

The Night and Day of Amihud's (2002) Liquidity Measure

Yashar H. Barardehi, Dan Bernhardt, Thomas G. Ruchti  
& Marc Weidenmier

**July 2019**

**No: 1211**

(Updated Sept 2020 & accepted for publication in, [The Review of Asset Pricing Studies](#))

**Warwick Economics Research Papers**

**ISSN 2059-4283 (online)**

**ISSN 0083-7350 (print)**

# The Night and Day of Amihud's (2002) Liquidity Measure\*

Yashar H. Barardehi

Dan Bernhardt

Thomas G. Ruchti

Marc Weidenmier

First draft: February 15, 2019

This draft: September 11, 2020

## Abstract

Amihud's (2002) stock (il)liquidity measure averages daily ratios of absolute close-to-close return to dollar volume, including overnight returns. Our modified measure uses open-to-close returns, matching return and trading volume measurement windows. It is more strongly correlated with trading-cost measures (by 8–37%); moreover, it better explains cross-sections of returns, doubling estimated liquidity premia. Using non-synchronous trading near close, we show overnight returns are primarily information-driven: including them in Amihud's proxy for price impacts of trading magnifies measurement error, understating liquidity premia. Our modification helps wherever Amihud's measure is required. Our measures are publicly available for 1964–2019, and can be updated.

---

\*We are grateful for comments from Yakov Amihud, Azi Ben-Rephael (discussant), Asaf Bernstein, Charles Calomiris, Terry Hendershott, Eric Hughson, Tim Johnson, Ivalina Kalcheva, Joonki Noh, Christine Parlour, Andriy Shkilko, Elvira Sojli, Konstantin Sokolov, Erik Theissen (discussant), Mitch Warachka, and conference participants at the Chapman Money and Finance, the 3<sup>rd</sup> SAFE Market Microstructure, and the 2020 Midwest Financial Association meetings. Weidenmier acknowledges financial support from Chapman University. Barardehi (barardehi@chapman.edu) is at the Argyros School of Business & Economics, Chapman University. Bernhardt (danber@illinois.edu) is at Department of Economics, University of Illinois and University of Warwick. Ruchti (ruchti@andrew.cmu.edu) is at the Tepper School of Business, Carnegie Mellon University. Weidenmier (weidenmi@chapman.edu) is at the Argyros School of Business & Economics, Chapman University and NBER. The data for open-to-close Amihud measures of liquidity are available to researchers at [https://www.dropbox.com/s/9chf0a0dvj9tz91/OCAM\\_OCAM\\_1964-2019.zip?dl=0](https://www.dropbox.com/s/9chf0a0dvj9tz91/OCAM_OCAM_1964-2019.zip?dl=0)

# 1 Introduction

The stock (il)liquidity measure proposed by Amihud (2002) is the most widely-used such measure in empirical financial economics.<sup>1</sup> Its key advantage stems from the fact that its simple construction only requires daily return and dollar volume data that are available for many markets and countries over long periods of time. Use of liquidity measures like Amihud’s is inevitable for applications where intraday trade and quote data are simply unavailable. Amihud’s measure has consistently produced evidence of priced idiosyncratic liquidity and liquidity risk,<sup>2</sup> and it has been found to be a reasonable proxy of institutional trading costs.<sup>3</sup> Both market microstructure experts and other scholars use this measure regularly.<sup>4</sup>

Our paper identifies a simple and intuitive improvement to Amihud’s liquidity measure, and documents its striking consequences. The original measure uses the ratio of absolute daily close-to-close return to dollar volume as a proxy for price impacts of trading, i.e., the amount a given trading volume moves market prices. Data and institutional details indicate that while there is nearly no trading volume outside regular trading hours, the corresponding overnight price movements make up a large share of close-to-close absolute returns. The literature has established that overnight (after-hours) price movements are typically driven by information arrival that is unrelated to the daily trading volume used in the denominator.<sup>5</sup>

---

<sup>1</sup>Amihud’s (2002) article has 2,487 citations by peer-reviewed published articles, including 297 in the top-three finance journals and 57 in the top-three accounting journals (Web of Science, accessed August 18, 2020).

<sup>2</sup>Kingsley et al. (2017) find Amihud’s (2002) measure to be unsurpassed as a “cost-per-dollar-volume” proxy for global research. Goyenko et al. (2009) show that it is a good proxy for the price impacts of trading.

<sup>3</sup>See Chordia et al. (2000), Pástor and Stambaugh (2003), Acharya and Pedersen (2005), Sadka (2006), Asparouhova et al. (2013), Drienko et al. (2017), Harris and Amato (2019), among others. Goyenko et al. (2009) show that Amihud’s measure remains priced post-decimalization. Anand et al. (2013) and Barardehi et al. (2019) provide evidence of strong time-series and cross-sectional correlations between Amihud’s liquidity measure and actual and estimated institutional trading costs. Studies such as Lipson and Mortal (2004) relate equity liquidity, captured by Amihud’s measure, and corporate finance decisions.

<sup>4</sup>For example, in corporate finance and accounting research where use of intraday data is rare, Amihud’s measure is also widely-used as an easy-to-construct proxy for liquidity.

<sup>5</sup>Boudoukh et al. (2019) find that 50% of overnight idiosyncratic volatility is due to public news, as opposed to revelation of private information through trading; this compares to only 12% during trading hours. Santosh (2016) finds that 71% of stock value shocks driven by after-hour earnings surprises are reflected in opening prices the following trading day. See also Stole and Whaley 1990, Cao et al. 2000, and

This means that including overnight returns in a proxy of price impacts creates a fundamental time mismatch between inputs: the denominator reflects trading volume during trading hours, while the numerator reflects the absolute return between the close of the previous day and the current day, including overnight price movements. This leads us to modify Amihud’s (2002) measure by using absolute daily open-to-close return in the numerator of the price impact proxy, thereby restoring the time match between numerator and denominator.

Our modified Amihud (2002) measure produces 50-120% larger estimates of liquidity premia in the cross-section of returns. Importantly, unlike the original Amihud measure, the modified measure is priced post 2001, even after adjusting for upward biases due to microstructure noise (Asparouhova et al. 2010); this finding is more pronounced for small-/mid-cap stocks. Our modification significantly increases the correlation coefficients between the Amihud measure and standard benchmarks of trading costs constructed at both high frequencies (e.g., effective/quoted spreads, estimated Kyle’s  $\lambda$ ) and low frequencies (e.g., effective costs estimates of Hasbrouck 2009). These increased correlations of 5–15 percentage points present themselves in significantly changed cross-sectional rankings of stocks according to the Amihud measure that translate into sharp improvements in our ability to explain the cross-section of returns. Exploiting non-synchronous trading before close as an instrument, we establish that the overnight returns excluded by our modification are largely information-driven and divorced from price impacts of trading. In sum, our modification eliminates liquidity measurement error that would otherwise bias estimated liquidity premia downward.

We motivate our open-to-close Amihud measure, *OCAM*, as a substitute for the original close-to-close Amihud measure, *CCAM*, based on multiple observations. Reported trading volumes on CRSP almost entirely reflect trade during regular trading hours.<sup>6</sup> As Figure 1

---

Barclay and Hendershott (2003) for the impact of information revelation at open.

<sup>6</sup>Overnight trading did not exist until 1991, and post 1991, CRSP measures of daily volume only include a tiny fraction of after-hours volume. According to CRSP’s Data Description Guide available on WRDS, it includes a fraction of after-hours automated executions on NASDAQ and ECNs. Pre-decimalization, over *all* hours, ECNs account for less than 3% of total trading volume (see, e.g.,

illustrates, this creates a time mismatch between the measures entering the numerator and denominator of *CCAM*. This mismatch is economically important. Overnight price movements are comparable with those in regular hours, but overnight trading volume comprises less than 2% of daily volume, even post decimalization (see Appendix A.1). Including large overnight price movements that are unrelated to trading volume and trading costs adds substantial noise to *CCAM*.<sup>7</sup> Indeed, there would still be time mismatch even if CRSP perfectly

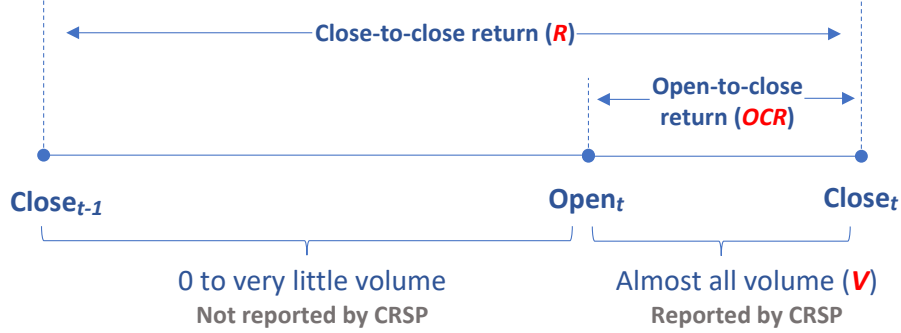


Figure 1: **Time mismatch between daily measures of trading volume and return.** This figure illustrates CRSP’s sampling of daily return and trading volume for a typical trading day. These measures are used to construct Amihud’s (2002) daily proxy of price impacts.

measured after-hours trading volume. Daily returns reflect closing prices, but daily volumes are measured over calendar days, misaligning after-hours volume and the associated return by one trading day. Our modification addresses this mismatch by removing overnight returns.

We establish that overnight price movements represent a major source of the observed cross-sectional and time-series variations in *CCAM*. Figure 2 plots over time the ratio of *OCAM* to *CCAM* for the median stock in each tercile of the ratio. Were there negligible differences between the two Amihud measures, this ratio would be close to one. In fact, the ratio is far less than one for the bottom tercile of stocks in the first forty years of the sample (ranging from below 0.2 to about 0.6 before rising to 0.8 and above by the mid-2000s), a pattern that cannot be explained by cross-stock variation in after-hours trading.<sup>8</sup> We find <https://www.sec.gov/news/studies/ecnafter.htm#exec>.

<sup>7</sup>This logic is also why measures like spreads or estimates of Kyle’s  $\lambda$  only use data from regular hours.

<sup>8</sup>After-hours trading was introduced in 1992, but it remained so limited that it motivated research to

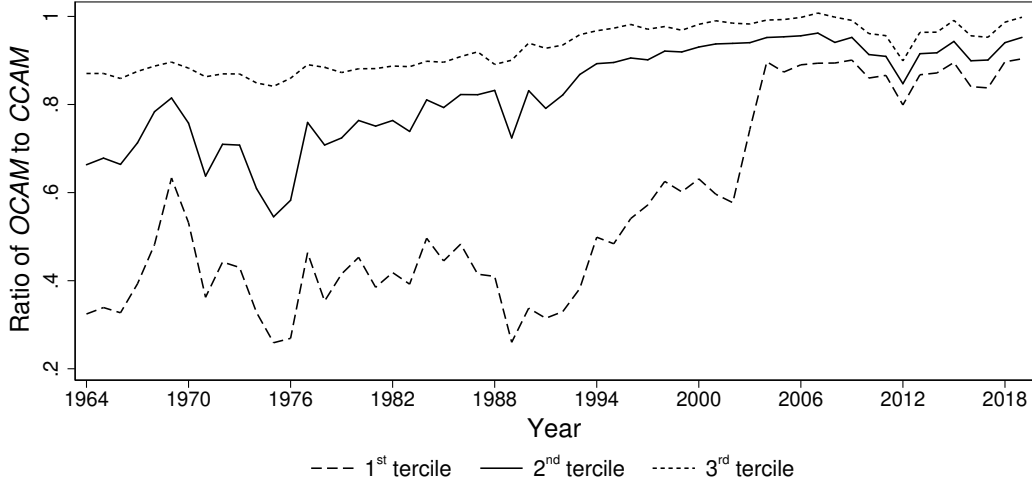


Figure 2: **Evolution of the ratio of  $OCAM$  to  $CCAM$ .** The figure plots the temporal changes in the cross-stock distribution of the  $OCAM$ -to- $CCAM$  for NYSE- and AMEX-listed common shares in the 1964-2019 period. For each stock  $i$  in year  $y$ , the ratio  $ROC'_{iy} = OCAM_{i,y-1}/CCAM_{i,y-1}$  is calculated. Stocks are sorted into terciles of  $ROC'_{iy}$  each year, and the year-specific medians in respective terciles of this ratio are plotted against time. See Section 2.2 for variable construction.

other key patterns regarding this new measure: (i) cross-stock variation tends to decline over time; (ii) while the ratio of  $OCAM$  to  $CCAM$  displays a generally positive trend, there are several episodes of sharp systematic decline such as that during the 2008 financial crisis, highlighting the economic relevance of  $OCAM$  even in modern financial markets; (iii) the ratio is less than one for almost 95% of stock-years; and (iv) reflecting the temporal stability of the ratio, the two measures display a strong time-series co-movement, suggesting that  $OCAM$  picks up temporal variation in price impacts similarly to  $CCAM$ .

To construct  $OCAM$ , we use the Global Financial Data (GFD) commercial database that reports historical open prices. Adopting Amihud's (2002) specification for the cross-section of expected returns, we show that liquidity premia based on  $OCAM$  are roughly double those obtained using Amihud's traditional  $CCAM$  measure, both for the entire 1964-2019 sample period, and in different sub-periods. Liquidity premia decline over time, consistent with find-  


---

 explain this anomaly (see, e.g., Belcourt (1996)). Moreover, years before after-hours trading on ECNs was introduced, differences between the two measures had already fallen sharply (close to current levels).

ings of Asparouhova et al. (2010), Ben-Rephael et al. (2013), and Harris and Amato (2019).

Qualitatively identical outcomes obtain when we only use CRSP data to construct *OCAM*. CRSP only begins to report open prices in 1992, which limits its historical reach. Nonetheless, using a CRSP-based 1993–2019 sample, we find that *OCAM* produces statistically significant liquidity premia of 3.4–4.9 bps per month for NYSE- and AMEX-listed stocks. In contrast, *CCAM* is not associated with statistically significant liquidity premia once we adjust for microstructure noise. We document qualitatively identical findings when we use GFD data to construct *OCAM* over the same period (see Appendix A.2).

Our findings are robust to other measurement, sample and specification choices. First, standardizing the measures to have zero mean and unit variance, we show that differences in the cross-sectional dispersions of the two versions of the Amihud measure do not underlie the differences in the magnitudes of liquidity premia. Second, excluding windows around earnings announcements reveals that trading around earnings release does not drive the differences. Third, qualitatively identical findings obtain for NASDAQ-listed common shares. Fourth, to maximize the cross-sectional coverage of our sample, we construct a hybrid sample that maximizes the number of stock-year observations, using open prices from GFD whenever CRSP open prices are unavailable. We document qualitatively identical results for this hybrid sample. Fifth, panel regression estimates that control for stock and time fixed effects also yield liquidity premia based on *OCAM* that are roughly double those for *CCAM* (see the online Appendix A.6). Hence, neither unknown fixed stock characteristics nor temporal changes in the composition of listed stocks explain the difference between *CCAM* and *OCAM*.<sup>9</sup>

Importantly, we establish that *OCAM* outperforms *CCAM* in capturing stock liquidity. We first show that *OCAM* is a more accurate proxy of trading costs. To do this, we compare

---

<sup>9</sup>In Appendix A.5, we also find that excluding the first five or ten minutes after open to eliminate impacts of open auctions eliminates noise in the construction of the liquidity measure, raises the correlation between the numerator and denominator of the Amihud measure, and results in greater estimated liquidity premia. However, this measure can only be constructed once TAQ data become available.

the cross-sectional correlations of *OCAM* and *CCAM* vis à vis standard trading cost proxies, such as effective/quoted spreads, estimates of Kyle’s  $\lambda$ , or effective trading costs estimates of Hasbrouck (2009). For each benchmark, we reject the null of *OCAM* and *CCAM* being similarly correlated in favor of *OCAM* being more strongly correlated with that benchmark. Reflecting these findings, *OCAM* outperforms *CCAM* in explaining the cross-section of returns. To evaluate the incremental information content of each measure, we decompose each version of Amihud’s (2002) measure into linearly orthogonal components with respect to the alternative. The residuals from regressing *OCAM* on *CCAM* significantly explain the cross-section of expected returns, even when we control for *CCAM*. In contrast, the converse is not true.

Our paper also informs a recent debate on the pricing of Amihud’s (2002) measure. Lou and Shu (2017, LS) argue that its pricing is driven by variation in the dollar volumes entering its denominator, and that the absolute returns entering the numerator are irrelevant. Amihud and Noh (2020, AN) show that LS’s decomposition of the Amihud measure omits the covariance between daily absolute return and the inverse of daily dollar volume, which is priced.<sup>10</sup> The core premise of microstructure models is that trading drives price impacts, so the association between the numerator and denominator of the price impact proxy is key. Our modification greatly magnifies the correlation between absolute returns and the inverse of dollar volumes, i.e., between the numerator and the denominator of the Amihud measure. Removing overnight returns from the numerator filters out noise that would otherwise lead to underestimated correlations. In fact, the primary source of the difference between *OCAM* and *CCAM* reflects this: *OCAM* and *CCAM* differ by more when our modification affects the link between the numerator and denominator by more. We reinforce AN’s basic findings using *OCAM*, showing that the covariance between *OCAM*’s numerator and denominator inputs is priced distinctly more strongly in the cross-section than that of *CCAM*.

Our final contribution is to uncover what drives the cross-sectional and temporal variation

---

<sup>10</sup>AN observe that the log-linearization used by LS to decompose Amihud’s (2002) measure into the sum of the log of mean absolute return and the log of the average of one over dollar volume is incorrect.



in the deviations of the *OCAM/CCAM* ratio from 1. In particular, we explain the ratio’s rise and convergence toward unity post decimalization, establishing that the improvements *OCAM* makes over *CCAM* reflect that its construction excludes (overnight) information-driven price movements that do not enter trading costs. To show this, we construct a proxy for overnight information arrival using the extent of non-synchronous trading near close. This proxy exploits the variation in the time distance between the last transaction of the day and 4:00pm EST. If a stock experiences a longer period of non-trading before close then it should see more accumulated information after the final trade, but before close. That accumulated information is impounded into price at open on the next trading day, leading to greater overnight price movement. As such, our design uses differential information arrival before close to proxy for total information arrival in the overnight hours.<sup>11</sup>

The cross-sectional variation in the extent of non-trading across stocks before close explains about 77% of the cross-sectional variation in the ratio of *OCAM* to *CCAM* post-1993. Indeed, after removing the variation associated with this proxy of overnight information arrival, the cross-sectional distribution of the *OCAM/CCAM* residual becomes stable over time—the temporal variation in the distribution of the ratio is almost entirely explained by that in the distribution of non-trading. After controlling for stock characteristics, including liquidity, along with stock and time fixed effects, an additional ten minutes of non-trading prior to close is associated with a 5-percentage-point decline in the *OCAM/CCAM* ratio. Thus, large differences in non-trading near close imply large differences in the two measures. This analysis ties the extent of greater measurement error in *CCAM* vis à vis *OCAM* to information-driven price movements that are unrelated to price impacts of trading, explaining why the association between the numerator and the denominator of the price impact proxy strengthens when overnight returns are excluded.

Our paper contributes to a literature that distinguishes price movements during trading

---

<sup>11</sup>The literature has previously identified the information content of non-synchronous trading at close (see Atchison et al. 1987) and its importance for driving large auto-correlations in the returns of stock indexes.

hours from those when markets are closed (e.g., Cliff et al. 2008; Barclay and Hendershott 2008; Hendershott et al. 2018; Boudoukh et al. 2019; see also French and Roll 1986). Our findings highlight that price movements that are unassociated with trading are likely irrelevant for liquidity measurement. In addition, our paper weighs in on the debate about why the Amihud measure is priced. Finally, our analysis suggests that overnight price movements largely reflect information arrival, extending insights of Jones et al. (1994), who find that price movements in non-trading episodes during regular trading hours are primarily information driven. These findings also indicate a broader reach of our analysis.

Post-decimalization, empirical analyses of expected returns in U.S. markets show that they have become more liquid. Moreover, high frequency liquidity measures are now available. Nonetheless, Amihud’s (2002) liquidity measure remains vital for measuring liquidity in international markets.<sup>12</sup> Our modification can be employed quite broadly, as data on open prices are available for many countries over long periods of time. For example, Data Stream reports daily open prices from exchanges in nearly 100 countries. For 35 of these countries, reporting of open prices begins well before CRSP’s coverage of daily open commences. The fact that our correction matters for the pricing of liquidity in the U.S. —the world’s leading financial market—strongly suggests that it would matter more in less liquid international markets. In unreported results, we find that the median *OCAM/CCAM* ratio in the Brazilian stock market is roughly half its NYSE counterpart in the 2008–2018 period, indicating the heightened and continuing economic importance of our correction for many international markets.

---

<sup>12</sup>Amihud’s (2002) measure is heavily used in studies of markets outside North America. Amihud et al. (2015) document positive liquidity premia across 45 countries in the 1990–2011 period. Lee (2011) documents priced liquidity risk in international markets. Hung et al. (2014) find weaker post-earnings announcement drifts in international markets that are more liquid according to Amihud’s measure. Boehmer et al. (2015) show that increased algorithmic trading across 42 international markets impacts Amihud’s measure. Chen et al. (2017) document evidence in developing markets that the initial enforcement of insider trading laws is more effective for firms whose stocks witness improvements in their Amihud (2002) liquidity measures. Lang et al. (2015) find that the Amihud (2002) measure is highly correlated with textual lengths of financial reports for firms in the 42 countries studied.

## 2 Data and variables

### 2.1 Amihud illiquidity and modified Amihud

The traditional Amihud (2002) stock liquidity measure, dubbed *CCAM* (for close-to-close Amihud measure), calculates the average of the daily absolute return per dollar traded over a given time period spanning  $D$  consecutive trading days, where daily returns reflect close-to-close returns, incorporating overnight price adjustments and dividend distributions. As in Amihud (2002),  $CCAM_{iy}$  uses  $D_{iy}$  daily observations of stock  $i$  in year  $y$ ,

$$CCAM_{iy} = \frac{1}{D_{iy}} \sum_{d=1}^{D_{iy}} \frac{|R_{idy}|}{DVOL_{idy}}, \quad (1)$$

where  $R_{idy}$  and  $DVOL_{idy}$ , respectively, are stock  $i$ 's return and dollar trading volume on day  $d$  in year  $y$ ; and  $D_{iy}$  is the number of days for which trading volume for stock  $i$  in year  $y$  is non-zero.<sup>13</sup> In our rolling regression analyses, we use an alternative construction that updates measures monthly. Thus, instead of using annual averages that we indexed by  $y$  in (1), we average the daily absolute return per dollar traded over the 12 months ending in month  $t$ :

$$CCAM_{it} = \frac{1}{D_{it}^{t-12}} \sum_{d=1}^{D_{it}^{t-12}} \frac{|R_{idt}^{t-12}|}{DVOL_{idt}^{t-12}}. \quad (2)$$

As such, the measure in month  $t$  uses stock  $i$ 's  $D_{it}^{t-12}$  daily observations, for days with non-zero trading volume, in the previous 12 months rather than from the previous calendar year.

Our open-to-close version, *OCAM*, instead uses the open-to-close absolute return to construct daily absolute returns per dollar traded. Stock  $i$ 's open-to-close return on day  $d$  is

$$OCR_{id} = \frac{P_{id}^c}{P_{id}^o} - 1, \quad (3)$$

---

<sup>13</sup>Thin trading may bias *OCAM* toward zero, attenuating estimates of liquidity premia. On a trading day when a stock trades only at the open auction, but nowhere else in the trading day, trading volume will be non-zero but open-to-close return will be zero. *OCAM* takes a value of zero on such stock-days, classifying the stock as being perfectly liquid, while thin trading likely indicates an illiquid stock. However, such events are quite rare.

where  $P_{id}^o$  and  $P_{id}^c$  are the open and close prices, respectively. As such, the analogues to the traditional Amihud’s measures defined in equations (1) and (2) are

$$OCAM_{iy} = \frac{1}{D_{iy}} \sum_{d=1}^{D_{iy}} \frac{|OCR_{idy}|}{DVOL_{idy}} \quad (4)$$

and

$$OCAM_{it} = \frac{1}{D_{it}^{t-12}} \sum_{d=1}^{D_{it}^{t-12}} \frac{|OCR_{idt}^{t-12}|}{DVOL_{idt}^{t-12}}. \quad (5)$$

## 2.2 Data and variable definitions

Our main sample runs from January 1, 1963 to December 31, 2019, and contains trade and price information, focusing on all NYSE- and AMEX-listed stocks. In robustness analyses, we extend the sample to include NASDAQ-listed stocks. We obtain daily closing prices, trading volumes, and dividend distributions from Daily CRSP.<sup>14</sup> We match these daily observations with open prices obtained from Global Financial Data (GFD).<sup>15</sup> For stock-days with open price observations in GFD, we match daily observations across CRSP and GFD using unique combinations of security identifiers PERMNO, PERMCO, and CIK. As a result, out of 29,225,292 unique daily observations in CRSP, we successfully match 25,468,167 observations with GFD.<sup>16</sup> This matching accounts for over 91% of stock-year sets of CRSP observations with GFD.<sup>17</sup> We obtain monthly returns, prices, dividend distributions, and number of

<sup>14</sup>We follow the procedure proposed by Gao and Ritter (2010) to adjust daily trading volumes that are overstated due to the Nasdaq trade recording routines.

<sup>15</sup>Open prices reported by GFD reflect transaction prices when there is a transaction at open; the mid-point of best bid and ask at open is reported when a transaction price is not available. As a robustness check of GFD opening prices, we also estimate our asset pricing model for 1993–2019 using opening price data from CRSP. Findings are unaffected by the data source.

<sup>16</sup>To control for potential data errors in CRSP or GFD, we use similarity in closing prices reported by CRSP and GFD, dropping a matched stock-day observation if its CRSP closing price deviates from that in GFD by more than 0.1%. For example, a stock day with a CRSP closing price of \$20.03 and a GFD closing price of \$20 is dropped. This filter binds for less than 3% of matched daily observations.

<sup>17</sup>The attributes of the matched GFD-CRSP set of stocks correspond very closely to those of the entire population of stocks covered by CRSP, indicating that the matched subsample is representative. The ratio of means for the GFD-CRSP sample to the CRSP population are all very close to one for key stock characteristics at the beginning of each year: 1.009 for market capitalization, 1.046 for share price, and 1.002 for volatility.

shares outstanding from Monthly CRSP. We match these monthly data with 1-month T-bill rates. We exclude a stock-year set of observations if that stock’s daily closing price is below \$1 on any day in the preceding calendar year.<sup>18</sup> We follow Amihud (2002) by excluding stock-month observations that are among the 1% least liquid in each month, according to Amihud’s measure. To address the convention of reporting the same price for high, low, open and close when data from historical records cannot be retrieved, we exclude a stock from the sample if its daily open and close prices are equal on all trading days of any given year.

We construct cross-sections of stock characteristics and merge them with cross-sections of monthly returns in two ways. First, following Amihud (2002), we calculate market betas of size portfolios (deciles of market capitalizations at the end of the previous year),  $\beta_{py}^{mkt}$ , using daily stock and equally-weighted market returns every year—we use  $\beta_{py}^{mkt}$  for  $\beta_{iy}^{mkt}$  if stock  $i$  is in portfolio  $p$  in year  $y$ . We then compile the following stock-specific measures at annual frequencies: Dividend yield,  $DYD_{iy}$ , is defined as the ratio of total dividend distributions in a year divided by the closing price at the end of the year. Annual measures of momentum are the returns over the last 100 days of the year,  $R100_{iy}$ , and the realized return over the earlier remaining days of the year,  $R100YR_{iy}$ . Annual return volatility is captured by the annual standard deviation of daily returns per year,  $SDRET_{iy}$ . Market capitalization,  $M_{iy}$ , is the product of shares outstanding and the closing price at the end of the year. We match these annual measures with each of the monthly return observations of the relevant stock over the following year, to construct an unbalanced monthly panel.

Our second approach addresses the possibility that using the same annual measures of stock characteristics to explain returns in each of the 12 monthly cross-sections in the following year adds noise to later month observations in a year—the previous year’s measure grows less germane. Hence, we follow Amihud and Noh (2020), and use a rolling regression approach, constructing stock characteristics at monthly frequencies, and then matching

---

<sup>18</sup>Our findings are robust to replacing the “penny-stock” filter with one that excludes stocks with end-of-previous-year’s closing prices below \$5, as in Amihud (2002).

them with monthly returns in the following months.<sup>19</sup> Momentum measures  $RET_{i,t-1}$  and  $RET_{i,t-2}^{t-12}$ , respectively, capture compound returns over the preceding month and the 11 months before that. Market capitalization,  $M_{i,t-2}$ , is the product of shares outstanding and the closing price two months earlier. We use Compustat to construct book value measures as the sum of stockholders' equity and deferred taxes at the end of each fiscal year. The book-to-market ratio,  $BM_{i,t-1}$ , is created by dividing the most recently reported book value by the market-capitalization at the end of month  $t - 1$ .<sup>20</sup> Relevant liquidity measures,  $OCAM_{i,t-2}$  and  $OCAM_{i,t-2}$  are constructed using daily observations from the 12-month period ending in month  $t - 2$ —see equations(2) and (5).

We use monthly TAQ data to collect time stamps of the last transaction on each trading day during regular trading hours (9:30am–4:00pm EST), for all NYSE- and AMEX-listed stocks in the 1993–2013 period.<sup>21</sup> We use the temporal distance between these time stamps and 4:00pm EST, in hours, to construct a measure of the extent of non-trading by stock-year. For a given year, we match these observations with our main sample, described above, using NCUSIP from CRSP and CUSIP from TAQ. For observations without such links, we match SYMBOL from TAQ with TSYMBOL from CRSP. From Daily Indicators by WRDS, we obtain daily measures of time-weighted average percentage quoted and trade-weighted average percentage effective spreads. We construct each transaction's effective spread in two ways, one based on the mid-point of prevailing quotes at the same second, and one based on the mid-point from the previous second. Daily estimates of Kyle's  $\lambda$  are constructed based on both suppressed and unconstrained price impact intercepts using price change and net order flow estimates from 5-minute intervals.

---

<sup>19</sup>See also Barardehi et al. (2019) or Lu and Shu (2017) for similar approaches.

<sup>20</sup>We use the “linktable” provided by Wharton Research Data Service to match stocks across CRSP, Compustat, and GFD. As such, a stock with no such links is excluded from the sample used in this analysis.

<sup>21</sup>The TAQ database is available for 1993 and after.

### 3 Liquidity, overnight returns, and stock attributes

Figure 2 on page 3, which plots the temporal evolution in the distribution of the ratio of open-to-close and close-to-close Amihud (2002),  $ROC_{iy} = \frac{OCAM_{iy}}{CCAM_{iy}}$ , by tercile, reveals that overnight price movements play a major role in driving both cross-sectional and temporal variation in the traditional Amihud (2002) measure,  $CCAM$ . Including overnight price movements also inflates the measure for most stocks. Importantly, as the evolution of the tercile-specific medians of this ratio indicate, the contamination driven by overnight price movements is time-varying and declining, but it does not disappear. Table 1 presents medians of several stock characteristics across six  $ROC$  quintiles, after sorting the sample on  $ROC$  on a year-by-year basis. Save for market capitalization and share price, stocks in different such categories do not seem to possess materially different characteristics, indicating that the cross-stock variability in the mismeasurement of liquidity is driven at least in part by factors other than these basic stock characteristics.

One might wonder whether the patterns documented in Figure 2 could reflect temporal variation in the composition of common stocks over our long sample period. One could posit that the disparity between  $OCAM$  and  $CCAM$  might be due to some unknown stock characteristics, and that the presence of stocks with small  $ROC_{iy} = \frac{OCAM_{iy}}{CCAM_{iy}}$  in the cross-section has varied over time. In fact, the number of publicly listed firms varies significantly over the past few decades (Kahle and Stulz 2017).<sup>22</sup> To preclude the possibility that results are driven by changes in sample composition, we show that the temporal variation in the cross-section of  $ROC$  is robust to sample composition. Figure 3 focuses on the sample featuring the 700 stocks with the largest market-capitalizations at the end of the previous year. The figure reveals patterns in the year-specific cross-stock terciles of  $ROC_{iy}$  that are qualitatively identical to those in Figure 2; there is a moderate upward shift in early decades in the bottom quartile and median, with minimal shifts for the top quartile (less than 2 percentage points).

---

<sup>22</sup>The median number of stocks per year in our main sample is 1285 across all years.

This finding is consistent with the result in Table 1 that *ROC* is largely unrelated to several key stock characteristics. As one might expect, in the sample of large-/mid-cap stocks, the cross-sectional variation shrinks a few years earlier than the full sample displayed in Figure 2.

Because our focus is on the measurement of liquidity, the patterns in Figure 2 are especially relevant for smaller and more thinly-traded stocks, which are generally perceived as relatively less liquid. We document evidence of this by examining how the ratio of *OCAM*<sub>*iy*</sub> to *CCAM*<sub>*iy*</sub>, *ROC*<sub>*iy*</sub>, varies with measures of size and turnover and other stock characteristics. We measure stock *i*’s size in year *y* by the natural log of its market capitalization at the end of year *y* − 1. Our turnover measure for stock *i* is the natural log of its average daily turnover<sup>23</sup> in year *y* − 1, i.e.,  $\ln(TR_{i,y-1})$ . We fit a panel of annual *ROC* measures against these stock characteristics for the time period 1964–2019, clustering standard errors at both stock and year levels to account for the possibility of inflated *t*-statistics driven by auto-correlated error terms (see Petersen 2009). We model the cross-sectional variation in *ROC*<sub>*iy*</sub> using the specification

$$ROC_{iy} = \delta_0 + \Delta_1 X_{i,y-1} + \text{fixed effects} + \epsilon_{iy}, \quad (6)$$

which includes both stock and year fixed effects.  $\Delta$  is the vector of coefficients, and vector *X* contains natural logs of market-cap and turnover, market beta, dividend yield, volatility, and share price. Because *OCAM* is closer to *CCAM* when *ROC*<sub>*iy*</sub> is larger, a positive coefficient in  $\Delta$  means that the two measures differ by less when the respective characteristic is larger.

Table 2 shows large, positive and robust significant relationships between *ROC* and market beta, stock size, and turnover. This finding indicates that potential “contamination” by overnight price movements is more pronounced for smaller and more-thinly traded stocks that tend to feature greater market exposure. Variation in stock characteristics explain over 33% of the variation in *ROC*. These associations remain after controlling for both stock and

---

<sup>23</sup>Daily turnover is defined as the ratio of the number of shares traded daily to the corresponding number of shares outstanding.



year fixed effects.

The fact that the impact of overnight price movements on measures of liquidity is greater for less liquid stocks suggests that *OCAM* may be priced differently than *CCAM* in the cross-section of stock returns. As preliminary evidence of this possibility, we show that the differences between *CCAM* and *OCAM* reflect quite different cross-sectional rankings of stocks based on the two liquidity measures. In particular, the differences do not represent a simple scaling effect: rank correlation statistics between *CCAM* and *OCAM* are well below one, especially for less liquid stocks and in earlier years of the sample. To quantify this, we sort stocks each year into top 30%, middle 40%, and bottom 30% liquidity according to *CCAM*. We then calculate Kendall’s  $\tau$  statistics every year within each liquidity group, and calculate the average statistic across different years in the entire sample period or sub-period.

*OCAM* and *CCAM* order stocks in the cross-section very differently. Table 3 shows that rank correlations of *OCAM* and *CCAM* over the entire sample period are far below one. These correlations are much smaller for less liquid stocks, going from 72% for the least liquid stocks to 94.9% for the most liquid ones. Consistent with the patterns in Figure 2, rank correlation statistics across all liquidity groups rise substantially over time.<sup>24</sup> These findings underscore that *CCAM* and *OCAM* measure cross-sectional differences in stock liquidity differently, and hint at potentially different pricing of the two measures in the cross-section.

## 4 Modified Amihud measure and liquidity premia

Our asset pricing analysis largely adopts the specification used by Amihud (2002). This lets us demonstrate the merits of our proposed correction to Amihud’s liquidity measure in the context where it was introduced. Using both Amihud’s classic close-to-close price impact measure, *CCAM*, and our open-to-close modification, *OCAM*, we reproduce the basic cross-sectional results in Amihud (2002)—in particular, the sixth, eighth, and ninth

---

<sup>24</sup>Unreported analyses verify qualitatively similar patterns based on Pearson and Spearman correlations.

columns of Table 1 (p. 41)—in our setting.<sup>25</sup> We then contrast findings based on *CCAM* with those obtained using *OCAM* to highlight the value of our modification. We model the cross-section of returns in month  $t$  of year  $y$ ,  $R_{ity}$  as in Amihud (2002), by estimating

$$R_{ity} = k_{ty}^0 + \sum_{j=1}^J k_{ty}^j X_{i,y-1}^j + \epsilon_{ity} \quad (7)$$

using the Fama-MacBeth approach. Here,  $X_{i,y-1}^j$  is stock  $i$ 's  $j^{th}$  characteristic, measured using data from year  $y - 1$ ;  $k_{ty}^j$  is the  $j^{th}$  characteristic's loading; and  $\epsilon_{ity}$  is an error term.

We correct for potential upward biases due to microstructure noise by employing weighted least squared (WLS) estimates. We follow Asparouhova et al. (2010) in weighting monthly observations using gross returns  $(1 + R_i)$  from the previous month.<sup>26</sup> We first follow the approach in Amihud (2002) of dividing each *CCAM* <sub>$iy$</sub>  and *OCAM* <sub>$iy$</sub>  observation by its respective sample mean across stocks in year  $y$ . This centers each liquidity measure to have a mean of one. Thus, the coefficients on liquidity measures reflect liquidity premia: the additional return investors require for holding the stock with average liquidity, compared to the idealized, fully-liquid stock. Centering in this way ensures that any differences in liquidity premia are not mechanically driven by the fact that *OCAM* is on average smaller than *CCAM*. We reinforce these findings by showing that qualitatively identical results obtain when we estimate liquidity premia using standardized liquidity measures, i.e., when we estimate the premia associated with a one standard deviation increase in each liquidity measure.

Table 4 shows that our findings regarding the relationship between expected returns and stock characteristics align with those in Amihud (2002). In particular, we find positive and significant coefficients on stock (il)liquidity and measures of momentum, but negative and significant coefficients on stock size and return volatility measures. Crucially, substituting

---

<sup>25</sup>This is not a replication. First, our sample focuses on NYSE- and AMEX-listed *common* shares, while Amihud (2002) studies *all* NYSE-listed stocks. Second, we can match GFD and CRSP for a little about 91% of stock years. Third, the sample period in our study is far longer.

<sup>26</sup>This correction is more necessary in the pre-decimalization era where the tick size was one-eighth, but it is likely very conservative in recent years with the penny tick and tiny bid-ask bounce.

*OCAM* for *CCAM* leads to liquidity premia that are nearly *double* those obtained using *CCAM*, but it leaves the coefficients on other stock characteristics essentially unchanged. Importantly, the larger coefficient on *OCAM* is not offset by an increase in standard errors. As a result, *OCAM* significantly explains the cross-section of returns even post 2001, while *CCAM* is not priced in recent years. The last two columns in Table 4 estimate the model using a subsample that excludes the largest 400 stocks based on previous year’s market-cap measures, reinforcing that liquidity premia based on Amihud measure still exist post 2001.<sup>27</sup>

Because we center each liquidity measure on its cross-sectional mean, the differences in estimated liquidity premia cannot be due to *CCAM* being, on average, larger than *OCAM*.<sup>28</sup> We reinforce this finding by estimating equation (7) using standardized liquidity measures  $CCAM^{std}$  and  $OCAM^{std}$  that are normalized to have 0 mean and unit standard deviation in each cross-section. The corresponding coefficients on the measures represent the liquidity premia associated with one standard deviation increases in *OCAM* and *CCAM*. This normalization accounts for the possibility that the differences in estimated liquidity premia in Table 4 may reflect differences in the cross-stock dispersion of the two measures. In particular, because *CCAM* includes overnight price movements, it picks up the close-to-close return volatility in its numerator; whereas *OCAM* only picks up volatility from regular hours. As a result, *CCAM* is slightly more dispersed than *OCAM* in the cross-section. Table 5 shows that even after adjusting for differences in the means and standard deviations of the two measures, liquidity premia based on  $OCAM^{std}$  still exceed those based on  $CCAM^{std}$  by 50-75%.

One might also wonder whether differences in the liquidity premia obtained using *CCAM* and *OCAM* are affected by cross-sectional variation in the timing of earnings announcement

---

<sup>27</sup>One reason to exclude large stocks is that Amihud’s measure over-aggregates offsetting buy and sell volume (see Barardehi and Bernhardt 2019 and Barardehi et al. 2019) for an analysis of the consequences of such over-aggregation for a host of phenomena at higher frequencies), a feature that has been exacerbated post-decimalization by the far higher trading volumes. Over-aggregation tends to be far more acute for larger stocks, introducing significant noise in the liquidity measures and biasing estimates toward zero.

<sup>28</sup>Centering is important because *ROC* is less than one for the vast majority of stock-years.

releases. Historically, some firms announced earnings during regular hours, whose impacts would affect both *CCAM* and *OCAM*, while other firms released earnings after hours, which would only impact *CCAM*. The number of firms releasing earnings during regular hours fell sharply in later years—by 2011, less than 2% of firms released earnings during regular hours (deHaan et al. 2015). To account for this, we exclude daily observations from the three-day windows around earnings announcement dates reported by Compustat before constructing “earnings-adjusted” illiquidity measures,  $CCAM^{ea}$  and  $OCAM^{ea}$ . Table 6 shows that the findings are essentially unchanged when we exclude earnings announcement windows: liquidity coefficients increase very slightly from those in Table 4.

Our results are related to the literature that examines temporal changes in liquidity premia. Our qualitative findings accord with diminishing liquidity premia among common stocks (e.g., Ben-Rephael et al. 2013). However, this literature typically finds that premia disappear completely in recent years, especially when one accounts for possible microstructure noise (e.g., Asparouhova et al. 2010). We find that the improvements embedded in *OCAM* mean that it significantly explains the cross-section of expected returns even in recent years.<sup>29</sup> We reinforce this point in Appendix A.2 where we rely solely on CRSP data from the 1992–2019 period. We find that while liquidity premia based on *CCAM* cease to exist after one adjusts for microstructure noise, premia based on *OCAM* remain statistically significant. We next show that, *OCAM* is, indeed, a superior (less noisy) measure of liquidity than *CCAM*.

## 5 Is *OCAM* a better liquidity measure than *CCAM*?

We establish that *OCAM* is more strongly associated with a wide array of benchmark measures of trading costs than *OCAM*. We select two sets of trading costs benchmarks, a set of “standard” high-frequency measures of trading costs, and a set that is publicly available for

---

<sup>29</sup>This finding becomes even more pronounced when one focuses on the set of smaller cap stocks, which are both less liquid, and have less turnover, resulting in reduced over-aggregation.

long histories that overlap with much of our sample period. The high-frequency measures include two versions of relative effective spreads, relative quoted spreads, and two versions of estimates of Kyle’s  $\lambda$ , all constructed annually for the 1993–2013 period. The long-history group contains measures of effective trading costs developed by Hasbrouck (2009), namely, two versions of moments estimates of costs plus Gibbs estimates of costs that are available at annual frequencies for the 1964–2013 period.<sup>30</sup>

For each benchmark, we test the null that *CCAM* and *OCAM* are similarly correlated with the benchmark. Every year, we calculate the correlation coefficients of *CCAM* and *OCAM* vis à vis a given benchmark. We then run difference-in-average correlation tests using correlation coefficient pairs obtained from the years falling in the respective sample period. Table 7 shows that the average correlations between *OCAM* and trading costs benchmarks are economically and statistically larger than the analogues for *CCAM*, with differences in correlations ranging from 5.6 to 14.4 percentage points. The differences in correlations are notably higher using Hasbrouck’s measures reflecting that *CCAM* deviates more significantly from *OCAM* in earlier years.

We next show that *OCAM*’s ability to better capture trading costs manifests itself in its stronger ability to explain the cross-section of stock returns; something that does not directly follow from the larger liquidity premia associated with *OCAM*. Due to the high correlation between *CCAM* and *OCAM*, using them both in the same regression leads to a multi-collinearity problem. Additionally, we are interested in the incremental explanatory power of one measure, with respect the other, for the cross-section of returns. we decompose each version of the Amihud measure into two linearly orthogonal components with respect to the other version.<sup>31</sup> We estimate

$$CCAM_{ity} = \lambda_{ty}^0 + \lambda_{ty}^1 OCAM_{ity} + Z_{ity} \quad (8)$$

---

<sup>30</sup>Annual data for this time period are available for download at Professor Joel Hasbrouck’s website: <http://people.stern.nyu.edu/jhasbrou/Research/GibbsEstimates2005/>.

<sup>31</sup>Lou and Shu (2017) and Barardehi et al. (2019) adopt similar approaches.

and

$$OCAM_{ity} = \gamma_{ty}^0 + \gamma_{ty}^1 CCAM_{ity} + \tilde{Z}_{ity} \quad (9)$$

every month, and store the corresponding residuals  $Z_{ity}$  and  $\tilde{Z}_{ity}$ . We then estimate Equation (7) using the Fama-MacBeth approach, augmenting the set of independent variables by either  $\tilde{Z}$  or  $Z$  to infer the incremental explanatory power of the alternative measure with respect to the baseline. When  $CCAM$  is used as the baseline liquidity measure, we add  $\tilde{Z}$  to the set of independent variables. Because  $CCAM$  and  $\tilde{Z}$  are, by construction, orthogonal, the coefficient on  $\tilde{Z}$  captures the change in liquidity premia as one augments the incremental information content in  $OCAM$  that is not included the baseline measure  $CCAM$ , i.e., the residuals from Equation (9). In the same fashion, when the baseline measure is  $OCAM$ , we add  $Z$  to the set of independent variables in Equation (7).

Table 8 shows that when we remove all information contained in  $CCAM$  related to variation in  $OCAM$ , the residual possesses no incremental information in addition to that contained in  $OCAM$ . In contrast, the residual from regressing  $OCAM$  on  $CCAM$ , i.e.,  $\tilde{Z}$ , is priced over the entire sample period, as well as in the four last decades of the previous millennium when  $CCAM$  serves as the baseline liquidity measure. In essence, little relevant information is lost when we take out the information contained in the classical  $CCAM$  Amihud measure. Phrased differently, including the noise in the form of close-to-open returns appears to add measurement error that attenuates estimates of liquidity premia.<sup>32</sup>

## 6 Amihud and Noh (2020) vs. Lou and Shu (2017)

Despite the widespread use of Amihud’s (2002) liquidity measure as a proxy of stock liquidity, the literature has paid little attention until recently to why it is priced in the cross-section of stock returns. Lou and Shu (2017, LS) initiated the debate by arguing that the sole source of

---

<sup>32</sup>In unreported robustness, we find qualitatively identical results when  $Z$  or  $\tilde{Z}$  is employed as the sole liquidity measure, i.e., when  $CCAM$  or  $OCAM$  are not included.

pricing of Amihud’s (2002) measure is the cross-sectional variation in average dollar-volume, dismissing the relevance of the liquidity measure for price impacts of trading. Amihud and Noh (2020, AN) identified an error in how LS decompose the Amihud liquidity measure into its components. They show that much of the pricing is due to a component LS missed: the correlation between the numerator, daily absolute price changes, and the denominator, dollar volumes. In this section, we first discuss the components of the Amihud liquidity measure, explaining the conditioning under which this measure may be able to capture price impacts of trading. We then highlight the effects of our proposed modification to the Amihud measure on its components, reinforcing AN’s argument and uncovering why *OCAM* outperforms *CCAM* in explaining the cross-sections of stock returns.

Amihud’s (2002) liquidity measure approximates the price impacts of trading at daily frequencies using the ratio of absolute daily return to daily dollar volume,  $\frac{|R|}{DVOL}$ , defining the liquidity measure as  $E \left[ \frac{|R|}{DVOL} \right]$ . This approximation relies on several assumptions. First, trading moves prices—consistent with classical asymmetric information models (e.g., Kyle 1985) or inventory models (e.g., Demsetz 1968) of microstructure. Hence, daily price changes, on average, correctly reflect order flow imbalances; i.e., the daily return is positive with positive net order flow, and negative with negative net order flow. This assumption is consistent with established regularities, including the positive association between volatility and trading volume, and the positive estimates of Kyle’s  $\lambda$ . Second, one can proxy for the variation in absolute net order flow using the variation in dollar volume. To approximate the price impact per-dollar traded, Amihud’s measure divides daily price change by the total dollar volume, rather than the (unobserved) net order flow. This assumption understates the price change per unit net order flow because total dollar volume always exceeds absolute net order flow. Third, daily aggregation provides a reasonable sampling frequency. This approximation is reasonable in times of modest trading activity; but it leads to over-aggregation issues once daily trading volumes explode post decimalization (see Barardehi et al. 2019),

which, in turn, biases estimates of liquidity premia downward toward zero.

A statistical decomposition of Amihud's liquidity measure, as defined above, implies that

$$CCAM = E \left[ \frac{|R|}{DVOL} \right] = E[|R|] E \left[ \frac{1}{DVOL} \right] + Cov \left( |R|, \frac{1}{DVOL} \right). \quad (10)$$

LS effectively assume that  $Cov \left( |R|, \frac{1}{DVOL} \right) = 0$  when they decompose the natural log of Amihud measure as

$$\ln \left( E \left[ \frac{|R|}{DVOL} \right] \right) = \ln(E[|R|]) + \ln \left( E \left[ \frac{1}{DVOL} \right] \right). \quad (11)$$

However, assuming that there is no association between absolute return and trading volume is completely at odds with the central premise that trading has price impacts. That is, prices should rise when net order flow is positive, and they should fall when net order flow is negative. Amihud's measure translates this positive correlation between net order flows and returns into a positive correlation between absolute return and trading volume that cannot be ignored. Pointing this out, AN show that the nonlinear approximation of the correlation between daily absolute returns and dollar volumes,<sup>33</sup>

$$DIF^{CCAM} = \ln \left( E \left[ \frac{|R|}{DVOL} \right] \right) - \ln(E[|R|]) - \ln \left( E \left[ \frac{1}{DVOL} \right] \right), \quad (12)$$

is priced in the cross-section of stock returns even when  $\ln(E[|R|])$  and  $\ln \left( E \left[ \frac{1}{DVOL} \right] \right)$  are present as controls.

We next provide insights into why replacing the close-to-close absolute return,  $R$ , by the open-to-close absolute return,  $OCR$ , improves the pricing ability of the Amihud measure and strengthens its associations with benchmark trading cost measures. A statistical decomposition of  $OCAM$  yields

$$OCAM = E \left[ \frac{|OCR|}{DVOL} \right] = E[|OCR|] E \left[ \frac{1}{DVOL} \right] + Cov \left( |OCR|, \frac{1}{DVOL} \right). \quad (13)$$

---

<sup>33</sup>See AN's Appendix analysis A1.



Because the inverse of dollar volume is common in Equations (10) and (13), any disparity between *OCAM* and *CCAM* must reside in the differences in the two covariance terms and the two absolute return terms. As such, we relate the patterns documented in Figure 2 to potential disparities between the above terms. This leads us to define the ratios

$$ROC^{|ret|} = \frac{E[|OCR|]}{E[|R|]}, \quad (14)$$

$$ROC^{Cov} \equiv \frac{Cov^{OCAM}}{Cov^{CCAM}} = \frac{Cov\left(|OCR|, \frac{1}{DVOL}\right)}{Cov\left(|R|, \frac{1}{DVOL}\right)}, \quad (15)$$

where the empirical inputs in each ratio are estimated using daily observations annually.

Figure 4 plots medians of the ratios in equations (14) and (15) by terciles of  $ROC_{iy}$  (see equation 6) over time. Figures 2 and 4 Panel A have the same scale, revealing that the variation in average absolute returns,  $ROC^{|ret|}$ , is a small fraction of that in  $ROC$ . Panel B reveals that using open prices in the construction of Amihud's measure sharply affects the association between its numerator and the denominator inputs.<sup>34</sup> The variation across terciles and over time in  $ROC^{Cov}$  is sizably larger than that of  $ROC^{|ret|}$ . This indicates that excluding overnight price movements from the construction of Amihud's measure does far more than simply shrink the volatility included in the measure that, in turn, scales down the measure. We next highlight this by documenting how  $Cov\left(|OCR|, \frac{1}{DVOL}\right)$  is priced.

AN decompose  $\ln(CCAM)$  into three components:  $\ln(|\overline{R}|)$ ,  $\ln(IDVOL) \equiv \ln\left(\frac{1}{DVOL}\right)$ , and  $DIF^{CCAM}$ . They also consider LS's version of *CCAM* measure that implicitly assumes  $Cov(|R|, \frac{1}{DVOL}) = 0$ ; we dub this measure  $\ln(LS^{CCAM}) \equiv \ln\left(|\overline{R}| \times \frac{1}{DVOL}\right)$ . We consider the analogues of these measures in the context of *OCAM*. That is, we produce  $\ln(OCAM)$ ,  $\ln(IDVOL)$ ,  $\ln(LS^{OCAM}) \equiv \ln\left(|\overline{OCR}| \times \frac{1}{DVOL}\right)$ , and  $DIF^{OCAM} \equiv$

---

<sup>34</sup>Co-variances are affected by the levels of the constituent variables; in Appendix A.4, we establish that the ratio of the analogous correlation coefficients, which are free of scale effects, exceeds one for the vast majority of stock-years. That is, the correlation between Amihud measure's numerator and denominator inputs rises when open-to-close returns, rather than close-to-close returns, are used in the numerator of the price impact proxy.

$\ln(OCAM) - \ln(LS^{OCAM})$ . We follow AN as closely as possible to obtain estimates that are the analogues of columns 1–3 in AN’s Table 1. We find qualitatively similar results.<sup>35</sup> Table 9 shows that for both versions of the Amihud measure, each of their sub-components explains the cross-section of expected returns. Most importantly, in the full, unrestricted specification (column (3)) in Panels A and B of Table 9, the nonlinear approximation of the correlation between daily absolute returns and dollar volumes, i.e.,  $DIF$ , is more strongly priced based on  $OCAM$  than  $CCAM$ . In particular, the increase in excess return associated with a one standard deviation increase in  $DIF^{OCAM}$  is 15.9 bps; whereas that for  $DIF^{CCAM}$  is only 8.3 bps. This suggests that the tighter link between the numerator and denominator of the  $OCAM$  measure largely underlies why liquidity premia based on  $OCAM$  exceed those based on  $CCAM$ .

We next reinforce this argument by showing that overnight price movements are, indeed, primarily information-driven, not trade-driven.

## 7 Information-driven overnight returns

We conclude our analysis by documenting the information-driven role of overnight price movements in distorting  $CCAM$  from capturing liquidity properly. To do this, we construct a proxy for overnight information arrival using the extent of non-synchronous trading near close that exploits the variation in the time distance between the last transaction of the day and the official close (4:00pm EST) of the trading day. We then show that the cross-stock variation in our proxy—the mean extent of non-trading before close—explains the vast bulk of the cross-stock variation in the differences between  $OCAM$  and  $CCAM$  post-1993.

The economics underlying our non-synchronous trading proxy is that if a stock experi-

---

<sup>35</sup>The slight quantitative differences likely reflect that (i) our sample period is 1964–2019 while AN analyze data from 1955–2016; and (ii) our cross-section of NYSE- and AMEX-listed firms is limited to those whose book-values are reported by Compustat, while AN obtain book values from Compustat *and* Moody’s databases so they have slightly larger cross-sections in the 1950s and 1960s.

ences longer periods of non-trading before close and expected information flows per unit time are similar across stocks, then it should see more accumulated information after the final trade, but before close. That information is impounded into price at open on the next trading day, leading to greater overnight price movement.<sup>36</sup> As such, our design exploits differential information arrival before close to proxy for total information arrival during overnight hours.

Cross-stock variation in non-trading, which may reflect a lack of liquidity and other market microstructure frictions, introduces heterogeneity in the mean stock of accumulated “overnight” information flow. Figure 5 illustrates the mapping between the extent of non-trading prior to close on a given trading day and the amount of accumulated information that is expected to be impounded in price the following trading day.

Using intraday transaction data from the TAQ database between 1993–2013, we measure the time distance, in hours, between the last transaction and close each trading day. We then average these distances every year for each stock to construct  $HTC_{iy}$  (hours to close), which measures the average extent of non-trading before close in a given year. We average over an entire year to mitigate the measurement error in the time distance measure. This measurement error can arise due to random variation in the time distance, or the fact that closing prices are determined in a special call auction that “is strongly associated with ETF ownership and institutional rebalancing” (Bogousslavsky and Muravyev 2019). Such rebalancing adds noise to the information content of closing prices, which could only weaken the ability of  $HTC_{iy}$  to proxy overnight information arrival.

Figure 6 shows that, pre-decimalization, non-trading exhibits remarkable cross-stock variation. For instance, in 1993, the first and third quartiles of the mean time distance between

---

<sup>36</sup>When there is no closing price available, i.e., when there is not a “closing cross” that corresponds to the final transaction at close, both CRSP and GFD report the midpoint of the best bid and ask prices at 4:00pm as the “close price.” Importantly, this type of closing price, which is not associated with trading, may differ from the price associated with the last transaction the same trading day. If anything, when such differences are meaningfully large, they introduce noise in our proxy of overnight information accumulation, i.e., the time distance between the last transaction of the day and 4:00pm, attenuating our results.

the last trade and close are roughly 8 and 50 minutes, respectively—the last transaction was well before close for most stocks. This variation begins to vanish as decimalization is implemented from 1997 to 2001, disappearing after 2007, once automated trading fully dominates markets. These patterns mirror those for the cross-sectional distribution of  $ROC$  in Figure 2, suggesting that variation in non-trading is related to variation in  $ROC$  (the extent of disparity between  $CCAM$  and  $OCAM$ ). Figure 7 plots  $ROC_{iy}$  against  $HTC_{iy}$ . In particular, the figure reveals a strong association between the extent of non-trading and differences between  $CCAM$  and  $OCAM$ :  $ROC$  declines with  $HTC$  with a slope of  $-1/3$ . Figures 6 and 7 reveal qualitatively similar results after controlling for changes in sample composition by focusing on the 700 largest stocks according to market-capitalizations at the end of the previous year.

To quantify the ability of  $HTC$  to explain variation in the disparity between  $CCAM$  and  $OCAM$ , we add  $HTC_{iy}$  to a regression of  $ROC_{iy}$  on stock characteristics. The panel regression includes stock fixed effects to account for time-invariant stock characteristics that may drive  $ROC$ . We also use year fixed effects to account for systematic temporal variation (e.g., varying sample composition) in  $ROC$ . Additionally, we control for either  $CCAM$  or  $OCAM$  to preclude the possibility that findings are influenced by the level of stock liquidity. Finally, we substitute  $HTC_{iy}$  by the lag of non-trading before close,  $HTC_{i,y-1}$ . Table 10 shows that  $HTC$  strongly explains the variation in the disparity between  $CCAM$  and  $OCAM$  even after controlling for stock characteristics and fixed effects. Variation in  $HTC$  explains 77% of the variation in  $ROC$ . In fact, adding other stock characteristics as well as stock and year fixed effects lead to a relatively modest increase in adjusted- $R^2$ , raising it from 0.77 to 0.85. The negative coefficient on  $HTC$  indicates that longer periods of non-trading before close are associated with greater disparities between  $CCAM$  and  $OCAM$ . Concretely, a one-hour increase in the average duration of non-trading before close is associated with roughly a 0.3 (30 percentage point) decline in  $ROC$ —i.e., a 5 percentage point decline in  $ROC$  for each additional 10 minutes of non-trading before close. The estimated slope coefficient is very

close to the visually discerned slope of  $-1/3$  from Figure 7.

To provide additional evidence of the co-variation between  $ROC$  and  $HTC$ , we document the temporal evolution of the cross-sectional variation in  $ROC$  after accounting for the variation explained by  $HTC$ . To do this, for each year, we obtain the residuals from cross-sectional regressions of  $ROC_{iy}$  on  $HTC_{iy}$ , and then investigate the temporal evolution of the cross-sectional distribution of these residuals. These cross-sectional regressions feature an average  $R^2$  of 74.5%, further underscoring the relevance of information-driven price movements for our results. Figure 8 highlights the ability of  $HTC$  to explain variation in  $ROC$ . In contrast to the large cross-sectional and temporal variations in  $ROC$  highlighted in Figure 6, the residuals of  $ROC$  on  $HTC$  are concentrated around zero (over half of the observations always fall between  $-0.05$  and  $0.05$ ) and are remarkably stable over time, indicating that much of  $ROC$ 's variation is explained by variation in the extent of non-trading near close, especially pre-decimalization and before automation of equity markets.<sup>37</sup>

The fact that accounting for overnight information arrival via our proxy removes essentially all temporal variation in the ratio of  $OCAM$  to  $CCAM$  indicates that  $HTC$  is a good proxy—that we average sufficiently to remove noise in the time-to-close measure—and that differences in the two measures are largely due to overnight information arrival. In turn, this means that the superior ability of  $OCAM$  to explain returns and the higher associated estimated liquidity premia indicate that including the overnight information-driven return in  $CCAM$  serves to add noise to the measure that biases liquidity premia toward zero.

Our findings also have implications for the interpretations of Bogousslavsky and Muravyev (2019). They highlight that closing prices are strongly influenced by non-informationally-based trades of institutional investors, reducing their information content. This means that overnight price movements will be affected, in part, by this noise. This effect only serves to attenuate the relationship between our proxy of overnight information arrival and the dif-

---

<sup>37</sup>Such stability of residuals does *not* obtain when one regresses  $ROC$  on year fixed effects only.

ferences between *CCAM* and *OCAM*. The fact that we find strong co-variation highlights the merits of our methodology and of our proposed correction to Amihud’s measure.

## 8 Conclusion

Amihud’s (2002) liquidity measure (*CCAM*) has been widely used by researchers to study the importance of (or control for) stock liquidity in many financial economics settings. Its usage, in part, reflects the measure’s simple construction using data that can be obtained for long histories and across different markets. The many insights based on it make its precise measurement crucial. Our paper develops and implements simple improvements to this measure.

Our *OCAM* modification uses open-to-close returns, rather than close-to-close returns, to address a time mismatch in the construction of *CCAM*. Our modified measure better explains the cross-section of returns, revealing that liquidity premia are substantially larger than previously believed. *OCAM* is also more strongly correlated with low- and high-frequency measures of trading costs than *CCAM*. Reflecting this, *OCAM* better explains the cross-section of expected returns. Liquidity premia based on *OCAM* are 50-120% larger than those based on *CCAM*.

Including overnight returns in the Amihud measure adds measurement error that sharply attenuates estimates of liquidity premia. This is evident by the finding that the association between the numerator and denominator inputs of the Amihud measure sharply strengthen once we exclude overnight returns. Amihud and Noh (2020) show that this association drives the pricing of *CCAM* in the cross-section. We show both that much of the differences between *OCAM* and *CCAM* are due to how the relationship between the denominator and numerator is affected by our modification, and that the covariance between the denominator and numerator is notably more strongly priced for *OCAM* than *CCAM*.

Finally, we exploit cross-stock and temporal variation in the extent of non-trading be-

fore close as a proxy for variation in information-driven price movements to understand the sources of differences between *CCAM* and *OCAM*. We find that this proxy explains most of the cross-stock variation in these differences. Overall, our paper highlights the importance of excluding information-driven price movements when constructing measures of stock liquidity.

## References

- Acharya, V. V., and L. H. Pedersen. 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77, 375–410.
- Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5, 31–56.
- Amihud, Y., A. Hameed, W. Kang, and H. Zhang. 2015. The illiquidity premium: International evidence. *Journal of Financial Economics*, 117(2), 350–368.
- Amihud, Y. and J. Noh. 2020. Illiquidity and stock returns – II: Cross-section and time-series effects. *Review of Financial Studies*, forthcoming.
- Anand, A., P. Irvine, A. Puckett, and K. Venkataraman. 2013. Institutional trading and stock resiliency: Evidence from the 2007-2009 financial crisis. *Journal of Financial Economics* 108, 773–797.
- Asparouhova, E., H. Bessembinder, and I. Kalcheva. 2010. Liquidity biases in asset pricing tests. *Journal of Financial Economics* 96, 215–237.
- Asparouhova, E., H. Bessembinder, and I. Kalcheva. 2013. Noisy prices and inference regarding returns. *Journal of Finance* 68, 655–714.
- Atchison M. D., K. C. Butler and R. R. Simonds. 1987. Nonsynchronous security trading and market index autocorrelation. *Journal of Finance*
- Barardehi, Y.H., D. Bernhardt, and R.J. Davies. 2019. Trade-time measures of liquidity. *Review of Financial Studies*. 32(1), 126–179.
- Barclay, M. J. and T. Hendershott 2003. Price discovery and trading after hours. *Review of Financial Studies* 16, 1041–73.
- Barclay, M. J. and T. Hendershott 2008. A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance* 15, 839–849.
- Ben-Rephael, A., O. Kadan, and A. Wohl. 2015. The diminishing liquidity premium. *Journal of Financial and Quantitative Analysis* 50, 197–229.
- Boudoukh, J., Feldman, R., Kogan, S., Richardson, M. (2019). Information, trading, and volatility: Evidence from firm-specific news. *Review of Financial Studies* 32(3), 992–1033.



- Brennan, M. J., T. Chordia, and A. Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49, 345–73.
- Belcourt, T. 1996. Three Essays in Applied Microeconomic Theory. PhD Thesis. *Queen's University*.
- Boehmer, E., K. Fong, and L. Wu. 2015. International evidence on algorithmic trading. Working Paper.
- Bogousslavsky, V. and D. Muravyev. 2019. Should We Use Closing Prices? Institutional Price Pressure at the Close. Working Paper.
- Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2012. Robust Inference with Multiway Clustering. *Journal of Business & Economic Statistics* 29, 238–249.
- Cao, C., E. Ghysels and F. Hatheway. 2000. Price discovery without trading: evidence from the Nasdaq pre-opening. *Journal of Finance* 55, 1339–1365.
- Chen, Z., Y. Huang, Y. Kusunadi, and K.J. Wei. 2017. The real effect of the initial enforcement of insider trading laws. *Journal of Corporate Finance*, 45, 687–709.
- Chordia, T., R. Roll, and A. Subrahmanyam. 2000. Commonality in liquidity. *Journal of Financial Economics* 56, 3–28.
- Cliff, M., M. J. Cooper, and H. Gulen. 2008. Return difference between trading and non-trading hours: like night and day. Working paper.
- deHaan, E., T. Shevlin, and J. Thornock. 2015. Market (in)attention and the strategic scheduling and timing of earnings announcements, *Journal of Accounting and Economics* 60, 36–55.
- Demsetz, H., 1968. The cost of transacting. *Quarterly Journal of Economics* 82, 33–53.
- Drienko, J., T. Smith, A. Von Reibnitz. 2017. A review of the return-illiquidity relationship. *Critical Finance Review*, forthcoming.
- Gao, X. and J.R. Ritter. 2010. The marketing of seasoned equity offerings. *Journal of Financial Economics* 97, 33–52.
- Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka. 2009. Do liquidity measures measure liquidity? *Journal of Financial Economics* 92, 153–81.
- Harris, L. and A. Amato. 2019. Illiquidity and Stock Returns: Cross-Section and Time-

- Series Effects: A Replication. *Critical Finance Review* 8, 173–202.
- Hasbrouck J. 2009. Trading costs and returns for U.S. equities: Estimating effective costs from Daily Data. *Journal of Finance* 64, 1445–77.
- Hendershott T., D. Livdan, and D. Rösch. 2020. Asset Pricing: A Tale of Night and Day. *Journal of Financial Economics*, forthcoming.
- Hung, M., X. Li, and S. Wang. 2014. Post-earnings-announcement Drift in Global Markets: Evidence from an Information Shock. *Review of Financial Studies*, 28(4), 1242–1283.
- Jones, C., G. Kaul, and M.L. Lipson. 1994. Information, Trading, and Volatility. *Journal of Financial Economics*, 36(1), 127–154.
- Kahle, K.M. and R.M. Stulz. 2017. Is the U.S. Public Corporation in Trouble?. *Journal of Economic Perspectives*, 31(3), 67–88.
- Kingsley Y.L.F., C.W. Holden and C.A. Trzcinka. 2017. What Are the Best Liquidity Proxies for Global Research?. *Review of Finance*, 21(4), 1355–1401.
- Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315–1335.
- Lang, M. and L. Stice-Lawrence. 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, 60(2-3), 110-135.
- Lee, K. H. 2011. The world price of liquidity risk. *Journal of Financial Economics*, 99(1), 136–161.
- Lipson M. L. and S. Mortal. 2009. Liquidity and capital structure. *Journal of Financial Markets* 12, 611–644.
- Lou, X. and T. Shu. 2017. Price impact or trading volume: Why is the Amihud (2002) measure priced? *Review of Financial Studies* 30, 4481–4520.
- Petersen, M.A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22, 435–480.
- Sadka, R. 2006. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics* 80, 309–349.
- Santosh S. 2016. The speed of price discovery: trade time vs clock time. Working paper.
- Stoll, H. and R. Whaley. 1990. Stock market structure and volatility. *Review of Financial Studies* 3, 37–71.

## Tables and Figures

Table 1: **Stock characteristics by levels of  $ROC$ .** This table presents medians of stock characteristics by year-specific quintiles of  $ROC = OCAM/CCAM$ . Medians of stock characteristics, based on observations from the previous year, are calculated by these quintiles.  $\beta^{mkt}$  is market beta,  $M$  is market capitalization in millions of dollars,  $DYD$  is dividend yield (%),  $SDRET$  is daily return volatility (%), and  $PRC$  is the end-of-year closing price.

|               | Quintile of $ROC$ |       |       |       |       |
|---------------|-------------------|-------|-------|-------|-------|
|               | 1                 | 2     | 3     | 4     | 5     |
| $ROC$         | 0.36              | 0.65  | 0.82  | 0.88  | 0.94  |
| $\beta^{mkt}$ | 0.94              | 1.03  | 1.05  | 1.05  | 1.04  |
| $M$           | 28.4              | 97.2  | 238.1 | 459.5 | 431.4 |
| $DYD$         | 1.40              | 1.31  | 1.48  | 1.74  | 1.74  |
| $SDRET$       | 2.35              | 2.38  | 2.34  | 2.25  | 2.32  |
| $PRC$         | 11.38             | 17.50 | 21.50 | 24.43 | 20.75 |

Table 2: **Association between *OCAM*-to-*CCAM* ratio and stock characteristics.** This table presents panel regression estimates when the ratio of open-to-close and close-to-close Amihud (2002) measures,  $ROC_{iy}$ , is regressed on stock characteristics, the natural logs of market-capitalization,  $\ln(M_{i,y-1})$ , and mean daily turnover,  $\ln(TR_{i,y-1})$ ; dividend yield,  $DYD_{i,y-1}$ ; daily return volatility,  $SDRET_{i,y-1}$ ; and share price (scaled by 1/100),  $PRC_{i,y-1}$ , as in equation (6), in the 1964–2019 sample. Specifications differ in the set of fixed effects introduced. Numbers in parentheses reflect standard errors of estimates, clustered at both stock and year levels. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|               | 1964–2019           |                     |                     |                     | 1964–1980           | 1981–2000            | 2001–2019           |
|---------------|---------------------|---------------------|---------------------|---------------------|---------------------|----------------------|---------------------|
| $\beta^{mkt}$ | 0.095***<br>(0.033) | 0.089***<br>(0.021) | 0.070**<br>(0.034)  | 0.077***<br>(0.021) | 0.235***<br>(0.038) | 0.024<br>(0.054)     | 0.108***<br>(0.019) |
| $\ln(M)$      | 0.061***<br>(0.004) | 0.040***<br>(0.003) | 0.069***<br>(0.006) | 0.047***<br>(0.005) | 0.068***<br>(0.007) | 0.057***<br>(0.007)  | 0.001<br>(0.003)    |
| $\ln(TR)$     | 0.071***<br>(0.005) | 0.073***<br>(0.006) | 0.080***<br>(0.006) | 0.077***<br>(0.005) | 0.111***<br>(0.006) | 0.098***<br>(0.003)  | 0.024***<br>(0.005) |
| $DYD$         | 0.006<br>(0.005)    | −0.004<br>(0.004)   | 0.011**<br>(0.005)  | 0.003<br>(0.003)    | 0.018<br>(0.011)    | −0.005<br>(0.006)    | 0.000<br>(0.003)    |
| $SDRET$       | 0.072***<br>(0.016) | 0.022<br>(0.014)    | 0.097***<br>(0.017) | 0.044***<br>(0.012) | 0.022<br>(0.021)    | 0.017<br>(0.016)     | 0.013<br>(0.014)    |
| $PRC$         | −0.019<br>(0.013)   | −0.007<br>(0.009)   | −0.025<br>(0.017)   | −0.015<br>(0.014)   | −0.097*<br>(0.052)  | −0.088***<br>(0.030) | 0.002**<br>(0.001)  |
| Stock FEs     | No                  | Yes                 | No                  | Yes                 | Yes                 | Yes                  | Yes                 |
| Year FEs      | No                  | No                  | Yes                 | Yes                 | Yes                 | Yes                  | Yes                 |
| $R^2$         | 0.33                | 0.51                | 0.35                | 0.52                | 0.38                | 0.70                 | 0.76                |
|               |                     | 69,474              |                     |                     | 24,424              | 26,935               | 18,115              |

Table 3: **Rank correlation statistics between *OCAM* and *CCAM*.** This table presents Kendall’s  $\tau$  statistics across *CCAM* and *OCAM* over time periods, 1964–2019, 1964–1980, 1981–2000, and 2001–2019. Each year, stocks are sorted into top 30%, middle 40%, and bottom 30% liquidity according to *CCAM*. Kendall’s  $\tau$  statistic is calculated every year within each liquidity group, and then averaged across different years in the entire sample period or in a sub-period.

| Liquidity group   | 1964–2019 | 1964–1980 | 1981–2000 | 2001–2019 |
|-------------------|-----------|-----------|-----------|-----------|
| Top 30% liquid    | 94.9%     | 91.5%     | 96.2%     | 96.5%     |
| Middle 40% liquid | 84.4%     | 69.7%     | 85.3%     | 96.7%     |
| Bottom 30% liquid | 72.4%     | 63.0%     | 64.4%     | 89.1%     |

Table 4: **Fama-MacBeth estimates of monthly returns on stock characteristics, NYSE- and AMEX-listed common shares, 1964–2019.** This table presents Fama-MacBeth estimates of Equation (7) for NYSE and AMEX-listed common shares in time periods 1964–2017, 1964–1980, 1981–2000, and 2011–2019. The last two columns present estimates after removing the largest 400 stocks in the 2011–2019 period. The dependent variable is the monthly stock return in percentage points. *CCAM* is the traditional Amihud (2002) liquidity measure, and *OCAM* is the Amihud (2002) measure after removing overnight price movements. We divide each *CCAM*<sub>*iy*</sub> and *OCAM*<sub>*iy*</sub> observation by its respective sample mean across stocks in year *y*, thereby centering each measure to have a mean of one, making coefficients across the two measures comparable.  $\beta^{mkt}$  is market beta estimated across ten size portfolios using daily observations from the last calendar year. *R100* is the compound return on a stock in the last 100 days of the previous calendar year, and *R100YR* is the compound return over the remaining trading days in the last calendar year.  $\ln(M)$  is the natural log of market capitalization at the end of the previous calendar year. *SDRET* is the standard deviation of daily returns over the previous calendar year. *DYD* is the ratio of total cash dividend distribution over the previous calendar year to the closing price at the end of that year, or dividend yield. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|               | 1964–2019            |                      | 1964–1980            |                      | 1981–2000           |                     | 2001–2019           |                     | 2001–2019<br>mid-/small-cap |                    |
|---------------|----------------------|----------------------|----------------------|----------------------|---------------------|---------------------|---------------------|---------------------|-----------------------------|--------------------|
| <i>CCAM</i>   | 0.064***<br>(0.016)  |                      | 0.105***<br>(0.041)  |                      | 0.068***<br>(0.022) |                     | 0.022<br>(0.018)    |                     | 0.028<br>(0.017)            |                    |
| <i>OCAM</i>   |                      | 0.126***<br>(0.025)  |                      | 0.205***<br>(0.065)  |                     | 0.137***<br>(0.034) |                     | 0.043*<br>(0.024)   |                             | 0.053**<br>(0.025) |
| $\beta^{mkt}$ | 0.412<br>(0.289)     | 0.656**<br>(0.286)   | 0.466<br>(0.482)     | 0.796<br>(0.511)     | 0.134<br>(0.618)    | 0.387<br>(0.592)    | 0.658*<br>(0.348)   | 0.813**<br>(0.353)  | 0.441<br>(0.584)            | 0.430<br>(0.593)   |
| <i>R100</i>   | 0.519**<br>(0.209)   | 0.505**<br>(0.207)   | 0.668**<br>(0.329)   | 0.676**<br>(0.329)   | 1.015***<br>(0.259) | 0.997***<br>(0.256) | −0.138<br>(0.454)   | −0.167<br>(0.449)   | −0.066<br>(0.506)           | −0.154<br>(0.493)  |
| <i>R100YR</i> | 0.226***<br>(0.082)  | 0.226***<br>(0.083)  | 0.363**<br>(0.149)   | 0.374**<br>(0.152)   | 0.409***<br>(0.128) | 0.410***<br>(0.127) | −0.090<br>(0.147)   | −0.100<br>(0.147)   | −0.246<br>(0.192)           | −0.274<br>(0.191)  |
| $\ln(M)$      | −0.095***<br>(0.029) | −0.081***<br>(0.028) | −0.165***<br>(0.058) | −0.148***<br>(0.054) | −0.038<br>(0.049)   | −0.024<br>(0.047)   | −0.092**<br>(0.040) | −0.080**<br>(0.041) | −0.072<br>(0.096)           | −0.011<br>(0.105)  |
| <i>SDRET</i>  | −1.54***<br>(0.326)  | −1.76***<br>(0.336)  | −0.68<br>(0.615)     | −1.04<br>(0.643)     | −2.42***<br>(0.476) | −2.68***<br>(0.498) | −1.38**<br>(0.601)  | −1.43**<br>(0.607)  | −1.24**<br>(0.583)          | −1.29**<br>(0.590) |
| <i>DYD</i>    | 0.268<br>(0.718)     | 0.175<br>(0.701)     | 1.343<br>(2.254)     | 1.038<br>(2.195)     | 0.367<br>(0.346)    | 0.330<br>(0.348)    | −0.794<br>(0.545)   | −0.757<br>(0.548)   | −0.663<br>(0.808)           | −0.549<br>(0.822)  |

Table 5: **Fama-MacBeth estimates of liquidity premia using standardized liquidity measures, 1964–2019.** This table presents Fama-MacBeth estimates of liquidity premia using Equation (7) for NYSE and AMEX-listed common shares in time periods 1964–2017, 1964–1980, 1981–2000, and 2011–2019. The last two columns present estimates after removing the largest 400 stocks in the 2011–2019 period. The dependent variable is the monthly stock return in percentage points.  $CCAM^{std}$  is the standardized traditional Amihud (2002) liquidity measure, and  $OCAM^{std}$  is the standardized Amihud (2002) measure after removing overnight price movements. The set of controls is identical to that in Table 4. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|              | 1964–2019           | 1964–1980           | 1981–2000           | 2001–2019         | 2001–2019<br>mid-/small-cap |
|--------------|---------------------|---------------------|---------------------|-------------------|-----------------------------|
| $CCAM^{std}$ | 0.137***<br>(0.035) | 0.180**<br>(0.072)  | 0.157***<br>(0.052) | 0.078<br>(0.060)  | 0.137<br>(0.084)            |
| $OCAM^{std}$ | 0.218***<br>(0.044) | 0.288***<br>(0.092) | 0.239***<br>(0.061) | 0.135*<br>(0.074) | 0.237**<br>(0.109)          |

Table 6: **Fama-MacBeth estimates of liquidity premia, 1964–2019: adjusted for earnings announcements.** This table presents Fama-MacBeth estimates of Equation (7) for NYSE and AMEX-listed common shares in time periods 1964–2017, 1964–1980, 1981–2000, and 2011–2019. The last two columns present estimates after removing the largest 400 stocks in the 2011–2019 period. The dependent variable is the monthly stock return in percentage points.  $CCAM$  is the traditional Amihud (2002) liquidity measure, and  $OCAM$  is the Amihud (2002) measure after removing overnight price movements. Both measures are constructed after removing 3-day windows around earnings announcements. We divide each  $CCAM_{iy}$  and  $OCAM_{iy}$  observation by its respective sample mean across stocks in year  $y$ , thereby centering each measure to have a mean of one, making coefficients across the two measures comparable. The set of controls is identical to that in Table 4. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|             | 1964–2019           | 1964–1980           | 1981–2000           | 2001–2019         | 2001–2019<br>mid-/small-cap |
|-------------|---------------------|---------------------|---------------------|-------------------|-----------------------------|
| $CCAM^{ea}$ | 0.066***<br>(0.016) | 0.104**<br>(0.041)  | 0.073***<br>(0.022) | 0.025<br>(0.017)  | 0.032*<br>(0.017)           |
| $OCAM^{ea}$ | 0.126***<br>(0.025) | 0.202***<br>(0.065) | 0.139***<br>(0.034) | 0.045*<br>(0.023) | 0.054**<br>(0.024)          |

Table 7: **CCAM and OCAM vs. standard measures of trading costs.** This table presents average cross-sectional correlations between each *CCAM* and *OCAM* vis à vis other high-frequency measures of stock liquidity. **Panel A** presents correlations against annually-constructed high-frequency measures of trading costs.  $EFSP_0$  and  $EFSP_1$  are averages of percentage effective spreads constructed with respect to mid-point price at, respectively, the same and the previous second of the corresponding transaction.  $QSP$  is the time-weighted average of percentage bid-ask spread, constructed annually.  $\lambda_0$  and  $\lambda_1$  are annual averages of estimates of Kyle's  $\lambda$ , estimated using five-minute observations at the daily level, with and without a no-intercept restriction, respectively. Every year, correlations between each measure with *CCAM* or *OCAM* are calculated in the 1993–2013 period, averages across years and differences are reported. **Panel B** presents correlations against annual measures of effective costs calculated by Hasbrouck (2009).  $cMdmLog$  and  $cMdmLogz$  are two version of Roll's measure of daily return auto-correlations, respectively, reflecting whether missing daily return observations are dropped or replace by zero.  $cLogMean$  reflects Gibbs estimates using a market-factor model applied to CRSP closing prices and dividends. Every year, correlations between each measure with *CCAM* or *OCAM* are calculated in the 1964–2003 period, averages across years and differences are reported. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

| <i>Panel A</i> : High-freq. measures: 1993–2013 |                  |                  |                         | <i>Panel B</i> : Effective costs measures: 1964–2003 |                  |                  |                          |
|---|------------------|------------------|-------------------------|--|------------------|------------------|--------------------------|
|   | <i>OCAM</i>      | <i>CCAM</i>      | Difference <sup>†</sup> |  | <i>OCAM</i>      | <i>CCAM</i>      | Difference <sup>††</sup> |
| $EFSP_0$  | 0.827<br>(0.009) | 0.751<br>(0.014) | 0.076***<br>(0.009)     | $cMdmLog$  | 0.546<br>(0.020) | 0.399<br>(0.017) | 0.148***<br>(0.007)      |
| $EFSP_1$  | 0.829<br>(0.007) | 0.752<br>(0.011) | 0.077***<br>(0.008)     | $cMdmLogz$   | 0.505<br>(0.023) | 0.390<br>(0.019) | 0.115***<br>(0.006)      |
| $QSP$   | 0.845<br>(0.006) | 0.773<br>(0.010) | 0.072***<br>(0.010)     | $cLogMean$   | 0.732<br>(0.017) | 0.615<br>(0.014) | 0.116***<br>(0.006)      |
| $\lambda_0$                                     | 0.816<br>(0.030) | 0.757<br>(0.028) | 0.059***<br>(0.009)     | <sup>††</sup> Degrees of freedom is 40.              |                  |                  |                          |
| $\lambda_1$                                     | 0.778<br>(0.040) | 0.722<br>(0.037) | 0.056***<br>(0.009)     |  |                  |                  |                          |

<sup>†</sup>Degrees of freedom is 20.

Table 8: **Fama-MacBeth estimates for pricings of orthogonally-decomposed measures, NYSE-listed common shares, 1964–2019.** This table presents Fama-MacBeth estimates of Equation (7) for NYSE- and AMEX-listed common shares in time periods 1964–2019, 1964–1980, 1981–2000, and 2011–2019. The last two columns present estimates after removing the largest 400 stocks in the 2011–2019 period. The dependent variable is monthly stock return in percentage points. Independent variables include those in Table 4 plus the linearly orthogonal component of the alternative liquidity measure with respect to the baseline measure. When *CCAM* is the baseline,  $\tilde{Z}$  is the additional independent variable; when *OCAM* is the baseline  $Z$  is added.  $Z$  and  $\tilde{Z}$  reflect the residuals from Equations (8) and (9), respectively. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in the parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|             | 1964–2019           | 1964–1980           | 1981–2000           | 2001–2019         | 2001–2019<br>mid-/small-cap |
|-------------|---------------------|---------------------|---------------------|-------------------|-----------------------------|
| <i>CCAM</i> | 0.085***<br>(0.018) | 0.144***<br>(0.048) | 0.088***<br>(0.024) | 0.031<br>(0.019)  | 0.037*<br>(0.019)           |
| $\tilde{Z}$ | 0.147***<br>(0.034) | 0.255***<br>(0.081) | 0.130***<br>(0.050) | 0.068*<br>(0.041) | 0.067*<br>(0.041)           |
| <i>OCAM</i> | 0.126***<br>(0.025) | 0.209***<br>(0.065) | 0.135***<br>(0.034) | 0.043*<br>(0.024) | 0.052**<br>(0.025)          |
| $Z$         | −0.018<br>(0.019)   | −0.040<br>(0.039)   | 0.006<br>(0.032)    | −0.023<br>(0.030) | −0.014<br>(0.029)           |



Table 9: **Pricing of the components of *CCAM* and *OCAM*: 1964–2019.** The table presents Fama-MacBeth cross-sectional estimates of the pricing of the different components of *CCAM* and *OCAM*. Equation (1) in Amihud and Noh (2020) is estimated using data for NYSE- and AMEX-listed stocks in the 1964–2019 period, producing analogous results to those in Table 1 from Amihud and Noh (2020). The dependent variable is monthly stock return in excess of monthly T-Bill rate. Independent variables include various components of *CCAM* or *OCAM* (described in Section 6; natural log of market-capitalization from the end of two months earlier,  $\ln(M_{i,t-2})$ ; the most recent book-to-market value,  $BM_{i,t-1}$ ; previous month’s return,  $R_{i,t-1}$ ; and the compound return from the 11 months before that,  $R_{i,t-2}^{t-12}$ . Stock-month observations are included only if their average closing price over the preceding 12 months falls between \$5 and \$1,000. The remaining filters are identical to those from the earlier analyses. To closely follow Amihud and Noh (2020), Newey-West standard errors using six lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

| <i>Panel A: Pricing of CCAM components</i> |                    |                     | <i>Panel A: Pricing of OCAM components</i> |                                |                   |                    |                     |
|--|--------------------|---------------------|--|--------------------------------|-------------------|--------------------|---------------------|
|  | (1)                | (2)                 | (3)  |                                | (1)               | (2)                | (3)                 |
| $\ln(CCAM)$                                | 0.095**<br>(0.048) |                     |  | $\ln(OCAM)$                    | 0.082*<br>(0.047) |                    |                     |
| $\ln(LS^{CCAM})$                           |                    | 0.109**<br>(0.048)  |  | $\ln(LS^{OCAM})$               |                   | 0.107**<br>(0.045) |                     |
| $\ln( \overline{R} )$                      |                    |                     | −0.390**<br>(0.191)                        | $\ln( \overline{OCR} )$        |                   |                    | −0.385**<br>(0.159) |
| $\ln(IDVOL)$                               |                    |                     | 0.079**<br>(0.039)                         | $\ln(IDVOL)$                   |                   |                    | 0.106**<br>(0.046)  |
| $DIF^{CCAM}$                               |                    | 0.714***<br>(0.234) | 0.530**<br>(0.234)                         | $DIF^{OCAM}$                   |                   | 0.231*<br>(0.139)  | 0.466***<br>(0.142) |
| $\Delta Y/\Delta(DIF)^\dagger$             |                    |                     | <b>0.083</b>                               | $\Delta Y/\Delta(DIF)^\dagger$ |                   |                    | <b>0.159</b>        |

<sup>†</sup>Measures the change in excess returns in response to one standard deviation increase in *DIF*.

Table 10: **Association between *OCAM*-to-*CCAM* ratio and non-trading before close.** This table presents panel regression estimates when  $ROC_{iy}$  is regressed on  $HCT$  using alternative specifications. Specifications differ in (i) the sets of fixed effects introduced; (ii) controlling for stock characteristics including the natural logs of market-capitalization,  $\ln(M_{i,y-1})$ , and mean daily turnover,  $\ln(TR_{i,y-1})$ ; dividend yield,  $DYD_{i,y-1}$ ; daily return volatility,  $SDRET_{i,y-1}$ ; and share price,  $PRC_{i,y-1}$ ; (iii) controlling for  $CCAM_{i,y-1}$  and  $OCAM_{i,y-1}$ ; and (iv) using lagged non-trading before close,  $HTC_{i,y-1}$ . Numbers in parentheses reflect standard errors of estimates, clustered at both stock and year levels. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

| Dep. Var. = $ROC$ |                      |                      |                      |                      |                      |                      |                      |                      |
|-------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| $HTC$             | -0.274***<br>(0.005) | -0.259***<br>(0.007) | -0.283***<br>(0.004) | -0.286***<br>(0.006) | -0.305***<br>(0.007) | -0.284***<br>(0.007) | -0.305***<br>(0.007) |                      |
| $HTC_{y-1}$       |                      |                      |                      |                      |                      |                      |                      | -0.187***<br>(0.009) |
| Stock FE          | No                   | Yes                  | No                   | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  |
| Year FE           | No                   | No                   | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  | Yes                  |
| Characteristics   | No                   | No                   | No                   | No                   | Yes                  | Yes                  | Yes                  | Yes                  |
| $CCAM_{y-1}$      | No                   | No                   | No                   | No                   | No                   | Yes                  | No                   | No                   |
| $OCAM_{y-1}$      | No                   | No                   | No                   | No                   | No                   | No                   | Yes                  | No                   |
| Adj- $R^2$        | 0.77                 | 0.83                 | 0.80                 | 0.85                 | 0.85                 | 0.85                 | 0.85                 | 0.76                 |
| Observations      |                      |                      |                      | 24,855               |                      |                      |                      | 21,152               |

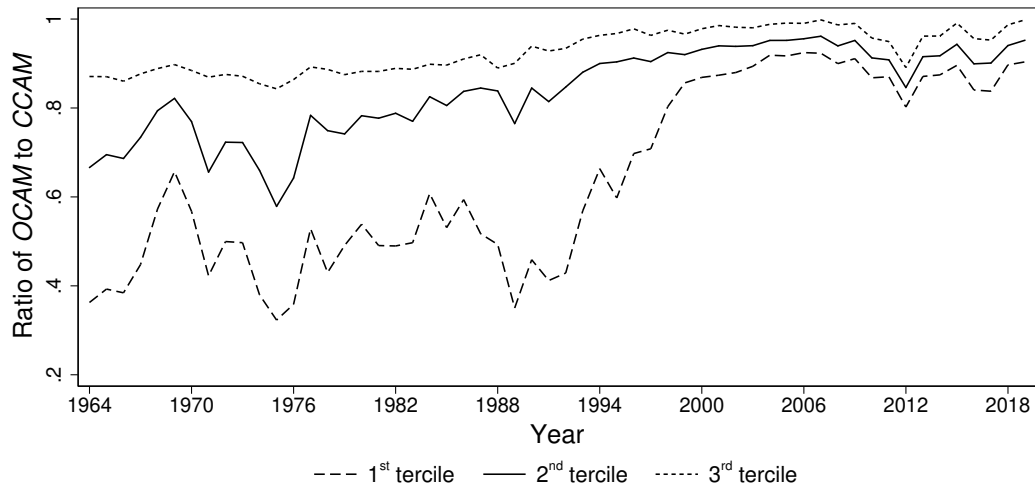
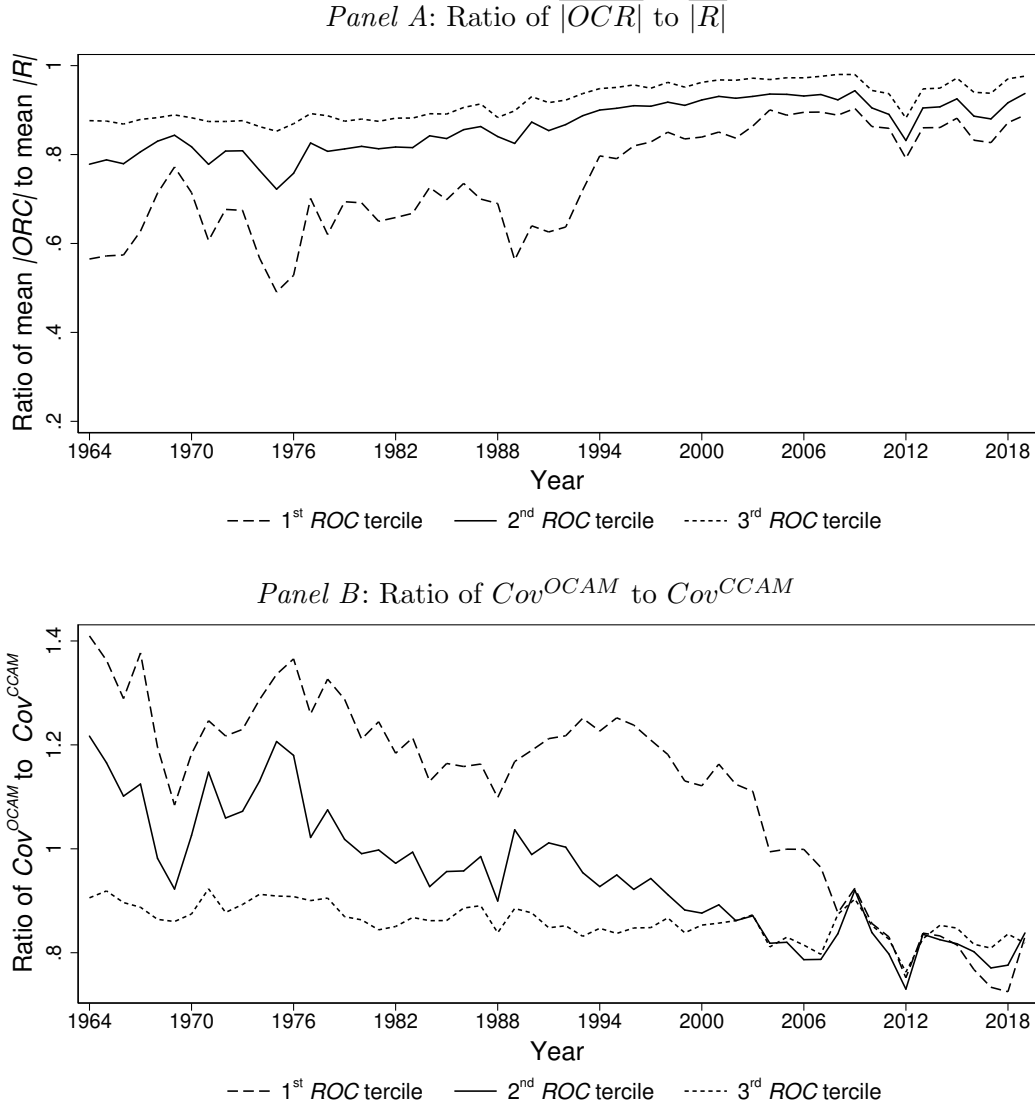


Figure 3: **Evolution of the ratio of *OCAM* to *CCAM* (700 largest firms).** The figure plots the temporal changes in the cross-stock distribution of the *OCAM*-to-*CCAM* among the 700 NYSE- and AMEX-listed common shares with largest market-capitalizations in the 1964-2019 period. Each year, stocks are sorted by market-capitalization to obtain the 700 firms featuring the largest market values. For each stock  $i$  in year  $y$ , the ratio  $ROC_{iy} = OCAM_{i,y-1}/CCAM_{i,y-1}$  is calculated. Stocks are sorted into terciles of  $ROC_{iy}$  each year, and the year-specific medians in respective terciles of this ratio are plotted against time



**Figure 4: Evolution of the ratio of OCAM to CCAM.** The figure plots the temporal changes in the cross-stock distribution of the OCAM-to-CCAM for NYSE- and AMEX-listed common shares in the 1964-2019 period. For each stock  $i$  in year  $y$ , the ratios  $ROC_{iy} = OCAM_{i,y-1}/CCAM_{i,y-1}$ ,  $ROC_{iy}^{Cov} = Cov_{i,y-1}^{OCAM}/Cov_{i,y-1}^{CCAM}$ , and  $ROC_{i,y}^{|ret|} = \overline{|OCR|}_{i,y-1}/\overline{|R|}_{i,y-1}$  are calculated. Stocks are sorted into tertiles of  $ROC_{iy}$  each year, and the year-specific medians of  $ROC^{|ret|}$  and  $ROC^{Cov}$  in the respective tertiles of  $ROC$  are plotted against time

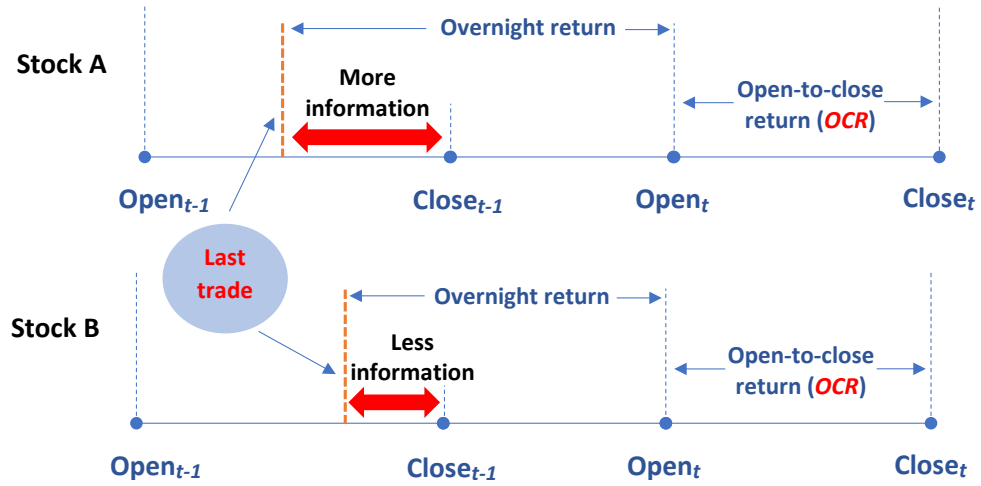


Figure 5: **Illustration of non-synchronous trading and the extent of overnight information accumulation.** The figure illustrates the relationship between the extent of non-trading before close on date  $t - 1$  and the amount of information contained in overnight returns.

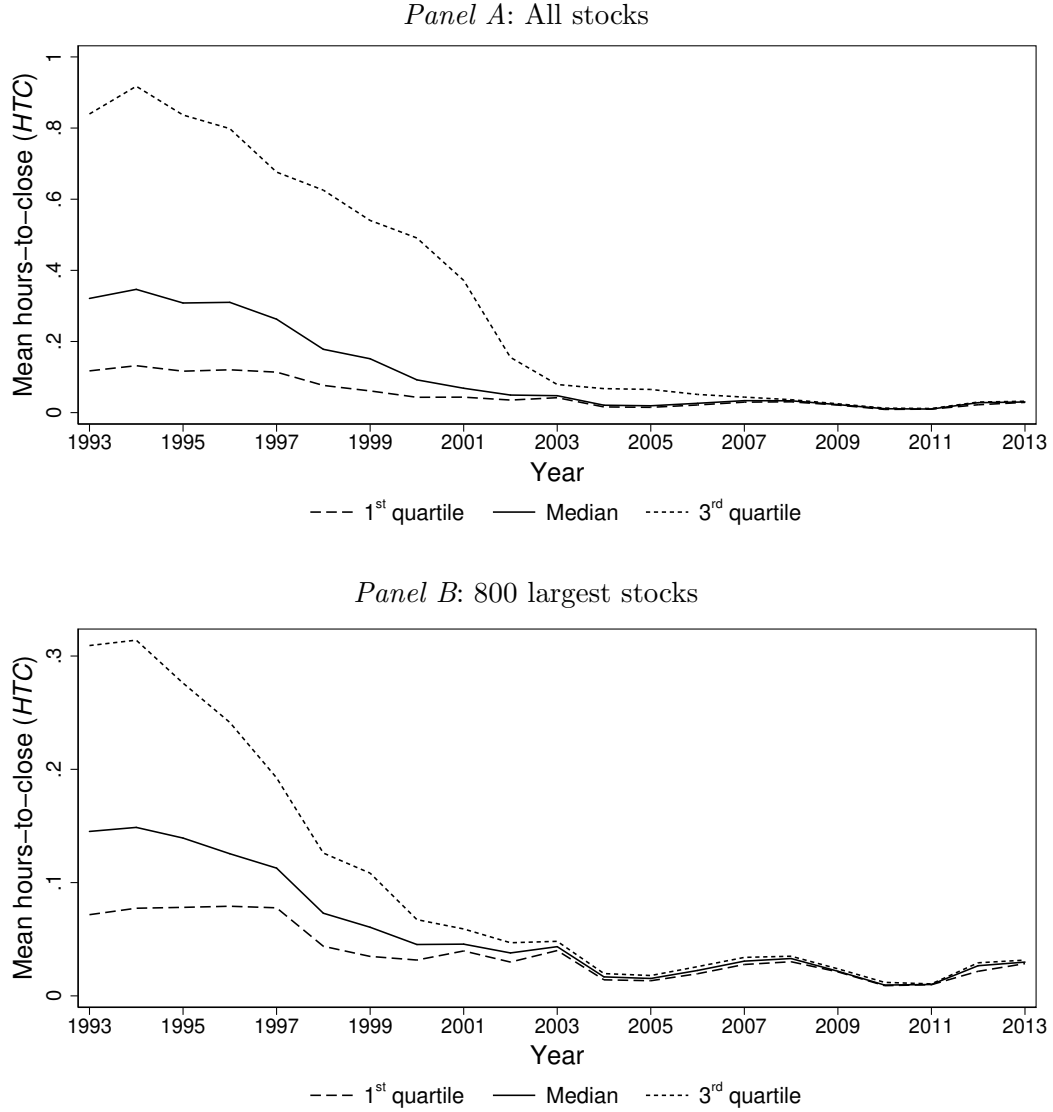


Figure 6: **Evolution of the the extent of non-trading.** This figure plots temporal changes in the cross-stock distribution of  $HTC$  in the 1993-2013 period. For each stock  $i$  in year  $y$ ,  $HTC_{iy}$  measures the average time distance, in hours, of the last transaction and close per trading day. The year-specific quartiles of  $HTC$  are plotted against time. Panel A presents statistics in the sample of all stocks, and Panel B presents the patterns for the sample of largest 700 stocks based on market-capitalizations at the end of the preceding year.

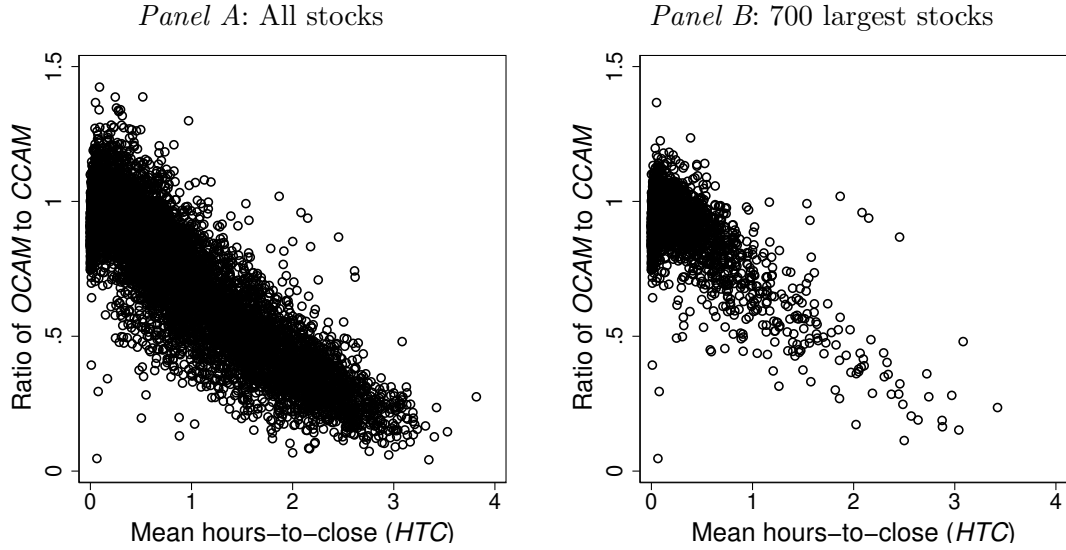


Figure 7: **Association between  $ROC$  and  $HTC$ .** The figure presents a scatter plot of  $ROC_{iy}$  against  $HTC_{iy}$  in the 1993-2013 period. For each stock  $i$  in year  $y$ ,  $ROC_{iy}$  is the ratio of open-to-close Amihud measure,  $OCAM$ , to the traditional Amihud measure,  $CCAM$ ; and  $HTC_{iy}$  measures the average time distance, in hours, of the last transaction and close (4:00pm EST) per trading day.

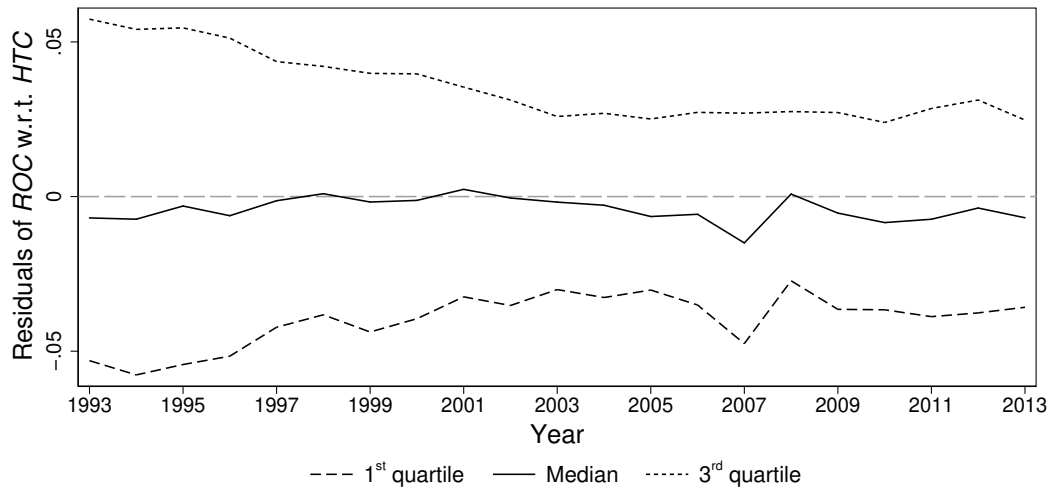


Figure 8: **Evolution of the  $OCAM$ -to- $CCAM$  ratio after controlling for non-trading.** The figure plots the temporal changes in the cross-stock distribution of the residuals from annual cross-sections of  $ROC_{iy}$  on  $HTC_{iy}$  in the 1993-2013 period. For each stock  $i$  in year  $y$ ,  $ROC_{iy}$  is the ratio of open-to-close Amihud measure,  $OCAM$ , to the traditional Amihud measure,  $CCAM$ ; and  $HTC_{iy}$  measures the average time distance, in hours, of the last transaction and close (4:00pm EST) per trading day. The year-specific quartiles of residuals are plotted against time.

## A Appendix

### A.1 Overnight trading volume vs. overnight price movement

In this section, we establish that even though overnight trading volume trade represents a very tiny share of total trading volume, overnight price movements comprise a large share of close-to-close price movements. Using Monthly TAQ data from the 1993–2013 period, we construct measures of 24-hour trading volume for all NYSE-listed common stocks. For each stock, we calculate the share of trading volume realized in windows open–close, close–4:15pm, and 4:15pm–open next day out of the 24-hour volume annually. We also collect open and close prices (both from CRSP), and the last transaction price before or at 4:15pm (from TAQ). We then calculate absolute returns over each window, adjusting for any overnight price adjustment or dividend distribution. For each stock, we calculate average absolute return in each window, and divide it by the corresponding average absolute close-to-close return.

Figure A.1 shows that between 4:15pm and the next open, the ratio of the mean absolute return to mean absolute close-to-close return exceeds 0.45 for a typical stock, even though the associated volume is less than 1.5% of total trading volume. The analogous ratio for the mean absolute open-to-close return is 0.91, and 95% of trading volume is realized during regular hours.

### A.2 Liquidity premia 1993-2019: CRSP data only

This section verifies the robustness of findings to the source of open prices. We focus on the time period when open prices are available by CRSP, so that *OCAM* can be constructed using CRSP or GFD. We also extend the analysis to NASDAQ-listed common shares for the 1993–2019 period, establishing qualitatively identical findings for NASDAQ-listed firms.

Table A.1 presents the Fama-MacBeth regression analogues to Table 4, reinforcing our earlier findings. Once more, liquidity premia based on *OCAM* are roughly double those based



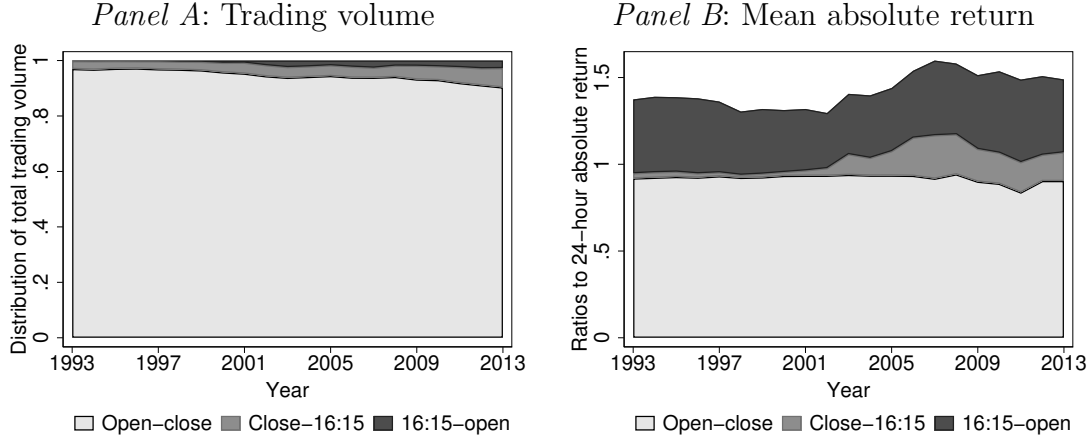


Figure A.1: **Empirical distributions of trading volume and average absolute return across 24 hours.** Panel A shows the composition of NYSE-listed stocks' annual trading volumes for time windows of open–close, close–4:15pm, and 4:15pm–open. Trading volumes are calculated using Monthly TAQ data, and average fractions of annual trading volumes in each window out of the 24-hour volume are plotted over time. Panel B plots the ratio of absolute returns over a window to absolute close–close returns. Using open price, close price, and the last transaction price by 4:15pm from each stock-day, absolute returns for each window are calculated (adjusted for overnight price adjustments and dividend distributions). The average absolute return in a window is calculated for each stock annually, and then divided by the corresponding average absolute close–close return. For each stock-year, cumulative ratios going from open–close to 4:15pm–open of the next day are constructed. Jensen's inequality implies that the sum of the three ratios must exceed one. The cross-stock medians of the resulting cumulative ratios are plotted over time.

on *CCAM* in the 1993–2013 period. More importantly, this finding does not vary with data source (although about 9% of the sample stocks differ by source). For NYSE-/AMEX-listed firms, we fit equation (7) using three constructions of *OCAM*: (i) GFD only; (ii) CRSP only; (iii) CRSP only when observations can be matched with GFD-only sample. In all three cases *OCAM*-based liquidity premia are over twice *CCAM*-based counterparts. Moreover, premia based on *OCAM* remain statistically significant in WLS fits while *CCAM*'s coefficients are no longer significant. Additionally, using CRSP data, we establish that liquidity premia based on *OCAM* are over twice those based on *CCAM* for NASDAQ-listed stocks.

Table A.1: **Fama-MacBeth estimates of monthly returns on stock characteristics, NYSE- and AMEX-listed versus NASDAQ-listed common shares, 1993–2019 (CRSP data only).** This table presents Fama-MacBeth estimates of Equation (7) for NYSE- and AMEX-listed common shares in the 1993–2019, contrasting findings when *OCAM* is constructed using GFD data to when it is constructed based CRSP. The data sources used to construct *OCAM* include: (i) GFD only; (ii) CRSP only; (iii) CRSP only when observations can be matched with GFD-only sample. Table also presents estimates of Equation (7) for NASDAQ-listed stocks using CRSP. Variable construction is identical to that in Table 4. Estimates are carried out using both ordinary least squares and weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

| Method        | Ordinary Least Squares |         |           |          |                 |          |           |          | Weighted Least Squares |          |           |          |                 |          |           |          |
|---------------|------------------------|---------|-----------|----------|-----------------|----------|-----------|----------|------------------------|----------|-----------|----------|-----------------|----------|-----------|----------|
| Exchange      | NYSE/AMEX              |         |           |          |                 |          | NASDAQ    |          | NYSE/AMEX              |          |           |          |                 |          | NASDAQ    |          |
| Data          | GFD only               |         | CRSP only |          | CRSP: GFD-match |          | CRSP only |          | GFD only               |          | CRSP only |          | CRSP: GFD-match |          | CRSP only |          |
| <i>CCAM</i>   | 0.026*                 | (0.014) | 0.022*    | (0.013)  | 0.025*          | (0.014)  | 0.074*    | (0.031)  | 0.021                  | (0.014)  | 0.017     | (0.013)  | 0.020           | (0.014)  | 0.061*    | (0.031)  |
| <i>OCAM</i>   |                        | 0.049** |           | 0.044**  |                 | 0.044**  |           | 0.157*** |                        | 0.038*   |           | 0.034*   |                 | 0.034*   |           | 0.135*** |
|               |                        | (0.019) |           | (0.018)  |                 | (0.019)  |           | (0.042)  |                        | (0.019)  |           | (0.018)  |                 | (0.019)  |           | (0.041)  |
| $\beta^{mkt}$ | 0.667                  | 0.801*  | 0.555     | 0.712*   | 0.667           | 0.791*   | −0.116    | −0.012   | 0.678                  | 0.776*   | 0.576     | 0.700*   | 0.681*          | 0.775*   | −0.206    | −0.107   |
|               | (0.421)                | (0.408) | (0.407)   | (0.391)  | (0.420)         | (0.406)  | (0.317)   | (0.304)  | (0.411)                | (0.403)  | (0.400)   | (0.389)  | (0.410)         | (0.400)  | (0.308)   | (0.297)  |
| <i>R100</i>   | −0.067                 | −0.082  | −0.041    | −0.049   | −0.071          | −0.081   | −0.113    | −0.120   | 0.002                  | −0.008   | 0.048     | 0.043    | 0.001           | −0.005   | −0.063    | −0.069   |
|               | (0.333)                | (0.331) | (0.324)   | (0.323)  | (0.333)         | (0.330)  | (0.239)   | (0.240)  | (0.331)                | (0.328)  | (0.320)   | (0.318)  | (0.330)         | (0.327)  | (0.230)   | (0.231)  |
| <i>R100YR</i> | 0.058                  | 0.054   | 0.041     | 0.039    | 0.058           | 0.056    | −0.005    | −0.003   | 0.076                  | 0.071    | 0.065     | 0.063    | 0.076           | 0.073    | 0.018     | 0.020    |
|               | (0.118)                | (0.118) | (0.113)   | (0.113)  | (0.119)         | (0.118)  | (0.057)   | (0.056)  | (0.119)                | (0.119)  | (0.113)   | (0.113)  | (0.119)         | (0.119)  | (0.057)   | (0.057)  |
| $\ln(M)$      | −0.058                 | −0.050  | −0.050    | −0.042   | −0.058          | −0.054   | 0.006     | 0.027    | −0.056                 | −0.049   | −0.048    | −0.041   | −0.057          | −0.053   | 0.019     | 0.035    |
|               | (0.037)                | (0.037) | (0.036)   | (0.036)  | (0.037)         | (0.037)  | (0.043)   | (0.042)  | (0.037)                | (0.037)  | (0.037)   | (0.037)  | (0.037)         | (0.037)  | (0.042)   | (0.042)  |
| <i>SDRET</i>  | −1.23**                | −1.30** | −1.20**   | −1.27*** | −1.24**         | −1.31*** | −1.13***  | −1.28*** | −1.26**                | −1.31*** | −1.20**   | −1.26*** | −1.27**         | −1.32*** | −1.25***  | −1.37*** |
|               | (0.494)                | (0.501) | (0.474)   | (0.481)  | (0.494)         | (0.502)  | (0.417)   | (0.432)  | (0.491)                | (0.498)  | (0.469)   | (0.476)  | (0.491)         | (0.499)  | (0.417)   | (0.433)  |
| <i>DYD</i>    | −0.223                 | −0.222  | −0.178    | −0.182   | −0.234          | −0.242   | −0.318    | −0.297   | −0.286                 | −0.283   | −0.221    | −0.222   | −0.298          | −0.302   | −0.469    | −0.457   |
|               | (0.332)                | (0.331) | (0.304)   | (0.307)  | (0.332)         | (0.330)  | (0.672)   | (0.677)  | (0.333)                | (0.332)  | (0.306)   | (0.308)  | (0.333)         | (0.331)  | (0.666)   | (0.670)  |
| Obs.          | 451,562                |         | 522,808   |          | 451,562         |          | 765,569   |          | 451,562                |          | 522,808   |          | 451,562         |          | 765,569   |          |

### A.3 Liquidity premia 1993-2019: The GFD-CRSP hybrid sample

In this section, we establish that our qualitative findings are unaffected if we pool GFD and CRSP data in order to maximize the cross-sections of available stocks post 1992. As in the analysis reported in Table 6, we construct *CCAM* and *OCAM* after taking out 3-day windows around earnings announcement dates. We then standardize the liquidity measures, reporting estimates that are the analogues to those in Table 5. Table A.2 shows that for this hybrid sample over the entire time period, liquidity premia based on standardized *OCAM* are 56% larger than those based on standardized *CCAM*.

Table A.2: **Fama-MacBeth estimates of liquidity premia using standardized liquidity measures, 1964–2019 GFD-CRSP hybrid sample.** This table presents Fama-MacBeth estimates of liquidity premia using Equation (7) for NYSE and AMEX-listed common shares in time periods 1964–2017, 1964–1980, 1981–2000, and 2011–2019. The last two columns present estimates after removing the largest 400 stocks in the 2011–2019 period. Data used to construct *OCAM* are obtained from GFD for 1964–1992 and from CRSP for 1993–2019. The dependent variable is the monthly stock return in percentage points. *CCAM<sup>std</sup>* is the standardized traditional Amihud (2002) liquidity measure, and *OCAM<sup>std</sup>* is the standardized Amihud (2002) measure after removing overnight price movements. 3-day windows around earnings announcements are excluded from the construction of the liquidity measures. The set of controls is identical to that in Table 4. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|                           | 1964–2019           | 1964–1980           | 1981–2000           | 2001–2019         | 2001–2019<br>mid-/small-cap |
|---------------------------|---------------------|---------------------|---------------------|-------------------|-----------------------------|
| <i>CCAM<sup>std</sup></i> | 0.140***<br>(0.034) | 0.183**<br>(0.072)  | 0.168***<br>(0.051) | 0.071<br>(0.053)  | 0.102<br>(0.064)            |
| <i>OCAM<sup>std</sup></i> | 0.218***<br>(0.042) | 0.290***<br>(0.092) | 0.260***<br>(0.060) | 0.110*<br>(0.068) | 0.150*<br>(0.82)            |

### A.4 Correlations between numerator and denominator terms

In this section we demonstrate that the correlation between the numerator and the denominator of the Amihud measure substantially increases when we use absolute open-to-close

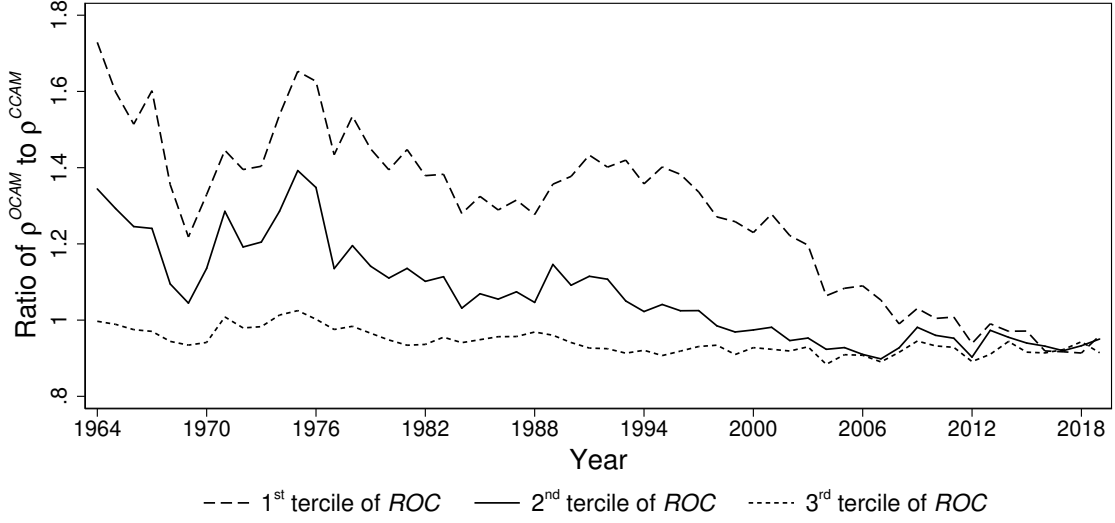


Figure A.2: **Evolution of the ratio of *OCAM* to *CCAM*.** The figure plots the temporal changes in the cross-stock distribution of the *OCAM*-to-*CCAM* for NYSE- and AMEX-listed common shares in the 1964-2019 period. For each stock  $i$  in year  $y$ , the ratios  $ROC_{iy} = OCAM_{i,y-1}/CCAM_{i,y-1}$  and  $ROC_{iy}^\rho = \rho_{i,y-1}^{OCAM}/\rho_{i,y-1}^{CCAM}$  are calculated. Stocks are sorted into terciles of  $ROC_{iy}$  each year, and the year-specific medians of  $ROC^\rho$  in the respective terciles of  $ROC$  are plotted against time

returns, instead of close-to-close returns, in the numerator of the price impact proxy. Equation (15) does not account for the fact that covariance is not a scale-free statistic. This means that  $ROC^{Cov}$  may be impacted by the scale effects embed in differences between the levels of  $|R|$  and  $|ROC|$ . As a result, to accurately measure the effect of removing overnight returns on the linear association between the numerator and  $1/DVOL$ , we next produce the analogue of Panel B in Figure 4 for the ratio of correlation coefficients,

$$ROC^{rho} \equiv \frac{\rho^{OCAM}}{\rho^{CCAM}} = \frac{Corr\left(|OCR|, \frac{1}{DVOL}\right)}{Corr\left(|R|, \frac{1}{DVOL}\right)}, \quad (16)$$

where  $Corr(.,.)$  stands for the correlation coefficient.

Figure A.2 shows that, consistent with theory, the correlation between the numerator and the denominator of the Amihud measure rises substantially when we filter out overnight price movements. More importantly, this feature is more pronounced for stocks whose *OCAM*

and *CCAM* differ by more, i.e., for stocks that belong to the lower terciles of *ROC*. Of note, this effect remains present until 2014.

## A.5 The effects of the open auction

In this section we show that variations driven by the open auction may not drive our findings. We fit equation (7) for the 1993–2013 period using four versions of *OCAM* reflecting “alternative” open prices. We find that excluding the first 5 and 10 minutes of trading day before constructing the Amihud measure greatly reduces the noise in capturing liquidity. That is, excluding the first five minutes results in an improved open to close Amihud measure. We reinforce this conclusion by showing that the correlation between the numerator and the denominator of the measure is often negative or close to zero in the first 5 minutes of trading, whereas it is high (and very similar across subperiods) for the rest of the trading day.

To complement the baseline *OCAM* (and *CCAM*) measure that includes official open and close prices as well as the dollar volume from the entire trading day, we use TAQ data to construct three alternative measures. *OCAM<sup>x</sup>* excludes trading volume from the first  $x$  minutes of the trading day and using the most recent transaction price prior to the  $x^{th}$  minute of the trading day as “open” price, with  $x \in \{5, 10, 15\}$ . For this smaller subset of stocks (which requires matching across both our sample and TAQ), neither *CCAM* nor *OCAM* is priced. This reflects the fact that matching TAQ with our main sample reduces the per-year average number of stocks in the cross-section from 1,708 to 1,436, and the cross-stock average market-cap rises 17% from \$4.8 billion to \$5.6 billion. The insignificant *OCAM* coefficient reflects the exclusion of smaller stocks that, on average, command larger liquidity premia, and are less subject to the over-aggregation bias that drives estimated liquidity premia down toward zero. By contrast, *OCAM<sup>5</sup>* and *OCAM<sup>10</sup>* are significantly priced—showing that excluding the open auction eliminates substantial noise and retrieves the measure’s ability to explain expected returns—although *OCAM<sup>15</sup>* is not.

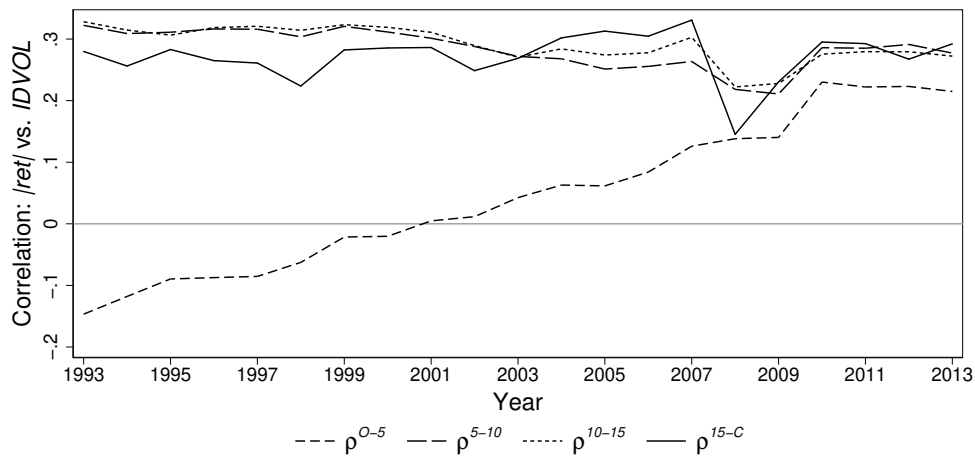
Table A.3: **Fama-MacBeth estimates of monthly returns on stock characteristics, NYSE- and AMEX-listed common shares, 1993–2013.** This table presents Fama-MacBeth estimates of Equation (7) for NYSE- and AMEX-listed common shares in the 1993–2013, contrasting findings given simple *CCAM* and *OCAM*, as well as three alternative versions of *OCAM* using TAQ data. *OCAM<sup>x</sup>* excludes trading volume from the first  $x$  minutes of the trading day and using the most recent transaction price prior to the  $x^{th}$  minute of the trading day as “open” price, with  $x \in \{5, 10, 15\}$ . Variable construction is identical to that in Table 4 in the paper. Estimates are carried out using weighted least squares, with lagged monthly gross (one plus) return used as weights, to correct for biases identified by Asparouhova et al. (2010). Newey-West standard errors using two lags are reported in parentheses. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

| Liquidity measure        | Coefficient<br>(St. Dev.) |
|--------------------------|---------------------------|
| <i>CCAM</i>              | 0.008<br>(0.016)          |
| <i>OCAM</i>              | 0.020<br>(0.021)          |
| <i>OCAM<sup>5</sup></i>  | 0.025**<br>(0.011)        |
| <i>OCAM<sup>10</sup></i> | 0.023*<br>0.013           |
| <i>OCAM<sup>15</sup></i> | (0.010)<br>0.016          |

Our findings indicate that the first few minutes of the trading day contain noisy trading and price dynamics that distort *OCAM* from capturing liquidity. To show this, we divide the trading day into four windows  $x$ : Open-9:35am, denoted by  $x = O - 5$ ; 9:35-9:40am, denoted by  $x = 5 - 10$ ; 9:40-9:45am, denoted by  $x = 10 - 15$ ; and 9:45am-close, denoted by  $x = 15 - C$ . Then, for each stock, we calculate the correlation coefficients between absolute return  $|ret|^x$  and dollar volumes  $DVOL^x$ , denoting them  $\rho^x$  for windows  $x \in \{O - 5, 5 - 10, 10 - 15, 15 - C\}$ . Recall that the first three windows are excluded cumulatively when we construct alternative versions of *OCAM*. Hence,  $\rho^{15-C}$  serves as a benchmark. Recall that, as Table 7 shows, the correlation between absolute return and dollar-volume, if obtained at the daily level, is

priced in the cross-section. The figure below shows the cross-sectional averages of these correlations over time.  $\rho^{O-5}$  is remarkably smaller than its analogues in the following windows. The other windows, however, contain useful information for liquidity in the context of the Amihud measure since  $\rho^{5-10}$ ,  $\rho^{10-15}$  and  $\rho^{15-C}$  are fairly close and stable across all years. As such, while excluding the first 5 minutes improve the Amihud measure, excluding minutes past 9:40am hurts more that it helps.

Figure A.3: **Correlation between absolute return and *DVOL* by time of day**



## A.6 Liquidity premia 1964-2019: panel regressions

In this section, we contrast the explanatory powers of *CCAM* and *OCAM* in panel regression settings, and document robustness of our main findings. To motivate the use of panel regressions, we first note that Fama-MacBeth regressions are not designed to control for unobserved temporally-fixed stock characteristics when estimating coefficients, and cannot account for fixed or varying auto-correlations in the error terms that may lead to inflated  $t$ -statistics (see Petersen 2009). Another possible concern with our previous estimates is that we match all monthly returns observations in year  $y$  with measures constructed using data from year  $y - 1$ . This matching suggests that investors only care about information from last calendar year rather than more recent information.

We estimate panels of monthly stock returns that are matched with stock characteristics constructed, mostly, from the most recent twelve months of data (see Section 2)—similar to Lou and Shu (2017) and Amihud and Noh (2020), we allow an additional month of gap for measures of liquidity, dividend yield, and volatility. Importantly, panel regressions facilitate (i) controlling for invariant stock and time characteristics by use of stock and month-year fixed effects; (ii) estimating clustered standard errors at the stock level to account for the fact that most of the independent variables in our analysis are auto-correlated. Both of these qualities are absent in Fama-MacBeth regressions. Hence, we estimate the panel of monthly returns

$$\begin{aligned}
R_{it} = & \alpha_0 + \alpha_1 LIQ_{i,t-2} + \alpha_2 \beta_{i,t-1}^{mkt} + \alpha_3 \beta_{i,t-1}^{hml} + \alpha_4 \beta_{i,t-1}^{smb} + \alpha_5 \beta_{i,t-1}^{umd} \\
& + \alpha_6 \ln(M_{i,t-12}) + \alpha_7 RET_{i,t-1} + \alpha_8 RET_{i,t-12}^{t-2} + \alpha_9 SDRET_{i,t-2} \\
& + \alpha_{10} DYD_{i,t-2} + \text{fixed effects} + u_{it},
\end{aligned} \tag{17}$$

with independent variables described in Section 2.<sup>38</sup> Table A.4 shows that qualitatively identical findings obtain when we estimate the model using panel regressions, constructing

---

<sup>38</sup>Also see the caption in Table A.4



Table A.4: **Panel regression estimates of monthly returns on stock characteristics, NYSE- and AMEX-listed common shares, 1964–2019.** This table presents GLS estimates of Equation (17) for NYSE-listed common shares in time periods 1964–2019, 1964–1980, 1981–2000, and 2011–2019. The dependent variable is the monthly return, in percentage points. *CCAM* is the traditional Amihud (2002) liquidity measure, and *OCAM* is the Amihud (2002) measure after removing overnight price movements. Both measures are constructed monthly, using daily absolute return per dollar observations from the 12-month period ending in  $t-2$ . Liquidity premium, in basis points, is the product of the coefficient on the liquidity measure and the measure’s average monthly standard deviation. Dividend yield,  $DY_{i,t-2}$ , divides total dividend distributions between months  $t-13$  and  $t-2$  by the closing price at the end of month  $t-2$ . Momentum measures  $RET_{i,t-1}$  and  $RET_{i,t-1}^{t-12}$ , respectively, capture compound returns over the preceding month and the eleven months before that. Return volatility  $SDRET_{i,t-2}$  is given by the standard deviation of daily stock returns over the 12-month ending in month  $t-2$ . Market capitalization,  $M_{i,t-12}$ , is the product of shares outstanding and the closing price at the end of the month, a year earlier. Panel regression estimates control for stock and year-month fixed effects. Standard errors are clustered at the stock level. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|                     | 1964–2017            |                      | 1964–1980            |                      | 1981–2000            |                      | 2001–2019            |                      |
|---------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| <i>CCAM</i>         | 0.139***<br>(0.015)  |                      | 0.098***<br>(0.018)  |                      | 0.395***<br>(0.056)  |                      | 0.342<br>(0.210)     |                      |
| <i>OCAM</i>         |                      | 0.604***<br>(0.054)  |                      | 0.549***<br>(0.065)  |                      | 2.00***<br>(0.220)   |                      | 1.22**<br>(0.529)    |
| $\beta^{mkt}$       | −0.068<br>(0.043)    | −0.059<br>(0.043)    | −0.043<br>(0.065)    | −0.032<br>(0.065)    | 0.187**<br>(0.086)   | 0.206**<br>(0.086)   | −0.388***<br>(0.098) | −0.380***<br>(0.098) |
| $\beta^{hml}$       | 0.058**<br>(0.024)   | 0.053**<br>(0.024)   | 0.085**<br>(0.034)   | 0.077**<br>(0.035)   | −0.095*<br>(0.051)   | −0.106**<br>(0.051)  | 0.065<br>(0.056)     | 0.066<br>(0.056)     |
| $\beta_{smb}$       | −0.101***<br>(0.027) | −0.106***<br>(0.028) | −0.107***<br>(0.039) | −0.111***<br>(0.039) | −0.010<br>(0.055)    | −0.011<br>(0.055)    | −0.225***<br>(0.068) | −0.219***<br>(0.068) |
| $\beta^{umd}$       | 0.043<br>(0.030)     | 0.048<br>(0.030)     | −0.012<br>(0.040)    | 0.001<br>(0.040)     | −0.196***<br>(0.055) | −0.190***<br>(0.055) | 0.419***<br>(0.074)  | 0.420***<br>(0.074)  |
| $\ln(M)$            | −1.27***<br>(0.038)  | −1.25***<br>(0.038)  | −2.21***<br>(0.069)  | −2.09***<br>(0.070)  | −1.75***<br>(0.070)  | −1.65***<br>(0.070)  | −1.58***<br>(0.081)  | −1.56***<br>(0.081)  |
| $R_{t-1}$           | −0.054***<br>(0.002) | −0.054***<br>(0.002) | −0.086***<br>(0.003) | −0.087***<br>(0.003) | −0.059***<br>(0.003) | −0.059***<br>(0.003) | −0.044***<br>(0.004) | −0.044***<br>(0.004) |
| $R_{t-12}^{t-2}$    | −0.002***<br>(0.000) | −0.002***<br>(0.000) | 0.000<br>(0.001)     | 0.001<br>(0.001)     | −0.007***<br>(0.001) | −0.007***<br>(0.001) | −0.014***<br>(0.001) | −0.014***<br>(0.001) |
| <i>SDRET</i>        | −0.133<br>(0.169)    | −0.284*<br>(0.171)   | −2.18***<br>(0.244)  | −2.56***<br>(0.249)  | −1.50***<br>(0.321)  | −1.86***<br>(0.326)  | 2.35***<br>(0.319)   | 2.35***<br>(0.318)   |
| <i>DYD</i>          | 0.005<br>(0.253)     | 0.007<br>(0.253)     | 0.613<br>(0.497)     | 0.607<br>(0.507)     | −0.144<br>(0.241)    | −0.138<br>(0.243)    | 0.504***<br>(0.119)  | 0.503***<br>(0.119)  |
| Premia <sup>†</sup> | 0.161                | 0.218                | 0.237                | 0.399                | 0.359                | 0.537                | 0.101                | 0.159                |

<sup>†</sup>Reflects the change in excess return in response to one standard deviation increase in illiquidity.

stocks characteristics at a higher granularity.

We also establish the robustness of the orthogonal-decomposition analysis in the context of panel regressions. Then we use the residuals obtained in Equations (8) and (9) to perform the analogous analysis to that in Section 5. Table A.5 shows that the residuals from regressing *OCAM* on *CCAM* provides incremental explanatory power when we control for *CCAM*. In sharp contrast, the residuals from regressing *CCAM* on *OCAM* do not add any incremental explanatory power when we control from *OCAM*.

**Table A.5: Panel regression estimates of monthly returns on stock characteristics and linearly-decomposed liquidity measures, NYSE- and AMEX-listed common shares, 1964–2019.** This table presents GLS estimates of Equation (17) for NYSE-listed common shares in time periods 1964–2019, 1964–1980, 1981–2000, and 2011–2019. The dependent variable is the monthly return, in percentage points. *CCAM* is the traditional Amihud (2002) liquidity measure, and *OCAM* is the Amihud (2002) measure after removing overnight price movements. *CCAM* (*OCAM*) is included along with residuals from Equation (9) (Equation (8)), as an additional independent variables. The rest of independent variables are identical to those in Table A.4. Panel regression estimates control for stock and year-month fixed effects. Standard errors are clustered at the stock level. Symbols \*, \*\*, and \*\*\* reflect statistical significance at 10%, 5%, and 1% type one error, respectively.

|             | 1964–2017           | 1964–1980           | 1981–2000           | 2001–2019          |
|-------------|---------------------|---------------------|---------------------|--------------------|
| <i>CCAM</i> | 0.148***<br>(0.015) | 0.113***<br>(0.018) | 0.465***<br>(0.057) | 0.371*<br>(0.211)  |
| $\tilde{Z}$ | 0.583***<br>(0.095) | 0.567***<br>(0.108) | 2.03***<br>(0.339)  | 2.72***<br>(0.926) |
| <i>OCAM</i> | 0.604***<br>(0.055) | 0.545***<br>(0.065) | 1.98***<br>(0.219)  | 1.11**<br>(0.528)  |
| <i>Z</i>    | −0.001<br>(0.026)   | −0.023<br>(0.029)   | −0.053<br>(0.085)   | −0.658*<br>(0.368) |