t(\cdot)$ be the conditional distribution of the daily return process r_t given $\Im_{t-1}$. The goal is to forecast the α-quantile of $F_t(\cdot)$ defined as $VaR_{t,\alpha} \equiv F_t^{-1}(\alpha)$ with $P(r_t \leq VaR_{t,\alpha}|\Im_{t-1}) = \alpha$.[7] The bivariate approach facilitates the VaR forecast at nominal coverage level α

$$VaR_{t,\alpha} = \mu_{t|t-1} + \sqrt{QV_{t|t-1}}F_\epsilon^{-1}(\alpha) \tag{12}$$

where $\mu_{t|t-1}$ and $QV_{t|t-1}$ are obtained from (6) and (7) by exploiting the daily return decomposition $r_t = r_{o,t} + r_{d,t}$, and $F_\epsilon^{-1}(\alpha)$ is the α-quantile of the skewed Student-t daily standardized innovation $\epsilon_t (= \epsilon_{o,t} + \epsilon_{d,t})$ which is estimated sequentially over rolling windows.

2.2 Univariate 'Bundling' Modeling Approach

When applied to individual stocks or equity cash indices that are only traded for part of the day, the realized variance estimator RV_t is based on prices observed from market open (O_t) to close (C_t). Hence, it can underestimate the true notional daily quadratic variation QV_t. To mitigate this problem, numerous volatility forecasting studies resort to 'bundling' the squared overnight return and daytime RV_t into an overnight-adjusted realized volatility measure that spans a full 24-hour period.

Hansen and Lunde (2005) propose the overnight-adjusted realized estimator variance

$$RV_t^{HL} \equiv \omega_1^* r_{o,t}^2 + \omega_2^* RV_t, \tag{13}$$

where ω_1^* and ω_2^* are weights that minimize the variance of RV_t^{HL} subject to $\omega_1\mu_1 + \omega_2\mu_2 = \mu_0$ with $\mu_1 \equiv E[r_{o,t}^2]$, $\mu_2 \equiv E[RV_t]$ and $\mu_0 \equiv E[QV_t]$, $t = 1, ..., T$. For applications, see Martin et al. (2009), Fleming and Kirby (2011), Fuertes and Olmo (2012), and Ahoniemi and Lanne (2013) inter alia.

Other studies employ more ad hoc overnight adjustments aimed towards scaling the RV_t measure upwards so that it spans a 24-hour period; for instance, the estimator $RV_t^{SC} \equiv \frac{\sum_{t=1}^T r_{d,t}^2 + \sum_{t=1}^T r_{o,t}^2}{\sum_{t=1}^T r_{d,t}^2} RV_t$ is employed by Koopman et al. (2005) and Angelidis and Degiannakis (2008) inter alia. Another ad hoc overnight adjustment treats the 17.5-hour overnight return (from 4:00pm to 9:30am) in the same way as each of the 5-minute intraday returns, $RV_t^{+ON} \equiv r_{o,t}^2 + r_{1,t}^2 + r_{2,t}^2 + ... + r_{M,t}^2 = r_{o,t}^2 + RV_t$, which amounts to using the naive weights $(\omega_1, \omega_2) = (1, 1)$ in equation (13) above; for applications, see Blair et al. (2001), Giot and Laurent (2004), and Bollerslev et al. (2009) inter alia.

[7] Our interest is in long trading positions. For short trading positions one would analyze instead the right tail, i.e. $F_t^{-1}(1-\alpha)$. Commercial banks are required to report VaR at confidence level 99% to regulators but most banks adopt the 95% level for internal backtesting. We consider both levels using $\alpha = \{0.01, 0.05\}$.

Ahoniemi and Lanne (2013) confront the above three overnight-adjusted realized variance estimators. By proxying the true daily volatility of the S&P 500 index with the daily squared return, they show that RV_t^{HL} stands out as the most accurate estimator in a purely statistical sense.

2.2.1 Univariate Daily Return Level and Volatility Forecasts

We parameterize the long memory behaviour of RV_t^{HL} with an ARFIMA$(0,d,1)$ equation. An AR(1) equation is fitted to a time-series of daily (close-to-close) returns under the assumption that all the dynamics in their conditional variance is ARFIMA type. This AR-ARFIMA model can be written as

$$r_t = c_0 + c_1 r_{t-1} + \sqrt{\sigma^2 \cdot RV_t^{HL}} \cdot \epsilon_t \tag{14a}$$

and

$$(1-L)^d(\ln RV_t^{HL} - \tau_0 - \tau_1 r_{t-1} - \tau_2 I_{t-1}^{-} r_{t-1}) = (1+\theta L)u_t, \tag{14b}$$

where ϵ_t is an i.i.d. standardized skewed Student-$t(0,1,\xi,\eta)$ innovation. This approach is relatively parsimonious but restrictive in that it does not acknowledge the different predictability of the *overnight* and *daytime* return processes, and therefore also precludes modeling their covariance.

2.2.2 Univariate Bundling-based VaR Forecasts

The forecasts obtained from the 'bundling' univariate approach serve as building blocks to construct 1-step-ahead VaR predictions for the daily return process as follows

$$VaR_{t|t-1,\alpha}^{bundle} = \widetilde{\mu}_{t|t-1} + \sqrt{\widetilde{QV}_{t|t-1}} F_{\epsilon}^{-1}(\alpha) \tag{15}$$

where $\widetilde{\mu}_{t|t-1}$ is the conditional mean forecast of the close-to-close return, obtained from the AR equation (14a) and $\widetilde{QV}_{t|t-1}$ is the close-to-close variance forecast from the ARFIMA equation (14b).

2.3 Alternative Bivariate (Without Covariance) Approaches

As noted earlier, one goal of this paper is to assess the validity of Hypothesis I, which maintains that modeling the covariance between *overnight* returns and *daytime* returns can be beneficial for setting equity-trading risk limits. This assessment cannot be carried out by comparing the bivariate and bundling approaches (outlined in Sections 2.1 and 2.2, respectively) since it would then not be possible to ascertain whether any improvements in VaR forecast accuracy afforded by the proposed bivariate approach stem from the separate modeling of overnight and daytime returns or from modeling the covariance. To disentangle these effects, we deploy two simpler bivariate modeling approaches.

Firstly, the 'bivariate without (w/o) covariance' modeling approach constructs the VaR predictions via equation (12), as before, with the only difference that the daily variance forecast $QV_{t|t-1}$, is obtained from a more restrictive version of equation (7) that excludes the conditional covariance component, i.e. $Cov(r_{o,t}, r_{d,t}|\ \Im_{t-1}) \equiv 0$. The second bivariate approach (also without covariance) redefines the overnight and daytime segments by moving the open price k minutes further into the trading day. The aim is to mitigate market microstructure effects at the market open such as (overnight) news spillover and price staleness. VaR forecasts from this 'bivariate $k$min further' approach are constructed with the individual components of (12) obtained from the AR-GARCH equations (8a)-(8b) and AR-ARFIMA equations (9a)-(9b) fitted to the redefined overnight and daytime returns, respectively.

3 VALUE-AT-RISK FORECAST EVALUATION

3.1 Predictive Ability Tests

Competing VaR models can be compared through Equal Predictive Ability (EPA) tests given that VaR measurement is an out-of-sample forecasting exercise. The main idea is to evaluate a predictive loss function for each of two competing models and gauge the significance of the difference. This literature was initiated by Diebold and Mariano (1995) and West and McCracken (1998) inter alia, and extended to a conditional framework by Giacomini and White (2006).

Let the in-sample (estimation) and out-of-sample (evaluation) periods comprise P and n days, respectively, with $P + n = T$. The forecasting object, $VaR_{t,\alpha}$, is the α-quantile of the conditional distribution of the daily return, r_t. A natural evaluation tool is the piecewise-linear 'check' loss function

$$L_t(e_t(\widehat{\theta}_{P,t-1})) = (\alpha - 1(e_t(\widehat{\theta}_{P,t-1}) < 0))e_t(\widehat{\theta}_{P,t-1}) \tag{16}$$

where $e_t(\widehat{\theta}_{P,t-1}) \equiv r_t - VaR_{t,\alpha}(\widehat{\theta}_{P,t-1})$ is the forecasting error, and $\widehat{\theta}_{P,t-1}$ is a rolling estimator of the vector θ_0 that collects the parameters of the conditional mean and variance equations (including the parameters that define the innovation distribution) used to compute the VaR forecasts.

We utilize the unconditional EPA test developed by Giacomini and White (2006) which extends the original Diebold and Mariano (1995) test by controlling for parameter uncertainty. This is accomplished by letting n go to infinity while P remains finite; thus, recursive forecasting is precluded. Following Giacomini and White (2006), we employ a rolling forecasting scheme which has the additional advantage of providing some shield against instability in the data generating process. It follows that, since the EPA test considers the effect of estimation method and in-sample size, it serves to assess the relative performance of VaR forecasting methods, not just VaR forecasting models.

In the present context, the null hypothesis of the unconditional EPA test can be expressed as

$$H_0 : E[L_t(e_t^{biv}(\widehat{\theta}_{P,t-1})) - L_t(e_t^{\mathcal{M}}(\widehat{\theta}_{P,t-1}))] = 0 \text{ a.s. for } t = P+1, \ldots, T \tag{17}$$

where e_t^{biv} and $e_t^{\mathcal{M}}$ are the errors associated with the VaR^{biv} and $VaR^{\mathcal{M}}$ forecasts, respectively; *biv* denotes the proposed bivariate (with covariance) model and $\mathcal{M}$ is any of the alternative models. This null hypothesis can be expressed more compactly as $H_0 : E[\Delta L_t] = 0$ with $\Delta L_t \equiv L_t(e_t^{biv}(\widehat{\theta}_{P,t-1})) - L_t(e_t^{\mathcal{M}}(\widehat{\theta}_{P,t-1}))$. The test is based on an out-of-sample t-statistic computed as $\sqrt{n}$ times the ratio of the sample mean of ΔL_t to its sample standard deviation using the Newey-West variance estimator to account for autocorrelation. Under H_0 the asymptotic distribution of the test statistic is $N(0, 1)$. We focus on the one-sided test version, $H_0 : E[\Delta L_t] \leq 0$ against $H_A : E[\Delta L_t] > 0$. EPA test rejection implies that on average over the out-of-sample period (n days) the loss associated with VaR^{biv} exceeds that of $VaR^{\mathcal{M}}$ and hence, the alternative model $\mathcal{M}$ at hand has superior predictive ability.

3.2 Correct Conditional Coverage Tests

We deploy two backtesting approaches to assess whether the following criterion of correct specification of an αth VaR model for daily portfolio returns r_t is satisfied out-of-sample

$$P(r_t \leq VaR_{t,\alpha}(\theta_0) \mid \Im_{t-1}) = \alpha, \text{ almost surely (a.s.)}, \alpha \in (0, 1), \forall t \in \mathbb{Z}. \tag{18}$$

This criterion, often referred to as Correct Conditional Coverage, is central to many theoretical discussions on VaR modeling; see e.g., Christoffersen et al. (2001), Engle and Manganelli (2004), and Fuertes and Olmo (2012). Let the out-of-sample demeaned hits or violations be denoted by $Hit_{t,\alpha}(\theta_0) \equiv 1(r_t \leq VaR_{t,\alpha}(\theta_0)) - \alpha$. If criterion (18) is met, then the expected value of $Hit_{t,\alpha}(\theta_0)$ conditional on the information set $\Im_{t-1}$ is α which, in turn, implies that $Hit_{t,\alpha}(\theta_0)$ is independent of any function of the variables in $\Im_{t-1}$. Intuitively, the VaR violations should be unpredictable; if future violations can be predicted then there is useful information that has not been incorporated in $\Im_{t-1}$.

Our first backtesting approach is the *dynamic quantile* (DQ) test of Engle and Manganelli (2004). It can be cast as a test for overall statistical significance of the linear probability regression

$$Hit_{t,\alpha}(\widehat{\theta}_{P,t-1}) = x_{t-1}(\widehat{\theta}_{P,t-1})\gamma + v_t, \ t = P + 1, \dots, T, \tag{19}$$

where γ is a $k \times 1$ parameter vector and v_t a zero-mean *iid* error sequence. The Correct Conditional Coverage hypothesis, $H_0 : E[Hit_{t,\alpha}(\widehat{\theta}_{P,t-1}) \mid x_{t-1}(\widehat{\theta}_{P,t-1})] = 0$ where $x_{t-1}(\widehat{\theta}_{P,t-1})$ is a vector of k regressors including a constant, can be stated as $H_0 : \gamma_0 = \gamma_1 = ... = \gamma_{k-1} = 0$. The Wald test statistic suggested by Engle and Manganelli (2004) can be expressed as

$$DQ = n\frac{\widehat{\gamma}_n'[M_n(\widehat{\theta}_P)]\widehat{\gamma}_n}{\alpha(1-\alpha)} \tag{20}$$

where $M_n(\widehat{\theta}_P) = \frac{1}{n}\sum\limits_{t=P+1}^{T} x_{t-1}'(\widehat{\theta}_{P,t-1})x_{t-1}(\widehat{\theta}_{P,t-1})$ is a $k \times k$ matrix; and $\widehat{\gamma}_n$ is the consistent and asymptotically normal OLS estimator of γ. Under H_0, it can be shown that $DQ \xrightarrow{d} \chi^2_k$ as $n \to \infty$.

The binary nature of the hits implies heteroskedasticity in regression (19) by construction. Nonlinear probit or logit regressions have been employed recently as refinements of the original DQ test since the asymptotic ML standard error formulae takes the heteroskedasticity into account naturally, and the fitted values (probabilities) are bounded between 0 and 1; see e.g., Berkowitz et al. (2011) and Dumitrescu et al. (2012). In our analysis, we adopt the *dynamic binary* (DB) probit regression

$$E[1(r_t \leq VaR_{t,\alpha}(\widehat{\theta}_{P,t-1})) \mid x_{t-1}(\widehat{\theta}_{P,t-1})] = \Phi(x_{t-1}(\widehat{\theta}_{P,t-1})\beta),\ t = P+1, \ldots, T \tag{21}$$

where $\Phi(\cdot)$ is the cumulative standard Normal distribution, and $\beta \equiv (\beta_0, \beta_1, ..., \beta_{k-1})'$ the parameter vector. In this probit setting, the Correct Conditional Coverage criterion (18) amounts to

$$E[1(r_t \leq VaR_{t,\alpha}(\widehat{\theta}_{P,t-1})) \mid x_{t-1}(\widehat{\theta}_{P,t-1})] = \Phi(\beta_0) = \alpha \tag{22}$$

or equivalently $\widetilde{H}_0 : \beta_0 = \Phi^{-1}(\alpha), \beta_1 = ... = \beta_{k-1} = 0$, and assessed via the likelihood ratio statistic

$$DB = 2(\mathcal{L} - \mathcal{L}_0) \tag{23}$$

with $\mathcal{L} = \sum_{t=P+1}^{T} [Hit_{t,\alpha}(\widehat{\theta}_{P,t-1}) ln\ \Phi(x_{t-1}(\widehat{\theta}_{P,t-1})\widehat{\beta}_n) + (1 - Hit_{t,\alpha}(\widehat{\theta}_{P,t-1})) ln\ (1 - \Phi(x_{t-1}(\widehat{\theta}_{P,t-1})\widehat{\beta}_n))]$ the log-likelihood of model (21), and $\widehat{\beta}_n$ is the consistent and asymptotically normal ML estimator of β. Under the null hypothesis it follows that $\mathcal{L}_0 = \sum_{t=P+1}^{T} [Hit_{t,\alpha}(\widehat{\theta}_{P,t-1}) ln\ \alpha + (1 - Hit_{t,\alpha}(\widehat{\theta}_{P,t-1})) ln\ (1-\alpha)]$, and it follows that $DB \xrightarrow{d} \chi^2_k$ as $n \to \infty$. Dumitrescu et al. (2012) show that the probit-based test has less size distortions and better power properties than the original DQ test for small samples.

We follow Engle and Manganelli (2004) and Dumitrescu et al. (2012) in adopting the regressor set

$$x_{t-1}(\widehat{\theta}_{P,t-1}) = (1, r_{t-1}, r^2_{t-1}, VaR_{t-1,\alpha}(\widehat{\theta}_{P,t-1}), Hit_{t-1,\alpha}(\widehat{\theta}_{P,t-1})),$$

as $\Im_{t-1}$ which allows the hit on day t to depend on the previous return, volatility, VaR and hit.

4 EMPIRICAL RESULTS

4.1 Data and Descriptive Statistics

Two stock market indices are chosen as well-diversified portfolios for our VaR forecasting analysis. One is the S&P 500 index, which is by far the most common benchmark for exchange traded, mutual, and pension funds that identify themselves as large cap. The other index is Russell 2000, a typical small-cap benchmark. The high-frequency prices are from *Disk Trading* at sampling frequency of 5 minutes from 9:30-16:00 Eastern Standard Time (EST) which amounts to $M = 78$ intraday intervals.[8] The closing price, $p_{C_t} \equiv p_{t,M}$, is the last price observed before 16:00. The jth intraday price, $p_{t,j}$ with $j = 1, ..., M-1$, is the last seen tick before the jth 5-minute mark. The opening price, $p_{O_t} \equiv p_{t,1}$, is the first index quote published after 9:30 when trading officially begins. The indices are observed over the 14 years from November 12, 1997 to September 30, 2011 ($T = 3491$ trading days).

In order to make the VaR forecasting task more challenging, the out-of-sample period begins on September 2, 2008 ($n = 778$ days), an important landmark of the late 2000s global financial crisis due to the market turbulence resulting from the Lehman Brothers bankruptcy and other events.[9] Figure 1 plots the time-series of squared overnight returns and realized variances, and a histogram (alongside the theoretical Normal density) of the logarithmic realized variances. Table 1 provides summary statistics for overnight and daytime returns over pre-Lehman and post-Lehman periods.

[Insert Figure 1 and Table 1 around here]

The standard deviation of daytime returns ($r_{d,t}$) is substantially higher than that of overnight returns ($r_{o,t}$), a finding well aligned with the wisdom that there is greater information flow during regular trading hours; both volatility measures increase dramatically post-Lehman. The relatively mild (negative) skewness and kurtosis of daytime (versus overnight) returns observed in the pre-Lehman period become notably more exacerbated post-Lehman. Overnight and daytime returns pertaining to the same day ($r_{o,t}, r_{d,t}$) are significantly positively correlated over the two sample periods; however, the correlation increases substantially post-Lehman. The overnight and daytime squared returns, as crude ex-post volatility proxies over the two segments of the day are summarized in Panel B of Table 1. To make the comparison more informative, we report the hourly volatility given by the mean squared return scaled by the total hours spanned by each segment of the day, 17.5 hours (overnight) or 6.5 hours (daytime). Higher daytime return volatility is thus confirmed.[10] The autocorrelation function of squared returns shows a slower decay (i.e., more persistence) at daytime than overnight. In sum, there are important contrasts between the *overnight* and *daytime* return generating processes.

High positive skewness and kurtosis of RV in Panel C of Table 1 corroborate that investors face non-normally distributed risks. The coefficient of variation is larger in the out-of-sample (post-Lehman) period indicating a decrease in the signal-to-noise ratio during the late 2000s financial crisis. The logarithmic RV time series for both S&P 500 and Russell 2000 are approximately Gaussian, both pre-

[8] *http://disktrading.is99.com.* For highly liquid assets, the 5-minute frequency is short enough for the daily volatility dynamics to be picked up with reasonable accuracy (small estimation error) and long enough for the adverse effects of market microstructure noise (e.g., bid-ask bounce, discrete price observations, irregular trading) not to be excessive.

[9] Fannie Mae and Freddie Mac, two U.S. government sponsored enterprises, owned or guaranteed nearly $5 trillion in mortgage obligations at the time they were placed into conservatorship by the U.S. government on September 7, 2008.

[10] Although unreported to preserve space, we also confirm that the overnight volatility (proxied by the squared overnight return) increases when the overnight segment is lengthened by waiting k mins after the market officially opens at 9:30 EST, for $k = \{5, 15, 30\}$. Likewise, the daytime volatility (proxied by either the mean squared open-to-close return or the mean realized variance) decreases when the daytime segment excludes the first k mins after the market opens.

and post-Lehman, as suggested by (unreported) skewness coefficients ranging from 0.17 to 0.50, and kurtosis coefficients ranging from 3.29 to 3.72.

Over the 14-year sample period, the amount of daily volatility ascribed to the non-trading hours according to the ratio $r_{o,t}^2/(r_{o,t}^2 + RV_t)$ is 5.95% for Russell 2000 and 2.83% for S&P 500. For Russell 2000, the ratio more than doubles from 4.70% in the pre-Lehman period (2713 days) to 10.28% in the post-Lehman period (778 days); the increase is milder for S&P 500 from 2.77% to 3.02%.

4.2 VaR Predictions Combining Overnight and Intraday Information

We first proceed with the *rolling* forecasting exercise which involves $n(=778)$ out-of-sample 1-day-ahead VaR predictions from each of the modeling approaches described in Section 2. The estimation window length is $P = T - n = 2713$ days. The first forecast is based on the model parameter vector estimated with data from day 1 to P, denoted $\widehat{\theta}_P$, the second forecast is based on $\widehat{\theta}_{P+1}$ from days 2 through $P+1$, and so forth. The construction of the 'bundled' overnight-adjusted measure RV^{HL} requires weights which are estimated in a way that avoids look-ahead bias and preserves the out-of-sample nature of the exercise; the weights to obtain $\{RV_t^{HL}\}_{t=1}^{P}$ are based on information up to day P, those to obtain $\{RV_t^{HL}\}_{t=2}^{t=P+1}$ exploit information from day 2 through day $P+1$, and so forth.[11]

Then we proceed with the formal evaluation of VaR models through the backtesting methods described in Section 3 which are deployed sequentially over windows of $\tilde{n} < n = 778$ forecasts (as opposed to backtesting all n forecasts at once). This dynamic evaluation approach allows us to gauge the extent to which changes in market conditions influence VaR performance. We employ $J = 279$ *overlapping* windows with $\tilde{n} = 500$ forecasts in each; the first window begins on September 2, 2008. In addition, we conduct the tests over $J = 10$ *non-overlapping* windows which implies a much smaller number of forecasts, $\tilde{n} = 78$, in each. Hence, there is a trade-off between the (non)overlapping aspect of the evaluation windows and the (small) large number of VaR forecasts available in each window.

Table 2 summarizes the unconditional EPA test that assesses the accuracy of VaR forecasts from the bivariate model vis-à-vis each of three competitors – the nested 'bivariate w/o covariance' model, the 'bivariate 15min further' model, and the 'bundling' RV^{HL} model.[12] The null hypothesis of the EPA test is that the forecasts from the proposed bivariate (with covariance) model are at least as good as those from the competitor at hand. Rejection rates are computed as $\sum_{j=1}^{J} w_j 1(p_j < 0.05)$ where p_j is the test p-value associated with the jth window, $j = 1, ..., J$, and $w_j \equiv 1/J$. Panel A and Panel B pertain to the rolling and non-overlapping evaluation exercises, respectively.

[Insert Table 2 around here]

[11] The estimation and forecasting are carried out in Oxmetrics 6 using the G@RCH 6.1 and ARFIMA 1.04 packages.

[12] The results for $k = \{5, 30\}$ are qualitatively similar to those for $k = 15$ and hence, we do not report them to preserve space. We also deployed the 'bundling' univariate model using RV_t^{+ON} and RV_t^{SC} but the resulting VaR measures had inferior predictive ability to those from RV_t^{HL}. This is in line with the statistical ranking of the three overnight-adjusted measures provided in Ahoniemi and Lanne (2013). Detailed results are available from the authors.

The rejection rates of the EPA test for Russell 2000 are 0% for all competitors except the 'bivariate w/o covariance' model for which they reach 11%. Thus the proposed bivariate model yields significantly more accurate VaR forecasts than any of the competing models for the small-cap portfolio. The evidence for the S&P 500 portfolio also favors the proposed bivariate model although the 'bivariate 15min further' model now emerges as the strongest competitor; the rejection rate of the EPA test in favor of the latter model reaches 24.7% over rolling windows and 30% over non-overlapping windows. Panels A2-A3 and B2-B3 of Table 2 suggest that only over the first 10% of the J out-of-sample windows denoted $1/10J$ (the first window begins on September 2008 when Lehman Brothers filed for bankruptcy) the 'bivariate 15min further' model dominates the proposed bivariate model; over the subsequent 90% of windows the proposed bivariate model strongly outperforms all the competitors.

To sum up, the rejection rates of the EPA test reported in Table 2 represent very little evidence against Hypothesis I, which leads us to conclude that there is considerable merit in modeling the covariance between overnight and daytime returns. Also, we observe that the VaR forecasts from the proposed bivariate model are clearly superior to those from the 'bundled' model. This evidence cannot refute Hypothesis II, and so we conclude that risk managers can benefit from modeling the overnight and daytime volatilities separately. This is confirmed by unreported EPA tests to assess the null that the bivariate w/o covariance model has superior predictive ability over the bundling model, which produce 0% rejection rates across both rolling and non-overlapping evaluation schemes. This is an important finding given the popularity of overnight-adjusted (bundled) estimators in the literature. The inferior performance of this coarse approach for tail risk forecasting can be ascribed to the fact that it assumes that the overnight and daytime returns are generated by the same process.

We now turn to the Correct Conditional Coverage tests, namely, the DQ test based on the linear probability regression and the counterpart test based on the nonlinear probit regression. Table 3 reports the VaR backtesting rejection rates (over J sequential windows) obtained as $\sum_{j=1}^{J} w_j 1(p_j < 0.05)$ where p_j is the estimated p-value of the corresponding test over the jth window and $w_j \equiv 1/J$.

[Insert Table 3 around here]

As shown in Panel A1 of the table, for the Russell 2000 portfolio the proposed bivariate model always produces the lowest VaR backtesting rejection rate over the $J = 279$ rolling windows. By contrast, the 'bivariate 15min further' model performs rather poorly, failing to meet the Correct Conditional Coverage criterion over most of the rolling windows. In the context of the S&P 500 portfolio, the proposed bivariate model does not always produce the lowest backtesting rejection rate. For 5% VaR, the 'bivariate 15min further' model and the 'bundling RV^{HL}' models alternate as those that meet the Correct Conditional Coverage criterion most often. However, their rejection rates are only slightly smaller than those of the proposed bivariate (with covariance) model.

In order to assess the significance of differences in backtesting rejection rates across models, we deploy a difference-in-proportions (DIP) test based on the statistic

$$\Pi_{J,P} \equiv \sqrt{J}\frac{\widehat{P}_{biv} - \widehat{P}_{\mathcal{M}}}{\sqrt{\widehat{P}_{biv}(1-\widehat{P}_{biv}) + \widehat{P}_{\mathcal{M}}(1-\widehat{P}_{\mathcal{M}})}}, \tag{24}$$

where P_{biv} and $P_{\mathcal{M}}$ are the population rejection rates of the bivariate model and a given competitor, respectively, estimated as (weighted) averages of the binary backtesting outcomes $1(p_j < 0.05)$ over $j = 1, ..., J$ rolling windows. The hypothesis of interest is $H_0 : P_{biv} \leq P_{\mathcal{M}}$ versus $H_A : P_{biv} > P_{\mathcal{M}}$. A thorny issue is that the binary backtesting outcomes are dependent across models and autocorrelated which invalidates the asymptotic approximation of the DIP test statistic distribution by the standard Normal. A well-established solution is to approximate the finite-sample distribution of the unstandardized test statistic by bootstrap techniques. More specifically, we construct the moving-block bootstrap distribution of $\sqrt{J}(\widehat{P}_{biv} - \widehat{P}_{\mathcal{M}})$ by resampling pairs of binary backtesting outcomes for each $(biv, \mathcal{M})$ combination. The block length is selected using Patton et al.'s (2009) data-driven method.[13]

In Panel A of Table 3, the rejection rate of criterion (18) by each competitor is marked with asterisks if it turns out to be significantly smaller (at the conventional 10%, 5% or 1% levels) than that of the proposed bivariate model according to the bootstrap DIP test. For Russell 2000, the unreported p-values of the DIP test range from 0.933 to 1.000 and so there are no rejections; hence, the proposed bivariate model is at least as good as any of the competitors. For the S&P 500 portfolio, the only significant DIP test statistic is associated with the probit-based backtesting of 5% VaR; in this case, the VaR forecasts from the 'bundling RV^{HL}' model fail to meet criterion (18) at a rate of 0.387 which is significantly smaller than the rejection rate of the bivariate model at 0.566. The same qualitative finding is shown in Panel A3. For the conservative 1% VaR, the proposed bivariate model meets the conditional coverage criterion (18) at least as often as any of the competitors.[14]

The rejection rates of the Correct Conditional Coverage criterion (18) reported in Panels B1-B3 pertain to the dynamic non-overlapping evaluation scheme. These rejection rates do not radically challenge the main evidence that modeling separately the daytime and overnight return processes as well as their covariance is likely to be beneficial for tail risk management. However, the number of binary backtesting outcomes, $1(p_j < 0.05)$, $j = 1, ..., J$, to estimate the rejection rates is too small ($J = 10$ non-overlapping windows) which obviates the DIP test because neither the asymptotic standard Normal distribution nor the moving-block bootstrap distribution are applicable.

As borne out by Panels A2 and B2 of Table 3, the period spanned by the first 10% windows of out-of-sample forecasts (from September 2008 onwards) represents a clear challenge for all VaR forecasting models in meeting the Correct Conditional Coverage criterion (18). This is particularly visible in Panel B2 where the rejection rates associated with the $1/10J$ windows (an evaluation period spanning the 78

[13]The Matlab code for implementing this method is available from Andrew Patton's web page.

[14]The specific p-values of the DIP test obtained for each pair of models $(biv, \mathcal{M})$ are available from the authors.

days inmediately post-Lehman) are 100% in all cases, which is unsurprising given that this subperiod epitomizes the brunt of the late 2000s financial crisis. The frequency with which the VaR modeling approach fails to deliver forecasts that meet criterion (18) drops sharply from the first $1/10J$ windows (Panels A2 and B2) to the subsequent $9/10J$ windows (Panels A3 and B3). This means that as more of the extremely volatile post-Lehman days are included in model estimation, the accuracy of the out-of-sample VaR forecasts improves. The 1980-2011 analysis of VaR forecasts in Frey et al. (2013) leads to a similar observation on the adverse influence of abrupt changes in market conditions on VaR forecast performance, despite important methodological differences with the present study.

The conditional coverage rejection rates examined thus far assign equal weights to all VaR backtesting outcomes. However, the actual *economic* loss function of a risk manager may penalize rejections according to the size and the sign of the coverage error. Panels A1 and B1 of Table 4 summarize the dynamic conditional coverage backtesting by weighting each rejection with the size of the coverage error, i.e. $w_j \equiv |\widehat{\alpha}_j - \alpha_j|$ for $j = 1, ..., J$, appropriately standardized so that the weights add to unity. Panels A2, A3, B2 and B3 take also into account the direction of the error. Panels A2 and B2 are relevant for active asset management when there is more aversion towards underpredicting downside tail risk (which entails uncovered losses) than towards overpredicting it (opportunity costs). Thus, the penalty is 'large' at $|\widehat{\alpha}_j - \alpha_j|$, $j = 1, ..., J_1$ for underpredictions and 'small' at $(\widehat{\alpha}_j - \alpha_j)^2$, $j = 1, ..., J_2$ for overpredictions ($J = J_1 + J_2$). Panels A3 and B3 reverse the above asymmetry, namely, overpredictions are weighed more heavily than underpredictions. This weighting scheme is broadly evocative of loss functions of large banks which have permission to calculate the capital they must hold against their trading books if they are less inclined to maintain idle capital than to bear out regulatory penalties.[15]

[Insert Table 4 around here]

As shown in Panels A1-A3 of Table 4, the proposed bivariate (with covariance) model still dominates the competitors for the Russell 2000 portfolio when the backtesting outcome over each *rolling* window is weighted by the magnitude of the losses in excess of the predicted VaR. With the *non-overlapping* windows (Panel B) there is also clear evidence from both the DQ and probit-based tests suggesting that the bivariate model yields superior 5% VaR predictions for the Russell 2000 portfolio.[16] However, the weighted backtesting results are less clearcut for the large-cap portfolio. The weighted rejection rates of the conservative 1% VaR for the S&P 500 portfolio still remain the lowest with the proposed bivariate model in Panel A but the results for 5% VaR tend to favor the 'bivariate 15min

[15] Details on the Basel trading book rules can be found at *http://www.bis.org/publ/bcbs158.pdf*.

[16] Following Andersen et al. (2011), we considered a specification for the overnight return volatility that exploits the immediately preceding realized variance, $h_{o,t} = \alpha_0 + \alpha_1 \varepsilon^2_{o,t-1} + \beta_1 h_{o,t-1} + \gamma_1 RV_{t-1} + \gamma_2 I^-_{t-1} RV_{t-1}$, instead of equation (8b). The main findings are unchallenged, namely, the Equal Predictive Ability tests and Correct Conditional Coverage tests still point to the VaR forecasts from the bivariate modeling approach as superior. Furthermore, we deployed the evaluation tests over post-June 2009 windows of forecasts and the results are qualitatively similar, which rules out the recent financial crisis as the driver of our findings. Details of both robustness checks are available from the authors.

further' model. The DIP test confirms that the rejection rates of the latter are significantly lower than those yielded by our model in four cases. With the non-overlapping windows the results are also tilted towards the 'bivariate 15min further' modeling approach for the S&P 500 portfolio although in this case the DIP test to assess significance of the difference is not feasible.

To further illustrate the relative merit of the proposed bivariate model for setting equity-trading risk limits, Figure 2 plots the sequential p-values of the DQ test (for the null hypothesis of correct conditional VaR coverage) obtained over J=279 rolling windows of $\tilde{n} = 500$ out-of-sample days.[17]

[Insert Figure 2 around here]

The figure reveals substantial instability in absolute and relative out-of-sample VaR forecast accuracy from the lens of the Correct Conditional Coverage criterion (18), which endorses our dynamic forecast evaluation scheme; see also Frey et al. (2013). For the small-cap portfolio, with the exception of the initial part of the out-of-sample period that captures the brunt of Lehman's debacle, the p-values of the DQ test tend to be the largest (leading to less rejections) for the proposed bivariate model; this finding is aligned with the relatively low rejection rates reported for this model in Table 3. For the large-cap S&P 500 portfolio, the graphs produce less clearcut evidence on the superior forecast accuracy of the bivariate model, particularly, for the 5% VaR measure.

Overall, the Russell 2000 portfolio analysis has produced stronger evidence in favor of the proposed bivariate (with covariance) model than the S&P 500 portfolio analysis. The contrast may stem from the distinct trading volumes of large- versus small-cap stocks, which has price discovery implications. The literature has documented greater efficiency of price discovery at the market open for high trading volume stocks; see e.g. Barclay and Hendershott (2008). The finding that the proposed bivariate approach appears particularly useful for setting small-cap equity trading limits indirectly confirms the presence of market microstructure effects (i.e., price staleness and news spillover) at the market open for relatively low trading volume stocks. In contrast, for large-cap stocks any news accumulated during non-trading hours are likely to be impounded into prices more rapidly as the market opens, which naturally dilutes the merit of modeling the covariance between overnight and daytime returns.

Thus, our findings are in line with the evidence in Barclay and Hendershott (2008). If indeed the opening price of large-cap stocks conveys more information than that of (relatively lower volume) small-cap stocks, then it is plausible to find that the proposed bivariate (with covariance) modeling approach clearly excels for setting equity trading limits with the Russell 2000 portfolio.

[17] The graphs of p-values for the probit variant of the test, available on request, are qualitatively similar.

5 CONCLUSION

Risk managers of banks and other financial institutions face the task of establishing mark-to-market loss limits for their different trading desks or business lines. Although accurate forecasts of downside risk cannot *per se* prevent losses, they provide risk managers of financial institutions with objective measures to act upon. Since the 1990s, commercial banks have routinely estimated daily Value-at-Risk (VaR) for this purpose. Given its practical relevance, various VaR modeling approaches have been put forward in the literature, but the issue of how to incorporate overnight information for predicting 1-day-ahead VaR has been largely neglected. This paper seeks to contribute towards filling this gap.

We propose modeling the dynamics of the overnight and daytime returns, and their covariance. Cross-dependencies between overnight and daytime returns can be theoretically motivated by market microstructure frictions, such as price staleness and news spillover, that are likely to occur at the beginning of the trading day and represent price discovery inefficiencies. The proposed model is confronted with simpler bivariate models that disregard the covariance, and with the widely-used univariate modeling approach that relies on an overnight-adjusted realized volatility measure.

The analysis is based on 14-year samples of intraday data for the S&P 500 and Russell 2000 indices. The first key finding is that modeling the covariance between daytime and overnight returns can be useful to set appropriate equity-trading risk limits, particularly, in the context of the small-cap Russell 2000 portfolio. The superior accuracy of the 1-day-ahead VaR forecasts for Russell 2000 from the proposed bivariate modeling approach, as borne out by dynamic Equal Predictive Ability tests and Correct Conditional Coverage tests, indicates that the overnight-daytime return covariance contains useful information for daily downside risk prediction. The contrasting findings for the two indices represent indirect evidence that the price discovery at market opening is more efficient for large-cap stocks. The contrast can be rationalized by the role of trading volume in price discovery. The low trading volume associated with small-cap stocks makes the impounding of overnight news into the opening price less efficient which, in turn, brings out the economic merit of modeling the covariance.

Our second key result is that the empirical evidence overwhelmingly favors the separate modeling of overnight and daytime return processes over the univariate modeling of overnight-adjusted realized variances for 1-day-ahead VaR prediction. This second result provides support for existing theoretical models of security returns that are built upon the premise that, in asset markets with periodic closures, the return generating process changes substantially from trading to non-trading hours implying different predictability. Therefore, in order not to compromise the solvency of trading positions monitored by VaR models, a recommendation for risk managers that arises from this analysis is to model the two segments of the day separately, and possibly account for their covariance. Finally, we confirm previous research in finding that all VaR models are seriously challenged when financial conditions change abruptly between the estimation and the forecasting periods.

References

[1] Ahoniemi, K., & Lanne, M. (2013). Overnight Stock Returns and Realized Volatility. International Journal of Forecasting, 29, 592–604.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X., & Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association, 96, 42–55.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. Econometrica, 71, 579–625.

[4] Andersen, T.G., Bollerslev, T. & Huang, X. (2011). A Reduced Form Framework for Modeling Volatility of Speculative Prices based on Realized Variation Measures. Journal of Econometrics 160, 176–189

[5] Angelidis, T., & Degiannakis, S. (2008). Volatility Forecasting: Intra-Day Versus Inter-Day Models. Journal of International Financial Markets, Institutions and Money, 18, 449–465.

[6] Barclay, M.J. & Hendershott, T. (2008). A Comparison of Trading and Non-Trading Mechanisms for Price Discovery. Journal of Empirical Finance, 15, 839–849.

[7] Basel Committee on Banking Supervision (1996). Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements. Bank for International Settlements, January 1996.

[8] Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating Value-at-Risk Models with Desk-Level Data. Management Science, 57, 2213–2227.

[9] Blair, B.J., Poon, S-H. & Taylor, S.J. (2001). Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High-frequency Index Returns. Journal of Econometrics, 105, 5–26.

[10] Bollerslev, T., Tauchen, G. & Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. Review of Financial Studies, 22, 4463–4492.

[11] Brownlees, C. T., & Gallo, G. M. (2010). Comparison of Volatility Measures: A Risk Management Perspective. Journal of Financial Econometrics, 8, 29–56.

[12] Chan, K., Chan, K.C. & Karolyi, G.A. (1991). Intraday Volatility in the Stock Index and Stock Index Futures Market. Review of Financial Studies, 4, 657–684.

[13] Christoffersen, P., Hahn, J., & Inoue, A. (2001). Testing and Comparing Value-at-Risk Measures. Journal of Empirical Finance, 8, 325–352.

[14] Clements, M.P., Galvão, A.B., & Kim, J.H. (2008). Quantile Forecasts of Daily Exchange Returns from Forecasts of Realized Volatility. Journal of Empirical Finance 15, 729–750.

[15] Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics 7, 174–196.

[16] de Gooijer, J.G., Diks, C.G.H. & Gatarek, L.T. (2012). Information flows around the globe: predicting opening gaps from overnight foreign stock price patterns. Central European Journal of Economic Modelling and Econometrics, 4, 23–44.

[17] Diebold, F.X. & Mariano, R.S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13, 253–263.

[18] Dumitrescu, E.I., Hurlin, C. & Pham, V. (2012). Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. Finance 33, 79–112.

[19] Engle, R.F. (2002). Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models. Journal of Business & Economic Statistics, 20, 339–350.

[20] Engle, R.F. & Manganelli, S. (2004). CAViaR Conditional Autoregressive Value-at-Risk by Regression Quantiles. Journal of Business & Economic Statistics, 22, 367–381.

[21] Fleming, J. & Kirby, C. (2011). Long Memory in Volatility and Trading Volume. Journal of Banking & Finance, 35, 1714–1726.

[22] French, K.R., & Roll, R. (1986). Stock Return Variances: The Arrival of Information and the Reaction of Traders. Journal of Financial Economics, 17, 5–26.

[23] Frey, R., Rodrigues, P., & Seeger, N. (2013). Out-of-Sample Performance of Jump-Diffusion Models for Equity Indices: What the Financial Crisis was Good for. SSRN Working Paper 2022909.

[24] Fuertes, A.-M. & Olmo, J. (2012). Optimally Harnessing Inter-Day and Intra-Day Information for Daily Value-at-Risk Prediction. International Journal of Forecasting, 29, 28–42.

[25] George, T.J., & Hwang, C-Y. (2001). Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing. Review of Financial Studies, 14, 979–1020.

[26] Giacomini, R., & H. White (2006). Tests of Conditional Predictive Ability. Econometrica 74, 1545–1578.

[27] Giot, P., & Laurent, S. (2004). Modeling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models. Journal of Empirical Finance 11, 379–398.

[28] Hansen, P.R., & Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. Journal of Financial Econometrics, 3, 525–554.

[29] Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. Journal of Finance, 46, 179–207.

[30] Hecq, A., Laurent, S. & Palm, F. (2012). Common Intraday Periodicity. Journal of Financial Econometrics, 10, 325–353.

[31] Hong, H. & Wang, J. (2000). Trading and Return under Periodic Market Closures. Journal of Finance 55, 297–354.

[32] Koopman, S.J., Jungbacker, B. & Hol, E. (2005). Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised, and Implied Volatility Measurements. Journal of Empirical Finance, 12, 445–475.

[33] Lockwood, L.J. & Lin, S. (1990). An Examination of Stock Market Return Volatility During Overnight and Intraday Periods 1964-1989. Journal of Finance, 45, 591–601.

[34] Martin, G.M., Reidy, A. & Wright, J. (2009). Does the Option Market Produce Superior Forecasts of Noise-Corrected Volatility Measures? Journal of Applied Econometrics, 24, 77–104.

[35] Masulis, R.W. & Ng, V.K. (1995). Overnight and Daytime Stock-Return Dynamics on the London Stock Exchange: The Impacts of "Big Bang" and the 1987 Stock-Market Crash. Journal of Business & Economic Statistics, 13, 365–378.

[36] Oldfield, G.S. & Rogalski, R.J. (1980). A Theory of Common Stock Returns Over Trading and Non-Trading Periods. Journal of Finance, 35, 729–751.

[37] Patton, A., Politis, D. N. & White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White, Econometric Reviews, 28, 372-375.

[38] Robinson, P.M. (1995). Gaussian Semiparametric Estimation of Long Range Dependence. Annals of Statistics, 23, 1630–1661.

[39] Slezak, S.L. (1994). A Theory of the Dynamics of Security Returns Around Market Closures. Journal of Finance, 49, 1163–1212.

[40] Stoll, H.R. & Whaley, R.E. (1990). Stock Market Structure and Volatility. Review of Financial Studies, 3, 37–71.

[41] Taylor, N. (2007). A Note on the Importance of Overnight Information in Risk Management Models. Journal of Banking & Finance, 31, 161–180.

[42] West, K. D., & McCracken, M. W. (1998). Regression-Based Tests of Predictive Ability. International Economic Review, 39, 817–840.

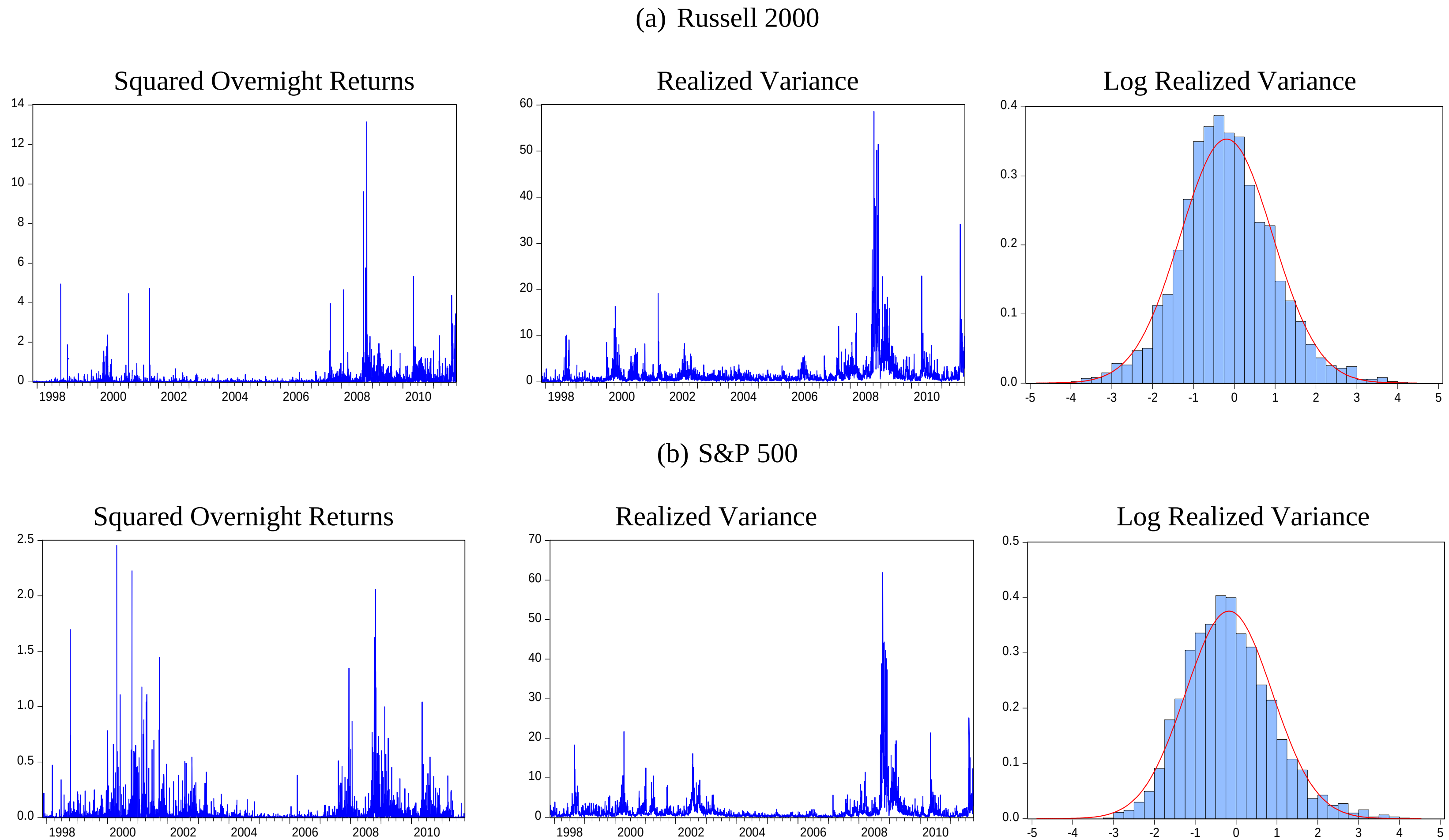

Figure 1. Overnight and daytime volatility of Russell 2000 and S&P 500 indices. The figure reports the squared previous-day-close to open log returns (first column), realized variance computed from 5-minute returns from 9:30 to 16:00 EST (second column) and histogram of logarithmic realized variance alongside theoretical Normal density (third column). The sample period begins on November 12, 1997 and ends on September 30, 2011. The out-of-sample period for VaR forecast evaluation begins on September 2, 2008.

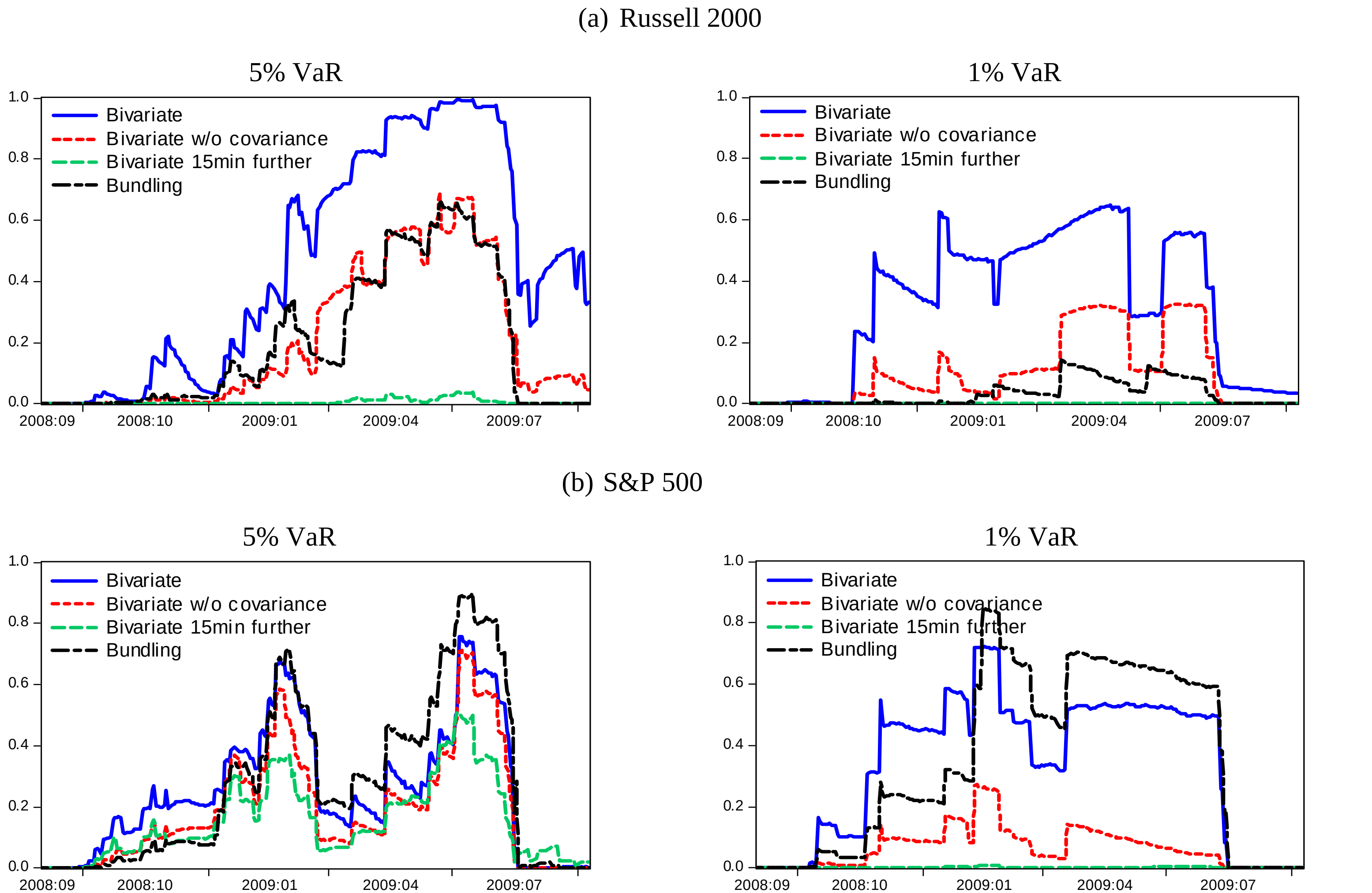

Figure 2. Dynamic quantile tests on the Correct Conditional Coverage of Value-at-Risk (VaR) forecasts. The figure reports p-values of the DQ test for correct conditional coverage over J=279 rolling windows of $\tilde{n}$=500 out-of-sample days each.

Table 1. Daily returns and variance during overnight and daytime.

The table summarizes the empirical distribution of daily overnight and daytime returns (Panel A) and the volatility measured as the squared return (Panel B) or realized variance (Panel C). The sample period is divided into a pre-Lehman (in-sample) period from November 12, 1997 to August 29, 2008, and a post-Lehman (out-of-sample) period from September 2, 2008 to September 30, 2011. Returns are in percentages. The table reports sample correlations between the overnight return on day *t* and day *t*-1, between the daytime return on day *t* and day *t*-1, and between the overnight return on day *t* and the daytime return on day *t*. Q_{20}(AC) is the Ljung-Box statistic for the null of no autocorrelation up to 20 lags. *, ** and *** indicate significance at the 10%, 5% or 1% levels. Mean (hourly) reports the mean squared return divided by the total hours spanned by each segment of the day. The daytime realized variance (RV) summarized in Panel C is computed as the sum of 5-minute squared returns from market open at 9:30am to market close at 4:00pm. *d* is the fractional integration parameter estimated via the Gaussian semi-parametric approach proposed by Robinson (1995).

	In-sample: Nov 1997 - Aug 2008				Out-of-sample: Sep 2008 - Sep 2011			
	Russell 2000		S&P 500		Russell 2000		S&P 500	
Summary statistics	$r_{o,t}$ night	$r_{d,t}$ day	$r_{o,t}$ night	$r_{d,t}$ day	$r_{o,t}$ night	$r_{d,t}$ day	$r_{o,t}$ night	$r_{d,t}$ day
				Panel A: returns				
Mean	0.020	0.000	0.007	0.005	-0.004	-0.014	-0.012	-0.004
Median	0.005	0.040	0.003	0.047	0.007	0.138	0.000	0.095
StDev	0.234	1.271	0.190	1.092	0.572	2.136	0.249	1.780
Skewness	-0.525	-0.074	-0.249	-0.031	-0.283	-0.358	-0.254	-0.362
Kurtosis	19.80	4.026	14.99	5.600	6.930	6.015	7.774	8.586
Correlation structure:								
$r_{o,t-1}$	-0.089 ***	-0.028	-0.092 ***	0.001	-0.149 ***	-0.064 *	-0.118 ***	-0.026
$r_{d,t}$	0.162 ***		0.202 ***		0.273 ***		0.390 ***	
$r_{d,t-1}$	0.041 **	0.015	-0.011	-0.038 **	-0.040	-0.085 **	-0.048	-0.122
Q_{20} (AC)	60.60	38.64	60.78	35.21	36.62	28.07	37.12	52.53
				Panel B: squared returns				
Mean (hourly)	0.003	0.248	0.002	0.183	0.019	0.701	0.004	0.487
Median	0.006	0.597	0.004	0.361	0.096	1.120	0.011	0.457
StDev	0.237	2.810	0.135	2.556	0.797	10.223	0.162	8.724
Skewness	14.163	4.661	11.578	7.089	8.880	5.307	6.353	5.905
Kurtosis	254.6	39.42	196.5	89.15	117.3	43.48	57.41	46.17
Correlation structure:								
$(r_{o,t-1})^2$	0.067 ***	0.175 ***	0.100 ***	0.077 ***	0.235 ***	0.199 ***	0.144 ***	0.218
$(r_{i,t})^2$	0.152 ***		0.130 ***		0.197 ***		0.366 ***	
$(r_{i,t-1})^2$	0.087 ***	0.132 ***	0.049 **	0.158 ***	0.103 ***	0.304 ***	0.170 ***	0.192 ***
Q_{20} (AC)	156.4	1697	247.9	1058	186.8	1151	441.3	1107
				Panel C: realized variance (RV)				
Mean		1.091		1.189		3.735		3.295
Median		0.658		0.736		1.590		1.109
StDev		1.372		1.512		6.211		6.389
Skewness		4.368		4.584		4.208		4.359
Kurtosis		34.78		37.35		26.17		26.75
d		0.476		0.492		0.493		0.495

Table 2. Rejection rates of unconditional Equal Predictive Ability tests.

This table summarizes the unconditional Equal Predictive Ability tests deployed sequentially over J=279 rolling windows of $\tilde{n}$=500 out-of-sample forecasts (Panel A) and J=10 non-overlapping windows of $\tilde{n}$=78 out-of-sample forecasts (Panel B). The first forecast is for September 2, 2008 and the last forecast for September 30, 2011. The proposed bivariate model is confronted with three competitors: two simpler bivariate models without (w/o) covariance – the nested bivariate model and a bivariate model for redefined overnight and daytime returns by moving the open price 15 minutes into the trading day – and the univariate `bundling' model based the close-to-close return alongside the overnight-adjusted RV^{HL} measure. The null hypothesis is that the proposed bivariate model is superior to a given competitor. The figures reported in Panel A and Panel B are rejection rates over the 279 rolling windows or the 10 non-overlapping windows, respectively; for instance, the figure 0.000 means that the null hypothesis is never rejected. Panels A2 and B2 report the rejection rates for the first 10% of the J windows of out-of-sample forecasts, which represents the height of the recent financial crisis. Panels A3 and B3 report the rejection rates for the remaining 90% of the J windows.

	Bivariate w/o covariance	Bivariate 15 min further	Bundling RV^{HL}	Bivariate w/o covariance	Bivariate 15 min further	Bundling RV^{HL}
	Russell 2000			S&P 500		
Panel A: Dynamic rolling evaluation (J=279 windows of $\tilde{n}$ =500 days)						
Panel A1: All J windows						
5% VaR	0.000	0.000	0.000	0.000	0.247	0.000
1% VaR	0.000	0.000	0.000	0.000	0.104	0.000
Panel A2: 1/10J windows						
5% VaR	0.000	0.000	0.000	0.000	1.000	0.000
1% VaR	0.000	0.000	0.000	0.000	1.000	0.000
Panel A3: 9/10J windows						
5% VaR	0.000	0.000	0.000	0.000	0.163	0.000
1% VaR	0.000	0.000	0.000	0.000	0.004	0.000
Panel B: Dynamic non-overlapping evaluation (J=10 windows of $\tilde{n}$ =78 days)						
Panel B1: All J windows						
5% VaR	0.000	0.000	0.000	0.100	0.100	0.000
1% VaR	0.100	0.000	0.000	0.200	0.300	0.200
Panel B2: 1/10J windows						
5% VaR	0.000	0.000	0.000	0.000	1.000	0.000
1% VaR	0.000	0.000	0.000	0.000	1.000	0.000
Panel B3: 9/10J windows						
5% VaR	0.000	0.000	0.000	0.111	0.000	0.000
1% VaR	0.111	0.000	0.000	0.222	0.222	0.222

Table 3. Rejection rates of Correct Conditional Coverage tests.

This table summarizes the dynamic quantile (DQ) and probit tests for the null hypothesis of Correct Conditional Coverage. The tests are deployed sequentially over J=279 rolling windows of $\tilde{n}$=500 out-of-sample forecasts (Panel A) and J=10 non-overlapping windows of $\tilde{n}$=78 out-of-sample forecasts (Panel B). The first forecast is for September 2, 2008 and the last forecast for September 30, 2011. The table reports rejection rates; for instance, the figure 1.000 means that the Correct Conditional Coverage criterion is rejected over all the sequential windows. Panels A1 and B1 report the rejection rates over all J out-of-sample windows. Panels A2 and B2 report the rejection rates over the first 10% of the J windows of out-of-sample forecasts, which represents the height of the recent financial crisis. Panels A3 and B3 report the rejection rates for the remaining 90% of the J windows. For each test the lowest rejection rate achieved across the four models is shaded. Asterisks in Panel A indicate significance of a bootstrap difference-in-proportions (DIP) test statistic where the null hypothesis is that the rejection rate of the proposed bivariate model is at least as small as that of a competitor model. *** denotes significance at the 1% level.

	Bivariate	Bivariate w/o covariance	Bivariate 15min further	Bundling RV^{HL}	Bivariate	Bivariate w/o covariance	Bivariate 15min further	Bundling RV^{HL}
	Russell 2000				S&P 500			
Panel A: Dynamic rolling evaluation (J =279 windows of $\tilde{n}$ =500 days)								
Panel A1: All J windows								
DQ test (5% VaR)	0.222	0.387	1.000	0.462	0.237	0.290	0.211	0.315
DQ test (1% VaR)	0.294	0.484	1.000	0.742	0.251	0.505	1.000	0.305
Probit test (5% VaR)	0.151	0.323	0.964	0.487	0.566	0.559	0.577	0.387***
Probit test (1% VaR)	0.068	0.373	1.000	0.577	0.835	1.000	1.000	0.878
Panel A2: 1/10J windows								
DQ test (5% VaR)	1.000	1.000	1.000	1.000	0.964	1.000	1.000	1.000
DQ test (1% VaR)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Probit test (5% VaR)	0.964	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Probit test (1% VaR)	0.679	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel A3: 9/10J windows								
DQ test (5% VaR)	0.135	0.319	1.000	0.402	0.155	0.211	0.124	0.239
DQ test (1% VaR)	0.215	0.426	1.000	0.713	0.167	0.450	1.000	0.227
Probit test (5% VaR)	0.060	0.247	0.960	0.430	0.518	0.510	0.530	0.319***
Probit test (1% VaR)	0.000	0.303	1.000	0.530	0.817	1.000	1.000	0.865
Panel B: Dynamic non-overlapping evaluation (J =10 windows of $\tilde{n}$ =78 days)								
Panel B1: All J windows								
DQ test (5% VaR)	0.200	0.200	0.300	0.300	0.200	0.300	0.200	0.300
DQ test (1% VaR)	0.300	0.300	0.700	0.400	0.300	0.300	0.300	0.400
Probit test (5% VaR)	0.200	0.200	0.200	0.200	0.300	0.400	0.400	0.400
Probit test (1% VaR)	0.700	0.600	0.600	0.500	0.700	0.800	0.600	0.700
Panel B2: 1/10J windows								
DQ test (5% VaR)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DQ test (1% VaR)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Probit test (5% VaR)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Probit test (1% VaR)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel B3: 9/10J windows								
DQ test (5% VaR)	0.111	0.111	0.222	0.222	0.111	0.222	0.111	0.222
DQ test (1% VaR)	0.222	0.222	0.667	0.333	0.222	0.222	0.222	0.333
Probit test (5% VaR)	0.111	0.111	0.111	0.111	0.222	0.333	0.333	0.333
Probit test (1% VaR)	0.667	0.556	0.556	0.444	0.667	0.778	0.556	0.667

Table 4. Weighted rejection rates of Correct Conditional Coverage tests.

This table summarizes weighted rejection rates of the dynamic quantile (DQ) and probit tests for the null hypothesis of Correct Conditional Coverage, Eq. (18), deployed sequentially over J=279 rolling windows of $\tilde{n}$=500 out-of-sample forecasts (Panel A) and J=10 non-overlapping windows of $\tilde{n}$=78 out-of-sample forecasts (Panel B). The out-of-sample period commences on September 2, 2008. Panels A1 and B1 report weighted rejection rates where the rejections are weighted by the absolute distance between the empirical conditional coverage probability and the nominal coverage, $|\hat{\alpha}-\alpha|$, appropriately standardized so that the weights add to unity. Panels A2 and B2 weight more heavily (by absolute distance) rejections for which $\hat{\alpha}$ is above α than rejections for which $\hat{\alpha}$ is below α (squared distance). Panels A3 and B3 weight more heavily (by absolute distance) rejections for which $\hat{\alpha}$ is below α than rejections for which $\hat{\alpha}$ is above α (squared distance). For each test the lowest rejection rate achieved across models is shaded. Asterisks in Panel A indicate significance of a bootstrap difference-in-proportions (DIP) test statistic where the null hypothesis is that the rejection rate of the proposed bivariate model is at least as small as that of a competitor model. ** and * denote significance at the 5% and 10% levels, respectively.

	Bivariate	Bivariate w/o covariance	Bivariate 15min further	Bundling RV^{HL}	Bivariate	Bivariate w/o covariance	Bivariate 15min further	Bundling RV^{HL}
	Russell 2000				S&P 500			
Panel A: Dynamic rolling evaluation (J =279 windows of $\tilde{n}$ =500 days)								
Panel A1: Weighted by Absolute Coverage Error								
DQ test (5% VaR)	0.191	0.245	1.000	0.803	0.381	0.442	0.181*	0.529
DQ test (1% VaR)	0.242	0.421	1.000	0.927	0.171	0.482	1.000	0.432
Probit test (5% VaR)	0.085	0.217	0.969	0.807	0.726	0.804	0.635*	0.628**
Probit test (1% VaR)	0.021	0.382	1.000	0.855	0.813	1.000	1.000	0.941
Panel A2: Weighted Asymmetrically Underprediction > Overprediction								
DQ test (5% VaR)	0.050	0.179	1.000	0.902	0.329	0.357	0.177*	0.457
DQ test (1% VaR)	0.243	0.418	1.000	0.948	0.143	0.460	1.000	0.375
Probit test (5% VaR)	0.045	0.168	0.969	0.919	0.939	0.810	0.633**	0.705*
Probit test (1% VaR)	0.021	0.390	1.000	0.873	0.807	1.000	1.000	0.932
Panel A3: Weighted Asymmetrically Overprediction > Underprediction								
DQ test (5% VaR)	0.318	0.736	1.000	0.630	0.447	0.736	0.316	0.757
DQ test (1% VaR)	0.184	0.489	1.000	0.580	0.749	0.831	1.000	0.765
Probit test (5% VaR)	0.120	0.582	0.976	0.607	0.482	0.795	0.783	0.411
Probit test (1% VaR)	0.003	0.141	1.000	0.568	0.949	1.000	1.000	0.997
Panel B: Dynamic non-overlapping evaluation (J =10 windows of $\tilde{n}$ =78 days)								
Panel B1: Weighted by Absolute Coverage Error								
DQ test (5% VaR)	0.232	0.316	0.489	0.543	0.261	0.579	0.268	0.446
DQ test (1% VaR)	0.133	0.118	0.791	0.482	0.159	0.168	0.146	0.536
Probit test (5% VaR)	0.232	0.316	0.463	0.531	0.727	0.595	0.653	0.461
Probit test (1% VaR)	0.562	0.537	0.451	0.485	0.618	0.632	0.154	0.987
Panel B2: Weighted Asymmetrically Underprediction > Overprediction								
DQ test (5% VaR)	0.235	0.322	0.489	0.545	0.268	0.589	0.268	0.450
DQ test (1% VaR)	0.131	0.120	0.798	0.489	0.161	0.171	0.147	0.547
Probit test (5% VaR)	0.235	0.322	0.463	0.545	0.771	0.590	0.653	0.451
Probit test (1% VaR)	0.557	0.537	0.446	0.489	0.618	0.628	0.147	1.000
Panel B3: Weighted Asymmetrically Overprediction > Underprediction								
DQ test (5% VaR)	0.134	0.253	0.662	0.676	0.202	0.577	0.228	0.348
DQ test (1% VaR)	0.028	0.018	0.747	0.443	0.060	0.060	0.048	0.447
Probit test (5% VaR)	0.134	0.253	0.658	0.653	0.834	0.601	0.801	0.370
Probit test (1% VaR)	0.518	0.479	0.463	0.447	0.613	0.631	0.057	0.984