
The Effect of Illiquidity on Bond Price Data: Some Symptoms and Remedies

*Oded Sarig
Arthur Warga*



The Research Foundation of
The Institute of Chartered Financial Analysts

The Research Foundation of The Institute of Chartered Financial Analysts

Board of Trustees

Chairman

James R. Vertin, CFA
Alpine Counselors

Vice Chairman

George W. Noyes, CFA
Standish, Ayer & Wood, Inc.

President

Alfred C. Morley, CFA
Association for Investment
Management and Research

Darwin M. Bayston, CFA
Association for Investment
Management and Research

Thomas A. Bowman, CFA
Association for Investment
Management and Research

Gary P. Brinson, CFA
Brinson Partners, Inc.

Charles D. Ellis, CFA
Greenwich Associates

Frederick L. Muller, CFA
Atlanta Capital Management

William F. Sharpe
Sharpe-Tint, Inc.

Eugene C. Sit, CFA
Sit Investment Associates, Inc.

Walter P. Stern, CFA
Capital Research Co.

Donald L. Tuttle, CFA
Indiana University

Review Board

Research Director

Charles A. D'Ambrosio, CFA
University of Washington

Gerald O. Bierwag
University of Arizona

Joanne M. Hill
PaineWebber, Inc.

Mark P. Kritzman, CFA
Windham Capital Management

Richard Roll
UCLA
Roll & Ross Asset Management

Gary G. Schlarbaum, CFA
Miller, Anderson & Sherrerd

Officers

Alfred C. Morley, CFA
President

Darwin M. Bayston, CFA
Executive Vice President

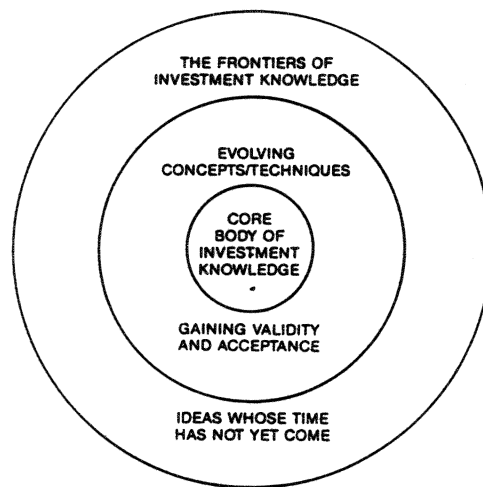
Thomas A. Bowman, CFA
Senior Vice President

Katrina F. Sherrerd
Vice President

Mission

The mission of the Research Foundation is to identify, fund, and publish research material that:

- expands the body of relevant and useful knowledge available to practitioners;
- assists practitioners in understanding and applying this knowledge; and
- enhances the investment management community's effectiveness in serving clients.



*The Research Foundation of
The Institute of Chartered Financial Analysts
P.O. Box 3668
Charlottesville, Virginia 22903
(804) 977-6600*

© 1990 The Research Foundation of the Institute of Chartered Financial Analysts.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright holder.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

From a Declaration of Principles jointly adopted by a Committee of the American Bar Association and a Committee of Publishers.

ISBN 10-digit: 0-943205-07-7 ISBN 13-digit: 978-0-943205-07-6

Printed in the United States of America.

Joni L. Tomal, *Editor*
Susan S. Brennan, *Production Editor*
Nina D. Hutchinson, *Typography/Layout*

6/15/90

The Effect of Illiquidity on Bond Price Data: Some Symptoms and Remedies

by

*Oded Sarig
Tel-Aviv University*

and

*Arthur Warga
Columbia University*

The Research Foundation of
The Institute of Chartered Financial Analysts

Table of Contents

Acknowledgements	ix
Foreword	xi
Introduction	xiii
Chapter 1. The Data	1
Chapter 2. Price Discrepancies and Data Filters	3
Chapter 3. Application: Model Estimation and Tests	11
Chapter 4. Application: Performance Evaluation	17
Chapter 5. Application: Estimating the Risk Structure of Interest Rates	19
Chapter 6. Conclusion	23
Appendix. Firms and Maximum Number of Bonds in Any Given Month (February 1985 to September 1987)	25
References	27

Acknowledgements

The authors are grateful to Charles D'Ambrosio, Robert Fuhrman, Avner Kalay, Richard Sweeney, and especially Greg Batey for helpful comments. Arthur Warga wishes to thank Howard Clark, Jr. for a grant, part of which was used to complete this work. Major funding was provided by the Research Foundation of the Institute of Chartered Financial Analysts.

Oded Sarig
Tel-Aviv University

Arthur Warga
Columbia University

Foreword

Bond analyses invariably contain the assumption that reported bond prices are true prices or close approximations of them. But what if they are not? If the difference between the prices assumed to be actual and the true ones are random, it may be of no consequence; the negatives usually offset the positives, provided large samples are used. If quoted bond prices differ systematically from actual prices, however, they are nonrandom, and data problems exist.

Sarig and Warga address these data problems. They evaluate the reliability of the CRSP (Center for Research in Security Prices) Government Bond File relative to the Shearson Lehman Brothers (SLB) Government Bond Price File. The average difference in the recorded prices between the two data sets is defined as liquidity-related price noise. The presence of such noise complicates analyses because one is never sure whether the information used is clean.

The authors find that the difference between the two price sets is not random. Rather, it is systematically related to bond liquidity. As one might suspect, the differences are greater for less-liquid bonds than for more-liquid ones. Even so, bond portfolio average returns are not systematically affected by liquidity-related price noise. In other words, even dirty data may be usable in a portfolio context. Nevertheless, the variance-covariance matrix of returns may be substantially affected—either negatively or positively—by these liquidity-related errors.

Not surprisingly, the authors discover that government bond prices are more accurate than corporate bond prices, a result they attribute to the higher liquidity and greater homogeneity of government bonds. The quality of reported price data deteriorates as bonds become less liquid, affecting measures of risk, return, and performance.

The authors filtered the data and discovered that age of bonds rather than their bid-ask spread was a more effective screen. This atypical approach is an interesting and useful insight.

The most important implications of this research affect portfolio management—particularly performance analysis. Anyone who has used data knows that their reliability is central to the results. For example, if the data contained in bond indexes are noisy, significant under- or overperformance may be found; if they contain systematic noisy biases, portfolio benchmarks that are based on those indexes may be suspect. Moreover, if the performance of portfolio managers is based on the single-factor market model, liquidity-related noise results in a downward bias in the risk estimates. Thus, a risk-adjusted performance benchmark may be flawed by the existence of noise in the data.

The Research Foundation is delighted to present this first-of-its-kind research to its constituency.

Charles A. D'Ambrosio, CFA
The Research Foundation of the
Institute of Chartered Financial Analysts

Introduction

Studies of bond-market data often assume that recorded bond prices are the true prices—i.e., the prices at which investors could buy (or sell) the underlying security at the close of the quotation day. In contrast, studies of stock prices explicitly recognize the importance of data problems, such as non-synchronous trading and bid-ask spreads, and attempt to overcome them.¹ Recently, McCulloch (1987) demonstrated that data problems may also affect studies of the bond market. He showed that Fama's (1984) rejection of the existence of a term premia in bond returns is an artifact of abnormally low bid-ask spreads of a particular instrument (nine-month Treasury bills). Here, we make a stronger point: Looking at the institutional side of the bond market, we show that certain bond market characteristics may cause quoted bond prices to differ from true prices in a systematic manner.

Understanding the nature of data problems is ostensibly important for users of this data—practitioners as well as researchers. This study focuses on the quality of the often-used CRSP Government Bond File, but it also demonstrates the importance of price errors to studies of corporate-bond data.

In Chapter 1 we compare CRSP bond-price data to prices reported in the Shearson Lehman Brothers (SLB) Government Bond Price File—both of which purportedly contain the same data, collected independently—to assess the accuracy of the recorded prices. Differences in recorded prices are considered to be liquidity-related price noise. This comparison does not presume that either data set is superior to the other. Rather, we analyze these price differences under the assumption that both quoted prices are of the true price plus independent noise terms.

As suspected, we find that discrepancies between the two data sets are systematically related to bond liquidity: As the liquidity of a bond decreases, the probability of large discrepancies between prices recorded in the CRSP Government Bond File and the SLB Government Bond Price File increases. On the other hand, these price discrepancies are approximately mean-zero and of a relatively small magnitude, as illustrated in Chapter 2.

Most users of bond-price data are not privy to multiple sources of data on all government bonds. Given that such nonrandom discrepancies exist, our second goal is to provide users of bond data with the means by which questionable observations may be screened out before any model is estimated. The suggested screens are based on observable variables and, hence, may be used by people who are not privy to an alternative data set. We evaluate the usefulness of the suggested screens in reducing the prevalence of price noise by the decrease in the average price discrepancy between the CRSP and SLB data sets achieved by their application. In Chapter 2, we show how applying the heuristic filters improves the quality of the data.

In Chapter 3 we show the impact of ignoring the differential quality of bond-price data on estimates of the means and variances of the distribution of monthly returns to constant-maturity portfolios. The results indicate that estimates of average rates of return to these portfolios are not systematically affected by the existence of liquidity-driven price noise. On the other hand, the results of a heuristic test that compares the variance-covariance matrix of returns with and without filtering the data indicate that the estimates of second moments may be substantially affected by liquidity-related price noise.

¹See, for example, Dimson (1979), Scholes and Williams (1977), or Roll (1984).

Bond market statistics are also used to assess the performance of bond portfolio managers. If one uses noisy data in forming the benchmark returns to which the manager's performance is compared, then spurious over- or underperformance may be found. Chapter 4 quantifies the magnitude of this potential problem by examining the properties of various performance measures estimated from filtered and unfiltered data.

Government bond prices are recorded more accurately than corporate bond prices because of higher liquidity and higher uniformity of the traded assets in the government bond market. This suggests that the magnitude of the effects of market illiquidity documented here for the government bond market should serve as lower limits for the respective suspected effects on corporate bond data. To illustrate these effects and the usefulness of sorting out suspect corporate bond data, in Chapter 5 we estimate the risk premia of corporate bonds of various ratings using filtered observations only. Chapter 6 concludes with a summary of our findings.

1. The Data

This chapter describes the features of the SLB and CRSP monthly government bond databases and compares their price records. Only non-callable government bond issues—including notes—that have no special tax features (e.g., no flower bonds) are included in the sample. Moreover, because of the limited availability of SLB data, comparisons of the two files are confined to the period of February 1981 through December 1985.^{1,2}

In the period studied, CRSP data was provided exclusively by Salomon Brothers.³ Because government bonds are not traded on a designated exchange, both Salomon Brothers and SLB price records are collected from trader quote-sheets at the close of trading on the last working day of each month. Traders for both firms are required to report prices for all the issues they are responsible for, regardless of whether there was trading activity in a particular bond.

There are several possible sources of price discrepancies between the two data sets. The first source is coding errors. Both companies expend extensive resources to try to prevent coding errors, however, so it is unlikely that such errors are prevalent in the data. Moreover, even if such errors do remain in the data, they should not be systematically related to any bond characteristic.

Second, the recording of the prices need not be done at the exact same time in both banks. The problem is particularly acute in periods of great price volatility, where price quotes recorded even minutes apart could differ substantially. In these periods, however, the timing difference may be offset by the high frequency of trading that is typical of such markets. Hence, we do not expect month-end prices and monthly rates of returns to differ substantially. More importantly, again, non-synchronous recording should not be correlated with any specific bond characteristic.

Probably the most common and systematic source of pricing discrepancy, and the one we analyze, is the existence of illiquid bonds—bonds that do not trade as often as other bonds. To fill the “quote-sheet” slots of these bonds, traders have to guess the price that would have cleared an active market if one existed.

An illiquid issue is usually priced using an interpolation scheme involving the prices of liquid issues having characteristics close to those of the bond in question (matrix pricing). Hence, the similarity of the “benchmark” bonds to the illiquid issue determines the quality of the resulting guess. When there is a lack of similar reference bonds, bond traders are more likely to come up with different prices for the illiquid bond. Furthermore, the guesses are likely to differ from the nonexistent—and therefore unobserved—market clearing price.

Unlike the preceding two reasons for price discrepancies, illiquidity is systematically related to certain bond characteristics. Specifically, because bonds are absorbed into investors’

¹Because of a change in SLB data collection procedures, we did not use SLB price records prior to 1981.

²The period under examination contained unusually high volatility in the bond market. To the extent that volatility and liquidity are correlated, the test period is characterized by an abnormally high level of liquidity. Therefore, the effects of illiquidity that we find should serve as lower bounds for less volatile periods.

³Recently, CRSP has started reporting Federal Reserve-collected prices, which are averages of price quotes of different houses. Our sample comprises single-source price records and, hence, our results are more relevant to users of single-source price data, such as traders of a given investment bank.

portfolios over time, a bond's liquidity tends to decrease monotonically with its age. Moreover, once a bond becomes illiquid because a large fraction of its amount outstanding is held by inactive investors, it tends to stay illiquid until it matures. For example, if a particular bond issue fits the needs of a dedicated pension fund's portfolio, the fund will buy—and take out of circulation—part of the total supply of this issue. Again, this effect on bonds' liquidity intensifies as the bonds age.

The age of bonds and their time to maturity at the issuance date are correlated. Therefore, the above analysis implies that illiquid bonds are more prevalent among long-maturity bonds than among short-maturity bonds. Because illiquid bonds are relatively poorly priced, price errors engendered by illiquidity do not occur at random. Rather, liquidity-related price errors are systematically related to identifiable bond characteristics.

We shall seek means to filter out observations suspected of being highly affected by liquidity-related noise. Given the above analysis, it is natural to employ measures of liquidity in screening out "bad" data. Unfortunately, we do not know of a direct measure of trading volume. Instead, we consider two proxies for bond liquidity: the bond's age and the bid-ask spread of price quotations. The first proxy is suggested by the discussion above. The second proxy is suggested by Roll's (1970) analysis as well as McCulloch's (1987) findings.

There are two reasons why the bid-ask spread should be a good proxy for liquidity. First, when a bond does not trade frequently, the trader, unsure of its true price, will require a higher spread to reduce the probability of trading at a loss. Second, traders are more likely to increase their bid-ask spreads when a bond is illiquid because of the difficulty of locating and delivering this bond when the bond is not in inventory (remember that no organized exchange exists for these bonds and locating a particular bond requires personal contacts with potential owners of the issue). We note, however, that because a high spread protects traders in volatile periods, this screening criterion may behave in a non-uniform manner over time.

2. Price Discrepancies and Data Filters

In this chapter, we examine the properties of liquidity-driven price noise. We also examine the relation between various symptoms of illiquidity and our suggested proxies for this variable.

As a preliminary step, we examine the most pronounced symptom of illiquidity: the occurrence of price runs. A price run is defined as a case where identical prices are reported for a bond in two consecutive month-ends. This may occur when the volume of trade in a particular bond issue is virtually nil. In this case, a trader may report the previous price as the current price, thereby creating a price run. Of course, true month-end prices do change, and it would be rare to find bonds with real price runs.¹ Hence, a price run is likely to be an unrecorded price change.

The frequency of price runs allows us to assess the severity of the problem with the quality of bond-price data. If the accuracy of the price-reporting process does not depend on the liquidity of the bond, the occurrence of a price run in an old (and therefore less-liquid) bond should be as likely as its occurrence in a young and more-liquid bond. On the other hand, if illiquid bonds tend to be less accurately priced, and if age is an acceptable proxy for the bond's liquidity, the frequency of unrecorded price changes should increase monotonically with the bond's age.

Tables 1 and 2 report the frequency of occurrence of price runs in the CRSP data grouped on the vertical axis by bond age, where age is the time elapsed in years from issuance to the quote day. The results are reported for all bid prices in the CRSP file for 1926 to 1985, and for the subperiod February 1981 to December 1985.

As suspected, a monotonic relation exists between bond age and the likelihood of finding a price run. A statistical test, reported in the note at the bottom of the tables, rejects the hypothesis of no correlation between age and runs at any reasonable significance level. This suggests that the accuracy of reported prices deteriorates as bonds age and become less liquid.

Although confined to the most severe expression of illiquidity, these results imply that the probability of observing price quotes that differ from the true prices is quite high. For example, in price quotes of bonds that are three or more years old, a price run occurs on average about once every 15 months (or in 416 observations out of 6,467) in the 1926-85 period.² This suggests that heuristic rules of filtering out suspected observations may be of value.

Given that liquidity-related price noise is systematically related to bonds' characteristics, we now turn, using the independently collected SLB price data file, to examine its nature. The SLB file includes only bid prices, whereas the CRSP file occasionally reports ask prices, transaction prices, or averages of bid and ask prices. Therefore, to eliminate a potential source of nonrandom error, we compared CRSP prices to SLB prices only when CRSP reported at least the bid price.

¹There were interest-rate regulations in some subperiods. If a ceiling on interest rates is binding for a particular bond, and if its coupon rate equals the ceiling, it is possible that price runs are more likely to occur in these subperiods. Even in this case, however, the fact that the coupon accrual accumulates linearly means that such runs will not occur in true prices. Furthermore, such runs should not be related to bond liquidity, which is the focus of our test.

²Note that in the last row of Table 1 there is some trailing-off of the probability of a price run as the bonds near maturity. It is possible that as a bond nears maturity, it can be priced more easily because it begins to behave like an actively traded Treasury bill. This may allow traders to price soon-to-mature bonds with relatively small error.

TABLE 1
Price Runs in CRSP Data
(January 1926 to December 1985)

<i>Years to Maturity at Issuance</i>	<i>Age (Years)</i>	<i>Less than 1</i>	<i>Between 1 and 2</i>	<i>Between 2 and 3</i>	<i>Between 3 and 4</i>	<i>Between 4 and 5</i>	<i>More than 5</i>
Less than 1	Observations	286					
	Runs	9					
	% Runs	3.15					
Between 1 and 2	Observations	2,103	385				
	Runs	101	24				
	% Runs	4.80	6.23				
Between 2 and 3	Observations	1,457	1,070	78			
	Runs	49	48	4			
	% Runs	3.36	4.49	5.13			
Between 3 and 4	Observations	822	676	574	146		
	Runs	20	28	31	13		
	% Runs	2.43	4.14	5.40	8.90		
Between 4 and 5	Observations	717	636	612	529	122	
	Runs	30	23	24	36	12	
	% Runs	4.18	3.62	3.92	6.81	9.84	
More than 5	Observations	2,766	2,320	2,124	1,932	1,648	2,090
	Runs	133	129	143	144	125	86
	% Runs	4.81	5.56	6.73	7.45	7.58	4.11

Notes: Under the null hypothesis that the probability of a run is independent of the bond's age, a cell has an equal probability of more or less runs observed in any cell to its right. Out of 175 possible comparisons, there are 142 cases where the observed frequency of runs is higher for the older bond's cell than for the younger bond's cell. Under the null, the expected number is 87.5 with a standard deviation of 6.6.

Source: The authors are the source of all tables and figures in this study.

Figure 1 plots the frequency of observed price discrepancies between matching prices reported in the CRSP and SLB files. As can be seen, the distribution of price discrepancies is centered around zero: The mean price discrepancy is \$0.0084 per \$100 face value. This is consistent with the traders in both investment banks erring in an unsystematic way, thereby introducing no upward or downward bias into the reported bond prices. The standard deviation of the price discrepancies is \$0.577 per \$100 face value, indicating that large discrepancies are also present in the data.

The fact that differences between price records have a zero mean implies that tests that use bond-market data to estimate mean returns, sample risk premia, and so forth, are not expected

TABLE 2
Price Runs in CRSP Data
(February 1981 to December 12, 1985)

<i>Years to Maturity at Issuance</i>	<i>Age (Years)</i>	<i>Less than 1</i>	<i>Between 1 and 2</i>	<i>Between 2 and 3</i>	<i>Between 3 and 4</i>	<i>Between 4 and 5</i>	<i>More than 5</i>
Less than 1	Observations	0					
	Runs	0					
	% Runs	0.00					
Between 1 and 2	Observations	231	153				
	Runs	0	3				
	% Runs	0.00	1.96				
Between 2 and 3	Observations	696	530	10			
	Runs	17	7	0			
	% Runs	2.44	1.32	0.00			
Between 3 and 4	Observations	370	310	257	55		
	Runs	3	7	9	2		
	% Runs	0.81	2.26	3.50	3.64		
Between 4 and 5	Observations	230	219	244	210	6	
	Runs	3	2	5	10	0	
	% Runs	1.30	0.91	2.05	4.76	0.00	
More than 5	Observations	1,051	842	762	677	566	1,009
	Runs	1	1	4	14	7	14
	% Runs	0.10	0.12	0.52	2.07	1.24	1.39

Notes: Under the null hypothesis that the probability of a run is independent of the bond's age, a cell has an equal probability of more or less runs observed in any cell to its right. Out of 175 possible comparisons, 108 cases are higher. Under the null, the expected number is 87.5 with a standard deviation of 6.6.

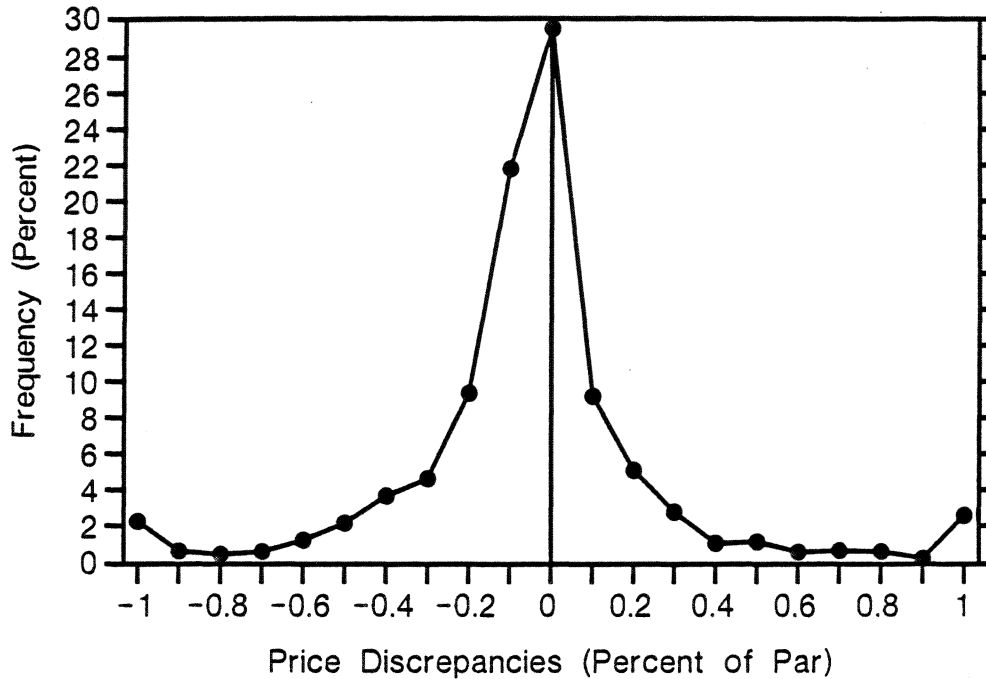
to be materially adversely affected by liquidity-driven price noise.³ On the other hand, estimates of variances and correlations, as well as statistical tests of significance, are more likely to suffer from the noisy recording of prices for illiquid bonds; Chapter 3 reports results that support this conclusion.

Figure 2 plots the distributions of price discrepancies using two subset of the data: bond prices versus note prices. For bonds (3,411 observations) the mean discrepancy is \$0.0037 with a standard deviation of \$0.7907, and for notes (5,502 observations) the mean discrepancy is \$0.0122 with a standard deviation of \$0.3941. Again, these figures are relative to a \$100 par value

³Given subsequent variance estimates and assuming that consecutive price errors are independent, a second-order Taylor series expansion of the rate of return calculation shows that liquidity-related price errors generate an upward bias in estimated mean returns (due to Jensen's inequality) of about 0.01 percent of the price-noise variance. This is a negligible effect for monthly returns.

FIGURE 1

Distribution of Price Discrepancies:
CRSP Versus SLB



Note: Values on the horizontal axis represent discrepancies in CRSP versus Shearson Lehman bond prices. Par value is \$100. Discrepancies that are greater than \$1.00 in absolute value are grouped together and account for the high tails.

for the bonds. Because notes are shorter maturity instruments than bonds, there are expected to be less old securities among the notes than among the bonds. If, as we conjectured, a security's age and its liquidity are negatively correlated, smaller price discrepancies should be observed for notes than for bonds. Figure 2 corroborates this prediction.

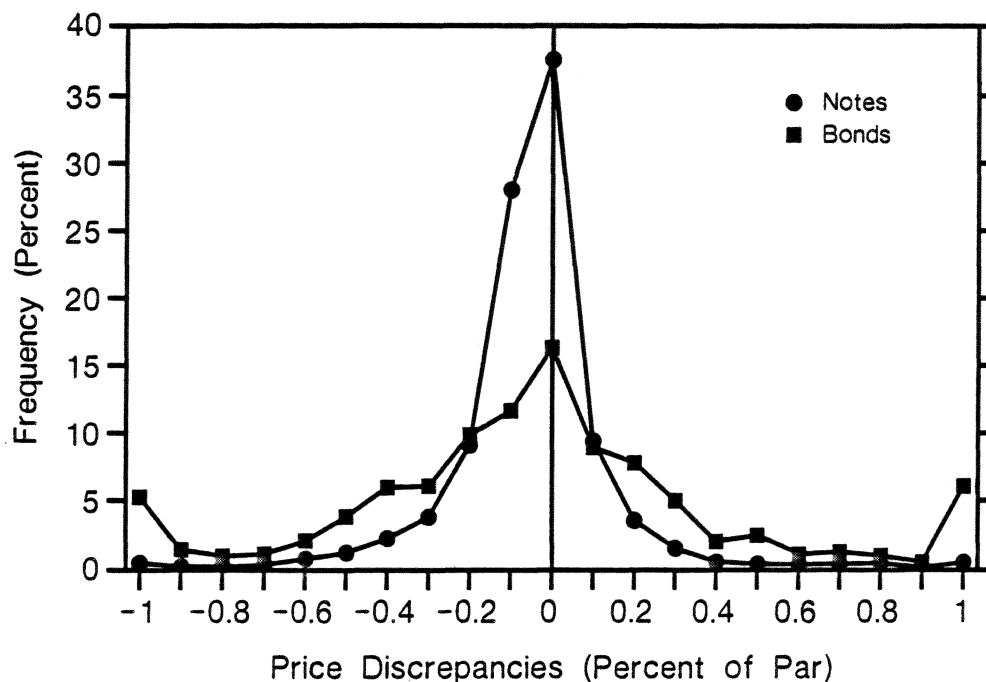
Next we examine the relation between price discrepancies and our proxies for bonds' liquidity—age and bid-ask spreads. To make these proxies good filters, we have to choose cut-off values that will eliminate most of the noise yet retain enough data to analyze models of the bond market. We eliminated price records of bonds whose age exceeded three years on the quotation date, or, alternatively, for which the bid-ask spread exceeded \$0.25 per \$100 par value. In Chapter 3, we construct constant-maturity portfolios, like those formed by previous users of the CRSP file (e.g., Fama 1984, and Gultekin and Rogalski 1985). The above-mentioned cut-off values are the most restrictive selection criteria that allowed us to calculate returns for these portfolios in almost every month.^{4,5}

⁴With these values, of the 600 portfolios' monthly returns estimated there were only six missing observations.

⁵See Sarig and Warga (1989a) for additional filters based on age and bid-ask spread, as well as filters based on the issue size of bonds.

FIGURE 2

**Distribution of Price Discrepancies
Bonds versus Notes**



Note: Values on the horizontal axis represent discrepancies in CRSP versus Shearson Lehman bond prices. Par value is \$100. Discrepancies that are greater than \$1.00 in absolute value are grouped together and account for the high tails.

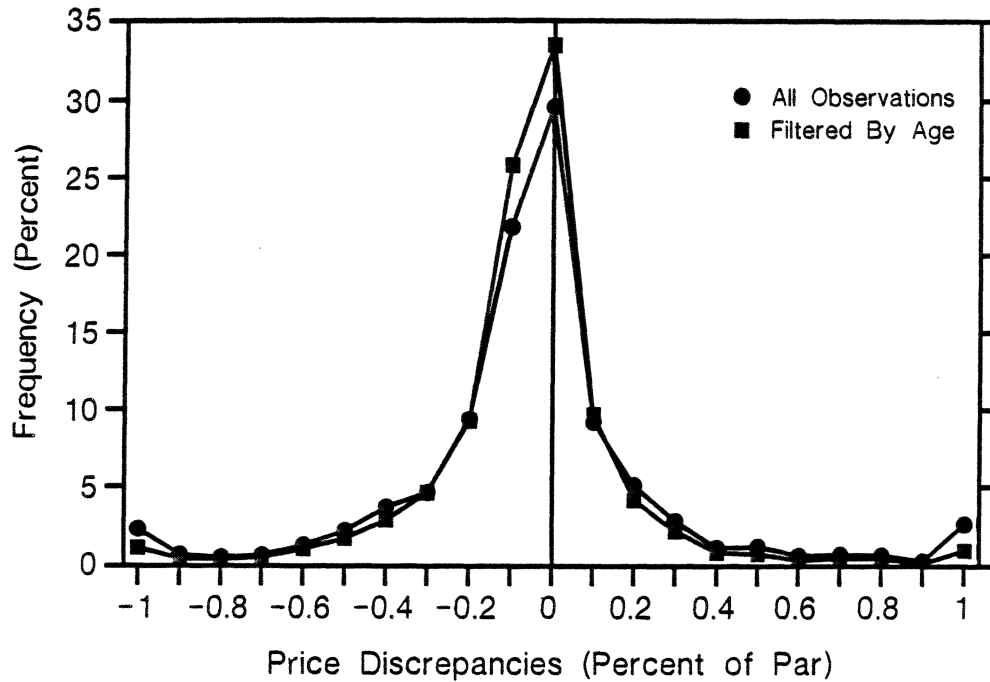
Note that the unscreened price discrepancies are unbiased (i.e., the price records differ by a mean-zero white noise). Hence, if age and bid-ask spread are to be useful as filters of bad data, their application should not introduce a bias. Furthermore, conditional on there being no bias, the quality of the filters can be measured by the extent to which they reduce the likelihood of observing large discrepancies.

Figures 3 and 4 present the distributions of price discrepancies after filtering suspected observations. In Figure 3, the distribution of unfiltered price discrepancies is compared to the distribution that results by filtering out prices of bonds older than three years (as of the quotation date). Figure 4 shows the comparison of the unfiltered distribution to the distribution of price discrepancies once price records of bonds with a bid-ask spread larger than \$0.25 were eliminated.

Similar to the distribution of all price differentials between the databases, the distributions of filtered price discrepancies (Figures 3 and 4) are symmetrically distributed at about zero. The mean discrepancy for the age-filtered data is \$0.022 per \$100 par value, and the mean discrepancy for the bid-ask spread-filtered data is \$0.013 per \$100 par value. This finding suggests that future

FIGURE 3

Distribution of Price Discrepancies:
The Age Filter



Note: Values on the horizontal axis represent discrepancies in CRSP versus Shearson Lehman bond prices. Par value is \$100. Discrepancies that are greater than \$1.00 in absolute value are grouped together and account for the high tails.

users of bond-market data, who will not be able to compare their data to an alternative data set, need not worry about introducing biases into price records by filtering the data.⁶

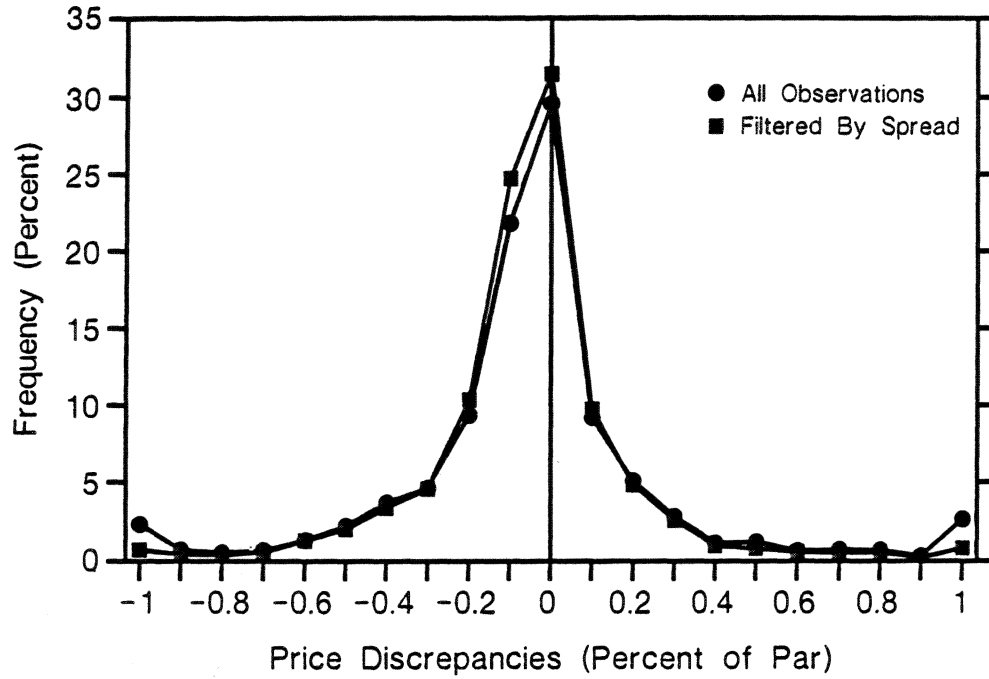
Both Figures 3 and 4 document the viability of the suggested filters as screens of suspected price records (i.e., price records that were found to be substantially different between the two files). This is because both filters reduce the noise in the data. In particular, the standard deviation of price discrepancies is reduced from \$0.5768 to \$0.4388 (per \$100 face value) by applying the age filter, and is reduced to \$0.4354 by applying the bid-ask spread filter.

It seems that the two filters equally improve the quality of data. We note, however, that the two filters do not eliminate the same records in all periods. In the next chapter we show that, although the overlap is high, the use of the age filter results in a more uniform elimination of price records over time than the spread filter does.

⁶Such biases may result if traders in different investment banks have tendencies to over- or underestimate bond prices to affect the value of their portfolios, which in turn affects their compensation.

FIGURE 4

**Distribution of Price Discrepancies:
The Bid-Ask Spread Filter**



Note: Values on the horizontal axis represent discrepancies in CRSP versus Shearson Lehman bond prices. Par value is \$100. Discrepancies that are greater than \$1.00 in absolute value are grouped together and account for the high tails.

3. Application: Model Estimation and Tests

In the preceding chapter we showed that liquidity-related price noise exists and that it is mean zero. Price data errors whose means are zero are not expected to materially bias monthly returns. Tests of equilibrium theories on the relation between risk and return may, however, be affected by the added noise to the extent that variance estimates are upwardly biased. This chapter assesses the importance of these effects by estimating moments of the returns to various bond portfolios. Virtually all tests of asset-pricing models (including bond-market data) test certain relations among moments of rates of return of various instruments. In this chapter we examine the extent to which the liquidity-driven price noise affects estimates of the means and variances of bond-portfolio returns by comparing statistical relations estimated using all of the CRSP data to those estimated using only filtered data.

Typically, bond market studies examine the time series properties of returns to portfolios of bonds with constant maturity. The use of constant-maturity portfolios, rather than buy-and-hold portfolios, is intended to alleviate the problem that, as time elapses, the nature of bonds and of bond portfolios changes systematically. To enable statistical estimation of market parameters (e.g., term premia), one has to insure that the salient characteristics of the examined portfolios, such as duration, are constant over time. Such constancy is achieved by using a portfolio-formation process that generates “constant maturity” portfolios. A byproduct of this process is that one obtains uninterrupted time series of returns, while the use of individual bonds creates broken return series.

Our portfolio-formation procedure involves averaging, in any month, data on bonds that have similar maturities. Bonds are considered of similar maturity if their maturity falls within six-month intervals for maturities of up to three years. Longer-maturity bonds were grouped into three- to four-year, four- to five-year, and five- to ten-year portfolios. This produces a total of nine portfolios. (See Fama (1984) for a more detailed description of the process.) The procedure generates fairly homogeneous and time-invariant portfolios of short maturities, but gives less satisfying results in longer maturities. In particular, a portfolio of bonds with maturities over 10 years exhibited the least degree of stability over time. As a result, we did not analyze the returns to portfolios of maturities over 10 years.

As explained in the preceding chapter, we chose cut-off values for the filters that constrain the data as much as possible while still yielding complete time series of returns for each of the constant-maturity portfolios. By so doing, we hoped to minimize the chance of incorporating suspicious data into our estimates, without reducing the number of observations for tests that focus on monthly returns to these portfolios.

The following filter cut-off values were used. For the age filter, we discarded monthly return data of bonds that were issued more than three years prior to the beginning of the quote month. For the bid-ask filter, we eliminated all bond returns that had a month-end bid-ask spread quote larger than \$0.25 (per \$100 par value).¹ For the filter of price discrepancies between databases,

¹When we could not make a positive identification of an excessive spread (e.g., when only a bid price was quoted in the CRSP file), we kept the observation.

observations with discrepancies between the CRSP and the SLB files that were larger than \$0.20 (per \$100 par value) were eliminated.²

Table 3 reports the means and *t*-statistics of the monthly rates of return of the nine constant-maturity portfolios (above the one-month Treasury-bill rate).³ Table 3 corroborates our conjecture that an error with a mean of zero will have little systematic effect on estimates of average rates of return. In particular, mean excess returns to all nine portfolios are little changed when any of the data filters are applied.

We next look at the effect of applying these filters on the estimated variance-covariance matrix between the returns to these nine portfolios. The first diagnostic indicating differences in covariance matrices is the F-statistic testing the joint hypothesis that the nine portfolio mean excess returns all equal zero.⁴

Table 4 reports F-statistics for the three filtered data subsets (labelled “constrained”) and for the three data subsets formed by using the data excluded by the filters (labelled “complement”). Because of the way we chose the cut-off values for the filters, the complement portfolios do not contain data in every month. Whenever an observation is missing for the complement portfolio return series, the corresponding constrained series’ observation is used.⁵ (The third column of Table 4 indicates the average number of months when the complement portfolio was not empty.)

Note that, uniformly, the constrained portfolios, which are less noise-contaminated, have higher levels of significance than the unconstrained portfolios. Although all F-statistics of the constrained portfolios are highly significant, the complement portfolios, which are more noise-contaminated, have lower significance levels.

To gain further understanding of the effects of liquidity-driven price noise on tests of pricing models, assume that the generating process for bond *i*’s return is given by:

$$r_{it} = E\{r_i\} + \beta_i d_t + e_{it} \quad (1)$$

where β_i is a *k*-vector of factor loadings and d_t is a *k*-vector of mean-zero risk factors (see, for example, Gultekin and Rogalski 1985). The existence of liquidity-driven price errors implies that the *observed* return series, R_{it} , will be given by:

$$R_{it} = r_{it} + q_{it} = E\{r_i\} + \beta_i d_t + e_{it} + q_{it} \quad (2)$$

where q_{it} denotes the liquidity-related price errors.

Equation (2) implies that liquidity-related price errors increase the total variability of observed return series. Therefore, a larger fraction of the total variability would be explained by the *n* (<*k*) first principal components (i.e., the estimates of the risk factors in equation (1)) when

²These filtering values did not actually succeed completely in providing constrained portfolios in every month. For the discrepancy filter, portfolios 3, 4, and 6 through 9 each had one month with a missing observation. The age and bid-ask spread portfolios were complete.

³The results reported are based on analysis of simple rates of returns. Similar results were obtained when continuously compounded rates of returns were used.

⁴These F-statistics are based on the inverse of the covariance matrix and thus are a way to discern if there can be substantive differences in the conclusions one might draw from filtered and unfiltered data.

⁵Warga (1989) tests for liquidity differences in bonds using techniques that allow for missing data in the time series of returns to constant duration portfolios.

TABLE 3

**Means of Monthly Portfolio Returns
Estimated Using Filtered and Unfiltered Data
In Percent Per Month
(February 1981 to December 1985)**

<i>Portfolio 1</i>	<i>Portfolio 2</i>	<i>Portfolio 3</i>
UN 1 = 0.1238 (4.28185)	UN 2 = 0.2004 (2.82741)	UN 3 = 0.2938 (2.8036)
SLB 1 = 0.1215 (4.37664)	SLB 2 = 0.1942 (2.9201)	SLB 3 = 0.3028 (2.8407)
AGE 1 = 0.1295 (4.42851)	AGE 2 = 0.2101 (3.00316)	AGE 3 = 0.3026 (2.84448)
B/A 1 = 0.1253 (4.35571)	B/A 2 = 0.1967 (2.76964)	B/A 3 = 0.2893 (2.75101)
<i>Portfolio 4</i>	<i>Portfolio 5</i>	<i>Portfolio 6</i>
UN 4 = 0.3134 (2.28647)	UN 5 = 0.3383 (2.02414)	UN 6 = 0.3721 (1.88804)
SLB 4 = 0.3156 (2.29342)	SLB 5 = 0.3425 (2.01563)	SLB 6 = 0.3544 (1.76671)
AGE 4 = 0.3127 (2.2822)	AGE 5 = 0.3326 (1.98681)	AGE 6 = 0.3834 (1.94204)
B/A 4 = 0.3141 (2.25357)	B/A 5 = 0.3466 (2.07283)	B/A 6 = 0.3713 (1.84654)
<i>Portfolio 7</i>	<i>Portfolio 8</i>	<i>Portfolio 9</i>
UN 7 = 0.4312 (1.86784)	UN 8 = 0.428 (1.57153)	UN 9 = 0.5233 (1.44474)
SLB 7 = 0.4201 (1.80803)	SLB 8 = 0.4226 (1.53594)	SLB 9 = 0.4808 (1.35269)
AGE 7 = 0.4224 (1.83618)	AGE 8 = 0.4303 (1.5924)	AGE 9 = 0.5158 (1.42919)
B/A 7 = 0.4205 (1.79426)	B/A 8 = 0.4308 (1.57666)	B/A 9 = 0.5359 (1.47514)

Notes: UN = unconstrained estimates
 SLB = estimates using data filtered by discrepancies
 AGE = estimates using data filtered by age
 B/A = estimates using data filtered by bid-ask spread
 () = *t*-statistic

TABLE 4

**F-Statistics for Joint Test of
Significance of Mean Excess Returns
(February 1981 to December 1985)**

$$\text{Unconstrained } F(9,50) = 4.03 \\ (0.0006)$$

<i>Filter</i>	<i>Constrained</i>	<i>Complement</i>	<i>Average Number of Distinct Observations</i>
SLB	F = 3.35 (0.003)	F = 1.90 (0.074)	39.0
AGE	F = 3.00 (0.006)	F = 1.84 (0.085)	58.2
B/A	F = 3.57 (0.002)	F = 1.92 (0.070)	36.6

Notes: *Constrained* refers to portfolios formed with a data filter (discrepancy, age, and bid-ask respectively). *Complement* refers to portfolios composed of data excluded by a filter. Missing observations for complement portfolio returns are filled in with the respective observations from the constrained return series. Column titled Average Number Of Distinct Observations indicates average number of months when complementary observations were not empty. Parentheses indicate significance level.

the covariance matrix is estimated using “clean” data, than the fraction of variance explained by these n principal components when the covariance matrix is estimated using “noisy” data.

Table 5 reports the percent of variance explained by the first three principal components, as well as the correlation between the estimated factor returns, using both the constrained portfolio returns (the “clean” data) and the complement portfolio returns (the “noisy” data). As expected, for all three data filters there is a significant increase in the explanatory power of the first factor when suspicious observations are removed. Moreover, the correlation between the estimated returns to the factors estimated with “clean” and with “noisy” data is very high, suggesting that liquidity-driven price noise is not correlated with the true factors (i.e., it is “white” noise).

Obviously, the factor analysis carried out above is only one example of the possible problem engendered by the use of unfiltered data. It is impossible to assess the magnitude of the effect on all potential tests, but it is encouraging to observe that the filters suggested earlier do go a long way toward eliminating the problem in this particular example.

Previously, we stated that the age filter provides a smoother screening of suspect data than the spread filter does. One manifestation of this is the fact that the age filter excludes observations in almost every month and does not exhibit time-varying effectiveness. On the other hand, the bid-ask spread filter is binding mostly in months of high bond market volatility. Hence, the spread filter excludes illiquid observations in fewer months than the age filter. Another way in

TABLE 5

**Percentage of Variance Explained by
First Three Principal Components
of the Matrix of Constant-Maturity Portfolios**

<i>Filter</i>		<i>Constrained</i>	<i>Complement</i>	<i>Correlation Between Principal Components</i>
SLB	PC1	97.31	91.71	0.993
	PC2	1.94	3.84	0.571
	PC3	0.27	2.18	0.160
AGE	PC1	97.53	93.84	0.993
	PC2	1.89	2.67	0.822
	PC3	0.26	1.10	0.405
B/A	PC1	97.40	91.19	0.988
	PC2	1.90	4.46	0.602
	PC3	0.30	1.78	0.302

Notes: *Constrained* refers to portfolios formed with a data filter (discrepancy, age, and bid-ask respectively). *Complement* refers to portfolios composed of data excluded by a filter. *Correlation* is between the nth ($n = 1,2,3$) principal component of the constrained matrix and the nth principal component of the complement matrix. Missing observations for complement portfolio returns are filled in with the respective observations from the constrained return series.

which the smoother screening of data by the age filter is revealed is that the age filter eliminates a similar fraction of the bonds in all constant-maturity portfolios. The bid-ask spread filter, on the other hand, has a greater impact in the long-maturity portfolios, and a smaller impact in the short-maturity portfolios. This may be a manifestation of a common practice to require a larger spread for long-maturity bond trades than for short-maturity bond trades.

The uneven nature of the bid-ask spread filter affects not only the number of the bonds excluded from the constant maturity portfolios, but also the characteristics of the bonds retained. To demonstrate this aspect of the bid-ask spread filter, we calculated value-weighted bond market indexes using all bond data, data filtered by age, and data filtered by the bid-ask spread. Table 6 provides descriptive statistics for these three indices.

Table 6 demonstrates the distortion induced by applying the bid-ask spread filter to the data: the duration of the market portfolio—approximated by the weighted average of the duration of the component bonds—is hardly affected by the application of the age filter (an average reduction of 2.4 days), whereas the application of the spread filter reduces (on average) the duration of the market portfolio by 62.7 days.

TABLE 6
Summary Statistics on Value-Weighted
Bond Indexes

	<i>Unconstrained</i>	<i>Type of Filter</i> <i>Age</i>	<i>Bid-Ask</i>
Annual mean rate of return (%)	14.1924	14.1516	14.0724
Standard deviation	17.3892	17.4708	16.5168
Average duration (days)	925.6	923.2	862.9
Range of monthly difference in duration of filtered data from duration of unfiltered data (days)		-45<X<31	-6<X<146

Because duration is a measure of price sensitivity to rate changes, the implication is that the nature of the bonds retained after the spread filter is applied is materially different from the nature of the bonds filtered out. Accordingly, the estimated mean returns and standard errors of returns are more significantly affected by the application of the spread filter than by the application of the age filter. Given the above-documented problems associated with the use of the bid-ask spread filter, the rest of our analysis will employ the age filter only.

4. Application: Performance Evaluation

In the preceding chapter, we documented the sensitivity of certain parameter estimates to the quality of the data used. The problem, however, is not confined to tests of bond market models: bond data are also used by professionals who manage or monitor managers of bond portfolios. In this chapter, we examine the extent to which the management and the performance evaluation of bond fund managers is affected by the quality of the data used.

Increasingly, bond portfolio management is done by creating indexed portfolios that attempt to mimic the behavior of a bond market index. Because the liquidity of a certain bond issue decreases with the age of the bond, one may ask whether bond managers who can only buy liquid bonds are limited in their ability to track the bond market performance.

The performance of index bond funds is measured by the proximity of the time series of their returns to the market index return. This is called tracking error. Frequently used measures of performance for index funds are the beta of that fund, measured statistically from a regression of their return series against the returns to the index that it purports to track, as well as the R^2 of this regression. The ideal values for an index fund are a beta and an R^2 of 1.0.¹

To understand the magnitude of this problem, we regressed the market index calculated using only returns for young bonds (i.e., bonds which were issued up to three years prior to the price quote rate) on the unconstrained market index. This is effectively a regression of the return to the most diversified portfolio of *liquid* bonds on the most-often-used benchmark portfolio. The estimated regression parameters are $\beta = 1.00428$, with a standard error = 0.004, and $\alpha = 0.00008$, with a standard error = 0.000074. The R^2 is 0.999.

The proximity of the estimated relation to the “line of perfect prediction” ($\alpha = 0$ and $\beta = 1.0$) and the very high R^2 suggest that the use of only liquid bonds in index funds does not seriously limit the ability of the fund to mimic a broad market index.

Although the benchmark for index-fund performance evaluation does not require a bond market model, the performance evaluation of *actively managed* funds requires a model of normal performance. Here one often employs the equivalent of the equity market’s model: Assuming a single factor explains returns, one estimates the beta of the evaluated fund using past data, and then analyzes the estimated intercept as well as the residuals of the fund whose performance is being evaluated. Given the preceding results, liquidity-driven price noise may affect the estimation of the risk exposure of managed bond portfolios, and, as a result, the valuation of their performance.

To analyze this problem, we used the return series to the nine age-filtered constant-maturity portfolios as a sample of actively managed portfolios. The returns to these portfolios represent the returns to managed portfolios invested in liquid bonds with various maturities chosen by their managers. Table 7 reports the beta estimates for these portfolios using either the unconstrained bond market index or the age-filtered bond market index as the independent variable.

Table 7 indicates the extent to which the beta estimates differ when a benchmark portfolio containing illiquid bonds is used in the estimation. Note that the differences between the

¹The R^2 statistic tells us what percentage of the total variation in the index bond fund return is explained by variation in the market index return.

TABLE 7

**Beta Estimates for
Constant-Maturity Portfolios**

<i>Portfolio Number</i>	<i>All Data</i>	<i>Filtered Data</i>	<i>Difference</i>
1	0.103	0.102	0.001
2	0.330	0.328	0.002
3	0.526	0.524	0.002
4	0.693	0.690	0.003
5	0.858	0.854	0.004
6	1.016	1.012	0.004
7	1.190	1.185	0.005
8	1.399	1.393	0.006
9	1.860	1.851	0.009

Notes: The estimates are derived by fitting a straight line to the data. Its statistical form is

$$r_{pt} = \alpha_p + \beta_p r_{mt} + e_t$$

where

r_{pt} is the return on the portfolio over the period being studied,

α_p is the intercept of the straight line being fitted to the data,

β_p is the slope of the line,

r_{mt} is the return on the market index over the period being studied, and

e_t accounts for all of the factors not included in r_{mt} .

estimated betas increase with the maturity of the portfolio. This may be a reflection of more prevalent and more pronounced liquidity-related price noise in long-maturity bond price records than in short-maturity bond price records. Consequently, the benchmark performance expected of high-maturity funds will be too low, whereas the benchmark performance expected of short-maturity bond funds will be too high. The differences are, however, statistically insignificant, so there does not seem to be much cause for concern by professional managers.

5. Application: Estimating the Risk Structure of Interest Rates

A few empirical studies (e.g., Fisher 1959, Johnson 1967, Cohan 1967, Silvers 1973, and Boardman and McEnally 1981) have examined the structure of risk premia of corporate bonds. Because of heterogeneity and lower liquidity in the corporate bond market, liquidity-related price noise is likely to be more prevalent and of a larger magnitude among corporate bond price records than among government bond price records. Hence, these studies are prone to be more affected by this noise than similar studies of the government bond market.

A number of researchers have attempted to control for liquidity as a determinant of bond prices in their empirical studies. For example, Fisher (1959), who proxies liquidity by the amount of an issue outstanding, found that liquidity was as important a determinant of risk premia as the operating earnings variability of the firm. In light of this evidence, we turn now to examining the risk premia of corporate bonds *using only filtered price records*.

Zero-coupon bond prices were collected from the data tapes of Shearson Lehman Brothers for the period February 1985 through September 1987. During that period, the set of pure discount bonds contained 119 U.S. Treasury strips and 137 corporate issues representing 42 companies.¹ The yields of corporate bonds above the yield on Treasury strips of the same maturity are the focus of the analysis in this chapter.

The sample period is dictated by the type of bonds examined. Because we analyze yield spreads between pure-discount corporate bonds and U.S. Treasury bonds, the sample must begin in February 1985 when the U.S. Treasury first issued strippable bonds. This does not severely restrict the use of corporate pure-discount bond data because most such instruments were issued beginning in 1983, with the bulk coming in late 1984 and early 1985.

With the corporate bond data we obtained, we could ascertain whether a particular record was a trader-quote price or a matrix price. To avoid using price estimates with the associated liquidity-related errors, we filtered out all price records that were based on matrix pricing. In addition, the data were filtered to eliminate all cases in which

- (1) a rating change occurred during the sample period; or
- (2) the reported price was below (above) the price of shorter (longer) maturity but otherwise identical bond; or
- (3) the bond was economically callable.

The last two filters require some explanation. If, in a given month, a bond, whose maturity is bounded by the maturity of otherwise identical instruments, had a price which was not bounded by the prices of the surrounding instruments, we deemed it likely that an error occurred in the recording of this instrument's price.

Although we wished to restrict ourselves to non-callable bonds, over half the sample of corporate bonds have call features. In our view, most of these call options were not likely to be

¹Stripped U.S. Treasury issues are zero-coupon government bonds. They are originally issued as coupon-paying bonds with each coupon carrying its own identifier (CUSIP number). The existence of separate identifiers for each coupon payment allows for trading in individual coupons. Hence, any holder of the original bond may strip it by selling its coupons to separate buyers. The separately sold coupons remain obligations of the U.S. government.

exercised, because the exercise price was par or above and these are pure-discount instruments. Curiously, the debentures of coupon-paying bonds issued by the firms in our sample were often used for zero-coupon issues without modification. This meant that an economically useless call option was a feature of many of the zero-coupon bonds. Therefore, the third filter screened out bonds that could be called in an “economically rational way,” i.e., callable for a call price below par.

In addition, it is worthwhile noting that all of the corporate and Treasury bonds in the sample are subject to identical Federal tax treatment. Unlike the corporate issues, however, the U.S. government issues are not subject to state and local taxes. For cross-sectional comparison purposes, however, local taxes may still be ignored because all yield spreads are subject to the same local tax.²

From each zero-coupon corporate bond in each month, we subtracted the yield to a zero-coupon government strip with identical maturity. If no government strip with identical maturity existed, the yield on the two strips with maturities most closely bounding the corporate bond were interpolated to obtain the appropriate risk-free zero-coupon yield. These yield differences were then averaged across all bonds of a given class in a given month and then across time.

Table 8 presents these average yield spreads for corporate zero-coupon bonds for the period February 1985 through September 1987 (32 months of observations) grouped by ratings and time to maturity.³ As expected, the risk premia are higher for lower-rated bonds. Nunn, Hill, and Schneeweis (1986) show that data sources can affect evidence such as that presented in Table 8. In a study of corporate bond returns covering the period December 1975 through June 1980, they find that institutionally based prices from Merrill Lynch exhibit increasing mean returns for lower-rated bonds (as we do), but returns from the less-liquid New York and American Bond Exchanges (provided by Moody's) do not.

²The estimates reported in Tables 1 and 2 and plotted in Figures 1 and 2 represent the risk premia for a tax exempt investor (e.g., a pension fund). Market participants claim that these are the dominant players (i.e., the marginal holders) in this segment of the bond market.

³Sarig and Warga (1989b) discuss how the evidence provided here is in accordance with existing theoretical models of the structure of corporate bond risk premia.

TABLE 8

**Yield Spreads for Corporate Zero-Coupon Bonds^{*}
In Percent per Annum
(February 1985 to September 1987)**

<i>Maturity</i>	<i>Rating</i>						
	<i>AAA</i>	<i>AA</i>	<i>A</i>	<i>BBB</i>	<i>BB</i>	<i>B/C</i>	<i>Unrated</i>
0.5 - 2.5 years							
Yield Spread	0.410	0.621	0.775	1.326	1.670	4.996	3.081
Number of bonds	21	74	123	48	64	41	38
2.5 - 4.5 years							
Yield spread	0.232	0.562	0.736	1.275	1.495	4.650	3.232
Number of bonds	11	99	251	152	79	117	96
4.5 - 6.5 years							
Yield spread	N/A	0.620	0.778	1.405	2.730	3.365	3.197
Number of bonds		114	221	59	58	125	119
6.5 - 8.5 years							
Yield spread	N/A	0.620	0.660	N/A	1.878	2.959	3.443
Number of bonds		96	138		51	80	119
8.5 - 10.5 years							
Yield spread	0.626	0.575	0.816	N/A	0.989	2.912	3.099
Number of bonds	24	69	97		10	10	88
10.5 - 12.5 years							
Yield spread	N/A	0.566	0.854	N/A	N/A	N/A	2.478
Number of bonds		64	110				64
12.5 plus years							
Yield spread	0.544	0.544	0.740	N/A	N/A	N/A	2.516
Number of bonds	64	501	510				278

Notes:

* Average yield spreads were calculated as follows. In each month the yield to an individual corporate bond has subtracted from it the yield to a zero-coupon government strip with identical maturity. If no government strip with identical maturity existed, the yield on the two strips with maturities most closely bounding the corporate bond were interpolated to obtain the appropriate risk-free zero-coupon yield. These yield differences were then averaged across bonds in a given month and then across time to produce the results reported for each cell.

6. Conclusion

This study has attempted to characterize and to assess the effect of market liquidity on the accuracy of records of bond prices. Using two independently collected data files, the CRSP Government Bond File and the Shearson Lehman Brothers Government Bond Price File, we investigated the magnitude and nature of price discrepancies between these supposedly identical files of government bond prices.

Our results come in the form of qualified good news. First, the results indicate that the quality of reported prices deteriorates as a bond becomes less liquid. For example, this deterioration was shown to take the form of an increase in the likelihood of no updating of the price record. Our results also indicate that liquidity-related price noise tends to an average of zero. Hence, estimates of means of rates of return to various bond portfolios are little affected by the added noise in price records. On the other hand, the added noise in price records manifests itself in a more pronounced manner in estimates of the covariance matrix. As a result, tests of models of the trade-off between risk and return are more susceptible to liquidity-related problems than are the estimates of the parameters of these models. This point was illustrated for the case of estimating a factor model for the bond market.

Observing that the added price-record noise is strongly correlated with our liquidity proxies, we applied these proxies as filters to the CRSP data and investigated the relative merits of these filters in eliminating suspect data. We found that an age-based filter provides a subset of the data that in several dimensions is better than the subset produced by a bid-ask spread-based filter and better than the unfiltered data.

As another illustration of the importance of liquidity-related price errors, we investigated the way the performance evaluation of managers of government bond portfolios is affected by the differential liquidity in the bond market. We found that managers who are restricted to liquid bonds are not subjected to severe biases in attempting to track published indexes. On the other hand, if the performance of portfolio managers is based on a single-factor (market) model, liquidity-related noise causes a downward bias in the risk estimates. The problem is more pronounced for long-maturity portfolios than for short-maturity portfolios, but is small in magnitude overall.

In contrast to the government bond market, the corporate bond market is significantly less liquid and less homogeneous (this is also true of the municipal sector). Hence, the magnitude of the effects of the problems documented in this paper should serve as a lower bound on the extent of these effects that can be expected to be encountered in the corporate bond market.

Previous studies of corporate bond data have revealed that different data sources can produce conflicting evidence as to whether or not lower-rated bonds have higher risk premia (they should). As an illustration, we estimated corporate bond risk premia for various rating groups using only "clean" data. Our estimates show that a lower rating implies a higher risk premium.

Appendix

Firms and Maximum Number of Bonds in Any Given Month (February 1985 to September 1987)

	<i>Firm</i>	<i>Maximum Number of Bonds</i>
1	Allied Corp	5
2	Amax Inc.	1
3	Archer-Daniels-Midland	2
4	Asian Development Bank	1
5	Associated Dry Goods	2
6	Avco Financial Services	1
7	Bally Manufacturing Corp.	3
8	Bankamerica Corp.	3
9	Barclays American Corp.	1
10	Cit Group Holdings	1
11	Chase Manhattan Corp.	1
12	Chrysler Financial	2
13	Cities Service	2
14	Commercial Credit	1
15	Conair Acquisition Corp.	6
16	Continental Ill. Corp.	1
17	Cook Inlet Communication	3
18	Credithrift Financial	1
19	Exxon Shipping Co.	1
20	Fluor Corporation	2
21	Ford Motor Credit	1
22	Franklin Savings Assoc.	1
23	General Motors Acpt. Corp.	3
24	Georgia Pacific	4
25	Greyhound Lease + Fin.	1
26	Hospital Corp. of America	6
27	IBM Credit Corporation	1
28	ITT Financial	6
29	Intel Overseas Corp.	1
30	Manufacturers Hanover Corp.	1
31	McDonalds Corp.	2
32	Metromedia Broadcasting	6
33	NRM Energy Co., L.P.	1
34	Owens Illinois Inc.	1
35	Penney, J. C.	2
36	Pepsico Cap. Res. Inc.	27
37	Price Communications Corp.	1
38	Prudential Realty Sec. I	1
39	Rapid American Corp.	20
40	Telemundo Group Inc.	2
41	Transamerica Financial	5
42	Turner Broadcasting Sys.	4
	Total	137

References

- Boardman, C.M. and R.W. McEnally. 1981. "Factors Affecting Seasoned Corporate Bond Prices." *Journal of Financial and Quantitative Analysis* 16 (June): 207-26.
- Cohan, A.B. 1967. "Yields on Corporate Debt Directly Placed." New York: National Bureau of Economic Research.
- Dimson, E. 1979. "Risk Measurement when Shares are Subject to Infrequent Trading." *Journal of Financial Economics* 7 (June): 197-226.
- Elton, E., M. Gruber, and P.G. Nabar. 1986. "Bond Returns, Immunization and the Return Generation Process." Unpublished manuscript, New York University.
- Fama, E. 1984. "Term Premiums in Bond Returns." *Journal of Financial Economics* 13 (December): 529-46.
- Fisher, L. 1959. "Determinants of Risk Premiums on Corporate Bonds." *The Journal of Political Economy* 67 (June): 217-37.
- Garbade, K.D. and W.L. Silber. 1976. "Price Dispersion in the Government Securities Market." *The Journal of Political Economy* 84:4 (August): 721-40.
- Gultekin, M.B. and R. Rogalski. 1985. "Government Bond Returns, Measurement of Interest Rate Risk, and the APT." *Journal of Finance* 40:1 (March): 43-62.
- Johnson, R.E. 1967. "Terms Structures of Corporate Bond Yields as a Function of Risk of Default." *Journal of Finance* 22 (May): 313-45.
- Litzenberger, R.H. and J. Rolfo. 1984. "Arbitrage Pricing, Transaction Costs and Taxation of Capital: A Study of Government Bonds with the Same Maturity Date." *Journal of Financial Economics* 13 (September): 337-51.
- McCulloch, J.H. 1987. "The Monotonicity of the Term Premium: A Closer Look." *Journal of Financial Economics* 18 (March): 185-92.
- Nunn, K., J. Hill, and T. Schneeweis. 1986. "Corporate Bond Price Data Sources and Return/Risk Measurement." *Journal of Financial and Quantitative Analysis* 21 (June): 197-208.
- Roll, R. 1970. *The Behavior of Interest Rates: The Application of the Efficient Market Model to U.S. Treasury Bills*. New York: Basic Books.
- . 1984. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance* 39 (September): 1127-40.
- Sarig, O., and A. Warga. 1989a. "Bond Price Data and Bond Market Liquidity." *Journal of Financial and Quantitative Analysis* 24 (September): 367-378.
-

_____. 1989b. "Some Empirical Estimates of the Risk Structure of Interest Rates." *Journal of Finance* 44 (December): 1351-60.

Scholes, M. and J. Williams. 1977. "Estimating Betas from Non-synchronous Data." *Journal of Financial Economics* 5 (December) 309-27.

Silvers, J.B. 1973 "An Alternative to the Yield Spread as a Measure of Risk." *Journal of Finance* 28 (September): 933-55.

Warga, A.D. 1989. "Bond Returns, Liquidity, and Missing Data." First Boston Working Paper, Series FB-89-01, Columbia University (January).