Finance Research Letters 000 (2017) 1–11

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

The effect of non-trading days on volatility forecasts in equity markets

Štefan Lyócsa[a,b,1], Peter Molnár[c,*]

[a] University of Economics in Bratislava, Faculty of National Economy, Bratislava, Slovakia
[b] Masaryk University, Faculty of Economics and Administration, Brno, Czech Republic
[c] University of Stavanger, UiS Business School, Stavanger, Norway

ARTICLE INFO

Article history:
Received 7 January 2017
Accepted 5 July 2017
Available online xxx

JEL classification:
C53
Q02
G17

Keywords:
Realized volatility
Volatility forecasting
Non-trading days

ABSTRACT

Weekends and holidays lead to gaps in daily financial data. Standard models ignore these irregularities. Because this issue is particularly important for persistent time series, we focus on volatility modelling, specifically modelling of realized volatility. We suggest a simple way of adjusting volatility models, which we illustrate on an AR(1) model and the HAR model of Corsi (2009). We investigate daily series of realized volatilities for 21 equity indices around the world, covering more than 15 years, and we find that our extension improves the volatility models—both in sample and out of sample. For HAR models and for consecutive trading days, the mean squared error decreased by 2.34% in average and for the QLIKE loss function by 1.41%.

1. Introduction

Daily time series are probably the most commonly used data in empirical finance. The main advantage of these data is that they usually represent the highest frequency freely available for various datasets. In addition, for many purposes, daily time series can be considered regularly spaced data, and standard time series techniques can be applied to them. However, most financial data exist only for trading days. As we show in this paper, considering this trading gap can significantly improve the performance of models in some cases.

Many time series data exhibit autoregressive properties, meaning that a variable's value today is related to its value tomorrow. If we can more precisely capture this relationship, our models will be more precise. However, in case of daily financial series, data usually exist for weekdays but not for weekends or holidays. Weekends can be considered a break in the data or days with missing data. In either case, the dependence between two consecutive trading days, for example, Wednesday and Thursday, can be intuitively expected to be stronger than the dependence between two trading days separated by a weekend or holiday, for example, Monday and Friday. If we ignore these differences and assume the same dependence throughout the week, we will likely underestimate the dependence between consecutive weekdays and overestimate the de-

* Corresponding author.
E-mail addresses: stefan.lyocsa@gmail.com (Š. Lyócsa), peto.molnar@gmail.com (P. Molnár).
[1] Lyócsa appreciates the support by the Slovak Research and Development Agency under contract No. APVV-14-0357 and by the Slovak Grant Agency under Grant No. 1/0406/17 and Grant No. 1/0257/18.

http://dx.doi.org/10.1016/j.frl.2017.07.002
1544-6123/© 2017 Published by Elsevier Inc.

pendence between Friday and Monday. We suggest that allowing the autoregressive coefficient to be dependent on whether a weekend separates the two observations sufficiently addresses this issue.

Obviously, this effect does not matter when very little dependence is reflected in the data. In the literature, volatility is understood to have a long memory (Bollerslev and Mikkelsen, 1996). In the empirical part of our paper, we thus focus on volatility. Volatility is not directly observable, but the concept of realized volatility (RV) calculated from high-frequency data (Andersen and Bollerslev, 1998) makes it observable for practical purposes. Models based on RV have been applied to stock markets (e.g., Christoffersen et al., 2010; Bugge et al., 2016), exchange rates (e.g., Andersen et al., 2001; Lyócsa et al., 2016), and commodities (Haugom et al., 2014; Birkelund et al., 2015; Lyócsa and Molnár, 2016). We thus focus on RV.

Long memory and the persistence of market volatility have led previous research to use fractionally integrated time series models (e.g., Bollerslev and Mikkelsen, 1996). However, the most popular model is the heterogeneous autoregressive (HAR) model of Corsi (2009), which captures the long-memory property as accurately as competing models and is easy to implement. We suggest a simple extension of this model, the NT-HAR model (non-trading days HAR), which allows the autoregressive coefficient to be dependent on whether a non-trading period occurs between two observations.

The data that we use in the empirical evaluation of our model are RV series for 21 equity indices around the globe, including equity indices for Brazil, Canada, China, France, Germany, Greece, India, Italy, Japan, Mexico, Singapore, South Korea, Spain, Switzerland, and the U.K.; three indices for U.S.; and one index for the Eurozone. We find that, in most cases, the NT-HAR model outperforms the benchmark HAR model, both in sample and out of sample.

The rest of the paper is organized as follows. Section 2 intuitively explains our model using a simulation example. Section 3 describes the data and the methodology; Section 4 presents the results; and Section 5 concludes.

2. Illustrative example

Our motivation can be intuitively explained as follows: given the measure of daily market volatility $RV_{i,t}$, the lagged market volatility $RV_{i,t-1}$ does not always precede the predicted volatility at time t by one calendar day. On average, in more than one-fifth of the cases (18.56%), the calendar-day difference between two consecutive days on equity markets is equal to or more than 3 (e.g., see column "*nC*" in Table 2). One consequence of non-equidistant observations could be that the lagged market volatility has a different information value with respect to subsequent volatility based on whether subsequent volatility appears on the following calendar day or, say, after the third calendar day, i.e., after the weekend. Naturally, this consequence should influence the autoregressive terms of predictive regressions. Let $\Delta(t, t-1)$ denote the calendar-day difference between two consecutive trading days. Intuitively, compared with the case when $\Delta(t, t-1) = 1$, the t–1 volatility should have a lower influence on the volatility in t if $\Delta(t, t-1) > 1$ because additional calendar days might introduce additional trading information and noise into the market, which is not observed.

To illustrate our motivation, we have simulated samples of 1000 daily volatilities following the volatility process of the specification introduced by Alizadeh et al. (2002):

$$\ln\sigma_t = \ln\bar{\sigma} + \rho(\ln\sigma_{t-1} - \ln\bar{\sigma}) + \mu e_{t-1} \tag{1}$$

where $\ln\bar{\sigma} = 2$, $\mu = 0.048$, and e_t follows the standard normal distribution. To eliminate the effect of the initialization of the simulation on the resulting time series, we removed the first 200 observations. How to study the effect of the weekend is unclear, as one might consider the weekend to represent one or two missing data points. From the remaining 800 observations, we have thus considered the removal of either (i) each sixth observation or (ii) each sixth and seventh observation to represent the non-trading days of Saturday and Sunday.

The second data-generating process was a simple first-order autoregressive model, AR(1):

$$\sigma_t = \alpha + \rho\sigma_{t-1} + \mu e_{t-1} \tag{2}$$

where α is set to 0.5 and, as before, e_t follows the standard normal distribution. Next, we fit two models:

$$\sigma_{i,t} = \beta_0 + \beta_1\sigma_{i,t-1} + e_{i,t} \tag{3}$$

$$\sigma_{i,t} = \beta_0 + \beta_1\sigma_{i,t-1} + \gamma_0 \times I[\Delta(t, t-1) \geq 3] + \gamma_1\sigma_{i,t-1} \times I[\Delta(t, t-1) \geq 3] + e_{i,t} \tag{4}$$

The first model (3) is a simple autoregressive model that does not take into account that we remove each sixth and seventh observation. The second model uses a simple dummy to account for the predicted volatility being observed after more than two calendar days. We denote this model as AR-NT, where NT refers to non-trading days. We let the size of the autoregressive coefficient ρ vary from 0.01 up to 0.99 by increments of 0.01. For each value of ρ, we have performed 100 iterations. We are interested in how accurately the ρ parameter is estimated via the estimates of the β_1 coefficients in specifications (3) and (4). In Figs. 1 and 2, we plot the difference between the true value ρ and the average value of the estimated autoregressive coefficient for the first and second data-generating processes. On the one hand, the standard *AR* model clearly leads to biased coefficient estimates, as the coefficient is underestimated. On the other hand, the *AR-NT* model leads to coefficients with negligible bias.

When the time series is not persistent, non-trading does not matter, as no information spills over from one trading day to another. When the time series is highly persistent, the non-trading day effect also diminishes, as even one or two missing

Please cite this article as: Š. Lyócsa, P. Molnár, The effect of non-trading days on volatility forecasts in equity markets, Finance Research Letters (2017), http://dx.doi.org/10.1016/j.frl.2017.07.002