Forecasting Bid- Ask Spreads

Axel Gross-Klussmann

Nikolaus Hautsch

Motivation from a practitioner's point of view

- Spreads are a measure of liquidity on stock markets: Trading is costly when the spread is large
- *Typical example in Trading:* Large orders split up into several smaller orders

  Placement of the orders depends on the spread

  Small spreads will be taken as signal for a cheap immediate trading opportunity
- Forecasts of the spread can help to reduce transaction costs

Motivation from a scientific perspective

- No study on the time series properties of intraday spread series available
- Virtually no relevant forecasting study available
- Reasons probably: spread distributions discrete and much more diffuse than e.g. volatility

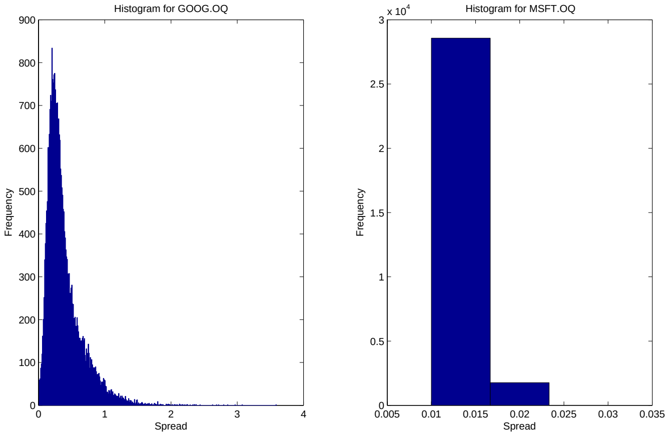
Outline

- Introduction
- Data
- Econometric Methodology
- Forecast Framework
- Results
- Conclusion
- (Appendix)

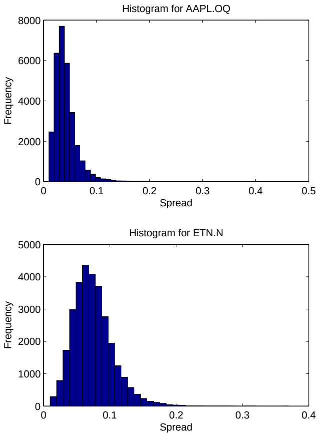
Data

- Spread data for 4 stocks from NYSE and NASDAQ in 2008
- Spread taken to be the spread at the endpoint of each interval at 5, 30, 60 to 300 second frequency
- Focus here: 30 seconds
- At this early stage of the study: Stocks with average spreads smaller than .02 are omitted

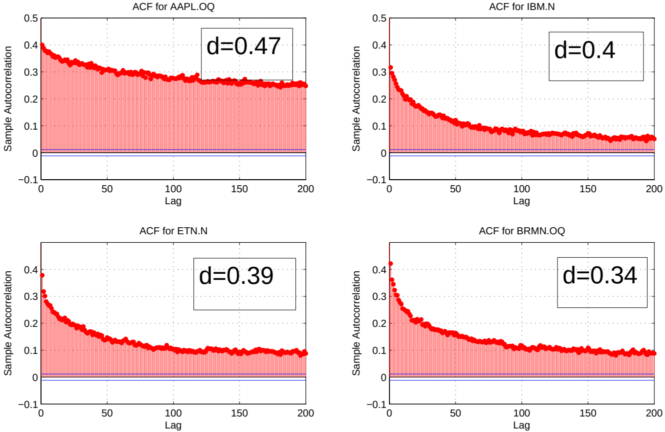
Two extreme cases



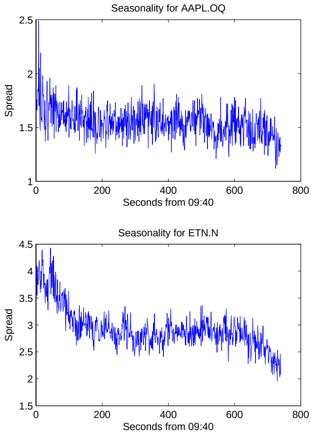
Histograms for the four chosen stocks at 30 second frequency



Autocorrelation functions for the four chosen stocks at 30 second frequency



Intraday seasonality pattern of the spread



Requirements for the Econometric Model

- Model has to be easily estimable to be employed in a rolling window setup
- Explanatory variables like seasonal adjustment terms can be included
- Framework should account for the discreteness in case of density forecasts
- Account for the long-range dependence

The Autoregressive Conditional Poisson Model (Heinen 2003)

- Very similar to the MEM framework, but $y_t = \mu_t \cdot \varepsilon_t$ does not hold
- Built around Poisson assumptions for the spread, e.g.

$$y_t | \mathcal{F}_{t-1} \sim Poi(\mu_t),$$

where μ_t is e.g. given by

$$\mu_t := E[y_t | \mathcal{F}_{t-1}] = \omega + \sum_{j=1}^{P} \alpha_j y_{t-j} + \sum_{j=1}^{Q} \beta_j \mu_{t-j}$$

Extensions of the ACP framework

- Long range dependence of the spread is addressed by

$$[1-\phi(L)](1-L)^d y_t = \omega + [1-\beta(L)]v_t,$$

where $v_t = y_t - \mu_t$ is the linear innovation and

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(1+j)} L^k$$

Stationarity and ergodicity requires $0 \leq d \leq 1$

- Some sort of non-linearities can be detected by

$$\ln \mu_t = \omega + \alpha\varepsilon_{t-1} + c|\varepsilon_{t-1} - 1| + \beta \ln \mu_{t-1}$$

with $\varepsilon_t = y_t/\mu_t$

Inclusion of additional variables

- Additional covariates x_t with corresponding parameters ϕ are easily included via an exponential link,

$$E[y_t|\mathcal{F}_{t-1}] := \mu_t^* = \mu_t \exp(x_t'\phi)$$

- Important additional predictor: seasonal adjustment terms,

$$s(t) = \delta^s \bar{t} + \sum_{j=1}^{Q} (\delta_{1,j}^s \cos(\bar{t} 2\pi j) + \delta_{2,j}^s \sin(\bar{t} 2\pi j)),$$

where $\delta^s, \delta_{1,j}^s$ and $\delta_{2,j}^s$ are parameters to be estimated

The Double Poisson Model

- The Double Poisson model allows for over- and underdispersion
- Prob. mass function (see Efron (1986)) given by

$$f(y_t, \mu_t, \gamma) = c(\mu_t, \gamma)\left(\gamma^{1/2} e^{-\gamma\mu_t}\right)\left(\frac{e^{-y_t} y_t^{y_t}}{y_t!}\right)\left(\frac{e\mu_t}{y_t}\right)^{\gamma y_t},$$

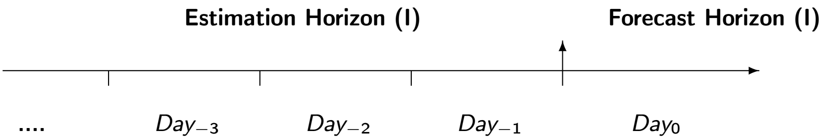
  for $\mu > 0$ and $\gamma > 0$
- Constant over- or underdispersion can be relaxed by $\gamma_t = \mu_t / h_t$ and

$$h_t := E[\sigma^2 | \mathcal{F}_{t-1}] = \omega^* + \alpha^*(y_{t-1} - \mu_{t-1})^2 + \beta^* h_{t-1}$$

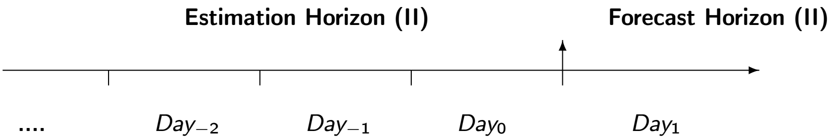
Forecast Setup

- Objective: One step ahead point and probability mass function forecasts of the spread obtained from rolling window setup
- Point forecasts from the means are rounded to match the discrete data
- Period: Jan, Feb 2008, forecast horizon: one day, estimation horizon: four days

The Rolling Window Principle



(1st Iteration)



(2nd Iteration)

Forecast Evaluation

- Point forecasts evaluated by the RMSE and the Diebold and Mariano (1995) framework
- Probability integral transforms (Diebold et al. (1998)) do not work directly for probability mass functions
- Therefore, construct for each t

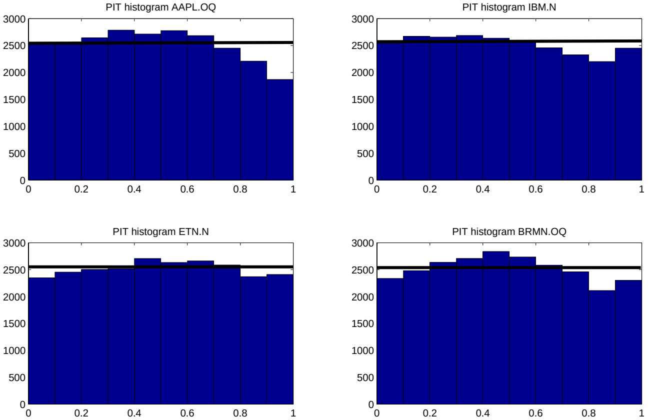
$$u_t^u := \widehat{P}(Y_t \leq y_t) \quad \text{and} \quad u_t^l := \widehat{P}(Y_t \leq y_t - 1)$$

- $u_t \sim \mathcal{U}(u_t^l, u_t^u)$ are iid $\mathcal{U}(0,1)$ under the null of a correct specification (Liesenfeld et al. (2006))

Root Mean squared prediction errors under different specifications

	AAPL.OQ *(0.038)*	IBM.N *(0.037)*	ETN.N *(0.073)*	BMRN.OQ *(0.038)*
Naive	0.0241	0.0211	0.0360	0.0263
EMA	0.0219	0.0198	0.0331	0.0221
LM ACP(1,1)	**0.0195**	**0.0185**	**0.0282**	0.0198
ACP(1,3)	0.0196	0.0186	0.0282	0.0197
ARMA(10,2)	0.0197	0.0187	0.0284	0.0198
LM MEM(1,1)	0.0196	0.0186	0.0286	**0.0197**

Histograms of the PIT for four stocks using the Double Poisson LM ACP



Important components of the forecasting models

	Point Forecasts	Density Forecasts
Double Poisson	-	X
time varying dispersion	-	-
"nonlinearity"	-	-
long memory	X	X
seasonality	(X)	(X)
additional predictors	-	-

Conclusion for the 30 second aggregates

- LM ACP models yield both the best density and the best point forecasts of spreads at the 30 second frequency
- The long range dependence in spreads matters
- Simple forms of non-linear models do not improve forecasts
- Market Microstructure theory does not help in forecasting

Thanks for your attention!

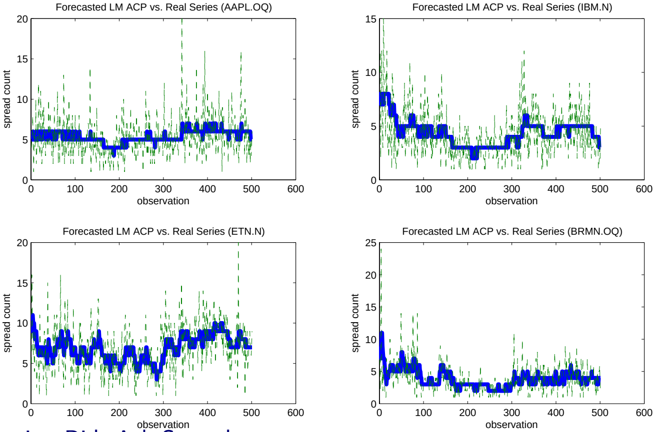
References

- Diebold, F.X. and R.S. Mariano (1995), Comparing Predictive Accuracy, *Working Paper* 169, NBER.
- Diebold, F.X., T.A. Gunther and A.S. Tay (1998), Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review* 39, 863-883.
- Efron, B. (1986), Double Exponential Families and their Use in Generalized Linear Regression, *Journal of the American Statistical Association* 81, 709-721.
- Heinen, A. (2003), Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model, *Working Paper*, SSRN.

References

- Liesenfeld, R., Nolte, I. and W. Pohlmeier (2006), Modeling financial transaction price movements: a dynamic integer count data model,*Empirical Economics* 30, 795-825.

300 observations of the spread and the forecasted series (ACP LM models)



Predictors from market microstructure theory

- Transaction component of the spread -> trading volume
- Adverse selection component -> volatility, size imbalances
- Liquidity component -> depth
- Variables or expectations of the variables can be included as additional predictors

Unconditional Mean and Variance of the ACP(1,1) model with Poisson assumption

- Unconditional mean given by

$$E[y_t] = \mu = \frac{\omega}{1-(\alpha_1+\beta_1)}. \tag{1}$$

- Overdispersion also possible under Poisson:

$$V[y_t] = \frac{\mu(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{1-(\alpha_1+\beta_1)^2} \geq \mu \tag{2}$$

Unconditional Variance of the ACP(1,1) model with Double Poisson assumption

- Unconditional variance of the ACP with Double Poisson pmf dependent on additional parameter γ

$$V[y_t] = \frac{1}{\gamma} \frac{\mu(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{1-(\alpha_1+\beta_1)^2} \geq \mu \tag{3}$$

if $\gamma \leq 1$