INTERNET APPENDIX

ASSET MISPRICING

Kurt F. Lewis
Francis A. Longstaff
Lubomir Petrasek

1. Additional Details about the Data

We use a confidential version of the Trade Reporting and Compliance Engine (TRACE) database that contains all over-the-counter trades in publicly traded U.S. corporate bonds, including those issued under the Debt Guarantee Program. We filter out erroneous and duplicate entries using the procedure described in Goldstein, Hotchkiss, and Sirri (2007), which removes 31 percent of TRACE entries that are trade reversals, cancellations, exact duplicates, or inconsistent with reporting guidelines. This version differs from the public version of TRACE in that it explicitly identifies the dealers involved in each transaction and includes the actual size of each transaction. In contrast, the public version of TRACE is subject to a dissemination cap of $5 million per transaction, and all transactions in excess of $5 million are disseminated as $5MM+. An important advantage of this is that we can directly infer the inventory holdings of each dealer in the market for each of the bonds in the sample. Furthermore, the TRACE data set also includes an indicator for whether the transaction is between a dealer and another dealer, or between a dealer and a customer. This allows us to identify both total customer trading volume and total interdealer trading volume for each of the bonds in the sample. The Appendix to the paper describes how dealer inventories and interdealer and customer trading volumes are estimated using the TRACE data. Table IA-1 reports additional summary statistics about the inventory holdings of guaranteed bonds for the individual primary dealers in the study.

We also use a confidential data set from the Federal Reserve Bank of New York that identifies the margin or haircut that each dealer must pay to obtain repo financing for corporate bonds. The haircut affects the amount of capital that a dealer needs to support inventory purchases and is an important determinant of the dealer's funding liquidity (see Brunnermeier and Pedersen (2009)). This data set consists of disaggregated data on haircuts on corporate bond collateral posted by individual dealers in the tri-party repo market. The tri-party repo market is a key source of short-term secured funding for securities dealers who typically use the market to finance their inventory purchases (see Copeland, Duffie, Martin, and McLaughlin (2012)). A third party, called a clearing bank, facilitates the settlement of tri-party repos by transferring securities and cash from dealers to cash lenders such as money market funds, securities lenders, and other institutional funding providers. The market value of the securities posted as collateral customarily exceeds the amount of cash financing received from the repo counterparty by an amount called the "haircut." A number of recent papers have focused on the issue of what determines the size of the haircuts that dealers face in the repo markets. Important examples of this literature include Gorton and Metrick (2012), Dang, Gorton, and Holmström (2013), Krishnamurthy, Nagel, and Orlov (2014), Eren (2014), Copeland, Martin, and Walker (2014),

and Infante (2019).

As discussed by Duffie and Liu (2001), Longstaff, Mithal, and Neis (2005), Blanco, Brennan, and Marsh (2005), and others, credit default swap (CDS) spreads reflect the market price of insuring against the default by the firm or entity underlying the CDS contract. This means that an increase in the cost of protecting against a default by the primary dealer maps directly into an increase in the primary dealer's cost of capital. The dealer CDS spread has been used as a measure of intermediary capital constraints in a number of other studies including Gilchrist and Zakrajsek (2012) and Copeland, Martin, and Walker (2014). We obtain daily market prices for five-year CDS contracts for the dealers that are identified as primary dealers. The source of this CDS data is Markit. We also obtain CDS spreads for the issuers of the individual bonds in the sample from Markit. We use this information to test whether the credit risk of the issuer is related to the mispricing of the guaranteed corporate bonds.

2. Liquidity Measures

We compute the daily effective bid-ask spread of the ith bond as the volume-weighted price difference between trades in which the jth dealer sells and buys the same bond on a given day t, acting as a principal:

$$BA_{i,t} = \sum_{j} w_{j,t}(P^{sale}_{i,j,t} - P^{buy}_{i,j,t}), \qquad (A1)$$

where $w_{j,t}$ is the jth dealer's share of the trading volume for the ith bond on day t, and $P_{i,j}$ is the clean, volume-weighted price for which the jth dealer sells or buys the ith bond on day t. We include the prices of all principal transactions in which the dealer transacts with a non-dealer client. In a principal transaction the dealer trades with the client against his own inventory. By buying low and selling high, the dealer effectively earns a bid-ask spread, which compensates him for inventory costs, asymmetric information, and any other costs such as clearing and settlement (e.g., Glosten and Harris (1988)). In contrast, agency transactions are trades in which the dealer passes a bond on to the customer's account from another dealer without taking on inventory risk. Agency transactions are not included because the dealer's compensation consists of a fixed commission rather than a bid-ask spread. We also exclude interdealer trades because they typically involve much smaller price concessions than dealer-customer trades.

To capture the price impact of trades, we compute the Amihud (2002) illiquidity measure, defined as the average price change per one million dollars traded, and estimated for each transaction and averaged by trading day:

$$\text{Amihud}_{i,t} = \sum_{k} \frac{1,000,000}{N} \times \frac{P_{i,k} - P_{i,k-1}}{\text{Volume}_{i,k}}, \qquad (A2)$$

where $P_{i,k}$ is the price associated with kth transaction in the ith bond on day t, and N is the total number of transactions on that day. Similar to the effective bid-ask spread, the Amihud measure is based on dealer-customer trades.

3. The Correlation Structure of Mispricing

To examine whether mispricing correlations are greater when financial intermediaries are constrained, we compute pairwise correlations using data for the 2008–2009 and 2010–2012 periods separately. We note from Figure 1 of the paper that dealer CDS spreads were generally higher during the earlier part of the sample. Thus, financial intermediaries were likely more constrained during the 2008–2009 period. The average pairwise correlations for the 2008–2009 and 2010–2012 periods are 69.29 percent and 35.25 percent, respectively.

As another way of exploring the commonality in mispricing, we conduct a principal components analysis based on the correlation matrix of mispricing for the bonds. Since not every bond trades daily, we sample the data at a standardized weekly frequency and focus on the subset of bonds that have the broadest overlap with other bonds. In particular, we focus on the 44 bonds that have at least 44 observations in common with the other bonds in the subset. While this condition is necessary for the sample correlation matrix to be positive definite, it is not by itself sufficient since the individual pairwise correlations are generally based on observations with slightly different timing. As a result, several of the eigenvalues of the sample correlation matrix are slightly negative. We adjust the correlation matrix by setting these eigenvalues to small positive values. This technical numerical adjustment has only a very minor effect on the estimated correlation matrix. We then estimate the sample correlation matrix and perform a standard principal components analysis. Table IA-2 summarizes the results of this principal components analysis. The first principal component accounts for 57.10 percent of the total variance, and the second principal component explains an additional 14.26 percent. Thus, a significant component of the mispricing appears common across bonds used in the sample. However, extending the analysis to include as many as the first seven principal components still explains only 90 percent of the variance. This indicates that mispricing is also affected by factors that are specific to the individual bonds.

4. Estimating the Marginal State Income Tax Rate

Table IA-3 reports the results from the panel regression of the level of the unadjusted yield spreads described in Section 5 on the coupon rates of the bond.

As described in the Appendix, the coefficient for the coupon rate represents the marginal effect of state income taxes. The regression also includes monthly fixed effects and the time to maturity of the bond as controls. The standard errors are clustered by bond. As shown, the marginal state income tax rate is 1.655 percent.

5. Instrumental Variables Tests

The discussion in Section 8.1 of the paper provides results for a test of whether haircuts and CDS spreads affect mispricing only through the inventory channel. Here we briefly describe the econometric test used to make this conclusion. In instrumental variable regressions, the test of overidentifying restrictions can be used to examine the question of whether the excluded instruments (the instrumental variables for the potentially endogenous regressor which, in the typical two-stage least squares model of IV regression are included in the first-stage regression but excluded from the second-stage regression) are "appropriately excluded." This notion of appropriate exclusion is precisely what we would like to investigate. An instrument satisfies the exclusion restriction when its effect on the dependent variable in the regression can *only* occur through the potentially endogenous independent variable for which it is being used as an instrument. This is part of the two-part test for whether instruments are valid. The first part, instrument relevance, is easy to check empirically by looking for strong significance in the first-stage F statistic, which confirms whether or not the instruments covary with the endogenous regressor for which they are meant to instrument. The second part, instrument exogeneity, includes the component of the exclusion restriction. This part is able to be tested using the Hansen J statistic, the test we perform here. Importantly, the test is conditional. It is a test where, under the null hypothesis, we are proposing that conditional on at least one of the instruments being appropriately excluded, all of the excluded instruments are appropriately excluded. For exactly identified models with a single endogenous regressor and a single excluded instrument, the J statistic is zero and the test has no meaning.

Table IA-4 shows the results of the test we run to reach the conclusion discussed in the paper. The first column of the table shows the result of a regression of mispricing on inventory where we use non-TLGP inventory changes as the sole excluded instrument for changes in TLGP inventory. Here, we make a purely intuition-based appeal for the exclusion restriction insofar as we cannot see a mechanism by which the change in non-TLGP inventory could possibly have an effect on the mispricing of TLGP bonds except through correlation between the inventory level of TLGP and non-TLGP bonds. That is, the only way the holdings of corporate bonds more generally can affect the mispricing of the TLGP bonds would be to the extent that changes in those holdings can be helpful in describing how changes in holdings of TLGP bonds themselves affect the prices of

those bonds. This gives us a strong grip however, to test for the appropriateness of excluding other variables. Because the non-TLGP holdings are relevant (F statistic of 28.14) and intuitively meeting the exclusion restriction, we can look at the other variables in Table 5 of the main text and ask: do these series, which we can see help explain changes in inventory, affect mispricing only through the inventory channel? Table IA-4 shows the results of six additional versions of the same regression. In each case, we investigate adding explanatory variables from Table 5 and look to see if we can conclude that they are appropriately excluded instruments and therefore that the effects of these variables on mispricing occur only through the inventory channel. Because the table shows the second-stage regression, the only substantial differences in the table are in the list of excluded instruments, and the statistics about the first stage shown toward the bottom of the table. By showing the second-stage regression we can see that the results for the impact of inventory changes on mispricing do not vary much with these changes.

The second column of Table IA-4 shows the results of what is referred to as the baseline, which uses the excluded instruments used in the exercise described in Section 8.1. This includes non-TLGP inventory, and the first three lags of TLGP inventory, which were used in Table 5 of the paper. From the statistics toward the bottom of the table, we can see that these instruments are both relevant (F statistic of 10.27) and appropriately excluded (J statistic of 7.2, which cannot be rejected as being drawn from a chi-squared distribution with three degrees of freedom at the five-percent level). From the baseline, we go on to examine specifications which include CDS and haircuts, both contemporaneously and lagged, one at a time (columns 3 through 6) and then all together (column 7). In each regression we are looking for the resulting J statistic, and the p-value showing the probability that the test statistic value was drawn from the appropriate distribution, meaning the distribution from which the statistic should be drawn under the null hypothesis that if one of the instruments was correctly excluded, all of them were correctly excluded—which, again, would mean that each variables effects on mispricing would be occurring only through the inventory channel. Because we can add a single variable at a time and rerun the test, columns 3 through 6 test contemporaneous and lagged CDS and haircuts individually. In each of these cases, we see that large J statistics occur in regressions when we add these variables to the list of excluded instruments. Contemporaneous CDS, when added to the list of excluded variables, creates a J statistic of 23.01, while for contemporaneous CDS we get a J statistic of 21.76. From the table we can see that whether we look at them individually as in columns 3 through 6, or jointly as in 7, it appears that we can conclusively reject the hypothesis that CDS and haircuts affect mispricing exclusively through the inventory channel.

6. Identifying the Source of Identification in the Panel Regression

To understand better the role of the individual explanatory variables in the panel regression reported in Table 3, we estimate two alternative specifications. The first includes bond fixed effects into the panel regression framework and can be viewed as a purely time series version of the regression. In doing this, however, it is important to recognize that some of the explanatory variables in Table 3 such as the coupon rate, issue size, etc. are bond specific. As a result, these variables are almost perfectly collinear with the bond fixed effects. For example, the regression of the coupon rate on the bond fixed effects has an R^2 of 0.99993; the regression of bond duration on the bond fixed effects has an R^2 of 0.99999; the regression of issue size on the bond fixed effects has an R^2 of 0.99962, etc. In light of this, the alternative specification with bond fixed effects includes only variables that are not collinear with these fixed effects.

Table IA-5 reports the results from this alternative specification. As shown, the dealer CDS and dealer haircut variables are positive and significant, while dealer centrality and the dealer share of volume are negative and significant.

In the second specification, we estimate the panel regression using a standard two-stage Fama and MacBeth (1973) approach. This specification can be viewed as providing a purely cross-sectional perspective. As shown in Table IA-6 which reports the estimation results, the duration, dealer haircut, and dealer centrality variables are significant.

Taken together, the results from these two tables, in conjunction with those in Table 3, suggest that the effects of dealer CDS, dealer haircuts, dealer centrality, and dealer share of volume are identified via time series variation, the effects of duration, dealer haircuts, and dealer centrality are identified via the cross-section, and the effects of the number of dealers are identified jointly via time series variation and the cross-section.

INTERNET APPENDIX REFERENCES

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

Blanco, R., Brennan, A., Marsh, I. W., 2005. An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. Journal of Finance 60, 2255-2281.

Brunnermeier, M., Pedersen, L. H., 2009. Market liquidity and funding liquidity. Review of Financial Studies 22, 2201-2238.

Copeland, A., Duffie, D., Martin, A., McLaughlin, S., 2012. Key mechanics of the U.S. tri-party repo market. FRBNY Economic Policy Review 18, 17-28.

Copeland, A., Martin, A., Walker, M., 2014. Repo runs: Evidence from the tri-party repo market. Journal of Finance 69, 2343-2380.

Dang, T. V., Gorton, G., Holmström, B., 2013. Haircuts and repo chains. Working paper, Yale University.

Duffie, D., Liu, J., 2001. Floating-fixed credit spreads. Financial Analysts Journal 57, 73-87.

Eren, E., 2014. Intermediary funding liquidity and rehypothecation as a determinant of repo haircuts and interest rates. Working paper, Stanford University.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy, 81 607-636.

Gilchrist, S., Zakrajsek, E., 2012. Credit spreads and business cycle fluctuations. American Economic Review 102, 1692-1720.

Glosten, L. R., Harris, L. E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123-142.

Goldstein, M. A., Hotchkiss, E. S., Sirri, E. R., 2007. Transparency and liquidity: A controlled experiment on corporate bonds. Review of Financial Studies 20, 235-273.

Gorton, G., Metrick, A., 2012. Securitized banking and the run on repo. Journal of Financial Economics 104, 425-451.

Infante, S., 2019. Liquidity windfall: The consequences of repo rehypothecation. Journal of Financial Economics 133, 42-63.

Krishnamurthy, A., Nagel, S., Orlov, D., 2014. Sizing up repo. Journal of Finance 69, 2381-2417.

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

Newey, W. K., West, K. D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Table IA-1

Summary Statistics for the Inventory Holdings of Primary Dealers. This table presents summary statistics for the inventory holdings of the 12 dealers identified as primary dealers. Number of bonds denotes the number of different bonds held in the dealer's inventory at some point during the sample period. Average, Median, and Standard Deviation denote the summary statistics taken across bonds of the dealer's inventory holdings as a percentage of total primary dealers' inventory holdings.

Dealer	Number of Bonds	Average	Median	Standard Deviation
1	60	11.36	8.62	10.29
2	60	10.62	0.41	18.44
3	60	10.22	6.78	10.61
4	59	6.17	3.52	6.97
5	60	9.67	7.29	10.38
6	61	26.68	25.28	15.25
7	61	11.26	9.61	6.28
8	58	3.89	2.85	3.37
9	61	5.58	3.39	8.33
10	58	1.91	0.69	2.96
11	57	2.28	1.72	2.79
12	46	0.36	0.09	0.50

Table IA-2

Commonality in Mispricing. This table presents the results from a principal components analysis of the correlation matrix of mispricing for a subset of 44 bonds. The subset consists of the bonds with more than 44 common observations with every other bond in the subset. Percentage denotes the percentage of total variation explained by the indicated principal component. Cumulative denotes the cumulative percentage of total variation explained using the indicated number of principal components. The correlations are based on weekly data from December 2008 to June 2012.

Principal Component	Percentage	Cumulative
1	57.10	57.10
2	14.26	71.36
3	7.20	78.56
4	4.42	82.98
5	3.62	86.60
6	2.04	88.64
7	1.39	90.03

Table IA-3

Cross-Sectional Regression Estimating the Marginal State Income Tax Rate. This table reports the results from a time-series panel regression of the unadjusted yield spreads described in Section 5 on the coupon rate of the bonds with controls for time to maturity and monthly fixed effects. Yield spreads are measured in basis points. Coupon rates are expressed as percentages. Time to maturity is measured in years. The t-statistics are based on robust standard errors clustered by bond. The superscripts * and ** denote significance at the ten-percent and five-percent levels, respectively. The data are monthly from December 2008 to June 2012.

Variable	Coefficient	t-Stat
Coupon Rate	1.655	3.95**
Time to Maturity	10.286	11.04**
Monthly Fixed Effects		Yes
R^2		0.849
Number of Observations		1,646

Table IA-4

Tests of Exclusion Restrictions for CDS and Haircuts as Instruments for Inventory. This table reports the results from an instrumental variables regression of mispricing on changes in inventory. The specifications are differentiated only by the alternative sets of instrumental variables for changes in inventory. The main two rows of results in the table show the IV regression coefficients which are largely the same across the specifications. The final eight rows of the table contain the information about which variables are used as excluded instruments for change in inventory. The three rows directly above the final eight rows give statistics about the quality of the instrument choices in each specification. The first stage F statistic describes the quality of the fit of the instruments given in the final eight rows with changes in inventory. Under the null hypothesis that **all** of the instruments are appropriately excluded, the Hansen J statistic should be distributed chi-squared with degrees of freedom $(K-1)$, where K is the number of excluded instruments show in the final eight rows of the table (e.g., for the baseline specification $K=4$, and thus the J statistic should be chi-squared with three degrees of freedom). The J statistic p-value shows the probability that a value as large as the corresponding J statistic under that specification was drawn from the chi-squared distribution specified by the null hypothesis. The t-statistics are based on robust standard errors clustered by bond. The superscripts * and ** denote significance at the ten-percent and five-percent levels, respectively. The sample is monthly from December 2008 to June 2012.

Variable	Single Excluded Instrument	Baseline	Baseline + CDS_t	Baseline + Haircut_t	Baseline + CDS_{t-1}	Baseline + Haircut_{t-1}	Baseline + All Four
Intercept	-1.9662^{**}	-1.7930^{**}	-1.7749^{**}	-1.7817^{**}	-1.8140^{**}	-1.7863^{**}	-1.7870^{**}
	(-9.55)	(-13.68)	(-13.51)	(-13.58)	(-13.82)	(-13.61)	(-13.73)
Δ Inventory	-1.1409^{**}	-0.5358^{**}	-0.6265^{**}	-0.5922^{**}	-0.4304^{**}	-0.5693^{**}	-0.5659^{**}
	(-2.94)	(-2.39)	(-2.98)	(-2.81)	(-1.82)	(-2.65)	(-2.75)
N	1,646	1,451	1,451	1,451	1,451	1,451	1,451
First Stage F	28.14	10.27	8.24	10.13	8.61	9.96	7.59
Hansen J Stat	0.000	7.20	23.01	21.76	20.02	12.46	30.75
J Stat p-Value	NA	0.0659	0.0001	0.0002	0.0005	0.0142	0.0001
Excl Inst 1	Non-TLGP Inv	Non-TLGP Inv	Non-TLGP Inv	Non-TLGP Inv	Non-TLGP Inv	Non-TLGP Inv	Non-TLGP Inv
Excl Inst 2		TLGP Inv_{t-1}	TLGP Inv_{t-1}	TLGP Inv_{t-1}	TLGP Inv_{t-1}	TLGP Inv_{t-1}	TLGP Inv_{t-1}
Excl Inst 3		TLGP Inv_{t-2}	TLGP Inv_{t-2}	TLGP Inv_{t-2}	TLGP Inv_{t-2}	TLGP Inv_{t-2}	TLGP Inv_{t-2}
Excl Inst 4		TLGP Inv_{t-3}	TLGP Inv_{t-3}	TLGP Inv_{t-3}	TLGP Inv_{t-3}	TLGP Inv_{t-3}	TLGP Inv_{t-3}
Excl Inst 5			CDS_t	Haircut_t	CDS_{t-1}	Haircut_{t-1}	CDS_t
Excl Inst 6							Haircut_t
Excl Inst 7							CDS_{t-1}
Excl Inst 8							Haircut_{t-1}

Table IA-5

Panel Regression Including Bond and Monthly Fixed Effects. This table reports the results from the panel regressions mispricing on the indicated variables. Mispricing is measured in basis points. Issuer and dealer CDS spreads are measured in basis points. Dealer haircut is expressed as a percentage. Number of dealers denotes the number of dealers executing trades in the bond during the month. Number of institutions denotes the number of financial institutions holding positions in the bond as of the end of the month. Dealer centrality denotes the fraction of total dealer inventory held by the primary dealer for the bond. Dealer share of volume denotes the trading volume of dealers divided by total trading volume. Age is expressed in years. Bid-ask spread is measured in cents per 100 dollar par amount. The t-statistics are based on robust standard errors clustered by bond. The superscripts * and ** denote significance at the ten-percent and five-percent levels, respectively. The sample is monthly from December 2008 to June 2012.

Category	Variable	Coeff.	t-Stat
Control	Issuer CDS	0.4591	0.94
Intermediary	Dealer CDS	0.0349	3.20**
	Dealer Haircut	1.3467	2.04*
Network	Number of Dealers	−0.0671	−1.03
	Number of Institutions	−0.0087	−0.30
	Dealer Centrality	−4.8020	−3.06**
	Dealer Share of Volume	−2.4231	−2.10**
Liquidity	Bid-Ask Spread	−3.0690	−0.39
	Amihud Measure	0.3076	0.79
Bond Fixed Effects			Yes
Monthly Fixed Effects			Yes
Adj. R^2			0.880
Number of Observations			1,727

Table IA-6

Fama-Macbeth Estimation of the Panel Regression. This table reports the results from the two-stage Fama-MacBeth regressions of mispricing on the indicated variables. Mispricing is measured in basis points. Coupon is expressed as a percentage. Issuer and dealer CDS spreads are measured in basis points. Duration is measured in years. Dealer haircut is expressed as a percentage. Number of dealers denotes the number of dealers executing trades in the bond during the month. Number of institutions denotes the number of financial institutions holding positions in the bond as of the end of the month. Dealer centrality denotes the fraction of total dealer inventory held by the primary dealer for the bond. Dealer share of volume denotes the trading volume of dealers divided by total trading volume. Age is expressed in years. Issue size denotes the logarithm of the total par amount of the bond outstanding expressed in billions of dollars. Bid-ask spread is measured in cents per 100 dollar par amount. The t-statistics are based on Newey-West (1987) standard errors (four lags). The superscripts * and ** denote significance at the ten-percent and five-percent levels, respectively. The sample is monthly from December 2008 to June 2012.

Category	Variable	Coeff.	t-Stat
Controls	Constant	6.2648	0.71
	Coupon	0.7931	0.57
	Issuer CDS	0.0055	1.61
Price Risk	Duration	7.9604	5.69**
Intermediary	Dealer CDS	−0.0062	−0.28
	Dealer Haircut	2.3319	3.74**
Network	Number of Dealers	−0.0365	−0.69
	Number of Institutions	0.0026	0.11
	Dealer Centrality	−4.2246	−1.74*
	Dealer Share of Volume	−2.9294	−1.18
Liquidity	Age	−27.3751	−1.04
	Issue Size	−0.7971	−1.54
	Bid-Ask Spread	2.3356	0.44
	Amihud Measure	0.2862	0.32
Average R^2			0.689
Number of Observations			1,727