The Cross-Section of Price Efficiency

Ion Lucas Saru*

Very preliminary, please do not circulate.

March 28, 2024

Abstract

Inventory management by market makers can result in quoted prices deviating from unobserved fundamental prices. In a setting where prices have a factor structure, inventory management implies that pricing errors of different securities are positively correlated if they load on the same risk factors. Using a state space model, I obtain estimates of 1-minute pricing errors for a balanced panel 1500 stocks for the period 2016 – 2022. Daily cross-sectional regressions of pricing error correlations reveal a negative relationship between pricing error correlations and the difference in factor betas. ETF flows, rather than holdings, are associated with higher pricing error correlations.

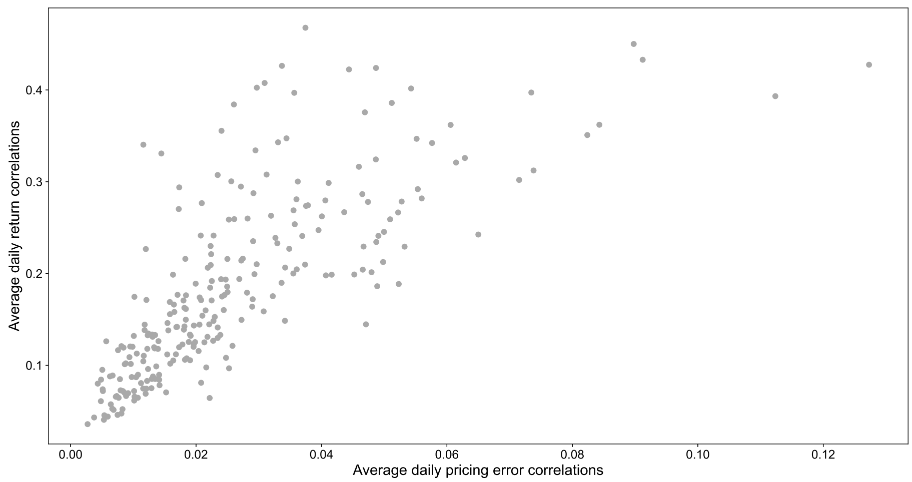
1 Introduction

In financial markets, trading prices can deviate from unobserved fundamental prices, or informationally efficient prices, due to pricing errors. The presence of such pricing errors can be motivated from a liquidity provider's inventory control problem (Foucault, Pagano, and Roell, 2013; Hendershott and Menkveld, 2014). Even though such pricing errors are transitory, they affect transaction costs of liquidity demanding traders. In this paper, I systematically analyze pricing errors in the cross-section stocks. Using a state space model, I identify pricing errors at the 1-minute frequency. Motivated by an inventory control model, I run cross-sectional regressions of correlations in pricing errors for stock pairs on the differences in their factor betas. Moreover, I analyze the relationship between correlations in pricing errors and liquidity demand by ETFs, motivated by a literature that shows that ETF flows are associated non-fundamental volatility

*Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands, saru.ionlucas@gmail.com, +31 20 598 6060, and Tinbergen Institute. I am grateful to Yacine Aït-Sahalia, Alexey Ivashchenko, and Albert Menkveld for their helpful comments. I thank SURF (www.surf.nl) for the support in using the National Supercomputer Snellius. Part of this research was conducted while I was visiting the Bendheim Center for Finance at Pricenton University.

Figure 1: Average unconditional daily return and pricing error correlations for large stocks in 2019

This figure plots average daily return correlations (vertical axis) and pricing error correlations (horizontal axis) for large stocks in 2019. Large stocks refers to stocks that are in the top three NYSE market capitalization deciles at the beginning of the year.



and correlations in daily illiquidity measures (Ben-David, Franzoni, and Moussawi, 2018; Agarwal et al., 2018).

Figure 1 plots average unconditional daily pricing error correlations on the horizontal axis and average daily return correlations on the vertical axis for large stocks in 2019. Large stocks are stocks that were in the top three deciles of NYSE market capitalization in the beginning of 2019. The relationship is positive: days on which pricing errors are on average more correlated are also days on which returns are more correlated. This suggests that pricing errors are systematic.

Institutional investors and mutual funds often trade a basket of stocks. Even though individual pricing errors in individual stocks may be small, such investors face pricing errors in transaction prices in multiple stocks. As Figure 1 suggests that pricing errors are systematic, individual pricing errors may not average out on aggregate. Moreover, investment mandates of institutional investors and mutual funds may cause them to trade in similar stocks in the sense that the stocks load on the same risk factors. This potentially aggravates the systematic impact of pricing errors.

To shed more light on the cross-section of pricing errors, I develop a simple inventory model based on Foucault, Pagano, and Roell (2013) in which market makers are setting quotes for stocks which prices, and therefore returns, are driven by a factor structure. As in standard inventory models, market makers adjust their quotes away from fundamental prices depending on the inventory they take, resulting in pricing errors. The underlying factor structure in asset prices connects different securities and affects the way market makers set prices in different securities. This results in pricing errors being correlated across securities to the extent that they load on the same risk factors.

Guided by the implications of the theoretical model, I empirically examine the correlations of pricing errors in the cross-section. Therefore, I construct a balanced panel of 1500 stocks spanning the years 2016 – 2022. Using a state space model (Menkveld, Koopman, and Lucas, 2007) I obtain estimates for 1-minute pricing errors for the stocks in my sample. For every stock pair I compute daily correlations in pricing errors. Daily Fama and MacBeth (1973) regressions of pricing error correlations on difference in the the stock's Fama and French (1993) factor betas reveal a negative relationship between pricing error correlations and the absolute difference in the stocks' factor betas. This suggests a systematic component in pricing error correlations. My results are robust to controlling for linear and nonlinear stock pair characteristics as well as differences in the microstructure of trading as captured by the difference in the stocks price and 5-minute return volatilities.

Motivated by a literature that documents a positive relationship between ETF flows and volatility and the co-movement of daily illiquidity measures(Ben-David, Franzoni, and Moussawi, 2018; Agarwal et al., 2018), I examine the role of liquidity demanding flows by ETFs for pricing error correlations. I focus both and common holdings by ETFs and the exposure of a stock pair to ETF flows and ETF creations and redemptions. Overall, the effect liquidity demand is an order of magnitude smaller than the effect of liquidity supply. While common holdings do not appear to move correlations in pricing errors, larger differences in relative ETF flows are associated with smaller pricing error correlations for a stock pair. This is intuitive as actual flows are part of the liquidity provider's maximization problem. Also, this is consistent with the finding of Antón and Polk (2014) that common ownership by funds is more relevant in periods of high flows.

Conrad and Wahal (2020) show that market risk is an important driver of inventory effects in assets prices. By systematically analyzing the cross-sectional effects of different risk factors on pricing errors, I contribute to the understanding of which risks drive intraday prices as well as price efficiency in securities.

My work relates to Rösch, Subrahmanyam, and Van Dijk (2017) who study systematic drivers of market efficiency over the time-series. In contrast, I study systematic drivers of price efficiency in the cross-section of stocks. My findings are consistent in that I find price efficiency to vary systematically in the cross-section of stocks. Relatedly, Chordia, Roll, and Subrahmanyam (2000) and Hasbrouck and Seppi (2001) study co-movement in liquidity.

My work is closely related to Hendershott and Menkveld (2014) who study price pressures induced by specialist inventories at a daily frequency. Rather than focusing on daily frequencies, I analyze intraday price pressures or pricing errors as a result of aggregate liquidity demand. Bogousslavsky and Collin-Dufresne (2023) study the impact of the volatility of order imbalances in the cross-section and time series.

Van Binsbergen et al. (2023) relate different asset pricing anomalies to the buildup and resolution of long-lasting price wedges. Their model defines price wedges as the deviation from a model-implied price. My research aims to identify pricing errors based on a top-down approach based on filtering at a high frequency.

By studying the effect of liquidity demanding ETF flows on correlations in pricing errors I relate to Lou and Polk (2021) who provide evidence that crowded trades by arbitrageurs drive prices away from fundamentals. Antón and Polk (2014) relate correlations in Fama and French (1993) – Carhart (1997) residuals to common ownership by funds. Ben-David, Franzoni, and Moussawi (2018) show that ETF flows increase volatility and Agarwal et al. (2018) relate common stock ownership by ETFs to co-movement in illiquidity.

The remainder of the paper is structured as follows. Section 2 presents a simple inventory model in which security prices are driven by a factor structure. Section 3 presents my methodology before I discuss the data in Section 4. I present my main empirical results in Section 5. Finally, Section 6 concludes.

2 A Simple Theoretical Model

In this section I present a simple inventory model for a market maker in a setting in which security prices, and therefore returns, are driven by a factor structure. The model follows the inventory model of Foucault, Pagano, and Roell (2013).

2.1 Setup

There are two periods, $t = 1$ and $t = 2$. I consider trading in N assets that load on M underlying risk factors. The assets have a payoff structure given by

$$V_i = \sum_{j=1}^{M} \theta_{ij} f_j \varepsilon_i, \quad i = 1, \dots, N \tag{1}$$

with

$$f_j = \mu_j + \eta_j, \quad j = 1, \dots, M. \tag{2}$$

The innovations $\varepsilon_i \sim N(0, \sigma_i^2)$, $\forall i$, $\eta_j \sim N(0, \sigma_{fj}^2)$, $\forall j$ are i.i.d. The payoff of each asset i has a factor structure in the sense that the payoff loads on the factors $f_1, \dots, f_M$. Moreover, there is an idiosyncratic component to the assets' given by ε_i. The payoff V_i of asset i is correlated with the payoff $V_{\tilde{i}}$, $\tilde{i} \neq i$, that is, the payoff of another asset, if any of the factor loadings θ_{ij}, $\theta_{\tilde{i}j}$ are not zero for at least one j.

There are K risk-averse dealers with CARA utility function and risk aversion ρ in the market. The dealer market is competitive. Dealers begin period $t = 1$

with inventory $z_i^k,\ i = 1, \ldots, N$ and a cash position c^k. The aggregate inventory position at the beginning of $t = 1$ is denoted by

$$Z_i = \sum_{k=1}^{K} z_i^k,\ i = 1, \ldots, N. \tag{3}$$

In period $t = 1$, trading takes place and each dealer trades with one client. Trading takes pace by each client submitting their order of quantity q_i. Each dealer posts a supply schedule $y_i^k(p)$. Market clearing requires that

$$\sum_{k=1}^{K} y_i^k(p_i) = q_i,\ i = 1, \ldots, N. \tag{4}$$

2.2 Dealer maximization

At date $t = 1$, each dealer chooses their supply schedule y_i^k for asset i that maximizes their expected utility at $t = 2$:

$$\mathbb{E}_1[-exp(-\rho W^k)] \tag{5}$$

with W^k being their final wealth at $t = 2$

$$W^k = c^k + \sum_{i=1}^{N} \left(V_i(z_i^k - y_i^k) + p_i y_i^k \right). \tag{6}$$

Given that the innovations ε_1, ε_2, η_1, and η_2 are normally distributed, this is equivalent to maximizing

$$\mathbb{E}_1[W^k] - \frac{\rho}{2}\mathbb{V}_1[W^k] \tag{7}$$

2.3 Equilibrium

In equilibrium, markets must clear such that

$$\sum_{k=1}^{K} y_i^k = q_{i0}, \quad i = 1, \ldots, N. \tag{8}$$

Dealer k's inverse supply schedule for asset i is derived from their maximization problem and given by

$$p_i = \sum_{j=1}^{M} \theta_{ij}\mu_j - \rho \left((z_i^k - y_i^k) \left(\sum_{j=1}^{M} \theta_{ij}^2 \sigma_{fj}^2 + \sigma_i^2 \right) + \sum_{\tilde{i} \neq i} (z_{\tilde{i}}^k - y_{\tilde{i}}^k) \left(\sum_{j=1}^{M} \theta_{ij}\theta_{\tilde{i}j}\sigma_{fj}^2 \right) \right) \tag{9}$$

As can be seen, the supply schedule for asset i depends not only on the dealer's inventory for asset i, but also on the dealer's inventory for the other assets. This is because all asset payoffs load on the same underlying factor structure and dealers care about the risks the individual assets are exposed to. The dependence of dealer k's inverse supply schedule for asset i on their inventory in asset $\tilde{i} \neq i$ is scaled by the factor loadings on the respective underlying factors in asset prices. In addition, the dealers' quoted price schedules differ since their starting inventories in $t = 1$ differ.

The dealer quotes a bid-ask spread per unit traded in asset i of

$$s = 2\rho \left(\sum_{j=1}^{M} \theta_{ij}^2 \sigma_{fj}^2 + \sigma_i^2 \right) \tag{10}$$

taking their inventory and the quantity in the other assets as given. Note that while the quoted price schedules of the dealers differ due to the inventories with which they enter period $t = 1$, spreads quoted by the different dealers are the same. This is the same results as in Foucault, Pagano, and Roell (2013) in the sense that spreads depend on the dealers' inventory at the begin of the period. Also in Bogousslavsky and Collin-Dufresne (2023) spreads are related to dealer inventory through order imbalance. The result differs from Hendershott and Menkveld (2014) in which spreads are orthogonal to dealer inventory.

Using the market clearing condition (8) yields that the equilibrium price is given by

$$\begin{aligned} p_i^* = \sum_{j=1}^{M} \theta_{ij}\mu_j - \bar{\rho} \Bigg(& (Z_i - q_i) \left(\sum_{j=1}^{M} \theta_{ij}^2 \sigma_{fj}^2 + \sigma_i^2 \right) \\ & + \sum_{\tilde{i} \neq i} (Z_{\tilde{i}} - q_{\tilde{i}}) \left(\sum_{j=1}^{M} \theta_{ij}\theta_{\tilde{i}j}\sigma_{fj}^2 \right) \Bigg) \end{aligned}$$

where Z_i and $Z_{\tilde{i}}$ are the aggregate inventories in assets i and $\tilde{i}$, respectively, as defined in (3), and $\bar{\rho} = \rho/K$ is the aggregate risk aversion coefficient of the dealers.

In equilibrium, dealer k trades

$$y^{k*}(p^*) = \frac{q}{K} + z^k - \frac{Z}{K}. \tag{11}$$

After trading took place, dealer k's inventory is therefore given by $\dfrac{Z}{K} - \dfrac{q}{K}$.

The midquote price for asset i before trading takes place can be defined as

$$m_i = \sum_{j=1}^{M} \theta_{ij}\mu_j - \bar{\rho}\left(Z_i \left(\sum_{j=1}^{M} \theta_{ij}^2 \sigma_{fj}^2 + \sigma_i^2 \right) + \sum_{\tilde{i} \neq i} Z_{\tilde{i}} \left(\sum_{j=1}^{M} \theta_{ij}\theta_{\tilde{i}j}\sigma_{fj}^2 \right) \right), \tag{12}$$

while the fundamental value of the fundamental value of the asset is equal to the conditional expected payoff of asset i, given by

$$m_i^* = \sum_{j=1}^{M} \theta_{ij}\mu_j. \tag{13}$$

The difference between the midquote and the fundamental value can be interpreted of the price pressure exerted by dealers on the price, or the pricing error in the observed quotes. This pricing error is given by

$$-\bar{\rho}\left(Z_i \left(\sum_{j=1}^{M} \theta_{ij}^2 \sigma_{fj}^2 + \sigma_i^2 \right) + \sum_{\tilde{i} \neq i} Z_{\tilde{i}} \left(\sum_{j=1}^{M} \theta_{ij}\theta_{\tilde{i}j}\sigma_{fj}^2 \right) \right) \tag{14}$$

As can be seen from (14), the pricing error is negative (the midquote is below the fundamental value of the asset) if dealers are on aggregate long in asset i, $Z_i > 0$. This is the standard results from the literature (for instance, Hendershott and Menkveld (2014)) that dealers adjust their quotes downward to facilitate selling if they have a long position in the asset. At the same time, the pricing error for asset i also depends on the dealer's aggregate position in the other assets $\tilde{i}$. The sign of this effect depends on the signs of the factor loadings $\theta_{ij}\theta_{\tilde{i}j}$, $j = i, \ldots, M$, $\tilde{i} \neq i$. The effect is always positive (that is, the pricing error increases) if all factor loadings are positive an the dealers have an aggregate long position in asset $\tilde{i}$, $Z_{\tilde{i}} > 0$.

Intuitively, in this case all assets load on the same risk factors and increase the dealers aggregate risk exposure. Therefore, the dealers want to facilitate selling asset i even more and set their midquote lower and make the pricing error less negative. If instead for some pair of factor loadings $\theta_{ij} > 0$, $\theta_{\tilde{i}j} < 0$ or $\theta_{ij} < 0$, $\theta_{\tilde{i}j} > 0$ (that is, the assets have opposite loadings on the risk factors), the dealers having an aggregate long position in asset $\tilde{i}$ increases the midquote and makes the pricing error less negative. This is because asset $\tilde{i}$ is a hedge for asset i regarding risk factor j. In this scenario, having a long position in asset $\tilde{i}$ decreases the overall risk exposure of the dealers to the risk factors.

Note that while the pricing error for asset i depends on both the dealers' positions all assets, the bid-ask-spread depends only on inventory in asset i. As a result, pricing errors in different assets co-move if they load on the same underlying risk factors and dealer inventory positions co-move, while spreads do not co-move with inventory positions. As a result, spreads in asset i are orthogonal to inventory in asset $\tilde{i}$.

3 Methodology

This section first describes the methodology used to identify pricing errors in Section 3.1, before turning the the empirical approach to model cross-sectional variation in pricing errors in Sections 3.2 and 3.3.

3.1 Identification of Pricing Errors

Central to the analysis is the identification of pricing errors at every point in time. Rather than relying on a model-based bottom-up approach to the identification of pricing errors as, for example, Van Binsbergen et al. (2023), I identify pricing errors using filtering based on the approach of Hasbrouck (1993) and the state space model of Menkveld, Koopman, and Lucas (2007) in a top-down way.

Observed (log) prices for stock i at time t can be decomposed into two latent components: a martingale efficient price component and a stationary pricing error:

$$p_{i,t} = m_{i,t} + s_{i,t} \tag{15}$$

$$m_{i,t} = m_{i,t-1} + w_{i,t} \tag{16}$$

where p_t are observed (midquote) prices, m_t are efficient prices, w_t are innovations in efficient prices, and s_t is the pricing error. To identify innovations in efficient prices I incorporate information on trade flow. This yields the model:

$$p_{i,t} = m_{i,t} + s_{i,t} \tag{17}$$

$$m_{i,t} = m_{i,t-1} + \kappa \tilde{x}_{i,t} + \mu_{i,t} \tag{18}$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi x_{i,t} + \nu_{i,t} \tag{19}$$

where (17) is the observation equation, (18) is the state equation for latent efficient prices, and (19) captures latent pricing errors. Furthermore, it is assumed that $\nu_{i,t} \sim \mathcal{N}(0, \sigma^2_{i,\nu})$ and $\mu_{i,t} \sim \mathcal{N}(0, \sigma^2_{i,\mu})$. As discussed in Hendershott and Menkveld (2014) and Menkveld and Saru (2024), innovations in order flow contain information while imbalances in order flow affect liquidity (Brandt and Kavajecz, 2004; Evans and Lyons, 2008). Therefore, including signed order flow allows identifying pricing errors in the model. The identifying assumption is that conditional on controlling for trade flow, $\nu_{i,t}$ and $mu_{i,t}$ are uncorrelated. This is in line with similar applications of the model such as Hendershott and Menkveld (2014) and Brogaard, Hendershott, and Riordan (2014), among others.

The model is estimated stock-day-by-stock-day at a 1-minute frequency. Midquote prices are used as observed proxy for efficient prices. By using midquote prices, bid-ask bounces are not part of the pricing error. Trade flow $x_{i,t}$ is the signed order flow for every 1-minute interval.[1] Innovations in trade flow $\tilde{x}_{i,t}$ are obtained as the residual from an AR model with 15 lags. Using aggregate

[1] I discuss details on trade signing in Section 4.

trade flow captures aggregate liquidity demand in the market. Furthermore, it can be interpreted as the aggregate inventory that all market makers in stock i absorb. This is aligned with the theoretical model presented in Section 2 in which midquotes and pricing errors depend on the aggregate position of all market makers. Moreover, using signed aggregated order flow captures order flow information that is in principle available to market participants subscribing to data feeds (George and Khoja, 2023).

The model is estimated by maximum likelihood and the Kalman filter is used to evaluate the likelihood function. The Kalman filter requires initial priors for the latent states characterized by a prior mean and a prior variance. I initialize the efficient price series with a diffuse prior. That is, the prior variance is set to κ with $\kappa \to \infty$. Pricing errors are initialized as stationary states with the prior variance set to the unconditional variance. After obtaining estimates for the coefficients, I obtain estimates of the latent state variables – the efficient price as well as the pricing error – as smoothed states of the model. These smoothed states are conditional on all observations on the respective trading day in stock i. The smoothed states for the pricing error equation (19) are central to my analysis of pricing errors in the cross section of stocks.

Even though I require that there are at least 200 trading days per year with non-zero volume, there are time intervals with zero or little volume as well as quotes in the sample. My proposed state space modeling naturally deals with missing observations through extrapolation (Kalman filter) and interpolation (for the smoothed states).

3.2 Factor Structure in Returns

The theoretical model in Section 2 predicts that pricing errors of different stocks are connected through the stocks' loadings on underlying risk factors. To test this implication empirically, I run factor regressions to estimate the factor betas of different stocks. In the baseline version, I use the factors proposed by Fama and French (1993): the market factor, the size factor (SMB), and the value factor (HML).

I estimate the factor betas by running daily rolling time-series regressions of excess returns on the factors. The last 500 available return observations in CRSP are used for the rolling regressions. If there is a missing return observation in CRSP for a trading day within the estimation window, the respective day is left out of the estimation but estimation is still performed for the overall window.

3.3 Cross-Sectional Variation in Pricing Error Correlations

Using the pricing error estimates obtained as smoothed states from the state space model, I compute the realized correlations in pricing errors for each stock pair in my sample, $\rho^s_{ij,t}$. Based on the factor betas for each stock pair, I compute the the absolute difference in their factor betas as $|\Delta\beta_{ij,t}|$. I then estimate daily cross sectional regressions for the realized correlations as

$$\begin{aligned}\rho^s_{ij,t} = a + b_0\rho^x_{ij,t} + b_1|\Delta\beta^M_{ij,t-1}| + b_2|\Delta\beta^{SMB}_{ij,t-1}| \\ + b_3|\Delta\beta^{HML}_{ij,t-1}| + \mathbf{X}\gamma + \varepsilon_{ij,t},\end{aligned} \tag{20}$$

where $\rho^x_{ij,t}$ is the correlation in order flow for stock pair i,j, $|\Delta\beta^M_{ij,t-1}|$, $|\Delta\beta^{SMB}_{ij,t-1}|$, and $|\Delta\beta^{HML}_{ij,t-1}|$ are the absolute differences in the loadings on the Fama and French (1993) factors estimated based on a rolling regression as described in Section 3.2. Other included control variables are denoted by $\mathbf{X}\gamma$. All independent variables are standardized to facilitate economic interpretability.

I implement the estimation in a Fama and MacBeth (1973) fashion. Therefore, I estimate the cross-sectional regressions day-by-day and report time-series averages of the daily coefficients. I report Newey and West (1987) standard errors that are robust to time-series autocorrelation in the cross-sectional estimates of 20 lags.

This implementation is similar to Antón and Polk (2014) who estimate monthly regressions for realized correlations in four-factor abnormal returns in a Fama and MacBeth (1973) fashion. Agarwal et al. (2018) estimate a pooled regression for quarterly correlations in stock liquidity. In my setting, daily regressions in are Fama and MacBeth (1973) style a preferred as market conditions potentially driving the realized correlations only change slowly. Moreover, I estimate both the state space model as well as the cross-sectional regressions on a day-by-day level. Aligning the frequencies alleviates concerns of mechanical correlations of the dependent and independent variables in equation (20) over different trading days. Moreover, the approach of obtaining smoothed states from the state space model an estimating their relationship with independent variables is consistent with Chordia, Green, and Kottimukkalur (2018).

Correlations in order flow for stock pair i,j on day t, $\rho^x_{ij,t}$, are included in the model as the order flow series identify pricing errors in the state space model presented in Section 3.1. The inclusion of the absolute differences of the factor betas is motivated by the theoretical model presented in Section 2. The model predicts that as the absolute difference in the factor betas increases, correlations should decrease. By controlling for correlations in the order flow series, the differences capture correlation in pricing errors above and beyond correlation in pricing errors that is due to a potential factor structure in the order flow series.

Additional control variables contained in $\mathbf{X}\gamma$ are the absolute difference in log midquotes of stock pair i,j on date $t-1$,[2] the 5-minute midquote return volatilities of stocks i and j on date $t-1$, and the factor betas of stocks i and j as well as squared factor betas. I refer to the latter as linear and nonlinear characteristics controls. Including the difference in the midquote prices captures differences in the participation of traders as well as differences in liquidity related

[2] I run additional regressions confirming that results do not change if both the midquotes of stocks i and j are included as control variables.

to the price level and the relative tick size (Weller, 2018; Li and Ye, 2023). 5-minute midquote return volatilities account for potential differences in liquidity as a result of volatility (Conrad and Wahal, 2020). Including the factor betas as well as their quadratic terms captures potential differences in the stocks that are related to their betas and follows Antón and Polk (2014).

The time convention of the independent variables in equation (20) capture which information is available to market makers on day t. As discussed before, correlations in order flow on day t, $\rho^x_{ij,t}$, are included as they mechanically drive correlations in pricing errors. The best estimate of the factor exposure of stock returns available to the market maker are the factor betas estimated including all return observations until day $t-1$. Therefore, these betas are included in the model. For the otehr control variables, observations for $t-1$ are included.

4 Data

The data used in the main analysis comes from TAQ and CRSP. I construct a balanced panel based on all common stocks listed on NYSE, Amex and NASDAQ from CRSP. The sample period spans January 1, 2016 – December 31, 2022.

For a stock to be included in the sample, I require a market capitalization of at at least USD 100 million and a stock price between USD 5 and USD 1000 at the end of the previous month. This is in line with, for example Hendershott and Menkveld (2014) or Bogousslavsky and Collin-Dufresne (2023). Furthermore, for a security to be included in the analysis I require the security to have at least 200 trading days displaying non-zero volume per year (Duarte, Hu, and Young, 2020). My final sample contains 1500 securities.

For each year, securities are assigned to market capitalization deciles based on the market capitalization at the end of the previous year. I use the market equity breakpoints for NYSE from Ken French's website for this.

TAQ data is cleaned according to the filters proposed by Holden and Jacobsen (2014). Trades are signed using the Lee and Ready (1991) algorithm given its documneted performance (Chakrabarty, Pascual, and Shkilko, 2015). Data from CRSP and TAQ is merged using the TCLINK linking files from WRDS. For the analysis, only observations within trading hours between 9:30 a.m. and 4:00 p.m. are kept. As my interest is in intraday pricing errors that are potentially driven by high-frequency order imbalances Bogousslavsky and Collin-Dufresne (2023), I work with a data frequency of 1 minute.

The factor regressions use daily return data from CRSP. The daily return factors as well as the risk-free rate come from Ken French's website.

Descriptive statistics for my sample by year and market capitalization are presented in Table 1. The descriptive statistics suggest that there is considerable variation in key variables both in the time-series as well as in the cross-section. It is notable that both for high market capitalization stocks as well as for medium market capitalization stocks effective spreads increase over the

Table 1: Descriptive statistics

This table presents descriptive statistics for the stocks in the sample by calendar year and market capitalization. High market capitalization refers to stocks that are in the top three NYSE market capitalization deciles at the beginning of the respective year. Medium market capitalization refers to stocks that are in the middle four NYSE market capitalization deciles at the beginning of the respective year. Similarly, low market capitalization refers to stocks that are in the bottom three NYSE market capitalization deciles at the beginning of the respective year. $midquote_{it}$ refers to the dollar midquote and is from TAQ. $shares\_outst_{it}$ refers to the number of shares outstanding (in million) and is from CRSP. $market\_cap_{it}$ refers to the market capitalization in millions of dollars and is from CRSP. $espread_{it}$ refers to the share-volume-weighted effective spread in basis points and is from TAQ. $volatility_{it}$ refers to the 5-minute midquote return volatility in basis points as is computed based on data from TAQ. $volume_{it}$ refers to the daily trading volume in millions of dollars and is from TAQ.

Panel A: Overall sample

	2016	2017	2018	2019	2020	2021	2022
$midquote_{it}$	50.65	60.81	69.12	71.60	75.15	101.82	89.67
$shares\_outst_{it}$	240.13	239.17	238.02	235.03	234.49	236.33	234.37
$market\_cap_{it}$	11 691.33	13 774.95	15 597.21	16 319.71	18 161.62	24 458.70	23 381.66
$espread_{it}$	18.06	16.22	17.08	15.33	19.67	16.66	16.30
$volatility_{it}$	20.00	17.07	20.24	18.61	31.51	21.31	23.67
$volume_{it}$	71.79	73.93	96.36	85.65	128.73	134.67	139.22

Table 1: – continued

Panel B: High market capitalization							
	2016	2017	2018	2019	2020	2021	2022
$midquote_{it}$	82.94	96.99	111.69	121.11	139.77	191.28	161.63
$shares\_outst_{it}$	707.63	700.03	670.15	653.24	642.22	653.67	623.42
$market\_cap_{it}$	37 336.60	43 567.86	47 673.79	49 684.72	55 965.70	76 861.23	70 348.92
$espread_{it}$	4.46	3.82	4.35	4.24	6.30	5.48	5.86
$volatility_{it}$	14.27	11.31	15.58	13.80	23.56	15.89	20.11
$volume_{it}$	217.27	218.61	282.51	247.61	382.22	405.80	408.16

Panel C: Medium market capitalization							
	2016	2017	2018	2019	2020	2021	2022
$midquote_{it}$	46.49	56.42	60.91	61.24	58.35	81.14	71.94
$shares\_outst_{it}$	83.67	84.08	82.38	79.74	83.54	93.50	86.94
$market\_cap_{it}$	2684.90	3303.66	3502.74	3352.47	3202.13	4902.85	4130.28
$espread_{it}$	10.10	9.71	10.23	9.38	13.09	11.52	11.06
$volatility_{it}$	19.30	16.17	19.87	18.39	31.45	21.26	23.82
$volume_{it}$	24.67	27.48	29.90	26.28	31.66	38.04	32.55

Table 1: – continued

Panel D: Low market capitalization							
	2016	2017	2018	2019	2020	2021	2022
$midquote_{it}$	27.70	35.03	38.97	34.73	30.56	43.34	38.42
$shares\_outst_{it}$	26.99	28.71	28.39	27.72	27.65	30.14	30.25
$market\_cap_{it}$	523.25	697.96	756.42	665.03	581.46	902.30	790.50
$espread_{it}$	39.77	34.58	38.53	36.29	45.58	35.11	34.79
$volatility_{it}$	25.87	23.06	25.23	24.09	40.51	26.74	27.25
$volume_{it}$	3.48	5.20	5.27	4.18	5.47	6.23	4.81

sample period. Dollar trading volume increases over the sample period for all market capitalizations.

Given the cross-sectional differences documented in Table 1, I report results from estimating the state space model by year and market capitalization and perform my cross-sectional analysis both for the overall sample as well as for only high market capitalization stocks. This alleviates concern that my results are driven by small, irregularly traded stocks.

5 Results

In this section I present the main results. First, I discuss the results from estimating the state space model in Section 5.1. Then, I discuss the cross-sectional results in Section 5.2 before turning to the relationship with liquidity demand in Section 5.3.

5.1 State Space Results

The theoretical model presented in Section 2 suggests that pricing errors are more correlated for stocks that load on the same underlying risk factors. In order to empirically test this prediction, I first need estimates of pricing errors for every point in time. As described in Section 3, I obtain pricing error estimates as smoothed states form a state space model. I present the estimation results from estimating the state space model in Table 2.

In light of the descriptive statistics presented in Section 4, I present results by year and market capitalization. The volatilities of efficient price innovations – permanent volatility – as well as pricing error innovations – residual pricing error volatility – increase for lower market capitalization deciles. This holds across all years in the sample and is consistent with the results of Brogaard, Hendershott, and Riordan (2014) and Hendershott and Menkveld (2014). Note that these are only one component of the volatility of efficient price innovations and pricing errors, respectively. Both are also driven by the volatility of order flow innovations and order flow, respectively.

Order imbalance innovations are informative across all market capitalizations for all years, as indicated by the the positive κ. In the setting, order imbalances can be interpreted as the aggregate order flow faced by liquidity providers on aggregate. That is, if the aggregate order imbalance is positive, a liquidity providers on aggregate take a short position. The positive κ suggests that liquidity providers are on aggregate subject to adverse selection. This result is consistent with Menkveld and Saru (2024) who show that clients are relatively more informed than intermediaries at lower intraday frequencies.

In addition, order imbalances are positively associated with pricing errors for high and medium market capitalization stocks (ψ is positive). This result suggests that, on aggregate, liquidity providers raise their midquotes relative to the latent efficient price in response to a positive order imbalance (that is, when they sell on aggregate). This result is consistent with the theoretical model

Table 2: State space model estimation results

This table presents estimation results for the state space model

$$p_{i,t} = m_{i,t} + s_{i,t}, \qquad m_{i,t} = m_{i,t-1} + \kappa\tilde{x}_{i,t} + \mu_{i,t}, \qquad s_{i,t} = \phi_i s_{i,t-1} + \psi x_{i,t} + \nu_{i,t}$$

for each stock-day using log midquote prices as observable prices ($p_{i,t}$). x_t is order flow and $\tilde{x}_t$ are surprises in order flow obtained as the residual from an AR(10) model. The table reports average coefficients by year and market capitalization. High market capitalization refers to stocks that are in the top three NYSE market capitalization deciles at the beginning of the respective year. Medium market capitalization refers to stocks that are in the middle four NYSE market capitalization deciles at the beginning of the respective year. Similarly, low market capitalization refers to stocks that are in the bottom three NYSE market capitalization deciles at the beginning of the respective year. σ_μ and σ_ν are in *bp* and κ as well as ψ in $bp/1,000,000$ USD. Standard errors are double clustered by stock and day and reported in parentheses. * denotes significance at the 1% level.

	2016			2017			2018		
	high	medium	low	high	medium	low	high	medium	low
σ_μ	4.621*	6.164*	7.767*	3.696*	5.230*	7.122*	5.280*	6.522*	7.767*
	(0.120)	(0.139)	(0.151)	(0.069)	(0.084)	(0.108)	(0.127)	(0.147)	(0.203)
σ_ν	1.708*	2.580*	5.040*	1.418*	2.236*	4.387*	1.814*	2.445*	4.874*
	(0.052)	(0.070)	(0.173)	(0.030)	(0.045)	(0.130)	(0.052)	(0.071)	(0.230)
ϕ	0.252*	0.221*	0.189*	0.214*	0.195*	0.183*	0.194*	0.176*	0.170*
	(0.006)	(0.005)	(0.004)	(0.007)	(0.004)	(0.003)	(0.008)	(0.005)	(0.004)
κ	8.925*	75.117*	521.425*	6.836*	62.034*	448.182*	7.783*	67.519*	492.269*
	(0.546)	(2.662)	(64.547)	(0.294)	(2.500)	(38.002)	(0.420)	(4.101)	(28.562)
ψ	3.994*	19.610*	−143.012*	3.285*	11.808*	−23.798	3.755*	14.523*	−74.780*
	(0.245)	(1.043)	(37.006)	(0.126)	(0.788)	(29.096)	(0.190)	(1.104)	(15.657)

Table 2: – continued

	2019			2020			2021		
	high	medium	low	high	medium	low	high	medium	low
σ_μ	4.598*	6.052*	7.339*	7.711*	10.045*	12.223*	5.304*	7.078*	8.362*
	(0.082)	(0.092)	(0.129)	(0.286)	(0.324)	(0.389)	(0.112)	(0.133)	(0.154)
σ_ν	1.686*	2.447*	4.878*	2.724*	4.080*	7.909*	1.873*	2.775*	5.457*
	(0.043)	(0.065)	(0.224)	(0.135)	(0.227)	(0.497)	(0.047)	(0.082)	(0.214)
ϕ	0.229*	0.191*	0.154*	0.253*	0.206*	0.162*	0.230*	0.180*	0.161*
	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)
κ	7.202*	63.608*	535.538*	11.392*	104.811*	1032.641*	7.533*	61.343*	496.751*
	(0.429)	(2.323)	(28.520)	(0.688)	(4.545)	(74.845)	(0.523)	(2.396)	(23.911)
ψ	3.588*	18.244*	−95.977*	5.641*	31.508*	−184.126*	3.770*	17.434*	−61.809*
	(0.226)	(0.795)	(18.991)	(0.299)	(1.959)	(40.247)	(0.175)	(0.940)	(10.659)

Table 2: – continued

	2022		
	high	medium	low
σ_μ	6.915*	8.041*	8.522*
	(0.138)	(0.144)	(0.151)
σ_ν	2.293*	2.742*	5.255*
	(0.061)	(0.067)	(0.218)
ϕ	0.228*	0.176*	0.146*
	(0.009)	(0.006)	(0.004)
κ	8.706*	72.991*	618.215*
	(0.486)	(2.326)	(33.424)
ψ	4.543*	20.383*	−85.716*
	(0.184)	(0.807)	(16.422)

presented in Section 2 as well as the literature, for instance Hendershott and Menkveld (2014). In response to taking a net short position, market makers want to facilitate selling to them and discourage more buying from them. They achieve this by raising both the bid and the offer price, driving up the midquote price.

It is notable that the impact of order imbalances on pricing errors is negative for small stocks across all years (ψ is negative). This is inconsistent with the previously described channel of market makers facilitating selling to them in response to taking a short position. Rather, a negative ψ suggests that market makers lower both their bid and offer prices in response to taking a short position. Two points are worth noting. First, there is relatively more heterogeneity in the stock-day estimates for low market capitalization stocks than for medium and high market capitalization stocks as can be seen by the relatively larger standard errors. Second, small stocks are traded less frequently than large and medium sized stocks (see Table 1). A potential explanation for the negative impact of order imbalances on pricing errors is an information channel as in Glosten and Milgrom (1985) rather than an inventory channel. Liquidity providers are subject to adverse selection and the direction of the order imbalance as well as the trading price contain information. As a results, liquidity providers update their quotes in the direction of the trade. Moreover, this is consistent with liquidity demanding trades executing against possibly stale quotes (Budish, Cramton, and Shim, 2015; Aquilina, Budish, and O'Neill, 2021).

While these results give an overview of the volatility of pricing errors in the cross-section as well as of the reaction of pricing errors to imbalances in order flow, they do not speak to the cross-sectional relationship between pricing errors. I address this in the next Section.

5.2 Cross-Sectional Results

Table 3 reports cross-sectional results for daily correlations in pricing errors. Based on the evidence presented in Tables 1 and 2, I report results for the full sample (Panel A) as well as for high market capitalization stocks only (Panel B).

In the first column, I present results for a reduced-form regression on specification (20), with only a constant and the correlation in signed order flow (that is, the correlation in the order imbalance) as independent variable. This specification serves as a benchmark for all additional specifications as daily correlations are mechanically a function of daily order flow correlations. This is because in the state space model the order flow series as well as innovations in order flow are used to identify pricing errors and efficient price innovations, as described in Section 3. For better comparability of the results, I standardize correlations in signed order flow. As expected, pricing error correlations are positively related to correlations in signed order flow. This effect is stronger for the full sample than for large market capitalization stocks.

In the second column, I add standardized differences between the factor betas for stock pair i, j. Consistent with the the theoretical model in Section 2,

Table 3: Cross-Sectional Results for Pricing Error Correlations

This table reports Fama and MacBeth (1973) estimates of daily cross-sectional regressions for correlations in pricing errors:

$$\rho^s_{ij,t} = a + b_0\rho^x_{ij,t} + b_1|\Delta\beta^M_{ij,t-1}| + b_2|\Delta\beta^{SMB}_{ij,t-1}| + b_3|\Delta\beta^{HML}_{ij,t-1}| + \mathbf{X}\gamma + \varepsilon_{ij,t}.$$

Pricing errors are obtained as smoothed states from estimating the state space model on the stock-day level. All independent variables are standardized to facilitate economic interpretability. $\rho^x_{i,j,t}$ is the correlation in the signed order flow series, $|\Delta\beta^M_{i,j,t-1}|$ is the absolute difference in the market betas of stocks i and j, obtained from a factor regression using the last 500 available return observations. Similarly, $|\Delta\beta^{SMB}_{i,j,t-1}|$ is the absolute difference in the SMB betas, and $|\Delta\beta^{HML}_{i,j,t-1}|$ is the absolute difference in the HML betas of stocks in i and j. $|\Delta m_{t-1}|$ is the absolute difference in the midquotes of stocks i and j on trading day $t-1$, and $\sigma(r^{5min}_{i,t-1})$ as well $\sigma(r^{5min}_{j,t-1})$ are the 5-minute midquote return volatilities for stocks i and j on trading day $-t1$. The specifications that control for linear and nonlinear characteristics include the (standardized) factors betas as well as squared terms of the factor betas for stocks i and j. I report Newey and West (1987) standard errors robust to autocorrelation of up to 20 lags in the cross-sectional estimates. Standard errors are reported in parentheses. * denotes significance at the 1% level.

	Panel A: Full Sample			
	(1)	(2)	(3)	(4)
Constant	0.019 08*	0.019 32*	0.019 30*	0.014 91*
	(0.000 74)	(0.000 76)	(0.000 76)	(0.000 76)
$\rho^x_{i,j,t}$	0.008 68*	0.002 96*	0.002 90*	0.002 65*
	(0.001 01)	(0.000 07)	(0.000 07)	(0.000 06)
$\|\Delta\beta^M_{i,j,t-1}\|$		−0.002 14*	−0.001 96*	−0.002 32*
		(0.000 11)	(0.000 10)	(0.000 10)
$\|\Delta\beta^{SMB}_{i,j,t-1}\|$		−0.001 63*	−0.001 45*	−0.002 06*
		(0.000 08)	(0.000 08)	(0.000 07)
$\|\Delta\beta^{HML}_{i,j,t-1}\|$		−0.000 52*	−0.000 19	−0.003 50*
		(0.000 18)	(0.000 19)	(0.000 15)
$\|\Delta m_{t-1}\|$			−0.000 78*	−0.000 63*
			(0.000 06)	(0.000 04)
$\sigma(r^{5min}_{i,t-1})$			−0.001 67*	−0.002 56*
			(0.000 13)	(0.000 10)
$\sigma(r^{5min}_{j,t-1})$			−0.001 66*	−0.002 45*
			(0.000 11)	(0.000 10)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes

Table 3: – continued

	Panel B: High Market Capitalization			
	(1)	(2)	(3)	(4)
Constant	0.027 26*	0.027 26*	0.027 26*	0.019 18*
	(0.001 24)	(0.001 24)	(0.001 24)	(0.001 36)
$\rho^x_{i,j,t}$	0.003 10*	0.002 88*	0.002 86*	0.002 63*
	(0.000 08)	(0.000 08)	(0.000 08)	(0.000 08)
$\lvert\Delta\beta^M_{i,j,t-1}\rvert$		−0.004 11*	−0.004 21*	−0.005 26*
		(0.000 20)	(0.000 20)	(0.000 23)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$		−0.002 51*	−0.002 46*	−0.003 31*
		(0.000 12)	(0.000 13)	(0.000 17)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$		−0.002 13*	−0.002 09*	−0.006 58*
		(0.000 27)	(0.000 28)	(0.000 27)
$\lvert\Delta m_{t-1}\rvert$			−0.000 45*	−0.000 49*
			(0.000 09)	(0.000 08)
$\sigma(r^{5min}_{i,t-1})$			−0.000 07	−0.002 66*
			(0.000 17)	(0.000 14)
$\sigma(r^{5min}_{j,t-1})$			−0.000 50*	−0.002 68*
			(0.000 16)	(0.000 14)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes

the daily pricing error correlation for stock pair i, j decreases as the difference in the Fama and French (1993) factor betas increases.

The effect is both economically and statistically significant. The coefficients on the correlation in signed order flow and the differences in factor betas are of a similar magnitude. A one-standard deviation increase in the correlation in signed order flow has a comparable effect to a one-standard deviation decrease in the difference between the market betas. For the full sample results, the effect is strongest for the market betas. The effect is weaker for the SMB betas as well as the HML betas. In addition, the effect is similar in magnitude to the effect of common fund ownership on the correlation in Fama and French (1993) – Carhart (1997) residuals documented by Antón and Polk (2014). The finding that the economic effect is the strongest for differences in market betas is in line with the findings of Conrad and Wahal (2020) that market risk drives inventory effects. I show that differences in the exposure to market risk drive price efficiency in the cross-section. In addition, market risk is not the only systematic risk factor driving inventory effects in the cross-section of stocks.

Comparing the full sample results with the results for high market capitalization stocks only reveals interesting differences. First, the results are overall stronger for the high-market capitalization stocks than for the full sample. Second, the market factor still appears to be the most relevant factor. At the same time, however, the effect for the SMB betas as well as for the HML betas increases, both in absolute terms as well as relative to the impact of correlations in signed order flow. These results alleviate the concern that my results are driven by small, irregularly traded stocks. In contrast, they appear to be driven by large, regularly traded stocks.

In the third column, I add controls for the differences in midquote prices on the previous trading day as well as the for the 5-minute return volatilities of stocks i and j. In addition, I add linear and nonlinear characteristic controls in column 4, following the specifications of Antón and Polk (2014). Overall, adding these controls leaves the results unchanged for both the full sample and for high market capitalization stocks only. If anything, the effect gets stronger once the full set of linear and nonlinear characteristic controls is added.

The literature suggests that different types of traders may be more active in a subset of stocks, dependent on the stocks price as the relative tick size decreases in the trade price (Weller, 2018; Li and Ye, 2023). For instance, Weller (2018) uses the previous stock price as an instrument for the activity of algorithmic and high-frequency traders. Suppose market making and liquidity provision are fragmented in the cross-section of stocks, market makers are limited in their risk bearing capacity, face capital constraints (as in Stoll (1978)), and/or are only imperfectly able to trade with each other as in the theoretical model in Section 2. Then, one would expect pricing error correlations to be larger among the subset of stocks in which the respective market makers are active. If this fragmentation occurs proportional to trading prices, this would be captured by the difference in the midquote prices. My result that an increase in the difference between the last midquote prices on the previous trading day is associated with a reduction in the correlation in pricing errors is consistent with this intuition.

Additionally, I include standardized midquote prices rather than the difference in midquote prices in untabulated results. I find that an increase in midquote prices of stock pair i, j is associated with a reduction in the pricing error correlation for the stock pair. This holds both for the overall sample as well as for high market capitalization stocks only. To the extent that the past stock price proxies for algorithmic and high-frequency trading activity, this is consistent with Brogaard, Hendershott, and Riordan (2014) and Hendershott, Jones, and Menkveld (2011) who document that high-frequency trading is associated with an improvement in liquidity and price efficiency.

The result that the effect is stronger for high market capitalization stocks is consistent across all specifications. A possible explanation is that as these stocks are more frequently traded, liquidity providers take positions more frequently. As a result, they adjust their quotes more frequently in response to their positions, resulting both more in frequent as well as more frequently changing pricing errors.

This results is relevant for institutional investors trading in a basket of these stocks. Even though these stocks are liquid as measured by the bid-ask spread, the pricing errors are correlated in the cross-section. Moreover, pricing errors are systematic as shown in Figure 1.

The finding that the difference in market betas has the strongest economic effect is consistent with the literature documenting common factors in the time-series of liquidity as well as price efficiency metrics (Conrad and Wahal, 2020; Rösch, Subrahmanyam, and Van Dijk, 2017). My results show that common factor also drive the cross-section of price efficiency.

5.3 Relationship with Liquidity Demand

The results in the previous sections are consistent with a liquidity supply channel as in the theoretical model presented in Section 2. However, Ben-David, Franzoni, and Moussawi (2018) show that ETFs increase the volatility in their underlying securities through an arbitrage channel. Agarwal et al. (2018) show that common ETF ownership increases stock-level illiquidity to co-move. This raises the question whether the findings documents before are due to a liquidity demand rather than liquidity supply channel. This could be the case if ETF flows have a strong factor structure and marker maker set their quotes both facing their current inventory positions as well as anticipating future flows due to ETF arbitrage.

In this channel I explore the possibility that the results are indeed driven by a liquidity demand channel. Given the findings documented in the previous literature, I focus particularly on ETF ownership. Therefore, I obtain daily data on ETF constituents and ETF flows from ETF Global. Starting from the universe of all ETFs in ETF Global, I keep only ETFs classified as equity ETFs and drop levered ETFs, active ETFs, and ETNs. On average, there are 1,213 unique ETFs per month in my sample, with an upward trend over the sample period. This is comparable to the coverage of Agarwal et al. (2018). ETF Global contains daily data on ETF constituents, the weight of the constituents in the ETF, the number of shares held in the constituent, and aggregate ETF flows. Constituents in ETF Global are identified by their CUSIP. I merge the ETF constituents to their CRSP PERMNO based on historical CUSIP information in CRSP MSENAMES.

I compute several measures of (differences in) stock exposure to ETF ownership. First, I compute a measure capturing the exposure of stocks i and j to ETF flows. Therefore, I compute the inflow (outflow) into (out of) stock i as a result of its exposure to ETF ownership, scaled by stock i's market capitalization:

$$FLOW_{i,t} = \frac{\sum_{f=1}^{F} w_{i,t}^{f} FLOW_{t}^{f}}{S_{i,t}P_{i,t}} \tag{21}$$

where $w_{i,t}^{f}$ denotes the weight of stock i in ETF f and $FLOW_{t}^{f}$ is the inflow (outflow) experienced by ETF f on day t. To capture the difference in exposure of stocks i and j to ETF flows, I compute the absolute difference

$$|\Delta FLOW_{ij,t}| = |FLOW_{i,t} - FLOW_{j,t}|. \tag{22}$$

As the other independent variables, I standardize $|\Delta FLOW_{ij,t}|$ for every day. In a separate specification, I include standardize absolute exposures of stocks i and j to ETF flows, $|FLOW_{i,t}|$ and $|FLOW_{j,t}|$.

Next, I adopt the measure of Antón and Polk (2014) to estimate the common ownership of stock pair i, j by the ETFs in my sample:

$$CAPij,t = \frac{\sum_{f=1}^{F}(S_{i,t}^{f}P_{i,t} + S_{j,t}^{f}P_{j,t})}{S_{i,t}P_{i,t} + S_{j,t}P_{j,t}} \tag{23}$$

where the numerator sums the total holdings of all ETFs in the sample in stock pair i, j and the denominator scales by the total market capitalization of stocks i and j. The number of shares held by the ETFs in my sample comes from ETF Global, while data on prices and total shares outstanding comes from CRSP. To facilitate economic interpretability, $CAP_{ij,t}$ is rank-transformed and standardized for every trading day. This measure is also used by Agarwal et al. (2018).

In addition, I compute a measure intended to capture changes in the number of shares held by ETFs, that is, ETF creations and redemptions. Therefore, I fist compute changes in ETF holdings in stock i, scaled by the number of shares outstanding in stock i

$$CREATIONS_{i,t} = \frac{\sum_{f=1}^{F}(S_{i,t}^{f} - S_{i,t-1}^{f})}{S_{i,t}}, \tag{24}$$

where $S_{i,t}^{f}$ is the number of shares ETF f holds in stock i on day t and $S_{i,t}$ is the number of shares outstanding of stock i on day t. This measure is positive for ETF creations and negative for ETF redemptions. Then, I compute a measure capturing the difference in the exposure of stocks i and j to ETF creations and redemptions:

$$|\Delta CREATION_{i,j,t}| = |CREATIONS_{i,t} - CREATIONS_{j,t}|. \tag{25}$$

Intuitively, scaling ETF creations and redemptions but the amount of shares outstanding captures the relative intensity of creations and redemptions.

To the extent that ETF ownership drives daily pricing error correlations, I expect common ETF ownership $CAPij,t$ to increase pricing error correlations (Antón and Polk, 2014; Agarwal et al., 2018). Similarly, I expect the coefficient on absolute ETF flows ($|FLOW_{i,t}|$ and $|FLOW_{j,t}|$) to be positive (Antón and Polk, 2014; Ben-David, Franzoni, and Moussawi, 2018). At the same time, I expect the coefficient on the difference in exposure to ETF flows ($|\Delta FLOW_{ij,t}|$) and the difference in exposure to ETF creations and redemptions ($|\Delta CREATIONSij,t|$) to be negative (Ben-David, Franzoni, and Moussawi, 2018).

According to Ben-David, Franzoni, and Moussawi (2018) ETF sponsors disseminate net-asset values at a 15-second frequency throughout the trading day. Since I analyze data at a lower frequency — at a 1-minute frequency – I am able to capture potential pricing errors arising as a consequence of ETF sponsor's arbitrage activity.

Results from estimating daily cross-sectional regressions for correlations in pricing errors including the above measures for exposure to ETF ownership

Table 4: Cross-Sectional Results Including Liquidity Demand Proxies

This table reports Fama and MacBeth (1973) estimates of daily cross-sectional regressions for correlations in pricing errors

$$\rho^s_{ij,t} = a + b_0\rho^x_{ij,t} + b_1|\Delta\beta^M_{ij,t-1}| + b_2|\Delta\beta^{SMB}_{ij,t-1}| + b_3|\Delta\beta^{HML}_{ij,t-1}| + \mathbf{X}\gamma + \varepsilon_{ij,t}$$

including proxies for liquidity demand. Pricing errors are obtained as smoothed states from estimating the state space model on the stock-day level. All independent variables are standardized to facilitate economic interpretability. $\rho^x_{i,j,t}$ is the correlation in the signed order flow series, $|\Delta\beta^M_{i,j,t-1}|$ is the absolute difference in the market betas of stocks i and j, obtained from a factor regression using the last 500 available return observations. Similarly, $|\Delta\beta^{SMB}_{i,j,t-1}|$ and $|\Delta\beta^{HML}_{i,j,t-1}|$ are the absolute difference in the SMB and HML betas of stocks in i and j, respectively. All specifications include controls for the absolute difference in the midquotes of stocks i and j on trading day $t-1$, 5-minute midquote return volatilities for stocks i and j on trading day $t-1$, and linear as well as nonlinear characteristics controls. I report Newey and West (1987) standard errors robust to autocorrelation of up to 20 lags in the cross-sectional estimates. Standard errors are reported in parentheses. * denotes significance at the 1% level.

	Panel A: Full Sample			
	(1)	(2)	(3)	(4)
$Constant$	0.014 92*	0.014 95*	0.014 87*	0.014 91*
	(0.000 76)	(0.000 76)	(0.000 76)	(0.000 76)
$\rho^x_{i,j,t}$	0.002 64*	0.002 64*	0.002 63*	0.002 64*
	(0.000 06)	(0.000 06)	(0.000 06)	(0.000 06)
$\lvert\Delta\beta^M_{i,j,t-1}\rvert$	−0.002 32*	−0.002 31*	−0.002 34*	−0.002 32*
	(0.000 10)	(0.000 10)	(0.000 10)	(0.000 10)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$	−0.002 05*	−0.002 07*	−0.001 99*	−0.002 06*
	(0.000 07)	(0.000 07)	(0.000 07)	(0.000 07)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$	−0.003 50*	−0.003 51*	−0.003 50*	−0.003 51*
	(0.000 15)	(0.000 15)	(0.000 15)	(0.000 15)
$\lvert\Delta FLOW_{ij,t}\rvert$	−0.000 04			
	(0.000 04)			
$\lvert FLOW_{i,t}\rvert$		0.000 45*		
		(0.000 04)		
$\lvert FLOW_{j,t}\rvert$		0.000 47*		
		(0.000 04)		
$CAP_{i,j,t}$			0.000 97*	
			(0.000 05)	
$\lvert\Delta CREATION_{i,j,t}\rvert$				−0.000 12*
				(0.000 03)

Table 4: – continued

	Panel B: High Market Capitalization			
	(1)	(2)	(3)	(4)
Constant	0.019 06*	0.019 16*	0.019 17*	0.019 14*
	(0.001 36)	(0.001 35)	(0.001 36)	(0.001 36)
$\rho^x_{i,j,t}$	0.002 63*	0.002 63*	0.002 63*	0.002 63*
	(0.000 08)	(0.000 08)	(0.000 08)	(0.000 08)
$\lvert\Delta\beta^M_{i,j,t-1}\rvert$	−0.005 21*	−0.005 26*	−0.005 26*	−0.005 25*
	(0.000 23)	(0.000 23)	(0.000 23)	(0.000 23)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$	−0.003 27*	−0.003 31*	−0.003 31*	−0.003 29*
	(0.000 17)	(0.000 17)	(0.000 17)	(0.000 17)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$	−0.006 54*	−0.006 59*	−0.006 59*	−0.006 57*
	(0.000 27)	(0.000 27)	(0.000 27)	(0.000 27)
$\lvert\Delta FLOW_{ij,t}\rvert$	−0.001 10*			
	(0.000 08)			
$\lvert FLOW_{i,t}\rvert$		0.000 13		
		(0.000 08)		
$\lvert FLOW_{j,t}\rvert$		0.000 22*		
		(0.000 07)		
$CAP_{i,j,t}$			−0.000 10	
			(0.000 08)	
$\lvert\Delta CREATION_{i,j,t}\rvert$				−0.000 77*
				(0.000 07)

are presented in Table 4. All specifications include controls for the absolute difference in the midquotes of stocks i and j on trading day $t-1$, 5-minute midquote return volatilities for stocks i and j on trading day $t-1$, and linear as well as nonlinear characteristics controls.

Comparing the results with the results for differences in the Fama and French (1993) betas only (Table 3), show that the main results are unchanged. Larger absolute differences in the factor betas are associated with lower daily pricing error correlations. This hold for both the full sample results as well as the results for high market capitalization stocks.

For the full-sample results, the signs of the coefficients on measures of (differences in) stock exposure to ETF ownership are in line with the hypotheses that emerge from the literature: common ETF ownership as well as absolute flows from ETFs into stocks i and j are associated with higher pricing error correlations. Larger differences in the exposure to ETF flows and ETF creations and redemptions are associated with lower pricing error correlations. Comparing the economic magnitudes of the coefficients of the differences in factor betas and and the exposure to ETF ownership reveals that the effect of ETF ownership is in general an order of magnitude smaller than the effect of differences in betas.

For high market capitalization stocks only, the signs of the coefficients on measures of (differences in) stock exposure to ETF ownership are again in line with the hypotheses that emerge from the literature, except for common ownership ($CAP_{i,j,t}$). However, the effect of common ownership is insignificant at all conventional significance levels and the point estimate is economically small, constituting a precise zero. At the same time, I find a relatively larger negative effect of the difference in exposures of stocks i and j to ETF flows ($|\Delta FLOW_{ij,t}|$). In comparison to the full sample results, the coefficient on $|\Delta FLOW_{ij,t}|$ is at least an order of magnitude larger. A possible explanation for this is that the large stocks in my sample are more common constituents of widely traded ETFs.

In addition, this result suggests the following: When analyzing the effect of ETF ownership at high (intraday) frequencies, it is not ETF ownership per se that drives prices. Rather, flows resulting from ETF ownership drive prices and pricing errors. This is intuitive as market makers/liquidity observe and react to the flows they observe, rather than to holdings by market participants (this is the case in inventory models as my model in Section 2 as well as in adverse-selection models such as Kyle (1985) and Glosten and Milgrom (1985)). Moreover, this is consistent with the result in Antón and Polk (2014) that common ownership is more relevant in periods of high (absolute) flows.

If the effect I document in Section 5.2 was solely driven by a liquidity demand channel, rather than liquidity supply as in the theoretical model in Section 2, I expect the coefficients on the measures of (differences in) stock exposure to ETF ownership to be of at least a similar magnitude as the coefficients on the differences in the factor betas. In addition, I expect the size of the coefficients on the differences in factor betas to decrease. This is neither the case in the full sample results, nor in the high market capitalization sample results.

Rather, my results suggest that intraday pricing errors as well as their daily correlations are driven by both liquidity supply as well as liquidity demand channels, with the economic effect of a liquidity supply channel being economically larger. With this I complement the literature that shows that correlations in daily liquidity metrics are driven by ETF ownership, and therefore liquidity demand (Agarwal et al., 2018). In this context, it should also be noted that I estimate my state space model which yields estimates of pricing errors on a stock-day level. By construction, this does not allow me to identify longer lasting pricing errors. At the same time, estimating the state space model on the stock-day level alleviates concerns of me capturing a mechanical relationship between pricing error correlations and the explanatory variables in my regressions over multiple days.

Bogousslavsky and Muravyev (2023) document an increase in trading during closing auctions related to the increasingly important role of ETFs in today's financial market architecture. Hendershott and Menkveld (2014) show that closing prices are pressured away from efficient prices as a result of specialist inventory positions. My findings complement these findings for higher frequencies. This is again in line with the literature showing the impact of common fund and ETF ownership on correlations lower-frequency liquidity metrics and factor residuals (Antón and Polk, 2014; Agarwal et al., 2018).

6 Conclusions

Motivated by an inventory model I study pricing errors in the cross-section of stocks. A simple inventory model in which prices load on risk factors predicts that pricing errors of different securities are correlated if they load on the same underlying risk factor. Using a state space model, I identify pricing errors at a 1-minute frequency and compute correlations in pricing errors for stock pairs for each trading day. My results indicate that absolute differences in factor betas are negatively related to correlations in pricing errors. The effect is economically sizeable.

In addition, I study the relationship between pricing error correlations and liquidity demanding flows from ETFs and well as ETF ownership. While my results reveal a null result for the relationship between common ownership by ETFs and pricing error correlations, I find that ETF flows contribute to pricing error correlations in the cross section of stocks.

My results reveal systematic drivers of price efficiency in the cross-section of stocks. This adds to the literature showing that market wide risks drive inventory effects in stocks (Conrad and Wahal, 2020).

References

Agarwal, Vikas, Paul Hanouna, Rabih Moussawi, and Christof W Stahel (2018). *Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?* Working Paper.

Antón, Miguel and Christopher Polk (2014). "Connected Stocks". In: *Journal of Finance* 69.3, pp. 1099–1127.

Aquilina, Matteo, Eric Budish, and Peter O'Neill (Sept. 2021). "Quantifying the High-Frequency Trading "Arms Race": A Simple New Methodology and Estimates". In: *The Quarterly Journal of Economics* 137.1, pp. 493–564.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi (2018). "Do ETFs Increase Volatility?" In: *Journal of Finance* 73.6, pp. 2471–2535.

Bogousslavsky, Vincent and Pierre Collin-Dufresne (2023). "Liquidity, Volume, and Order Imbalance Volatility". In: *Journal of Finance* 78.4, pp. 2189–2232.

Bogousslavsky, Vincent and Dmitriy Muravyev (2023). "Who trades at the close? Implications for price discovery and liquidity". In: *Journal of Financial Markets* 66.C.

Brandt, Michael W. and Kenneth A. Kavajecz (2004). "Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve". In: *Journal of Finance* 59.6, pp. 2623–2654.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014). "High-Frequency Trading and Price Discovery". In: *Review of Financial Studies* 27.8, pp. 2267–2306.

Budish, Eric, Peter Cramton, and John Shim (2015). "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response". In: *The Quarterly Journal of Economics* 130.4, pp. 1547–1621.

Carhart, Mark M (1997). "On Persistence in Mutual Fund Performance". In: *Journal of Finance* 52.1, pp. 57–82.

Chakrabarty, Bidisha, Roberto Pascual, and Andriy Shkilko (2015). "Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm". In: *Journal of Financial Markets* 25.C, pp. 52–79.

Chordia, Tarun, T Clifton Green, and Badrinath Kottimukkalur (2018). "Rent Seeking by Low-Latency Traders: Evidence from Trading on Macroeconomic Announcements". In: *Review of Financial Studies* 31.12, pp. 4650–4687.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2000). "Commonality in Liquidity". In: *Journal of Financial Economics* 56.1, pp. 3–28.

Conrad, Jennifer and Sunil Wahal (2020). "The Term Structure of Liquidity Provision". In: *Journal of Financial Economics* 136.1, pp. 239–259.

Duarte, Jefferson, Edwin Hu, and Lance Young (2020). "A Comparison of Some Structural Models of Private Information Arrival". In: *Journal of Financial Economics* 135.3, pp. 795–815.

Evans, Martin D.D. and Richard K. Lyons (2008). "How is Macro News Transmitted to Exchange Rates?" In: *Journal of Financial Economics* 88.1, pp. 26–50.

Fama, Eugene F. and Kenneth R. French (1993). "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56.

Fama, Eugene F. and Kenneth R. French (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.

Fama, Eugene F and James D MacBeth (1973). "Risk, Return, and Equilibrium: Empirical Tests". In: *Journal of Political Economy* 81.3, pp. 607–636.

Foucault, Thierry, Marco Pagano, and Ailsa Roell (2013). *Market Liquidity: Theory, Evidence, and Policy*. Oxford University Press.

George, Thomas J. and Mozzam A. Khoja (2023). *Estimating Price Impact and its Components: Evidence from HFT Around Earnings Announcements*. Working Paper.

Glosten, Lawrence R. and Paul R. Milgrom (1985). "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders". In: *Journal of Financial Economics* 14.1, pp. 71–100.

Hasbrouck, Joel (1993). "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement". In: *Review of Financial Studies* 6.1, pp. 191–212.

Hasbrouck, Joel and Duane J. Seppi (2001). "Common Factors in Prices, Order Flows, and Liquidity". In: *Journal of Financial Economics* 59.3, pp. 383–411.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld (2011). "Does Algorithmic Trading Improve Liquidity?" In: *Journal of Finance* 66.1, pp. 1–33.

Hendershott, Terrence and Albert J. Menkveld (2014). "Price Pressures". In: *Journal of Financial Economics* 114.3, pp. 405–423.

Holden, Craig W. and Stacey Jacobsen (2014). "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions". In: *Journal of Finance* 69.4, pp. 1747–1785.

Kyle, Albert S (1985). "Continuous Auctions and Insider Trading". In: *Econometrica* 53.6, pp. 1315–1335.

Lee, Charles M C and Mark J Ready (1991). "Inferring Trade Direction from Intraday Data". In: *Journal of Finance* 46.2, pp. 733–746.

Li, Sida and Mao Ye (2023). *Discrete Prices, Discrete Quantities, and the Optimal Price of a Stock*. Working Paper.

Lou, Dong and Christopher Polk (2021). "Comomentum: Inferring Arbitrage Activity from Return Correlations". In: *The Review of Financial Studies*.

Menkveld, Albert J., Siem Jan Koopman, and Andre Lucas (2007). "Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods". In: *Journal of Business & Economic Statistics* 25, pp. 213–225.

Menkveld, Albert J. and Ion Lucas Saru (2024). *Who Knows? Information Differences Between Trader Types*. Working Paper.

Newey, Whitney K and Kenneth D West (1987). "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Rösch, Dominik M., Avanidhar Subrahmanyam, and Mathijs A. Van Dijk (2017). "The Dynamics of Market Efficiency". In: *Review of Financial Studies* 30.4, pp. 1151–1187.

Stoll, Hans R (1978). "The Supply of Dealer Services in Securities Markets". In: *Journal of Finance* 33.4, pp. 1133–1151.

Van Binsbergen, Jules H., Martijn Boons, Christian C. Opp, and Andrea Tamoni (2023). "Dynamic Asset (Mis)Pricing: Build-up versus Resolution Anomalies". In: *Journal of Financial Economics* 147.2, pp. 406–431.

Weller, Brian M. (2018). "Does Algorithmic Trading Reduce Information Acquisition?" In: *Review of Financial Studies* 31.6, pp. 2184–2226.

A Cross-Sectional Results with Additional Factors

In this Section I present additional results for different factor models. Table 5 presents results for the Fama and French (1993) – Carhart (1997) factors and Table 6 presents results for the Fama and French (2015) factors. As before, I report results for the full sample as well as high-market capitalization stocks only.

Comparing with the results for the Fama and French (1993) factors shows consistent results. As before, the results are stronger for high-market capitalization stocks only, compared to the full sample results. This alleviates the concern that the results are driven by small, irregularly traded stocks. In addition, differences in market betas consistently appear to be the most important driver of differences in daily pricing error correlations. The economic magnitude of differences in the additional factors (UMD, RMW, and CMA) is overall smaller than the magnitude of differences in the Fama and French (1993) factors. Especially the CMA factor appears to have little traction in the overall sample. In the high-market capitalization sample, however, the coefficients on the CMA factor are in line with the intuition of the theoretical model presented in Section 2.

Table 5: Cross-Sectional Results for Pricing Error Correlations for Fama and French (1993) – Carhart (1997) Factors

This table reports Fama and MacBeth (1973) estimates of daily cross-sectional regressions for correlations in pricing errors and Fama and French (1993) – Carhart (1997) factors. Pricing errors are obtained as smoothed states from estimating the state space model on the stock-day level. All independent variables are standardized to facilitate economic interpretability. $\rho^{x}_{i,j,t}$ is the correlation in the signed order flow series, $|\Delta\beta^{M}_{i,j,t-1}|$ is the absolute difference in the market betas of stocks i and j, obtained from a factor regression using the last 500 available return observations. Similarly, $|\Delta\beta^{SMB}_{i,j,t-1}|$ is the absolute difference in the SMB betas, $|\Delta\beta^{HML}_{i,j,t-1}|$ is the absolute difference in the HML betas, and $|\Delta\beta^{UMD}_{i,j,t-1}|$ is the absolute difference in the UMD (momentum) betas of stocks in i and j. $|\Delta m_{t-1}|$ is the absolute difference in the midquotes of stocks i and j on trading day $t-1$, and $\sigma(r^{5min}_{i,t-1})$ as well $\sigma(r^{5min}_{j,t-1})$ are the 5-minute midquote return volatilities for stocks i and j on trading day $-t1$. The specifications that control for linear and nonlinear characteristics include the (standardized) factors betas as well as squared terms of the factor betas for stocks i and j. I report Newey and West (1987) standard errors robust to autocorrelation of up to 20 lags in the cross-sectional estimates. Standard errors are reported in parentheses. * denotes significance at the 1% level.

	Panel A: Full Sample			
	(1)	(2)	(3)	(4)
$Constant$	0.019 08*	0.019 32*	0.019 29*	0.014 37*
	(0.000 74)	(0.000 76)	(0.000 76)	(0.000 77)
$\rho^{x}_{i,j,t}$	0.008 68*	0.002 96*	0.002 89*	0.002 62*
	(0.001 01)	(0.000 07)	(0.000 07)	(0.000 06)
$\|\Delta\beta^{M}_{i,j,t-1}\|$		−0.002 27*	−0.002 11*	−0.002 38*
		(0.000 11)	(0.000 10)	(0.000 09)
$\|\Delta\beta^{SMB}_{i,j,t-1}\|$		−0.001 69*	−0.001 54*	−0.001 98*
		(0.000 09)	(0.000 09)	(0.000 08)
$\|\Delta\beta^{HML}_{i,j,t-1}\|$		−0.000 42	−0.000 17	−0.003 42*
		(0.000 17)	(0.000 17)	(0.000 15)
$\|\Delta\beta^{UMD}_{i,j,t-1}\|$		0.000 04	0.000 39*	−0.000 93*
		(0.000 06)	(0.000 07)	(.000 07)
$\|\Delta m_{t-1}\|$			−0.000 83*	−0.000 66*
			(0.000 06)	(0.000 04)
$\sigma(r^{5min}_{i,t-1})$			−0.001 74*	−0.002 48*
			(0.000 13)	(0.000 10)
$\sigma(r^{5min}_{j,t-1})$			−0.001 70*	−0.002 39*
			(0.000 11)	(0.000 10)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes

Table 5: – continued

	Panel B: High Market Capitalization			
	(1)	(2)	(3)	(4)
Constant	0.027 26*	0.027 26*	0.027 26*	0.017 96*
	(0.001 24)	(0.001 24)	(0.001 24)	(0.001 39)
$\rho^x_{i,j,t}$	0.003 10*	0.002 88*	0.002 86*	0.002 61*
	(0.000 08)	(0.000 08)	(0.000 08)	(0.000 07)
$\lvert\Delta\beta^M_{i,j,t-1}\rvert$		−0.004 23*	−0.004 31*	−0.005 42*
		(0.000 19)	(0.000 19)	(0.000 21)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$		−0.002 35*	−0.002 32*	−0.003 18*
		(0.000 12)	(0.000 13)	(0.000 17)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$		−0.001 57*	−0.001 52*	−0.006 59*
		(0.000 28)	(0.000 29)	(0.000 26)
$\lvert\Delta\beta^{UMD}_{i,j,t-1}\rvert$		−0.001 30*	−0.001 17*	−0.002 39*
		(0.000 15)	(0.000 14)	(0.000 16)
$\lvert\Delta m_{t-1}\rvert$			−0.000 54*	−0.000 61*
			(0.000 09)	(0.000 09)
$\sigma(r^{5min}_{i,t-1})$			−0.000 08	−0.002 56*
			(0.000 16)	(0.000 16)
$\sigma(r^{5min}_{j,t-1})$			−0.000 52*	−0.002 58*
			(0.000 15)	(0.000 14)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes

Table 6: Cross-Sectional Results for Pricing Error Correlations for Fama and French (2015) Factors

This table reports Fama and MacBeth (1973) estimates of daily cross-sectional regressions for correlations in pricing errors and Fama and French (2015) factors. Pricing errors are obtained as smoothed states from estimating the state space model on the stock-day level. All independent variables are standardized to facilitate economic interpretability. $\rho^x_{i,j,t}$ is the correlation in the signed order flow series, $|\Delta\beta^M_{i,j,t-1}|$ is the absolute difference in the market betas of stocks i and j, obtained from a factor regression using the last 500 available return observations. Similarly, $|\Delta\beta^{SMB}_{i,j,t-1}|$, $|\Delta\beta^{HML}_{i,j,t-1}|$, $|\Delta\beta^{RMW}_{i,j,t-1}|$, and $|\Delta\beta^{CMA}_{i,j,t-1}|$ are the absolute difference in the SMB, HML, RMW, and CMA betas of stocks in i and j, respectively. $|\Delta m_{t-1}|$ is the absolute difference in the midquotes of stocks i and j on trading day $t-1$, and $\sigma(r^{5min}_{i,t-1})$ as well $\sigma(r^{5min}_{j,t-1})$ are the 5-minute midquote return volatilities for stocks i and j on trading day $-t1$. The specifications that control for linear and nonlinear characteristics include the (standardized) factors betas as well as squared terms of the factor betas for stocks i and j. I report Newey and West (1987) standard errors robust to autocorrelation of up to 20 lags in the cross-sectional estimates. Standard errors are reported in parentheses. * denotes significance at the 1% level.

	Panel A: Full Sample			
	(1)	(2)	(3)	(4)
$Constant$	0.019 08*	0.019 32*	0.019 30*	0.013 35*
	(0.000 74)	(0.000 76)	(0.000 76)	(0.000 75)
$\rho^x_{i,j,t}$	0.008 68*	0.002 97*	0.002 91*	0.002 61*
	(0.001 01)	(0.000 07)	(0.000 07)	(0.000 06)
$\lvert\Delta\beta^M_{i,j,t-1}\rvert$		−0.002 13*	−0.001 91*	−0.002 17*
		(0.000 11)	(0.000 10)	(0.000 10)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$		−0.001 64*	−0.001 49*	−0.002 03*
		(0.000 08)	(0.000 08)	(0.000 07)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$		−0.000 20	−0.000 06	−0.003 59*
		(0.000 13)	(0.000 13)	(0.000 14)
$\lvert\Delta\beta^{RMW}_{i,j,t-1}\rvert$		−0.000 50*	−0.000 02	−0.001 15*
		(0.000 08)	(0.000 06)	(0.000 05)
$\lvert\Delta\beta^{CMA}_{i,j,t-1}\rvert$		0.000 20*	0.000 17*	−0.001 03*
		(0.000 06)	(0.000 06)	(0.000 06)
$\lvert\Delta m_{t-1}\rvert$			−0.000 83*	−0.000 63*
			(0.000 07)	(0.000 04)
$\sigma(r^{5min}_{i,t-1})$			−0.001 71*	−0.002 49*
			(0.000 12)	(0.000 10)
$\sigma(r^{5min}_{j,t-1})$			−0.001 66*	−0.002 40*
			(0.000 11)	(0.000 10)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes

Table 6: – continued

	Panel B: High Market Capitalization			
	(1)	(2)	(3)	(4)
Constant	0.027 26*	0.027 26*	0.027 26*	0.017 02*
	(0.001 24)	(0.001 24)	(0.001 24)	(0.001 38)
$\rho^{x}_{i,j,t}$	0.003 10*	0.002 89*	0.002 87*	0.002 59*
	(0.000 08)	(0.000 08)	(0.000 08)	(0.000 07)
$\lvert\Delta\beta^{M}_{i,j,t-1}\rvert$		−0.003 90*	−0.003 95*	−0.004 48*
		(0.000 21)	(0.000 21)	(0.000 25)
$\lvert\Delta\beta^{SMB}_{i,j,t-1}\rvert$		−0.001 89*	−0.001 85*	−0.002 89*
		(0.000 12)	(0.000 12)	(0.000 15)
$\lvert\Delta\beta^{HML}_{i,j,t-1}\rvert$		−0.001 07*	−0.001 08*	−0.006 69*
		(0.000 24)	(0.000 23)	(0.000 26)
$\lvert\Delta\beta^{RMW}_{i,j,t-1}\rvert$		−0.001 82*	−0.001 74*	−0.002 33*
		(0.000 19)	(0.000 17)	(0.000 12)
$\lvert\Delta\beta^{CMA}_{i,j,t-1}\rvert$		−0.000 84*	−0.000 80*	−0.003 11*
		(0.000 13)	(0.000 13)	(0.000 16)
$\lvert\Delta m_{t-1}\rvert$			−0.000 59*	−0.000 66*
			(0.000 08)	(0.000 08)
$\sigma(r^{5min}_{i,t-1})$			−0.000 07	−0.002 40*
			(0.000 15)	(0.000 14)
$\sigma(r^{5min}_{j,t-1})$			−0.000 47*	−0.002 55*
			(0.000 14)	(0.000 14)
Linear characteristics	No	No	No	Yes
Nonlinear characteristics	No	No	No	Yes