Siem Jan Koopman, Rutger Lit, André Lucas and Anne Opschoor,
"Dynamic Discrete Copula Models for High Frequency Stock Price Changes",
Journal of Applied Econometrics, 2018, 33, 7, 966–985.

Data source and preparation

In this document we provide the steps we took to obtain and prepare the data. We use data from the Trades and Quotes (TAQ) database of the New York Stock Exchange at a one-second frequency. VU University Amsterdam has a subscription to the Wharton Research Data Services (WRDS) which offers among other data sources, access to the TAQ database. It contains intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE) and American Stock Exchange (AMEX), as well as Nasdaq National Market System (NMS) and SmallCap issues. We have selected 10 companies from different industries and with different trade intensities for all trading days in 2012. The following corresponding Ticker symbols are USB, AIG, COF, MS, AXP, BAC, GS, WFC, C and JPM. We analyze the tick-by-tick data without the "odd-lots" that represent trades with volumes less than 100 and that are not recorded on the consolidated tape; see for example O'Hara, Yao, and Ye (2014). The data require standard pre-processing. For a review of high-frequency data cleaning procedures; see for example Falkenberry (2002). We apply the cleaning algorithm of Brownlees and Gallo (2006) after applying a rudimentary filter corresponding to the cleaning steps P1, P2, P3 and T1, T2, T3 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008, p. 8). In cleaning step T2, only trades with the sale condition {blank, @, *, E, F} are kept, see the TAQ user guide for details. Descriptive statistics are presented in Table 2 of the paper.

References

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Realised kernels in practice: Trades and quotes. *Econometrics Journal 4*, 1–33.

Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis 51*, 2232–2245.

Falkenberry, T. N. (2002). High frequency data filtering. Technical report, Tick Data.

O'Hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *Journal of Finance 69(5)*, 2199–2236.