The risky bond is traded in a competitive bond market with noisy supply s_t, which is a public knowledge. The market is populated with two classes of investors, $i = 1, 2$, with relative population weights ω and $1 - \omega$. The investors are identical within each class, and each investor's initial endowment of the risky bond is set to 0 for simplicity. Type 1 investors are informed; they observe g_t. Type 2 investors do not observe g_t but learn it from the bond price using the Bayes rule. In addition, Type 1 investors have a random exposure z_t to some non-traded asset that generates a log return of n_{t+1} in the subsequent period.[25] Type 2 investors do not know the exposure of type 1 investors to the non-traded risk. Overall, the information set of the informed investors at time t is $\{d, p, n, f, s, g, z\}_{0,\dots,t}$ while the information set of the uninformed investors is $\{d, p, n, f, s\}_{0,\dots,t}$.

I assume that n_t, g_t, and z_t are time-independent zero-mean normally distributed random variables with variances σ_n^2, σ_g^2, σ_z^2 respectively. I further assume that f_t is also time-independent and normally distributed with the mean $m_f = \kappa + c(1 - \theta)$ and the variance σ_f^2.[26] All of n_t, g_t, z_t, and f_t are contemporaneously uncorrelated except for n_t and f_t that have a time-invariant negative covariance, which means that default losses are low when non-traded asset returns are high. This implies a constant positive covariance between r_t and n_t that equals σ_{rn}. Finally, the supply of the risky bond follows an AR(1) process

$$s_{t+1} = \delta s_t + \epsilon_{t+1}, \tag{5}$$

where $|\delta| < 1$ and ϵ_t is normally distributed with zero mean and variance σ_s^2; it is independent over time and is independent from n_t, g_t, z_t, and f_t.

The investors of both types $i = 1, 2$ maximize the next period conditional expected utility $\mathbb{E}_t\left[-e^{-W_{t+1}^{(i)}}\right]$ (derived from the next period wealth $W_{t+1}^{(i)}$ by choosing the demand $X_t^{(i)}$ for the risky bond.[27] To keep the model tractable I need to take the log-linear approximation

[25]Here I follow Llorente et al. (2002) assuming for simplicity that only one type of investors has income from a non-traded asset. It is enough to generate price reversals due to liquidity trading.

[26]The mean of f_t is chosen such that the long-term mean of the log bond price is 0 and the contributions of coupons and public news about future defaults to returns cancel one another on average.

[27]As in Llorente et al. (2002), the risk aversion is set to 1 since it only enters the expressions for investors' demands as the multiple of the variances of all exogenous shocks. Hence, one can implement higher or lower risk aversion in the model by proportionally scaling variances of all shocks up or down.

of the wealth dynamics, which under the assumptions of the model is

$$W_{t+1}^{(1)} \approx W_t^{(1)} + X_t^{(1)} r_{t+1} + z_t(1 + n_{t+1}),$$
$$W_{t+1}^{(2)} \approx W_t^{(2)} + X_t^{(2)} r_{t+1}.$$

The model setup is different from Llorente et al. (2002) in two ways. First, I work with log returns approximated in (4) around $\bar{p} \equiv 0$ and linearized wealth dynamics instead of dollar returns and non-linearized wealth dynamics. Second, more importantly, I assume noisy supply (5) instead of a constant zero supply. Noisy supply allows me to decompose the trading volume in the model into two components: the first one is related to trading between informed and uninformed investors and exogenous changes in asset supply drive the second one. Empirical counterparts of these two components are respectively the volume of corporate bonds purchased by clients matched by client sales in a given period and net changes in broker-dealer inventory.

B. Model equilibrium

I solve for the rational expectations equilibrium of the model assuming a linear pricing function for the log bond price. Define the log price adjusted for the publicly known credit loss component as $\tilde{p}_t \equiv p_t + (f_t - m_f)$ and assume it is linear with respect to g_t, z_t, and s_t:

$$\tilde{p}_t = -a(g_t + bz_t + es_t). \tag{6}$$

Observe that the steady-state level of log bond price is 0 as in the linear approximation of log return (4).

Given the pricing function (6), the equation for returns (4) re-writes as:[28]

$$r_{t+1} = -\theta\left(f_{t+1} - m_f\right) + \theta\tilde{p}_{t+1} - \tilde{p}_t - g_t. \tag{7}$$

The expression for conditional expected returns follows from (7):

$$\mathbb{E}_t^{(i)}\left[r_{t+1}\right] = -\tilde{p}_t - \mathbb{E}_t^{(i)}\left[g_t\right] - ae\theta\delta s_t.$$

[28] In what follows, I replace an approximate equality in (4) with the exact one.

The informed investors know g_t, hence $\mathbb{E}_t^{(1)}[g_t] = g_t$. The uninformed investors observe $\tilde{p}_t$ and s_t and estimate $\mathbb{E}_t^{(2)}[g_t|\tilde{p}_t, s_t]$. I show in Appendix A that

$$\mathbb{E}_t^{(2)}[g_t|\tilde{p}_t, s_t] = \gamma(g_t + bz_t), \tag{8}$$

where $\gamma = \frac{\sigma_g^2}{\sigma_g^2 + b^2\sigma_z^2} > 0$. One can further show that conditional return variances for two types of investors are constant over time.

With conditional expected return linear in g_t, z_t, and s_t and conditional return variance constant for both types of investors, the demand for risky bonds, $X_t^{(1)}$ and $X_t^{(2)}$, is also linear in g_t, z_t, and s_t[29]. The market for risky bonds clears:

$$\omega X_t^{(1)}(g_t, z_t, s_t) + (1-\omega)X_t^{(2)}(g_t, z_t, s_t) = s_t,$$

which must hold for any values of g_t, z_t, and s_t, implying a system of three non-linear equations for yet undetermined coefficients a, b, and e. One can show that if the parameters of the model are such that the system has real-valued solutions then it must be that a, b, and e are all positive, moreover, $\omega + \gamma - \omega\gamma < a < 1$ and $b = \sigma_{rn}$. I demonstrate in Appendix A that under mild restrictions on the parameters (that boil down to σ_s^2 being not 'too big') the model always has real-valued solutions, of which a unique triple of $\{a^*, b^*, e^*\}$ has economically reasonable values.

C. *Trading volume in the model*

Consider the aggregate difference in risky bond holdings in the economy at time t

$$\omega \Delta X_t^{(1)} + (1-\omega)\Delta X_t^{(2)} = \Delta s_t.$$

Using the equilibrium conditions one can decompose it as

$$\omega \Delta X_t^{(1)} + (1-\omega)\Delta X_t^{(2)} = \underbrace{V_{c,t}^{(1)}(\Delta g_t, \Delta z_t) + V_{c,t}^{(2)}(\Delta g_t, \Delta z_t)}_{=0} + \underbrace{V_{s,t}^{(1)}(\Delta s_t) + V_{s,t}^{(2)}(\Delta s_t)}_{=\Delta s_t},$$

where

$$V_{c,t}^{(1)}(\Delta g_t, \Delta z_t) \;=\; V_{c,t}^{(2)}(\Delta g_t, \Delta z_t) \;= |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)|, \tag{9}$$

[29] See Appendix A.

and $\alpha = {}^{\omega(a-1)}/_{\sigma_r^2}$. Here, $V_c^{(1)}$ and $V_c^{(2)}$ represent the volume of trading *between* informed and uninformed investors. This trading volume is due to changes in a private signal about credit loss Δg (information-driven trading) and the position in a non-traded asset Δz (liquidity-driven trading). $V_c^{(1)}$ and $V_c^{(2)}$ always have opposite signs but are equal in the absolute value. For the convenience of notation, I will denote this trading volume $v_{c,t} = |\alpha (\Delta g_t + \sigma_{rn} \Delta z_t)| \geqslant 0$. An econometrician observing bond trading records in the TRACE database can compute what the client buy volume matched by the client sell volume was at time t.[30] It is an empirical proxy for $v_{c,t}$.

Two other components, $V_s^{(1)}$ and $V_s^{(2)}$, represent trading due to changing bond supply. One can show that in equilibrium these two components are always of the same sign and they represent the proportion in which two types of agents absorb additional bond supply Δs. By construction, a change in bond supply is the buy volume that was not matched by the sell volume of the opposite sign. Its absolute value is equal to the absolute value of a change in aggregate dealers' inventory. The latter is an empirical counterpart of $v_{s,t} \equiv |\Delta s_t|$. What the model assumes is that $v_{c,t}$ and Δs_t are independent since the latter is uncorrelated with Δg and Δz that drive the former. Table III has demonstrated that this assumption largely holds in the data. The key takeaway of this paragraph is that I assume that an econometrician knows $v_{c,t}$ and $v_{s,t}$, and these two quantities are defined within the model as stated above.

D. *Volume-return relationship and information asymmetry*

Assume an econometrician observes the time-series of bond returns r_t and two types of volume, $v_{c,t}$ and $v_{s,t}$, as discussed above. Then the conditional expectation of future returns given current returns and volume can be approximated as

$$\mathbb{E}_t \left[r_{t+1} | r_t, v_{c,t}, v_{s,t} \right] \approx \left(\beta_1 + \beta_2 v_{c,t}^2 + \beta_3 v_{s,t}^2 \right) r_t, \tag{10}$$

[30] All records in TRACE represent trading *between* a broker-dealer and a client and can be of two types only: a purchase by a client from a dealer or a sale to a dealer.

the derivation is presented in Appendix A. This volume-return relationship is a theoretical counterpart of equation (2) estimated in the empirical part of the paper. Unlike equation (2), equation (10) contains squared volumes. In the data, squared volumes are extremely right-skewed, hence from an econometric standpoint, it is reasonable to estimate the volume-return relationship as in (2) with volume entering the equation without a square (Llorente et al. 2002 follow the same approach). It does not change an economic interpretation of volume-return coefficients.[31]

Now, I would like to discuss how coefficients β_1, β_2, and β_3 change in the model as the extent of informed trading changes. In the benchmark model Llorente et al. (2002), both β_1 and β_2 are negative, but β_1 is decreasing and β_2 is increasing with the extent of information asymmetry proxied by σ_g^2. β_1 measures the first return autocorrelation, and negative β_1 decreasing with σ_g^2 means that for two equally risky bonds returns will *revert more* for the one with more information asymmetry. β_2 measures the impact of volume on the first autocorrelation, and negative β_2 increasing with σ_g^2 means that for two equally risky bonds returns will *revert less following high-volume days* for the one with more information asymmetry. These theoretical results find empirical support in the U.S. stock market, as Llorente et al. (2002) shows.

Unlike in the benchmark model, I can not make a general statement about the signs of volume-return coefficients and their dependence on σ_g^2; I need to solve the model numerically first. In Figure 6, I present the relationships between information asymmetry σ_g^2 and β coefficients for the model calibrated to an average bond in TRACE. The bond has a coupon rate of 5%, high persistence of a supply shock $\delta = 0.95$, and a daily standard deviation of returns of 1%.[32] The latter stays fixed in all numerical solutions; this is an additional

[31] Since an econometrician knows the sign of inventory changes, she could write an analog of equation (10) conditioning additionally on this piece of knowledge. It would change the form of the equation slightly, and the loadings on two types of volume would become incomparable. An important part of my empirical analysis consists of a direct comparison of coefficients β_2 and β_3, and for that, I need to condition in (10) on the absolute value of inventory changes.

[32] In Figure 6, I set $\delta = 0.95$ which roughly corresponds to $\text{Corr}(\Delta s_t, \Delta s_{t-1}) = -0.03$ because in the model $\text{Corr}(\Delta s_t, \Delta s_{t-1}) = -\frac{1}{2}(1-\delta)$. In the model, δ measures the persistence of supply, which is roughly the persistence of inventory. $\delta = 0.95$ implies the half-life of broker-dealer inventory of about 13 days. Further

constraint I impose on the solutions of the model.[33] Figure 6 represents the cross-section of bonds with the same unconditional risk but different contributions of public, private, and liquidity shocks to return variance.

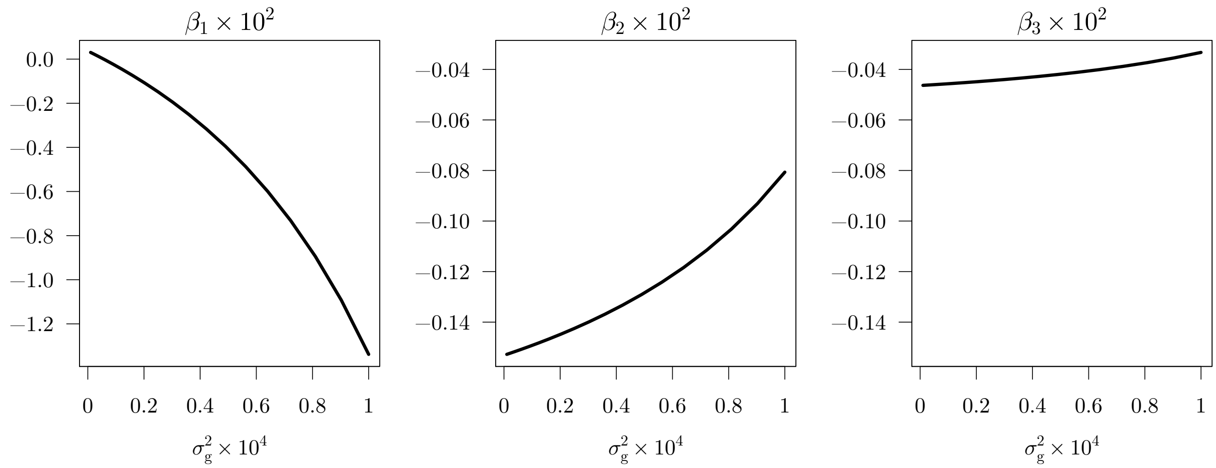

Figure 6. Dependence of β_1, β_2, and β_3 on information asymmetry σ_g^2 holding total return variance fixed. Each point on the curves is a numerical solution of the model. I obtain the relationships between σ_g^2 and β coefficients by varying σ_g from 0 to 1% holding an unconditional standard deviation of returns at 1%, which is a daily standard deviation of bond returns in the TRACE data. I choose the following parameters of the model to match a median bond in sample: coupon rate $C = 5\%$, the persistence of a supply shock $\delta = 0.95$. The fraction of informed investors is $\omega = 0.05$, the correlation between traded and non-traded asset returns is $\sigma_{rn} = 0.3$, the variance of the supply shock is $\sigma_s^2 = 0.1$. I first solve the model for a very small value of σ_g, 5 b.p. here. Then, I hold the equilibrium value of a fixed in all subsequent solutions for $\sigma_g > 5$ b.p; I allow e to change. Thus, the comparative statics plotted here is a collection of solutions of the system of equations of three variables (σ_z^2, σ_f^2, and e): two model equilibrium equations plus an additional restriction on the total return variance.

The left and central panels in Figure 6 deliver the same message as the benchmark model. With more informed trading, returns tend to revert more, but less so following days when investors trade a lot with each other. On the left panel, which presents reversals following no-volume days, there is no reversal when σ_g is zero, and returns are due to public news that is fully priced within the same period. As σ_g increases, no-volume reversals intensify

(unreported) estimations show, in line with the results of Dick-Nielsen and Rossi (2018), that dealers revert deviations from their target inventory faster post-crisis.

[33] Llorente et al. (2002) impose the same restriction on the total unconditional variance of returns.

due to a greater impact of uninformed investors' errors in estimating g_t on returns.[34] On the central panel, the reversal following high-volume days is the strongest when σ_g is zero because the entire trading volume between informed and uninformed investors represents, in this case, liquidity trading. Liquidity trading has price impact but does not reveal any new information about the asset payoff; hence, the price reverts the next period. As σ_g increases, it's more and more likely that some part of the between-investors trading volume comes from Δg and conveys the information about future returns; hence the reversal tends to decrease (β_2 tends to increase). The right panel in Figure 6 shows that β_3 that measures an additional component of reversals following days when inventory changes a lot is relatively insensitive to σ_g. It does not look surprising given that Δs in the model is uncorrelated with other motives for trading. One would expect β_3 to be flat with respect to σ_g in such case; a slightly upward sloping line on the right panel of Figure 6 is due to equilibrium e (price impact of inventory-changing trades) changing with σ_g.

The shape of the lines in Figure 6 matches closely the shape of their empirical counterparts presented in Figure 3. In the model, as it is in the data, β_1 decreases, and β_2 increases with information asymmetry, while β_3 is insensitive to information asymmetry. It gives additional support for the premises of the model: client-to-client trading volume may be due to private information, but client-to-dealer trading volume is likely driven by liquidity needs only.

As in Llorente et al. (2002), the limitation of my extended model is that β_2 stays negative for all reasonable model calibrations and does not turn positive (same applies to β_3 which is not the part of the benchmark model). In reality, as Section III has shown, β_2 is positive for most corporate bonds. It does not undermine the main idea suggested by the model and tested in the empirical part of the paper. As the extent of informed trading increases,

[34] Here is the intuition for this result. With no volume, time t returns are not driven by liquidity shocks since Δz_t and Δg_t must be zero. Assume $z_{t-1} > 0$ and informed investors are net sellers of bonds. From (7) and (8) one finds that r_t is negative when $\frac{a}{\gamma}\mathbb{E}^{(2)}_{t-1}[g_{t-1}] < g_{t-1}$ other things being equal, i.e., when actual losses in default are higher than previously expected by uninformed investors. But that means that in $t-1$ informed investors' demand for bonds was lower than required by their hedging needs; so it is in t since the volume is zero. Hence, time t price is low and time $t+1$ expected return is high. Higher information asymmetry amplifies this effect.

returns following high-volume days are less likely to revert, especially when dealers are not trading from their inventory capacity.

VII. Conclusion

In this paper, I estimate a dynamic volume-return relationship for individual bonds and explore the determinants of estimated volume-return coefficients in a cross-section of bonds. A particular focus of my analysis is on the impact of information asymmetry on volume-return coefficients.

The hypotheses that I test arise from a stylized theoretical model of competitive bond trading with asymmetric information and non-traded risks. In the model, trading between investors is due to liquidity needs (hedging of the non-traded risk) or private information. Also, investors in the model absorb random bond supply shocks; their empirical counterpart is the change in aggregate bond inventory. The model suggests that bonds with high information asymmetry have stronger price reversals than bonds with low information asymmetry, but less so following high-volume days when dealers' inventory *does not change*, and investors are essentially trading with each other. Conversely, following days with substantial changes in dealers' inventory, the difference in reversals between high- low-asymmetry bonds remains. In the model, this result emerges because changes in inventory (supply shocks) are assumed independent from the arrival of private news.

I find strong empirical support for model predictions in the data. Bonds with high information asymmetry exhibit stronger price reversals than low-asymmetry bonds, but less so following days when trading volumes are high, but dealers' inventory does not change at the end of the day (clients purchases equal client sales). High-asymmetry bonds in my analysis are the bonds that are owned by few mutual funds and intermediated by few dealers, have smaller outstanding amounts and issued by smaller firms with no actively traded CDS contract on the issuer and high stock return volatility.

In particular, I find that a typical bond with high information asymmetry has the first autocorrelation of returns close to -0.4 following average-volume trading days. Following two standard deviations above-average volume day when dealers' inventory does not change, the first autocorrelation reduces to -0.18. A similar bond with *the same* average realized bid-ask spread, return volatility, credit spread, and volume autocorrelation, but low information asymmetry has the first return autocorrelation of -0.2, which increases only by 0.05 to -0.15 following high-volume inventory-neutral days.

If one considers, instead, the reversals following days when trading volume is high, but it is due to substantial changes in dealers' inventory, then the difference in reversals between bonds with high and low information asymmetry remains at the average-volume day level. These results are consistent with the assumption that trading volume in high-asymmetry bonds is more likely to come from investors who possess private information. Since dealers typically know their clients well and might be able to detect informed investors, they let other investors provide liquidity for such trades. Overall, my results suggest that there might be informed trading in corporate bonds, but when it happens, dealers are not providing liquidity and are not adversely selected.

My findings have implications for the design of investment strategies exploiting corporate bond reversals. In particular, I show that long-reversal portfolios of high-asymmetry bonds outperform long-reversal portfolios of low-asymmetry bonds both before and after adjustment for trading costs. Hence, illiquidity does not fully explain reversal returns. Moreover, reversal portfolios of high-asymmetry bonds outperform the corporate bond market after trading cost adjustment. An investor considering an implementation of a bond reversal strategy might profit from additionally sorting bonds on information asymmetry proxies.

My results also relate to a recent policy debate about corporate bond market transparency. I find that bonds with less transparent valuations tend to have stronger price reversals when trading is purely liquidity-driven, and fundamental values of the bonds likely remain unchanged. Stronger liquidity-driven reversal is just another name for non-fundamental price volatility that is often regarded as an undesirable feature of a well-functioning financial market. From this standpoint, a proposed reduction in corporate bond market transparency (TRACE delayed trade dissemination pilot project) might not be optimal.

REFERENCES

Adrian, Tobias, Nina Boyarchenko, and Or Shachar, 2017, Dealer balance sheets and bond liquidity provision, *Journal of Monetary Economics* 89, 92 – 109.

Asquith, Paul, Andrea S. Au, Thomas Covert, and Parag A. Pathak, 2013, The market for borrowing corporate bonds, *Journal of Financial Economics* 107, 155 – 182.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2019, Common risk factors in the cross-section of corporate bond returns, *Journal of Financial Economics* 131, 619 – 642.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2018, Long-term contrarian strategies in the corporate bond market, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978861.

Bao, Jack, Maureen O'Hara, and Xing (Alex) Zhou, 2018, The Volcker Rule and corporate bond market making in times of stress, *Journal of Financial Economics* 130, 95 – 113.

Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *The Journal of Finance* 66, 911–946.

Barbon, Andrea, Marco Di Maggio, Francesco A. Franzoni, and Augustin Landier, 2018, Brokers and order flow leakage: Evidence from fire sales, Swiss Finance Institute Working Paper no. 17-61, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2991617.

Berndt, Antje, and Yichao Zhu, 2018, Dealer inventory, short interest and price efficiency in the corporate bond market, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3189980.

Bessembinder, Hendrik, Stacey Jacobsen, William Maxwell, and Kumar Venkataraman, 2018, Capital commitment and illiquidity in corporate bonds, *The Journal of Finance* 73, 1615–1661.

BlackRock, 2018, The next generation bond market, SEC Fixed Income Market Structure Advisory Committee (FIMSAC) meeting memo, Available at https://www.sec.gov/spotlight/fixed-income-advisory-committee/blackrock-next-generation-bond-market-fimsa-011118.pdf.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905–939.

Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *The Journal of Finance* 43, 661–676.

Choi, Jaewon, and Yesol Huh, 2018, Customer liquidity provision: Implications for corporate bond transaction costs, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2848344.

Chordia, Tarun, Amit Goyal, Yoshio Nozawa, Avanidhar Subrahmanyam, and Qing Tong, 2017, Are capital market anomalies common to equity and corporate bond markets? An empirical investigation, *Journal of Financial and Quantitative Analysis* 52, 1301–1342.

Citi, 2018, Developments in credit market liquidity, SEC Fixed Income Market Structure Advisory Committee (FIMSAC) meeting memo, Available at https://www.sec.gov/spotlight/fixed-income-advisory-committee/citi-developments-in-credit-market-liquidity-fimsa-011118.pdf.

Dick-Nielsen, Jens, 2014, How to clean Enhanced TRACE data, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337908.

Dick-Nielsen, Jens, and Marco Rossi, 2018, The cost of immediacy for corporate bonds, *The Review of Financial Studies* 32, 1–41.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815–1847.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar, 2007, Corporate bond market transaction costs and transparency, *The Journal of Finance* 62, 1421–1451.

Feldhütter, Peter, and Thomas K. Poulsen, 2018, What determines bid-ask spreads in over-the-counter markets?, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286557.

FINRA, 2019, Proposed pilot program to study recommended changes to corporate bond block trade dissemination, Regulatory notice no. 19-12, Available at https://www.finra.org/industry/notices/19-12.

Friewald, Nils, and Florian Nagler, 2019, Over-the-counter market frictions and yield spread changes, *The Journal of Finance* .

Gilchrist, Simon, and Egon Zakrajšek, 2012, Credit spreads and business cycle fluctuations, *American Economic Review* 102, 1692–1720.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71 – 100.

Goldstein, Michael A., and Edith S. Hotchkiss, 2019, Providing liquidity in an illiquid market: Dealer behavior in us corporate bonds, *Journal of Financial Economics* .

Han, Song, and Xing Zhou, 2014, Informed bond trading, corporate yield spreads, and corporate default prediction, *Management Science* 60, 675–694.

Hanson, Samuel G, Robin Greenwood, and Gordon Y Liao, 2018, Asset price dynamics in partially segmented markets, *The Review of Financial Studies* 31, 3307–3343.

Hendershott, Terrence, Roman Kozhan, and Vikas Raman, 2019, Short selling and price discovery in corporate bonds, *Journal of Financial and Quantitative Analysis* 139.

Ivashchenko, Alexey, and Artem Neklyudov, 2018, (In)frequently traded corporate bonds, Working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3306124.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang, 2002, Dynamic volume-return relation of individual stocks, *The Review of Financial Studies* 15, 1005–1047.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *The Journal of Finance* 39, 1127–1139.

Schestag, Raphael, Philipp Schuster, and Marliese Uhrig-Homburg, 2016, Measuring liquidity in bond markets, *The Review of Financial Studies* 29, 1170–1219.

Wang, Jiang, 1994, A model of competitive stock trading volume, *Journal of Political Economy* 102, 127–68.

Zawadowski, Adam, and Martin Oehmke, 2016, The anatomy of the CDS Market, *The Review of Financial Studies* 30, 80–119.

Appendix A. Aspects of the model

Log-linear approximation of returns

Consider a homogeneous portfolio of perpetual defaultable bonds with invoice price P_t and coupon rate C. Its next period return R_{t+1} is:

$$1 + R_{t+1} = \frac{(1 - D_{t+1})(P_{t+1} + C)}{P_t},$$

where $D_{t+1} = h_{t+1}L_{t+1}$, and h_{t+1} represents a default rate and $L_{t+1} \in [0, 1]$ represents loss given default for bonds in the portfolio at time $t+1$.[35] Define $r_t \equiv \log(1 + R_t)$, $p_t \equiv \log(P_t)$, $c \equiv \log(C)$, and $-d_t \equiv \log(1 - D_t)$. Then

$$\begin{aligned} r_{t+1} &= -d_{t+1} - p_t + \log(P_{t+1} + C) \\ &= -d_{t+1} - p_t + p_{t+1} + \log\left(1 + \frac{C}{P_{t+1}}\right) \\ &= -d_{t+1} - p_t + p_{t+1} + \log\left(1 + e^{c - p_{t+1}}\right) \end{aligned}$$

Notice that the first-order Taylor expansion of $\log\left(1 + e^{c-x}\right)$ around $c - \bar{x}$ yields:

$$\log\left(1 + e^{c-x}\right) \approx \log\left(1 + e^{c-\bar{x}}\right) + \frac{e^{c-\bar{x}}}{1 + e^{c-\bar{x}}}\left((c - x) - (c - \bar{x})\right).$$

Then the expression for returns becomes:

$$\begin{aligned} r_{t+1} &= -d_{t+1} - p_t + p_{t+1} + \underbrace{\log\left(1 + e^{c-\bar{p}_{t+1}}\right) + \frac{e^{c-\bar{p}}}{1 + e^{c-\bar{p}}}(c - p_{t+1}) - \frac{e^{c-\bar{p}}}{1 + e^{c-\bar{p}}}(c - \bar{p})}_{\text{Call } \theta = \frac{1}{1+e^{c-\bar{p}}} \Rightarrow \frac{e^{c-\bar{p}}}{1+e^{c-\bar{p}}} = 1 - \theta} \\ &= -d_{t+1} - p_t + p_{t+1} - \log\theta + (1 - \theta)(c - p_{t+1}) - (1 - \theta)(c - \bar{p}) \\ &= \theta p_{t+1} - p_t - d_{t+1} + (1 - \theta)c + \underbrace{\left(-\log\theta - (1 - \theta)\log\left(\theta^{-1} - 1\right)\right)}_{\equiv \kappa}, \end{aligned}$$

which is equation (4). I set $\bar{p} = 0$ (the steady-state bond price is par), then $\theta = \frac{1}{1+C}$.

[35] With probability $1 - h_{t+1}$ the bond pays $P_{t+1} + C$ and with probability h_{t+1} it pays $(1 - L_{t+1})(P_{t+1} + C)$.

Learning by uninformed investors

The uninformed investor is a Bayesian agent learning about g_t and z_t at time t by observing $\tilde{p}_t$ and s_t. Recall that

$$\tilde{p}_t = -a(g_t + bz_t + es_t).$$

Hence, the agent knows $g_t + bz_t$ and an estimate of g_t immediately gives an estimate of z_t. The conditional distribution of $\tilde{p}_t$ given g_t and s_t is

$$\tilde{p}_t | g_t, s_t \sim N\left(-a(g_t + es_t), a^2 b^2 \sigma_z^2\right).$$

The unconditional distribution of g_t is $N(0, \sigma_g^2)$. Bayes theorem implies that $g_t | \tilde{p}_t, s_t$ is also Normal with a PDF $f_{g|\tilde{p},s}$:

$$f_{g|\tilde{p},s} \propto \exp\left(\underbrace{-\frac{(\tilde{p}_t + a(g_t + es_t))^2}{2a^2 b^2 \sigma_z^2} - \frac{g_t^2}{2\sigma_g^2}}_{\equiv -\frac{1}{2}K}\right).$$

Expanding the square and collecting terms, one gets:

$$K = \frac{g_t^2 - 2g_t\left[-\frac{a\sigma_g^2 \tilde{p}_t + a^2 \sigma_g^2 e s_t}{a^2\left(\sigma_g^2 + b^2 \sigma_z^2\right)}\right] + \Lambda(\tilde{p}_t, s_t)}{\frac{b^2 \sigma_z^2 \sigma_g^2}{\sigma_g^2 + b^2 \sigma_z^2}},$$

where $\Lambda(\tilde{p}_t, s_t)$ does not depend on g_t. Plug in the expression for the pricing function $\tilde{p}_t = -a(g_t + bz_t + es_t)$ to get:

$$\mathbb{E}_t^{(2)}\left[g_t | \tilde{p}_t, s_t\right] = \underbrace{\frac{\sigma_g^2}{\sigma_g^2 + b^2 \sigma_z^2}}_{\equiv \gamma}(g_t + bz_t),$$

$$\mathbb{V}_t^{(2)}\left[g_t | \tilde{p}_t, s_t\right] = (1 - \gamma)\sigma_g^2.$$

Optimal demands

The informed investor is solving the following problem:

$$\max_{X_t^{(1)}} \mathbb{E}_t\left[e^{-\left(W_t^{(1)} + X_t^{(1)} r_{t+1} + Z_t(1 + n_{t+1})\right)}\right],$$

where the distributions of r_{t+1} and n_{t+1} given the informed investor's information set at time t are both Normal with means $\mathbb{E}_t^{(1)}[r_{t+1}]$ and 0, and variances $\mathbb{V}_t^{(1)}[r_{t+1}]$ and σ_n^2 correspondingly. The covariance between r_{t+1} and n_{t+1} is time-invariant and equals σ_{rn} by assumption. The solution of the informed investor's optimization problem is

$$X_t^{(1)} = \frac{\mathbb{E}_t^{(1)}[r_{t+1}] - \sigma_{rn} Z_t}{\mathbb{V}_t^{(1)}[r_{t+1}]}.$$

The optimization problem for the uninformed investor (who does not own the non-traded asset by assumption) is the same up to Z_t component in the wealth dynamic and yields

$$X_t^{(2)} = \frac{\mathbb{E}_t^{(2)}[r_{t+1}]}{\mathbb{V}_t^{(2)}[r_{t+1}]}.$$

Conditional variances $\mathbb{V}_t^{(1)}[r_{t+1}]$ and $\mathbb{V}_t^{(2)}[r_{t+1}]$ are constant:

$$\begin{aligned} \mathbb{V}_t^{(1)}[r_{t+1}] &= \theta^2(\sigma_f^2 + \sigma_{\tilde{p}}^2), \\ \mathbb{V}_t^{(2)}[r_{t+1}] &= \theta^2(\sigma_f^2 + \sigma_{\tilde{p}}^2) + (1-\gamma)\sigma_g^2, \end{aligned}$$

Now, call $\sigma_r^2 \equiv \theta^2(\sigma_f^2 + \sigma_{\tilde{p}}^2)$ and plug in the expressions for conditional expected returns and variances into the expressions for optimal demand to get:

$$\begin{aligned} X_t^{(1)} &= \frac{a-1}{\sigma_r^2} g_t + \frac{b(a-1)}{\sigma_r^2} z_t + \frac{ae(1-\theta\delta)}{\sigma_r^2} s_t, \\ X_t^{(2)} &= \frac{a-\gamma}{\sigma_r^2 + (1-\gamma)\sigma_g^2} g_t + \frac{b(a-\gamma)}{\sigma_r^2 + (1-\gamma)\sigma_g^2} z_t + \frac{ae(1-\theta\delta)}{\sigma_r^2 + (1-\gamma)\sigma_g^2} s_t. \end{aligned}$$

Existence of the equilibrium

The equilibrium conditions imply the following system of three non-linear equations in a, b, and e:

$$\begin{aligned} \tfrac{\omega(a-1)}{\sigma_r^2} + \tfrac{(1-\omega)(a-\gamma)}{\sigma_r^2+(1-\gamma)\sigma_g^2} &= 0, \\ \tfrac{\omega(ab-\sigma_{rn})}{\sigma_r^2} + \tfrac{(1-\omega)(a-\gamma)b}{\sigma_r^2+(1-\gamma)\sigma_g^2} &= 0, \\ \tfrac{\omega ae(1-\theta\delta)}{\sigma_r^2} + \tfrac{(1-\omega)ae(1-\theta\delta)}{\sigma_r^2+(1-\gamma)\sigma_g^2} &= 1. \end{aligned}$$

The second equation immediately implies that $b = \sigma_{rn}$ is the only possible solution for b. The system of two remaining equations for a and e can be re-written as

$$0 = \phi_1(a,e) \equiv (a-\bar{a})(\sigma_r^2 + \omega(1-\gamma)\sigma_g^2) - (1-\bar{a})\omega(1-\gamma)\sigma_g^2,$$
$$0 = \phi_2(a,e) \equiv ae(1-\theta\delta)\omega(1-\gamma) - \sigma_r^2(a-\gamma),$$

where $\bar{a} = \omega + \gamma - \omega\gamma > \gamma > 0$. Observe from the first equation that $\phi_1(\bar{a}, e) < 0$ and $\phi_1(1, e) > 0$. Hence, if the solution a^* exists, it must be that $a^* \in (\bar{a}, 1)$. Then, take the derivative of the first equation with respect to a treating e as a function of a:

$$\frac{d}{da}\left[\phi_1(a, e(a))\right] = \sigma_r^2 + \omega(1-\gamma)\sigma_g^2 + (a-\bar{a})(\sigma_g^2 + b^2\sigma_z^2 + \sigma_s^2 e^2 + \sigma_s^2 ae\frac{d}{da}[e(a)]),$$

which is positive for $a \in (\bar{a}, 1)$ if $e^*(a)$ that solves the second equation $0 = \phi_2(a, e)$ grows in a. In this case we would have a unique positive solution $a^* \in (\bar{a}, 1)$. Now, I am going to establish the conditions under which this is indeed the case.

The second equation can be re-written as a quadratic equation with respect to e:

$$0 = \phi_2(a,e) = \left(a^2(a-\gamma)\theta^2\sigma_s^2\right)e^2 - \left(a(1-\theta\delta)\omega(1-\gamma)\right)e + (a-\gamma)\theta^2(\sigma_f^2 + a^2(\sigma_g^2 + b^2\sigma_z^2)).$$

Since $a^* > \bar{a} > \gamma$, it must be that $\phi_2(a, 0) > 0$, and if the solution e^* exists it must be that $e^* > 0$. Two candidate solutions of the quadratic equation can be written as:

$$e^*(a) = v(a) \pm v(a)k(a) \text{ where}$$
$$v(a) \equiv \underbrace{\frac{(1-\theta\delta)(1-\gamma)\omega}{2\theta^2\sigma_s^2}}_{\equiv 1/B}\frac{1}{a(a-\gamma)},$$
$$k(a) \equiv \sqrt{1 - B^2\psi(a)},$$
$$\psi(a) \equiv (a-\gamma)^2\left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a^2\right)$$

and for $a \in (\bar{a}, 1)$ $v > 0, v' < 0, 0 < k < 1, k' < 0, \psi > 0, \psi' > 0$. For the solutions to exist it must be that $\psi < B^{-2}$ for $a \in (\bar{a}, 1)$. Observe that

$$\psi = (a-\gamma)^2\left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_f^2}{\sigma_s^2}a^2\right) < (1-\gamma)^2\left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_f^2}{\sigma_s^2}a^2\right) \text{and}$$
$$B^{-2} = \frac{(1-\theta\delta)^2(1-\gamma)^2\omega^2}{4\theta^4\sigma_s^4}.$$

So, it is suffice to impose the following restriction on model parameters:

$$\frac{(1-\theta\delta)^2\omega^2}{4\theta^4}\frac{1}{\sigma_s^2\left(\sigma_f^2+\sigma_g^2+b^2\sigma_z^2\right)}>1,$$

to guarantee that the discriminant is non-negative and the quadratic equation for e has solutions. The condition is easy to obey since the shocks in the left-hand side denominator are small numbers. From now on I assume that the condition is satisfied.

Of the two roots of the quadratic equation for e, I am going to focus on the smaller one, $e^*(a) = v(a) - v(a)k(a)$. First, it is the root that guarantees that $e^*(a)$ grows with a when $a \in (\bar{a}, 1)$ as I am about to prove. Second, for reasonable parameters values $v(a)$ is a fairly large number (in a numerical example in Section VI it is around 60) and a positive root $v(a) + v(a)k(a)$ does not make much economic sense.

The smaller root $e^*(a) = v(a) - v(a)k(a)$ grows with $a \in (\bar{a}, 1)$ if $\frac{d}{da}\left[e^*(a)\right] > 0$, i.e.:

$$\begin{aligned}
v' - v'k - vk' > 0 &\Leftrightarrow \\
v'(1-k) > vk' &\Leftrightarrow \\
\frac{v'}{v} > \frac{k'}{1-k} &\Leftrightarrow \\
\frac{v'}{v} > \frac{k'(1+k)}{1-k^2} &\Leftrightarrow \\
\frac{v'}{v} > \frac{-\frac{1}{2k}B^2\psi'(1+k)}{B^2\psi} &\Leftrightarrow \\
\frac{v'}{v} > -\frac{1}{2}\frac{\psi'\left(1+\frac{1}{k}\right)}{\psi} &\Leftrightarrow \\
-\frac{2a-\gamma}{a(a-\gamma)} > -\frac{1}{2}\frac{\psi'\left(1+\frac{1}{k}\right)}{\psi} &\Leftrightarrow \\
-\frac{2a-\gamma}{a(a-\gamma)} > -\frac{\frac{\sigma_f^2}{\sigma_s^2}(a-\gamma)+\frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a(a-\gamma)(2a-\gamma)}{(a-\gamma)^2\left(\frac{\sigma_f^2}{\sigma_s^2}+\frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a^2\right)}\left(1+\frac{1}{k}\right) &\Leftrightarrow \\
2-\frac{\gamma}{a} < \frac{\frac{\sigma_f^2}{\sigma_s^2}+\frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a(2a-\gamma)}{\frac{\sigma_f^2}{\sigma_s^2}+\frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a^2}\left(1+\frac{1}{k}\right) &\text{ and observe that}
\end{aligned}$$

$$2 - \frac{\gamma}{a} < 2 < 1 + \frac{1}{k} < \frac{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2} a(2a-\gamma)}{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2} a^2} \left(1 + \frac{1}{k}\right),$$

which is indeed true.

To sum up, under the condition

$$\frac{(1-\theta\delta)^2\omega^2}{4\theta^4} \frac{1}{\sigma_s^2 \left(\sigma_f^2 + \sigma_g^2 + b^2\sigma_z^2\right)} > 1$$

the equation $0 = \phi_2(a, e)$ always has a root $e^*(a) > 0$ that grows with $a \in (\bar{a}, 1)$, and it leads to the unique solution $a^* \in (\bar{a}, 1)$ of $0 = \phi_1(a, e^*(a))$.

Derivation of the volume-return relationship

Plug in the expression for the pricing function $\tilde{p}_t = -a(g_t + bz_t + es_t)$ into (7) to get

$$r_t = -\theta(f_t - m_f) - a\theta g_t - a\theta b z_t - a\theta e s_t + (a-1)g_{t-1} + abz_{t-1} + aes_{t-1}.$$

Assume an econometrician also observes $v_{c,t} = |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)|$ and $v_{s,t} = s_t - s_{t-1}$. Now, the goal is to compute $\mathbb{E}_t\left[r_{t+1} | r_t, v_{c,t}, v_{s,t}\right]$.

Call, for the sake of convenience of notations, $x \equiv r_{t+1}$, $y \equiv r_t$, $v \equiv \alpha(\Delta g_t + \sigma_{rn}\Delta z_t)$, and $u \equiv v_{s,t}$. The unconditional distribution of (x, y, v, u) is Gaussian:

$$(x, y, v, u)' \sim \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}\right),$$

where $\Sigma_{11} = \sigma_{xx}$, $\Sigma_{12} \equiv [\sigma_{xy} \quad \sigma_{xv} \quad \sigma_{xu}]$ and

$$\Sigma_{22} \equiv \begin{bmatrix} \sigma_{yy} & \sigma_{yv} & \sigma_{yu} \\ \sigma_{yv} & \sigma_{vv} & 0 \\ \sigma_{yu} & 0 & \sigma_{uu} \end{bmatrix}.$$

The projection theorem for multivariate Normal distributions implies:

$$\mathbb{E}\left[x | y, v, u\right] = \beta_{xy} y + \beta_{xv} v + \beta_{xu} u,$$

where $(\beta_{xy} \quad \beta_{xv} \quad \beta_{xu}) = \Sigma_{12}\Sigma_{22}^{-1}$.

Now consider $\mathbb{E}\left[x|y,|v|,u\right]$. First, apply the law of iterated expectations:

$$\begin{aligned}\mathbb{E}\left[x|y,|v|,u\right] &= \mathbb{E}\left[\mathbb{E}\left[x|y,v,u\right]|y,|v|,u\right] \\ &= \mathbb{E}\left[\beta_{xy}y + \beta_{xv}v + \beta_{xu}u|y,|v|,u\right] \\ &= \beta_{xy}y + \beta_{xv}\mathbb{E}\left[v|y,|v|,u\right] + \beta_{xu}u.\end{aligned}$$

Notice that $\mathbb{E}\left[v|y,|v|,u\right] = \mathbb{E}\left[v|y,|v|\right]$ since $\sigma_{vu} = 0$. Now, use the fact that for any random variable Q with a PDF $f_Q(q)$:

$$\mathbb{E}\left[Q||q|\right] = |q|\frac{f_Q(|q|) - f_Q(-|q|)}{f_Q(|q|) + f_Q(-|q|)}.$$

In this case, it implies:

$$\mathbb{E}\left[v|y,|v|\right] = |v|\frac{f_{v|y}(|v|) - f_{v|y}(-|v|)}{f_{v|y}(|v|) + f_{v|y}(-|v|)},$$

where

$$v|y \sim \mathcal{N}\left(\frac{\sigma_{yv}}{\sigma_y y}y, \sigma_{vv} - \frac{\sigma_{yv}^2}{\sigma_{yy}}\right)$$

After straightforward algebra, one finds that

$$\mathbb{E}\left[v|y,|v|\right] = |v|\frac{e^{\rho|v|y} - e^{-\rho|v|y}}{e^{\rho|v|y} + e^{-\rho|v|y}} \approx \rho_{yv}|v|^2 y$$

for small values of v, where $\rho_{yv} = \frac{\sigma_{yv}}{\sigma_{vv}\sigma_{yy} - \sigma_{yv}^2}$.

Assembling altogether:

$$\mathbb{E}\left[x|y,|v|,u\right] \approx \left(\beta_{xy} + \rho\beta_{xv}|v|^2\right) y + \beta_{xu}u.$$

Since v and u are assumed independent, an additional conditioning on $|u|$ in the expectation sign is straightforward:

$$\mathbb{E}\left[x|y,|v|,|u|\right] \approx \left(\beta_{xy} + \rho_{yv}\beta_{xv}|v|^2 + \rho_{yu}\beta_{xu}|u|^2\right) y,$$

which is the analogue of (10). Above, $\rho_{yu} = \frac{\sigma_{yu}}{\sigma_{uu}\sigma_{yy} - \sigma_{yu}^2}$. To compute the coefficients in this relationship given model parameters one needs to compute the covariance matrix Σ. Direct calculations yield:

$$\sigma_{xx} = \theta^2\sigma_f^2 + \left((a\theta)^2 + (a-1)^2\right)\sigma_g^2 + (ab)^2(\theta^2+1)\sigma_z^2 + \frac{(ae)^2(\theta^2+1-2\theta\delta)}{1-\delta^2}\sigma_s^2;$$

$$\sigma_{xy} = (1-a)a\theta\sigma_g^2 - (ab)^2\theta\sigma_z^2 + \frac{(ae)^2(\theta\delta(1-\delta)+\delta-\theta)}{1-\delta^2}\sigma_s^2;$$

$$\sigma_{xv} = \alpha(a(\sigma_g^2 + b^2\sigma_z^2) - \sigma_g^2);$$

$$\sigma_{xu} = \frac{ae(1-\theta\delta)}{1+\delta}\sigma_s^2;$$

$$\sigma_{yy} = \sigma_{xx};$$

$$\sigma_{yv} = \alpha(1-a(1+\theta))\sigma_g^2 - \alpha ab^2(1+\theta)\sigma_z^2;$$

$$\sigma_{yu} = -\frac{ae(1+\theta)}{1+\delta}\sigma_s^2;$$

$$\sigma_{vv} = 2\alpha^2(\sigma_g^2 + b^2\sigma_z^2);$$

$$\sigma_{uu} = \frac{2}{1+\delta}\sigma_s^2.$$

Appendix B. Data and sample

Sample selection

I apply some filters to the TRACE database *after* cleaning it as in Dick-Nielsen (2014). Here are the criteria I use to select the bonds in the sample:

- The bond is nominated in USD;
- It is a fixed coupon (including zero-coupon), non-asset backed, non-convertible, non-enhanced bond;
- Not privately issued and not issued under Rule 144A;
- Of one of the following types according to the Mergent FISD classification: CMTN (US Corporate MTN), CDEB (US Corporate Debentures), CMTZ (US Corporate MTN Zero), CZ (US Corporate Zero), USBN (US Corporate Bank Note), PS (Preferred Security), UCID (US Corporate Insured Debenture);
- The interest is paid 1, 2, 4, or 12 times a year, or the bond is zero-coupon;
- The quoting convention is 30/360.

Four additional criteria must be jointly satisfied to keep a trade record in the sample:

- The trade is executed between Jan 1, 2010, and Jun 30, 2017;
- Executed at eligible times (time stamps of the trades are between 00:00:00 and 23:59:59; there is a small number of trades in TRACE with misreported times that do not fall into this range, I remove them from the sample);
- Executed on NYSE business days;
- Executed on or after the dated date of the bond (the date when the interest starts to accrue).

Agency transactions with commissions are retained in the sample.

Winsorization

To ensure that my results are not driven by extreme observations, I winsorize some variables. In particular, in the original bond-day panel (before active periods are determined) I winsorize:

- C-to-C trading volume at 99%;
- C-to-D trading volume at 1% and 99%;
- Credit spread at 99.9%;
- Bid-ask spread at 99.9%;
- Total daily returns at 0.1% and 99.9%.

SEC N-Q forms and holdings data

Adapted from Ivashchenko and Neklyudov (2018).

SEC N-Q forms submitted by mutual funds are available online through the SEC EDGAR system. I recover the number of mutual fund owners from machine-read N-Q forms. Mutual funds have a lot of discretion in how they fill in their N-Q forms, which makes it challenging to process the data. Below I describe the main steps I take.

Funds do not report bond CUSIP numbers in N-Q forms. I identify bond holdings in N-Q forms by issuer name, maturity, and coupon rate. I attempt to find N-Q records matching the CUSIPs in my sample. Several possibilities arise. If there is no match, I assign a value of zero to the number of mutual funds that hold the bond. If there is a match, it may or may not be unique. Even if the match is unique (which is the dominant case), it may refer to a not-in-sample bond with the same coupon rate and maturity. To ensure that I identify a correct bond I compute the cosine text similarity measure between an issuer name from the FISD database and an issuer name I recover from N-Q forms.[36] Table B1 provides some examples. Table B1a shows a record with a uniquely identified bond, while Table B1b shows

[36] I experimented with different similarity measures and did not observe much difference in results.

a record with double matching: one bond is the true bond I am looking for, another bond is mortgage-backed security with the same coupon rate and maturity. Regardless of whether the match is unique or not, I keep a record in the sample only if the cosine similarity measure is above 0.45.

cusip_id	issuer	maturity	rate	report	CIK	what	similarity
22541LAL7	credit suisse first boston (usa) inc	2009-01-15	3.88	2005-01-31	0000933996	credit suisse fb usa inc	0.67

(a) Unique maturity and coupon rate pair

cusip_id	issuer	maturity	rate	report	CIK	what	similarity
36158FAA8	ge global ins hldg corp	2026-02-15	7.00	2005-01-31	0000933996	ge global insurance holding	0.56
36158FAA8	ge global ins hldg corp	2026-02-15	7.00	2005-01-31	0000933996	fhlmc pool	0.17

(b) Non-unique maturity and coupon rate pair

Table B1. Examples of records with a unique and non-unique combination of maturity and coupon rate. First four columns (CUSIP number, issuer name, maturity, and coupon rate) are the data from Mergent FISD. The next three columns (report date, investment fund identifier CIK, and 'what') are the data from a machine-read N-Q filing matched to the FISD data by maturity and coupon rate. 'Similarity' is a cosine similarity between 'issuer' and 'what' fields.

Funds report N-Q forms twice every fiscal year, which is fund-specific. I use the 'last observation carried forward' approach to fill business days between two reporting dates for every fund, i.e., I assume that a fund holds all recently reported bonds every business day between the most recent and the previous N-Q report.[37] Once I have fund-bond-day holding indicators, I compute the number of funds that hold the bond in a given period and use this variable as the 'number of mutual fund owners' in my analysis.

Actively traded CDS contracts

DTCC publishes a list of 1000 most actively traded single-name CDS contracts quarterly since June 2009.[38] It includes both American and European, sovereign, and corporate issuers. I machine-read the data from these quarterly DTCC reports and remove all sovereign and all non-American reference entities. The DTCC reports contain some aggregate information

[37] It introduces a timing error when a fund opens a new or closes an existing bond position. It should not be critical for my results because I compute the average number of mutual funds that hold the bond in active periods, which are quite prolonged by construction.

[38] See DTCC website.

on CDS transactions like the total number of clearing dealers and average daily notional amount. In this paper, I use only the fact that an entity is listed among 1000 most actively traded contracts and do not use additional characteristics provided by DTCC.

The reference entities in DTCC reports are text strings; other firm IDs are not provided. I match text strings from DTCC reports to issuer names from Mergent FISD database (after some usual text cleaning) by computing Jaro-Winkler distance and keeping all name pairs where the distance is less than 0.2. Then I manually check all matched pairs to ensure that I do not have any false matches. All the entities that were not matched or were not mentioned in the DTCC report in a given quarter are assigned the CDS dummy value of 0. All matched entities are assigned the value of 1 for all days in a given quarter. Among 1000 U.S. firms mentioned at least once in DTCC reports from 2010 to 2017, I match a bit more than 800. I might have some 'true negatives' in the final sample (the firms that were not matched due to some text processing errors), but it should not affect my results as long as 'false positives' (wrongly matched firms) are absent.

Appendix C. Additional Tables and Charts

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
Issue size, mln USD	655.24	500.00	708.38	0.61	9.40	250.00	850.00	2000.00	15000.00	5746678
Rating	7.97	7.33	3.27	1.00	4.00	6.00	10.00	14.00	21.00	5746678
Age, years	4.93	3.58	4.63	0.00	0.33	1.67	6.75	15.50	62.42	5746678
Maturity, years	9.37	6.50	8.05	1.00	1.50	3.50	12.08	27.33	29.92	5746678
Duration	6.75	5.57	4.49	0.84	1.41	3.20	9.00	15.86	27.93	5746678
Total return, %	0.03	0.03	1.25	-8.19	-1.85	-0.36	0.43	1.90	8.49	5746678
Credit spread, %	2.55	1.90	2.84	0.00	0.69	1.28	2.98	6.24	88.70	5746678
Average bid-ask, %	1.14	0.74	1.16	0.00	0.08	0.31	1.62	3.37	19.99	2308138
No. trades per day	6.45	3.00	11.17	1.00	1.00	2.00	7.00	22.00	2540.00	5746678
No. days since last trade	2.33	1.00	7.25	1.00	1.00	1.00	2.00	7.00	1436.00	5735632
C-to-C volume, % of size	0.50	0.00	1.97	0.00	0.00	0.00	0.08	2.50	15.99	5746678
C-to-D volume, % of size	0.01	0.00	3.52	-19.67	-4.35	-0.22	0.33	4.29	17.91	5746678
|C-to-D volume|, % of size	1.52	0.28	3.18	0.00	0.00	0.05	1.31	7.86	19.67	5746678

Table C1. Summary statistics of the unfiltered bond-day panel. This is a counterpart of Table I that shows how sample characteristics change in the full unfiltered bond-day panel (no restriction on the number of days since the previous trade).

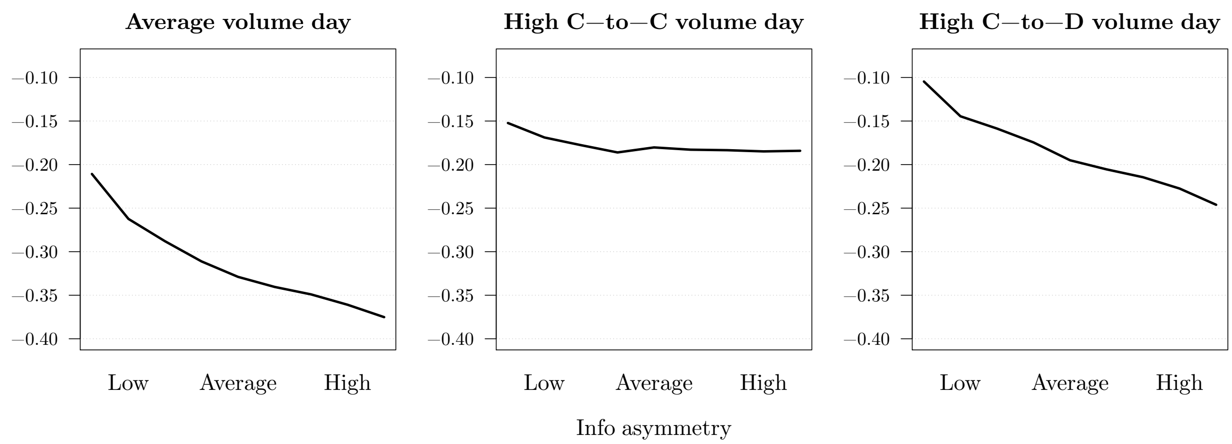

Figure C1. Point estimates for high-volume day reversals. The calculations are based on models (8) from Tables VI-VIII. On the x-axis from left to right are the deciles of information asymmetry proxies. For instance, 'Low asymmetry' bond is the one that has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 90-th percentile and stock volatility in the 10-th percentile. 'High asymmetry' bond has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 10-th percentile and stock volatility in the 90-th percentile. All other covariates from the regression models (average bid-ask spread, volume correlations, return volatility, and credit spread) are fixed at the median level. High C-to-C volume day is the day with C-to-C volume 2 standard deviations above the average (and average C-to-D volume); its reversal is $\mathbb{E}[\hat{\beta}_1|\text{covariates}] + 2\mathbb{E}[\hat{\beta}_2|\text{covariates}]$. High C-to-D volume day is the day with C-to-D volume 2 standard deviations above the average (and average C-to-C volume); its reversal is $\mathbb{E}[\hat{\beta}_1|\text{covariates}] + 2\mathbb{E}[\hat{\beta}_3|\text{covariates}]$. The reversal on the average volume day is simply $\mathbb{E}[\hat{\beta}_1|\text{covariates}]$.

	IG	HY	IG	HY	IG	HY
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$	
Intercept	−0.407***	−0.425***	0.108***	0.103***	0.107***	0.102***
	(0.008)	(0.023)	(0.011)	(0.019)	(0.008)	(0.019)
Average bid-ask	−0.084***	−0.085***	0.018**	−0.012	−0.038***	−0.050***
	(0.007)	(0.013)	(0.008)	(0.014)	(0.006)	(0.012)
C-to-C vlm corr.	0.136***	0.124**	0.092***	0.165***	−0.010	0.029
	(0.017)	(0.052)	(0.019)	(0.046)	(0.016)	(0.046)
C-to-D vlm corr.	−0.088***	−0.024	−0.126***	−0.110**	−0.025	0.004
	(0.016)	(0.043)	(0.022)	(0.050)	(0.019)	(0.049)
No. funds	0.0002***	0.0003*	−0.0002**	−0.0002*	0.0002**	0.00003
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
CDS dummy	0.005*	−0.006	−0.007*	0.005	0.006**	−0.009
	(0.003)	(0.009)	(0.004)	(0.008)	(0.003)	(0.008)
Issue size	0.058***	0.089***	−0.009**	−0.024*	−0.001	0.013
	(0.005)	(0.019)	(0.004)	(0.014)	(0.004)	(0.015)
No. dealers	0.001***	0.002***	−0.001***	0.0001	−0.0002	0.0003
	(0.0002)	(0.001)	(0.0002)	(0.0004)	(0.0002)	(0.0005)
Issuer size	0.00005***	0.00003	0.00003	−0.00000	−0.0001***	−0.0003***
	(0.00002)	(0.0001)	(0.00002)	(0.0001)	(0.00002)	(0.0001)
–Equity volatility	−0.541*	1.119**	−1.143***	0.492	0.544	−0.097
	(0.320)	(0.485)	(0.422)	(0.453)	(0.356)	(0.407)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	3,971	710	3,971	710	3,971	710
R^2	0.443	0.381	0.038	0.042	0.089	0.117
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table C2. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$; investment-grade and high-yield bonds separately. Each model is an OLS regression with heteroscedasticity-consistent standard errors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. 'No. funds' is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. 'No. dealers' is the average number of unique dealers that intermediate trades in each bond. Stock return volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
$\hat{\beta}_1$	-0.3287	-0.3429	114	9649	0	8700	9763
$\hat{\beta}_2$	0.0712	0.0618	7079	2684	1671	185	9763
$\hat{\beta}_3$	0.0584	0.0564	6894	2869	2043	345	9763

Table C3. Summary statistics for the cross-section of volume-return coefficients (estimated controlling for the market return in the first-step regression). This is a counterpart of Table IV, but the first-step regression here is $R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \beta_4 R_t^{\text{mkt}} + \epsilon_{t+1}$. The market return R_t^{mkt} is the return on Barclays IG Corporate Bond index.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Intercept	−0.438***	−0.448***	0.111***	0.099***	0.078***	0.089***
	(0.005)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)
Average bid-ask	−0.044***	−0.052***	0.017***	0.017***	−0.046***	−0.042***
	(0.006)	(0.005)	(0.006)	(0.007)	(0.005)	(0.005)
C-to-C vlm corr.	0.077***	0.074***	0.111***	0.101***	0.021	0.018
	(0.016)	(0.016)	(0.019)	(0.019)	(0.016)	(0.017)
C-to-D vlm corr.	−0.059***	−0.069***	−0.119***	−0.120***	−0.013	−0.021
	(0.014)	(0.014)	(0.020)	(0.021)	(0.017)	(0.018)
No. funds	0.001***	0.0004***	−0.0001**	−0.0001**	0.0003***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
CDS dummy	−0.001	−0.002	−0.006*	−0.006*	0.004	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Issue size	0.024***	0.028***	−0.012***	−0.011***	−0.009**	−0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
No. dealers	0.001***	0.001***	−0.001***	−0.001***	−0.0004**	−0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Issuer size		−0.0001***		0.00000		−0.0001***
		(0.00002)		(0.00002)		(0.00002)
–Equity volatility		−0.782**		−0.974***		0.656**
		(0.323)		(0.340)		(0.279)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	4,985	4,656	4,985	4,656	4,985	4,656
R^2	0.260	0.274	0.030	0.033	0.077	0.075
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table C4. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ (market return included in the first-step regression). Each model is an OLS regression with heteroscedasticity-consistent standard errors. Volume-return coefficients are averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. 'No. funds' is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. 'No. dealers' is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

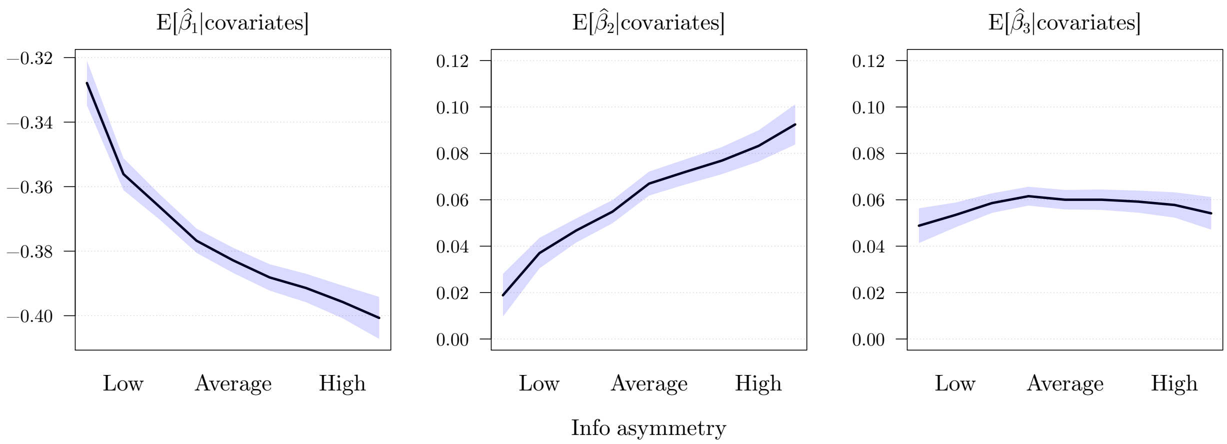

Figure C2. Point estimates and confidence intervals for the expected values of volume-return coefficients (market return included in the first-step regression). This figure is a counterpart of Figure 3, but the volume-return coefficients are estimated controlling for market return in the first-step regression (see Table C3).

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
$\hat{\beta}_1$	-0.3112	-0.3252	179	9644	12	8302	9823
$\hat{\beta}_2$	0.1126	0.0798	7414	2409	1948	157	9823
$\hat{\beta}_3$	0.0778	0.0732	7304	2519	2405	289	9823

Table C5. Summary statistics for the cross-section of volume-return coefficients (estimated controlling for volumes in the first-step regression). This is a counterpart of Table IV, but the first step equation here is $R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \beta_4 \tilde{V}_t^{(c)} + \beta_5 \tilde{V}_t^{(s)} + \epsilon_{t+1}$.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Intercept	−0.395***	−0.398***	0.170***	0.164***	0.104***	0.117***
	(0.007)	(0.007)	(0.011)	(0.013)	(0.007)	(0.008)
Average bid-ask	−0.091***	−0.103***	0.027***	0.027***	−0.055***	−0.050***
	(0.006)	(0.006)	(0.009)	(0.010)	(0.006)	(0.006)
C-to-C vlm corr.	0.122***	0.119***	0.022	0.016	−0.010	−0.007
	(0.017)	(0.017)	(0.026)	(0.026)	(0.017)	(0.017)
C-to-D vlm corr.	−0.074***	−0.075***	−0.099***	−0.072*	0.019	0.007
	(0.016)	(0.017)	(0.037)	(0.039)	(0.019)	(0.020)
No. funds	0.0003***	0.0003***	−0.0003***	−0.0003***	0.0003***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
CDS dummy	0.002	0.002	−0.004	−0.004	0.005*	0.002
	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)
Issue size	0.050***	0.046***	−0.016***	−0.014**	−0.016***	−0.010**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)
No. dealers	0.001***	0.001***	−0.001***	−0.001***	−0.0003*	−0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Issuer size		0.00004**		−0.00003		−0.0001***
		(0.00002)		(0.00003)		(0.00002)
–Equity volatility		0.453		−0.330		0.645**
		(0.286)		(0.472)		(0.298)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	5,018	4,691	5,018	4,691	5,018	4,691
R^2	0.356	0.368	0.029	0.027	0.076	0.076
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table C6. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ (volumes included in the first-step regression). Each model is an OLS regression with heteroscedasticity-consistent standard errors. Volume-return coefficients are averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. 'No. funds' is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. 'No. dealers' is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

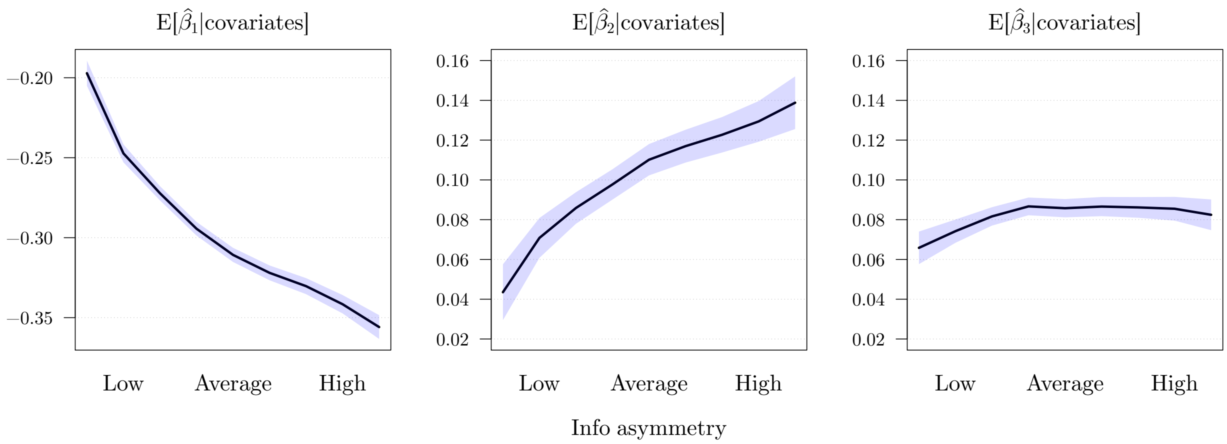

Figure C3. Point estimates and confidence intervals for the expected values of volume-return coefficients (volumes included in the first-step regression). This figure is a counterpart of Figure 3, but the volume-return coefficients are estimated controlling for trading volumes in the first-step regression (see Table C5).