Dynamics in Systematic Liquidity

By

Björn Hagströmer, Richard G. Anderson, Jane M. Binner, Birger Nilsson*

RP 0923

Björn Hagströmer
Aston University
SW11, Aston Triangle,
B4 7ET BIRMINGHAM, UK.
E-mail: hagstrob@aston.ac.uk

Jane Binner
Aston Business School
Birmingham

Richard G. Anderson
Federal Re-serve Bank of St. Louis

Birger Nilsson
Lund University, Sweden.

Dynamics in Systematic Liquidity

Björn Hagströmer, Richard G. Anderson, Jane M. Binner, Birger Nilsson[*]

May 26, 2009

Abstract

We develop the principal component analysis (PCA) approach to systematic liquidity measurement by introducing moving and expanding estimation windows. We evaluate these methods along with traditional estimation techniques (full sample PCA and market average) in terms of ability to explain (1) cross-sectional stock liquidity and (2) cross-sectional stock returns. For several traditional liquidity measures our results suggest an expanding window specification for systematic liquidity estimation. However, for price impact liquidity measures we find support for a moving window specification. The market average proxy of systematic liquidity produces the same degree of commonality, but does not have the same ability to explain stock returns as the PCA-based estimates.

Keywords: systematic liquidity, market liquidity, commonality, dynamic principal component analysis, robust PCA

1 Introduction

It is well-established that liquidity (the ease of trading) affects equity prices and returns (Amihud, Pedersen and Mendelson 2005, survey this literature). It is now also well-documented that the comovement of stocks' liquidity, referred to as *commonality in liquidity*, is driven by unobservable underlying market liquidity factors, or *systematic liquidity factors* (Chordia, Roll and Subrahmanyam 2000, Huberman and Halka 2001, Hasbrouck and Seppi 2001). It has been shown in many studies that these factors affects individual-asset returns (Pastor and Stambaugh 2003, Acharya and Pedersen 2005, Korajczyk and Sadka 2008). Hence, systematic liquidity factors are relevant as risk factors to any investor.

To manage liquidity risk, accurate measurement of systematic liquidity is crucial. In this paper we use the familiar principal components estimation technique to derive the factors. We extend previous research by introducing dynamic estimation windows in this setting, allowing us to test the temporal stability (robustness) of the systematic liquidity measure.

---

[*]Björn Hagströmer and Jane M. Binner are from Aston University. Richard G. Anderson is from Federal Reserve Bank of St. Louis. Birger Nilsson is from Lund University, Sweden. Correspondence to Björn Hagströmer: Address: Aston University SW11, Aston Triangle, B4 7ET BIRMINGHAM, UK. Tel: +44(0)7758 355 883. Fax: +44(0)121 2043326. E-mail: *hagstrob@aston.ac.uk*

1

Principal component analysis (PCA) is a popular method for explaining the covariation between many variables in terms of a small number of common factors. Most investigations use a static specification of PCA, based on analysis of the covariance matrix of variables of interest, which is assumed to be time-invariant. In the case of liquidity, where both systematic liquidity and commonality in liquidity are known to vary over time (Chordia, Roll and Subrahmanyam 2001, Kamara, Lou and Sadka 2008), this assumption may be counterfactual. Still, all studies using PCA to derive systematic liquidity choose this static approach.[1] We test the appropriateness of this assumption by introducing moving and expanding (recursive) estimation windows for PCA.[2] We compare our results to those obtained from PCA with static window (i.e., where the full sample is used); and those from a cross-sectional (equal weighted) liquidity average, another common approximation of systematic liquidity (Chordia et al. 2000); giving us four methods to approximate systematic liquidity.

No consensus has been reached on how to measure liquidity. We investigate eight different liquidity measures, including quoted spread, effective spread, turnover, Amihud's (2002) illiquidity measure, and Sadka's (2006) measures of market making costs (inventory cost and adverse selection cost). We utilize the Trade and Quote database (TAQ) to calculate monthly liquidity measures for the S&P500 stocks between 1995 and 2007.

For each liquidity measure we evaluate the four different methods for deriving systematic liquidity factors, using two different evaluation criteria: (A) ability to explain cross-sectional variation in stock liquidity (i.e., commonality in liquidity) and (B) ability to explain cross-sectional variation in stock returns. We argue that high commonality is an indicator of systematic liquidity measurement accuracy. PCA is based on the estimated covariance matrix. If the estimated covariance matrix differs over different estimation horizons it can be due to either sampling error or actual time-variation in the underlying covariance matrix. If it is due to noise, a longer estimation window should yield higher commonality, as the sampling error is smaller with more data. If there is time-variation in the underlying covariance matrix, the PCA with moving estimation window should be able to capture this time-variation and therefore produce a higher degree of commonality. Hence, by running commonality tests we investigate which estimation method yields factors that best summarize cross-sectional liquidity variation.

---

[1] Hasbrouck and Seppi (2001); Beltran-Lopez, Giot and Grammig (2006); Chen (2007); Chollete, Næs and Skjeltorp (2007); Chollete, Næs and Skjeltorp (2008); Kempf and Mayston (2008); Kopp, Hütl, Loistl and Prix (2008); Korajczyk and Sadka (2008).

[2] The only related liquidity study that does not stay with the static window (full sample) PCA is Chen (2007) who uses an expanding window specification; however she does not discuss the implications of this and the constant covariance assumption remains present.

2

Many studies find a relation between stock liquidity and returns. Therefore, if a systematic liquidity measure can explain cross-sectional stock liquidity, it should be able to explain returns as well. As the relation between liquidity and returns is central to the application of systematic liquidity factors, we make this ability our second evaluation criterion in our assessment of estimation methods for systematic liquidity. We run an extended market model including systematic liquidity factors, and study the improvement in return variation explanation relative to a standard market model.

Our first finding is that time-series properties of the liquidity covariance matrix differ across liquidity measures. When considering liquidity measured by the bid-ask spread, turnover or the Amihud measure, the accuracy of estimated systematic liquidity is not improved by allowing the covariance matrix to vary over time. The static, expanding and moving window approaches produce roughly the same average commonality. This suggests that the covariance matrices of these measures are quite stable over time. For the price impact measures by Sadka (2006), however, we find that the moving window approach improves the accuracy of estimated systematic liquidity, i.e., the average commonality is consistently higher when applying the moving window PCA. Hence, there appears to be time-variation in the underlying covariance matrix for these measures. Our investigation of ability to explain stock returns point in the same direction as that of stock liquidity. For most liquidity measures, different estimation window specifications for PCA do not make any difference for explaining returns. Transitory fixed market maker costs, which is Sadka's (2006) measure of inventory cost, is an important exception to this. This liquidity measure, relative to the other measures considered, has a good ability in both explaining stock liquidity and stock returns (highest of all measures), and it is best estimated using the moving window PCA.

A second finding is that the market average proxy of systematic liquidity, suggested by Chordia et al. (2000), produces a commonality that on average is in line with the various PCA factors for all measures considered (with the exception of the Amihud measure, which is plagued by outliers). For turnover and spread measures of liquidity, the cross-sectional average can be used as proxy for systematic liquidity. For the other measures, time series properties in commonality produced by the market average differ substantially from those of PCA measures. Furthermore, when applied to stock return variation, the PCA-based measures are superior to the market average regardless of the choice of liquidity measure.

Thirdly, we find that the properties of static window (full sample) PCA and the expanding window (recursive) PCA in terms of ability to explain both stock liquidity and stock returns

3

are converging over time. The latter is computationally more expensive, but the former is suffering from a forward looking bias that undermines its applicability in practice. Our findings indicate that the static window PCA can be replaced by expanding window PCA without loss of accuracy.

In the next section we present the framework applied to estimate systematic liquidity factors with PCA, along with a discussion on how to deal with missing data, outliers, and computational burden for this kind of problems. In section 3 we present details of the data, its processing, and the individual liquidity measures that we apply. In section 4 we perform commonality tests and analyze results thereof. In section 5 we evaluate systematic liquidity in terms of explanatory power with respect to stock returns. Finally, in section 6 we summarize the main implications of the study and point out potential future research.

2 Systematic liquidity derivation

We assume that the data-generating process of liquidity for a given stock is driven by some underlying market liquidity factor and an idiosyncratic liquidity variable. In a market of N firms for which we consider τ time periods, yielding a stock liquidity matrix L^* of dimension $(N \times \tau)$, this process can be described as

$$L^* = \mu\iota' + BF + U. \tag{1}$$

The underlying market liquidity factor is described by K row vectors in the matrix F $(K \times \tau)$ and each stock's sensitivity to these vectors is given in the matrix B $(N \times K)$. U is the matrix of idiosyncratic liquidity shocks $(N \times \tau)$.

The average liquidity across time for each stock is given in the vector μ of length N, ι is a vector of ones of length τ. In order to estimate F with equal influence of each stock's variance, we standardize the liquidity measure to have unit variance and zero mean over the τ time periods.[3] Denoting the vector of stock liquidity standard deviation for each firm σ, the standardized liquidity matrix takes the form

$$L_{i,t} = (L^*_{i,t} - \mu_i)/\sigma_i. \tag{2}$$

---

[3] Standardization of the data is customary for PCA, but the exact methodology to do this has varied in the liquidity literature (and has often not been disclosed). For example, Korajczyk and Sadka (2008) are using an expanding estimation window to calculate market-wide mean and standard deviation. We standardize on a stock-by-stock basis using means and standard deviations calculated for the τ time periods considered, as market-wide standardization may give disproportionate weight to stocks with high liquidity volatility.

The factor model described in Equation 1 can then be expressed as

$$L = \beta F + \epsilon, \tag{3}$$

where $\beta_{i,k} = B_{i,k}/\sigma_i$ and $\epsilon_{i,t} = U_{i,t}/\sigma_i$. From this model, we can estimate the underlying market factor $\hat{F}$ using PCA. The principal components are the eigenvectors of the covariance matrix of L multiplied by L, and sorted into row $k = 1, 2, ..., K$ of $\hat{F}$ by its corresponding eigenvalues, starting with the highest.

2.1 PCA with dynamic estimation window

In the systematic liquidity literature, τ has typically been set to the sample size T, yielding a static $\hat{F}$. This implies an assumption of constant covariances in the cross-section over the sample size T. We challenge the validity of this assumption by running a moving window PCA, utilizing only the τ latest observations. If covariances are time-variant, this will exclude old observations that may yield biased covariance estimates. Formally, this implies a structure with time indices $t = 1, 2, ..., T$ on each of the variables in Equation 3 and a time span $(t - \tau + 1) : t$ considered for estimation in each period t. The estimation is repeated for $t = \tau, \tau + 1, \tau + 2, ..., T$. We also consider expanding window PCA (or recursive PCA), where the problem dimension is growing with t, so we have $\tau = t$. This specification is appropriate in problems where the covariances are believed to be time-invariant, and where the usage of future data should be avoided. Such a specification has earlier been considered in a systematic liquidity study by Chen (2007). Earlier economics applications of PCA with dynamic estimation window have appeared in literature on integration of equity markets (Volosovych 2005, Gilmore, Lucey and McManus 2008) and macroeconomic forecasting (Heij, van Dijk and Groenen 2008), and more commonly outside economics, in particular process monitoring (Li, Yue, Valle-Cervantes and Qin 2000, Wang and Xia 2002). We label these methods *PCA with dynamic estimation window*. Another type of dynamics is the dynamic PCA (or generalized dynamic factor model), introduced by Forni, Hallin, Lippi and Reichlin (2000, 2005), where leads and lags are considered in the estimation. We do not pursue this technique, but a systematic liquidity application is available in Hallin, Mathias, Pirotte and Veredas (2009).

For the expanding window PCA we start the estimation at $\tau = 36$ and for the moving window PCA we keep the sample size constant at that size throughout the sample. As a comparison to previous studies, we also provide the traditional version of PCA, estimated over the whole

sample, which in our notation corresponds to $\tau = T$ (we refer to this as *static window PCA*). This implies that we run PCA with dynamic estimation window of sizes $\tau = [36, t, T]$. For further comparison, we evaluate these factors together with an equal-weighted cross-sectional average of liquidity.

2.2 Estimation methodology: Robust Asymptotic PCA

PCA has some technical complications that have rarely been discussed in the applications to liquidity. When applying PCA with dynamic estimation windows, where PCA is run hundreds of times, these complications become important for the robustness of the results. As PCA is based on finding eigenvectors in the covariance matrix of the underlying variables, it is important that the covariances are robustly estimated. In this subsection we discuss problems associated with that estimation in some detail, including issues with missing values, outliers, negative eigenvalues, and computational burden. For a more general discussion of PCA and technical considerations thereof, see e.g., Jolliffe (2002).

Firstly, the estimation of the covariance matrix needs to be adapted for missing values. Typically, when dealing with stock market data there is an unbalanced panel of observations as companies rise and fall and go through mergers and acquisitions. As exclusion of companies that do not exist throughout the sample would cause material information losses, a method for dealing with long series of missing values is needed. With a moving window specification, stocks with long series of missing values can be excluded until they have enough observations. This partially sidesteps the missing value problem, but temporary cases of missing values remain. To estimate covariances in the presence of missing values, we use element-wise estimation, using all pairs of observations available for the two variables (referred by Jolliffe 2002, Ch. 13). For a stock to be included in the analysis, we demand that more than 12 observations are available in the time span considered.[4]

Secondly, outliers can inflate or deflate variances and covariances to an extent that they mislead the PCA substantially (see Jolliffe 2002, Ch.10). In our setting, large idiosyncratic stock liquidity shocks are easily imaginable and these should not be allowed to influence the estimation of systematic liquidity. To deal with this, we use the robust covariance matrix estimation technique by Mehrotra (1995). This technique is based on using medians rather than means when calculating slopes needed in variance and covariance estimation. This is a computationally expensive technique, but it yields a high stability in the principal components.

---

[4]For two alternative methods dealing with missing values in a static setting, see Korajczyk and Sadka (2008).

Thirdly, a problem appearing when adjusting the covariance matrix as explained above, is that it may no longer be positive semi-definite. To limit this problem, we use asymptotic PCA as suggested by Connor and Korajczyk (1986). With this methodology, the covariances between time periods rather than the covariances between stocks are applied as the basis for PCA. Connor and Korajczyk show that components derived using this method are asymptotically equivalent to those of PCA, and our investigation shows that in our setting of 500 stocks, the methods are equivalent. When using asymptotic PCA, the problem of negative eigenvalues is much smaller than for PCA, though it is still present. Furthermore, the computational burden of PCA is in this way limited, as the covariance matrix has dimension τ rather than N.

3 Liquidity measures and data

We limit this study to the stocks that were in S&P500 by the end of 2007. The maximum time frame considered is 1995-2007, but many stocks have shorter series due to name and/or ticker changes and company restructuring. To measure liquidity with high accuracy we need observations on transactions and order book entries. We use the TAQ database which contains all transactions on major US stock exchanges and frequently updated quotes of best bid and best ask price from the order books. It also holds data on lot size, shares outstanding and share types. S&P500 stocks have their primary listing on NYSE (most), Nasdaq (about 100), and Amex (few). Accordingly, we restrict our data set to equities from NYSE, Nasdaq, Amex and Nasdaq-ADF.[5] In the liquidity literature it is common to exclude Nasdaq trades, as their reporting of trades differs from other exchanges, which can lead to exaggerated trade volumes. As we want to study the S&P500 index liquidity, we do not follow this convention. About one fifth of the stocks in the index have their primary listing on Nasdaq, so excluding them would give a biased picture of the index.

The liquidity of a market has many aspects. Kyle (1985) summarize these aspects in three points: (1) *Tightness:* How much it costs to turn over a position in a short time; (2) *Depth:* The ability of the market to absorb quantities without large price impacts (which relates to Kyle's λ); (3) *Resiliency:* The market's ability to quickly return to the underlying value of a security, e.g., after an uninformative price shock. This multi-dimensionality of liquidity has triggered a voluminous literature on what is the most appropriate measurement methodology. Depending on the research problem different researchers have found different measures more or

---

[5]Nasdaq-ADF is an automatized trading platform.

less appropriate and no consensus on measurement has been reached. As it is not the purpose of this article to pursue that measurement discussion further, we are using a set of eight liquidity measures in the subsequent analysis.

The first four liquidity measures are calculated using the expressions presented in Table 1. All variables presented in this section are firm specific (where the opposite is not noted), but the firm index i $(i = 1, 2, ..., N)$ has been dropped for brevity of exposition. Note that most of these measures can be derived using daily data as well. To achieve highest possible quality we use the high frequency data for all our measures. All measures are of monthly frequency.

The most straightforward measure of liquidity is the bid-ask spread, which provides a measure of tightness in the market for a security. The spread is an intuitive measure of the cost of an immediately executed roundtrip trade of a stock (for a single trade the half spread can be used). According to Amihud and Mendelson (1989) the measure is related to characteristics associated with liquidity such as the number of investors, the transaction volume, the information availability, and the size of the firm. We consider both the *quoted* and the *effective* spreads, the latter being adjusted for the fact that trades often are executed within or outside the quoted spread both at the NYSE and Nasdaq (because of e.g., bettered quotes or hidden limit orders, see discussion in Huang and Stoll 1996 and Petersen and Fialkowski 1994). In order to be cross-sectionally comparable both spread measures are divided by the spread midpoint ($mp_{k,t}$).

Turnover is the ratio of traded volume and shares outstanding, which is an intuitive proxy of liquidity as its inverse is the average holding period of the stock, which should be longer for illiquid stocks (see Amihud and Mendelson 1986).

Amihud (2002) suggest a liquidity measure that is a proxy of the adverse selection cost (the likelihood of trading with informed traders, see Akerlof 1970; Bagehot 1971), related to the depth of the market and described by λ in the model by Kyle (1985). This measure, the average daily ratio of absolute returns and dollar volume, proxies the price impact of trading using daily data. This is attractive when intraday data is unavailable, particularly for asset pricing studies where long time series are needed.[6] Due to its popularity in the literature, we include the *Amihud* measure even though we have access to intraday data.

Glosten and Harris (1988) summarize the modeling of how market maker costs motivate the bid-ask spread in what is called *the asymmetric information model*. Following Kyle (1985), the order flow is divided between information-based trading and noise trading. The observed price

---

[6]The Amihud measure is adjusted for market capitalization as suggested by Acharya and Pedersen (2005), on the basis that dollar volume tends to grow over time, indicating inflation rather than increasing liquidity. Note that market capitalization ($mcap$) is a market-wide measure – the same factor is used throughout the cross-section of firms. We use January 1993 as reference date.

Name	Derivation
Quoted Spread	$L_t^1 = \frac{1}{n_t} \sum_{k=1}^{n_t} \frac{Ask_{k,t} - Bid_{k,t}}{mp_{k,t}}$
Effective Spread	$L_t^2 = \frac{1}{n_t} \sum_{k=1}^{n_t} \frac{|p_{k,t} - mp_{k,t}|}{mp_{k,t}}$
Turnover	$L_t^3 = \frac{1}{SO_t} \sum_{j=1}^{d_t} vol_{j,t}$
Amihud	$L_t^4 = \frac{mcap_{ref}}{mcap_t} \sum_{j=1}^{d_t} \frac{|r_{j,t}|}{dvol_{j,t}}$

Table 1: Liquidity measurement

Notation: $Ask_{k,t}$ and $Bid_{k,t}$ are the best ask and best bid prices prevailing at the time of trade k in month t, and $mp_{k,t}$ is the midpoint which is defined as the average of the Ask and Bid prices. The transaction price of trade k in month t is denoted $p_{k,t}$, and the number of trades in month t is n_t. The volume of stocks traded on day j in month t is denoted $vol_{j,t}$, and SO_t is shares outstanding in month t. Return on day j in month t is defined $r_{j,t} = (P_{j,t}/P_{j-1,t}) - 1$ where $P_{j,t}$ is the dividend adjusted daily price. Monthly returns are defined by the same principle. Dollar volume traded on day j in month t is denoted $dvol_{j,t}$, and d_t is the number of trading days in month t. The Amihud measure is adjusted for market capitalization as suggested by Acharya and Pedersen (2005), denoting market capitalization of S&P500 in month t $mcap_t$ and at a reference date $mcap_{ref}$ (note that these measures are not firm specific). We use January 1993 as our reference date.

process at the time of trade k, P_k, is equal to the true price, m_k, adjusted for trade-specific costs that the market maker carries. They model this relation as

$$P_k = m_k + D_k C_k, \tag{4}$$

where C_k is a cost component containing fees and inventory costs incurred to the market maker (see Stoll 1978), but excluding costs of asymmetric information. This is called the *transitory spread component*. D_k is a direction of trade dummy which is set to +1 (-1) when the transaction price is higher (lower) than the prevailing spread midpoint, implying that the trade is buyer-initiated (seller-initiated).

The true price of the stock is assumed to be affected only by changes in information related to the company. Information is reaches the market either trough a public information flow, y_k, or through information-based trading that reveals private information. Hence, they model the true price process as

$$m_k = m_{k-1} + y_k + D_k Z_k, \tag{5}$$

where Z_t is the cost component that is due to information-based trading. Z_t is called the adverse

selection spread component, or the *permanent spread component*, as the dynamic specification of the true price process makes the impact of adverse selection costs permanent.

By taking the first difference of Equation 4, i.e., $\Delta P_k = P_k - P_{k-1}$, and inserting Equation 5, Glosten and Harris (1988) retrieve

$$\Delta P_k = D_k Z_k + D_k C_k - D_{k-1} C_{k-1} + y_k. \tag{6}$$

Both spread components are modelled to be linear functions of volume of trade k, V_k, which inserted in Equation 6 yields Equation 9, given below.

$$C_k = \bar{\Psi} + \bar{\lambda} V_k, \tag{7}$$

$$Z_k = \Psi + \lambda V_k. \tag{8}$$

$$\Delta P_k = \Psi D_k + \lambda D_k V_k + \bar{\Psi} \Delta D_k + \bar{\lambda} \Delta(D_k V_k) + y_k. \tag{9}$$

In estimating this relationship Glosten and Harris (1988) find that $\bar{\lambda} = \Psi = 0$, implying that inventory cost is the effect of changes in trade direction $D_k - D_{k-1}$, and that the adverse selection cost is the effect of signed trading volume $D_k V_k$. Hasbrouck (1988, 1991) later on points out that adverse selection costs can only appear in response to unexpected order flows, which has been implemented in the model framework by Foster and Viswanathan (1993) and Brennan and Subrahmanyam (1996) (by using the trade innovations from a model with lagged order flows and lagged price changes). Sadka (2006) incorporates this in the Glosten and Harris's (1988) model by writing the signed order flow function as

$$Z_k = \Psi \left(1 - \frac{E_{k-1}[D_k]}{D_k} \right) + \lambda \left(V_k - \frac{E_{k-1}[D_k V_k]}{D_k} \right), \tag{10}$$

where E_{k-1} denotes conditional expectations. Inserting this in the framework of Glosten and Harris (1988), yields on expression of changes in observed stock prices where it is obvious that adverse selection costs are incurred from unexpected trade flow changes:

$$\Delta P_k = \Psi(D_k - E_{k-1}[D_k]) + \lambda(D_k V_k - E_{k-1}[D_k V_k]) + \bar{\Psi} \Delta D_k + \bar{\lambda} \Delta(D_k V_k) + y_k. \tag{11}$$

Sadka estimates the unexpected signed order flow using an AR(5) model and denotes it $\epsilon_{\lambda,k} = D_k V_k - E_{k-1}[D_k V_k])$. By assuming a normal distribution of $\epsilon_{\lambda,k}$ with constant variance, he is

able to derive the unexpected order flow *direction*, which he denotes $\epsilon_{\Psi,k} = D_k - E_{k-1}[D_k]$, using the cumulative density function. Using the estimated time series of $\epsilon_{\lambda,k}$ and $\epsilon_{\Psi,k}$, he is able to analyze the relationship described in Equation 11 using OLS.

We estimate Equation 11 for each firm and each month, retrieving estimates of the coefficients in the transitory inventory cost (and other market making cost) function ($\bar{\Psi}$ and $\bar{\lambda}$) and of the coefficients in the adverse selection cost function (Ψ and λ). Following Brennan and Subrahmanyam (1996) and Sadka (2006), the coefficients are scaled by the end of month stock price, in order to reflect relative rather than absolute costs. In the exposition below, we refer to the four coefficients as *price impact coefficients*.

In Table 2 basic statistical properties of the eight monthly liquidity measures are presented. For comparative purposes, monthly returns are also included. Monthly returns have a mean of 1.65%, which reflects the bull markets characterizing the time period considered. It should be noted though that the sample also includes the Asian crisis (Fall 1997), the Russian and Long Term Capital Management crisis (Fall 1998), the burst of the dot-com bubble (Spring 2000), and the beginning of the credit crunch (Summer/Fall 2007). The development over time for each of our liquidity measures, on average across stocks, is shown in Figure 3. As seen in Panel A, turnover is increasing in liquidity, whereas the other measures are falling in liquidity. Quoted and effective spreads appear to be correlated, but the former is more than double the magnitude of the latter. This is due to that the effective spread measures the half spread, and that many trades are closed within the quoted spread. As seen in Panel B, each of the price impact coefficients are falling in liquidity, so falling trends over the sample periods are expected, which holds for each coefficient except $\bar{\lambda}$. This measure is also negative on average throughout the sample, which may seem peculiar, but this is in line with the findings of Sadka (2006, pp.317–318), in which a set of explanations for this is provided. Note that this series offsets the λ series. All coefficient series have negative values in their distributions, which is to be expected when a large number of regressions are run. Overall, the values of the price impact coefficients are smaller than what Sadka (2006) found, which reflects the fact that liquidity in general is known to be higher in our sample than in his, both with respect to firms and time periods considered (he investigated 1983-2001, with smaller firms on average). Our results correspond well to Sadka in that the transitory fixed costs are much higher than the permanent fixed costs (by a factor of 10); and that the variable costs are dominated by the permanent adverse selection component (which is 0.37 on average, as compared to -0.01 for the transitory variable cost). The finding of

		Mean	Median	Std.Dev	Max	Min
L_t^1	QSpread (%)	0.37	0.26	0.41	9.99	0.02
L_t^2	ESpread (%)	0.11	0.07	0.14	3.50	0.01
L_t^3	Turnover (%)	10.49	7.70	11.41	556.58	0.122
L_t^4	Amihud $\times 10^6$	0.10	0.02	0.70	47.77	0.00
L_t^5	Ψ (%)	0.006	0.004	0.014	0.779	-0.591
L_t^6	$\lambda \times 10^6$	0.37	0.16	1.38	63.98	-53.42
L_t^7	$\bar{\Psi}$ (%)	0.06	0.03	0.12	3.96	-0.58
L_t^8	$\bar{\lambda} \times 10^6$	-0.12	-0.02	0.84	71.89	-46.58
r_t^m	Returns (%)	1.65	1.37	10.90	292.59	-72.70

Table 2: Descriptive statistics of liquidity measures

The liquidity measures are calculated monthly for 1995-2007 as specified in Table 1 and Equation 11. Ψ and $\bar{\Psi}$ are permanent and transitory fixed costs; λ and $\bar{\lambda}$ are permanent and transitory variable costs of trading.

Glosten and Harris (1988) that $\Psi = \bar{\lambda} = 0$ is however not supported.[7]

4   Commonality in liquidity

The degree of commonality is the proportion of stock liquidity variation that is due to systematic liquidity variation. This can be measured by estimating the relationship

$$\Delta L_{i,l}^* = \gamma_{1,l,\tau} + \gamma_{2,l,\tau}\Delta\hat{F}_{1:k,l,\tau} + u_{i,l,\tau}, \tag{12}$$

which is closely related to Equation 1. Stock liquidity is here regressed on the first k rows of the estimated systematic liquidity factor (i.e., the k first principal components). The residuals, $u_{i,l,\tau}$ are taken to be idiosyncratic liquidity shocks. In accordance with Chordia et al. (2000), the proportion of variation in stock liquidity explained by systematic liquidity, R^2, averaged across stocks, is taken as the degree of liquidity commonality. As indicated by Δ, we follow Chordia et al. (2000) and run the commonality test on innovations in liquidity. These are retrieved by fitting autoregressive models to liquidity measures and systematic liquidity factors.[8] In some previous studies, the commonality test has been run in levels, but we experience substantial non-stationarity problems with such a specification. Note that in addition to the firm index i, we add the subscripts l and τ to distinguish the different liquidity measures and PCA specifications that we evaluate in this framework. Note also that we use non-standardized liquidity data, as

---

[7]The t test null hypothesis of each of the coefficients is rejected in more than 50% of the cases for almost all months throughout the sample. Towards the end of the sample the rejection rate is higher than 75%.

[8]Autoregressive order is chosen using Schwarz Bayesian information criterion, with up to two lags allowed. We allow the lag length to differ across liquidity measures. The model is run using a 36 months moving window.
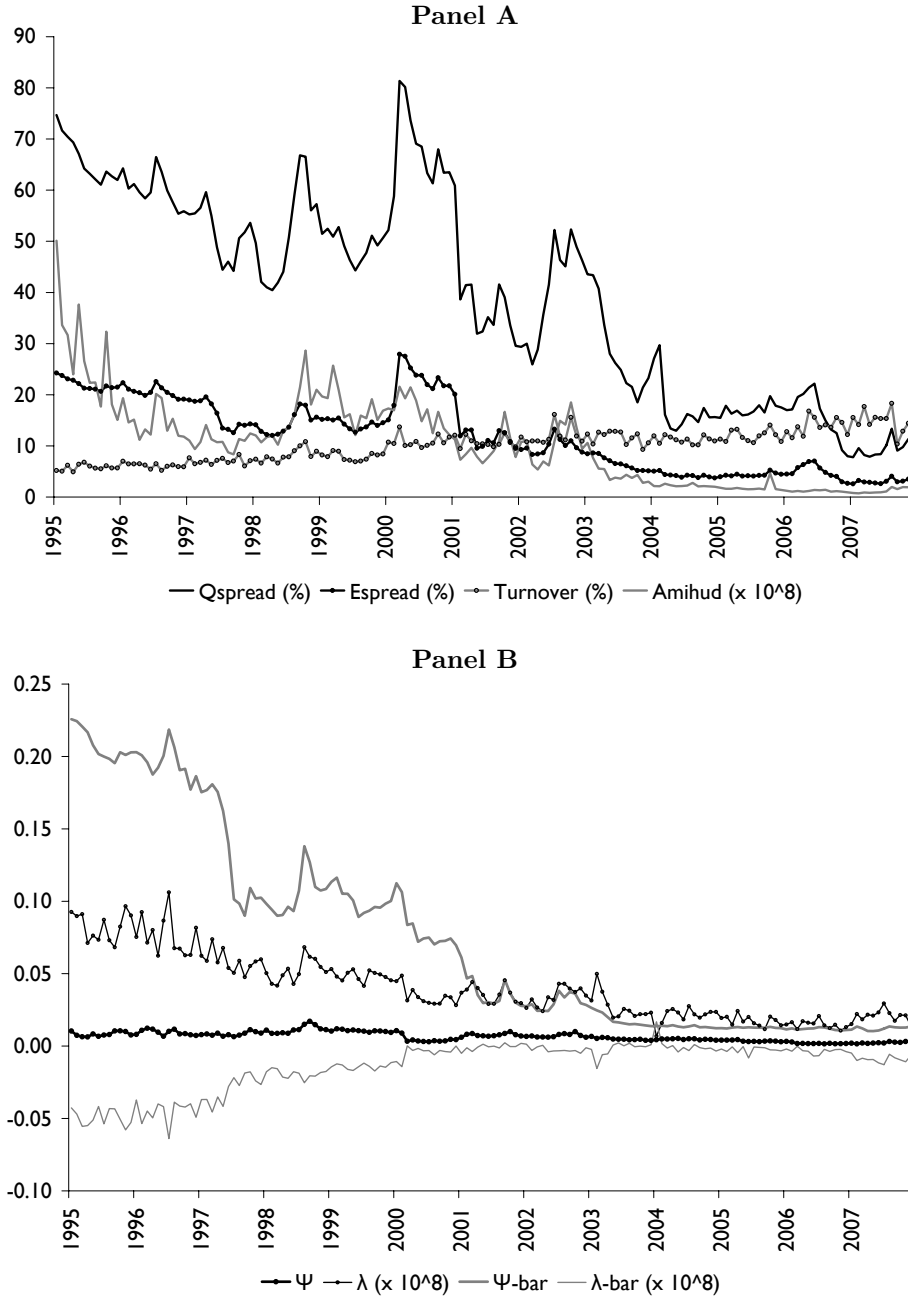
Panel A

Panel B

Figure 1: Average liquidity 1995-2007

This figure shows each of our eight liquidity measures averaged across stocks for the time frame 1995-2007. The expressions for calculating the measures in Panel A are given in 1. The measures in Panel B are coefficients estimated from Equation 11. Ψ and λ are coefficients in the adverse selection cost function, with permanent influence on stock prices. $\bar{\Psi}$ and $\bar{\lambda}$ are coefficient in the inventory cost function and have a transitory impact on stock prices. All measures in both panels are cross-sectional averages.

denoted by L^*.[9]

We investigate commonality in liquidity for $k = 1, 2, 3$. This means that all three first systematic liquidity factors are included in the $k = 3$ case, implying that R^2 is growing with k. As we are interested in the time-variation of commonality, we run this regression using a moving estimation window. To capture potential short-term trends in the underlying covariance structure, we use a three year window for our regressions, the same as we use for the moving window specification of systematic liquidity. We have experimented with different window sizes, finding that a two year window yields a less stable systematic liquidity measure (with lower commonality), whereas slightly longer horizons yield results similar to the three year window. A larger estimation window has been considered for the commonality regressions as well, but that naturally makes the differences between moving and expanding windows less significant. We require more than 12 stock liquidity observations in the specified time frame for a firm to be considered in this analysis. To investigate how commonality depends on the estimation of the underlying liquidity factor estimation, the procedure is repeated for our different settings of τ.[10]

Before analyzing these differences between systematic liquidity factors, it is useful to ask whether a high degree of commonality is good. Does high commonality imply that the systematic liquidity measure is more accurate? In the regression commonality test we use systematic liquidity factors that have been estimated using different window specifications τ, but we evaluate them over the same time span. PCA is based on extraction of eigenvalues and eigenvectors of the covariance matrix. Estimated covariance matrix differences between short and long estimation windows can be due to either estimation error (sampling error) or actual variation over time in the underlying covariance matrix. If the latter is true, the systematic liquidity factor utilizing a short estimation window should be better suited to capture commonality. If the variation is due to noise; however, a long estimation horizon should yield higher accuracy. Based on this discussion, we argue that high commonality is desirable; and more specifically, that high commonality is a sign of accuracy in the systematic liquidity measurement.

---

[9]An alternative way of measuring commonality when using PCA is to study the magnitude of the covariance matrix eigenvalues. The variation explained by each component vector is equal to the ratio of its corresponding eigenvalue to the sum of all eigenvalues. This method for liquidity commonality measurement was introduced by Hasbrouck and Seppi (2001) and has also been applied by e.g., Chen (2007), Kempf and Mayston (2008) and Kopp et al. (2008). We do not apply it here, as we study shocks of PCA, not PCA directly.

[10]E.g., Kamara et al. (2008) have looked at commonality in terms of the coefficient $\gamma_{2,l,\tau}$. When using dynamic PCA, we have a sign indeterminacy in the factors, making that regression coefficient unstable. It is also difficult to interpret the absolute size of $\gamma_{2,l,\tau}$. Hence, in our commonality test, we focus on R^2, which is unaffected by these problems. The problem of sign indeterminacy across liquidity measures was discussed for the static window PCA case by Korajczyk and Sadka (2008). In the dynamic setting the same problem is added in the time dimension.

	τ		36			t			T		
	k	1	2	3	1	2	3	1	2	3	$\bar{L}$
L_t^1	Qspread	0.11	0.19	0.26	0.16	0.22	0.26	0.15	0.21	0.27	0.23
L_t^2	Espread	0.12	0.21	0.27	0.20	0.24	0.27	0.20	0.25	0.28	0.29
L_t^3	Turnover	0.06	0.13	0.20	0.10	0.15	0.20	0.08	0.17	0.21	0.20
L_t^4	Amihud	0.08	0.16	0.21	0.15	0.21	0.25	0.16	0.22	0.25	0.14
L_t^5	Ψ	0.05	0.10	0.14	0.04	0.08	0.12	0.04	0.08	0.12	0.11
L_t^6	λ	0.05	0.10	0.15	0.04	0.08	0.12	0.04	0.08	0.12	0.07
L_t^7	$\bar{\Psi}$	0.11	0.20	0.26	0.14	0.18	0.22	0.13	0.16	0.20	0.26
L_t^8	$\bar{\lambda}$	0.05	0.10	0.15	0.04	0.08	0.11	0.04	0.08	0.11	0.07
r_t^m	Returns	0.11	0.18	0.24	0.10	0.17	0.23	0.10	0.16	0.23	0.22

Table 3: Average degree of commonality in liquidity

The table shows the degree of commonality for eight different measures of liquidity. Degree of commonality is defined as R^2 of regression run on Equation 12 run with a moving estimation window of 36 monthly observations. Numbers in the table are averaged across stocks and time periods. The estimation window size used for deriving systematic liquidity factors with PCA is given by τ, and the number of systematic liquidity factors considered is given by k. The same procedure is run using cross-sectional equal-weighted average liquidity, denoted $\bar{L}$. Ψ and λ are coefficients in the adverse selection cost function, with permanent influence on stock prices. $\bar{\Psi}$ and $\bar{\lambda}$ are coefficient in the inventory cost function and have a transitory impact on stock prices. See Equation 11 for details. The routine is repeated for returns, measuring to what degree common factors in returns can explain variation in individual stock returns.

The commonality regression as specified above allows us to investigate time dynamics of commonality for different liquidity measures and different measurement methods. We look at the dynamics over time below, but first we study overall differences between systematic liquidity measures by looking at commonality averaged both cross-sectionally and over time. These results are presented in Table 3.

We include the return measure in the table to get an intuition of what different magnitudes of commonality imply. It is well known that the market as an aggregate to a large extent drives returns in individual stocks. Using exactly the same methodology as for liquidity measures, we find that the degree of commonality in returns at $k = 3$ is 20 to 25% (i.e., variation in common factors of stock prices can explain more than 20% of the variation in individual stock price variations). For the bid-ask spread measures, turnover, the Amihud measure, and the fixed (transitory) inventory cost, we find commonality of about the same degree as for returns. The other measures register lower degrees of commonality. Before turning to differences across τ, which is our main interest, it is interesting to compare our results to some previous studies. In the case of returns, a commonality of the same level have been found by both Hasbrouck and Seppi (2001) and Korajczyk and Sadka (2008). Commonality tests of previous literature are not directly comparable to our results, as specification of commonality tests, liquidity variables, as well as data frequencies differ. Regardless, a brief review of their findings is useful. Chordia

et al. (2000) found commonality in quoted and effective spreads. Kamara et al. (2008) identified commonality in the illiquidity measure by Amihud. Korajczyk and Sadka (2008) performed level commonality tests on the same measures as we use. They find high commonality in the bid-ask spread (quoted and effective), the Amihud measure, and $\bar{\Psi}$; and low commonality in Ψ, λ and $\bar{\lambda}$. This corresponds well to our results, though the magnitudes of commonality they find are higher due to their commonality specification. Their finding of weak commonality in *turnover* is not reflected in our results.

The differences across our specifications of PCA with regard to commonality are on average small, but some tendencies can be seen. The degree of commonality found in the first four liquidity measures is slightly higher when using an expanding window rather than a moving window. For the four price impact coefficients the tendency is the opposite. In accordance with our interpretation above, the lower degree of commonality registered for the moving window factors for the first four measures means that the underlying market liquidity is disturbed by sampling error when a short estimation window is applied. If the covariance matrix is time-varying that sampling error should be counter-acted by an ability to capture that variation. For the four price impact coefficients, such short-term effects appear to be present, as the moving window consistently generates higher commonality than the expanding window specification.

We also run the commonality test using the cross-sectional equal-weighted average liquidity ($\bar{L}$), which is another popular proxy for systematic liquidity (Chordia et al. 2000, Kamara et al. 2008). This can be expressed as

$$\Delta L_{i,l}^* = \gamma_{1,l,L} + \gamma_{2,l,L}\Delta\bar{L}_l + u_{i,l,L}, \tag{13}$$

where L replaces τ as subscript to distinguish the notation from Equation 12. Results of this regression are given in the rightmost column of Table 3. Performance of this average liquidity approximation of the systematic liquidity factor varies widely across liquidity measures. Comparing it to PCA with the first three factors considered ($K = 3$), it appears unsuited to capture underlying market liquidity in terms of the Amihud measure and quoted spreads as well as some of the price impact measures. In the case of the Amihud measure, we believe that the low commonality is due to that the mean is driven by some large outlier observations, particularly in the first half of the sample.

The static window PCA ($\tau = T$) has often been used in previous systematic liquidity literature. With this method the full time sample is considered in the estimation, implying that

future information is involved in the systematic factors when we evaluate them in the commonality test. When looking at liquidity in retrospect this is not a problem, but if these factors are applied to explain (or even forecast) asset prices, there will be a forward-looking bias.[11] We note here that the static window PCA factor registers degrees of commonality similar to those of the expanding window. If the methods can be regarded as equivalent in terms of outcome, it would be appropriate to use expanding window PCA rather than static, but before concluding this we have to see if the findings are consistent over time.
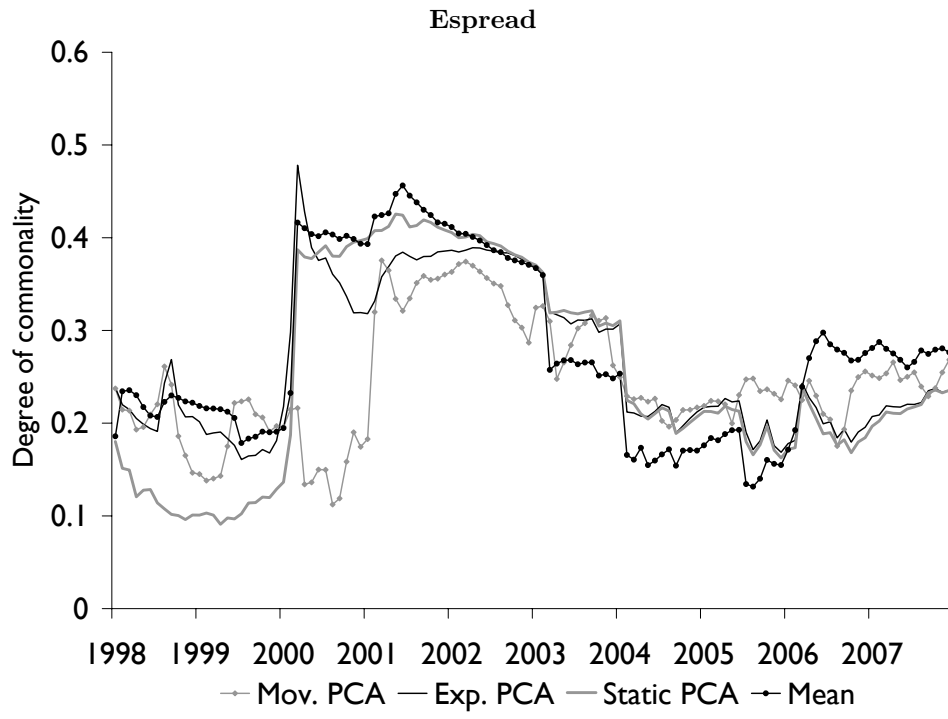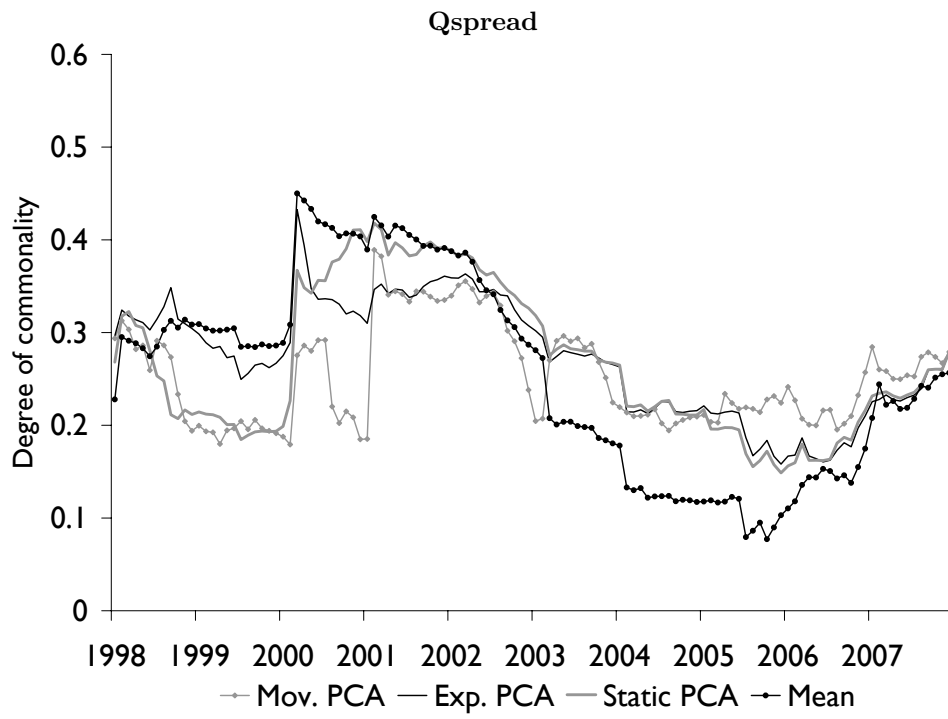
4.1 Dynamics in liquidity commonality

We now turn to the commonality dynamics over time. Figure 2 displays the time dynamics of commonality for different τ (within each panel) and different liquidity measures (one in each panel). The only difference between the curves within each panel is the horizon used when estimating systematic liquidity factors.[12] Commonality depicted is based on the three first systematic liquidity factors and regressions run on 36 observations.
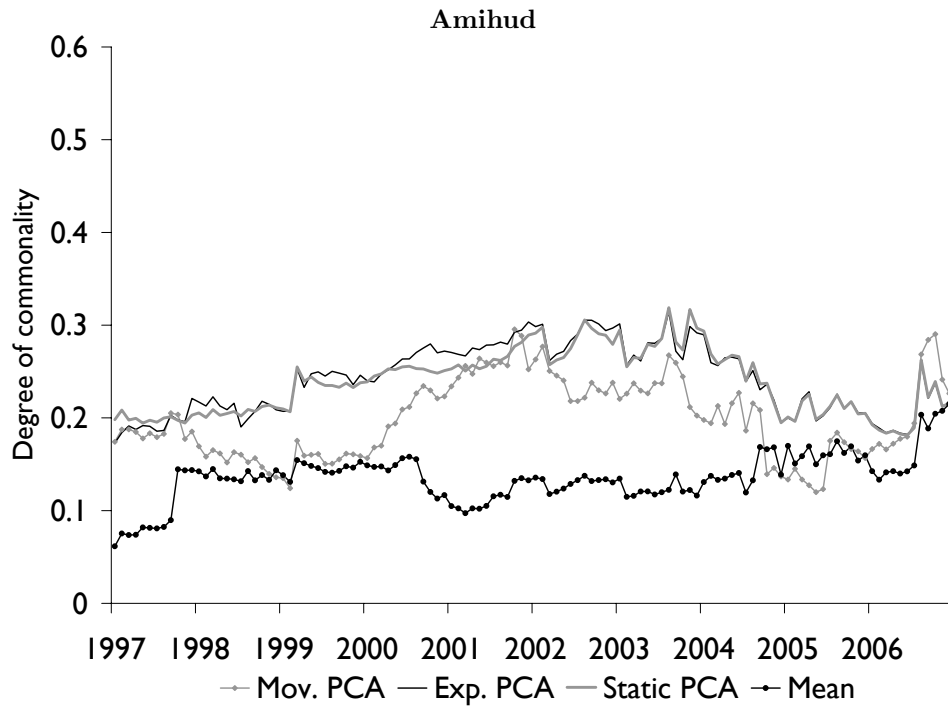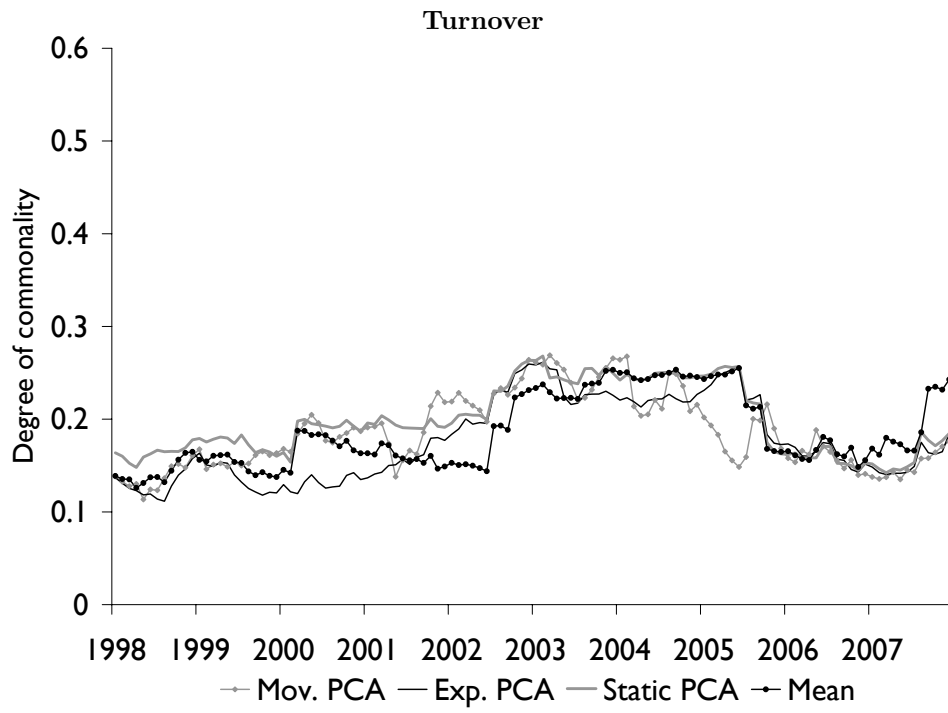
The graphs reveal that commonality varies over time, and for some liquidity measures it varies substantially. The largest variations are seen in the spread measures and the temporary fixed cost measure. For these measures the period chosen for measuring commonality will matter substantially for what answer is retrieved. In many previous studies of commonality measurement data has been chosen by availability making the point of measurement arbitrary. Furthermore, the graphs uncover some cases of substantial differences between measurement methodology. The difference between the most commonly used proxies of systematic liquidity, static window PCA and average liquidity, can be large (differences of > 0.1 in degree of commonality can be seen temporarily for most liquidity measures).

Analyzing the dynamics of the individual liquidity measures. For the Amihud measure we see that the expanding window consistently yields higher commonality than the moving window. This implies that there are long term covariances between asset liquidities that are not revealed when using a short estimation window, and that these are more important than potential short term changes in the covariance matrix. For the Amihud measure, average liquidity appears to be a consistently worse predictor of firm liquidity than PCA proxies. This is notable as $\bar{L}$ was the proxy used in a recent study of commonality in Amihud's illiquidity measure (Kamara

---

[11]Several studies have used the static window PCA for such applications.

[12]This implies that the moving and the expanding window PCA factors are by definition the same in the first time period at $t = 36$ (Dec. 1997, not shown), and that the expanding and static window factors are by definition the same in the last period (Dec. 2007).

Qspread

Espread

Turnover

Amihud

Permanent fixed cost (Ψ)

Mov. PCA — Exp. PCA — Static PCA — Mean

Permanent variable cost (λ)

Mov. PCA — Exp. PCA — Static PCA — Mean

Transitory fixed cost ($\bar{\Psi}$)

Mov. PCA — Exp. PCA — Static PCA — Mean

Transitory variable cost ($\bar{\lambda}$)

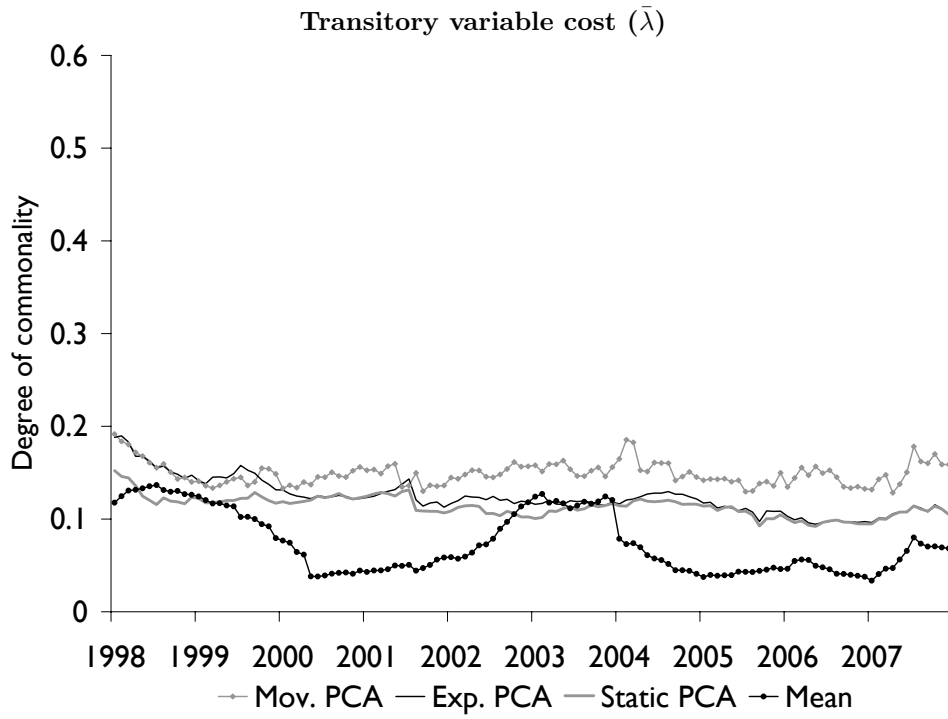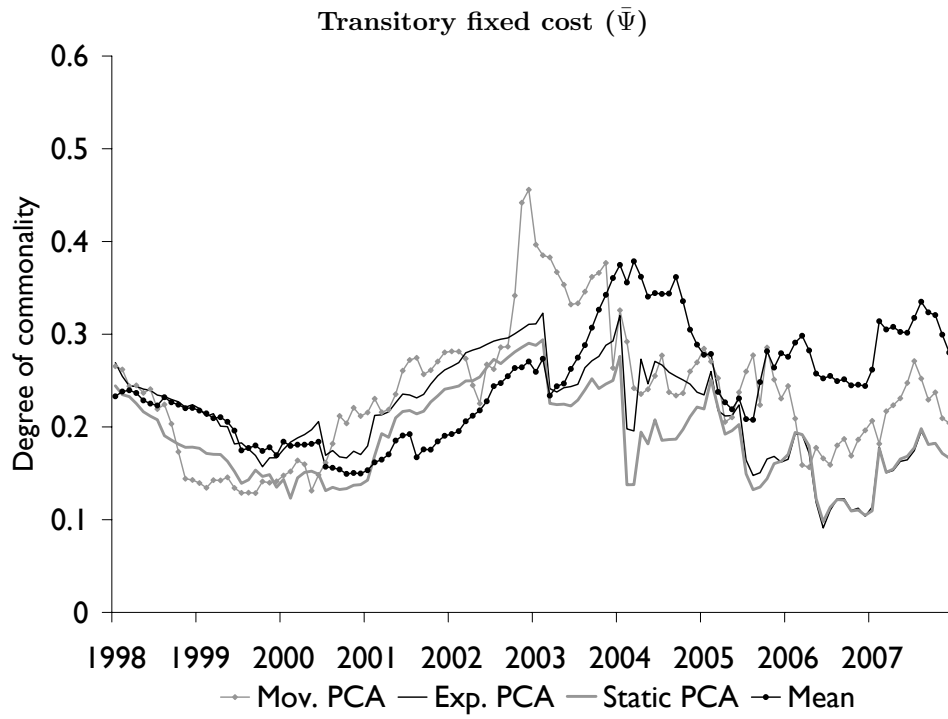Mov. PCA — Exp. PCA — Static PCA — Mean

21
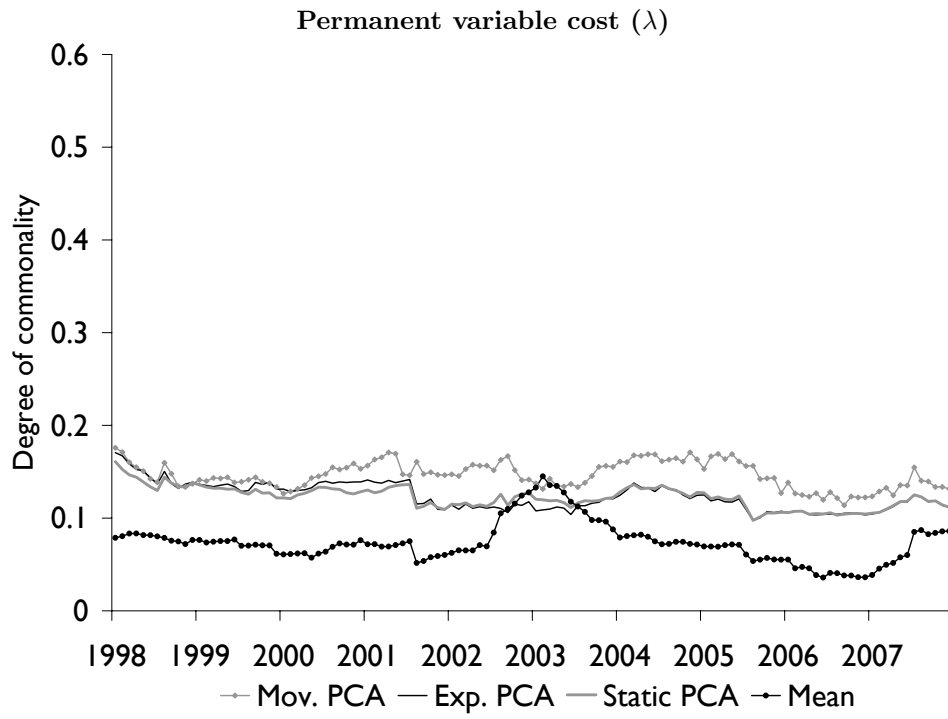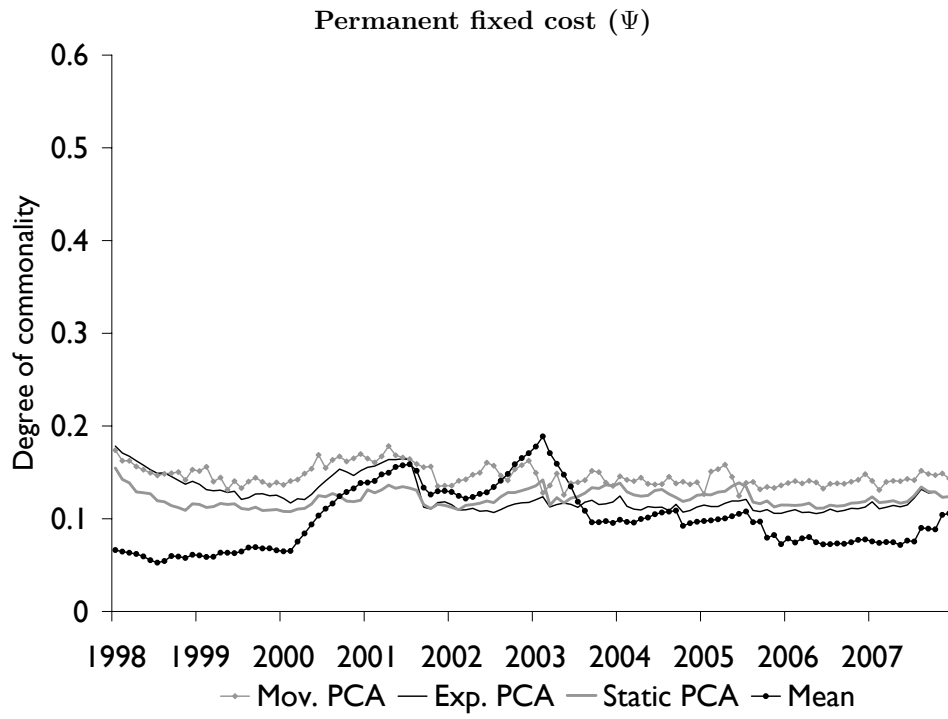
Figure 2: Commonality dynamics

In each panel time dynamics of the degree of regression commonality is graphed for 1998-2007. Degree of commonality displayed is the R^2 value of Equation 12 estimated with $k = 3$. Commonality measures are averages across stocks, estimated for the time period (t-35):t. In each panel, the curves represent different settings of $\tau = [36, t, T]$, denoted `MOV36`, `EXP`, `STATIC` in the legend for each panels. Also, commonality as predicted by the cross-sectional average is given, estimated as in Equation 13. This curve is denoted `MEAN`. Each panel represent commonality in different liquidity measures: A) Amihud; B) Turnover; C) Quoted spread; D) Effective spread.

et al. 2008). As noted above, a reason for the poor performance of the average liquidity is that it can be driven by large outliers. For turnover, the degree of commonality is relatively stable over time and no consistent differences between systematic liquidity measures are seen. The average liquidity proxy can hence be used without loss of precision. For the two measures of the bid-ask spread, on the other hand, we see significant variation over time. The spike in commonality in March 2000 coincides with the burst of the dot-com bubble. It is interesting to see that this causes a sharp increase in commonality of spreads. This can be related to discussion by e.g., Amihud, Mendelson and Wood (1990) about strong liquidity comovement in times of crisis. For the spread measures, no recommendation on which systematic liquidity measurement technique that should be used can be given. There appear to be important short-term trends captured by the moving window PCA during parts of the sample, but during other periods this method is less efficient in capturing trends appearing in the long term sample.

For the price impact liquidity measures, the highest commonality is consistently recorded by the moving window PCA. This implies time-varying covariance matrices – revealing significant short term trends. In general we see low commonality in the price impact measures, except for the transitory fixed cost $\bar{\Psi}$, which is exactly the same finding as in Korajczyk and Sadka (2008). This is the inventory cost that the market maker has to carry. When it is large the market maker has to charge wider spreads in order to avoid losses. Systematic liquidity in terms of $\bar{\Psi}$ is also well proxied by its cross-sectional average, though the moving window specification appears to be more consistent. Sadka (2006) and Glosten and Harris (1988) find that this transitory fixed cost ($\bar{\Psi}$) along with the permanent variable cost (λ) (i.e., the volume related adverse selection cost) are more important than the other two price impact measures, and the magnitudes shown in Section 3 point in the same direction for our study. Our commonality regressions; however, which are more about variation than magnitudes in the underlying liquidity measures, show that the other price impact measures are as efficient as the permanent variable cost in explaining liquidity variation. This is another indication of that these (Ψ and $\bar{\lambda}$) are non-zero.

In general, the expanding window PCA and the static window PCA yield very similar degrees

of commonality. In most cases they have converged around 2003, implying that an eight year window is enough to capture long term trends in systematic liquidity. If such a time frame is available, the forward-looking bias of static window PCA can be avoided by applying the expanding window specification, as was done by Chen (2007), without loss of accuracy.

4.2   Correlation between estimated commonality

Although similar on average, commonality series for one liquidity measure often have quite different time-series properties, i.e., are only weakly correlated, depending on the method used for systematic liquidity estimation. This is especially so for the cross-sectional mean when compared to the PCA-based estimates of systematic liquidity. Table 4 shows correlations for the estimated time-series of commonality for each liquidity measure (i.e., correlations between the series graphed in Figure 2).

	Mov-Exp	Mov-Stat	Mov-Mean	Exp-Stat	Exp-Mean	Stat-Mean
Qspread	0.52	0.72	0.43	0.82	0.91	0.77
Espread	0.43	0.62	0.37	0.81	0.90	0.82
Turnover	0.75	0.80	0.61	0.86	0.81	0.82
Amihud	0.70	0.68	-0.08	0.96	-0.01	-0.09
Ψ	0.68	0.33	0.33	0.31	0.07	0.46
λ	0.38	0.52	0.29	0.90	0.00	0.30
$\bar{\Psi}$	0.48	0.58	0.30	0.91	-0.03	-0.05
$\bar{\lambda}$	0.09	0.15	0.33	0.84	0.45	0.21

Table 4: Correlation between estimated commonality

For each liquidity measure, Pearson correlation coefficients between pairs of commonality series in terms of R^2 of the models described by Equations 12 and 13 are calculated. Each series contains monthly observations of the years 1998–2007. Mov is based on Equation 12 and $\tau = 36$; Exp is based on Equation 12 and $\tau = t$; $Stat$ is based on Equation 12 and $\tau = T$; $Mean$ is based on Equation 13. Ψ and λ are coefficients in the adverse selection cost function, with permanent influence on stock prices. $\bar{\Psi}$ and $\bar{\lambda}$ are coefficient in the inventory cost function and have a transitory impact on stock prices. See Equation 11 for details.

Apparently, the liquidity measures can be divided in two groups. The first group contains quoted spread, effective spread, and turnover. For this group, correlations are relatively high for all estimation methods. The lowest correlation, 0.37, is between the cross-sectional mean and the moving window (for efficient spread) and the highest correlation 0.91 between the expanding and static window (for quoted spread). For this group of measures, time series properties of systematic liquidity are in general similar across estimation methods, implying that average liquidity is a good proxy of systematic liquidity in terms of these measures.

The second group contains the price impact measures: Amihud, permanent market maker costs (adverse selection costs; Ψ and λ), and transitory market maker costs (inventory costs;

$\bar{\Psi}$ and $\bar{\lambda}$). For this group, correlations vary considerably. Correlations between PCA-based estimates are in most cases high, whereas correlations between the cross-sectional average and the PCA-based estimates are much lower. For the Amihud measure estimated by the cross-sectional mean, point estimates of correlation with the static, expanding and moving window are in all cases negative: -0.09, -0.01 and -0.08, respectively. In other words, the cross-sectional average estimate of systematic liquidity produces a commonality with very different time-series properties than the PCA-based estimates of systematic liquidity. Given the fact that the cross-sectional mean yields an average commonality that is much lower than PCA, this suggests that the time-series properties of systematic liquidity are not fully captured by the cross-sectional mean for the Amihud measure. This is interesting as many studies use the cross-sectional average to estimate systematic liquidity in terms of the Amihud measure.

Turning to the price impact measures of Sadka (2006) similar results hold, i.e., commonality estimated using the cross-sectional mean is often in principle uncorrelated with commonality estimated using different specification of PCA. For example, for the fixed inventory costs measure, which produces the highest commonality among Sadka's four measures, correlations between commonality for the cross-sectional average and the PCA-based estimates of systematic liquidity are -0.05, -0.02 and 0.30. This is in spite of average commonality being at the same level for PCA- and mean-based measurement. Again, this indicates that time-series properties of estimated systematic liquidity are very different depending on the method used for estimation.

5   Liquidity factors and stock prices

Above we have argued that commonality is an indicator of liquidity factor measurement accuracy. A different matter, still of great interest, is whether the liquidity factor is useful for explaining stock prices. As it is well-known that individual stock liquidity affects prices (Amihud et al. 2005) it is sensible to believe that if systematic liquidity explains stock liquidity, it should to some extent be priced in the cross-section of stocks. This hypothesis has been tested and confirmed by Pastor and Stambaugh (2003), Acharya and Pedersen (2005), and Korajczyk and Sadka (2008).

In the preceding section we studied correlations between the commonality time series retrieved with different systematic liquidity derivation techniques. High correlations between these series would indicate that they also have similar abilities in explaining stock returns. In our investigation above we found many correlations being low, implying fundamentally different time series properties of the underlying market factors. If that is the case, the different sys-

tematic liquidity estimation techniques may lead to differences in the ability to explain stock returns too. This is the topic of investigation in the current section.

We adopt an extended market model to evaluate the explanatory power of our systematic liquidity measures. We let

$$r_{i,t} = \alpha_{i,l,\tau} + \beta_{i,l,\tau} r_{M,t} + \Phi_{i,l,\tau} \Delta \hat{F}_{1:k,l,\tau} + \epsilon_{i,t,l,\tau}, \tag{14}$$

where $r_{i,t}$ is again the return of stock i in time t, and $r_{M,t}$ is the market return in time t. Return sensitivities to systematic liquidity are given by $\Phi_{i,l,\tau}$ and returns unexplained by the market returns and systematic liquidity factors are denoted $\epsilon_{i,t,l,\tau}$. As indicated by the indexes l and τ, the regressions are repeated for our eight liquidity measures and our three different PCA-based systematic liquidity measurement techniques. We set $k = 3$. In the same fashion as above, we run the regression using a 36 months moving estimation window. From these regressions we are primarily interested in the explanatory power of the systematic liquidity factors with respect to returns. Hence, we record the R^2 of each run of the regression and take the cross-sectional average, which gives us time series of the explanatory power of liquidity. To isolate the effect of liquidity, we run three versions of this model: (I) full regression; (II) setting $\Phi_{i,l,\tau} = 0$; (III) setting $\beta_{i,l,\tau} = 0$. In Table 5 we present averages across time of the full model (1) in Panel A; the excess explanatory power of the liquidity factor [(1)-(2)] in Panel B; and finally the explanatory power of the liquidity factor alone (3) in Panel C.[13] Also, as with the commonality regressions we consider the alternative relation

$$r_{i,t} = \alpha_{i,l,L} + \beta_{i,l,L} r_{M,t} + \phi_{i,l} \Delta \bar{L}_l + \epsilon_{i,t,l,L}, \tag{15}$$

simply replacing the systematic liquidity factor with the cross-sectional mean liquidity and changing the subscripts to L.

The results in Table 5 show that the systematic liquidity factor estimated using PCA can have a substantial explanatory power for stock returns. When studied in isolation (Panel C), it ranges from 10.9% up to 16.4%. When accounting for the market return factor (Panel B), the excess explanatory power of systematic liquidity ranges from 7.9% to 10.9%. Cross-sectional mean liquidity has a substantially lower explanatory power regardless of what liquidity measure we look at.

---

[13]As mentioned in conjunction to the commonality tests, we are unable to interpret the size and the sign of $\Phi_{i,l,\tau}$ due to the sign indeterminacy of PCA factors. This is the reason that we focus on R^2 values.

Panel A: Average R^2 market return and liquidity factor

l		$\tau = 36$	$\tau = t$	$\tau = T$	L
1	Qspread	0.286	0.286	0.284	0.235
2	Espread	0.287	0.287	0.289	0.235
3	Turnover	0.291	0.290	0.293	0.234
4	Amihud	0.292	0.296	0.293	0.236
5	Ψ	0.290	0.290	0.289	0.239
6	λ	0.283	0.290	0.292	0.238
7	$\bar{\Psi}$	0.313	0.306	0.306	0.241
8	$\bar{\lambda}$	0.285	0.291	0.284	0.234

Panel B: Average R^2 liquidity factor in excess of market return

l		$\tau = 36$	$\tau = t$	$\tau = T$	L
1	Qspread	0.081	0.082	0.080	0.031
2	Espread	0.083	0.083	0.085	0.031
3	Turnover	0.087	0.086	0.088	0.030
4	Amihud	0.087	0.091	0.089	0.032
5	Ψ	0.086	0.085	0.085	0.035
6	λ	0.079	0.086	0.088	0.034
7	$\bar{\Psi}$	0.109	0.102	0.102	0.037
8	$\bar{\lambda}$	0.081	0.087	0.080	0.029

Panel C: Average R^2 of liquidity factor alone

l		$\tau = 36$	$\tau = t$	$\tau = T$	$\bar{L}$
1	Qspread	0.120	0.129	0.127	0.054
2	Espread	0.128	0.128	0.124	0.059
3	Turnover	0.123	0.109	0.119	0.036
4	Amihud	0.112	0.117	0.114	0.038
5	Ψ	0.131	0.129	0.120	0.044
6	λ	0.111	0.122	0.125	0.063
7	$\bar{\Psi}$	0.164	0.143	0.135	0.107
8	$\bar{\lambda}$	0.127	0.140	0.154	0.038

Table 5: Systematic liquidity factor ability to explain returns

Panel A shows, for each liquidity measure l, R^2 averaged across stocks and across time for the model described in Equation 14 and 15; Panel B shows R^2 averaged across stocks and across time for the model described in Equation 14 and 15 minus the corresponding number from the same model but with $\Phi = 0$; Panel C shows R^2 averaged across stocks and across time for the model described in Equation 14 and 15 with $\beta = 0$. τ indicates how the estimation window is specified in the systematic liquidity derivation in Equation 12. Ψ and λ are coefficients in the adverse selection cost function, with permanent influence on stock prices. $\bar{\Psi}$ and $\bar{\lambda}$ are coefficient in the inventory cost function and have a transitory impact on stock prices. See Equation 11 for details.

The results for the static window PCA systematic liquidity measure are, as was found for commonality, similar to those of the expanding window PCA. The difference between moving window and expanding window estimation for systematic liquidity is in general small, but some cases where they differ can be noted. For the Amihud measure, the expanding window is slightly better than the moving, which is in accordance with our findings on commonality in Section 4. Among the price impact liquidity measures, an expanding estimation window specification appears to be better for variable costs, whereas a moving is better for fixed costs.

The spread measures yielded highest commonality of all measures, but do not stand out as strong measures in terms of explanatory variables for returns. Hence, high commonality does not necessarily imply high explanatory power of returns. The measure that recorded the second highest commonality, transitory fixed costs ($\bar{\Psi}$), has much higher pricing power than all the other liquidity measures (10.9% in B and 16.4% in C, when using moving window specification). Sadka (2006) argued that out of the four $\bar{\Psi}$ and λ should be most important for pricing, but our results show that they are all on par with the other liquidity measures in this regard. Again, the covariation rather than the magnitude of the underlying liquidity measure is what matters for the investigation.

5.1 Time series properties of systematic liquidity pricing power

In our analysis above we established that the two different dynamic estimation window PCA specifications on average resulted in systematic liquidity factors with rather similar explanatory power on returns. As we did for commonality time series in the previous section, we now look at correlation between time series on ability to explain returns. These correlations are presented in Table 6. In Panel A, correlations between R^2 in excess of the market returns are given; and in Panel B correlations between R^2 series of models estimated using liquidity factors as the only explanatory variable are presented.

Similar to the case of commonality, the correlation between the moving window and expanding window PCA specifications is in general high and positive. We interpret this as a sign that the two methods explain the same pricing patterns. For Amihud, Turnover and permanent variable costs (λ) that correlation is lower, which can be interpreted as meaning that the two methods capture different aspects of return variation. That implies that there are both short and long term trends that contribute to the return data generating process, captured by moving and expanding estimation windows respectively. When comparing PCA specifications to the mean liquidity, correlations are in many cases very low or even negative. This shows that mean

Panel A: Systematic liquidity pricing effect in excess of market returns

l		Mov-Exp	Mov-Stat	Mov-Mean	Exp-Stat	Exp-Mean	Stat-Mean
1	Qspread	0.67	0.48	0.20	0.73	0.42	0.55
2	Espread	0.57	0.48	0.19	0.67	0.32	0.35
3	Turnover	0.29	-0.12	-0.31	-0.31	0.10	0.41
4	Amihud	-0.19	-0.30	0.04	0.77	0.57	0.63
5	Ψ	0.76	0.73	0.56	0.85	0.39	0.58
6	λ	0.31	0.34	0.06	0.80	0.80	0.80
7	$\bar{\Psi}$	0.85	0.89	0.85	0.95	0.93	0.93
8	$\bar{\lambda}$	0.55	0.54	0.50	0.82	0.77	0.55

Panel B: Systematic liquidity pricing effect alone

l		Mov-Exp	Mov-Stat	Mov-Mean	Exp-Stat	Exp-Mean	Stat-Mean
1	Qspread	0.86	0.86	0.34	0.95	0.50	0.54
2	Espread	0.53	0.46	0.49	0.20	0.71	-0.20
3	Turnover	0.24	0.56	-0.08	-0.06	0.12	0.03
4	Amihud	0.44	0.24	0.27	0.75	0.38	0.17
5	Ψ	0.88	0.81	-0.18	0.89	-0.21	0.00
6	λ	0.41	0.41	0.48	0.96	0.48	0.48
7	$\bar{\Psi}$	0.14	0.48	0.34	0.83	0.86	0.78
8	$\bar{\lambda}$	0.74	0.77	-0.09	0.89	0.01	-0.18

Table 6: Systematic liquidity factor pricing power: correlation between estimation techniques

For each liquidity measure l, Pearson correlation coefficients between pairs of R^2 series of the models described by Equations 14 and 15 are calculated. Panel A shows correlations between R^2 averaged across stocks for the model described in Equation 14 and 15 minus the corresponding number from the same model but with $\Phi = 0$; Panel B shows correlations between R^2 averaged across stocks for the model described in Equation 14 and 15 with $\beta = 0$. Each series contains monthly observations of the years 1998–2007. Mov is based on Equation 14 and $\tau = 36$; Exp is based on Equation 14 and $\tau = t$; $Stat$ is based on Equation 14 and $\tau = T$; $Mean$ is based on Equation 15. Ψ and λ are coefficients in the adverse selection cost function, with permanent influence on stock prices. $\bar{\Psi}$ and $\bar{\lambda}$ are coefficient in the inventory cost function and have a transitory impact on stock prices. See Equation 11 for details.

liquidity explains other pricing patterns than the PCA measures.

The correlation between the static window PCA and expanding window PCA in terms of pricing power is again found to be high. In the previous section we found that the two methods also achieved equivalent levels of explanatory power. Hence, we conclude that the expanding window PCA is a good substitute for the static window PCA, which has the important drawback that it uses future information for its estimation.

6   Concluding remarks

We improve the measurement of systematic liquidity, a factor describing market variation of liquidity that has been shown to be an important risk factor in asset pricing. Traditionally systematic liquidity is either proxied by the market average liquidity or derived using PCA. By running PCA it is implicitly assumed that the covariance matrix of cross-sectional liquidity is constant. As this assumption may be unrealistic, we introduce PCA with dynamic estimation window, which re-estimates the covariance matrix for each time period. This can be done either with a moving or an expanding estimation window. Which one is appropriate depends on the properties of the covariance matrix over time. We assess these two specifications of dynamic estimation window PCA together with the two traditional systematic liquidity estimation techniques.

The evaluation of the four estimation techniques is run in terms of (A) ability to explain cross-sectional stock liquidity and (B) ability to explain cross-sectional stock returns. The first criterion is the explicit purpose of systematic liquidity factors, whereas the second is important for its application in asset pricing. For both cases, we find support for using the dynamic estimation window specification of PCA.

Which measurement technique that is most appropriate for systematic liquidity is dependent on what liquidity measure is considered. For measuring illiquidity as defined by Amihud (2002), a long time frame is beneficial both in terms of liquidity and return variation, so an expanding window PCA estimation is appropriate. For turnover and bid-ask spread measures of liquidity, no single measurement technique dominate the other when measuring commonality in liquidity. The simplest possible proxy, such as a cross-sectional average can therefore be used. For explaining stock returns; however, all the PCA-based measures are substantially better. For price impact coefficients as estimated by Sadka (2006), the highest commonality is retrieved when using a moving window to estimate systematic liquidity, implying a time-varying covari-

ance matrix. The most important factor from Sadka's price impact regression, both in terms of liquidity and return variability, is the transitory fixed cost variable, which is associated with the inventory cost that a liquidity provider carries. Systematic liquidity in this variable is best estimated using a moving window PCA. Using that methodology, 26% of the cross-sectional variation in liquidity can be explained, and 16% of the cross-sectional variation in asset returns. This is higher than what we find for all the other liquidity measures tested.

Our investigations also show that the expanding window PCA yield results equivalent to those of the traditional static window PCA, both for explaining asset liquidity and for explaining asset returns. Hence, where possible, expanding window PCA should replace static window PCA, as the latter utilizes future information for its estimation.

References

Acharya, V. and Pedersen, L.: 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* **77**(2), 375–410.

Aitken, M. and Comerton-Forde, C.: 2003, How should liquidity be measured?, *Pacific-Basin Finance Journal* **11**(1), 45–59.

Akerlof, G.: 1970, The market for lemons: Quality uncertainty and the market mechanism, *Quarterly journal of economics* **84**(3), 488–500.

Amihud, Y.: 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* **5**(1), 31–56.

Amihud, Y. and Mendelson, H.: 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* **17**(2), 223–250.

Amihud, Y. and Mendelson, H.: 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, *The Journal of Finance* **44**(2), 479–486.

Amihud, Y., Mendelson, H. and Wood, R.: 1990, Liquidity and the 1987 stock market crash, *Journal of Portfolio Management* (SPRING 1990).

Amihud, Y., Pedersen, L. and Mendelson, H.: 2005, *Liquidity and Asset Prices*, Now Publishers Inc.

Bagehot, W.: 1971, The only game in town, *Financial Analysts Journal* **27**(2), 12–14.

Bauer, W.: 2004, Commonality in liquidity in pure order-driven markets, *Technical report*, University of Zrich Working Paper.

Beltran-Lopez, H., Giot, P. and Grammig, J.: 2006, Commonalities in the order book, *Universite Catholique de Louvain Working Paper* .

Brennan, M. and Subrahmanyam, A.: 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics* **41**, 441–464.

Brownlees, C. and Gallo, G.: 2006, Financial econometric analysis at ultra-high frequency: Data handling concerns, *Computational Statistics and Data Analysis* **51**(4), 2232–2245.

Chen, J.: 2007, *Understanding Equity Returns*, PhD thesis, Columbia University.

Chollete, L., Næs, R. and Skjeltorp, J.: 2007, What captures liquidity risk? a comparison of trade and order based liquidity factors, *Norges Bank Working Paper* .

Chollete, L., Næs, R. and Skjeltorp, J.: 2008, The risk components of liquidity, *Norges Bank Working Paper* .

Chordia, T., Roll, R. and Subrahmanyam, A.: 2000, Commonality in liquidity, *Journal of Financial Economics* **56**(1), 3–28.

Chordia, T., Roll, R. and Subrahmanyam, A.: 2001, Market liquidity and trading activity, *The Journal of Finance* **56**(2), 501–530.

Connor, G. and Korajczyk, R.: 1986, Estimating pervasive economic factors with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* **15**, 323–346.

Engle, R. and Patton, A.: 2004, Impacts of trades in an error-correction model of quote prices, *Journal of Financial Markets* **7**(1), 1–25.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2000, The generalized dynamic-factor model: Identification and estimation, *Review of Economics and Statistics* **82**(4), 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2005, The generalized dynamic factor model, *Journal of the American Statistical Association* **100**(471), 830–840.

Foster, F. and Viswanathan, S.: 1993, Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models, *Journal of Finance* pp. 187–211.

Gilmore, C., Lucey, B. and McManus, G.: 2008, The dynamics of central european equity market comovements, *Quarterly Review of Economics and Finance* **48**(3), 605–622.

Glosten, L. and Harris, L.: 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* **21**(1), 123–142.

Hallin, M., Mathias, C., Pirotte, H. and Veredas, D.: 2009, Market liquidity as dynamic factors, *ECORE Discussion Paper 2009-004* .

Hasbrouck, J.: 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* **22**(2), 229–52.

Hasbrouck, J.: 1991, Measuring the information content of stock trades, *Journal of Finance* pp. 179–207.

Hasbrouck, J. and Seppi, D.: 2001, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* **59**(3), 383–411.

Heij, C., van Dijk, D. and Groenen, P.: 2008, Macroeconomic forecasting with matched principal components, *International Journal of Forecasting* **24**(1), 87–100.

Henker, T. and Wang, J.: 2006, On the importance of timing specifications in market microstructure research, *Journal of Financial Markets* **9**(2), 162–179.

Huang, R. and Stoll, H.: 1996, Dealer versus auction markets: A paired comparison of execution costs on nasdaq and the nyse, *Journal of Financial Economics* **41**(3), 313–357.

Huberman, G. and Halka, D.: 2001, Systematic liquidity, *Journal of Financial Research* **24**(2), 161–178.

Jolliffe, I.: 2002, *Principal component analysis*, Springer New York.

Kamara, A., Lou, X. and Sadka, R.: 2008, The divergence of liquidity commonality in the cross-section, *Journal of Financial Economics* **89**, 444–466.

Kempf, A. and Mayston, D.: 2008, Liquidity commonality beyond best prices, *Journal of Financial Research* **31**(1), 25–40.

Kopp, E., Hütl, M., Loistl, O. and Prix, J.: 2008, Systematic liquidity in the xetra order book: A multi-stage approach, *Vienna Univeristy of Economics and Business Administration, Working paper* .

Korajczyk, R. and Sadka, R.: 2008, Pricing the commonality across alternative measures of liquidity, *Journal of Financial Economics* **87**, 45–72.

Kyle, A.: 1985, Continuous auctions and insider trading, *Econometrica* **53**(6), 1315–1336.

Lee, C. and Ready, M.: 1991, Inferring trade direction from intraday data, *Journal of Finance* **46**(2), 733–746.

Li, W., Yue, H., Valle-Cervantes, S. and Qin, S.: 2000, Recursive pca for adaptive process monitoring, *Journal of Process Control* **10**(5), 471–486.

Mehrotra, D.: 1995, Robust elementwise estimation of a dispersion matrix, *Biometrics* pp. 1344–1351.

Pastor, L. and Stambaugh, R.: 2003, Liquidity risk and stock returns, *Journal of Political Economy* **11**, 642–685.

Petersen, M. and Fialkowski, D.: 1994, Posted versus effective spreads: Good prices or bad quotes, *Journal of Financial Economics* **35**(3), 269–292.

Sadka, R.: 2006, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics* **80**(2), 309–349.

Stoll, H.: 1978, The pricing of security dealer services: An empirical study of nasdaq stocks, *The Journal of Finance* **33**(4), 1153–1172.

Vergote, O.: 2005, How to match trades and quotes for nyse stocks?, *Catholic University of Leuven (KUL) - Department of Economics WP* .

Volosovych, V.: 2005, Measuring financial market integration over the long run: Is there a u-shape?, *University of Houston Working Paper* .

Wang, S. and Xia, Y.: 2002, *Portfolio selection and asset pricing*, Springer New York.

A   Appendix: Filtering and matching of trades and quotes

For quotes and trades we follow most of the filtering conventions applied in previous literature. Specifically, trades that are recorded to be cancelled, corrected, out of sequence, or have special conditions attached to them are excluded.[14] Trades recorded outside the interval 9.30am - 4.05pm are also excluded. Quotes are filtered out if the spread is negative or larger than five dollars or larger than 25% when the spread midpoint is less than twenty dollars.

For the latter years it is common to find numerous quote observations for the same stock within one second. As TAQ does not distinguish time units smaller than one second, all quotes can not be considered. There is not much guidance in the literature on how to deal with this problem. Brownlees and Gallo (2006) recommend using the median quote but also say that the first or last observation could be used (the latter is the choice of e.g., Korajczyk and Sadka (2008)). There does not appear to be any theoretical foundation for either, so we make an arbitrary choice and use the last observation each second. Trades in the same stock that are recorded within one second are aggregated to one trade with the volume equalling the sum of all trades in that second. For prices, Brownlees and Gallo (2006) again recommend using the median, whereas Engle and Patton (2004) use the first observation. We choose to calculate a volume-weighted average price for aggregation of trades.

Another filtering is needed to avoid outliers. For trades, we follow the recommendation of Brownlees and Gallo (2006) on how to deal with trade price inconsistencies. For each day, trades that deviate more than three standard deviations from the delta-trimmed mean are deleted.[15] Finally, we also exclude outliers in the data on shares outstanding (used in the turnover measure). These series feature jumps back and forth between set levels in a way that is not reflected in firm histories. To avoid these reversals, we exclude SO observation during time periods following a change in shares outstanding of more than 25% that is reversed within 10 months. This is an

---

[14]In TAQ language, only trades with $CORR \leq 1$ and $COND$ values of "*", "E" or blank, and $0.5 < PRICE < 15000$ are retained. This excludes trades with the conditions: Cash-only basis, Bunched, Cash Sale, Next Day Settlement Only, Bunched Sold, Rule 127, Rule 155, Sold Last, Next Day, Opened Last, Prior Reference Price, Seller, Split Trade, Pre-/Post-Market Trades, Average Price Trades, Opened After Trading Halt, Sold Sale, Crossing Section.

[15]The specific rule for exclusion is the following (see Brownlees and Gallo (2006), p.2237): $(|p_i - \bar{p}_i(k)|) > 3s_i(k) + \gamma$, where $\bar{p}_i(k)$ is the δ-trimmed mean price, and $s_i(k)$ is the standard deviation in the neighbourhood of k observations. A granularity parameter γ is added to deal with the cases where there is no price variability in the neighbourhood. Of all trades in our sample, 0.08% were affected by this filter. Brownlees and Gallo suggest that k should be chosen in accordance with the trading activity in the stock. They choose $k = 60$ for their sample, which corresponds to a trading intensity of 270*k per day. We use this factor (270) to calculate k for each day for each stock, setting 10 as a minimum value for k. For days with less than k observations, all observations are used. The granularity coefficient should be chosen as a multiple of the minimum price variation (MPV). Brownlees and Gallo choose $\gamma = 0.02$, which is a multiple of 2 of decimal (MPV) used in their sample. As MPV decreases substantially over the years of our dataset, we calculate MPV for each stock in each month based on the first 1000 non-aggregated trades. δ is set to 10%.

arbitrary choice based on our overview of the data set – no guidance on this have been found in the literature.

To measure effective spread, we need to approximate what the bid-ask spread was just before the trade was executed. Most of the liquidity literature has used the findings of Lee and Ready (1991) as guidance on how to match trades and quotes for this approximation. Their recommendation is to use the latest quote recorded at least five seconds before the trade. This is based on the delay in trade reporting prevailing at the exchanges almost 20 years ago when trades were still reported manually. According to Henker and Wang (2006) the reporting process underwent automatization during the period 1994–2001 leading to much shorter reporting time. Henker and Wang (2006) and Vergote (2005) have investigated what time lag is appropriate today, finding one and two seconds respectively. Both studies are limited to NYSE, Nasdaq matching is more involved. In this study we follow the one second rule recommended by Henker and Wang (2006), as that study was performed on the S&P500 index stocks with primary listing on NYSE (Vergote uses a much smaller set of stocks). Another outlier check is performed after the matching: When the trade price is more than 10% outside its prevailing spread, the trade is not used in price impact regressions, but the data is retained for measures where matching is not needed.