

Empirical Finance: Methodology and Application

Oliver Linton and Alexey Onatskiy

August 31, 2018

There are several subthemes to our ongoing research, which each contain both econometric methodology and empirical application

1. Market Microstructure
2. Big Data

We currently have two postdoctoral researchers: Marco Valerio Geraci (hired in 2017) and Z. Merrick Li (Starting in September 2018), and around 10 PhD students actively working in the areas considered here.

1 Outputs

1. **Implications of high-frequency trading for security markets** (Linton, O. and Mahmoodzadeh, S.), (2018) *Annual Review of Economics* 2018. 10:237–59. <https://doi.org/10.1146/annurev-economics-063016-104407>. Also issued as CWPE1802. High-frequency trading (HFT) has grown substantially in recent years due to fast-paced technological developments and their rapid uptake, particularly in equity markets. This review investigates how HFT could evolve and, by developing a robust understanding of its effects, identifies potential risks and opportunities that HFT could present in terms of financial stability and other market outcomes such as volatility, liquidity, price efficiency, and price discovery. Despite commonly held negative perceptions, the available evidence indicates that HFT and algorithmic trading may have several beneficial effects on markets. However, these types of trading may cause instabilities in financial markets in specific circumstances. Carefully chosen regulatory measures are needed to address concerns in

the shorter term. However, further work is needed to inform policies in the longer term, particularly in view of likely uncertainties and lack of data. This work will be vital in supporting evidence-based regulation in this controversial and rapidly evolving field.

2. **Semiparametric ultra-high dimensional model averaging of non-linear dynamic time series** (Chen, J., Li, D., Linton, O., and Lu, Z.), (2018) *Journal of the American Statistical Association, Theory and Methods*, Vol 113, pp 919-932. doi = 10.1080/01621459.2017.1302339. We propose two semiparametric model averaging schemes for nonlinear dynamic time series regression models with a very large number of covariates including exogenous regressors and autoregressive lags. Our objective is to obtain more accurate estimates and forecasts of time series by using a large number of conditioning variables in a nonparametric way. In the first scheme, we introduce a Kernel Sure Independence Screening (KSIS) technique to screen out the regressors whose marginal regression (or auto-regression) functions do not make a significant contribution to estimating the joint multivariate regression function; we then propose a semiparametric penalized method of Model Averaging MArginal Regression (MAMAR) for the regressors and auto-regressors that survive the screening procedure, to further select the regressors that have significant effects on estimating the multivariate regression function and predicting the future values of the response variable. In the second scheme, we impose an approximate factor modelling structure on the ultra-high dimensional exogenous regressors and use the principal component analysis to estimate the latent common factors; we then apply the penalized MAMAR method to select the estimated common factors and the lags of the response variable that are significant. In each of the two schemes, we construct the optimal combination of the significant marginal regression and auto-regression functions. Asymptotic properties for these two schemes are derived under some regularity conditions. Numerical studies including both simulation and an empirical application to forecasting in action are given to illustrate the proposed methodology
3. **The Behaviour of Betting and Currency Markets on the Night of the EU Referendum** (Auld, T. and Linton, O.), (2017) CWPE1750 Forthcoming in *International Journal of Forecasting*. We study the behaviour of the Betfair betting market and the sterling/dollar exchange rate (futures price) during 24 June 2016, the night of the EU referendum. We investigate how the two markets responded to the an-

nouncement of the voting results. We employ a Bayesian updating methodology to update prior opinion about the likelihood of the final outcome of the vote. We then relate the voting model to the real time evolution of the market determined prices. We find that although both markets appear to be inefficient in absorbing the new information contained in vote outcomes, the betting market is apparently less inefficient than the FX market. The different rates of convergence to fundamental value between the two markets leads to highly profitable arbitrage opportunities.

4. **The 2016 sterling flash crash: when liquidity disappeared from one of the worlds most liquid markets** (with O. Tobek, J. Noss, L. Crowley-Reidy, and Lucas Pedace) Bank of England Working paper no 687, <https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2017/the-october-2016-sterling-flash-episode> This paper provides an in-depth analysis of the evolution of liquidity during the flash episode in sterling during the early hours of 7 October 2016. It examines a number of estimates both of the cost of trading, and the price impact of executed transactions. These include a variant of the ‘volatility over volume’ measure of liquidity based on transaction data, which provides a better proxy of illiquidity — as given by measures based on high-frequency limit order book data — than other summary measures of price impact. The paper also shows that the fall in the value of sterling during the initial part of the flash episode was consistent with the estimated impact on prices of a large number of individually small — but in aggregate large — volume of orders to sell sterling during a normally quiet period of the trading day. However, the subsequent change in price was larger than that consistent with the estimated impact on prices of observed orders to sell sterling. This might support the suggestion, which was included in the report on the episode provided by the Bank for International Settlements, that the move in sterling may have been amplified by the pause in trading on the CME futures exchange.
5. **The Corporate bond purchase scheme and trading behaviour in the sterling corporate bond market** (with Y. Baranova, L. Boneva, D. Elliott, I. Kaminska, B. Morley, and N. McLaren) Bank of England working paper shortly. In August 2016, the Bank of England announced a Corporate Bond Purchase Scheme (CBPS) to purchase up to \$10bn of sterling corporate bonds. To investigate the impact of these purchases on liquidity, we create a novel dataset that combines transaction-level data from the secondary corporate bond market

with proprietary bid-level data from the CBPS auctions. Identifying the impact of central bank asset purchases on liquidity is potentially impacted by reverse causality, because liquidity considerations might impact purchases. But the bid-level data allow us to construct proxy measures for the BoE’s demand for bonds and auction participants’ supply of bonds, meaning that we can control for the impact of liquidity on purchases. Across a range of liquidity measures, we find that CBPS purchases improved the contemporaneous liquidity of purchased bonds. In addition, we find that the liquidity of sterling-denominated bonds improved relative to euro-denominated bonds over the purchase period, consistent with a portfolio rebalancing channel.

6. **An Empirical Analysis of Circuit Breakers on the London Stock Exchange** (with James Brugler, Joseph Noss, and Lucas Pedace) To appear as BoE working paper. This paper uses transaction data to estimate how single stock circuit breakers on the London Stock Exchange affect the market quality of stocks that remain in continuous trading. This ‘spillover’ effect is estimated by calculating the effect of a trading halt on the market quality of stocks that remain in continuous trading, but whose absolute returns are of a magnitude near sufficient to trigger a trading halt. We find that circuit breakers lead to a significant improvement in the liquidity, and reduction in the volatility, of stocks that remain in continuous trading. This might suggest that – at least over the period covered by our data – single stock circuit breakers play an important role in reducing the spillover of poor market quality across stocks.
7. **A New Semiparametric Estimation of Large Dynamic Covariance Matrix with Multiple Conditioning Variables** (with J. Chen and D. Li) *Journal of Econometrics*. This paper studies the estimation of dynamic covariance matrices with multiple conditioning variables, where the matrix size can be ultra large (divergent at an exponential rate of the sample size). We introduce an easy-to-implement semiparametric method to estimate each entry of the covariance matrix via model averaging marginal regression, and then apply a shrinkage technique to obtain the large dynamic covariance matrix estimation. Under some regularity conditions, we derive the asymptotic properties for the proposed estimators including the uniform consistency with general convergence rates. We further consider extending our methodology to deal with the scenarios: (i) the number of conditioning variables is divergent as the sample size increases, and (ii) the large covariance ma-

trix is conditionally sparse relative to contemporaneous market factors. We provide a simulation study that illustrates the finite-sample performance of the developed methodology. We also provide an application to financial portfolio choice from daily stock returns.

8. **A Coupled Component DCS GARCH model for intraday and overnight volatility** (with J. Wu) Working paper to be submitted to Journal of Econometrics shortly. We propose a semi-parametric coupled component GARCH model for intraday and overnight volatility that allows the two return series to have different properties. We adopt a dynamic conditional score model with t-distributed innovations that captures the very heavy tails of overnight returns. We propose a several-step estimation procedure that captures the nonparametric slowly moving components by kernel estimation and the dynamic parameters by estimated maximum likelihood. We establish the consistency, asymptotic normality, and semiparametric efficiency of our semiparametric estimation procedures. We extend the modelling to the multivariate case where we allow time varying correlation between stocks. We apply our model to the study of Dow Jones industrial average component stocks, CRSP size-based portfolios, and size-based portfolios in four large international markets over the period 1993-2017. We show that the ratio of overnight to intraday volatility has actually increased in importance for Dow Jones stocks during the last two decades. This ratio has also increased for large stocks in the CRSP database, but decreased for small stocks in CRSP. Notably, the slope increases monotonically from the smallest-cap decile to the largest-cap decile. This pattern also exists in other international markets. The multivariate model shows that overnight and intraday correlations have both increased, but overnight correlations have increased more substantially during recent crises than intraday correlations.
9. **Do Consumption-Based Asset Pricing Models Explain Serial Dependence in Stock Returns?** (with M. Ashby) Available at SSRN: <https://ssrn.com/abstract=3173586> or <http://dx.doi.org/10.2139/ssrn.317358>. We show that the Bansal-Yaron, Campbell-Cochrane and Cecchetti-Lam-Mark models of asset prices cannot explain the serial correlation structure of stock returns. We show this by estimating these models and deriving expected returns from them and then testing whether the difference between observed and expected returns is a martingale difference sequence. We use variance ratio and rescaled range tests which we modify to account for the expected returns being functions of esti-

mated parameters. We also use a weighted quantilogram test based on a bootstrap procedure robust to this estimation. The evidence against the Bansal-Yaron and Campbell-Cochrane models is significant. While the evidence against the Cecchetti-Lam-Mark model is not in general significant, our point estimates strongly suggest its residuals are not a martingale difference sequence. Furthermore, a semi-parametric maximal predictability test suggests there is some evidence that the three models' state variables struggle to explain the degree of predictability observed in the market return. A timing strategy designed to exploit predictability in the market can significantly outperform the market in certainty equivalent terms under the Bansal-Yaron model. The timing strategy may underperform the market by less than it ought to under the Campbell-Cochrane model.

10. **Testing in High-dimensional Spiked Models** (Johnstone, I. M. and Onatski, A.), (2018) *Annals of Statistics*, forthcoming. We consider the five classes of multivariate statistical problems identified by James (1964), which together cover much of classical multivariate analysis, plus a simpler limiting case, symmetric matrix denoising. Each of James' problems involves the eigenvalues of $E^{-1}H$ where H and E are proportional to high dimensional Wishart matrices. Under the null hypothesis, both Wisharts are central with identity covariance. Under the alternative, the non-centrality or the covariance parameter of H has a single eigenvalue, a spike, that stands alone. When the spike is smaller than a case-specific phase transition threshold, none of the sample eigenvalues separate from the bulk, making the testing problem challenging. Using a unified strategy for the six cases, we show that the log likelihood ratio processes parameterized by the value of the sub-critical spike converge to Gaussian processes with logarithmic correlation. We then derive asymptotic power envelopes for tests for the presence of a spike.
11. **Alternative Asymptotics for Cointegration Tests in Large VARs** (Onatski, A. and Wang, C.), (2018) *Econometrica*, forthcoming. Johansen's (1988, 1991) likelihood ratio test for cointegration rank of a Gaussian VAR depends only on the squared sample canonical correlations between current changes and past levels of a simple transformation of the data. We study the asymptotic behavior of the empirical distribution of those squared canonical correlations when the number of observations and the dimensionality of the VAR diverge to infinity simultaneously and proportionally. We find that the distribution almost

surely weakly converges to the so-called Wachter distribution. This finding provides a theoretical explanation for the observed tendency of Johansen's test to find "spurious cointegration". It also sheds light on the workings and limitations of the Bartlett correction approach to the over-rejection problem. We propose a simple graphical device, similar to the scree plot, for a preliminary assessment of cointegration in high-dimensional VARs.

12. **Extreme canonical correlations and high-dimensional cointegration analysis** (Onatski, A. and Wang, C.), (2018) CWPE1805. The simplest version of Johansen's (1988) trace test for cointegration is based on the squared sample canonical correlations between a random walk and its own innovations. Onatski and Wang (2017) show that the empirical distribution of such squared canonical correlations weakly converges to the Wachter distribution as the sample size and the dimensionality of the random walk go to infinity proportionally. In this paper we prove that, in addition, the extreme squared correlations almost surely converge to the upper and lower boundaries of the support of the Wachter distribution. This result yields strong laws of large numbers for the averages of functions of the squared canonical correlations that may be discontinuous or unbounded outside the support of the Wachter distribution. In particular, we establish the a.s. limit of the scaled Johansen's trace statistic, which has a logarithmic singularity at unity. We use this limit to derive a previously unknown analytic expression for the Bartlett-type correction coefficient for Johansen's test in a high-dimensional environment.