Introduction to the Issue on High Frequency Econometrics

Lukas BAUER[1], Roxana HALBLEIB[2], Richard OLSEN[3], Torben G. ANDERSEN[4], and Ingmar NOLTE[5]

Over the past decades, advances in computer technology have facilitated the collection of vast amounts of financial data, leading to the emergence of high-frequency (HF) financial econometrics. A critical boost to the area was the Olsen & Associates (O&A) decision to share HF data with the academic community and to host the first conference on the topic in 1995.

In May 2022, Roxana Halbleib, Winfried Pohlmeier, and Sandra and Ingmar Nolte organized a conference on Intrinsic Time in Finance at Hegne Abbey by Lake Constance, inspiring this themed issue. The conference gathered scholars and practitioners in econometrics, mathematical statistics, and finance to discuss advances in the intrinsic time perspectives of HF financial econometrics.

In the following two sections, we first give a brief account of O&A's role in the development of HF financial econometrics, almost 30 years after the original O&A conference, and we then provide an overview of the articles included in the themed issue, edited by Torben G. Andersen, Christian Gourieroux, Roxana Halbleib, Ingmar Nolte, and George Tauchen.

O&A Data Sharing

O&A, founded in 1985 by Dr. Richard B. Olsen, was a research institute focused on developing an information system (IS) for financial markets providing real-time forecasts and trading recommendations based on tick-by-tick market data. Since data storage was expensive and data vendors did not provide tick data, O&A scraped data from video screens of Reuters IS without any programming interface. They initially focused on foreign exchange (FX), the largest financial market with trading around the clock. Later, the service was expanded to short-term interest rate markets.

[1]Chair of Statistics and Econometrics, Institute of Economics, University of Freiburg, Germany.
[2]Chair of Statistics and Econometrics, Institute of Economics, University of Freiburg, Germany.
[3]Lykke, Olsen Ltd, Switzerland, richard.olsen@lykke.com
[4]Kellogg School, Northwestern University, USA; NBER; Canadian Derivatives Institute, Montreal; and Center for Research in Energy, (CoRE), Aarhus.
[5]Lancaster University Management School, Lancaster, United Kingdom.

From the start, O&A interacted with the academic community, including Benoit Mandelbrot, Charles Goodhart, Richard Baillie, Tim Bollerslev, Frank Diebold, Rob Engle, Joel Hasbrouck, Michael Melvin and Maureen O'Hara. Their support helped render the O&A conferences on High Frequency Data in Finance in 1995 and 1998 a success. In connection with these events, O&A made three HF datasets of 110+ Mio FX price quotes (HFDF-1993, HFDF-1996 and HFDF-2000), (freely) available. This large data sample was both a revelation and a challenge. The initial O&A event was the first financial big data conference with 200 scholars exploring the same data set from many different perspectives. The conferences inspired foundational publications in the field of high-frequency financial econometrics: e.g., Karolyi and Stulz (1996); Andersen and Bollerslev (1997, 1998*a*,*b*); Müller et al. (1997); Taylor and Xu (1997); Madhavan et al. (1997); Comte and Renault (1998); Gallant et al. (1999); Huang and Masulis (1999); Ray and Tsay (2000); Andersen et al. (2001); Martens (2001); Kirman and Teyssière (2001); Barndorff-Nielsen and Shephard (2001, 2002); Maheu and McCurdy (2002); Andersen et al. (2003, 2005).

O&A also conducted extensive research, summarized in the book 'Introduction to High Frequency Finance' (Dacorogna et al., 2001), with topics including the intraday volatility pattern, the intrinsic ϑ*-time* used to deseasonalize HF data; the *scaling law* between absolute returns and underlying time spans, reinforcing the notion that FX market breathes in *intrinsic* time, the strong *negative autocorrelation* and *fat tails* linked to short-term price overshoots, the notion of *market heterogeneity* driven by traders acting at different time horizons; the heterogenous autoregressive conditional heteroskedastic (HARCH) volatility model (Müller et al., 1997), and the *directional-change* time operator (Guillaume et al., 1997) and associated scaling laws.

The pioneering O&A data sharing initiatives, conferences, and research findings provided important empirical evidence on HF market features that were partially known, but little exploited. Along with the increasing computational power, the O&A activities opened new avenues in understanding price formation as well as measurement and forecasting of market intensity and riskiness.

Overview of the Themed Issue

In this section, we introduce the 14 papers included in the themed issue on High Frequency Econometrics.

The papers of Laurent, Renò & Shi and of Kolokolov, Renò & Zoi deal with non-trivial *drifts in HF returns*. In particular, the first allows for both the drift and volatility components in a univariate diffusion representation to be locally explosive. It shows that such joint drift and volatility processes impact the bias, variance, and convergence rate of realized volatility (RV) estimators. In this setting, the *realized drift* can readily be estimated from a fixed time span using the realized autocovariance (RAC) at first order. In contrast, drift estimators within standard settings require a long span of data and remain noisy. An empirical application shows that augmenting some existing volatility forecast models with past values of RAC and interaction terms between RAC and realized quarticity significantly improvees predictive performance.

Differently, Kolokolov, Renò and Zoi test for non-negligible *drifts* in HF returns. They develop unbiased block estimators with uniformly minimum variance for a fixed block, i.e., the Block Uniformly Minimum Variance Unbiased (BUMVU) estimators as an extension of existing HF non-overlapping block UMVU estimators. These are useful in estimating time-varying parameters when using data blocks of fixed size, e.g., for estimating the lower variance bounds for block estimators of volatility functionals computed from HF data or for estimation of homoskedastic nonparametric regressions with varying mean. The paper shows, among other results, that RV is BUMVU, but there exists no BUMVU estimators for volatility powers above 2. The test for a non-negligible drift, built on the difference between RV (biased by the drift) and the BUMVU estimator, has good properties in small samples, also under the presence of jumps, stochastic volatility and many zero observations. Applied to the NASDAQ index, the test reveals many examples of non-negligible drifts in HF data and it separates constant drift from short-time drift periods within the trading day.

The papers of Oh, Kim & Wang and Li & Linton seek to estimate integrated measures from HF data. The former estimates *integrated beta* from HF data contaminated with endogenous and autocorrelated market microstructure noise (MMN), thus concentrating on generating useful estimates for integrated and spot volatility in the presence of MMN. Therefore, like Kolokolov, Renò & Zoi, they focus on integrated quantities of time-varying measures involving a (co)volatility functional, in this case *spot beta*. Their realized integrated beta (RIB) estimator has desirable asymptotic properties, such as consistency and a limiting Gaussian distribution. Their paper also proposes an autoregressive framework for predicting future daily beta values, and their empirical exercise demonstrates the effectiveness of the robust RIB estimator in constructing market-neutral portfolios.

Closely related, Li & Linton focus on accurately *estimating spot and integrated measures from HF data* under complex, asymptotically vanishing MMN structures, such as price-dependency, autocorrelation, and nonstationarity, while also allowing for random observation times, rendering the *spot volatility* estimator available at the tick frequency and robust to MMN. The paper extends the standard discretized diffusion price framework through the asymptotically vanishing MMN and builds on prior moment estimators of MMN to correct the bias in the underlying RV estimator. The infill asymptotics of the new RV estimators are generally valid only when accounting for the serial correlation in MMN and random observation times. Likewise, the well-known weak "U" shape and intraday reverse "J "shape of volatility are empirically detectable only from the irregular observation scheme. The estimators proposed are computationally efficient and exploit the huge information content of the tick data, while being robust to MMN.

Dealing with complex MMN structures and *time endogeneity* when *estimating integrated volatility* (IV) is also addressed by Cui et al. (2024). They decompose MMN into an explicative part driven nonparametrically by the market's trading information and a residual induced by market complexity. They develop a two-step Laplace estimator of daily IV computed from the observed prices filtered by the explicative MMN component, estimated from limit order book data under general assumptions, such as beta-mixing dependency in the trading information, jumps with infinite variation and *endogenous* observation times. Thus, non-return information is used and irregular sampling is accounted for. Their results show that the information in the limit order book and the time endogeneity significantly improves the efficiency of the variance estimator along with its convergence rate. They also provide new empirical evidence on the reversed "J" shape of intraday volatility.

The next two contributions further demonstrate the benefits of *time endogeneity in estimating RV*. Dimitriadis, Halbleib, Polivka, Rennspies, Streicher & Wolter assume asset prices follow a stochastic diffusion that is time-changed by the transaction process; the tick-time stochastic volatility (TTSV) model. It allows for the jumps/ticks to exhibit stochastic variance as empirically observed throughout the trading day. In contrast to the discretized diffusion model, it allows them to distinguish between time-varying trading intensity and time-varying tick variance and, thus, between alternative *intrinsic time* sampling schemes. Moreover, it also explores the use of sampling driven by absolute price changes, i.e., *hitting times*, when studying the finite sample properties of daily RV. The paper departs from the classical goal of determining the optimal sampling frequency. Instead, it focuses on the optimal allocation of sampling points for a given frequency. The paper finds that the endogenous sampling schemes outperform calendar sampling, with the hitting time or newly introduced realized business time sampling (using observed trades and estimated tick variance) providing the most efficient RV estimators.

Li, Li, Nolte, Nolte & Yu also pursue a *hitting time or price duration* sampling (PDS) scheme in their development of a jump test. The idea is that, for a specific sampling interval and a given threshold, jumps trigger disproportionally larger threshold overshoots than continuous price increments. The test statistic is based on the ratio between the sample moments of uncensored and threshold censored PDS returns normalized by the barrier width defining the threshold. To mitigate weak dependencies due to MMN, the approach uses return pre-averaging as first step and a wild bootstrap as second step procedures, ensuring observations display a limiting Gaussian random walk dynamic. Simulation results show that the test compares favorably to existing ones, both in terms of size and power, and especially when MMN is substantial. Applied to a set of NYSE stocks, the test indicates presence of jumps for all stocks for up to 15% of the trading days.

Andersen, Bondarenko, Gousgounis & Onur stipulate the existence of a general *HF trading market equilibrium* that trades off business-time risk exposure (return variation per standard transaction), trading costs (bid-ask spread) and market depth at top of the order book. This Trading Invariance (TI) relationship generalizes prior work emphasizing the pronounced short-term correlation between return variation and trading activity, while also extending HF specifications of the market microstructure invariance hypotheses. In the application to a number of FX futures contracts, tick size changes serve as natural experiments that confirm the asserted interaction among the specified risk and liquidity components, while existing invariance-style hypotheses fail. The paper also stresses the inferential biases associated with the traditional regression-based approach in existing empirical studies and thus develops a robust GMM-based alternative. Most strikingly, the coefficient values of the log-linear TI representation are confirmed to be nearly identical across currencies and time, corroborating the hypothesis of a stable and invariant trading market equilibrium. The findings have implications for market design and surveillance.

Todorov & Zhang theoretically identify and estimate the ”U” or reversed ”J” shape of the *intraday volatility pattern* from short-dated *options*. Hence, they characterize the HF pattern in spot volatility from options and not HF returns. The identification of the periodicity in the intraday volatility pattern is obtained from the ratio of nonparametric estimates of the conditional risk-neutral expectation of future return variation over the two tenors and by the fact that the volatility and jump intensity components are not expected to change over a short time span. The advantage of the option data in this setting is that it provides direct observation of the expected future return variation and may exploit short-term asymptotics. The rate of convergence of the estimator depends on the mesh of the options strike grids and tenors, while the simulations show that the precision of the estimator depends on the time-of- day with less noisy estimates towards the market close.

Patton & Zhang also find *Intraday Volatility* patterns to play a central role in tailoring the weights of the intraday returns in the classical RV estimator according to the degree of improvement in volatility forecasts obtained by supervised learning. Hence, they contribute to the emerging field of using Machine Learning techniques to estimate RV by allowing the weights to vary within the trading day. Comprehensive evidence from more than 800 U.S. stocks, a broad range of RV forecasting models and numerous multi-step ahead exercises, show that the weights of the newly introduced *bespoke RV* estimator load heavier on recent lags and increase unevenly during the day, with a sharp boost within the last two trading hours.

Archakov, Hansen & Lunde likewise extend the basic (multivariate) RV estimator to improve *variance forecasts* for the integrated covariance matrix. Their multivariate generalization of Realized GARCH exploits the DCC decomposition and a vectorization of the matrix log-transformation of the correlation matrix. While the individual conditional volatilities are modeled by univariate realized GARCH to reduce the curse of dimensionality, the vector of log-transformed variables of the correlation matrix follows a dynamic factor structure with the factors governed by their realized counterparts. An alternative dimension reduction approach imposes a block structure (e.g., sector-wise) on the correlation matrix. Their model parameters are readily estimated by a simple two-step Maximum Likelihood procedure, and the model outperforms purely low frequency counterparts when forecasting the covariance matrix of nine assets.

In modeling and forecasting the covariance matrix based on realized measures, the choice of underlying *matrix distribution* is critical. Stollenwerk provides a comprehensive and unified framework of matrix probability distributions for covariance matrices, including the Wishart, F- and Riesz distributions along with new proposals, such as the t- or inverse F-Riesz family. Built on stochastic decompositions using lower (upper) random matrices with chi-squared distributions on diagonals and Gaussians on the off-diagonals, these families allow for dependencies and distinct degrees of fat-tailedness among the covariance elements. In an application exploiting the so-called scalar-BEKK dynamic representation, the t-Riesz distribution proves to be the best fit for covariance series.

Li, Chen & Linton, like Archakov et al., deal with the curse of dimensionality in multivariate settings through *factor models*. Their focus on HF return modeling, stipulating a dual factor structure, and estimation through a Double PCA (DPCA) procedure. The aim is to discriminate between factors specific to efficient prices and factors associated with MMN. A first PCA is applied to observed HF returns, while a second PCA is applied to the cumulative sum of the common components from the first step. Exploiting the nonstationarity of the frictionless efficient price process and stationary of the MMN, a split is obtained into factors for the efficient and noise components, respecitvely. The procedure is consistent when the number of assets and sampling frequency diverge, and it is convenient in dealing with substantial MMN and Epps effects. Their application to S&P 500 components shows that MMN exhibit common factors, albeit smaller in magnitude than the common systemic risk factors. Thus, the MMN structure is harder to detect, but they conclude it may offer substantial improvement in portfolio management.

Chen, Feng, Mykland & Zhang also deal with *high-dimensional* modeling using HF data. They develop three *regression coefficient tests* allowing for the number of regressors to exceed the number of observations and even to diverge asymptotically. A particular focus is HF regressions with time-varying coefficients that have found broad applications in asset pricing and risk management when, e.g., estimating and testing *beta* loadings or estimating large covariance matrices. The different tests are robust to either dense or sparse alternatives or to both, and they are shown to be empirically useful in identifying HF factors with incremental information and time-varying loadings within the classical Fama–French 3-factor model setting.

References

Andersen, T. G. and Bollerslev, T. (1997), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**(2-3), 115–158.

Andersen, T. G. and Bollerslev, T. (1998*a*), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**(4), 885–905.

Andersen, T. G. and Bollerslev, T. (1998*b*), 'Deutsche mark–dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies', *The Journal of Finance* **53**(1), 219–265.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **96**(453), 42–55.
URL: *https://doi.org/10.1198/016214501750332965*

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T. and Meddahi, N. (2005), 'Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities', *Econometrica* **73**(1), 279–296.
URL: *http://www.jstor.org/stable/3598946*

Barndorff-Nielsen, O. E. and Shephard, N. (2001), 'Non–Gaussian Ornstein–Uhlenbeck–based models and some of their uses in financial economics', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **63**(2), 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **64**(2), 253–280.

Comte, F. and Renault, E. (1998), 'Long memory in continuous–time stochastic volatility models', *Mathematical finance* **8**(4), 291–323.

Cui, W., Hu, J. and Wang, J. (2024), 'Nonparametric estimation for high-frequency data incorporating trading information', *Journal of Econometrics* **240**(1), 105690.

Dacorogna, M., Gençay, R., Müller, U. A., Pictet, O. and Olsen, R. B. (2001), *An introduction to high-frequency finance*, Academic Press, San Diego.

Gallant, A. R., Hsu, C.-T. and Tauchen, G. (1999), 'Using daily range data to calibrate volatility diffusions and extract the forward integrated variance', *The Review of Economics and Statistics* **81**(4), 617–631.
URL: *http://www.jstor.org/stable/2646711*

Guillaume, D. M., Dacorogna, M., Davé, R., Müller, U. A., Olsen, R. B. and Pictet, O. (1997), 'From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets', *Finance and Stochastics* **1**, 95–129.

Huang, R. D. and Masulis, R. W. (1999), 'Fx spreads and dealer competition across the 24-hour trading day', *The Review of Financial Studies* **12**(1), 61–93.
URL: *http://www.jstor.org/stable/2645987*

Karolyi, G. A. and Stulz, R. M. (1996), 'Why do markets move together? an investigation of u.s.-japan stock return comovements', *The Journal of Finance* **51**(3), 951–986.
URL: *http://www.jstor.org/stable/2329228*

Kirman, A. and Teyssière, G. (2001), 'Microeconomic models for long memory in the volatility of financial time series', *Studies in Nonlinear Dynamics & Econometrics* **5**(4).
URL: *https://doi.org/10.2202/1558-3708.1083*

Madhavan, A., Richardson, M. and Roomans, M. (1997), 'Why do security prices change? a transaction-level analysis of nyse stocks', *The Review of Financial Studies* **10**(4), 1035–1064.
URL: *http://www.jstor.org/stable/2962338*

Maheu, J. M. and McCurdy, T. H. (2002), 'Nonlinear features of realized fx volatility', *The Review of Economics and Statistics* **84**(4), 668–681.
URL: *http://www.jstor.org/stable/3211725*

Martens, M. (2001), 'Forecasting daily exchange rate volatility using intraday returns', *Journal of International Money and Finance* **20**(1), 1–23.
URL: *https://www.sciencedirect.com/science/article/pii/S0261560600000474*

Müller, U. A., Dacorogna, M., Davé, R., Olsen, R. B., Pictet, O. and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions—analyzing the dynamics of market components', *Journal of Empirical Finance* **4**(2-3), 213–239.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions — analyzing the dynamics of market components', *Journal of Empirical Finance* **4**(2), 213–239. High Frequency Data in Finance, Part 1.
URL: *https://www.sciencedirect.com/science/article/pii/S0927539897000078*

Ray, B. K. and Tsay, R. S. (2000), 'Long-range dependence in daily stock volatilities', *Journal of Business & Economic Statistics* **18**(2), 254–262.
URL: *https://www.tandfonline.com/doi/abs/10.1080/07350015.2000.10524867*

Taylor, S. J. and Xu, X. (1997), 'The incremental volatility information in one million foreign exchange quotations', *Journal of Empirical Finance* **4**(4), 317–340.