Memo:
Liquidity Factors

April 23, 2025

1 Data

Observations start on April 1, 2019 and end on December 31, 2024. Starting March 2023 we currently only have DAX weights available for selected months. We temporarily resolve this issue by assuming constant weights until we observe a new value.

1.1 Stocks

Stocks are the 49 stocks that in this period were at some point part of the DAX. We use computed λ_t^i and $\hat{\lambda}_t^i$ where i denotes a stock and t denotes time measured in days. Any data we use that has frequency higher than daily has been aggregated to daily data by using the last non-missing value of the day for prices, the sum over intraday trades for volumes, and the average for liquidity measures ℓ, λ, $\hat{\lambda}$.

1.2 Bonds

We downloaded daily data on DAX and Bond yields from Bloomberg for April 2019 – December 2024. In particular we use *GETB1*, the one year government bond yield, for the risk free rate r^f. Alternatively, we also have 2, 5, 10, and 30 year yields available (GETDEMxYR) with $\{x = 2, 5, 10, 30\}$.

In particular, we use:

$$\begin{aligned} r_t^m &= \frac{DAX_t - DAX_{t-1}}{DAX_{t-1}}, \\ r_t^f &= GETB1_t/100. \end{aligned}$$

2 Factor Models

Consider the regression proposed by Francioni and Egloff (with the red term added here):

$$r_t^i - r_t^f = \alpha^i + \beta^i \left(r_t^m - r_t^f \right) + \phi^i \lambda_t^i(T, R) + \varphi^i \frac{r_t^m - r_t^f}{\lambda_t^i(T, R)} + \epsilon_t^i \tag{1}$$

For λ_t constant this would be a multicollinear problem. However, for time-varying λ_t the blue part can be added to an otherwise standard factor regression. ϕ would then have the interpretation of a liquidity discount. The particular specification and how different $\lambda(T, R)$ would enter is open.

Equation (1) is estimated separately for each of the 49 stocks. For each stock we start out with the most general specification, including the red and the blue term for both λ and $\hat{\lambda}$ for bid and ask separately. For price range/time horizon in this initial run we chose the pairs $s = (T = 10, R = 5)$ and $l = (T = 3600, R = 200)$.

We perform for each stock a stepwise regression with inclusion probability of 10%. Table 1 and figure 1 provide an overview of how often each of the liquidity measures was included in the final model by providing descriptive statistics of the t-values for each liquidity measure across the 49 stepwise regressions. We see that the ask λ entering linearly for the $T = 3600$ time period was included in the final regression specification for all 49 stocks with an average t-value of 7.4, where the minimal t-value was 2.00. The same λ entered into 28 final specifications in the ratio form, with an average t-value of 3.5 and a minimum of 1.66.

What I conclude from this preliminary exercise is a strong correlation between long and short term liquidity measure λ and excess returns with some evidence for nonlinearity as indicated by liquidity measures entering as ratios in addition to linearly. Liquidity measured by $\hat{\lambda}$ does not seem to have particularly strong information content. For all measures the wider range, longer horizon measures entered more often than the narrower, short horizon measures.

Table 1: Summary statistics for |t-values| of factor model regression.

	N	Mean	Std. Dev.	Min	P10	Median	P90	Max
t_ask_ls	39	3.90	1.58	1.71	1.92	3.68	6.08	7.48
t_bid_ls	36	4.18	1.61	1.68	2.20	3.83	6.78	8.74
t_ask_ll	49	7.40	2.92	2.00	3.58	7.56	11.20	12.69
t_bid_ll	48	7.87	2.94	2.28	3.92	8.13	11.78	13.45
t_ask_lsh	15	2.58	0.96	1.71	1.77	2.09	3.99	4.68
t_bid_lsh	14	2.56	0.81	1.68	1.80	2.25	3.69	3.98
t_ask_llh	23	2.63	0.60	1.70	1.85	2.61	3.37	4.10
t_bid_llh	28	2.74	0.81	1.65	1.92	2.57	4.19	4.27
t_ask_rs	29	3.49	2.45	1.65	1.90	2.72	5.52	13.78
t_bid_rs	29	2.96	0.98	1.66	1.86	2.71	4.24	5.94
t_ask_rl	28	3.50	1.80	1.70	1.80	3.21	6.16	9.04
t_bid_rl	29	3.04	1.06	1.69	1.73	3.09	4.39	5.66
t_ask_rsh	18	3.93	2.90	1.65	1.90	3.08	8.51	13.78
t_bid_rsh	0	.	.	.				.
t_ask_rlh	0	.	.	.				.
t_bid_rlh	0	.	.	.				.

For the first letter, l denotes that the regressor entered directly into the regression, r denotes that the regressor entered in the denominator with the excess return in the numerator.
For the second letter, s denotes short time period of 10 seconds. l denotes long time period of 3600 seconds.
The h as the third letter indicates $\hat{\lambda}$, i.e. the φ coefficient, absence of h indicates λ, i.e. the ϕ coefficient.

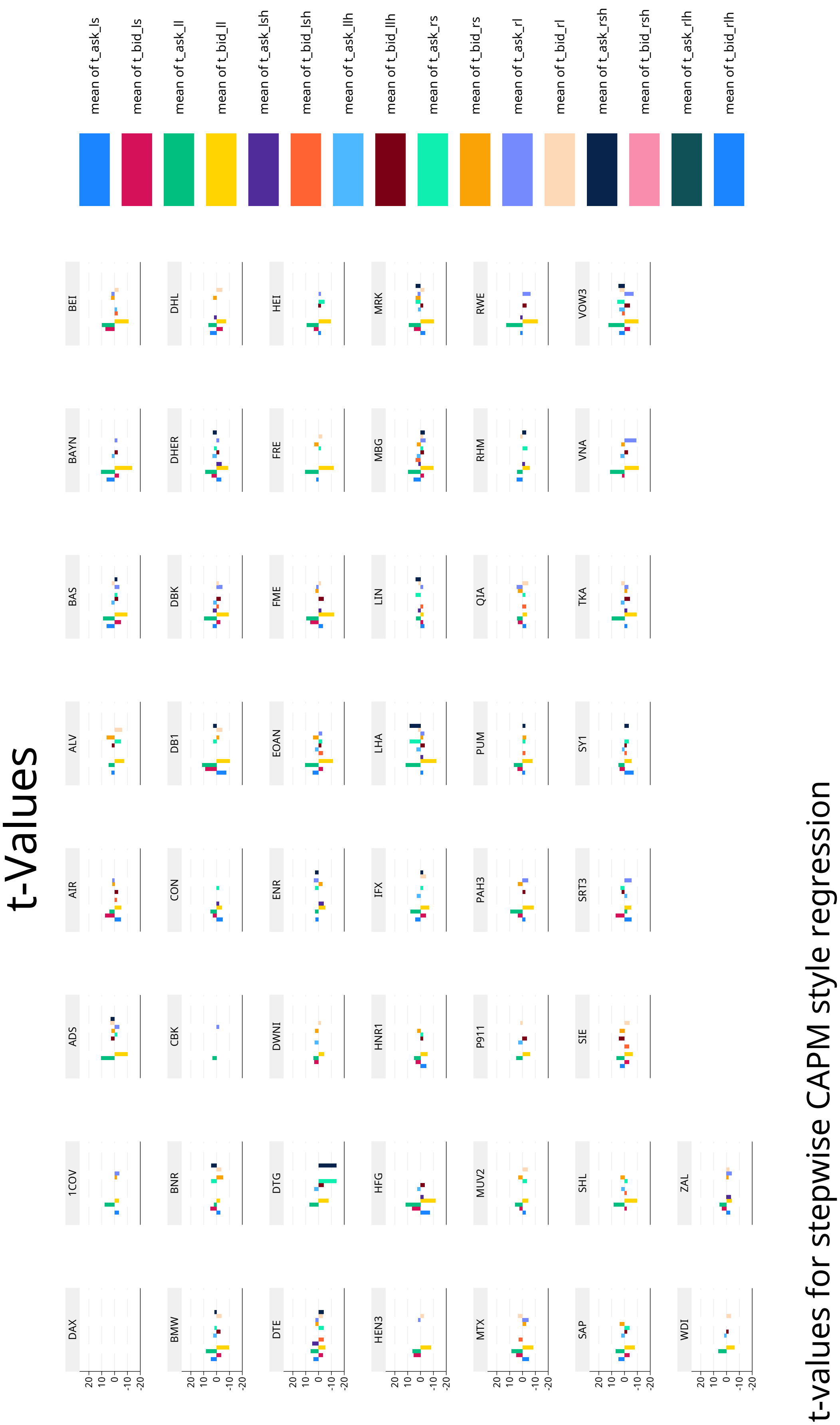

Figure 1: t-values for the factor model regression described in (1).