All Things both Great and Small: Transaction Cost Persistence in Corporate Bonds *

Redouane Elkamhi†
University of Toronto

Jonatan Groba‡
Lancaster University Management School

Ingmar Nolte§
Lancaster University Management School

January 13, 2017

PRELIMINARY
COMMENTS WELCOME

Abstract

This paper presents a framework to study deviations from the efficient yield curve (yield cost) for trades with different sizes. The yield cost depends on small, medium, and large trade factors with intuitive interpretations that are time varying and persistent. The resulting model is capable of accounting for the different levels of persistence and uncertainty in costs faced by different transaction sizes. We find that the transaction costs deviate rates significantly from the efficient term structure, especially for small trades. Yields of small trades deviate 19 bps on average, however during illiquid periods the short-term yields deviate 22 bps in the case of large trades and 90 bps in the case of small trades. This is because retailers suffer from more persistent transaction costs than institutional investors.

Keywords: Bond price, transaction costs, trade size, big data.
JEL Classification: C10, G12

*The authors would like to thank Torben Andersen, Sudipto Dasgupta, Ralph de Haas, Vasso Ioannidou, Albert J. Menkveld, Stephen Taylor, Ayako Yasuda, Josef Zechner, and seminar participants at Lancaster University Management School.

†Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON, M5S 3E6, Canada. E-mail: Redouane.Elkamhi@rotman.utoronto.ca

‡Lancaster University Management School, Department of Accounting and Finance, Lancaster LA1 4YX, United Kingdom. E-mail: j.groba@lancaster.ac.uk

§Lancaster University Management School, Department of Accounting and Finance, Lancaster LA1 4YX, United Kingdom. E-mail: i.nolte@lancaster.ac.uk

1 Introduction

In a market without frictions, corporate bonds compensate investors for the expected default losses. It is well known that corporate bond market participants face important transaction costs that get incorporated into credit spreads. However, current asset pricing literature does not provide a framework for measuring transaction cost risk. This paper addresses this issue by designing a framework that easily allows accounting for transaction costs that depend on the size of the transaction. Using data on corporate bond transactions we show that three transaction cost factors (small, medium, and large trades) are relevant time-varying and persistent risks in corporate bonds.

The outstanding volume of $8 trillion in US corporate bond securities allow participants to earn a higher yield than otherwise equivalent Treasury bonds. The average trading volume in the US corporate market has significantly increased since the financial crisis under a period of ultra-low Treasury rates. The investors traded on average $27 billion in 2015 in comparison to the $ 18 billion in the early 2000s.[1]

This high increase in trading activity after the financial crisis indicates that there has been a change in transaction costs across all the market. It has also been argued that the new Basel III framework and the Volcker rule may increase immediacy costs in the bond market as banks are taking less risk. Understanding the dynamics of the corporate bond market transaction costs is important for liquidity risk assessment and monitoring. It is at the center of regulators' agenda to enhance liquidity risk management of financial institutions.

Broker-dealers are required by FINRA to engage in liquidity risk management practices to enhance investor protection during times of stress where the trading costs may become prohibitively expensive.[2] Recently, in October 2016, the SEC adopted a ruling that requires

[1] Source: SIFMA

[2] See FINRA regulatory notice: https://www.finra.org/industry/notices/10-57 and https://www.finra.org/industry/notices/15-33 .

investment companies and open-ended funds to disclose information about the liquidity risk of holdings.[3] The new rule will allow funds to use "swing pricing" to pass on to the purchasing and redeeming shareholders the trading costs. This protects existing investors from the performance dilution they may suffer as a result of trading costs by other investors.

The new regulatory and electronic trading environment is headed towards a continuous assessment of liquidity in OTC markets. The availability of real-time information in bond markets also brings new challenges for participants to assess default risk effects and transaction costs. The current tools available do not allow for pricing in a high-frequency setting as new information arrives. Here is why we believe that our approach provides new perspective in how to incorporate transaction-specific information into current pricing models. The events of the last decade allow us to study the risk and persistence of transaction costs under such stressed scenarios.

This article presents a new approach to modelling the term structures of defaultable securities taking into account transaction costs associated with the size of a transaction. Thus, our paper closely follows the key findings by Edwards et al. (2007) and Feldhütter (2012) who uncover transaction size as an important determinant of transaction costs. Our approach is different than existing models because it associates a time-varying risk factor to the transaction costs of transactions at different transaction sizes.

Specifically, we assume that current transaction costs reflect the expected instantaneous transaction costs until a bond matures. First, the model assumes that there are Markov factors that generate uncertainty in the transaction costs. Second, we assume that the instantaneous transaction cost is an affine function of the transaction cost factors. Third, we provide a specific parametrization of the linear combination of transaction cost factors that is able to map any transaction side (Buy or Sell) and any transaction size to an instantaneous transaction cost. Fourth, we provide closed form solutions for the expected instantaneous

[3]The SEC rule press release can be found here: https://www.sec.gov/news/pressrelease/2016-215.html .

transaction costs. Finally, we consider a Kalman filter with unequally spaced observations to extract the unobserved factors.

In summary, this paper makes four contributions. First, it describes a new approach for pricing corporate bonds with time-varying transaction costs. Second, we design a transaction cost function in a reduced-form setting that is tractable, and has intuitive interpretations. Our model accounts for three factors to explain the bond returns for different sizes of trades: costs for large trades, the difference between large and small trades, and a medium size factor. Third, we show that the transaction costs are a time-varying source of risk for corporate bonds and that this risk is priced. Fourth, the empirical results provide evidence that the transaction costs of retail trades are more persistent than the transaction costs of institutional trades. This implies that the corporate market is segmented and retailers are more exposed to illiquidity shocks that may persist much longer.

2 Contribution to Existing Literature

2.1 Pricing models with liquidity

Reduced-form models are convenient for the valuation of contingent claims subject to default risk. Early on, Duffie and Singleton (1999) allow for a stochastic liquidity component in the short-rate process. Identifying liquidity risk from corporate bond prices (or yields) is challenging. When pricing corporate bonds using reduced-form models, there are several existing approaches to identify liquidity risk. Within this framework, Driessen (2005) adds a liquidity factor that is based on the yield spread of low-age and high-age portfolios and is an important determinant of credit spreads. Longstaff et al. (2005) use information on credit default swaps to extract the nondefault related components from corporate bond spreads.

They find that this nondefault component ranges from about 20-100 basis points and it is strongly related to the bid-ask spread.

Research on Treasury markets has extensively studied liquidity effects on Treasury markets. A well-known feature of this market is the on-the-run phenomenon. Goyenko et al. (2011) uses quoted bid and ask prices to study the time-series dynamics of liquidity for on-the-run and off-the run securities. They find that liquidity is significantly affected by macroeconomic conditions and show that short-term become more liquid than long-term securities during recessions. Fontaine and Garcia (2012) builds a term structure model for Treasury securities with a liquidity premium that increases with maturity and decays with the age of a bond.

Our approach builds on the existing models that account for a liquidity factor. However, we model explicitly instantaneous transaction costs that consider the side of a transaction (Buy or Sell) and the size of the transaction.

2.2 Transaction costs and transparency

As the market becomes more transparent with all the bond trades being recorded and publicly disseminated, new research has been able to measure the effective transaction costs. Schultz (2001) pioneered the research on corporate bond transactions and documented that trading costs are smaller for larger trades. Later both Bessembinder et al. (2006) and Edwards et al. (2007) showed that trading costs decreased as the market became more transparent and provided further evidence that small trades pay much larger percentage trading costs than large trades.

Despite the strong evidence showing that transaction size is a relevant key factor behind corporate bond transaction costs, asset pricing literature has not being able to accommodate transaction costs into pricing models. There is one exception to this. Feldhütter (2012)

develops a search-based model and argues that the difference between the prices of small trades and the prices of large trades reflect selling pressures. Under this search model investors trade with a dealer at different search intensities, and the transaction size proxies for the search intensity. Small trades represent unsophisticated investors (low search intensity), and large trades are performed by sophisticated investors (high search intensity). In this setting, the prices of large trades react stronger to selling pressure than small trades and liquidity shocks are measured as months when the bid price difference between large trades and small trades decreases.

Our approach is able to take into account the non-linear relationship between transaction size and transaction costs as in Edwards et al. (2007) and further creates uncertainty in this relationship, allowing transaction costs to be time-varying. More interestingly, the model differentiates between transaction sizes and is able to directly measure the underlying stochastic factor that drives the difference between the small and the large transactions that has been identified by Feldhütter (2012) as a measure of selling pressures in the market.

3 Transaction costs

In this section, we describe the pricing model for bonds including transaction costs. P_t is the transaction price and V_t is the unobserved "true value" of the bond at the time t of the trade. Similarly to Edwards et al. (2007), we consider proportional transaction costs. As a novelty feature, we consider customer costs c_t and dealer costs d_t that are time-varying and uncertain. The variables Q and D indicate the direction of a trade. Customers may buy ($Q = 1$) or sell ($Q = -1$) securities from dealers. And the direction of interdealer trades may also be a buy ($D = 1$) or a sell ($D = -1$). In the case of an interdealer trade $Q = 0$, and in the case of a customer trade $D = 0$.

$$P_t = V_t + Q_t P_t c_t + D_t P_t d_t = V_t \left(\frac{1}{1 - Q_t c_t}\right)\left(\frac{1}{1 - D_t d_t}\right) \quad (1)$$

If P_t is the price of a zero coupon bond, its yield to maturity would be given by equation 2. We can further decompose the yield into smaller components. The *yield cost* component would reduce (increase) the bond yield due to the transaction cost of a customer bond purchase (sale).[4]

$$yield = -\frac{log(P)}{T-t} = -\frac{log(V)}{T-t} + \frac{log(1-Qc)}{T-t} + \frac{log(1-Dd)}{T-t} \quad (2)$$

To add flexibility and intuition to the cost function, we create an alternative model based on the Nelson-Siegel-type function. An added advantage of this approach is that the estimated coefficients have an easier interpretation. We want to arrive at a three-factor affine model that matches a yield cost function of the form

$$\begin{aligned} yield\,cost = -\frac{log(1-Qc)}{T-t} &= X_{t,Large} Q_t \left[\frac{1 - e^{-\kappa_1 (T-t)}}{\kappa_1 (T-t)}\right] \\ &+ X_{t,Small} Q_t \left[\frac{1 - e^{-\kappa_2 (T-t)}}{\kappa_2 (T-t)}\right]\left[\frac{1 - e^{(\frac{-S_t}{\varsigma_1})}}{\frac{S_t}{\varsigma_1}}\right] \\ &+ X_{t,Med} Q_t \left[\frac{1 - e^{-\kappa_3 (T-t)}}{\kappa_3 (T-t)}\right]\left[\frac{1 - e^{(\frac{-S_t}{\varsigma_1})}}{\frac{S_t}{\varsigma_1}} - e^{(\frac{-S_t}{\varsigma_1})}\right] \\ &- \frac{A(t,T)}{T-t} \end{aligned} \quad (3)$$

[4]Throughout the paper, we will focus on customer trades as dealer trades can be modelled equivalently. Notice that transaction data does not identify the direction of a trade between two dealers.

The factors X represent the instantaneous transaction cost. $X_{t,Large}$ may be viewed as the cost for large investors. $X_{t,Small}$ represents the cost difference between large and small investors. And $X_{t,Med}$ may be viewed as a factor for medium investors. The parameter ς sets the boundary between the different types of investors. The correlation between the factors captures price pressure spillovers. If the correlation between $X_{t,Large}$ and $X_{t,Small}$ is highly negative, a decrease in transaction cost for large investors leads to an increase in transaction costs for small investors.

The ratio $(1-e^{-\kappa(T-t)})/\kappa$ tends to 0 as we approach maturity and tends to $1/\kappa$ when we get further from maturity. This captures the stylized fact shown in Edwards et al. (2007) that transaction costs decline with the bond maturity, as bonds close to maturity have lower proportional transaction costs than do other bonds due to the imminent redemption payment and the low liquidity risk. The yield cost coefficient $(1-e^{-\kappa(T-t)})/(\kappa(T-t))$ is close to 0 for long maturities and approaches 1 near maturity. The parameter κ moderates the sensitivity of the yield to instantaneous transaction costs as a function of the time to maturity. As we will show later, the parameter κ measures the persistence of the state factor. Factors with high values of κ exhibit a strong tendency to revert towards the mean. Small κ make short-term securities to be more affected by an instantaneous transaction cost X_t that is above its mean than for long-term securities. In other words, long-term securities are "immune" to periods of high instantaneous transaction costs when they are likely to rapidly revert to their mean. And highly persistent state factors with small κ make short- and long-term securities equally affected by the level of the state factor.

The most prominent feature of our model is that it incorporates a coefficient that depends on the size of the transaction S_t and a threshold level ς. Since the coefficient $(1-e^{(-S_t/\varsigma_1)})/(S_t/\varsigma_1)$ does not depend on the time to maturity, it captures the instantaneous cost of a transaction; as it will become clearer later once we delve into the model. It is straightforward to see that this coefficient tends to 1 for a smaller transaction size,

and becomes 0 for large transactions. The speed at which the coefficient moves towards 1 depends on the parameter ς_1. Transactions smaller than ς_1 will lead to a coefficient closer to 1 and a larger impact on the yield cost. Since the first factor $X_{t,Large}$ does not depend on the transaction size, it represents the costs for large transactions. The second factor $X_{t,Small}$ becomes relevant for transactions smaller than ς, and the third factor reaches its maximum in medium transactions above ς. Similarly to a Nelson-Siegel function used for yields, the coefficients will capture different yields for different transactions —as opposed to different maturities. The attractive feature of this model is that it captures non-linearities between transaction size and transaction costs similarly to Edwards et al. (2007) whereas adding interpretation to the non-linearities. Additionally, these coefficients are flexible and accommodate different designs. These coefficients could be replaced by similar non-linear transformations of the transaction size shown in Edwards et al. (2007), with the caveat that the underlying factors lose the intuitive interpretation of our model. Next, we show in detail how to arrive to such yield cost function in an affine arbitrage-free setting.

The yield-adjustment $\frac{A(t,T)}{T-t}$ does not depend on the level of the state factors and it will be defined later.

3.1 The transaction cost model

We define the transaction costs c_t that depend on the size of the trade S. For each size, there is an associated latent discount factor ℓ that decreases or increases the efficient price V depending on whether the intention is to buy ($Q = 1$) or sell ($Q = -1$).

Definition 1. *The time-t proportional transaction costs c_t are given by*

$$1 - Q_t c_t(Q, S, t, T) \quad = \quad E_t^{\mathbb{Q}} \left[exp\left(- \int_t^T \ell_u \, \mathrm{d}u \right) | \ell_t, Q_t, S_t \right] \quad , \tag{4}$$

where $E^{\mathbb{Q}}$ denotes expectation under the risk-neutral measure. An n-factor affine term structure model is obtained under the assumption that the instantaneous transaction cost ℓ is an affine function of a vector of the state variables X_t

$$\ell_t \quad = \quad \rho_0(Q,S) + \rho_1(Q,S)^{\intercal} X_t \quad , \tag{5}$$

where $\rho_0(Q,S) \in \mathbf{R}$ and $\rho_1(Q,S) \in \mathbf{R}^n$ are time-independent functions. These function allow us to map any transaction side Q and transaction size S into an instantaneous transaction cost. The state variable X_t is assumed to be an "affine diffusion" Markov process that solves the stochastic differential equation (SDE) under the risk-neutral dynamics,

$$\mathrm{d}X_t \quad = \quad K^{\mathbb{Q}}[\theta^{\mathbb{Q}} - X_t]\mathrm{d}t + \Sigma D(X_t)\mathrm{d}W_t^{\mathbb{Q}} \tag{6}$$

$W_t^{\mathbb{Q}}$ is an n-dimensional vector containing n independent standard Brownian motions under Q, the vector $\theta^{\mathbb{Q}} \in \mathbf{R}^n$ contains the drift parameters, the mean reversion matrix $K^{\mathbb{Q}} \in \mathbf{R}^{n\times n}$ and the volatility matrix $\Sigma \in \mathbf{R}^{n\times n}$ are matrices which may be non-diagonal and asymmetric. While $D \in \mathbf{R}^{n\times n}$ is a diagonal matrix that depends on the vector $\gamma \in \mathbf{R}^n$ and $\delta \in \mathbf{R}^{n\times n}$ such that the jth diagonal element is given by

$$[D(X_t)]_{jj} \quad = \quad \sqrt{\gamma^j + \delta^j X_t} \quad ,$$

where the upper index j lists the rows and the lower index lists the columns in γ_1^j and δ_i^j. In this framework, Duffie and Kan (1996) provide solutions to exponentially affine functions

of the state variables.

$$E_t^{\mathbb{Q}}\left[exp\left(-\int_t^T \ell_u \,\mathrm{d}u\right)|\Theta\right] \;=\; exp(B(t,T)^{\intercal}X_t + A(t,T)) \tag{7}$$

The principal payment at maturity does not contain any transaction cost associated $1 - Qc_T(Q,S,T,T) = 1$, which implies that $B(T,T) = 0$ and $A(T,T) = 0$. Parametric assumptions are used to get an explicit solution to this PDE. And $B(t,T)$ and $A(t,T)$ are solutions to the system of ordinary differential equations (ODEs or Riccati equations)

$$\begin{aligned}
\frac{\mathrm{d}B(t,T)}{\mathrm{d}t} &= \rho_1 + (K^{\mathbb{Q}})^{\intercal}B(t,T) \\
&\quad -\frac{1}{2}\sum_{j=1}^{n}[\Sigma^{\intercal}B(t,T)B(t,T)^{\intercal}\Sigma]_{jj}(\delta^j)^{\intercal} \quad , \quad B(T,T) = 0 \qquad (8)\\
\frac{\mathrm{d}A(t,T)}{\mathrm{d}t} &= \rho_0 - B(t,T)^{\intercal}K^{\mathbb{Q}}\theta^{\mathbb{Q}} \\
&\quad -\frac{1}{2}\sum_{j=1}^{n}[\Sigma^{\intercal}B(t,T)B(t,T)^{\intercal}\Sigma]_{jj}\gamma^j \quad , \quad A(T,T) = 0 \qquad (9)
\end{aligned}$$

The coefficients $B(t,T)$ and $A(t,T)$ can be easily computed numerically using Runge-Kutta methods. However, they can only be computed in closed-form for specific cases. In this paper we propose a parametrization of the equations 6 and 5 that leads to a yield cost function of the form in equation 3. We model three unobservable state factors $X_{t,Large}$, $X_{t,Small}$, and $X_{t,Med}$ that affect the transaction discount factor ℓ depending on the transaction side Q and the transaction size S. The instantaneous transaction cost ℓ depends linearly on the state variables $X_{t,Large}$, $X_{t,Small}$, and $X_{t,Med}$, and provides non-linear relationship to the transaction costs of large, small, and medium trades.

Then we can solve for the transaction costs under the assumption that they are exponentially affine to the transaction state variables.

Proposition 1. *The instantaneous transaction cost is*

$$\ell_t = X_{t,Large}Q + X_{t,Small}Q\left[\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}}\right] + X_{t,Med}Q\left[\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}} - e^{(\frac{-S}{\varsigma_1})}\right] \tag{10}$$

with coefficients

$$\rho_0(Q,S) = 0 \qquad and \qquad \rho_1^{\intercal}(Q,S) = \left[\begin{array}{ccc} Q & Q\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}} & Q\left(\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}} - e^{(\frac{-S}{\varsigma_1})}\right) \end{array}\right]$$

and the system of SDEs for the state variables under the risk-neutral measure follow a multivariate Ornstein-Uhlenbeck process

$$\left(\begin{array}{c} \mathrm{d}X_{t,Large} \\ \mathrm{d}X_{t,Small} \\ \mathrm{d}X_{t,Med} \end{array}\right) = \left(\begin{array}{ccc} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{array}\right)\left[\left(\begin{array}{c} \theta_1^{\mathbb{Q}} \\ \theta_2^{\mathbb{Q}} \\ \theta_3^{\mathbb{Q}} \end{array}\right) - \left(\begin{array}{c} X_{t,Large} \\ X_{t,Small} \\ X_{t,Med} \end{array}\right)\right]\mathrm{d}t + \Sigma\left(\begin{array}{c} dW_{t,1}^{\mathbb{Q}} \\ dW_{t,2}^{\mathbb{Q}} \\ \mathrm{d}W_{t,3}^{\mathbb{Q}} \end{array}\right)$$

The solution to the system of ODEs is

$$\begin{aligned} B^1(t,T) &= -\rho_1^1\frac{1-e^{-\kappa_1(T-t)}}{\kappa_1} \\ B^2(t,T) &= -\rho_1^2\frac{1-e^{-\kappa_2(T-t)}}{\kappa_2} \\ B^3(t,T) &= -\rho_1^3\frac{1-e^{-\kappa_3(T-t)}}{\kappa_3} \\ A(t,T) &= \int_t^T B(s,T)^{\intercal}K^{\mathbb{Q}}\theta^{\mathbb{Q}}\,\mathrm{d}s + \frac{1}{2}\int_t^T\sum_{j=1}^{n}[\Sigma^{\intercal}B(s,T)B(s,T)^{\intercal}\Sigma]_{jj}\gamma^j\,\mathrm{d}s \end{aligned}$$

As a result, the following relationship holds for the transaction costs c_t

$$\begin{aligned} log(1-Qc_t) &= -X_{t,Large}Q\left[\frac{1-e^{-\kappa_1(T-t)}}{\kappa_1}\right] \\ &\quad -X_{t,Small}Q\left[\frac{1-e^{-\kappa_2(T-t)}}{\kappa_2}\right]\left[\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}}\right] \\ &\quad -X_{t,Med}Q\left[\frac{1-e^{-\kappa_3(T-t)}}{\kappa_3}\right]\left[\frac{1-e^{(\frac{-S}{\varsigma_1})}}{\frac{S}{\varsigma_1}}-e^{(\frac{-S}{\varsigma_1})}\right]+A(t,T) \end{aligned}$$

Proof. See Appendix A ■

The transaction costs are a weighted average of the three unobserved state factors and a yield-adjustment $A(t,T)$ with closed-form solutions provided in the Appendix. Note that we have chosen a Nelson-Siegel-type specification for the function $\rho_1(Q,S)$ that maps the state factors into transaction costs c depending on the transaction side Q and the transaction size S.

Even though the factors are not directly observable, with the loading structure designed in our model, they can be clearly identified as immediate transaction costs for large, small, and medium trades. The Figure 1 plots our chosen Nelson-Siegel-type rho weights of our unobserved factors. The large factor is always present for any transaction size, and is the only one that affects the costs of large transactions. The weight coefficient for the small factor approaches a value of 1 for small transactions and decays for large transactions. And the medium size factor reaches its maximum after the threshold ς_1. In our empirical section, we choose a threshold of \$100,000 for retail trades. Our framework also allows the use of other transformation function ρ_1. For instance, one could consider polynomials or the non-linear transformations of Edwards et al. (2007). With the caveat that using other transformation functions will modify the interpretation of the extracted underlying factors $X_{t,j}$.

[INSERT FIGURE 1 HERE]

3.2 Defaultable bond pricing model

Let's consider a customer trade at time t

$$P(t,T) \;=\; V_t\left(\frac{1}{1-Qc_t}\right) \tag{11}$$

Now that we have modeled the transaction cost c_t, we need to characterize the "efficient" price $V(t,T)$ of a bond. The price $V(t,T)$ is generic and can accommodate either structural or reduced-form pricing models. In this article we adopt the default intensity approach of Lando (1998) and Duffie and Singleton (1999) since our main focus is on transaction costs dynamics and not the underlying default mechanism.

The price of a default-free zero-coupon bond that matures at time T is given by

$$B(t,T) \;=\; E_t^{\mathbb{Q}}\left[exp\left(-\int_t^T r_u\,\mathrm{d}u\right)\right] \tag{12}$$

We assume the existence of three latent dynamic factors similarly to Diebold et al. (2005), Diebold and Li (2006), and Christensen et al. (2011). Specifically, we follow the independent-factor arbitrage-free Nelson-Siegel (AFSN) of Christensen et al. (2011) where the risk-free term structure depends on three factors. Namely, level ($Level_t$), slope ($Slope_t$), and curvature ($Curv_t$). This model captures in a tractable way three important empirical features of Treasury yields and it is shown to have a better predictive performance than comparable models. Under this model, the three state variables are independent Vasicek processes under the P-measure and restrictions are imposed under the Q-measure such that the risk-free spot rate matches the dynamic Nelson-Siegel model of the form

$$
\begin{aligned}
spot_f(t,T) &= Level_t + Slope_t \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right] \\
&\quad + Curv_t \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right] - \frac{\tilde{A}(t,T)}{T-t}
\end{aligned} \tag{13}
$$

Risky securities require a credit spread. In the reduced-form model of Duffie and Singleton (1999), the default event is determined by the first jump of a Poisson process with stochastic intensity $h_t^{\mathbb{P}}$. And in case of a default event, the bond experiences a loss of $L\%$. The value of a risky bond is discounted by the risk-free rate and an instantaneous spread s_t

$$
V(t,T) = E_t^{\mathbb{Q}}\left[exp\left(-\int_t^T (r_u + s_u)\,\mathrm{d}u\right)|\Theta\right] \tag{14}
$$

This spread is the "risk-neutral mean-loss rate" $s_t \equiv h_t^{\mathbb{Q}} L$ which accounts for both the probability of default and the default loss. As our interest is not on the credit spread, s_t can be parameterized directly without considering h_t and L_t separately as it is done in Duffee (1999) and Driessen (2005). In our empirical setting, we are fitting an efficient price for each bond and not for a set of maturities for the same firm. Having only one process for the credit spread and estimating the spread term structure by bond reduces significantly the modelling pressure on the efficient price V_t and simplifies the estimation. As long as the spread follows an affine model of the form in equation 6, it can easily be solved with the well-known exponentially-affine model solutions summarized in equations 8-9.

The resulting model combines the effects of changes in the default-free interest rate, the risk-neutral mean loss rate, and most importantly the transaction cost dynamics for different transaction sizes.

$$P(t,T) \;=\; E_t^{\mathbb{Q}}\left[exp\left(-\int_t^T (r_u + s_u)\,\mathrm{d}u\right)\right] E_t^{\mathbb{Q}}\left[exp\left(-\int_t^T -\ell_u\,\mathrm{d}u\right)\right] \tag{15}$$

This mimics the "default and liquidity-adjusted" model of Duffie and Singleton (1999) that adds a stochastic carrying cost ℓ directly in the short-rate process s_t. The added advantage is that we have given a structure to ℓ that depends on transaction state factors and exploits transaction characteristics that are inferred from historical data. Our transaction costs ℓ could also be embedded into the short-rate process. And in the case of a mean-loss rate conditionally independent from the transaction costs, both approaches would be equivalent.

If we assume that the three components (risk-free rate, default and costs) are conditionally independent, the expectation in equation 15 is the product of three expectations for each the three underlying components. The instantaneous interest rate r_t, the default intensity together with the loss-given default $h_t L$, and the transaction cost ℓ_t are assumed to be an affine function of state variables. In the case of coupon bonds, the efficient value discounts each cash-flow by the risk-free and the mean-loss rate, and afterwards the proportional transaction cost is applied onto the efficient value (See Appendix B).

It is widely acknowledged that a large component of credit spreads does not only reflect credit risk information. The current literature finds it difficult to reconcile credit risk models with the observed credit spreads (see e.g. Huang and Huang, 2012). Elton et al. (2001) report that almost half of the credit spread accounts for a non-default systematic risk component. Collin-Dufresne et al. (2001) find that a single factor independent from credit and liquidity factors is a relevant component of credit spread changes. In contrast, Longstaff et al. (2005) use Credit Default Swap spreads to extract the default component in bond spreads and find

that more than 50% of the corporate spread is due to default risk, whereas the non-default component is strongly mean reverting and highly related to bond-specific bid-ask spreads.

In light of extensive studies showing the strong presence of illiquidity in credit spreads, we believe that our transaction cost component can help understand to a higher extent the time-varying properties of liquidity frictions faced by large as well as small investors. Using quarterly bid-ask quotes from Bloomberg for large sample of bonds, Chen et al. (2007) provide strong evidence that illiquidity explains half of the cross-sectional variation in credit spreads. Using transaction data, Bao et al. (2011) explain that bid-ask spreads do not capture to a full extent the illiquidity of corporate bonds. More interestingly, they find that an aggregate illiquidity measure —constructed at monthly frequency as the cross-sectional median price reversal— largely explains the time variation in yields spreads.

Similarly to Bao et al. (2011) and Hendershott and Menkveld (2014) we understand the spread s_t as the fundamental or efficient spread in the absence of transaction frictions. This efficient spread may embed other risks other than credit risk. We do not focus on the components of the efficient spread, and we shift our interest to the component ℓ_t that reflects the time-varying transaction costs that deviate bond rates from the efficient spreads.

4 Data and Estimation Method

4.1 Data

We download all bond trades for S&P500 firms. We merge TRACE, FISD, and COMPUSTAT. The dataset starts in July 2002 and finishes in July 2013. First, we identify the tickers members of the S&P500. Second, we download all the corporate bond transactions of the corresponding companies. We filter the corporate bond transactions to account for cancelled and corrected trades. TRACE allows us to identify Buy trades, Sell trades, and interdealer

trades, and also records the price and the size of each transaction. See Appendix D for a detailed description of the database.

4.2 State-space formulation for the transaction cost model

We consider the following nonlinear state-space system:

$$X_{t+h} = F(X_t, \eta_{t+h}) \tag{16}$$

$$y_{t+h} = G(X_{t+h}, \Theta) + \varepsilon_{t+h} \quad , \qquad Var(\varepsilon_{t+h}) = H_{t+h} \tag{17}$$

where X_{t+h} is the unobserved state of the system, y_{t+h} is the observed market price, and h is the time interval between two observations. The first equation in this state-space model is the transition equation. This transition equation reflects the discrete time evolution of the state variables and represents a discretized equivalent of equation 6. The second equation is the measurement equation is determined by the state vector and its measurement function G is specified by our pricing model. The observation noise ε_{t+h} is normal with zero mean and covariance matrix denoted by H_{t+h}. We take into account that the observation error increases with the time between observations and we assume that $H_{t+h} = h\sigma_\varepsilon^2 I_{N_{t+h}}$, where N_{t+h} is the number of transactions observed at a given time.

The transition equation approximates the state dynamics using the two first conditional moments. This setup allows for QML estimation based on the Kalman filter. The conditional mean and variance of the continuous-time affine model given in equation 6 are[5]

[5]More details about how to obtain the conditional moments of the affine model can be found in Fisher and Gilles (1996), Fackler (2000), Duffee (2002), and the online appendix of Christoffersen et al. (2014). De Jong (2000) shows how to calculate the mean and variance in the case of a diagonal mean-reversion coefficient

$$E_t(X_{t+h}|X_t) = (I - e^{-Kh})\theta + e^{-Kh}X_t \quad (18)$$

$$Var(X_{t+h}|X_t) = \int_0^h e^{-K(h-s)}\Sigma \; diag(\gamma + \delta E_t(X_{t+s}))\; \Sigma^\intercal e^{-K^\intercal(h-s)}\mathrm{d}s \quad (19)$$

We assume that the state noise η_{t+h} is normal with zero mean and covariance matrix denoted by Q_{t+h}. In particular, given a process with expected zero mean in equation 19, the state covariance becomes $Q_{t+h} = \int_0^h e^{-K(h-s)}\Sigma \; diag(\gamma)\; \Sigma^\intercal e^{-K^\intercal(h-s)}\mathrm{d}s$. The transition equation over a discrete time interval, with length equal to h units of time, can be written as

$$X_{t+h} = (I - exp(-Kh))\theta + exp(-Kh)X_t + \eta_{t+h} \quad , \qquad Var(\eta_{t+h}) = Q_{t+h} \quad (20)$$

4.3 Estimation method

First we estimate the default-free bond model. Second, we estimate the defaultable bond price with transaction costs. For this second step, we take the default-free estimates as given.

We use the Kalman filter approach to estimate the underlying stochastic factors. In particular, we use the Unscented Kalman filter. The discretized state transition equation is Gaussian, but the measurement equation is non-linear due to the presence of coupon payments before maturity. The Extended Kalman filter approximates the measurement equation using a Taylor expansion. This approach becomes particularly undesirable in our setting, because there might be a long time between two observations worsening the performance of the Taylor approximation. Instead we use the Unscented Kalman filter (see Julier and Uhlmann, 1997, 2004) that directly allows for nonlinearities. This method is computationally efficient

matrix. For the Gaussian model, the mean and variance are provided in Dai and Singleton (2002). We calculate the mean and variance using the general analytical solutions provided by Fackler (2000).

as we do not need to calculate derivatives for the linear approximation. In particular, we use the square-root unscented Kalman filter of Van der Merwe and Wan (2001) that is numerically stable and guarantees a positive definite state covariance matrix. For applications of the UKF, we have as examples Chen et al. (2008), Fontaine and Garcia (2012), Doshi et al. (2013), Christoffersen et al. (2014) and Elkamhi et al. (2014).

The Kalman filter also allows evaluating the log-likelihood function.

$$logL \quad = \quad \sum_{k=1}^{K} -\frac{N_k}{2} log(2\pi) - \frac{1}{2} log|F_k| - \frac{1}{2}(Y_k - \hat{Y}_k^-)^\intercal F_k^{-1}(Y_k - \hat{Y}_k^-) \quad , \tag{21}$$

where K is the sample size, N_k is the number of observations, the conditional covariance matrix F_k is obtained with the product $S_{\hat{Y}_k} S_{\hat{Y}_k}^\intercal$ obtained from the square-root unscented Kalman filter, and the difference $Y_t - \hat{Y}_k^-$ measures the prediction error.

We maximize this log-likelihood to estimate the parameters and extract the underlying state factors. An interesting and relevant feature of our Kalman filter and the log-likelihood is that we need to consider at each step the time between two observations. This is because the observations are not equally spaced. Thus, the exact time between two observations is used to generate the expected state in the next step, and the exact time to maturity is used to generate the equilibrium price. As each bond contains a large number of observations, maximizing the likelihood function directly is computationally very expensive. Instead, we maximize the likelihood function indirectly using a Recursive Least Squares approach. We design a Dual Kalman filter where the parameters are a stationary process with noise, and we learn the value of the parameter vector iteratively with each observation. Under this approach we search for the parameter vector that minimizes the prediction error. We run the Dual Kalman filter multiple times, and ultimately we choose the set of parameters that yield the best log-likelihood. This approach significantly reduces the computational burden

as the likelihood gradient vector does not need to be estimated. See Appendix C for a detailed formulation.

5 Empirical Results

5.1 The risk-free term structure

Before turning to the estimation of the transaction costs, we first need to measure the risk-free discount curve that is going to be used for each observation taking into account the time until the remaining cashflows. For the risk-free curve, we collect daily data for the constant maturity 1-month, 3-month, 6-month, 1-year, 2-year, 3-year, 5-year,7-year, 10-year, 20-year, and 30-year rates from the Federal Reserve. We have a total of 11 par rates. The bills below 1-year maturity are assumed to pay one single coupon at maturity, and the remaining securities pay coupons semiannually. By design, these securities are at par value and as a result our measurement value is always 100 for all the maturities. At each day, each coupon payment is discounted using the independent-factor AFNS model with the objective that the present value equals 100. Our procedure consists on using a Dual Kalman filter to minimize the mean-square error recursively. Table 1 reports the estimates of the risk-free term structure. The estimated parameters displayed in Panel A. Similarly to Christensen et al. (2011) the curvature factor exhibits the highest mean-reversion and volatility, whereas the level factor exhibits the lowest mean-reversion and volatility. However, our estimates are different since we are using a more recent sample period. The model performs remarkably well in the daily dataset as we can observe in Panel B. The prediction errors are quite small and below 1.6% of par value for all maturities. Short term securities show the lowest pricing errors. For instance 1-month bills have a pricing error of \$0.035 of par value and 30-year Treasuries have a pricing error of \$1.532 of par value. Naturally the longer maturities exhibit

a larger prediction error because they have a larger duration and thus they are more sensitive to small changes in the discount rate.

[INSERT TABLE 1 HERE]

5.2 Transaction-based model estimates

To illustrate how the transaction costs have evolved over the last decade, we select the CUSIP=459200BA8, one of the most liquid bonds issued by IBM. This is a 4.5% semi-annual coupon paying bond that matures in November, 29th 2012.

The Figure 2 plots 34 thousand transactions recorded on this particular bond. Each circle represents a transaction, and the size of the circle depends on the transaction sizes, with larger circles identifying larger transactions. We can clearly observe that Buy prices are frequently over the Sell prices. A common concern when calculating liquidity from historical transactions is that transactions are recorded at different times. Thus, using prices recorded at different times to measure liquidity might be misleading, especially in the case of illiquid bonds with infrequent trades. As the price difference between distant trades may reflect default information, and not liquidity, one should consider the time between observations.

We can also appreciate how the price difference between buys and sells is larger at the beginning of the sample. This separation becomes much smaller near maturity, when the security has a small number of coupons left and it converges rapidly to par value. The size also seems to be a relevant characteristic as smaller trades tend to suffer extreme prices.

[INSERT FIGURE 2 HERE]

The empirical identification of the transaction cost factors depends on having enough observations with different transaction sizes spread along the sample period. The observed costs may be downward biased if the decision to trade depends on the transaction costs.

Large investors may avoid large trades in times of high transaction costs and break up a large trade into smaller trades whenever smaller trades face smaller costs. The summary statistics in Table 2 show that the retail transactions account for a 79% of the trades, whereas the large institutional trades represent a 7%. The remaining 14% fall into the medium size category. Even though larger trades are a smaller proportion of the sample, the large factor can always be identified from any transaction as it captures the baseline transaction cost of any trade ($\rho_1^1 = \pm 1$). The factors could also be biased if there were only buy trades for large transactions and sell trades for small transactions, however, our sample has a fair balance of buys and sells across different size buckets.

[INSERT TABLE 2 HERE]

The timing of the observation could also affect the state estimates if large trades and smaller trades were observed at different time periods. This is not the case in our sample. On average small trades are observed every 3.3 hours and large trades are observed every 1.58 days. To make sure that there is a fair representation of different sizes at different times, we plot the volume of trades per quarter in Figure 3. Large trades are regularly observed in each quarter. Small trades (27,007 trades) outnumber large trades (2,340 trades) but they represent a much lower Dollar Volume. The retail trades account for \$616 million whereas the large trades represent \$6.5 billion. Most of the volume traded by institutional investors takes place near issuance. In the first quarter of 2013 only, near \$700 million were exchanged by large investors, which exceeds the amount traded by small investors during the whole life of the bond. The empirical evidence shows that large trades were not replaced by retail trades, or otherwise we would have seen a large Dollar Volume increase in small trades.

[INSERT FIGURE 3 HERE]

We estimate our model as described in the previous section, and as a result, the dynamics of the state vector is revealed in Figure 4. The model correctly takes into account the time

between observations, the time to maturity, and the transaction costs for different transaction sizes. We observe a large factor that is small at the beginning of the sample, but it increases during the financial crisis. The small factor adds more costs for small transactions. The small investors have suffered price pressures since the bond inception and during the financial crisis. The medium factor takes its minimum during the early 2000s period of low monetary policy interest rates.

[INSERT FIGURE 4 HERE]

One of the advantages of the instantaneous transaction costs is that we are able to calculate the cost curve at any point in time using the estimated state values. The Figure 5 plots the equilibrium cost-size relationship. Similarly to the findings of Edwards et al. (2007) the costs decrease with the size of the transaction. However, we are able to depict a time-varying cost c_t. Comparing 2005 with 2010 we can see that the costs decrease as we approach maturity, consistent with the model predictions and preliminary evidence. It is not a surprise to see that the crisis year of 2008 suffered some of the highest costs, with a half-spread of 140 bps for a small transaction and 60 bps for a large transaction.

[INSERT FIGURE 5 HERE]

6 Transaction costs and the term structure

How relevant are the transaction costs to explain the term structure of interest rates? To motivate the relevance of transaction costs in analyzing the term structure we can simply return to equation 3. Broadly speaking we can consider two main effects driving the term structure of transaction costs. The first one is a short-term effect that depends on how distant the state factor is from its average. Periods of high levels of the state factors translate into high transaction costs at short-term maturities. The second effect is a long-term effect

captured by $A(t,T)$ that becomes more relevant for periods distant to maturity and mainly captures the volatility of the state until maturity.

To illustrate the effect of transaction costs, we show in Figure 6 the term structure of a representative investment grade company with an upward sloping yield curve. We choose a set of parameters that summarize stylized empirical findings on transaction costs for the IBM 4.75% bond and compare the costs faced by customer buys/sells under two scenarios. A scenario of average transaction costs and a scenario of high transaction costs.

[INSERT FIGURE 6 HERE]

When the state vector lays on its average, the transaction costs tend to increase with maturity, as the short-term effect is small and the long-term effect prevails. In this case, the yield curve can deviate 19 bps from the efficient yield curve in the case of small trades. However, in a period of high transaction costs, the short-term securities suffer much larger transaction costs than long-term securities. In our example, the transaction costs can deviate the yield by 90 bps from the efficient yield curve in the case of small buy trades of short-term securities. The gap size between the market and the efficient yield mostly depends on the persistence of transaction costs. Highly persistent transaction costs increase the deviation from the efficient yields and the deviation propagates to higher maturities as it is more likely that costs remain high for a longer time. In the case of low persistence, shocks are short-lived and would have a minimum effect only on short-term yields.

More interestingly, the yield curve is diverse and depends on the size of a given transaction. With smaller trades generally leading to higher transaction costs and larger deviations from the efficient yield curve. The yield costs may be relative similar across maturities, but not its absolute cost. The last two graphs show the price proportional costs as a function of size and time to maturity. The absolute cost decreases with the transaction size, but more importantly it depends on the time to maturity. There can be a difference of 100 bps

between 10-year bonds and 5-year bonds in small trades. And during illiquid periods, the transaction costs may become 12% larger.

7 Conclusion

This article models the transaction costs faced by different trade sizes in the corporate bond market. Our model allows us to take into account the uncertainty in transaction costs faced by any trade size. To our knowledge, transaction cost risk is yet unexamined in the asset pricing literature. Our methodology specifies how transaction cost risk for different trade sizes translates to asset prices.

We show evidence for the existence of three transaction cost factors faced by investors. The first factor captures the costs faced by large investors. The second factor captures the difference between small and large transactions. And the third factor accounts for the movements in costs for medium trades.

Using a large dataset with past historical transactions we measure the underlying factors and the model parameters. This can be used by market participants to create expectations on transaction costs and predict future prices. Also, this allows creating a yield curve for any specific transaction size.

The evidence shows that the underlying factors are highly time-varying. We find that the liquidity is significantly affected by the market environment, particularly during the financial crisis. The results are consistent with previous research showing that the transaction costs are smaller for large transactions and for bonds nearer maturity. However, our costs are dynamic and arbitrage-free. Most interestingly, retail trades suffer from costs that are more persistent than for large trades. In other words, shocks have a longer lasting effect in the case of retail trades. Thus, empirical work that value-weights trades will tend to neglect retailers who are exposed to more damaging transaction cost risk.

Given the large size of our dataset, we are currently estimating the model for each bond in the sample, and we will provide detailed statistics about the large, small, and medium factors in future versions of the paper.

References

Back, K. and K. Crotty (2014). The informational role of stock and bond volume. *The Review of Financial Studies 28*(5), 1381–1427.

Bao, J., J. Pan, and J. Wang (2011). The illiquidity of corporate bonds. *The Journal of Finance 66*(3), 911–946.

Bessembinder, H., W. Maxwell, and K. Venkataraman (2006). Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. *Journal of Financial Economics 82*(2), 251–288.

Chen, L., D. A. Lesmond, and J. Wei (2007). Corporate yield spreads and bond liquidity. *The Journal of Finance 62*(1), 119–149.

Chen, R.-R., X. Cheng, F. Fabozzi, and B. Liu (2008). An explicit, multi-factor credit default swap pricing model with correlated factors. *Journal of Financial and Quantitative Analysis 43*(1), 123–160.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics 164*(1), 4–20.

Christoffersen, P., C. Dorion, K. Jacobs, and L. Karoui (2014). Nonlinear kalman filtering in affine term structure models. *Management Science 60*(9), 2248–2268.

Collin-Dufresne, P., R. S. Goldstein, and J. S. Martin (2001). The determinants of credit spread changes. *The Journal of Finance 56*(6), 2177–2207.

Dai, Q. and K. J. Singleton (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics 63*, 415–441.

De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics, 18*(3), 300–314.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*(2), 337–364.

Diebold, F. X., M. Piazzesi, and G. D. Rudebusch (2005). Modeling bond yields in finance and macroeconomics. *The American Economic Review 95*(2), 415–420.

Doshi, H., J. Ericsson, K. Jacobs, and S. M. Turnbull (2013). Pricing credit default swaps with observable covariates. *Review of Financial Studies 26*(8), 2049–2094.

Driessen, J. (2005). Is default event risk priced in corporate bonds? *The Review of Financial Studies 18*(1), 165–195.

Duffee, G. R. (1999). Estimating the price of default risk. *The Review of Financial Studies 12*(1), 197–226.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance 57*(1), 405–443.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*(4), 379–406.

Duffie, D. and K. J. Singleton (1999). Modeling term structures of defaultable bonds. *The Review of Financial Studies 12*(4), 687–720.

Edwards, A. K., L. E. Harris, and M. S. Piwowar (2007). Corporate bond market transaction costs and transparency. *The Journal of Finance 62*(3), 1421–1451.

Elkamhi, R., K. Jacobs, and X. Pan (2014). The cross section of recovery rates and default probabilities implied by credit default swap spreads. *Journal of Financial and Quantitative Analysis 49*(1), 193–220.

Elton, E. J., M. J. Gruber, D. Agrawal, and C. Mann (2001). Explaining the rate spread on corporate bonds. *The Journal of Finance 56*(1), 247–277.

Fackler, P. L. (2000, June). Moments of affine diffusions. Working Paper, North Carolina State University.

Feldhütter, P. (2012). The same bond at different prices: Identifying search frictions and selling pressures. *The Review of Financial Studies 25*(4), 1155–1206.

Fisher, M. and C. Gilles (1996, April). Term premia in exponential-affine models of the term structure. Working Paper, Federal Reserve Board.

Fontaine, J.-S. and R. Garcia (2012). Bond liquidity premia. *The Review of Financial Studies 25*(4), 1207–1254.

Goyenko, R., A. Subrahmanyam, and A. Ukhov (2011). The term structure of bond market liquidity and its implications for expected bond returns. *Journal of Financial and Quantitative Analysis 46*(1), 111–139.

Hendershott, T. and A. J. Menkveld (2014). Price pressures. *Journal of Financial Economics 114*(3), 405–423.

Huang, J.-Z. and M. Huang (2012). How much of the corporate-treasury yield spread is due to credit risk? *Review of Asset Pricing Studies 2*(2), 153–202.

Julier, S. J. and J. K. Uhlmann (1997). New extension of the kalman filter to nonlinear systems. *SPIE Proceedings 3068*, 182–193.

Julier, S. J. and J. K. Uhlmann (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE 92*(3), 401–422.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research 2*, 99–120.

Longstaff, F. A., S. Mithal, and E. Neis (2005). Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market. *The Journal of Finance 60*(5), 2213–2253.

Rossi, M. (2014). Realized volatility, liquidity, and corporate yield spreads. *Quarterly Journal of Finance 4*(1).

Schultz, P. (2001). Corporate bond trading costs: A peak behind the curtain. *The Journal of Finance 56*(2), 677–698.

Van der Merwe, R. and E. A. Wan (2001). The square-root unscented kalman filter for state and parameter-estimation. *ICASSP 2001, IEEE International Conference 6*, 3461–3464.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control 23*(3), 395–404.

Figure 1: Transaction cost factor coefficients

The graph shows the coefficients for each transaction cost factor as a function of transaction size and assuming a size threshold of $\varsigma_1 = \$100,000$.

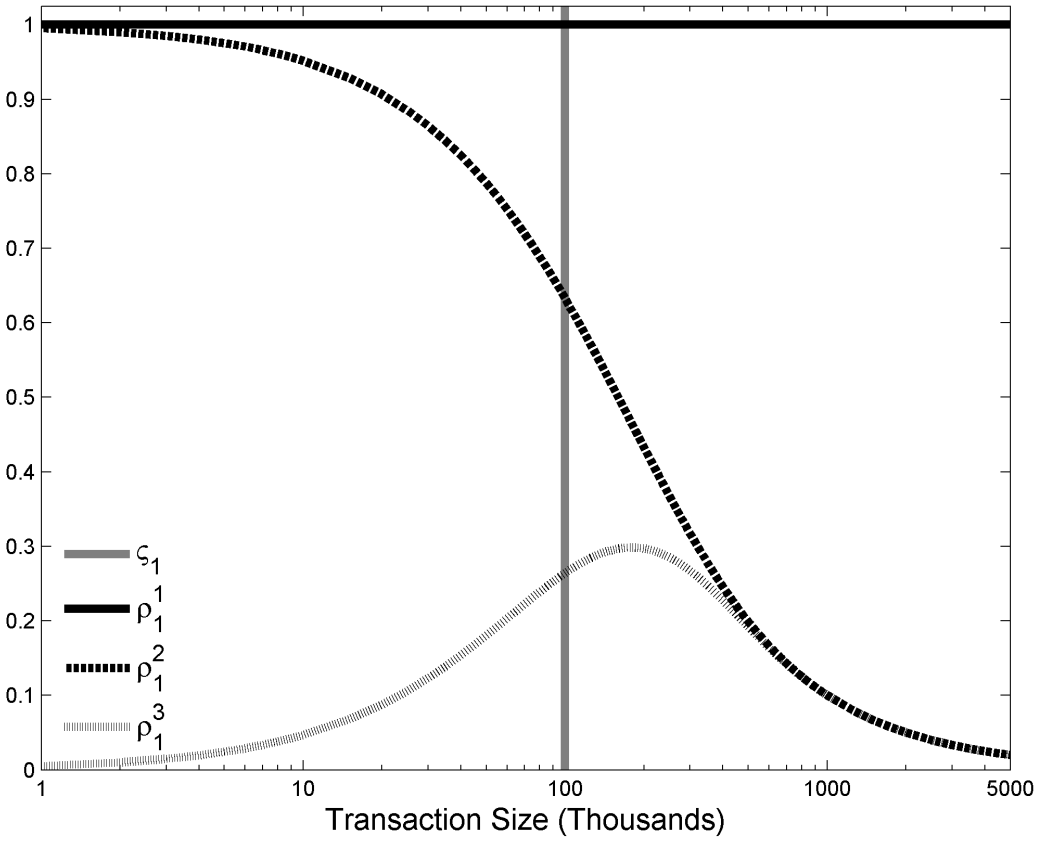


Figure 2: IBM transaction sizes

The graph shows Buy and Sell Trades for bond IBM 4.75% 29/11/2012 with CUSIP=459200BA8. There are a total of 34 thousand customer transactions.

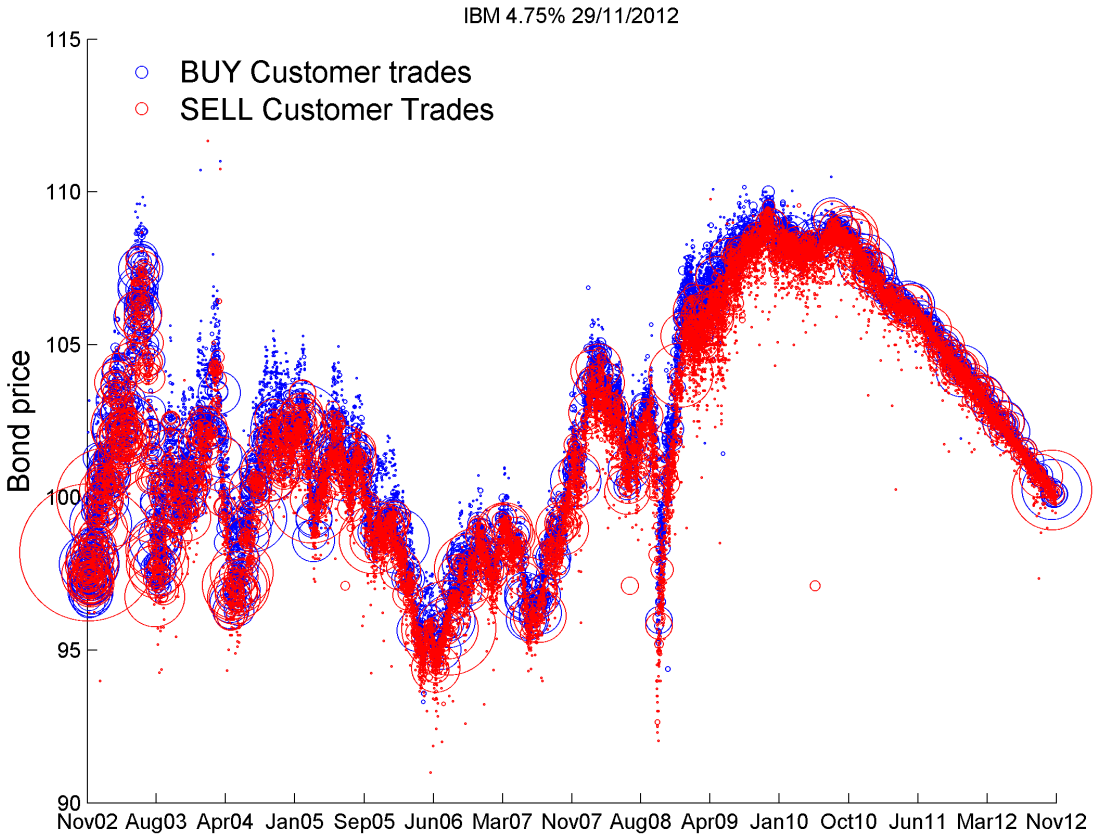
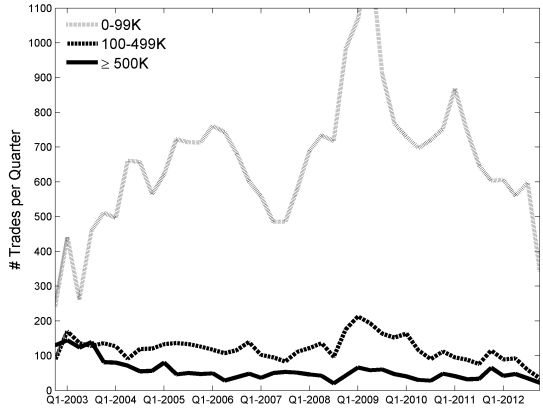


Figure 3: IBM transactions per quarter

The graphs shows the number and volume of trades per quarter by size categories for bond IBM 4.75% 29/11/2012 with CUSIP=459200BA8. There are a total of 34 thousand customer transactions.

(a) # Trades



(b) Dollar Volume (millions)

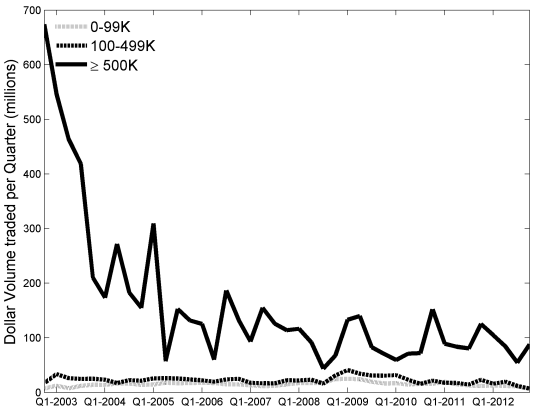


Figure 4: IBM transaction cost factors

The graph shows the underlying transaction cost factors for large, small and medium trades using a total of 34 thousand customer transactions.

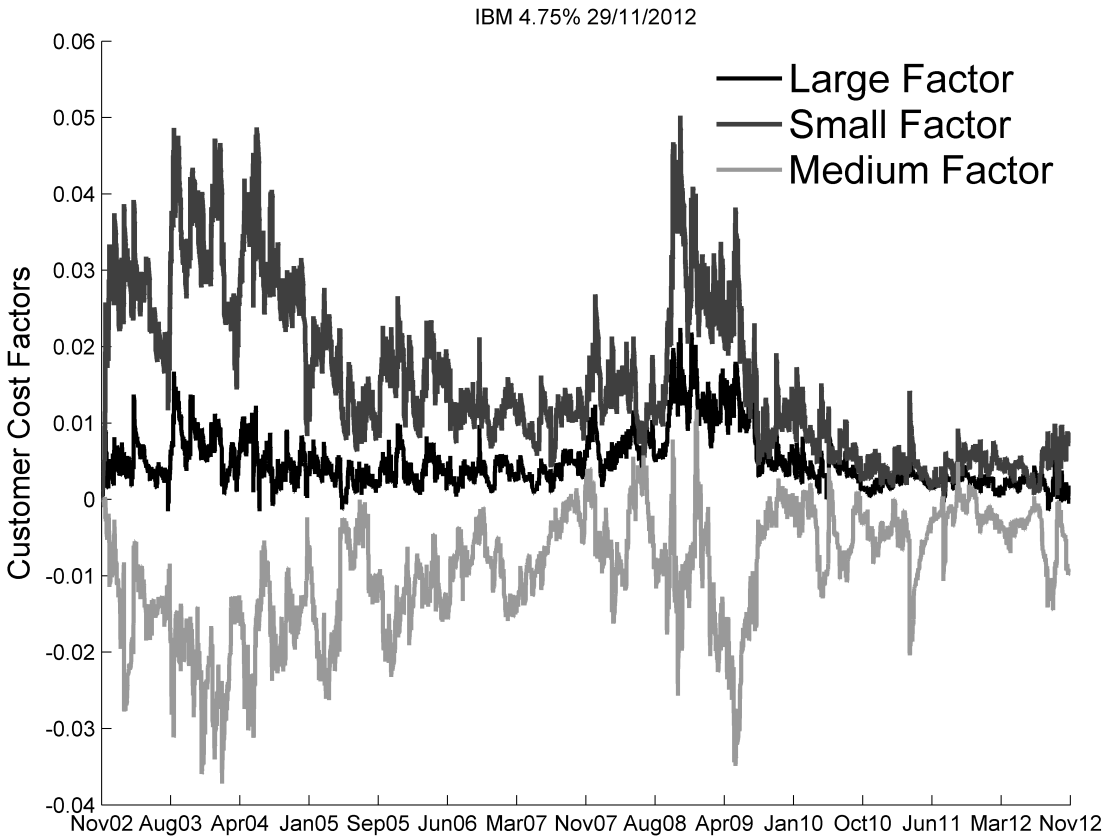


Figure 5: IBM transaction cost function

The graph shows the cost function for three dates.

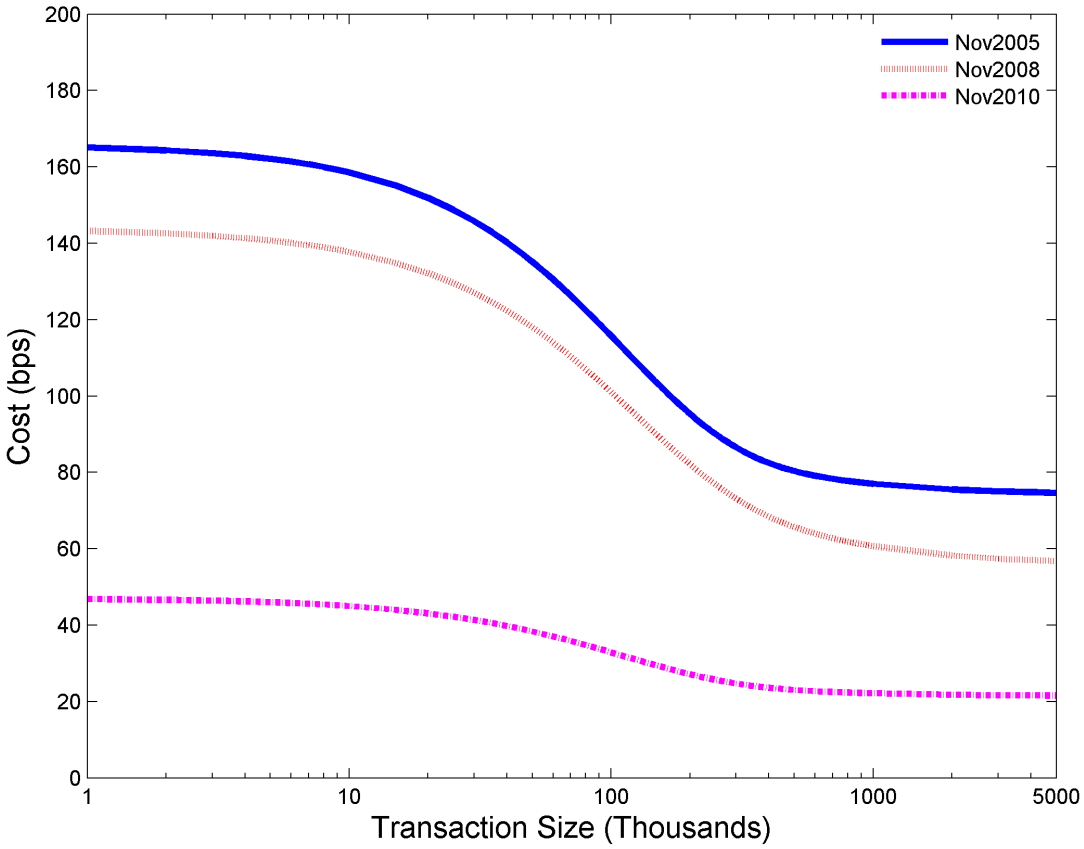
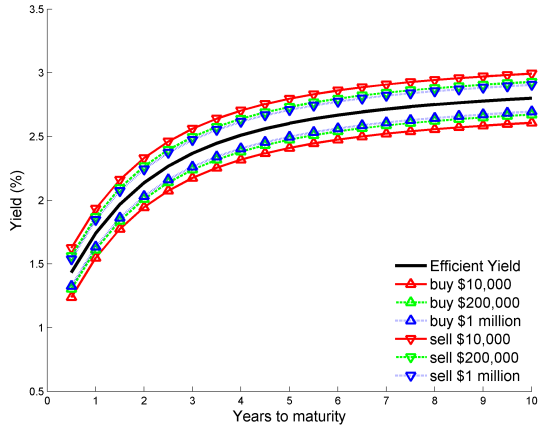


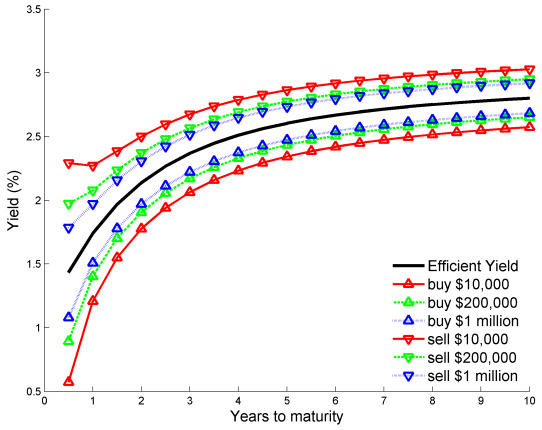
Figure 6: Representative Investment Grade Bond

This graph shows the yield curve and the proportional costs for a set of maturities and transaction sizes under two scenarios: an scenario of average transaction costs, and an scenario of high transaction costs. We assume θ_1=10 bps, θ_1=10 bps, θ_1=-5 bps, κ_1=10, κ_2=8, κ_3=7, σ_i=1 bps. Graphs (a) and (c) show a period of average costs where the state variables are at their average $X_t = \theta$. Graphs (b) and (d) show a period of high costs where the "large" factor is 100 bps higher than its average $X_{t,Large} = \theta_1 + 0.01$, and the "small" factor is 200 bps higher than its average $X_{t,Small} = \theta_2 + 0.02$.

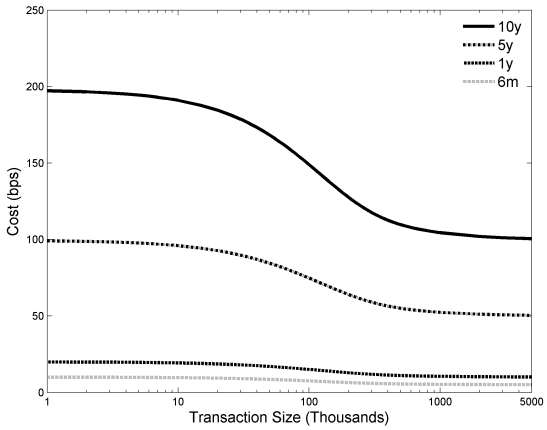
(a) Yield Curve with average costs



(b) Yield Curve with high costs



(c) Cost function with average costs



(d) Cost function with high costs

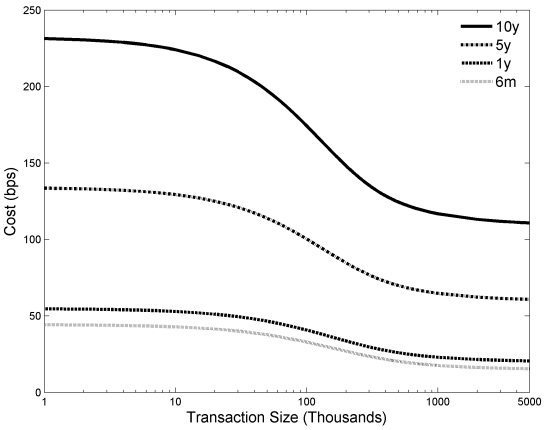


Table 1: The risk-free term structure

Panel A contains estimates for the independent-factor arbitrage-free Nelson-Siegel model of Christensen et al. (2011). Panel B reports the root-mean-squared prediction errors (RMSEs) and the measurement error standard deviations (ME SD). The model is estimated on daily data for Treasury securities from July 2002 to June 2016 using the Dual Square-root Unscented Kalman filter approach described in Appendix C. It is assumed that all interest rates are observed with i.i.d. measurement errors.

Panel A - Parameter estimates

	λ	$\kappa^{\mathbb{P}}_{Level}$	$\kappa^{\mathbb{P}}_{Slope}$	$\kappa^{\mathbb{P}}_{Curv}$	$\theta^{\mathbb{P}}_{Level}$	$\theta^{\mathbb{P}}_{Slope}$	$\theta^{\mathbb{P}}_{Curv}$	σ_{Level}	σ_{Slope}	σ_{Curv}
	0.1995	0.4552	0.4889	1.3935	0.0360	-0.0263	0.0127	4×10^{-4}	131×10^{-4}	145×10^{-4}

Panel B - Model performance

	1 mth	3 mth	6 mth	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	20 Yr	30 Yr
RMSE (% par value)	0.0353	0.0816	0.1122	0.1600	0.5053	0.9441	1.4097	1.5458	1.4547	1.5858	1.5238
ME SD (% par value)	0.0343	0.0792	0.1050	0.1551	0.3858	0.6846	1.0629	1.2420	1.0439	1.1709	0.6363
Log Lkl							-5.70×10^4				

Table 2: Transactions summary

Summary statistics of transactions for the bond IBM 4.75% 29/11/2012 with CUSIP=459200BA8.

				Time between transactions (days)	
	Volume (mil. $)	# Trans.	% Trans.	Average	Std. Dev.
Buy 0-99K	308.6330	11177	32.6288	0.3315	0.7173
Sell 0-99K	307.5155	15829	46.2093	0.2339	0.6072
0-99K	616.1485	27007	78.8410	0.1373	0.4507
Buy 100-499K	592.3020	3155	9.2103	1.1738	1.6579
Sell 100-499K	328.6450	1776	5.1846	2.0746	2.4797
100-499K	920.9470	4932	14.3979	0.7509	1.1581
Buy ≥500K	3,113.0	1300	3.7951	2.8482	3.9172
Sell ≥500K	3,436.6	1039	3.0331	3.5600	4.5157
≥500K	6,549.6	2340	6.8311	1.5827	2.4153

A Proof of Proposition 1

We set a constant volatility ($\delta^j = 0$) and the ODEs for $B(T,t)$ in equation 8 simplify to:

$$\frac{\mathrm{d}B(t,T)}{\mathrm{d}t} = \rho_1 + (K^{\mathbb{Q}})^\intercal B(t,T) \quad , \quad B(T,T) = 0$$

This system of linear ODEs can be solved using the matrix exponential. First, we rearrange the ODEs with the constant on the right-hand side. Second, we multiply the equation by the matrix exponential $e^{(K^{\mathbb{Q}})^\intercal (T-t)}$. Third, it can be shown that $\frac{\mathrm{d}}{\mathrm{d}t}(e^{(K^{\mathbb{Q}})^\intercal (T-t)} B(t,T)) = \left(\frac{\mathrm{d}B(t,T)}{\mathrm{d}t} - (K^{\mathbb{Q}})^\intercal B(t,T)\right) e^{(K^{\mathbb{Q}})^\intercal (T-t)}$. Fourth, we integrate the equation. Fifth, we solve the integral in the left-hand side. Last, we move $B(T,T)$ to the right-hand side.

$$\begin{aligned}
\frac{\mathrm{d}B(t,T)}{\mathrm{d}t} - (K^{\mathbb{Q}})^\intercal B(t,T) &= \rho_1 \\
\left(\frac{\mathrm{d}B(t,T)}{\mathrm{d}t} - (K^{\mathbb{Q}})^\intercal B(t,T)\right) e^{(K^{\mathbb{Q}})^\intercal (T-t)} &= e^{(K^{\mathbb{Q}})^\intercal (T-t)} \rho_1 \\
\frac{\mathrm{d}}{\mathrm{d}t}(e^{(K^{\mathbb{Q}})^\intercal (T-t)} B(t,T)) &= e^{(K^{\mathbb{Q}})^\intercal (T-t)} \rho_1 \\
\int_t^T \frac{\mathrm{d}}{\mathrm{d}s}\left(e^{(K^{\mathbb{Q}})^\intercal (T-s)} B(s,T)\right) \mathrm{d}s &= \int_t^T e^{(K^{\mathbb{Q}})^\intercal (T-s)} \rho_1 \, \mathrm{d}s \\
B(T,T) - e^{(K^{\mathbb{Q}})^\intercal (T-t)} B(t,T) &= \int_t^T e^{(K^{\mathbb{Q}})^\intercal (T-s)} \rho_1 \, \mathrm{d}s \\
-e^{(K^{\mathbb{Q}})^\intercal (T-t)} B(t,T) &= -B(T,T) + \int_t^T e^{(K^{\mathbb{Q}})^\intercal (T-s)} \rho_1 \, \mathrm{d}s
\end{aligned}$$

Next, we solve for B(t,T). First, we left-multiply the equation above by $-e^{-(K^{\mathbb{Q}})^\intercal (T-t)}$. Second, we use the boundary condition. Third, we simplify by including the matrix exponential inside the integral.

$$
\begin{aligned}
B(t,T) &= e^{-(K^{\mathbb{Q}})^{\intercal}(T-t)}B(T,T) - e^{-(K^{\mathbb{Q}})^{\intercal}(T-t)}\int_t^T e^{(K^{\mathbb{Q}})^{\intercal}(T-s)}\rho_1\,\mathrm{d}s \\
B(t,T) &= -e^{-(K^{\mathbb{Q}})^{\intercal}(T-t)}\int_t^T e^{(K^{\mathbb{Q}})^{\intercal}(T-s)}\rho_1\,\mathrm{d}s \\
B(t,T) &= -\int_t^T e^{(K^{\mathbb{Q}})^{\intercal}(t-s)}\rho_1\,\mathrm{d}s
\end{aligned}
$$

If $(K^{\mathbb{Q}})^{\intercal}$ is nonsingular and ρ_1 is time-independent, then the solution to the integral is

$$
B(t,T) = \left((K^{\mathbb{Q}})^{\intercal}\right)^{-1}\left(e^{-(K^{\mathbb{Q}})^{\intercal}(T-t)} - I\right)\rho_1
$$

When $(K^{\mathbb{Q}})$ is a nonsingular diagonal matrix with diagonal elements κ_i, the solution becomes:

$$
K^{\mathbb{Q}} = \begin{pmatrix} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{pmatrix} \implies B(t,T) = \begin{pmatrix} -\rho_1^1\frac{1-e^{-\kappa_1(T-t)}}{\kappa_1} \\ -\rho_1^2\frac{1-e^{-\kappa_2(T-t)}}{\kappa_2} \\ -\rho_1^3\frac{1-e^{-\kappa_3(T-t)}}{\kappa_3} \end{pmatrix} ,
$$

which proofs our Proposition. Solutions can also be calculated when the matrix $(K^{\mathbb{Q}})^{\intercal}$ is nonsingular using the Jordan form of $(K^{\mathbb{Q}})^{\intercal}$. See the Appendix in Christensen et al. (2011) for a particular example where $(K^{\mathbb{Q}})^{\intercal}$ is singular.

We can also solve for A(t,T) integrating the ODE in equation 9. First, we use the boundary condition to solve for the constant C. Second, we replace the value of C to find the definite integral that solves for A(t,T). Third, we use the integral on the ODE. Fourth, we express the last term in matrix notation using the *trace*.

$$
\begin{aligned}
A(t,T) &= \int \frac{\mathrm{d}A(s,T)}{\mathrm{d}s}\,\mathrm{d}s + C \\
A(t,T) &= -\int_t^T \frac{\mathrm{d}A(s,T)}{\mathrm{d}s}\,\mathrm{d}s \\
A(t,T) &= -(T-t)\rho_0 + \int_t^T B(s,T)^\intercal K^{\mathbb{Q}}\theta^{\mathbb{Q}}\,\mathrm{d}s + \frac{1}{2}\int_t^T \sum_{j=1}^{n} [\Sigma^\intercal B(s,T)B(s,T)^\intercal \Sigma]_{jj}\gamma^j\,\mathrm{d}s
\end{aligned}
$$

Before we calculate the first integral in this equation, we need the indefinite integral of $B(t,T)^\intercal$

$$
\int B(s,T)^\intercal\,\mathrm{d}s = \rho_1^\intercal \left(e^{-(K^{\mathbb{Q}})(T-s)}(K^{\mathbb{Q}})^{-1} - sI\right)\left(K^{\mathbb{Q}}\right)^{-1}
$$

Now we can replace the indefinite integral in the equation

$$
\begin{aligned}
A(t,T) &= -(T-t)\rho_0 + \left[\rho_1^\intercal \left(e^{-(K^{\mathbb{Q}})(T-s)}(K^{\mathbb{Q}})^{-1} - sI\right)\left(K^{\mathbb{Q}}\right)^{-1}\right]_t^T K^{\mathbb{Q}}\theta^{\mathbb{Q}} + \frac{1}{2}\mathcal{I} \\
A(t,T) &= -(T-t)\rho_0 + \rho_1^\intercal \left(\left(I - e^{-K^{\mathbb{Q}}(T-t)}\right)(K^{\mathbb{Q}})^{-1} - (T-t)I\right)\theta^{\mathbb{Q}} + \frac{1}{2}\mathcal{I} \qquad (22)
\end{aligned}
$$

Last, the second integral $\mathcal{I}$ can be calculated. First, we change the notation from summation $\sum$ to *trace* using a diagonal matrix $diag(\gamma)$ whose diagonal elements are γ. Second, we replace $B(s,T)$ and $B(s,T)^\intercal$ with the solution previously calculated. Third, the $\Sigma^\intercal$ can be moved to the end using the cyclic permutation property of the trace of a matrix. Last, we can move the integral inside the trace, leaving the time-independent matrix $\Sigma diag(\gamma)\Sigma^\intercal$ outside the integral.

$$
\begin{aligned}
\mathcal{I} &= \int_t^T trace\left(\Sigma^\intercal B(s,T)B(s,T)^\intercal \Sigma diag(\gamma)\right)\,\mathrm{d}s \\
&= \int_t^T trace\left(\Sigma^\intercal \left(e^{-(K^{\mathbb{Q}})^\intercal (T-s)} - I\right)\left((K^{\mathbb{Q}})^\intercal\right)^{-1}\rho_1\rho_1^\intercal\left(K^{\mathbb{Q}}\right)^{-1}\left(e^{-(K^{\mathbb{Q}})(T-s)} - I\right)\Sigma diag(\gamma)\right)\,\mathrm{d}s \\
&= \int_t^T trace\left(\left(e^{-(K^{\mathbb{Q}})(T-s)} - I\right)\Sigma diag(\gamma)\Sigma^\intercal\left(e^{-(K^{\mathbb{Q}})^\intercal (T-s)} - I\right)\left((K^{\mathbb{Q}})^\intercal\right)^{-1}\rho_1\rho_1^\intercal\left(K^{\mathbb{Q}}\right)^{-1}\right)\,\mathrm{d}s \\
&= trace\left(\int_t^T \left(e^{-(K^{\mathbb{Q}})(T-s)} - I\right)\Sigma diag(\gamma)\Sigma^\intercal\left(e^{-(K^{\mathbb{Q}})^\intercal (T-s)} - I\right)\mathrm{d}s\,\left((K^{\mathbb{Q}})^\intercal\right)^{-1}\rho_1\rho_1^\intercal\left(K^{\mathbb{Q}}\right)^{-1}\right) \\
&= trace\left(\left(\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4\right)\left((K^{\mathbb{Q}})^\intercal\right)^{-1}\rho_1\rho_1^\intercal\left(K^{\mathbb{Q}}\right)^{-1}\right)
\end{aligned}
$$

Inside this last expression there are four integrals that can be evaluated individually.

$$
\begin{aligned}
\mathcal{I}_1 &= \int_t^T \left(e^{-(K^{\mathbb{Q}})(T-s)}\right)\Sigma diag(\gamma)\Sigma^\intercal\left(e^{-(K^{\mathbb{Q}})^\intercal (T-s)}\right)\,\mathrm{d}s \\
\mathcal{I}_2 &= -\int_t^T \left(e^{-(K^{\mathbb{Q}})(T-s)}\right)\Sigma diag(\gamma)\Sigma^\intercal\,\mathrm{d}s \\
\mathcal{I}_3 &= -\int_t^T \Sigma diag(\gamma)\Sigma^\intercal\left(e^{-(K^{\mathbb{Q}})^\intercal (T-s)}\right)\,\mathrm{d}s \\
\mathcal{I}_4 &= \int_t^T \Sigma diag(\gamma)\Sigma^\intercal\,\mathrm{d}s
\end{aligned}
$$

The integrals $\mathcal{I}_2$, $\mathcal{I}_3$ and $\mathcal{I}_4$ can be easily solved.

$$
\begin{aligned}
\mathcal{I}_2 &= \left(K^{\mathbb{Q}}\right)^{-1}\left(e^{-K^{\mathbb{Q}}(T-t)} - I\right)\Sigma diag(\gamma)\Sigma^\intercal \\
\mathcal{I}_3 &= \Sigma diag(\gamma)\Sigma^\intercal\left(e^{-(K^{\mathbb{Q}})^\intercal (T-t)} - I\right)\left((K^{\mathbb{Q}})^\intercal\right)^{-1} \\
\mathcal{I}_4 &= (T-t)\Sigma diag(\gamma)\Sigma^\intercal
\end{aligned}
$$

However, $\mathcal{I}_1$ is the product of two matrix exponentials and the symmetric matrix $Q_c = \Sigma diag(\gamma)\Sigma^\intercal$. If Q_c is symmetric and positive semidefinite, we can use the solution provided by Van Loan (1978). First, we substitute the limits of the integration to match the integral in Van Loan (1978). Second, we define a block triangular matrix Z as a function of $K^{\mathbb{Q}}$ and Q_c. Third, the integral can be calculated as the product of two submatrices of $e^{Z(T-t)}$. Here $\left[e^{Z(T-t)}\right]_{1:n,n+1:2n}$ represents the upper-right block matrix of $e^{Z(T-t)}$.

$$
\begin{aligned}
\mathcal{I}_1 &= \int_0^{T-t} \left(e^{-K^{\mathbb{Q}}s}\right) \Sigma diag(\gamma)\Sigma^\intercal \left(e^{-(K^{\mathbb{Q}})^\intercal s}\right) \mathrm{d}s \\
Z &= \begin{bmatrix} K^{\mathbb{Q}} & \Sigma diag(\gamma)\Sigma^\intercal \\ 0 & -(K^{\mathbb{Q}})^\intercal \end{bmatrix} \\
\mathcal{I}_1 &= e^{-K^{\mathbb{Q}}(T-t)} \left[e^{Z(T-t)}\right]_{1:n,n+1:2n}
\end{aligned}
$$

Now we can replace all the integrals $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$ and $\mathcal{I}_4$ into $\mathcal{I}$ to finally obtain the complete expression for $A(t,T)$

$$
\begin{aligned}
A(t,T) &= -(T-t)\rho_0 + \rho_1^\intercal \left(\left(I - e^{-K^{\mathbb{Q}}(T-t)}\right)(K^{\mathbb{Q}})^{-1} - (T-t)I\right)\theta^{\mathbb{Q}} \\
&\quad + \frac{1}{2} trace\Bigg(\Big(e^{-K^{\mathbb{Q}}(T-t)} \left[e^{Z(T-t)}\right]_{1:n,n+1:2n} \\
&\quad + \left((K^{\mathbb{Q}})^\intercal\right)^{-1} \left(e^{-(K^{\mathbb{Q}})^\intercal(T-t)} - I\right) \Sigma diag(\gamma)\Sigma^\intercal \\
&\quad + \Sigma diag(\gamma)\Sigma^\intercal \left(e^{-K^{\mathbb{Q}}(T-t)} - I\right) \left(K^{\mathbb{Q}}\right)^{-1} \\
&\quad + (T-t)\Sigma diag(\gamma)\Sigma^\intercal\Big) \\
&\quad \Big(\left((K^{\mathbb{Q}})^\intercal\right)^{-1} \rho_1 \rho_1^\intercal \left(K^{\mathbb{Q}}\right)^{-1}\Big)\Bigg)
\end{aligned}
$$

B Coupon Bonds

In the case of a coupon bond, we assume each of the cashflows C_{t_i} received at time t_i can be considered as a zero-coupon bond. We can calculate the value of a single risky cash-flow V_{t_i} using the approach described in Section 3.2. Thus, the final value V of the coupon bond is the sum of each of the coupon values.

$$V(C_{t_1}, \ldots, C_{t_n}) = \sum_{t_i=t_1}^{t_n} V_{t_i} C_{t_i} \tag{23}$$

C Kalman Filter Setup

Here we explain the estate estimation of the dynamic system in equations 16-17. We adapt the Square-Root Unscented Kalman Filter (SRUKF) of Van der Merwe and Wan (2001) to take into account that the conditional moments and the measurement equation need to be modified in each iteration to adapt to the characteristics of each individual trade (transaction side, transaction size, time between observations, and time to maturity).

In our setting, having a large number of transactions significantly increases the computational time to search for the optimal parameters in the QML estimation. In order to speed up the search for the optimal estimates, we proceed as follows.

1. We run a rolling-window regression for each bond to initialize the parameter vector $\hat{\Theta}^0 = (\hat{K}^0, \hat{\theta}^0, \hat{\Sigma}^0)$.

2. We employ a Dual Kalman filter for learning the parameters. We run the Dual Kalman multiple times recursively using as initial parameters those obtained in a previous step. With each run we obtain improved estimates $\hat{\Theta}^1$.

3. We choose from the previous step the set of parameters $\hat{\Theta}^{QML}$ that yield the best likelihood.

4. We extract the transaction cost factors using the final estimates $\hat{\Theta}^{QML}$ employing a Kalman filter for state estimation.

C.1 Initializing the parameter vector

Taking logs on both sides of expression 1 we can calculate the log return r_{t+h} between a customer trade at time $t+h$ and a previous customer trade at time t. We then replace $log(1-Qc)$ with the solution provided in Proposition 1.

$$
\begin{aligned}
r_{t+h} &= log(V_{t+h}/V_t) - log(1 - Q_{t+h}c_{t+h}) + log(1 - Q_t c_t) \\
r_{t+h} &= log(V_{t+h}/V_t) - B(t+h,T)^{\intercal} X_{t+h} + B(t,T)^{\intercal} X_t + A(t,T) - A(t+h,T) \\
r_{t+h} &= log(V_{t+h}/V_t) - \rho_{1,t+h}^{\intercal} \left(\left(e^{-K(T-t-h)} - I \right) K^{-1} \right) X_{t+h} + \rho_{1,t}^{\intercal} \left(\left(e^{-K(T-t)} - I \right) K^{-1} \right) X_t \\
&+ A(t,T) - A(t+h,T) \\
r_{t+h} &= log(V_{t+h}/V_t) - \rho_{1,t+h}^{\intercal} \left(\left(e^{-K(T-t-h)} - I \right) K^{-1} \right) X_{t+h} + \rho_{1,t}^{\intercal} \left(\left(e^{-K(T-t)} - I \right) K^{-1} \right) X_t \\
&+ A(t,T) - A(t+h,T) \qquad (24)
\end{aligned}
$$

To obtain initial parameters we create simple discrete approximation of the log-returns in equation 24. This approximation ignores arbitrage-free principles. The only purpose of this approximation is to measure initial values for the parameter vector and save computational time in the QML estimation.

When we compare two adjacent trades, a large portion of the price difference will come from the transaction side (buy or sell) and the transaction size, i.e. from the observable $\rho_{1,t}$

vector of a given trade. As time span between two adjacent trades is generally small ($h \to 0$) we can approximate the log-return by

$$r_{t+h} \approx log(V_{t+h}/V_t) + \left[\rho^{\intercal}_{1,t+h} - \rho^{\intercal}_{1,t}\right] \left(\left(I - e^{-K(T-t)}\right) K^{-1}\right) X_t + A(t,T) - A(t+h,T)$$

The coefficient $A(t,T)$ mainly captures long-term effects and affects the long-term yields. Since the coefficient usually decreases smoothly as we approach maturity, the difference $A(t,T) - A(t+h,T)$ tends to be small for two adjacent trades. Thus, we approximate it by a constant and an error term. The efficient value log-return needs a cumbersome calculation in the case of coupon bonds that can be simplified. For simplicity, we assume that the value growth can be replaced with the growth $(r_f + CS) \cdot (h)$ of an equivalent zero-coupon bond with average yield to maturity of $r_f + CS$. The yield to maturity is decomposed into a risk-free rate r_f and a credit spread CS and the excess log-return approximation becomes

$$r_{t+h} - r_f(h) \approx \alpha + CS(h) + \left[\rho^{\intercal}_{1,t+h} - \rho^{\intercal}_{1,t}\right] \left(\left(I - e^{-K(T-t)}\right) K^{-1}\right) X_t + \varepsilon_{t+h}$$

or equivalently the following regression

$$r_{t+h} - r_f(h) = \alpha + CS(h) + \left[\rho^{\intercal}_{1,t+h} - \rho^{\intercal}_{1,t}\right] \beta + \varepsilon_{t+h}$$

were $\beta_i = \frac{1-e^{-\kappa_i(T-t)}}{\kappa_i} X_{t,i}$. We run this regression using non-overlapping rolling-windows and obtain a time series of $\widehat{CS}$ and $\hat{\beta}$ estimates. It is straightforward to see that the beta estimates contain information about the level of the state factors as well as the mean-reversion and the time to maturity. Thus, an empirical model based on this regression will

lead to betas that decrease as we approach maturity, consistent with empirical evidence found in the related literature.

We invert the betas, and represent the state values as a function of the betas.

$$\begin{aligned} X_{i,t+h} &= \beta_{i,t+h}\kappa_i \frac{1}{1-e^{-\kappa_i(T-t-h)}} \\ X_{i,t} &= \beta_{i,t}\kappa_i \frac{1}{1-e^{-\kappa_i(T-t)}} \end{aligned}$$

To construct a first estimate of the mean-reversion parameters, we start from the discretization of the state vector in equation 20. We replace the estate value with the betas, and we get an expression that depends on the time series of betas.

$$\begin{aligned} X_{i,t+h} &= (1-e^{-\kappa_i h})\theta + e^{-\kappa_i h} X_t + \eta_{i,t+h} \\ \beta_{i,t+h}\kappa_i \frac{1}{1-e^{-\kappa_i(T-t-h)}} &= (1-e^{-\kappa_i h})\theta + e^{-\kappa_i h}\beta_{i,t}\kappa_i \frac{1}{1-e^{-\kappa_i(T-t)}} + \eta_{i,t+h} \end{aligned}$$

where $\eta_{i,t+h}$ are iid normal with zero mean and variance $\sigma_\eta^2 = \sigma^2(1-e^{-2\kappa_i h})/(2\kappa_i)$.

We organize the terms in the last expression and we obtain an autoregressive representation of the betas

$$\begin{aligned} \beta_{i,t+h} &= (1-e^{-\kappa_i h})\theta + e^{-\kappa_i h}\beta_{i,t}\kappa_i \frac{1}{1-e^{-\kappa_i(T-t)}} + \eta_{i,t+h} \\ \beta_{i,t+h} &= \underbrace{\frac{1-e^{-\kappa_i(T-t-h)}}{\kappa_i}(1-e^{-\kappa_i h})\theta}_{b_0} + \underbrace{e^{-\kappa_i h}\frac{1-e^{-\kappa_i(T-t-h)}}{1-e^{-\kappa_i(T-t)}}}_{b_1}\beta_{i,t} + \underbrace{\frac{1-e^{-\kappa_i(T-t-h)}}{\kappa_i}\eta_{i,t+h}}_{\varepsilon_{i,t+h}} \end{aligned}$$

We then start the search for the parameter vector using

$$
\begin{aligned}
\kappa_i^0 &= b_1^{-1}(\hat{b_1}) \\
\theta_i^0 &= \hat{b}_0 \frac{\kappa_i^0}{(1-e^{-\kappa_i^0(T-t-h)})(1-e^{-\kappa_i^0 h})}
\end{aligned}
$$

C.2 Dual Kalman filter for state and parameter estimation

In a second step, the Dual Kalman filter is used to learn about the parameters $w \in \Theta$. The Dual Kalman filter requires setting a new state-space system for the parameters. We run sequentially two UKFs: one for estate estimation at time k using previous parameter estimates $\hat{w}_{k-1}$, and other for parameter learning using the current state estimate $\hat{X}_k$.

$$
\begin{aligned}
X_k &= F(X_{k-1}, \eta_k) \\
y_k &= G(X_k, \hat{w}_{k-1}) + \varepsilon_k
\end{aligned}
$$

In the parameter UKF, the parameters w_k are a stationary process with noise r_k, and the observation y_k is a nonlinear transformation of the parameters given an estimate of the state vector.

$$
\begin{aligned}
w_k &= w_{k-1} + r_k \\
y_k &= G(\hat{X}_k, w_k) + e_k
\end{aligned}
$$

We employ the Square-Root Unscented Kalman filter outlined in the next section C.3 for both the state and the parameter systems. The time-update step in the case of the SRUKF

for parameter estimation greatly simplifies the computation because the system is linear. Particularly, the predicted parameter is $\hat{w}_k = \hat{w}_{k-1}$ and the noise square-root covariance becomes $S_k^- = \sqrt{\lambda_{RLS}} S_{w_{k-1}}$, where λ_{RLS} is a scalar weighting factor slightly smaller than 1 that accounts for the dependence from the previous covariance.

C.3 Square-root Unscented Kalman filter

This section outlines the Square-Root Unscented Kalman filter employed for state and parameter estimation.

First, we initialize with the unconditional moments from equations 18 and 19 using an initial set of parameters $\hat{\Theta}$. In the case of the Dual Kalman filter for parameter learning we start with $\hat{\Theta}^0$ and iteratively obtain a new set of improved estimates. In the final step, we would start with the optimal parameters $\hat{\Theta}^{QML}$ run the following filter.

$$\hat{X}_0 = E[X_0] \qquad S_0 = chol\left\{E[(X_0 - \hat{X}_0)(X_0 - \hat{X}_0)^\intercal]\right\} \tag{25}$$

For $k \in \{1, \ldots, \text{last observation}\}$,

The process noise covariance and the measurement noise covariance in each iteration are:

$$R_k^v = Q_{t+h} \tag{26}$$

$$R_k^n = H_{t+h} \tag{27}$$

Calculate sigma points:

$$\mathcal{X}_{k-1} = [\hat{X}_{k-1} \quad \hat{X}_{k-1} + \gamma S_k \quad \hat{X}_{k-1} - \gamma S_k] \tag{28}$$

Time update:

$$\begin{aligned}
\mathcal{X}^*_{k-1} &= F[\mathcal{X}_{k-1}] && (29)\\
\hat{X}^-_k &= \sum_{i=0}^{2L} W_i^{(m)} \mathcal{X}^*_{i,k|k-1} && (30)\\
S^-_k &= qr\left\{\left[\sqrt{W_1^{(c)}}(\mathcal{X}^*_{1:2L,k|k-1} - \hat{X}^-_k) \quad \sqrt{R^v_k}\right]\right\} && (31)\\
S^-_k &= cholupdate\left\{S^-_k \quad , \quad \hat{X}^*_{0,k} - \hat{X}^-_k \quad , \quad W_0^{(c)}\right\} && (32)\\
\mathcal{X}_{k|k-1} &= [\hat{X}^-_k \quad \hat{X}^-_k + \gamma S^-_k \quad \hat{X}^-_k - \gamma S^-_k] && (33)\\
\mathcal{Y}_{k|k-1} &= H[\mathcal{X}_{k|k-1}] && (34)\\
\hat{Y}^-_k &= \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}^*_{i,k|k-1} && (35)
\end{aligned}$$

Measurement update equations:

$$\begin{aligned}
S_{\hat{Y}_k} &= qr\left\{\left[\sqrt{W_1^{(c)}}(\mathcal{Y}_{1:2L,k} - \hat{Y}^-_k) \quad \sqrt{R^n_k}\right]\right\} && (36)\\
S_{\hat{Y}_k} &= cholupdate\left\{S_{\hat{Y}_k} \quad , \quad \mathcal{Y}^-_{0,k} - \hat{Y}^-_k \quad , \quad W_0^{(c)}\right\} && (37)\\
P_{X_kY_k} &= \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{X}_{i,k|k-1} - \hat{X}^-_k][\mathcal{Y}_{i,k|k-1} - \hat{Y}^-_k]^\intercal && (38)\\
\mathcal{K}_k &= (P_{X_kY_k}/S^\intercal_{\hat{Y}_k})/S_{\hat{Y}_k} && (39)\\
\hat{X}_k &= \hat{X}^-_k + \mathcal{K}_k(Y_k - \hat{Y}^-_k) && (40)\\
U &= \hat{X}_k S_{\hat{Y}_k} && (41)\\
S_k &= cholupdate\left\{S^-_k \quad , \quad U \quad , \quad -1\right\} && (42)
\end{aligned}$$

D Databases

We briefly describe how we have created the dataset used in this paper using TRACE, FISD, and COMPUSTAT. The stock tickers provided in these datasets cannot be used to merge COMPUSTAT with TRACE. There are usually discrepancies across datasets due to mergers, spin offs, and other reasons. Another difficulty is that the CUSIP number from each bond may not identify the company, but a subsidiary. As a result, many companies have issued bonds under different CUSIPs.

The companies are manually merged. First, we download the firms members of S&P by the stock Ticker. Second, we manually select from TRACE and FISD the bond CUSIPs that belong to each stock Ticker. Third, we manually find the PERMNO that identifies each stock in CRSP. Fourth, we manually find the GVKEY that identifies each company in COMPUSTAT.

- **TRACE database**. We download all the corporate bond transactions available for the firms members of the S&P500. We use all trades in TRACE Enhanced until the end of 2012, and we use TRACE database from 2013 onwards. Similarly to Rossi (2014) and Back and Crotty (2014), we follow the next steps to clean the dataset:

    1. Clean same-day corrections and cancelations. The database identifies each transaction with a message sequence number that is unique in each reporting day. The correction and cancelation reports also include a message sequence number that refers to the original report. We delete the original trade report identified by the message sequence number, delete the cancelation reports, and replace the original reports with the matched correction report.
    2. Reversal trades. A Reversal is a cancellation of a trade report originally submitted on a previous day. We remove the reversal trades and the matching original report. We identify the Reversal trades and the original trade matching trades with the

same CUSIP, Quantity trade, Price, Execution Date, Execution Time, Buy/Sell Indicator, and Contra Party Indicator. Often, trades cancelled with a Reversal trade are then replaced with an As-of trade. We keep the 'As-Of' trades that correct cancelled trades when available. The remaining 'As-Of' trades that have not been matched are removed because they have originated after the execution date.

3. Delete the buying side of Inter-dealer transactions and keep the Inter-dealer Sells, because both reflect two sides of the same trade.

4. Delete unusual trades. We eliminate trades that include commission, Halt trades, trades reported after hours, delayed dissemination trades, reversals of delayed dissemination trades, special trade condition trades, trades that were not disseminated and trades on a 'When issued' basis. We keep trade corrections that were reported after hours.

5. Eliminate extreme prices. First, we keep observations with prices between \$1 and \$500. Second, we delete observations of a bond with a price that deviates more than 20% from the median price of the day. Third, we eliminate observations that within a day are preceded and followed by a 20% change in price. Fourth, we keep trades that satisfy

$$|p - med(p,k)| \leq 5 * MAD(p,k) + \$1$$

where med and MAD are the centered median and median absolute deviation of $k = 20$ observations with prices surrounding a given transaction price p.

6. Keep bonds traded on 20 distinct days.

- **FISD database**. We use the files MATOUT, AOUTHIST, COUPINFO, ISSCUSIP, ISSUE, ISSUER, ISUCUHST, NAMEHIST, RATING, and RATNGHST.

1. We merge the filtered bond transactions with FISD files using the CUSIP
2. Drop convertible debt, putable debt, redeemable debt, and variable coupon debt.
3. We calculate Accrued Interest. We calculate the accrued interest until the settlement date. Often the bonds trade with T+3 settlement, meaning that the settlement takes place 3 business days after the trade execution. Business days are from Monday to Friday with the exception of the holidays and observations announced by FINRA.
4. We calculate the average rating by firm. We translate the ratings to a numeric scale with a score of 21 for the highest rating (Aaa and AAA) and a score of 1 for the lowest rating (C, D, DD and DDD). First, we calculate the daily average rating of a bond using the ratings from Moody's, S&P and Fitch. Second, we calculate the average rating of a firm weighting the bond ratings with the outstanding amount. Third, we apply a single-exponential smoothing with an alpha of 10% on the weighted average rating to alleviate the large jumps in the average credit rating due to new bonds entering or old bonds exiting the portfolio of outstanding bonds of a firm.

- **COMPUSTAT database**. We merge TRACE identifiers with COMPUSTAT to obtain accounting information from issuing firms.