Bayesian Analysis (2014) **9**, Number 4, pp. 809–818

Comment on Article by Windle and Carvalho

Enrique ter Horst * and German Molina †

Abstract. The article by Windle and Carvalho introduces a fast update procedure for covariance matrices through the introduction of higher frequency sources of information for the underlying process, demonstrated with a financial application. This discussion focuses on outlining the assumptions and constraints around their use in financial applications, as well as an elicitation of some key choices made for comparison with traditional benchmarks, that may ultimately affect the results.

Keywords: Stochastic Volatility, Financial application, EWMA, Covariance update

1 Introduction

We congratulate Windle and Carvalho for a very flexible and empirically sound contribution, where the authors build a bridge between traditionally used practical approaches, and the more theoretically sound models. Their approach, termed (UE) in the paper and this discussion, is a practical solution for updating the covariance matrix through an additional source of information, namely higher frequency data (Kyj 2008), usually available for the most liquid assets, while maintaining a tractable exponentially-weighted approach for the update, and keeping a coherent, scalable statistical model underneath. As such, it builds a link between traditional purely statistical models, relatively popular within the sell side arena in finance, and more ad hoc, "quick-and-dirty" approaches, namely the Exponentially Weighted Moving Averages (EWMA) updates of covariances, that are still relatively widespread (Shephard 2005), mainly within the buy side arena in finance (Fabozzi 2008). Since the authors propose this solution from the point of view of a financial application, we will focus our discussion on the pros and cons of its use for financial applications in general, as well as their particular application. There will be many other cases where their approach is a natural fit, but some aspects in the use for financial assets should be stressed, either as potential additions to future research, or as topics to consider prior to using their approach.

2 Constraints for financial applications

The problem addressed is the fast and efficient update of (potentially high-dimensional) unconstrained covariance matrices with new observations, while preserving a statistically sound model underneath, incorporating information available at a different granularity, and producing the best possible forecasts.

*Professor of Finance, Centro de Finanzas, CESA, Bogotá, Colombia, and IESA, Caracas, Venezuela enrique.terhorst@cesa.edu.co

†Head of Quantitative Trading, Idalion Capital Group german@germanmolina.com

 DOI:10.1214/14-BA917

The authors propose an approach that relies on the use of data collected at a higher frequency (HF) than that of the underlying model, for the updates at the lower frequency (LF). The approach proposed, together with the HF information, brings several constraints, such as (1) data availability at HF for all the assets; (2) sufficient frequency of updates to obtain reliable estimates; (3) full coverage of the LF range through HF data, equally-spaced data; (4) synchronizable data trading on similar timezones; (5) HF data coming from an approximately locally constant, common covariance matrix; (6) researcher not interested in analysis of the higher frequencies; (7) non-stationary forecasts; (8) assets not experiencing widely known features like leverage; and (9) data sufficiently clean from large sources of noise at snapshot times, which provides good estimates of the LF relationships.

In practice, there may be several reasons why these assumptions will be violated for financial applications, and it would be relevant to cover in this discussion the potential costs of the use of the (UE) approach, not losing perspective of the benefits it may bring, which are well covered in the paper.

2.1 Availability of HF data

The benefits of their approach are centered around applications where HF data is readily available, yet not of immediate interest beyond use as inputs. This may be a burden for the more illiquid assets, for those that are not tradeable on exchanges (for example real estate prices, hedge fund returns, or economic data, among others) or simply for those where the researcher is interested in the highest frequency available for forecasting.

2.2 Sufficiently large number of HF updates

The authors do not cover in sufficient depth the minimum requirements for k for this model to be advantageous, if any, nor the effects of different values of k on the relative performance, absolute and against the benchmark. What is a reasonable k? How sensitive are the results, in absolute and relative terms, under different values of k? What is the model relative performance when $k = 1$? We anticipate that the realized covariance will rely heavily on the abundance and quality of HF data, but how much? For high-volume trading data this is not an issue. However, many other financial assets/data will not always be available at any preferred frequency.

2.3 Full asset variability versus high frequency data

The authors seem to sidestep the issue of availability of HF data only during market trading hours in their application, and most applications outside multi-exchange or over-the-counter 24-hour markets cases. In Appendix C, they cover the case of realized covariance matrices for assets that trade continuously over the full LF period. However, they only use intraday data for the assets in the DJIA, which, for the case of the NYSE, only trade from 09:30 until 16:00 EST. We will cover this discrepancy in more detail in

Section 3.

2.4 Synchronized data availability across assets

Although this is an issue suffered by both the (UE) model and the Factor Stochastic Volatility (FSV) benchmark, this approach will suffer in scenarios where the data is not available for all assets at nearby timepoints (for example, global portfolios that include assets traded in non-synchronous exchanges or holiday periods that differ between countries). Additionally, events like limit ups/downs of particular assets, where they may not trade for the remainder of the trading session, may hinder their approach, although the use of the FSV in those cases may also be questionable, unless a restart of the session happens intraday.

2.5 Common, locally constant, covariance matrix

This assumption is usually based on discretization of data at the highest frequency available, at which, due to lack of further information, a constant covariance matrix must be assumed. However, if data is available at a higher frequency, maintaining the assumption at that HF is an artifact not driven by either the underlying continuous time model, nor by the application chosen, and deserves further justification. However, we believe that this assumption could be relaxed by allowing market microstructure information in the distributional assumptions of the $r_{t,i}$ common (or from a common distribution) across time periods, while maintaining X_t as the lower frequency covariance (defining it instead as average covariance between $t-1$ and t, and allowing $r_{t,i} \sim N(0, f(i)X_t)$ rather than being constant and common for all i within a given t in $r_{t,i}$). The information content of each HF observation being variable (Berg et al. 2004) would be a natural addition to fit well-known features (Ait-Sahalia et al. 2012) of market microstructures (not pure "noise", as characterised by the authors in Appendix D). The impact of the authors' assumption of iid HF returns on the results is unclear, since there is no direct comparison in the paper with an approach that does not rely on this assumption. For example, even if ignoring market microstructures, the authors could have used the HF data directly with the FSV on that frequency, and then forecast k steps ahead until covering the one-step ahead, LF period of interest. Is the model better because it uses the HF return data for the LF covariance predictions, or would the FSV defined at the higher frequency and operated with HF returns data have performed better? For example, could a 30-minute data FSV approach with k =10-step ahead (cumulative) forecast be a better benchmark?

2.6 The frequency of interest

As the authors point out, the approach finds its space in LF applications. However, the practitioner may find that using the data for modelling of the LF overshadows the benefits of a quicker update of the covariance matrix. In fact, users may find the quick update of the covariance a benefit especially useful in the HF, where an "online" decision-making is more likely to occur. Otherwise, the most tangible benefit comes mainly from the quality of the forecasts.

2.7 Stationarity

The paper defines outperformance exclusively based on 1-step ahead predictions. This is relatively simplistic, and does not offer a full picture of the forecasting quality, especially if suffering from non-stationarity. It would have been interesting to see how the model behaves against the FSV and other alternatives for longer-term forecasts.

2.8 The impact of the leverage effect

The leverage effect, defined as the negative correlation between the volatility of the particular asset and its returns, has been well documented in Bollerslev et al. (2006); Nakajima (2012), and is a feature that has a strong influence on the correct estimation of covariance matrices (see Bollerslev et al. (2006) or Ishihara et al. (2012) for a recent multivariate example). The particular combination of HF data and stock prices may be especially sensitive to leverage effects (Kalnina and Xiu 2013), and their impact on this application is unexplored. Other asset classes, like foreign exchange, may be less affected, where the asset in itself does not necessarily have to reflect a net beta to the market, and, as such, suffer from penalization for risk increases (Omori et al. 2007; Xu and Li 2010). Would the addition of leverage be feasible in the (UE) approach? Note that the leverage should, therefore, appear at both frequencies used in the paper, as it is a feature of the returns at any frequency. Some challenges in the use of leverage in the Bayesian framework can be found in Yu (2005). Should the leverage be the same across frequencies?

2.9 Data releases: News, jumps and asset price adjustments

There are many reasons, deterministic and stochastic, why the intraday volatility during the trading period may not be constant (Ait-Sahalia et al. 2012, 2010). For example, trading volumes are consistently higher during the early and later parts of the trading session, which has been found to be a factor in explaining intraday volatility differences (Andersen and Bollerslev 1997). Also drivers of volatility, like news releases, will happen often at pre-specified times of the day.

During (and outside of) the trading session there are many factors that may induce jumps in the underlying price process, including market or asset-specific news (Darrat et al. 2007), as well as jumps in the initial observation of the day (Ait-Sahalia et al. 2010). Jumps can come from different sources (Eraker et al. 2003; Tankov and Cont 2003). The proposed use of HF data may not take into account the potential for jumps, as it assumes normality of the returns $r_{t,i}$. Those jumps are more relevant at higher frequencies, where the ratio of jump size to expected volatility becomes the largest. What is the expected impact on the estimates for different choices of k? If r_t may show fat tails, should you not expect/model even fatter tails for $r_{t,i}$? Here we discuss four sources of jumps:

- Trading session (re)open: Even if assets trade on a 24-hour market, the Monday after a holiday reopen will induce a jump from the previous closing level. In the stock example that the authors use, there would be jumps in the HF data on the reopen of the NYSE, which will encompass more than 17 hours of new information (Ait-Sahalia et al. 2012). These jumps are stochastic and usually not predictable a priori outside of the "beta" of the assets to the index via futures markets. The authors appear to propose in their application (unlike the Appendix) first to ignore the jump from close-to-open by dismissing that information, and, additionally, smooth the data for those jumps (jitter end points) to obtain better estimates. Additionally, there are cases where the markets close intraday (e.g. commodity futures market with electronic and pit sessions), which would induce multiple opens/closes per day if all the available intraday data was to be used.
- Dividend distribution: Dividend-producing assets, like the subset chosen by the authors, will produce predictable jumps on the ex-dividend date. How would the authors propose to approach the ex-dividend HF returns assuming all HF data is used? We understand that the natural choice would be to assign the dividend jump adjustment to the period where the market is closed.
- Sectorial news or asset-specific news: Assets may experience stochastic jumps coming from news that affect them or their sector, as documented in Darrat et al. (2003). These jumps involve random shocks of information, and can be non-quantitative (e.g. CEO renouncing, lawsuit on the company, and other non-quantifiable events). However some are driven by the fundamentals of the company and can be accounted for, like a news release of the Earnings per Share or Dividend announcements. This type of information affects the realized covariance matrix through intraday jumps, but these jumps can themselves be modelled through the introduction of a larger set of information that contains shocks in expectations versus released data. Would this type of addition be viable in the (UE) approach?
- Absolute and relative liquidity: Illiquid assets do not trade over long periods of time. During those, the price may be adjusting, but there could be also unrealistic spreads, or either bid/ask may not be available, nor an actual trade. A jump in the intraday return would be visible only when there is a new trade or bid/ask update. The impact of this, larger if using HF data, does not seem to have been covered in sufficient depth in Appendix D.

3 Discussion on the application

3.1 Data and Adjustments

We outline here our understanding of the application used by the authors, since some of the points were unclear. We assume that the HF data just covers the open-to-close

NYSE trading session (09:30 until 16:00 EST), rather than the full day, since the authors mention that "r_t is the vector of open to close log-returns on day t", and talk about timestamps of trades in this particular exchange. As such, it appears that $r_t \neq p_t - p_{t-1}$ and $k \neq \frac{1}{\Delta_t}$.

Jitter end adjustments

Are the jitter end adjustments done on the boundary initial and final pairs of prices (of the total of k in the trading session)? For example, if price snapshots were to be taken every 30 minutes, is the first price snapshot relabelled to be the average between the 9:30am price (open level) and the 10:00am price (first intraday snapshot)? We understand that the averaging must be over something else, since otherwise, wouldn't the first (and last) returns constructed this way be defined over a different (larger) time scale, with a larger volatility than the others ($r_1, r_k \sim N(0, c\frac{X_t}{k})$), with $c \neq 1$? If this was the case, do you adjust for this in Section 5 for the calculation of Y_t, as all returns $r_{t,i}$ are assumed to be iid? Also how would you propose to adjust this if you were to include the return from close-to-open as an additional source of intraday returns?

Choice of sample period

The authors utilize data from stocks spanning the period from Feb 27th 2007 until Oct 29th, 2010. On Feb 27th, 2007, first day of the sample, the DJIA experienced the largest return in absolute value since March 24th, 2003, with a -3.3% move open-to-close, as well as the largest over the following 12 months. The choice, therefore, does not seem to be made at random. This starting point of the sample provides a better representation of the volatility prevailing in the market during the out-of-sample period, both during and post financial crisis. The initial variance of the 50-day open-to-close (index) return is 67% higher by choosing this starting date versus a single day later. It would be interesting to get the perspective of the authors as to the reasoning for this choice, since 2007 marks the bottom of the market volatility, and, as such, this starting point is not representative of the market behavior in surrounding times, but highly representative of the supposedly unknown out-of-sample period. This choice is also relevant for the calibration of λ, as detailed in Figure 3, and, therefore, may have long-lasting effects in their results.

Choice of assets

Their sample consists of only highly liquid assets that trade over the same trading hours. It is the best possible choice for this approach to perform well, but not necessarily representative of the asset universe where these models are used. It would be interesting to explore how the quality of the estimates decays as the frequency of HF updates and liquidity of the assets diminishes, and the HF data becomes noisier or less representative of the LF structure.

Choice of the sampling frequency

An element that would be of interest is the reliance of the results on the quality and availability of such data. The authors utilize $k = 10$ returns per trading day for the construction of each covariance matrix. Would the results improve/worsen for different sampling frequencies? Is there a sampling frequency at which the proposed approach fails to be superior to the FSV? A good understanding of this threshold would help the practitioner decide which approach is more likely to provide better results, since k is not always a parameter that the researcher can choose.

Choice of snapshot times

The choice of snapshot times can have a large influence on the results. If they coincide with points of release of news, the quality of the adjustments may vary significantly. For example, if the data was to be sampled every 30 minutes (the sampling frequency in the paper was unclear, since it depends on the jitter point methodology questions above), it would coincide, for example, with the ISM report release (10:00am once a month), which could bring non-synchronous jumps at that snapshot time. Any excess noise due to (potentially uneven at first) jumps could have a larger impact on HF data, sampled at that snapshot point, than on LF data, and reduce the quality of their approach.

3.2 Impact of jittering on the model assumptions

The effect of the jitter end point adjustment on the outcome of the models is unclear. Will this adjustment provide an additional advantage to the (UE) over the FSV/FSVE? For example, could it be the case that random, very non-representative, market open levels induced smaller estimates of the correlations, which could hurt the FSV (not the FSVE) the most for forecasting, since the (UE) would have several other intraday points to "wash out" this effect.

3.3 Full information set versus HF data

Single stock price updates are only available intraday for 6.5 hours for the NYSE. Are the authors ignoring all the information contained in more than 17 hours/day plus weekends of actual price movements/comovements of the assets, latent until the reopen? For example, Alcoa, used in Figure 1, has a ratio larger than 1 of close-to-open variance to close-to-close variance.

If we assume that, under the model proposed, $r_t = \sum_{i=1}^{k} r_{t,i}$, and r_t is the total return for the asset between period $t-1$ and period t, as described in Appendix C (needed for the discretization of geometric Brownian motion), then, (1) Under the LF underlying model, the volatility during any subperiods of equal length between $t-1$ and t is constant and nonzero (as discussed, an unrealistic assumption in itself), but (2) under the HF part of the model (Wishart distribution for $r_{t,i}$), the assumption seems to be zero contribution to the volatility of the periods outside of trading hours, to make this coherent with the assumptions in Appendix C. What is X_t in section 5? A representation of the variability for the (co)movements over 6.5 hours or over 24 hours? If it's the former, how does this tie up with the discretization of the geometric Brownian motion? If it's the latter, how does it tie up with the definition of Y_t in section 5? In any case, how do you account for the variability during non-trading hours?

This issue is induced by the use of partial intraday data in (UE), which is by no means necessary for the FSV approaches, if they were to use close-to-close low-frequency available information. As such, the FSVE may not be a fair benchmark. Would HF returns coming at non-regular intervals or not available over large periods (close-to-open) of the trading period t be an issue for the (UE) approach, even if still assuming a constant volatility over that period? While the authors attempt to incorporate new information available intraday, they may be sacrificing a very large portion of information available as well. We understand that the assumption for the example of equally-spaced intraday data may not be necessary, but would ask the authors to provide more detail about any concerns in this regard with incorporation of close-to-open market information in their example and in their methodology.

3.4 Portfolio optimization using incomplete HF information

News releases on earnings and dividends happen often outside of the regular trading hours, to avoid irrational impacts on the asset prices, inducing those moves in the close-to-open period. These moves will have an impact on the correlations at LF when using close-to-close data, but not at HF if using open-to-close data only. Do the authors consider that the correct correlation is that from using HF open-to-close data exclusively, ignoring these effects? Can we expect the correlation matrix to be equivalent?

4 Benchmarking the model

4.1 FSV as a benchmark

The authors propose a single benchmark with two different variants (amount of information included and factors included), and an incomplete set of data that adjusts to their needs (open-to-close). This seems rather restrictive, given the extensive existing literature of alternative options in GARCH and SVM spaces. What is the authors' viewpoint as the closest benchmark to their approach from the set of statistical approaches produced in the literature (Nakajima 2012), and could any such approach be potentially superior to theirs? Additionally, it seems that the 2 factor FSV performs better than the 1 factor for the MVP approach. Did the authors explore a larger set of factors?

4.2 EWMA as benchmark

Since the authors argue that EWMA outperforms SVM in their exploratory analysis as motivating background, it would have been interesting to add EWMA as a benchmark. How will their approach compare with a similarly-calibrated decay parameter in an EWMA approach, using either the LF or the HF data, or both, to construct the covariance matrices over a similar in-sample and out-of-sample period?

4.3 Minimum variance portfolio measures

The authors use the one-step ahead forecasts to assess the relative performance of their approach against others. Ignoring the fact that a single dataset was used to indicate superiority of their methodology, it would be interesting to know, for the out-of-sample performance measurement (not for fitting the model) whether the authors used only the out-of-sample open-to-close returns of the assets or the full asset returns, which would be the actual returns in a portfolio that is not unwound on a daily basis, which would be an unrealistic assumption for LF holding periods, the type implied by the underlying frequency of forecast.

4.4 Fair play with information sets

The authors have a clever use of the augmented data to improve the quality of the estimates. However, the benchmarking is still done in terms of the proposed method, where the HF data is used only as a source for better LF covariance estimation, not as the returns in a pure FSV approach. Would it be possible and, if so, more fair to let FSV use those HF returns directly, rather than through the proposed tweak (FSVE)? Would it have performed better than the proposed approach?

5 Conclusion

As our closing remark, we would like to emphasize that the authors' contribution is a great addition to the literature, especially because of its practical implications, simplicity and the addition of higher granularity in a seamless way. However, we believe that a thorough discussion of its limitations and further detail on the choices made for the comparison exercise are necessary to outline the risks of a widespread use across different applications.

References

Ait-Sahalia, Y., Fan, J., and Xiu, D. (2010). "High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data." *Journal of the American Statistical Association*, 105(492): 1504–1517. 812

Ait-Sahalia, Y., Jacod, J., and Li, J. (2012). "Testing for jumps in noisy high frequency data." *Journal of Econometrics*, 168: 207–222. 811, 812, 813

Andersen, T. G. and Bollerslev, T. (1997). "Intraday periodicity and volatility persistence in financial markets." *Journal of Empirical Finance*, 4: 115–158. 812

Berg, A., Meyer, R., and Yu, J. (2004). "Deviance Information Criterion for Comparing Stochastic Volatility Models." *Journal of Business and Economic Statistics*, 22(1): 107–120. 811

Bollerslev, T., Litvinova, J., and Tauchen, G. (2006). "Leverage and Volatility Feedback Effects in High-Frequency Data." *Journal of Financial Econometrics*, 4(3): 353–384. 812

Darrat, A., Rahman, S., and Zhong, M. (2003). "Intraday trading volume and return volatility of the DJIA stocks: A note." *Journal of Banking & Finance*, 27: 2035–2043. 813

Darrat, A., Zhong, M., and Cheng, L. T. (2007). "Intraday volume and volatility relations with and without public news." *Journal of Banking & Finance*, 31: 2711–2729. 812

Eraker, B., Johannes, M., and Polson, N. (2003). "The Impact of Jumps in Volatility and Returns." *Journal of Finance*, LVIII(3): 1269–1300. 812

Fabozzi, F. J. (2008). *Handbook of Finance: Valuation, Financial Modelling, and Quantitative Tools*. Wiley & Sons. 809

Ishihara, T., Omori, Y., and Asai, M. (2012). "Matrix exponential stochastic volatility with cross leverage." Working paper, Faculty of Economics, University of Tokyo. 812

Kalnina, I. and Xiu, D. (2013). "Model-Free Leverage Effect Estimators at High Frequency." Chicago Booth Research Paper No. 13-83. Available at SSRN: http://ssrn.com/abstract=2360256 or http://dx.doi.org/10.2139/ssrn.2360256. 812

Kyj, L. M. (2008). "Estimating Realized Covariance using High Frequency Data." Ph.D. thesis, Rice University. 809

Nakajima, J. (2012). "Bayesian analysis of multivariate stochastic volatility with skew distribution." Discussion paper number 2013-01. Department of Statistical Science, Duke University. 812, 816

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference." *Journal of Econometrics*, 140: 425–449. 812

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. UK: Oxford University Press, 1st edition. 809

Tankov, P. and Cont, R. (2003). *Financial Modelling with Jump Processes.*. Chapman and Hall/CRC Financial Mathematics Series. 812

Xu, D. and Li, Y. (2010). "Empirical Evidence of the Leverage Effect in a Stochastic Volatility Model: A Realized Volatility Approach." No 1002, Working Papers from University of Waterloo, Department of Economics. 812

Yu, J. (2005). "On Leverage in a Stochastic Volatility Model." *Journal of Econometrics*, 127: 165–178. 812