Measuring Limits of Arbitrage in Fixed-Income Markets

Jean-Sébastien Fontaine Guillaume Nolin
Bank of Canada

April 2019

Abstract

An emerging literature relies on an index of limits of arbitrage in fixed-income markets. We analyze the benefits of an index that is model-free, robust and intuitive. This new index strengthens the evidence that limits of arbitrage proxy for risks priced in the cross-section of returns. Trading simulations show that the new index improves identification of limits of arbitrage because it bypasses a noisy estimation step. Relative value indices in the US, the UK, Japan, Germany, Italy, France, Switzerland and Canada exhibit strong commonality and high correlations with local volatility and funding conditions. The indices are updated regularly and available publicly.

Keywords: Limits of Arbitrage, Fixed-Income, Sovereign Bonds

JEL Classification: G12

We thank Sébastien Bétermier, Bruce Chen, Jens Christensen, Darrell Duffie, Bruno Feunou, Meredith Fraser-Ohman, Étienne Giroux-Léveillé, Toni Gravelle, Sermin Gungor, Madhu Kalimipalli, Natasha Khan, Gitanjali Kumar, Stéphane Lavoie, James Pinnington, Francisco Rivadeneyra, Adrien Verdhelan, Harri Vikstedt, Jun Yang, Jing Yang and Bo Young Chang for comments suggestions. We also thank James Pinnington for invaluable research assistance. Email: jsfontaine@bankofcanada.ca

1 Introduction

Bonds from the same issuer and with the same cash flows should have the same prices. This is the law of one price. Yet, deviations from the law of one price are pervasive in fixed-income markets. Panels (A) and (B) of Figure 1 show the yields to maturity for every outstanding bond issued by the United States Treasury for two days: one in 2008 when deviations were large, and one in 2014 when deviations were small.

Deviations like those in Figure 1 are common, persistent and significant, even after accounting for transaction costs (Amihud and Mendelson, 1991). In an important study, Hu et al. (2013) (HPW, hereafter) show that a broad measure of these deviations—the noise index—reveals information about the quantity of capital available for arbitrage deviations.[1] One key insight in HPW is that a measure of limits of arbitrage can reveal a risk premium in the cross-section of asset returns, across a range of market conditions and beyond fixed-income markets.

Estimates of the noise index now play an important role in an emerging literature studying the impact of limits of arbitrage. Significant examples include Malkhozov et al. 2016 in international markets; Musto et al. 2018 in the case of investor clienteles; D'Amico and Pancost 2017 in repo markets; and Goldberg and Nozawa 2018 in identifying liquidity and demand shocks. A noise index is also used in broader contexts to discuss the impact of limits of arbitrage or capital scarcity on market conditions (e.g., Bisias et al. 2012; Adrian et al. 2015; Dudley 2016).

The construction of a noise index involves two steps. The first step is to choose and estimate a smooth yield curve. Figure 1 shows one estimate of the yield curve.[2] The second step is to calculate the noise index by aggregating the differences between the

[1]The effectiveness of arbitrage capital can be limited by funding frictions and market segmentation (see e.g. Duffie 1996, Vayanos and Weill 2008, Vayanos and Vila 2009, and Fontaine and Garcia 2015 for a review).

[2]Figure 1 shows the yield curve introduced by Gürkaynak et al. 2007 based on Svensson 1994, which is available at the Federal Reserve Board website.

yield of each bond and the curve estimated in the first step (i.e., the root mean square of these fitting errors). This is an intuitive approach that can be explained easily. Yet, the first step introduces measurement errors because it is prone to estimation and specification risk. Measurement errors introduce a well-known attenuation bias that lowers OLS estimates and reduces the power of tests. This weakens the evidence of limits of arbitrage and underestimates their impact. These measurement errors have the potential to be large, since implementation of the first step can appear arbitrary. Judgment is needed to choose a curve-fitting method and to include or exclude bonds for estimation.[3]

In this paper, we propose a different construction of the limits-to-arbitrage index that is model-free and intuitive.[4] To be clear, our goal is to provide a robust measure that researchers can use to broaden and strengthen the evidence in HPW. For this purpose, the key is to eliminate the measurement errors introduced by yield curve estimation. To do this, we build on the idea that if there are no limits of arbitrage, a bond and a replicating portfolio of bonds with the same risk should offer the same expected returns. Conversely, if frictions and transaction costs limit the forces of arbitrage, the replicating portfolio provides a yardstick to measure the relative value of a bond.[5]

To compute relative value, our approach uses a replicating portfolio inspired by the butterfly trade, a common and simple strategy used by traders to exploit deviations. The bond-level relative value is the yield difference between the bond and its replicating portfolio. The relative value index then aggregates bond-level relative

[3]See Bliss 1996 or Bolder and Gusba 2002 for a review. For instance, HPW exclude bonds with maturities less than 1 year or more than 10 years, while Gürkaynak et al. 2007 exclude bonds with less than 3 months to maturity as well as recently issued bonds (on-the-run bonds).

[4]This contrasts with Fontaine and Garcia 2012 and D'Amico and Pancost 2017, who use dynamic no-arbitrage models to identify deviations from no-arbitrage.

[5]Alternatively, a relative value trade could involve the perfect replication of cash flows. However, this is impractical and costlier than replicating interest rate risk, as it requires extensive trading in the relatively illiquid market for bond strips.

values every day. We use the label "relative value index" for clarity, to distinguish it from implementations using a yield curve model, which we label "curve-based index". We construct the relative value indices for the US, the UK, Japan, Germany, Italy, France, Switzerland and Canada. The indices are available publicly and updated regularly.[6]

We illustrate the benefits of the relative value index in three applications. First, we test whether the relative value index is a better proxy than the curve-based index for risks that are priced in the cross-section of asset returns. In the cross-section of de-levered call and put S&P 500 option returns (Constantinides et al., 2013), we augment the CAPM with either the relative value or the curve-based index, or both. The price of risk estimates are significant and essentially the same, close to -0.17. The combination of both indices also yields a significant estimate for relative value. By contrast, the curve-based index becomes insignificant and the sign flips. The differences are starker in the cross-section of corporate bond returns. Augmenting the CAPM with relative value raises the R^2 to 64 percent but only to 11 percent with the curve-based index. The price of risk estimate is -0.16 for relative value—similar to estimates in option returns—but the estimate is much smaller and insignificant for the curve-based index. The combination of both indices again flips the sign for the curve-based index.

Overall, the results using relative value provides stronger evidence to support the insight in HPW that a limits of arbitrage index can proxy for risks that are priced in several asset classes. For both asset classes, a change by two standard deviations of βs increases average returns by around 6 percent annually.[7] The question is why the

[6] See the Bank of Canada's website `https://www.bankofcanada.ca/2017/10/staff-working-paper-2017-44/`.

[7] The price of risk is negative, as expected. An increase of limits-to-arbitrage index is associated with lower returns on impact (β is negative), and options with lower β are riskier and offer higher average returns.

curve-based index is producing weaker results. Consistent with measurement errors, the combination of curve-based and relative value index in asset pricing tests produces robust and significant estimates for relative value but not for the curve-base index.[8] In addition, the R^2 does not increase, showing that both indices contain overlapping information.

Our second application evaluates a pseudo-trading strategy that invests in the portfolio of trades that uses bond-level relative value or curve-based signals. We expect that if a signal correctly identifies limits of arbitrage, its portfolio should generate positive profits, precisely because the pseudo-trading strategy ignores other risks and constraints that arbitrageurs face. The returns from this portfolio are an objective measure of economic content: higher returns mean that costs and the risks of arbitrage were larger. We find that relative value produces an average monthly return of 0.09 percent between 1988 and 2017. This accounts for bid-ask spreads, and we assume a conservative level of capital at risk.[9] This result provides new evidence that the insight in HPW has economic content. By contrast, we obtain an average monthly return of −0.19 percent when using a curve-based signal. Comparing portfolio returns in every two year sub-sample produces similar results. This result suggests that using a curve-based index introduces measurement errors and blurs the evidence.

The trading simulation provides a direct way of checking that small measurement errors at the bond level explain the difference between indices. To check this, we repeat the trading exercise but increase the threshold that a signal must meet before we implement new trades. We find that a small threshold on the curve-based index

[8]Small measurement errors in the construction of the index can have a large impact in asset pricing tests. We follow HPW and use the common approach to use the first difference of the index as a risk factor. Under classical assumptions, first-difference amplifies the impact of measurement errors if the index is persistent.

[9]The scale of the monthly average returns is determined by the choice of capital at risk in the simulation, but the ranking is not.

(a few basis points) quickly produces positive returns. Yet, relative value still outperforms for any level of the threshold, in large part because the thresholds eliminate less profitable trades.

Finally, our third application shows that the relative value is robust and can be applied easily in different markets. In particular, we do not need to check the adequacy of yield curve models in different markets. We construct relative value indices for the US, Canada, the UK, Germany, France, Italy, Switzerland and Japan. First, we repeat the trading simulation in each country. Relative value outperforms the curve-based signal in every country. Then, we check that each country's index is correlated with local proxies for limits of arbitrage. We find that the relative value index in each country is highly correlated with volatility in the local stock market (e.g., the VIX in the US). We also find that the relative value index in each country is correlated with the local funding liquidity (e.g., LIBOR-OIS spread for the US). In the cross-section of countries, we find that relative value indices are highly correlated with each other as well as with US proxies for volatility and funding rates. This pattern suggests that the limits of arbitrage are largely common for the sovereign bond markets of these large advanced economies, and proxy for a common factor related to global sources of risk.

Overall, the results confirm that using relative value identifies economically meaningful limits of arbitrage. Using relative value over the curve-based method can provide an improved proxy for risks priced in the cross-section of asset returns and an improved means to detect apparent profit opportunities in a trading simulation.

Our measure of relative value is the yield difference between a bond and its replicating portfolio. This paper is structured as follows. Section 2 describes the data and the method to calculate the model-free relative value measure for individual bonds, as well as the construction of relative value indices. Section 3 describes the results of asset pricing tests using CAPM augmented by relative value and curve-based indices. Section 4 compares relative value and the curve-based indices using a pseudo-trading strategy. Section 5 compares relative value indices internationally and correlates them to local proxies for volatility and funding liquidity.

2 Methodology

The model-free measure of relative value is inspired by the butterfly bond trading strategy, used by market participants to profit from deviations between the prices of similar bonds. The butterfly strategy typically involves a combination of bonds with similar durations, which are the wings of the butterfly: one bond with a lower duration and another with a higher duration, with weights to match the duration of the target bond.

For our measure, we use three bonds to match the duration, convexity and par value of the target bond to neutralize differences in maturities and coupon rates. Hence the portfolio replicates the risk of the target bond. Assuming the absence of arbitrage, the target bond and the replicating portfolio should have the same expected returns. But with limits of arbitrage, the difference between their yields is a proxy for costs and frictions preventing arbitrageurs from driving deviations to zero.

2.1 Data

We compute a daily relative value index for bonds issued by Canada, France, Germany, Italy, Japan, Switzerland, the UK and the US. As benchmarks, we compute a curve-based index along the lines of HPW in each country. All indices are updated regularly and available publicly.[10]

[10]See the Bank of Canada's website `https://www.bankofcanada.ca/2017/10/staff-working-paper-2017-44/`.

We use end-of-day clean prices and yields to maturity. For comparability with HPW, we include bonds that have between 1 and 10 years remaining to maturity, and we exclude inflation-linked bonds, variable-rate bonds and bonds with embedded options. Other restrictions and filters are described in the Appendix. For the US, we use prices and yields of Treasury bonds data between 1988 and December 2017, from the Center for Research in Security Prices database (CRSP). Before 1987, the different tax treatments of coupons and principal payments affected the relative valuation of bonds.[11] For Canada, we use daily prices between 1994 and December 2017. Similar to the US, coupon and capital gains in Canada had different tax treatments before 1994.[12] For bonds issued by the British, French, German, Italian, Japanese and Swiss governments, we use daily data from Markit between 2005 and December 2017. Table 1 reports the number of observations for each country as well as summary statistics for the number of bonds available daily. Japan and the US have the largest number of bonds, Switzerland the lowest.

2.2 Calculating Portfolio Weights

The replicating portfolio combines three bonds from the same issuer. To select these bonds, we sort all outstanding bonds by duration. The first bond that we include has the closest duration lower than the target bond. The second bond has the closest duration but higher than the target bond. That is, the first two bonds are closest to the target bond but on opposite sides of it on the duration axis. The third bond is next to the bond with the smallest absolute difference in duration with the target bond (of the first two bonds included). The bonds in the replicating portfolio are

[11] In the US, the Tax Reform Act of 1986 introduced equal tax treatment of coupon and principal payments. Before that, tax deductions for coupon payments made higher coupon bonds more expensive. See Green and Ødegaard 1997 for a detailed discussion.

[12] We use data from StatPro before April 2009 and from FTSE TMX thereafter. The dataset available from FTSE TMX was formerly known as PC-Bond and DEX. In Canada, tax deductions for principal payments made higher coupon bonds cheaper before 1994, when the lifetime capital gains deduction was eliminated for property acquired after February 22, 1994

thus the closest substitutes to the target bonds.

We label the target bond with b and label the bonds in the replicating portfolio with $i = 1, 2, 3$. The duration and convexity of the bonds are denoted by d_i and c_i respectively. We compute modified duration and convexity using mid yields. Every day and for every bond, the weights w_i in the replicating portfolio as a fraction of par value are chosen to satisfy:

$$d_1 w_1 + d_2 w_2 + d_3 w_3 = w_b d_b \tag{1}$$

$$c_1 w_1 + c_2 w_2 + c_3 w_3 = w_b c_b \tag{2}$$

$$w_1 + w_2 + w_3 = w_b = 1. \tag{3}$$

This system of equations requires that the replicating portfolio has the same duration (Equation 1), the same convexity (Equation 2) and the same par value as the target bond (Equation 3). Without loss of generality, we normalize the weight w_b to 1.

To eliminate extreme portfolio weights, we impose the following conditions on the weights:

$$0 \leq w_i \leq 2/3 \qquad \forall i \in \{1, 2, 3\}. \tag{4}$$

The left-hand side of this constraint means that the replicating portfolio does not mix short and long positions. The right-hand side of this constraint means that relative value averages information from the prices of several neighbouring bonds. If the inequality in Equation 4 is binding, the system defined by Equations (1 to 3) may have no solution. In this case, we relax the convexity condition in Equation 2, but we minimize the difference in convexity between the target and the replicating portfolio:

$$\underset{w_1, w_2, w_3}{\operatorname{argmin}} (c_1 w_1 + c_2 w_2 + c_3 w_3 - c_b), \tag{5}$$

subject to Equations 1, 3 and 4.[13] We have verified that differences in convexity do not drive our results.

2.3 Computing Relative Value with Bid and Ask Prices

Every day, we compute the yield to maturity of the replicating portfolio using the weights w_1, w_2, w_3 as well as the bid, mid, and ask yields (y_i^{bid}, y_i^{mid} and y_i^{ask}) of bond i:

$$\ddot{y}^{bid} = y_1^{bid} w_1 + y_2^{bid} w_2 + y_3^{bid} w_3 \tag{6}$$

$$\ddot{y}^{mid} = y_1^{mid} w_1 + y_2^{mid} w_2 + y_3^{mid} w_3 \tag{7}$$

$$\ddot{y}^{ask} = y_1^{ask} w_1 + y_2^{ask} w_2 + y_3^{ask} w_3. \tag{8}$$

When the yield for the target bond is higher than that of the replicating portfolio, the arbitrage strategy to profit from this deviation is to buy the target and short the replicating portfolio. The opposite strategy is profitable when the observed yield for the target is lower than that of the replicating portfolio. The difference in expected returns between the target and the replicating portfolio are unaffected by common movements in yields (up to a second-order approximation). Limits of arbitrage are likely to be driving the wedge in expected returns, which is also the intuition behind the yield curve method in HPW. However, this portfolio is sensitive to large or uneven changes in yields. This is mitigated by the fact that all the bonds included in the portfolio have very similar maturities.

Therefore, relative value is $rv_b^+ = y_b^{ask} - \ddot{y}^{bid}$ when the target bond is expensive relative to its replicating portfolio, while relative value is $rv_b^- = -y_b^{bid} + \ddot{y}^{ask}$ when the target bond is cheap relative to its replicating portfolio. Note that rv_b^+ and rv_b^-

[13] We use the `lsqlin` function with the interior-point algorithm in MATLAB to solve the constrained linear least-squares problem. The initial point for the solution process is a vector with equal weights 1/3.

cannot be strictly positive at the same time: the target cannot be simultaneously expensive and cheap relative to its replicating portfolio. However, rv_b^+ and rv_b^- can be negative at the same time because of bid-ask spreads. The relative value of target bond b is defined as:

$$rv_b \equiv \begin{cases} rv_b^+ & \text{if} \quad rv_b^+ > 0 \\ 0 & \text{if} \quad rv_b^+ < 0 \;\; \text{and} \;\; rv_b^- < 0 \\ -rv_b^- & \text{if} \quad rv_b^- > 0, \end{cases} \tag{9}$$

which implies that rv_b takes negative value $-rv_b^-$ when the target is relatively cheap. Note that rv_b is set to zero if rv_b^+ and rv_b^- are negative, since the absence of arbitrage holds in this case once we account for transaction costs. An alternative definition simply uses mid yields:

$$\tilde{rv}_b = -y_b^{mid} + \ddot{y}^{mid}.$$

This definition produces larger relative values than our base case and overstates the presence of limits of arbitrage. We also implement the curve-based measure $crv_{b,j}$, which is simply the distance between bond j and an estimate of the yield curve. To estimate this curve, we closely follow the implementation in HPW. However, we introduce an adjustment for the bid-ask spread so that this benchmark and the relative value measure are comparable (see Appendix A.3).

2.4 Two Examples

To help understand how relative value and curve-based measures differ, Figure 2 shows two specific examples. In each case, we present observed yields as well as the fitted yields implied by relative value or the curve-based measures. We focus on one date in August 2017 and one date in December 2017, and for each date we zoom in on a short section of the curve to make the differences easier to see. We label these Cases I and II for parsimony.

Panel (A) shows the observed yields and the yields fitted using relative value in Case I, zooming in on durations between two and four years. The curve is upward sloping and two features stand out. There is a small downward hump for the yields of bonds with duration between two and three years, and there is an area where the curve appears dislocated for durations close to four years. The fitted yields follow the small hump closely, and only a handful of yields seem to offer potential opportunities for arbitrage. The dislocated part of the curve indicates another handful of yields that may offer potential arbitrages.

Panel (C) shows the observed yield and the curve-fitted yields. There is a visible contrast in the way the curve-based measure interprets the data. The estimated curve mostly misses and stays above the small downward hump. This generates a large number of deviations that are systematically negative. At the other end, the estimates draw the curve too low, below most of the dislocated section, which generates several positive deviations.

Case II offers another useful illustration. Panel (B) shows the observed yields and yields fitted using relative value, zooming in on durations between six and close to nine years. It shows a smooth curve. Relative value identifies only a handful of potential arbitrage opportunities. Surprisingly, the curve-based method Panel (D) identifies many deviations. An inspection of the entire fitted curve suggests that parameter estimation involves a trade-off between the fit of different sections of the curve. Specifically, we find that a bump in the shorter maturities produces a series of negative deviations similar to those observed in Case I.

To summarize, Case I shows that the parametric curve can be too restrictive to fit some areas of the curve, which introduces measurement errors. In addition, Case

II shows that a parametric model can introduce fitting errors even in sections of the curve that seem smooth. This arises if the estimation balances between fitting errors across different sections of the curve. The illustrations in Figure 2 suggest that using a fitted curve introduces measurement errors with limited economic content. The empirical section will confirm this.

2.5 The Relative Value Index

The relative value index rv^j for country j is the root mean square of the individual relative values $rv_{b,j}$ for this country:

$$rv^j = \sqrt{\frac{1}{N^j}\sum_{b=1}^{N} rv_{b,j}^2}, \tag{10}$$

where N^j is the number of bonds in our sample for country j on a given day. Note that bonds with $rv_b^+ < 0$ and $rv_b^- < 0$ are effectively excluded from this index, since then $rv_b = 0$. These are bonds with quotes within the bid-ask spread from the yield of their replicating portfolio. Using root mean square parallels how HPW computes the index. Also following HPW, the index excludes observations where rv_b is greater than four times the sample standard deviation. We also implement a curve-based index of limits-to-arbitrage:

$$crv^j = \sqrt{\frac{1}{N^j}\sum_{b=1}^{N} crv_{b,j}^2}. \tag{11}$$

Figure 3 shows the relative value index and the curve-based index constructed according to Equations 10 and 11, respectively. Since these two indices measure similar limits of arbitrage, it is not surprising to see that they track each other very closely. The correlation between the indices is 0.86. One visible difference is the higher level of the curve-based index for most of the sample. That is what we should expect if relying on a parametric curve adds measurement errors. Another difference is that the correlation of monthly changes (0.60) and the correlation of daily changes (0.33) are much lower than the correlations between the levels. This will be important in asset pricing tests below.

3 Asset Pricing Tests

Following HPW, we check that the relative value index proxies for risks that are priced in the cross-section of returns: whether βs match average returns. For our purpose, this test asks whether relative value provides stronger evidence for limits of arbitrage than a curve-based index. This establishes that relative value is a useful proxy for risk: it describes an important risk-return relationship embedded in market prices.

3.1 Test Assets

We first consider the cross-section of option returns. Constantinides et al. 2013 create portfolios of call and put options that are de-levered. Removing the effect of leverage in option contracts makes these portfolios appropriate in linear asset pricing tests. More importantly, Constantinides et al. 2013 find that the returns of de-levered options—especially put options returns—are puzzling from the perspective of the simple CAPM model, but that a proxy for illiquidity goes a long way toward resolving the puzzle. Thus, option returns are good candidates for our tests.

We also consider the cross-section of corporate bonds. Corporate bonds are good candidates for our test as well, since corporate bonds are notoriously illiquid and because our index of limits of arbitrage specifically focuses on fixed-income markets. Bao et al. 2011 find that a proxy for aggregate liquidity explains individual bond yield spreads with large economic significance. Fontaine and Garcia 2012 find that bonds with lower ratings have larger exposures to their funding liquidity measure.

There is no standard corporate bond portfolio for asset pricing tests. We double-sort corporate bonds on their exposures to market-wide illiquidity and to market-wide volatility.[14] The exposure of each corporate bond is measured by the β of monthly bond returns. The βs are estimated using a three-year rolling window. At the beginning of each year, we form $5 \times 5 = 25$ portfolios. We then track their returns for one year before rebalancing the portfolios again at the start of the next year.

3.2 Results

Table 2 reports estimates of the prices of risk obtained for relative value Δrv and for the curve-based index Δcrv. In all cases, we include market returns as a risk factor and as a test asset to discipline the estimate for the price of market risk (see Lewellen et al. 2010). Panel (A) reports results using option portfolios. Results are tabulated for put and call options, puts only, and calls only. A simple message emerges: the price of risk for relative value is large, negative and significant. Augmenting the CAPM with relative value raises the cross-section R^2 from 24 to 85 percent for puts and from 56 to 85 for calls.[15] The inspection of the portfolio reveals negative βs: an increase of the relative value index is associated with lower returns (unreported). Combined with a negative price of risk, moving from a β of 0 to the -0.5—the β of a portolio of puts in the augmented CAPM—increases average monthly returns by around 1.5 percent.

If we compare Δcrv and Δrv, the relative value index provides some improvements, especially for put options. Augmenting the CAPM with Δcrv only increases

[14]Following Bai et al. 2015 closely, we combine several databases to construct a long sample of monthly individual corporate bond returns. To sort assets, market-wide illiquidity is the daily Amihud 2002 price impact measures averaged across stocks and over each month. Market-wide volatility is the monthly realized volatility computed from daily squared returns for the stock index.

[15]Consistent with Constantinides et al. 2013, we find a wider dispersion of average returns for put options, which was precisely the puzzling observation from the perspective of a simple CAPM.

the R^2 to 55 percent in the cross-section (compare with 85 percent for Δrv). When we augment the CAPM with both Δrv and Δcrv, we find tell-tale signs that relative value subsumes the information content of the curve-based index and offers a better signal. The R^2 in the cross-section of returns does not increase relative to using Δrv on its own. The estimated price of risk for relative value remains negative and significant, but, in contrast, the estimate for the curve-based index becomes insignificant and sometimes flips sign.

Panel (B) of Table 2 reports results using corporate bond portfolios as test assets. Using relative value or a curve-based index produces different conclusions. The price of risk for relative value is essentially the same as in Panel (A) for options. By contrast, the price of risk for the curve-based index is small, not statistically significant and changes sign in a specification that includes both Δrv and Δcrv. The R^2 in the cross-section of bond returns is 64 percent for Δrv but only 11 percent for Δcrv.

Overall, the results using relative value provide strong evidence for the insight in HPW that an index of limits of arbitrage can proxy for important risk priced in financial markets. The estimates for the price of risk are more robust. The dispersion of relative value βs is also similar in the cross-section of bonds and options, so that the spread in returns explained by relative value is similar for bonds and options. The evidence based on the curve-based index seems blurred by measurement errors. We check this possibility in the next application.

4 Relative Value in Trading Simulations

We use a pseudo-trading strategy to measure the economic content of the relative value index. The idea is that a signal that correctly identifies limits of arbitrage should generate positive profits, but that a signal weakly correlated or uncorrelated with limits of arbitrage should generate no profit. The pseudo-trading strategy ignores some of the risks and constraints that arbitrageurs faced. This is a feature and not a bug. Higher returns mean that arbitrageurs faced more risks or tighter constraints. Note that Duarte et al. 2007 show that these strategies amount to more than "picking up nickels in front of a steamroller" and that a patient, unconstrained investor should earn a positive return over time when following this strategy. Therefore, the trading simulation provides objective economic criteria to evaluate the quality of the bond-level relative value measures that constitute the index in Equation 10. The results show that relative value provides improvements relative to a curve-based measure. The results also show that the benefits of the relative value index here and in Section 3 are due to measurement errors in a curve-based index. It is important to note that we are not looking for the optimal trading strategy.

4.1 Initiating Trades

Every day, we use either relative value or the curve-based measure as a signal to determine whether a bond becomes the target of a new trade. If it does, we implement the same trading strategy irrespective of the signal. We then aggregate trades in a portfolio and track portfolio returns. The only difference between the portfolios is the signal used to generate trades.

We initiate a new trade when the absolute value of the signal is greater than some threshold: $|rv_b| > \tau$ or $|crv_b| > \tau$. If the bond is a target, we implement the same butterfly strategy irrespective of which signal is used. Each trade initially has zero duration, near-zero convexity, and thus carries little interest rate risk.[16] Note that we adapt the curve-based measure to account for the bid-ask spreads (see Appendix A.3). Otherwise, the curve-based measure would be at a disadvantage as a trading signal.

If $rv_b > \tau$, the target bond appears expensive. The trading strategy is therefore

[16]Trades are initiated when there are more than five days of pricing data remaining for the target bond and each of the constituents of the replicating portfolio.

to buy one unit of this bond and to sell one unit of the replicating portfolio (with weights given by Equations 1 to 4). The opposite strategy is implemented when the target appears cheap. Consistent with the composition of the index, this bond will not be targeted again before this position is liquidated. We repeat the same exercise using $crv_b > \tau$ as the signal. We report results for $\tau = 0, \ldots, 40$ basis points (bps).

The case $\tau = 0$ is not meant as a realistic strategy. But it is the relevant baseline, since both the curve-based index and our relative value index include all bonds that exceed this threshold ($|crv_b| > 0$ and $|rv_b| > 0$, respectively). Note that trades that satisfy this threshold offer potential profits at least as large as the bid-ask spread. As we will show, larger values of $\tau > 0$ are more realistic and lead to larger potential profits.

4.2 Liquidating Trades

Following Duarte et al. 2007, all trades are held until convergence, which is when the mid of the target bond y_b^{mid} is equal to the mid of the replicating portfolio $\ddot{y}^{mid}$, in the case of relative value, or when y_b^{mid} is equal to the fitted value $\hat{y}_b$ in the case of the curve-based signal. Positions are also liquidated (i) when the target bond matures in fewer than three months, (ii) when data become unavailable, (iii) after one year or (iv) at the end of the sample.

4.3 Computing Trade Returns

We track the performance of each trade over time. It is very important to account for the cost of carry in the computation of returns. The carry varies across trades due to different initial investments, due to different coupon cash flows, and due to interest earned or spent.

We follow Duarte et al. 2007 to compute returns. For this purpose, it is useful to think of a specialized fund that targets a single bond. The equity value E_t of this fund at time t is given by:

$$E_t = E_0 + B_t + P_t, \tag{12}$$

where E_0 is the initial equity or margin set aside to cover for potential losses; B_t is the fund's cash position (net borrowing if $B_t < 0$); and P_t is the marked-to-market value of the target bond and its replicating portfolio, accounting for the bid-ask spread.[17]

We can follow the equity E_t of this fund given the initial equity E_0, the evolution of its cash position B_t and the market value of bonds. For $t > 0$, the marked-to-market value P_t is computed as if the positions were liquidated at end-of-day bid and ask prices, accounting for accrued interest. At $t = 0$, the market value P_0 is recorded at the bid and ask prices of the initial transactions.

The evolution of the cash position B_t is given by:

$$B_{t+1} = B_t(1 + rf_t) + C_{t+1}, \tag{13}$$

with initial cash position $B_0 = -P_0$, where rf_t is the overnight rate, and C_{t+1} are subsequent coupon received or due.[18] Duarte et al. 2007 choose the initial equity E_0 such that the sample volatility of returns is 10 percent on average across funds. For simplicity and comparability across signals, we choose $E_0 = 10$, which is enough to cover any loss. The volatility implied by this choice varies across countries. We

[17] We use the following standard settlement convention: one business day for US Treasury bonds and those of other countries except Canada, two days for Government of Canada bills and bonds with fewer than three years before maturity, three days for Government of Canada bonds with more than three years remaining before maturity.

[18] We use the Actual/360 basis to calculate the daily interest paid. For simplicity, we use the target overnight inter-bank rate or its equivalent. In the US, we use the effective Fed funds rate if available. Otherwise, we use the mid-point of the Federal Reserve official target range when available. For Canada, we use the Bank of Canada target rate. For the UK, we use the Bank of England's official Bank rate. For euro area countries, we use the European Central Bank's main refinancing operation rate. For Switzerland, we use the Swiss National Bank's official LIBOR target. For Japan, we use the Bank of Japan's policy rate.

verified that the volatility is around 10 percent or less.

Given the evolution of the equity value $E_{i,t}$, we can compute daily returns $R_{i,t+1}$ for each fund:

$$R_{i,t+1} = \frac{E_{i,t+1}}{E_{i,t}} - 1. \tag{14}$$

Following Duarte et al. 2007, the index returns $\bar{R}_{t+1}$ is given by:

$$\bar{R}_{t+1} = \frac{1}{I}\sum_{i}^{I} R_{st+1} \quad if \quad I \geq 1 \tag{15}$$

$$\bar{R}_{t+1} = rf_t \quad if \quad I = 0, \tag{16}$$

where I is the number of funds in the index. We compute the index returns separately for trades initiated with relative value and a curve-based signal. The index returns correspond to a portfolio with equal weights across funds and that is rebalanced daily. If no fund is active on a given day, the index return for that day is the risk-free rate. The initial value of the index is set to $V_0 = 100$ and the evolution of the index is simply $V_{t+1} = V_t \times (1 + \bar{R}_{t+1})$.

4.4 Portfolio Returns

Table 3 reports summary statistics of portfolio returns. The average monthly returns of the portfolio of trades based on relative value and the curve-based index is 0.09 percent and -0.19 percent, respectively. Panel (A) of Figure 4 reports the cumulative index returns. Using estimated curves generates returns that are consistently negative. Between 1988 and 2017, this difference in returns accumulates and creates a large difference in outcomes. We checked that using estimated curves produces negative returns for every non-overlapping two-year period between 1988 and 2017 (unreported).

The portfolios experience negative returns around the financial crisis in 2008 (see

Figure 4). This period contributes to the negative skewness and the kurtosis reported in Table 3. However, the value of the relative value portfolio quickly recovers and exhibits substantial gains in the post-crisis period. The daily rebalancing is one reason behind this resilience during the crisis and the subsequent recovery. The portfolio keeps equal weights on arbitrage trades. Hence, as the crisis deepens and new apparent opportunities arise, rebalancing involves reducing the size of existing positions with negative returns to establish positions in new opportunities.

The relative value and curve-based portfolios have similar skewness (-0.32 and -0.34), while Koijen et al. 2018 report a small put positive skewness (0.47) for the carry strategy in the US Treasury market. The portfolio based on relative value signals has a lower kurtosis than the portfolio based on the curve-based signal (11.75 and 16). Koijen et al. 2018 report a kurtosis of 10.46 for the carry portfolio, and they also report average monthly returns of 0.46 percent between 1983 and 2012.

The baseline results that we report have lower returns, more negative skewness and more kurtosis than a carry strategy implemented in the US Treasury market. However, our baseline strategy uses $\tau = 0$, which is not meant as a realistic trading strategy. Instead, these portfolios mirror the construction of the relative value and curve-based indices. This analysis therefore provides an evaluation of the economic content of each index (Equation 10). Section 4.7 shows that total returns from each strategy increase substantially with $\tau > 0$.

The ranking between portfolios is not driven by transaction costs. Panel (B) of Figure 4 reports the cumulative returns if we ignore trading costs due to the bid-ask spread. Again, there is a stark difference between the cumulative returns of the relative value and curve-based portfolios. The ranking is not driven by the choice of the initial equity E_0. Lowering the initial equity increases leverage, produces higher average returns and volatility, but does not change the ranking.

The ranking is not driven by a difference in duration or convexity, which could arise because we do not rebalance trades after initiation. We find that portfolios have slightly positive duration and convexity, but the differences between portfolios are small.[19] The average difference in duration is 0.004 years, which is tiny (slightly more than one day). The average difference in convexity is close to the convexity of a six month T-Bill, which is too small to affect average returns meaningfully (see e.g., Ilmanen 2000).

4.5 Correlations with Risk Proxies

We check whether the ranking could be explained by differences in the exposures to aggregate risk factors. Figure 5 reports the correlations between monthly portfolio returns and several widely used proxies for risk. The broad message is that correlations are fairly low and similar between the portfolios.

Panel (A) reports the correlations with bond market variables: the on-the-run premium, bond yield volatility, the default spread, and the spread between supranational bonds and Treasury bonds. Overall, the portfolios exhibit similar correlations. In particular, the correlations with changes in the supranational spread are around -0.7 and around -0.4 for the relative value and the curve strategy, respectively. Hence, the portfolios perform poorly when the supranational spread widens. This suggests that relative value trades between Treasuries or between Treasuries and supranational bonds have a large exposure to a common risk factor.

Panel (B) reports the correlations with money market variables: changes in repo rates, the T-Bill rate and the TED spread. The portfolios exhibit very similar correlations.

[19] One reason why duration and convexity remain small is that most trades are liquidated within 30 days. Another reason is that we use bonds that are very close substitutes in the replicating portfolios.

Finally, Panel (C) reports the correlations with stock market variables: returns from the market, value, momentum, time-series momentum Asness et al. 2013, betting-against-beta Frazzini and Pedersen 2014 and changes in the VIX volatility index. If anything, the relative value portfolio has a higher correlation with momentum and betting-against-beta, while the curve-based portfolio has a higher correlation with market returns and the VIX. But the correlations are low. The key take-away from Figure 5 is that differences between risk exposures are unlikely to explain the ranking in Table 3.

4.6 Cross-Section

To understand the difference between the portfolios, we look at the distribution of individual trades. Overall, the results show that the curve-based measure generates more trades with small negative returns and low volatility. In contrast, relative value trades have positive average returns.

Panel (B) of Table 3 provides summary statistics for the cross-section of trade returns. Both signals generate a similar number of trades, but relative value generates trades with higher total returns than a curve-based measure on average. In fact, the mean and median are negative for the curve-based measure (-0.15 and -0.11 percent weekly, respectively). The cross-section also has a wide dispersion: the standard deviation is 3.84 percent (weekly) for relative value and 9.23 percent when using an estimated curve.

Panel (A) of Figure 7 shows the distribution of dollar gains or losses from each trade in a histogram. For both signals, most trades have total profit between -50 and $+50$ cents; however, a curve-based signal generates more trades with small negative losses. Panel (B) reports the histograms for the volatility of daily trade returns. The curve-based signal produces more trades with very low volatility. Overall, the evidence is consistent with measurement errors in the curve-based signal.

In other dimensions, the composition of trades is similar. Panel (C) shows that the duration of trades is similar. Panel (D) shows that the number of active trades is also similar over time. If anything, the curve-based signal generates a larger number of trades in a handful of episodes. We also checked that ignoring bid-ask spreads only shifts the distribution of returns (unreported).

4.7 Portfolio Returns Using Higher Thresholds

Measurement errors due to model fitting or model specification can explain why a curve-based signal produces a larger share of trades with small negative losses and lower volatility. Measurement errors produce false-positive signals, leading to trades that are eventually liquidated at a small loss when the errors are reversed. The trading simulation provides one way to check whether these trades drive the difference between portfolio returns. To do so, we compare the returns of trades initiated using a higher threshold than in our baseline $\tau = 0$ bps. Indeed, we find that for low values of the threshold $\tau > 0$, a curve-based signal becomes profitable and outperforms a risk-free investment. This is a natural check of the idea that small measurement errors blur a curve-based proxy for limits of arbitrage.

Figure 6 presents the results for successively higher thresholds: $\tau = 1, \ldots, 40$ bps. Panel (A) reports the terminal value of the portfolios at the end of 2017. For $\tau = 0$, the terminal values correspond to the results in Figure 5A (recall the initial value is normalized to 100). As τ increases, the curve-based portfolio becomes quickly profitable ($\tau = 1$ bps) and exceeds a risk-free investment ($\tau = 4$ bps). Therefore, filtering some of the measurement errors eliminates unprofitable trades from the portfolio. Panel (B) shows the decline in the number of trades as the threshold increases. For low values of τ, the number of curve-based trades drops much more rapidly than for relative value. This indicates that the curve-based method generates many smaller signals. Together, the results show that using a threshold eliminates a larger number of small and unprofitable curve-based trades.

Panel (A) also shows that, for both signals, raising the threshold further improves performance because it concentrates the portfolio in fewer but more profitable trades. This is true until τ reaches between 10 and 15 bps, where performance reaches a plateau. Of course, excluding trades with small signals is a more realistic trading strategy. For our purpose, we note that relative value retains its edge over a curve-based signal for every level of the threshold τ.

5 International Relative Value Indices

5.1 Commonality

The relative value index can be easily implemented outside of the US Treasury market. In the following, we compute and study the noise index in the US, the UK, Canada, Germany, France, Italy, Switzerland and Japan between January 2005 and December 2017. To motivate the construction of relative value indices, we check that bond-level relative value is highly correlated across bonds in each country. Table 4 reports a principal components analysis of the panel of bonds in each country.[20] Indeed, relative values are highly correlated: we find that the share of variations explained by the first two principal components ranges between 40 and 80 percent depending on the country.

Table 5 reports the correlations between the relative value indices of different countries. The indices for Germany, Switzerland, the UK and Canada are all corre-

[20]We apply principal component analysis to balanced panels of relative values in each country, where the number of bonds in the panel for each country is calibrated to the period with the smallest number of outstanding issues. At each date in our sample, we populate the panel starting with the bonds with remaining time to maturity closest to 10 years and adding bonds with shorter maturity until the fixed size of the panel is reached. The results are robust to other specifications.

lated with the US dispersion index. The correlation is 0.86 with the UK index and 0.48 with Canada's index. Panel (A) of Figure 8 shows the relative value indices for Canada, Germany, Switzerland, the US and the UK. As expected, the relative value indices exhibit very similar patterns: a calm period with low relative value before 2007, followed by a gradual increase in 2007 with a dramatic peak in 2009 and smaller variations since then.

This pattern suggests that the limits of arbitrage are largely common for the sovereign bond markets of these large advanced economies. However, the indices for Italy and France are correlated with each other but not with the indices of other countries. Panel (B) of Figure 8 compares the relative value indices of France and Italy. The index for Italy exhibits a striking peak at the height of the sovereign debt crisis of the euro area, and the index for France also increases during that period, suggesting that the euro area debt crisis had a large impact on limits of arbitrage in these two markets. Finally, the relative value index for Japan appears to be largely idiosyncratic.

5.2 Relative Value, Funding Costs and Volatility

Relative value indices are strongly correlated with measures of volatility and funding conditions. This is consistent with existing results for similar proxies (e.g., Fontaine and Garcia 2012, Hu et al. 2013). Figure 9 shows that the relative value indices in the US, the UK, Germany and Canada are strongly related to measures of equity market volatility in each country (Panels A–D). We use the VIX in the US, the Euro Stoxx 50 volatility index, the FTSE 100 volatility index and the Nikkei volatility index. These volatility indices are forward-looking. Table 6 shows that the correlation between each index and the local measure volatility is high. The correlation between each index and the US VIX is also very high. Both observations are consistent with tighter limits of arbitrage when financial markets are more volatile. The correlations are lower for France and Italy, presumably because the euro area measure of volatility does not accurately reflect limits of arbitrage in these markets. The correlation is very low in Japan, which stands out as an exception.

Table 6 also reports the correlations of relative value indices with proxies for funding conditions in each country. We use the spread between interbank borrowing rates and overnight indexed swap (OIS) rates, computed locally and for the US. As expected, the relative value indices are positively correlated with funding conditions in essentially all countries. Japan stands out again as the exception.

5.3 Economic Content

We also evaluate the economic value of the relative value index across the US, the UK, Canada, Germany, France, Italy, Switzerland and Japan. Table 7 reports summary statistics for the time series of index returns in each country, reported in monthly percentage terms. Again, relative value has more economic content than a curve-based signal in every country. The difference in average monthly returns ranges from 0.1 percent to 0.4 percent. Overall, the results confirm that using relative value identifies economically meaningful limits of arbitrage and provides improvements as a proxy for limits of arbitrage in fixed-income markets.

It is also true in every country that the curve-based measure generates a higher share of trades with small negative returns and low volatility. Table 8 reports summary statistics for the cross-section of trades across countries in weekly percentage. The same message emerges. Relative value trades have higher average returns than curve-based trades in every country and generally have lower variance.

6 Conclusion

We introduce the relative value measure of limits of arbitrage in fixed-income markets. Relative value is model-free, simple and intuitive. Relative value eliminates measurement errors due the estimation of a parametric yield curve model. We provide three applications to illustrate how relative value can strengthen the estimated impact of limits of arbitrage in several countries. These relative value indices are available publicly and will be regularly updated. We hope that these indices will help answer a number of research questions about limits of arbitrage. In addition, future research could apply our methodology to supranational, sub-national or corporate bond markets.

References

Adrian, T., M. Fleming, D. Stackman, and E. Vogt (2015). Has U.S. treasury market liquidity deteriorated? Liberty Street Economics, Federal Reserve Bank of New York.

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

Amihud, Y. and H. Mendelson (1991). Liquidity, maturity and the yields on U.S. Treasury securities. *Journal of Finance 46*, 1411–1425.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. *The Journal of Finance 68*(3), 929–985.

Bai, J., T. G. B. Turan, and Q. Wen (2015). Do the distributional characteristics of corporate bonds predict their future returns? Working paper, Georgetown McDonough School of Business.

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

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics 4*(1), 255–296.

Bliss, R. R. (1996). Testing term structure estimation methods. Working paper 96-12a, Federal Reserve Bank of Atlanta.

Bolder, D. and S. Gusba (2002). Exponentials, polynomials, and Fourier series: More yield curve modelling at the Bank of Canada. Staff Working Paper No. 2002-29, Bank of Canada.

Constantinides, G. M., J. C. Jackwerth, and A. Savov (2013). The puzzle of index option returns. *Review of Asset Pricing Studies 3*(2), 229–257.

D'Amico, S. and A. Pancost (2017). Special repo rates and the cross-section of bond prices. Working paper, UT Austin.

Duarte, J., F. A. Longstaff, and F. Yu (2007). Risk and return in fixed-income arbitrage: Nickels in front of a steamroller? *Review of Financial Studies 20*(3), 769–811.

Dudley, W. C. (2016). Market and funding liquidity: An overview. Remarks at the Federal Reserve Bank of Atlanta 2016 Financial Markets Conference. Federal Reserve Bank of New York.

Duffie, D. (1996). Special repo rates. *The Journal of Finance 51*(2), 493–526.

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

Fontaine, J. and R. Garcia (2015). *Handbook of Fixed Income*, Chapter Recent Advances in Old Fixed-Income Topics: Liquidity, Learning, and the Lower Bound.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Goldberg, J. and Y. Nozawa (2018). Liquidity supply and demand in the corporate bond market. Working paper, Federal Reserve Board.

Green, R. C. and B. A. Ødegaard (1997). Are there tax effects in the relative pricing of US government bonds? *The Journal of Finance 52*(2), 609–633.

Gürkaynak, R., B. Sack, and J. H. Wright (2007). The US Treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*(8), 2291–2304.

Hu, G., J. Pan, and J. Wang (2013). Noise as information for illiquidity. *The Journal of Finance 68*(6), 2341–2382.

Ilmanen, A. (2000). Convexity bias and the yield curve. *Advanced Fixed-Income Valuation Tools 61*, 25.

Koijen, R., T. Moskowiz, L. Pedersen, and E. Vrugt (2018). Carry. *Journal of Financial Economics 127*(2), 197–225.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Malkhozov, A., P. Mueller, A. Vedolin, and G. Venter (2016). International illiquidity. Working paper, London School of Economics.

Musto, D., G. Nini, and K. Schwarz (2018). Notes on bonds: Illiquidity feedback during the financial crisis. *Review of Financial Studies 31*(8), 29833018.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Working Paper 4871, National Bureau of Economic Research.

Vayanos, D. and J. L. Vila (2009). A preferred-habitat model of the term structure of interest rates. Working Paper 15487, National Bureau of Economic Research.

Vayanos, D. and P.-O. Weill (2008). A search-based theory of the on-the-run phenomenon. *The Journal of Finance 63*(3), 1361–1398.

A Appendix

A.1 Other Data

Local interbank borrowing rates are USD LIBOR for the US, GBP LIBOR for the UK, Canadian Dollar Offered Rate (CDOR) for Canada, CHF LIBOR for Switzerland, JPY LIBOR for Japan and Euribor for euro area countries. OIS rates are USD OIS for the US, Sterling Overnight Index Average (SONIA) for the UK, CAD OIS for Canada, CHF OIS for Switzerland, Tokyo Overnight Average Rate (TONAR) for Japan and Euro OverNight Index Average (EONIA) for euro area countries. Local volatility indices are VIX for the US, VFTSE for the UK, VNKY for Japan and VSTOXX for the euro area countries.

A.2 Data Filter

For each country, we include all bullet government bonds with a remaining term to maturity between 1 and 10 years. To calculate the relative value measure for the shortest or longest maturity bonds, we rely on bills or bonds with slightly shorter or longer maturities. We exclude days when the yield to maturity is available for less than four bonds. We also exclude inflation-linked bonds, callable bonds, fungible bonds, strips, perpetual bonds (UK), retail bonds (Japan) and flower bonds (US).

From an original sample of 2210 bonds in the CRSP US Treasury dataset (1988–2017), the number of bonds considered is 1203, for a total of 573,860 daily observations. We do not apply further filters to the CRSP dataset. For data sources covering other countries, we check for errors in the reported yield to maturity and compute a new yield from the quoted price. Every day for every bond, we calculate the median level of mid yields for the six bonds with the nearest remaining term to maturity, as well as the absolute median deviation from this median. If the reported mid yield of the bond is an outlier, we recalculate the bid, ask and mid *yields* to maturity using the quoted bid, ask and mid *prices*. We also exclude extreme outliers.

We check for errors in the reported bid-ask spread and compute a new bid-ask spread based on nearby observations. Every day for every bond, we check bid-ask spreads for outliers. For each outlier, we calculate the median bid, ask and mid yields for the six bonds with the nearest remaining term to maturity. We identify if either the bid or ask yield is the outlier. If both bid and ask prices and yields are outliers, we drop the observation. If only the bid (ask) yield is the outlier, we replace the bid (ask) yield and price by the ask (bid) yield plus (minus) the median bid-ask spread of the neighbouring six bonds; we also replace the bid (ask) price by the ask (bid) price minus (plus) the median bid-ask spread of the neighbouring six bonds.

A.3 Adjusting the Curve-based Measure to Incorporate Bid-Ask Spreads

The adjustment cannot be the same as in Equation 9, since there is no bid-ask for the estimated yield curve. Instead, the distance between the mid quote and the estimated curve, accounting for the bid-ask spread is given by:

$$\epsilon_b^+ \equiv \hat{y}_b - y_b^{mid} - (y_b^{ask} - y_b^{bid}) \tag{17}$$

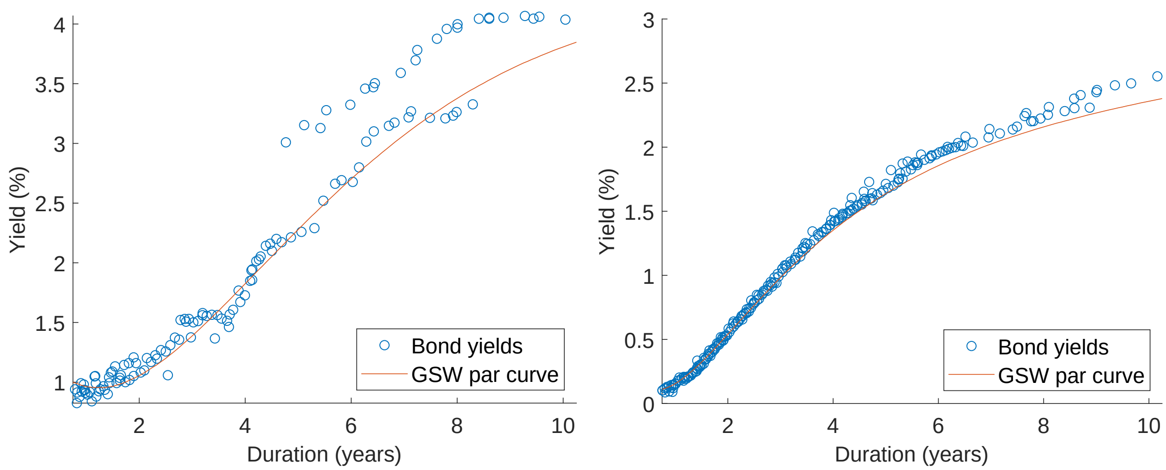
$$\epsilon_b^- \equiv y_b^{mid} - \hat{y}_b - (y_b^{ask} - y_b^{bid}) \tag{18}$$

where $(y_b^{ask} - y_b^{bid})$ is the bid-ask spread for the target bond. Similar to the relative value, the curve-based measure is then defined as:

$$crv_b = \equiv \begin{cases} \epsilon_b^+ & if \quad \epsilon_b^+ > 0 \\ 0 & \text{if} \quad \epsilon_b^+ < 0 \text{ and } \epsilon_b^- < 0 \ . \\ -\epsilon_b^- & if \quad \epsilon_b^- > 0 \end{cases} \tag{19}$$

Figure 1: Dispersion of US Treasury Yields to Maturity

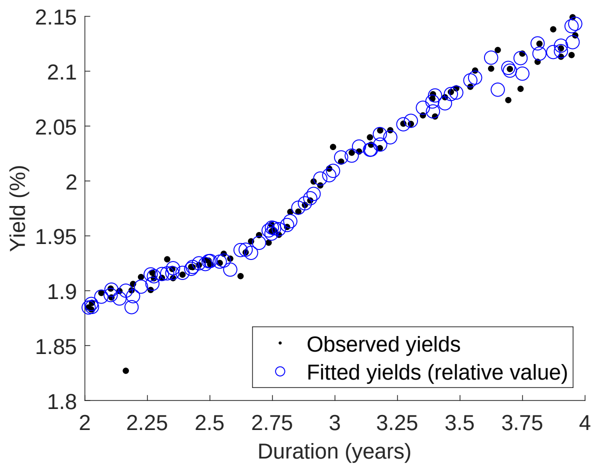
Yields to maturity for US Treasury securities from the CRSP database plotted against each security's duration. The Gürkaynak et al. (2007) parametric par curve (GSW) is plotted for comparison.



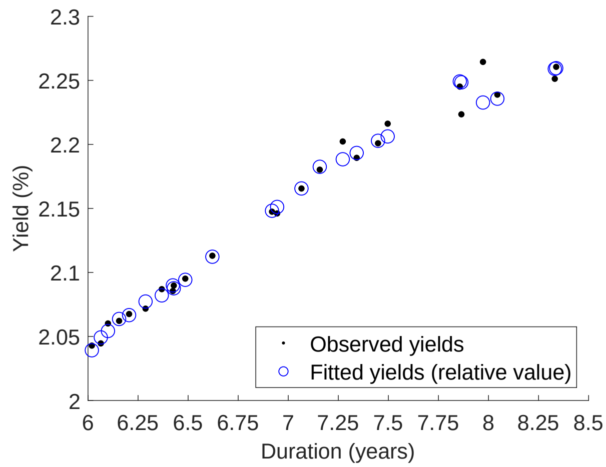
(A) 24 November 2008 **(B)** 24 November 2014

Figure 2: Comparing Relative Value and Curve Measures

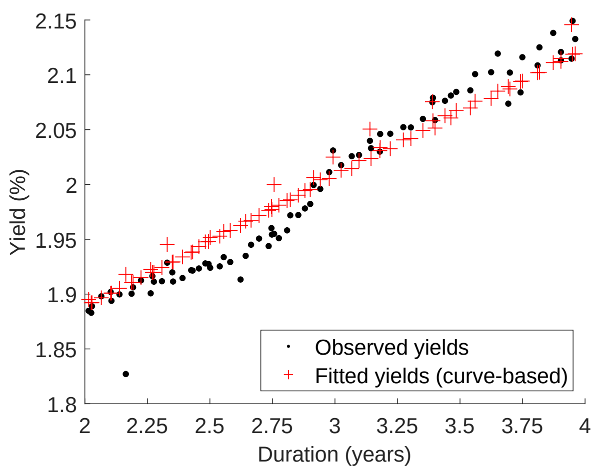
Observed yields and fitted yields using relative value and curve-based measures, zooming in on specific days and sections of the yield curve.



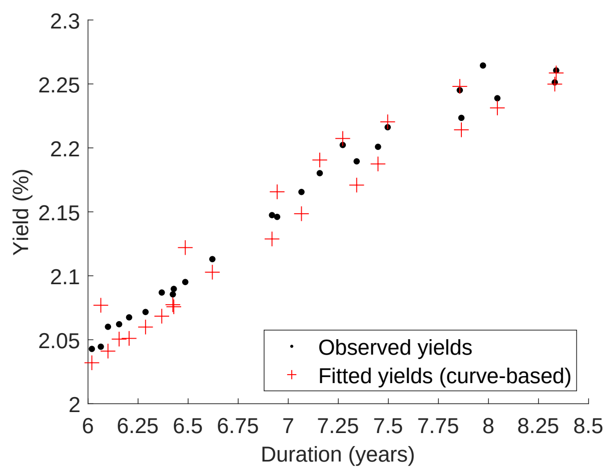
(A) 29 December 2017–Relative value



(B) 15 August 2017–Relative value



(C) 29 December 2017–Curve-based



(D) 15 August 2017–Curve-based

Figure 3: Two Measures of Limits of Arbitrage for US Treasury Bonds
Relative value and curve-based indices for US Treasury bonds with a remaining maturity between one and ten years.

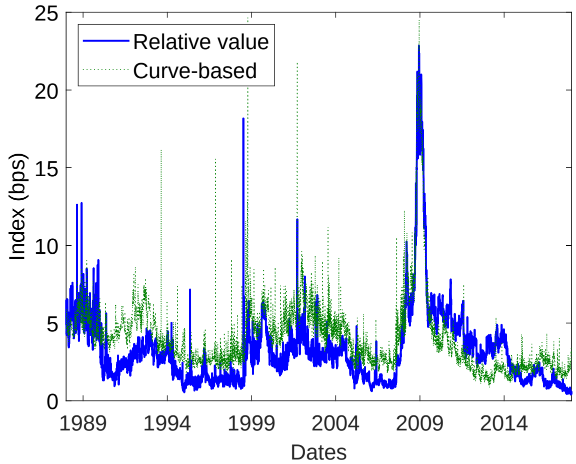
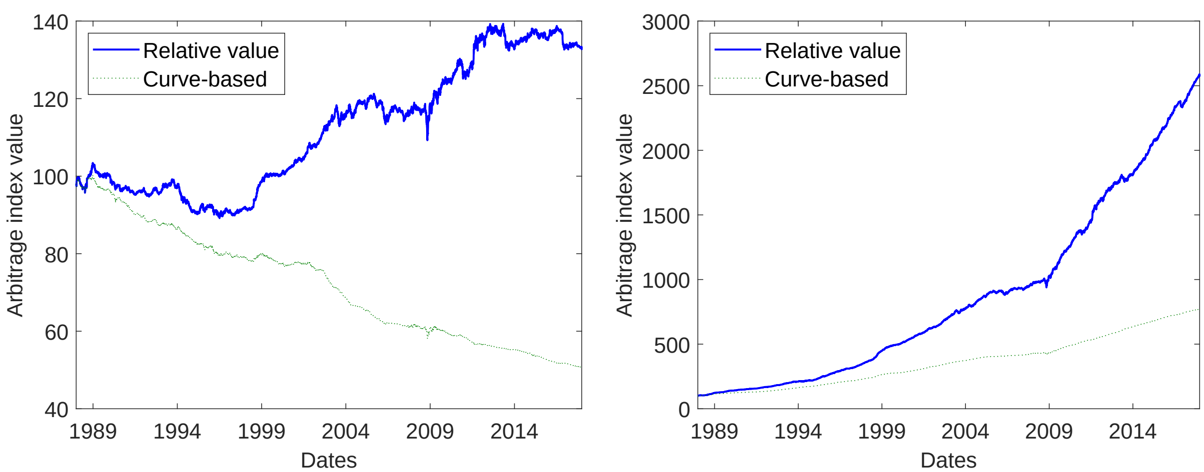


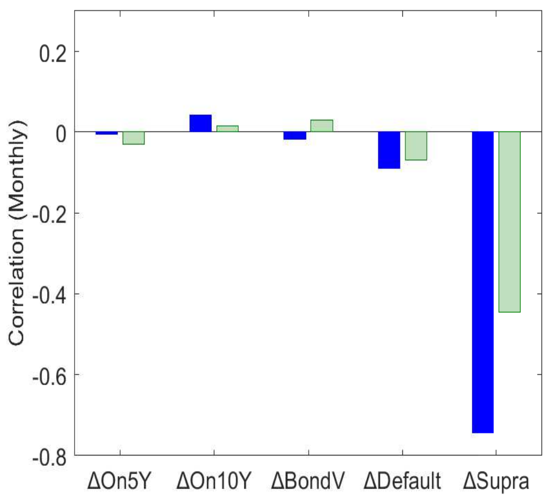
Figure 4: Cumulative Portfolio Returns in the US (1988–2017)
Cumulative returns for the indices of equally weighted portfolios of trades generated using relative value and the curve-based measure, including (A) or excluding (B) the bid-ask spread. The indices are set to 100 on January 1st, 1988.



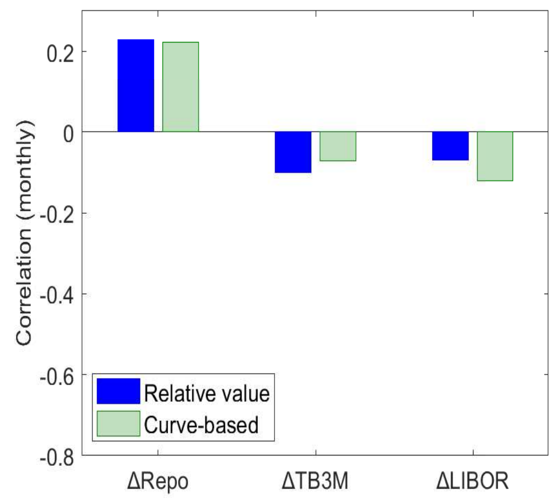
(A) US (including bid-ask spread) **(B)** US (excluding bid-ask spread)

Figure 5: Portfolio Exposures

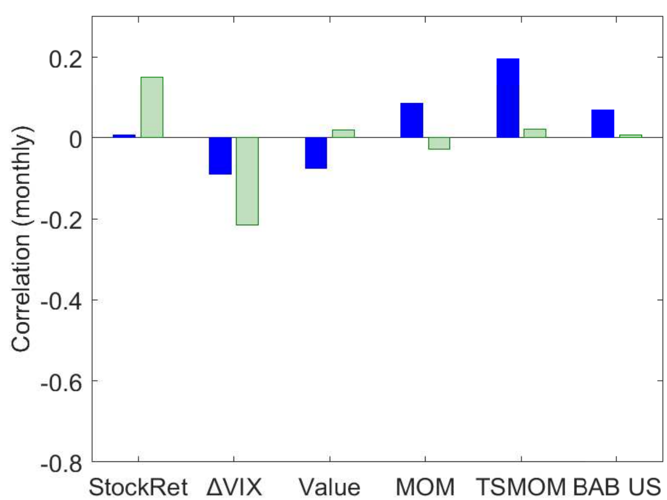
Correlations between monthly portfolio returns and proxies for risk. Panel A: changes in the 5-year on-the-run premium (ΔOn5Y), 10-year on-the-run premium (ΔOn10Y), 5-year US Treasury Note yield volatility (ΔBondV, 20-day average), Moody's Baa-Aaa spread (ΔDefault), and the 5-year Supranational-Treasury spread (ΔSupra). Panel B: changes in the USD Overnight GC Government Repo rate (ΔRepo), 3-month Treasury Bill rate (ΔTB3M) and 3-month LIBORTreasury Bill spread (ΔLIBOR). Panel C: the CRSP Value-Weighted Return including dividends (StockRet), the CBOE Volatility Index (ΔVIX), and the Value, Momentum (MOM), Time-Series Momentum (TSMOM) and US betting against beta (BAB).



(A) Bond markets



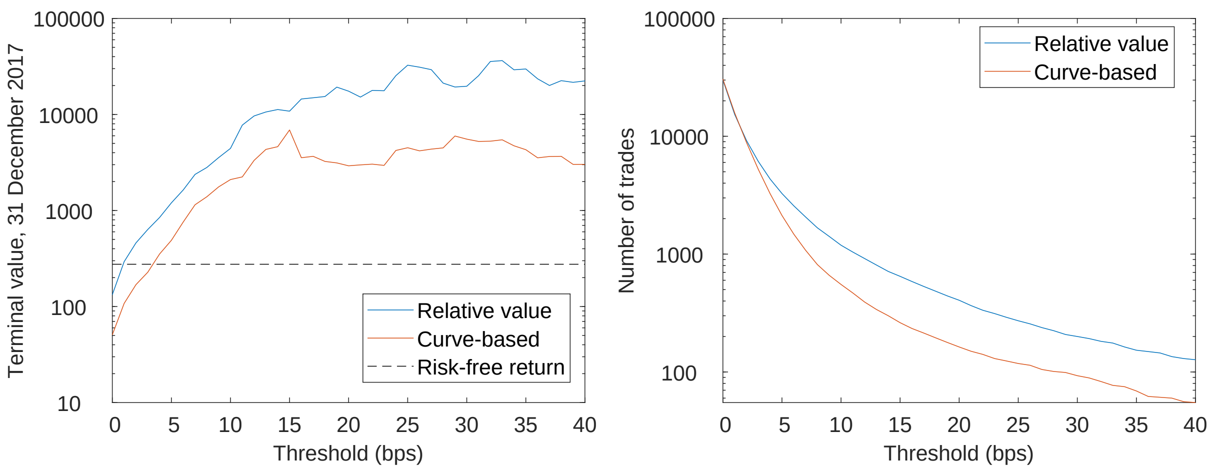
(B) Money markets



(C) Equity markets

Figure 6: Trading Signals Using Thresholds (US 1988–2017)

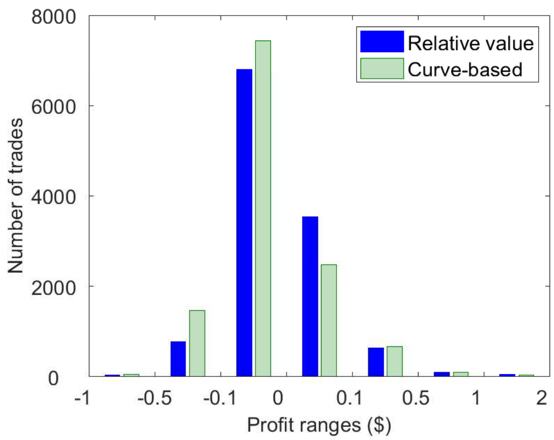
Results from trading strategies in the US Treasury market using the relative value and curve-based signals where implementation threshold ranges between $\tau = 1, \ldots, 40$ bps. Panel A: values of the indices of cumulative returns on December 31st, 2017, by threshold and measure, including the bid-ask spread. Panel B: total number of trades initiated during the sample period, by threshold and measure. The indices are set to 100 on January 1st, 1988.



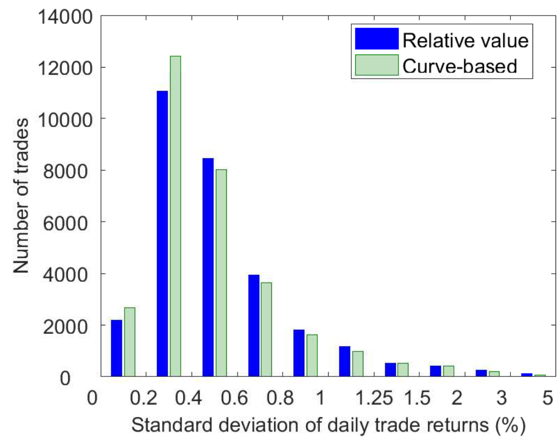
(A) End of sample index value **(B)** Total number of trades

Figure 7: Individual Trades (1988–2017)

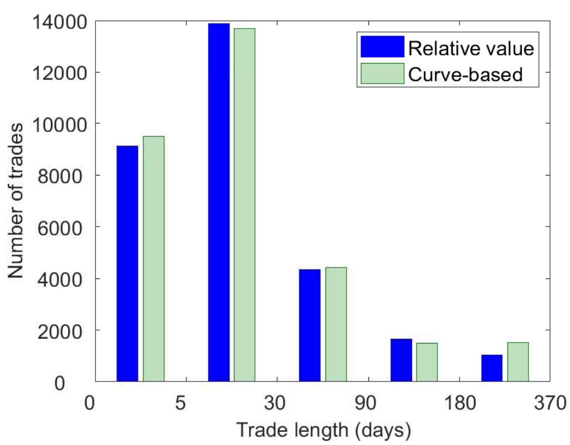
Panel A: histogram of the total profits or losses of trades. Panel B: histogram of the volatility of daily trade returns. Panel C: histogram of the number of days between the implementation and liquidation. Panel D: plot of the number of active trades generated.



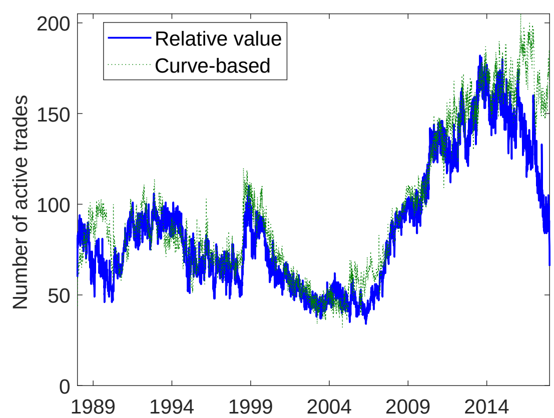
(A) Profits and losses



(B) Volatility



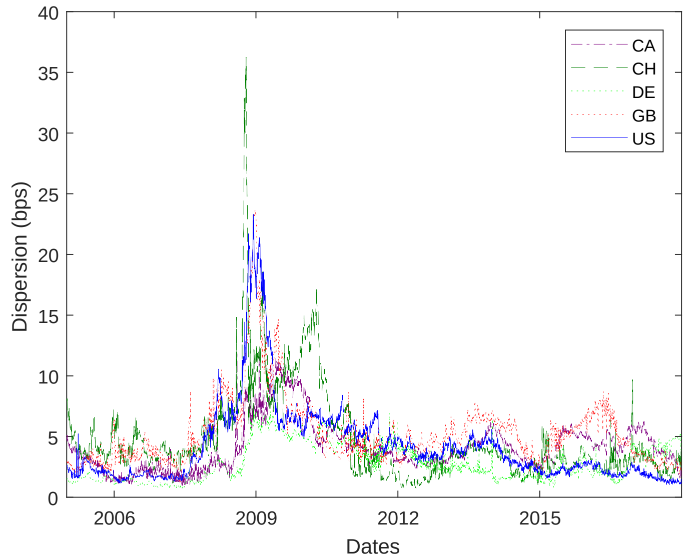
(C) Trade length



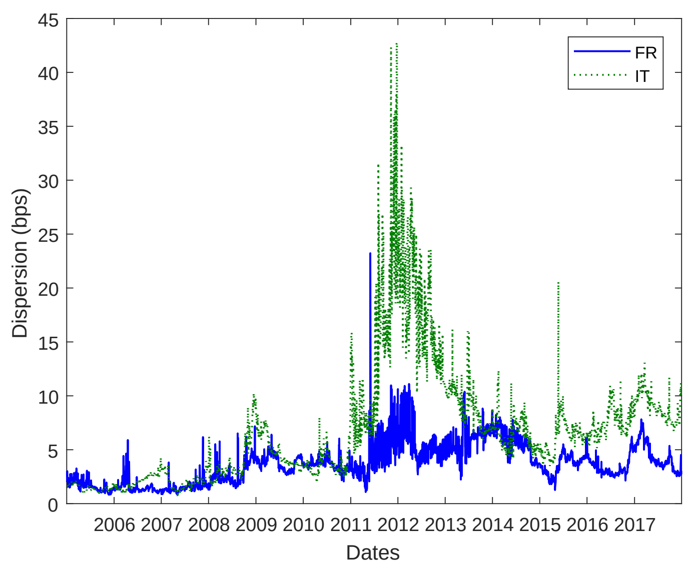
(D) Active trades

Figure 8: International Relative Value Indices

Relative value indices for sovereign issuers in advanced economies.



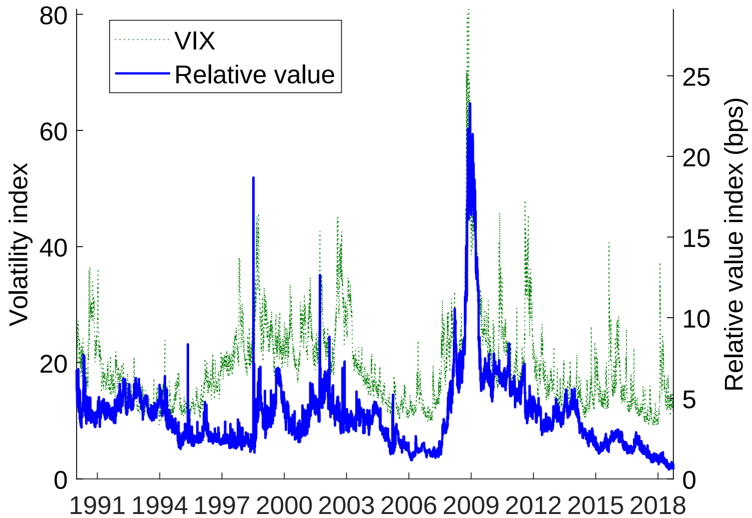
(A) US, UK, Canada, Germany and Switzerland



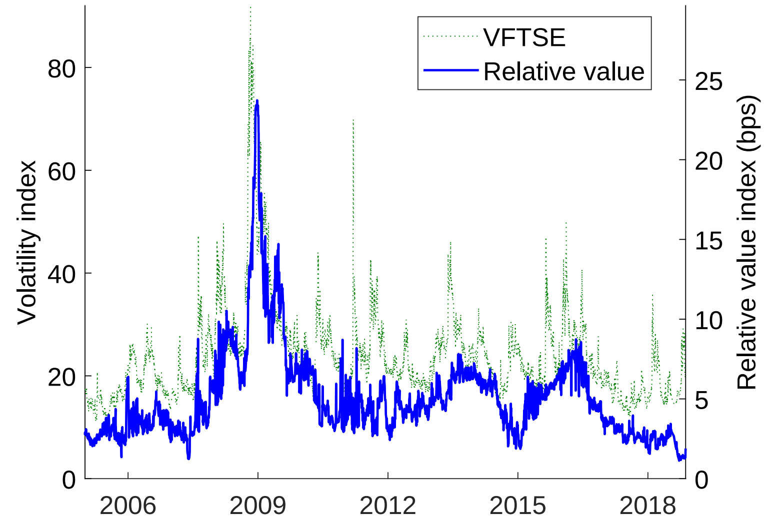
(B) France and Italy

Figure 9: Relative Value Index and Volatility

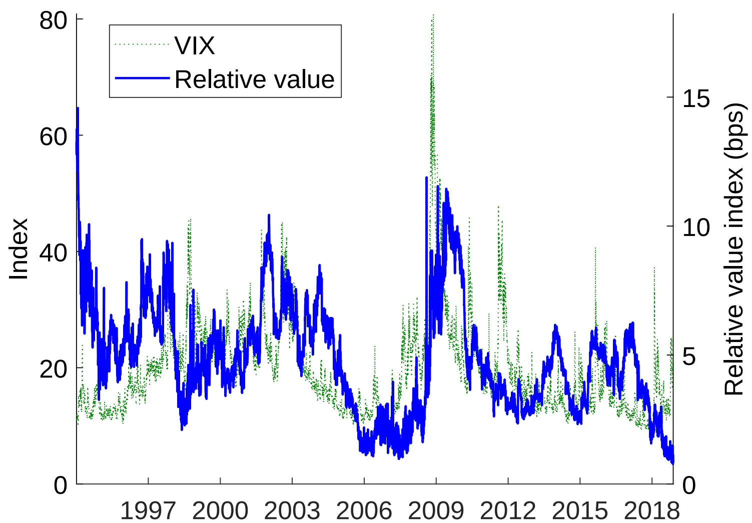
Comparison of the sovereign issuer's relative value index and the local stock market volatility index. Daily values.



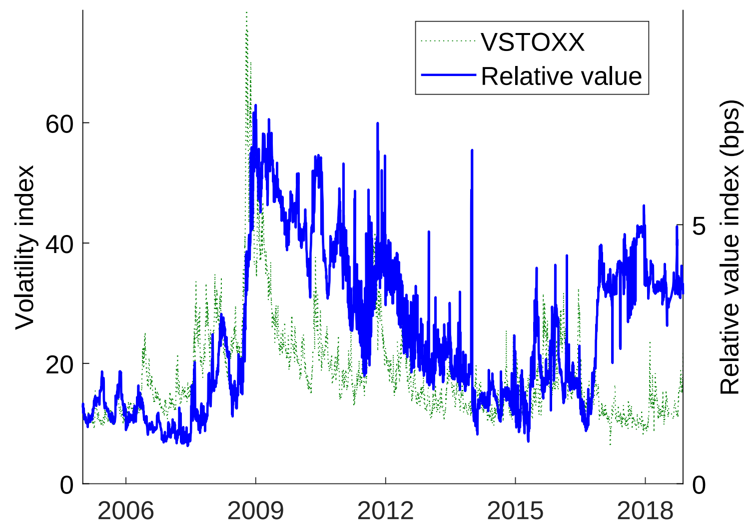
(A) US



(B) UK



(C) Canada



(D) Germany

Table 1: **Summary Statistics for Sovereign Issuers**

Summary statistics by sovereign issuer. Sample period is 1987–2017 for the US, 1994–2017 for Canada, and 2005–2017 for other countries.

	US	**GB**	**CA**	**CH**	**DE**	**FR**	**IT**	**JP**
# Observations	502371	59881	94605	35456	116231	85606	130994	473412
# Bonds, median	166	18	29	11	36	30	41	139
# Bonds, 25th percentile	103	16	26	10	28	18	26	131
# Bonds, 75th percentile	206	19	30	11	39	31	51	146
Average duration	3.62	4.28	3.56	4.95	4.17	4.51	3.93	4.39
Average std dev of duration	1.91	2.04	1.96	2.23	2.21	2.18	2.09	2.43
Average coupon	3.32	4.80	4.69	2.93	3.12	3.92	3.80	1.32
Average std dev of coupon	1.99	1.94	2.92	0.87	1.20	1.79	1.34	0.86

Table 2: **Relative Values as Proxy for Risks: Asset Pricing Tests**

Cross-sectional asset pricing tests based on two-stage Fama-MacBeth regressions. The prices of risk are annualized (multiplied by 12). Standard errors and Shanken-corrected standard errors are reported in parentheses. The market returns is included as test assets at estimation. Monthly data, January 1986–December 2015.

Panel (A) S&P500 options

	Call & put				Put				Call			
MKT	0.215	0.068	0.088	0.079	0.109	0.059	0.076	0.061	0.085	0.068	0.065	0.066
t-FM	(4.28)	(2.10)	(2.72)	(2.45)	(3.30)	(1.87)	(2.39)	(1.91)	(2.59)	(2.14)	(2.05)	(2.07)
t-Sh	(4.11)	(2.04)	(2.66)	(2.31)	(3.30)	(1.86)	(2.29)	(1.91)	(2.56)	(2.07)	(2.05)	(2.07)
RV		-0.169		-0.277		-0.266		-0.362		-0.310		-0.207
t-FM		(4.26)		(5.79)		(5.44)		(6.74)		(3.27)		(4.71)
t-Sh		(2.74)		(2.41)		(2.53)		(1.94)		(1.34)		(2.55)
Noise			-0.171	0.139			-0.543	0.304			-0.185	-0.075
t-FM			(3.74)	(3.45)			(4.71)	(6.00)			(2.62)	(0.98)
t-Sh			(2.43)	(1.44)			(1.21)	(1.73)			(1.62)	(0.52)
R^2	0.72	0.91	0.82	0.94	0.24	0.85	0.55	0.88	0.56	0.85	0.82	0.87

Panel (B) Corporate bonds

MKT	0.078	0.077	0.074	0.077
t-FM	(2.68)	(2.63)	(2.56)	(2.65)
t-Sh	(2.68)	(2.60)	(2.54)	(2.64)
RV		-0.162		-0.162
t-FM		(3.02)		(2.94)
t-Sh		(1.94)		(1.89)
Noise			-0.063	0.001
t-FM			(1.38)	(0.02)
t-Sh			(1.25)	(0.01)
R^2	0.00	0.64	0.11	0.64

Table 3: **Summary Statistics for US Treasury Trade Returns (1988–2017)**
Panel (A): summary statistics of portfolio returns, in monthly percentage points. Panel (B): summary statistics for the cross-section of total trade returns, in weekly percentage points. Both panels account for the bid-ask spread and use the baseline threshold $\tau = 0$.

Panel (A) Time series of portfolio returns

	Mean	Std dev	Min	Max	Skew	Kurt
Relative value	0.09	0.87	-5.72	5.10	-0.32	11.75
Curve-based	-0.19	0.47	-3.63	2.47	-0.34	16.00

Panel (B) Cross-Section of Trade Returns

	Mean	Median	Std dev	# Trades
Relative value	0.10	-0.04	3.84	30030
Curve-based	-0.15	-0.11	9.23	30590

Table 4: **Principal Components of Relative Value Measures**
Shares of the variations in a balanced panel of bonds' relative value explained by the first two principal components, in each country. The sample period is from 2005 to 2017.

	US	**GB**	**CA**	**CH**	**DE**	**FR**	**IT**	**JP**
1st	0.39	0.40	0.24	0.72	0.24	0.26	0.32	0.13
2nd	0.06	0.15	0.16	0.11	0.14	0.16	0.12	0.06

Table 5: **International Relative Value Indices: Correlations**
Correlation between the daily relative value indices in each country. The sample period is from 2005 to 2017.

	US	**GB**	**CA**	**CH**	**DE**	**FR**	**IT**	**JP**
US	1.00	0.80	0.53	0.64	0.60	0.15	0.03	-0.15
GB	0.80	1.00	0.59	0.52	0.44	0.16	-0.00	-0.10
CA	0.53	0.59	1.00	0.56	0.66	0.29	0.02	-0.31
CH	0.64	0.52	0.56	1.00	0.44	-0.11	-0.32	-0.35
DE	0.60	0.44	0.66	0.44	1.00	0.33	0.29	-0.20
FR	0.15	0.16	0.29	-0.11	0.33	1.00	0.62	0.21
IT	0.03	-0.00	0.02	-0.32	0.29	0.62	1.00	0.43
JP	-0.15	-0.10	-0.31	-0.35	-0.20	0.21	0.43	1.00

Table 6: **International Relative Value Indices, Money-Market Rates and Volatility**
Correlation between the daily value of relative value indices with local or international measures of volatility or funding stress. Appendix A.1 details the additional data. The sample period is from 2005 to 2017.

	US	**GB**	**CA**	**CH**	**DE**	**FR**	**IT**	**JP**
3M Local Interbank-OIS spread	0.66	0.66	0.33	0.52	0.34	0.18	0.20	-0.14
3M US LIBOR-OIS spread	0.70	0.62	0.15	0.58	0.28	0.04	0.05	-0.03
3M Interbank-OIS spread (intl average)	-0.65	0.07	0.34	0.08	-0.15	-0.26	-0.26	-0.24
Volatility index (local)	0.68	0.72	N/A	N/A	0.41	0.06	0.04	0.07
VIX	0.68	0.66	0.33	0.58	0.50	0.10	0.07	-0.06
Volatility index (intl average)	0.65	0.66	0.43	0.56	0.44	0.05	0.09	-0.00

Table 7: **Summary Statistics: Monthly Index Returns (2005–2017)**
Summary statistics for the time series of monthly portfolio returns, in percentage terms.

		Mean	Std dev	Min	Max	Skew	Kurt
US	Relative Value	0.08	1.00	-5.72	5.10	-0.44	13.61
	Curve	-0.17	0.48	-3.63	2.47	-1.14	26.57
GB	Relative value	0.37	1.00	-2.70	3.30	0.23	4.00
	Curve	0.25	1.09	-2.79	4.57	1.15	5.81
CA	Relative value	0.08	1.10	-3.47	3.41	-0.41	4.29
	Curve	-0.31	0.64	-2.57	1.10	-0.70	4.39
DE	Relative value	0.38	0.80	-2.31	2.59	-0.44	4.88
	Curve	0.26	0.60	-1.14	2.39	0.97	4.91
CH	Relative value	0.55	1.93	-6.58	11.37	1.69	15.16
	Curve	0.40	2.41	-6.47	11.46	1.88	10.28
FR	Relative value	0.86	1.92	-6.78	6.69	0.26	6.24
	Curve	0.73	1.49	-4.86	6.84	0.18	6.15
IT	Relative value	1.39	3.41	-2.47	31.81	5.83	47.14
	Curve	0.99	1.87	-2.37	10.93	2.91	14.41
JP	Relative value	0.49	0.98	-1.15	4.60	1.79	6.93
	Curve	0.19	0.70	-0.93	3.60	1.98	9.11

Table 8: **Summary Statistics: Cross-Section of Trade Returns for All Countries (2005–2017)** Mean, median and standard deviation in the cross-section of trade returns, from implementation to liquidation. Weekly returns, in percentage terms.

		Mean	Median	Std dev	# Trades
US	Relative value	-0.02	-0.04	1.29	17429
	Curve	-0.21	-0.09	1.45	17691
GB	Relative value	0.77	0.04	3.50	909
	Curve-based	0.76	0.05	2.86	931
CA	Relative value	-0.05	-0.03	1.12	1408
	Curve-based	-0.60	-0.22	1.41	2504
DE	Relative value	1.00	0.04	4.13	2310
	Curve-based	0.70	-0.01	5.19	2365
CH	Relative value	1.34	0.06	5.24	373
	Curve-based	0.66	0.03	3.97	518
FR	Relative value	1.80	0.07	9.34	1191
	Curve-based	1.74	0.09	8.85	1670
IT	Relative value	3.20	0.08	12.47	3911
	Curve-based	1.84	0.01	10.89	3275
JP	Relative value	0.13	0.01	1.78	7191
	Curve-based	0.09	-0.00	1.88	9402