Shape Evolution of the Interest Rate Term Structure

BIWEI CHEN[*]

Houghton College

This paper adopts a novel approach to studying the evolution of interest rate term structure over the U.S. business cycles and to predicting recessions. Applying an effective algorithm, I classify the Treasury yield curve into distinct shapes and find the less frequent shapes intrinsically linked to the recessions in the post-WWII data. In forecasting recessions, the median-short yield spread trumps the long-short spread for horizons up to 17 months ahead and the yield curve shape is nearly impressive as the median-short spread. Overall, the yield curve shape is an informative but more succinct indicator than the spreads in studying the term structure.

Keywords: Business cycle, recession forecast, U.S. Treasury yield curve, yield spreads.

JEL Classifications: E32, E43, E47, G12, C52, C53, C82

1 Introduction

Pioneering studies on the interest rate movements date back to Macaulay (1938) at the National Bureau of Economic Research (NBER). About three decades later, the classical monograph of Kessel (1965) became the first work to characterize term structure patterns over the U.S. business cycle after the WWII and he proposed the liquidity preference and expectations hypothesis as the theory underneath. Since then, term structure has been the research hotspot in finance and macroeconomics. While most of the theoretical and empirical inquiries were directed to the expectations hypothesis, applying the yield curve to macroeconomic forecasting had not shown its power until the late 1980s (Laurent, 1988; Harvey, 1988; Estrella and Hardouvelis, 1991). Since then, the predictive strength of the yield spread has become a "stylized fact" among macroeconomists and market participants (Wheelock and Wohar, 2009). The literature is still growing and beyond enumeration.

Along the lines, this paper studies the U.S. term structure fluctuations and its predictive power of recessions, contributing to the literature in a number of initiatives. Methodologically, it designs an effective classification algorithm for the Treasury yield curve and makes available a new indicator of the term structure. Second, it traces the evolution of the yield curve and characterizes its business cycle patterns. Third, it applies the yield curve shape in forecasting recessions and evaluates its performance relative to the term spreads.

Aside from measuring cross-sectional yields directly, this paper condenses the entire yield curve information into a categorical shape variable and traces its evolution over the business cycles. Since the upward-sloping yield curve is the most frequent shape, other yield curves can be defined as the minor shapes of less frequencies. The principal findings are: 1) The minor shapes tend to cluster before and in recessions; 2) the recession-related shape signals are time-varying in their timing and strength; 3) the duration of a recession is positively related to the density of the minor shapes; 4) the pre-recession shape signals become more monotonic and evident after the 1980s. Feeding the Treasury yield curve shape into the Probit model in recession forecasting, I find that its predictive performance is better than the models using the long-short yield spread in the near- and median-term forecasts by various measures, especially after the early 1980s. Without fully exploring the full information of the term structure, the conventional approach based on the long-short yield spread can miss many positive signals and downplay the predictive power the entire yield curve.

The remaining paper proceeds as follows: Section two describes the data and the classification algorithm. Section three examines the evolutionary patterns of the Treasury yields and shapes. Section four explores the links between the shape and recessions. Section five applies the yield curve shape in forecasting recessions and evaluates its performance. Section six concludes. Non-essential details and evidence are collected in the Appendix.

2   Data and Methodology

This section introduces some basic concepts in understanding the yield curve, describes the dataset on the U.S. Treasury yields, proposes an effective algorithm for classifying its shapes, and reports the key statistics for the classification results.

2.1   Treasury yields

The most fundamental concept in yield curve study is the yield to maturity, henceforth YTM. It measures annualized rate of return if the investor holds the bond to maturity. This interest rate is implied from the market price of the corresponding bond. For simplicity, consider a zero-coupon bond with face value FV and maturity T, its theortical price will satify the discounted present value formula $P = FV(1 + Y)^{-T}$, where Y is the zero-coupon bond YTM. Relating bond

yield (YTM) to its maturity T constitutes the term structure analysis (or yield curve analysis).[1] However, the U.S. Department of the Treasury does not create or publish any zero-coupon rates. Instead, Treasury yield curve is based on securities that pay coupon interest on a semiannual basis, the yields are considered "bond-equivalent" yields and the yield curve is considered a par yield curve, which corresponds to securities traded close to or at its face value.[2] The U.S. Treasury market yields are calculated from composite of quotations obtained by the Federal Reserve Bank of New York and current Treasury daily yield curve is fitted by a quasi-cubic hermite spline function.[3] The methodology and techniques to construct more consistent market yield curves have been improving significantly over the past decades. A more technical approach to fitting the U.S. Treasury yield curve can be found in Gürkaynak, et. al. (2007).

As described in Table 1, the Federal Reserve Board data download program provides the most comprehensive yield data on U.S. Treasury securities in H.15 statistics since 1953. At the short end, Treasury bills are money market assets with maturities of 1 year or less, sold at a discount from par and do not bear periodic interest payments. Treasury notes are median term coupon securities with maturities from 2 to 10 years. Treasury bonds have maturities more than 10 years. Treasury notes and bonds are capital market assets carrying periodic coupon payments. Yields derived from all these securities are therefore regarded as short term rates, median term rates and long term rates, respectively. In the H.15 statistics, only the 1-year, 3-year, 5-year, and 10-year yields are fully available over the entire sample from April, 1953 to present.

Another source little explored in the literature is the NBER Macrohistory database, which records monthly Treasury yields further back to 1920. Rather than reporting the yield per instrument, this dataset contains yield series averaging over close maturities. And the observations in this dataset are more unevenly distributed over various sample periods. It should also be noted that other inconsistency exist in its data collection and revision process. For these reasons, this paper only utilizes the NBER dataset from January, 1945 to March, 1953 to graph the yield movements.

---

[1]Bond yields, while inversely related to the price, are influenced by many other risk factors such as credit default, liquidity, call option, tax treatment, and coupon payment schemes. In practice, it is never possible to hold all other factors constant and isolate the relationship between yields and their maturities. Therefore, a pure term structure of interest rates never exists and is unobservable. For instance, bonds within the same credit rating trade with various degrees of liquidity. More often, due to the differences in coupon payment, securities with the same maturity can carry different yields (the higher the coupon rate, the higher the price, all else equal), let alone different maturities. Thus, while economists can model the pure term structure through zero coupon bonds in theory, practitioners need to "bootstrap" the discount factors successively from the coupon bearing bonds of different maturities to ensure less pricing errors.

[2]https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics

[3]https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics/treasury-yield-curve-methodology

Table 1: Federal Reseve Board H.15—Treasury yields

Maturity	Instruments	Availability	Obs
1-month	Treasury bills	200107 : present	234
3-month	Treasury bills	198201 : present	468
6-month	Treasury bills	198201 : present	468
1-year	Treasury bills	195304 : present	813
2-year	Treasury notes	197606 : present	535
3-year	Treasury notes	195304 : present	813
5-year	Treasury notes	195304 : present	813
7-year	Treasury notes	196907 : present	618
10-year	Treasury notes	195304 : present	813
20-year	Treasury bonds	195304 : 198612	732
		199310 : present	
30-year	Treasury bonds	197702 : 200202	480
		200602 : present	

Note: Observations as of December, 2020.
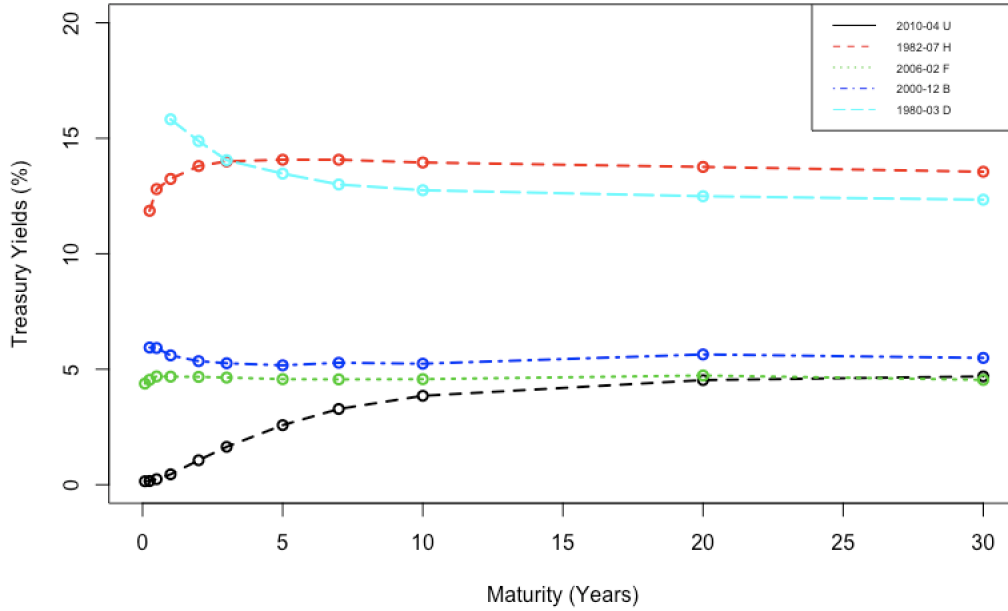
2.2 Shape classification

Despite its popularity, yield curve classification remains at a heuristic level and few has applied any rigorous metric, not to mention effective computer algorithms. News and media aside, academic research and economic textbooks provide some examples. Malkiel (1966) constructs the ascending, flat, descending, and humped yield curves. Campell (1995) describes yield curve shape as upward-sloping, downward-sloping (inverted), hump shaped, or even trough shaped (inverted hump shaped). Mishkin's textbook (2018) classifies yield curves as upward-sloping, fat, and downward-sloping. In Fabozzi et al. (2019), historical shapes observed for the Treasury yield curve are defined as positively sloped (upward-sloping), inverted (downward-sloping), and flat.[4]

Figure 1 displays five typical shapes of the U.S. Treasury yield curve on different dates. Analogous to literature mentioned above, I denote them upward (U), downward (D), flat (F), hump (H), and bowl (B). Here, the B-shape is equivalent to the inverted hump-shape. Noticeably, yield curves differ by their level, slope, and curvature.[5] Therefore, our yield curve

---

[4]The authors further distinguish between a normal and a steep yield curve by applying a threshold of 300 basis points to the yield spread between the 30-year bond and the six-month bill.

[5]In statistical modeling, the level, slope and curvature refer to the first three principle components, or the common factors, of bond yields (Litterman and Scheinkman, 1991). Jointly, they can explain over 95%

Figure 1: U.S. Treasury yield curve: observed shapes

classification can first condense all cross-sectional yields into these three measures from which then derived the shape. Given the Treasury yield data, I provide a summary of the corresponding measures and calculations in Table 2. The 10-year yield is calculated into the average long yield here whereas it is in practice considered as a Treasury note rate. There are two reasons for this adjustment. First, the 10-year Treasury yield has been chosen in the literature most frequently as the long end when computing the slope of the yield curve. Second, the data on 10-year yield have been available throughout the entire sample and it can balances the calculations of the yield levels by contributing to the average long yield.

A classification algorithm is a formal metric of observations. Based on the measures, an effective classification algorithm follows three principles: First, the shape being classified must be visually recognizable and exactly identified. Second, the classified shape provides a sufficient summary of the full information on cross-sectional yields. Third, the time series of classified shapes must be mutually exclusive and exhaustive for any given sample.

Summarized in Table 3, the algorithm reduces data dimension from eleven U.S. Treasury yields to three average yields: Y_s, Y_m, Y_l, which are the simple average of T-bill, note, and bond yields. Accounting for missing observations, they only average over available data points in respective maturity sector. Based on the spread and curvature relations, the algorithm defines five shapes. To separate the B-shape and D-shape, it sets the flat yield curve as a benchmark with

of the variation in yields. Notice, however, they are unobservable factors, not the shape per se.

Table 2: U.S. Treasury yield curve: measures and calculations

Measure		Definition	Calculation
Level	Y_s	The average of short-term yields	$(Y_{1m} + Y_{3m} + Y_{6m} + Y_{1y})/4$
	Y_m	The average of median-term yields	$(Y_{2y} + Y_{3y} + Y_{5y} + Y_{7y})/4$
	Y_l	The average of long-term yields	$(Y_{10y} + Y_{20y} + Y_{30y})/3$
Spread	S_{ms}	The difference between Y_m and Y_s	$Y_m - Y_s$
	S_{lm}	The difference between Y_l and Y_m	$Y_l - Y_m$
	S_{ls}	The difference between Y_l and Y_s	$Y_l - Y_s$
Curvature	C_{urv}	The difference between $2Y_m$ and $Y_s + Y_l$	$2Y_m - (Y_s + Y_l)$

Table 3: Shape classification algorithm for the Treasury yield curve

Yield curve shapes	Term structure relations with 0.1 percent threshold	Sample
Upward (U)	$(Y_m - Y_s > 0.1$ & $Y_m \leqq Y_l)$ or $(Y_s \leqq Y_m$ & $Y_l - Y_m > 0.1)$	2010-04
Hump (H)	$(Y_m - Y_s > 0.1$ & $Y_m > Y_l)$ or $(Y_s < Y_m$ & $Y_m - Y_l > 0.1)$	1982-07
Flat (F)	$\|Y_m - Y_s\| \leqq 0.1$ and $\|Y_l - Y_m\| \leqq 0.1$	2006-02
Bowl (B)	$(Y_s - Y_m > 0.1$ & $Y_m < Y_l)$ or $(Y_s > Y_m$ & $Y_l - Y_m > 0.1)$	2000-12
Downward (D)	$(Y_s - Y_m > 0.1$ & $Y_m \geqq Y_l)$ or $(Y_s \geqq Y_m$ & $Y_m - Y_l > 0.1)$	1980-03

Note: Y_s, Y_m, Y_l represent average short, median, and long yields. Source: Chen (2019).

a threshold of ten basis points for the differences among averaged yields. The H-shape yield curve is high in the middle but low on both ends; the B-shape yield curve is the opposite—low in the middle but high on both ends. While the F-, U-, and D-shape yield curves allow weak inequality relations among yields, the yield relations of the H- and B-shape are strictly unequal. The algorithm is effective in producing a sequence of yield curve shapes, applicable to any data frequency (Chen, 2019). A complete visualization of all classified U.S. Treasury yield curves and sensitivity analysis are in the Appendix.[6]

The key statistics of classified monthly yield curve are shown in Table 4. In addition to their counts and frequency, reported are the mean and standard deviation of the average yields and their corresponding spreads, conditional on each type of the yield curve.

---

[6]Admittedly, an exact flat yield curve can never exist in the data. A smaller threshold classifies less F-shape but a larger one renders the classification less precise as it groups more U-, B- and H-shape into it. The choice of 10% threshold strikes a balance between separation (B and H) and over-classification (F).

Table 4: U.S. Treasury yield curve: shape classification statistics (1953.04-2020.12)

Shapes	Counts	Y_s	Y_m	Y_l	S_{ls}	S_{ms}	S_{lm}	C_{urv}
Full sample	813 (100%)	4.62	5.28	5.85	1.22	0.65	0.57	0.09
		(3.29)	(3.12)	(2.80)	(1.31)	(0.71)	(0.72)	(0.57)
Upward (U)	598 (73.6%)	3.75	4.69	5.52	1.77	0.94	0.83	0.11
		(2.78)	(2.88)	(2.65)	(1.03)	(0.55)	(0.64)	(0.59)
Hump (H)	78 (9.6%)	6.60	6.89	6.69	0.09	0.29	-0.20	0.49
		(3.06)	(3.17)	(3.13)	(0.32)	(0.28)	(0.13)	(0.30)
Down (D)	70 (8.6%)	9.29	8.72	8.30	-0.99	-0.57	-0.42	-0.15
		(3.57)	(3.22)	(3.04)	(0.65)	(0.47)	(0.29)	(0.44)
Bowl (B)	41 (5.0%)	5.14	4.85	5.09	-0.06	-0.30	0.24	-0.54
		(2.03)	(1.96)	(1.88)	(0.34)	(0.23)	(0.16)	(0.21)
Flat (F)	26 (3.2%)	5.43	5.45	5.46	0.04	0.02	0.01	0.01
		(1.75)	(1.71)	(1.73)	(0.09)	(0.06)	(0.07)	(0.08)

Notes: Y_s, Y_m, Y_l are the average yields of Treasury bill, note, and bond yields in percentage, respectively. Based on the classification algorithm, reported values are the sample means with standard deviations in the parentheses. Source: The original monthly yields data are downloaded from the Federal Reserve Board H.15 Treasury nominal yield statistics.

For the full sample, the U.S. Treasury yield curve slopes up with a long-short spread of 123 basis points. It is concave but not statistically significant. The short yields are more volatile than the long yields and the yield spreads are much less volatile than the levels. In a descending order of frequency, we observe the upward-sloping yield curve (U 76.2%), the humped (H 8.64%), the downward-sloping (D 7.75%), the bowl (B 4.54%), and the flat (F 2.88%). These frequencies represent the unconditional distribution of the shape in the sample period. Since the upward-sloping is the dominant shape in the sample, whereas others together account for less than 30%, this paper groups the four less frequent shapes into one category, the minor shapes, and use this terminology hereafter. Interestingly, the Treasury yield curve not only distinguishes themselves by their shapes but also levels.[7]

3   Business Cycle Panoramas

To investigate the yield movement and shape evolution over different stages of the U.S. business cycle, their recurring patterns are examined together with statistical evidence provided for their underlying linkages.

---

[7]In principle, the yield curve can display any shape independent of its level. In reality, however, the shape is not independent of its levels, as shown in the classification results.

3.1 Co-movement and inversion

Figure 2 displays the yield movements over the past century, with shaded areas indicating the U.S. recessions.[8] Over the whole sample, all yields march ups and downs in a remarkably synchronous manner. From 1945 to present, the large swing in yields implies the mean-reverting nature of interest rates in the long cycle.[9]

Normally, the long-term yields are higher than the short-term yields. Hence, we observe the upward-sloping yield curve most often. Occasionally, the short yields can overtake the long yields, presenting a downward-sloping yield curve (e.g., in 2000). Very rare but not unlikely, all yields can cross and generate a flat yield curve (e.g., in 2006). Barely visible in the graph, the median yield can shoot above or below both ends of the yield curve, forming the hump-shape (e.g., in 2000) and bowl-shape yield curves (e.g., in 2019).

Over the business cycles, yield levels tend to rise in expansions and fall in recessions. Although their turning points are not in line with the NBER reference dates, the local interest rate maxima are more likely to precede the business cycle peaks while the local interest rate minima tend to follow business cycle roughs. Furthermore, the short yield tend to rise slightly above the long yield around the business cycle peaks, rendering the downward-sloping yield curve. Around and months after the business cycle troughs, the short yield tend to fall much below the long yield, featuring the steeply upward-sloping yield curve.

To better capture and visualize the less common patterns, the time series of three yield spreads and the curvature are shown in the Figure 3. Yield inversion can be defined as the process in which the slope of the yield curve switches from positive to negative. A narrowing of the yield spreads vividly characterizes the transition. In terms of shape change, it begins with an upward-sloping yield curve, then flattens out, and finally inverts to downward-sloping. And vice versa. Prior to the business cycle peaks, yield spreads tend to hit their local bottoms and then bounce back. Around and shortly after the business cycle troughs, yield spreads tend to reach their local maxima. The turning points of the spreads are not symmetric in size: on average, the slope of the upward yield curve is much bigger than the slope of the downward yield curve in absolute value.

---

[8]According to the National Bureau of Economic Research (NBER)'s business cycle dating committee, the U.S. economy is either in expansion or contraction, measured by the level of real economic activities. Between any two successive recessions, the economy is in expansion. Contractions (recessions) start at the peak of a business cycle and end at the trough. Refer to the NBER's Business Cycle Dating Procedure for more details. https://www.nber.org/cycles/recessions.html

[9]Historically, the U.S. interest rate had hit zero twice in the aftermath of the two most severe economic recessions, with the Treasury 3-month bill rate 0.01% in October of 1932 and the Treasury 1-month bill rate 0.00% in December of 2008. History repeats in 2020, in the midst of the COVID-19 pandemic, the short-term Treasury 1-, 2,-, 3-month rates dropped to zero on March 25, 2020. Along their paths, the long yields also fell to historical lows in each of this scenarios. At the other extreme, the peak of interest rate was recorded in the second half of 1981 when the economy was wrestling with runaway inflation (e.g., the Federal funds rate 22.36% on July 22 and the Treasury 1-year bill rate 17.31% on September 3).

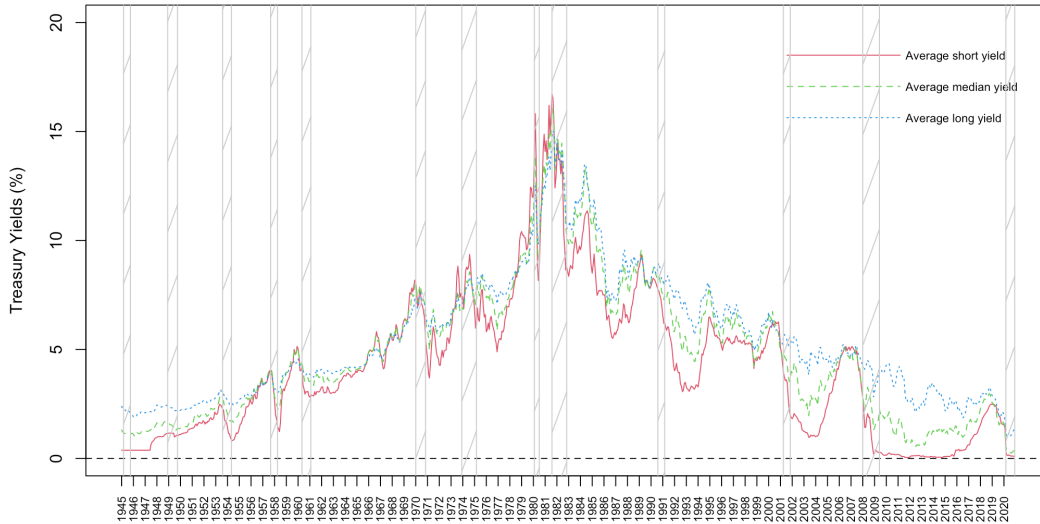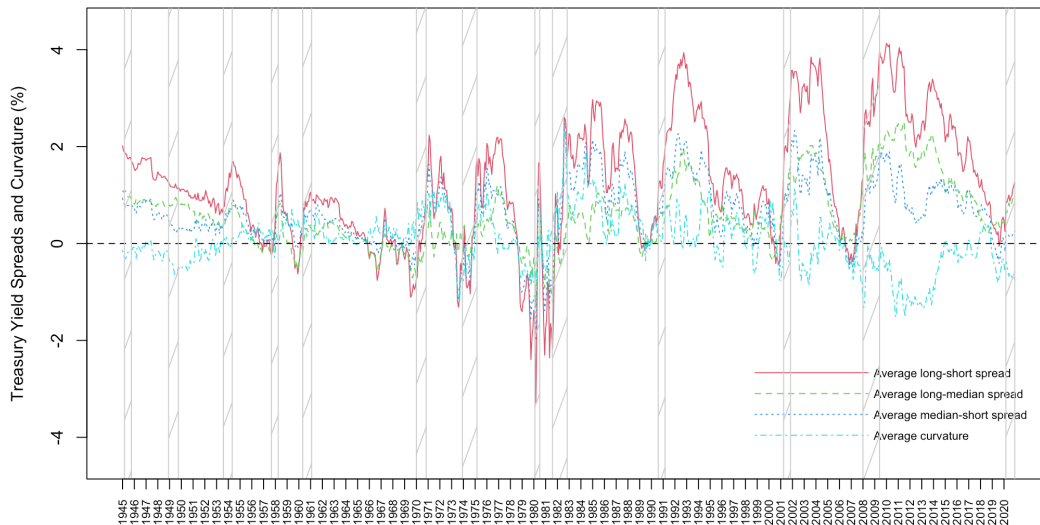Figure 2: U.S. Treasury yields: co-movements (1945.01-2020.12)

Figure 3: U.S. Treasury yields: inversions (1945.01-2020.12)

Compared with the levels and spreads, the yield curve curvature seems to drift like a white noise with little detectable link to the business cycle. Indeed, the sample correlations between the curvature and the levels are the weakest (below 0.2) among all. Although the correlations between the yields (and spreads) over the entire sample are strong and positive, their signs can turn negative or become insignificant in relatively short time frames. While the short yields are the dominant forces driving most of the negative yield curve inversions, the long yields tend to

rise during the inversions as well (see the Appendix for details).

3.2 Yield and shape evolution

To explore their linkages, I divide the U.S. business cycle into four stages and provide summary statistics for the corresponding yields and shapes. Since the shape data can only be created when three average yields are commonly available, our analysis starts from January, 1945 to March, 2020. The four business cycle stages are grouped into the 12-month periods before recessions, in-recession periods, 12-month periods after recessions, and otherwise. For two consecutive recessions not further apart, the period in between is considered as a pre-recession period in this study, which applies only to the 12 months between the 1980 recession and the 1981 recession.[10]

Table 5: Treasury yield statistics over the U.S. business cycle (1953.04–2020.12)

Stage I: all pre-recession 12-month periods (124 months)
Stage II: all in-recession periods (121 months)
Stage III: all post-recession 12-month periods (108 months)
Stage IV: all otherwise periods (460 months)

Yield measures		I	II	III	IV	All
Level	Y_s	6.59 (3.51)	5.37 (4.43)	3.89 (2.75)	4.07 (2.69)	4.62 (3.29)
	Y_m	6.41 (3.23)	5.91 (4.23)	5.07 (2.77)	4.86 (2.71)	5.28 (3.12)
	Y_l	6.34 (2.98)	6.38 (3.77)	5.86 (2.60)	5.57 (2.44)	5.85 (2.80)
Spread	S_{ls}	-0.25 (0.74)	1.00 (1.27)	1.96 (1.10)	1.50 (1.18)	1.22 (1.31)
	S_{ms}	-0.18 (0.48)	0.53 (0.70)	1.18 (0.50)	0.79 (0.61)	0.65 (0.71)
	S_{lm}	-0.07 (0.39)	0.47 (0.72)	0.78 (0.67)	0.71 (0.70)	0.57 (0.72)
Curvature	C_{urv}	-0.12 (0.46)	0.07 (0.63)	0.40 (0.46)	0.08 (0.58)	0.09 (0.57)

Note: Reported statistics are the sample mean with the standard deviation in parenthesis. Y_s average short yield, Y_m average median yield, Y_l average long yield, S_{ls} long-short spread, S_{ms} median-short spread, S_{lm} long-median spread, C_{urv} curvature. Sources: Federal Reserve Board H.15 Treasury nominal yield statistics.

Conditional on different stages of the business cycle, I summarize the first two moments of the yield levels, spreads, and curvature in Table 5. On average, yields hit their plateau during the 12-month periods before recession, remain at a relatively high level during the recessions,

---

[10]The 12-month interval is chosen arbitrarily for the convenience of framing the event window for shape variation. Similar patterns are observed when a 18-month interval is selected.

and enter the trough in the post-recession 12-month periods. Yield spreads turn negative prior to recessions, bounced back to positive in recessions, and widen in the post-recession 12-month periods. Also note that yield spreads are on average smallest before recession, corresponding to downward-sloping yield curves; they are largest during the post-recession recovery periods, equivalent to steeply upward-sloping yield curves.

Similarly, yield curvature becomes negative prior to recessions, return to positive in recessions, and shows a sizable increase when entering economic expansions. The yield curve tends to increase its concavity (hump) as the business cycle moves from pre-recession to post-recession. This is illustrated by the empirical mean curvature which increases from -0.12 in stage I to 0.4 in stage III.

Compared with the recession-related three stages, yield levels are less volatile in non-recession periods, as measured by the standard deviation. In particular, volatility of all yields increases preceding recessions, becomes most volatile in recession periods, and then dwindles after recessions. Similar volatility patterns are observed for the spreads and curvature.[11]

Table 6: Treasury yield curve shape over the U.S. business cycle (1953.04–2020.12)

Stage I: all pre-recession 12-month periods (124 months)
Stage II: all in-recession periods (121 months)
Stage III: all post-recession 12-month periods (108 months)
Stage IV: all otherwise periods (460 months)

	I	II	III	IV	All
Upward (U)	24 (19.4%)	81 (66.9%)	105 (97.2%)	388 (84.3%)	598 (73.6%)
Hump (H)	24 (19.4%)	20 (16.5%)	3 (2.78%)	31 (6.74%)	78 (9.6%)
Down (D)	41 (33.1%)	13 (10.7%)	0 (0.00%)	16 (3.48%)	70 (8.6%)
Bowl (B)	25 (20.2%)	7 (5.78%)	0 (0.00%)	9 (1.96%)	41 (5.0%)
Flat (F)	10 (8.06%)	0 (0.00%)	0 (0.00%)	16 (3.48%)	26 (3.2%)
All shapes	124 (15.3%)	121 (14.9%)	108 (13.3%)	460 (56.6%)	813 (100%)

Note: Reported statistics are the sample mean with the standard deviation in parenthesis. Data sources: Federal Reserve Board H.15 Treasury nominal yield statistics.

Table 6 provides a summary for the conditional shape counts and frequency over the stages of business cycles. In the last column of the table, the full-sample statistics serve as our com-

---

[11]In the Appendix, a visual comparison of the yield volatility over the different stages of business cycle is displayed in bar plots. Both standard deviation and median absolute deviation measures are applied, which establishes the robustness of the result. In addition, real-time volatility of the yields is estimated over the entire sample and some largest jumps in yields are ordered.

parison benchmark. The bottom row calculates the relative frequency of each stage over the business cycle. In the past 68 years, the U.S. economy was in recessions for about 10 years, or 14.9% of the time. Compared to the benchmark, the relative frequency of the four minor shapes more than doubled (F and H) and some even quadrupled (B and D) in the pre-recession stage. Accordingly, the upward-sloping yield curve shrank in frequency by more than 50%. Among the four minor shapes, the D-shape occurred relatively more often in the pre-recession stage than others. In recessions, the B-, D- and H-shape were observed slightly more active than their full sample counterparts. In the post-recession stage, the upward-sloping yield curve almost took over, leaving only three appearances of the H-shape. Lastly, despite their remarkable densities in the pre-recession stage, the four minor shapes were rarely present in the stage distant from the recessions.

By nature, yield and shape are two perspectives of the same term structure. While multiple yield values provide direct measures, the categorical shape condenses the entire yield curve into a single letter. The advantage of shape measure is obvious: it tracks its evolution over time in the most concise manner.

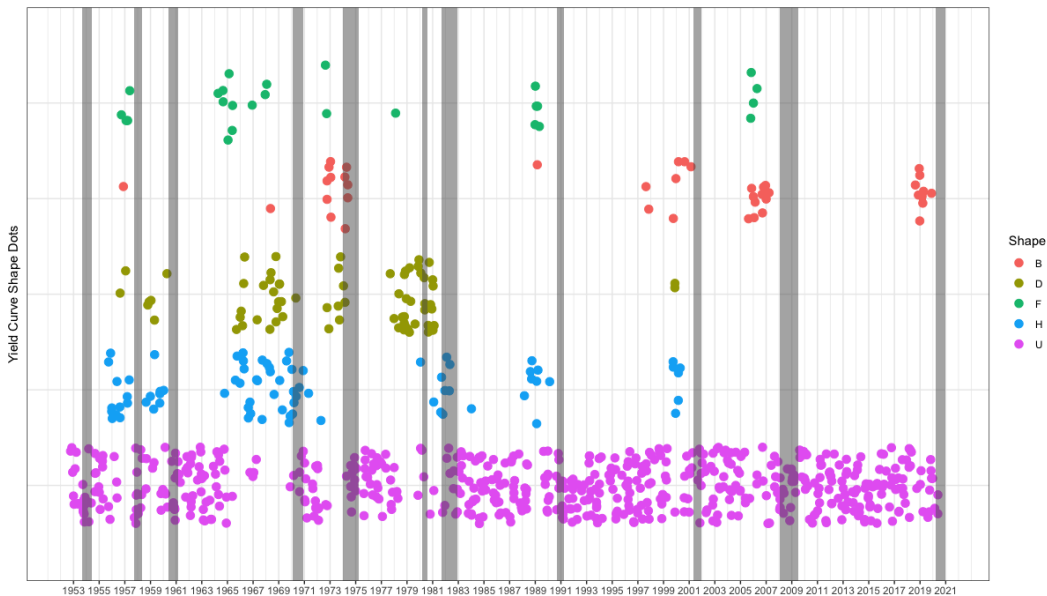Figure 4: U.S. Treasury yield curve: shape evolution (1953.04-2020.12)

Figure 4 displays the full picture since 1953.[12] The most frequent shape is plotted in the

---

[12]To differentiate the same shape observed multiple times in the same year, resampling methods are used in plotting the shapes for each year. Therefore, the dots clustered in any given year do not necessarily indicate their occurence over time.

bottom and the least on top. Evidently, the four minor shapes tend to cluster before the rercessions. More surprisingly, all minor shapes only show up before the recessions since late 1980s. Most alarming, the shape signals have become gradually more uniform to detect, as manifested by the pure B-shape in 2019, prior to the 2020 recession. A puzzle hence emerges: what theory can account for the time-varying ocurrences of these seemingly recession-related shape signals?

Despite these stunning patterns, no shape signal has been detected in the 1945, 1949, and 1954 recessions. In hindsight, the signals around 1966 could be wrong in foretelling a recession whereas actually it never arrived. Noises aside, whether such recession predictability exists and to what extent will be investigated in detail. A word of caution: neither can any shape cause a recession nor an expansion; they are merely parts of the business cycle fluctuations.

4 Shapes and Recessions

Can the shape of the yield curve predict recession? This section investigates the predictability ex-post along a number of dimensions and provides descriptive answers.

4.1 Timing, duration, and strength

Aforementioned, the four minor shapes tend to occur more often in the pre-recession stage than other periods. If we focus on a time frame spanning the recessions, the event study approach is well suited to catch their exact timing and duration.[13] To study the recession predictability of the yield curve shape, however, no cause and effect is present in its nature.

The beginning month of each recession is defined as the starting point of this event study. For the past ten successive recessions before which the shape of yield curve moves abnormally (relative to the normal shape), a window width of 18 months is chosen to cover the event in both directions.[14] Figure 5 summarizes the results with each colored dot representing one shape at a time. The vertical dashed line marks the beginning of a recession. On the whole, the shape "effects" to its left dramatically differ from those to the right, allowing for the overlapping 12-month when the 1980 post-recession met the 1981 pre-recession periods.

The timing of the shape is interesting but perplexing. The minor shapes were spotted as early as 18 months before the recessions, some even earlier. While the H-, D-, and F-shapes were the major forerunners in the earlier recessions, the B-shape played a more central role in the recent three recessions. In terms of effects, never had any other shape signaled more intensely than the downward-sloping yield curve in the early 1980s recessions. Historically, although rampant inflation was responsible for the formation of D-shape, the underlying economic factors casting

---

[13]In finance, an event study is a statistical method to assess the impact of an event on the value of a firm. Refer to World Bank DIME Wiki Event Study.

[14]The choice of 18 months suffices to cover more than 95% of the minor shapes. A 24-month window analysis is shown in the Appendix.

the H- and B-shape yield curves were little examined.

Figure 5: Treasury yield curve shape timing over recessions
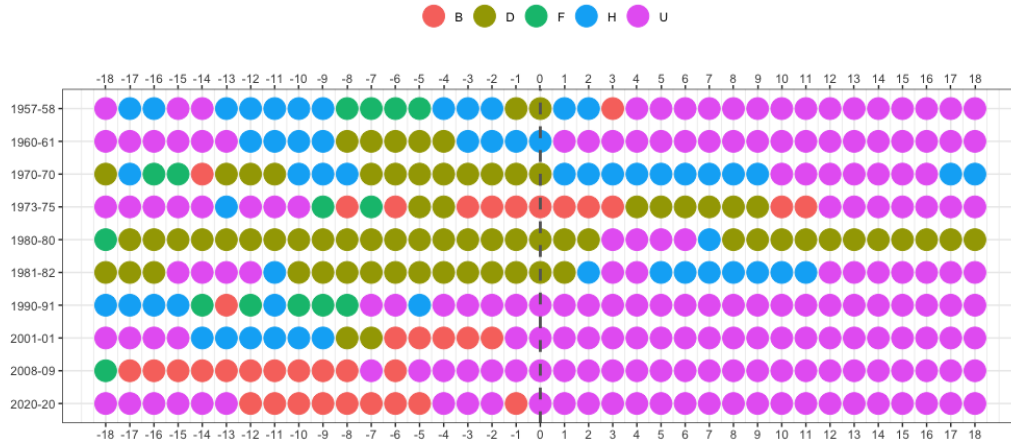
Another complexity in understanding the shape timing is the "noises" in the midst of recessions. Specifically, in the midst of the 1970 and 1974 recessions, were the H-, B-, and D-shapes signaling more economic downturns or displaying their remanent effects from earlier on? If they were trying to foretell an upcoming recession, why was it not coming in other cases? If they were just the remanent effects, why were there no such pattern in the recent recessions?

When it comes to their duration, each minor shape tend to cluster in batches. For the ten past recessions since 1957, Figure 6 plots shape counts within the 18-month window before each recession such that the relative importance of each shape is measured by its length in each bar. The horizontal dashed line sets a 50% benchmark for comparison. In stark contrast to its full-sample frequency, the upward-sloping yield curve declined to less than 50% in the 18-month event window prior to each recession. In the 1970 and 1980 pre-recession periods, the upward-sloping yield curve completely disappeared. Conversely, the four minor shapes took over and became unusually active in this 18-month window with duration much longer and frequencies much higher than other stages in the business cycle. Judging by the duration alone, however, not a single shape can dominate others over the entire sample in foretelling recessions, because the duration of each shape varied over time. Even the most dazzling D-shape observed in 1970s and 1980s was completely shadowed by the emerging B-shape in recent decades. As a result of the time-varying feature, it is not obvious to rank and access the relative performance of each minor shape in forecasting recessions.

Since the less frequent shapes become more active in the event window of the recessions, are more severe recessions associated with more intense signals of the shape? To answer this

Figure 6: Treasury yield curve shape duration before recessions
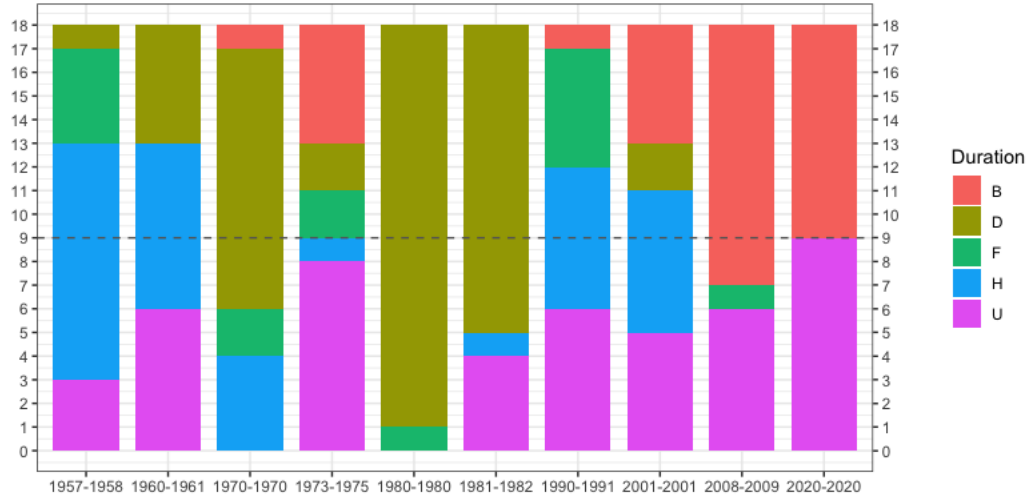
question, it is necessary to classify recessions by some degree of severity and measure the strength of the shape signals.

For the past ten recessions preceded by the less frequent shapes since 1957, their length varied and most of the minor shapes were concentrated in the pre-recession stage. With all available data, I classify the ten recessions into four tranches in terms of their respective length: the mild recessions of the 1980 (less than two quarters), the moderate of the 1957-1958, 1990-1991, 2001-2001 (more than two quarters but less than three quarters), the serious of 1960-1961, 1970-1970, and 2020 (more than three quarters but less than one year), and the 1973-1975, 1981-1982, 2008-2009 severe recessions (longer than a year). Overall, this grouping balances the number of recessions in each tranche.[15]

In similar fashion, the strength of shape signals can be measured by their respective counts within the recession-related windows. However, two adjustments are made to filter out "noises" and amplify signals. Notice that the 1980 post-recession overlapped the 1981 pre-recession for 12 months; the minor shape signals were consistently strong in between. It is more convincing to consider these occurrences as signaling the 1981-1982 severe recession rather than some remaining effects from the 1980 mild recession. Another adjustment is due to the unusually long lasting signals observed before the 1970 serious recession: a continuum of 30 months

---

[15]Based on the NBER business cycle chronology from 1957 to present, the shortest recession endures for six month (1980), the longest 18 months (2008-2009), three for eight months (1957-1958, 1990-1991, 2001-2001), two for 16 months (1973-1975, 1981-1982), one for ten months (1960-1961), and one for eleven months (1970-1970). As of December, 2020, the U.S. economy has been in recession for ten months since March. https://www.nber.org/cycles.html

of minor shapes were persistently present since July, 1967. Even though the 18-month event window was chosen to "receive" signals before recessions, it is not convincing to consider the extra 12 months as "noises" and discard them in this particular case. Thus, I relate them to the 1970 recession.

Figure 7: Shape signal strength and recession duration (1953-2020)
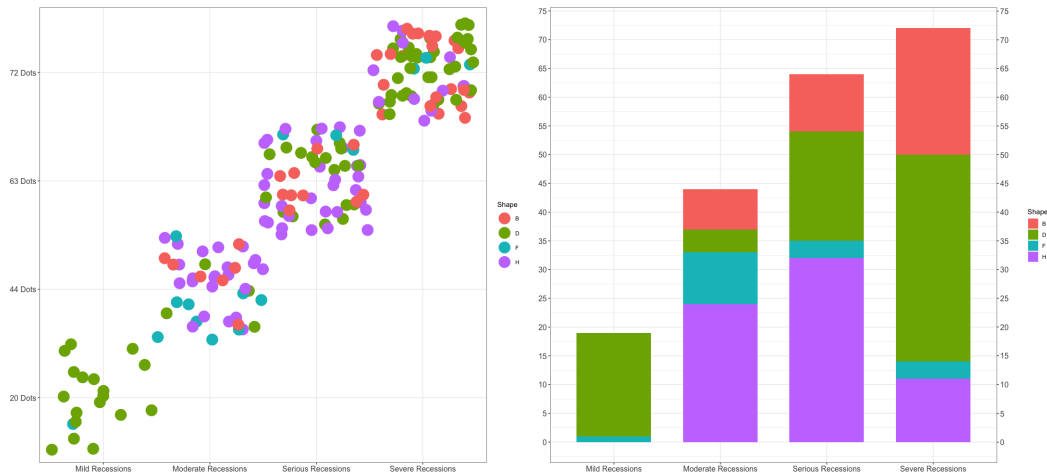
Matching the shape strength to recession duration, Figure 7 displays their association. Those colored dots in the left graph represent the minor shapes and their clustered density mounts in conjunction with the duration of recession. In the right graph, the height of each bar is the total counts of the minor shapes and it rises as a recession lasts longer.

4.2 The wrong and missing signals

Undoubtedly, a strong link exists between the shape and recession. Nonetheless, exceptions can be found. The most obvious outliers are a series of strikingly persistent minor shape signals from the end of 1964 to the beginning of 1967. Specifically, from December, 1964 to February, 1967, a total of 22 minor shapes were present in 27 months. Were they related to any recession in the interval? Was the wolf coming or just a false alarm?[16]

According to the data from the World Bank, the U.S. economy was geared down to a much slower growth rate of 2.5% in 1967 from 6.5% in the previous year, then it bounced back to 4.8% in 1968.[17] It turns out that, for the post-WWII U.S. economy, an annual decline of 4%

---

[16]From July, 1967 to October, 1970, a continuum of 40 months of minor shapes were present. The 1970 recession started from January to November. Compared with the other recessions, the signals prior to the 1970 recession were unusually long and remained strong toward the end of the recession. Despite the "noises," this batch is believed to contain true signals which forecast the 1970 incoming recession.

[17]https://www.macrotrends.net/countries/USA/united-states/gdp-growth-rate

real GDP growth rate is the sharpest in record without triggering a recession, whereas the NBER business cycle dating committee only considers more than two consecutive quarters of negative growth rate in real economic activities as a recession. Other than that, no obvious "mistake" was detected in the sample when evaluating the false positive signals.[18] Hence, the seemingly wrong signals were indeed a reliable indicator of future economic downturns.

On the false negative side, the obvious exceptions appear in three consecutive recessions after 1945. Why was there no signal before the wolf was coming? A short answer is the U.S. fiscal and monetary policy then facilitating low-cost wartime finance. According to the 2016 FOMC document "Targeting the yield curve: The experience of the Federal Reserve, 1942-51," during the nine-year period from early 1942 until the Treasury-Federal Reserve Accord of 1951, the yield on long-term Treasury bonds was capped at 2.5 percent, and ceilings were also imposed at several other points. In addition, the short-term yields were pegged at 3/8 percent up until July 1947. The yield curve control policy exerted long-lasting influence on the market yields: It was until after 1950, the short and long term yields rose gradually and the yield spreads started to narrow. And later the first post-WWII yield curve inversion arrived in 1957.

When the short and long yields were targeted at their respective values, the shape of yield curve was fixed on both ends. As long as the median yields were in line with the policy comittment, an upward-sloping yield curve was the result. Therefore, shape signals were muted before recessions under the policy of interest rate controls. In the post-WWII data, provided that the Fed did not peg the interest rates, the Treasury bond market has been reliably signaling recessions or downturns. This is critical in understanding the prediction of yield curve as recession signals. The next step is to apply model estimation to assessing its predictive power.

5   Recession Predictability

Every now and then, it is alarmingly reported in the news and media that a flatting or inverted yield curve precedes most recessions. Since the late 1980s, academic literature have begun investigating the power of the yield curve as a recession predictor in the post-WWII data. Estrella et al. (2003) find that term spread models predicting recessions are stable over time. Wheelock and Wohar (2009) provide a comprehensive literature survey. It is further shown that the yield spread model dominates the professional forecasters in real-time recession forecast (Rudebusch and Williams, 2009; Croushore and Marsten, 2016). However, existing research relies exclusively on a particular slope, or term spread, of the yield curve. Other parts of information in the yield curve had been largely ignored. Since the shape indicator distills all essential information of the entire yield curve, would it be more useful in predicting recessions than the spreads?

---

[18]That said, a single isolated H-shape was spotted in September, 1984, but was 21 months away from the 1981-1982 severe recession. Should it be a false signal, but don't be too hard on the hero.

5.1 Spreads vs shapes

The starting point of applying the yield curve slope or yield spread to recession forecast can be featured by tracking the size of the yield curve inversion over the business cycle. Shown in Figure 3, preceding the past ten recessions, the three yield spreads almost all narrowed down and became negative. On the whole, the long-short spread exhibited the most sizable inversions among the three. Does it send more powerful signals than others? Moreover, does more negative values of the spreads deliver more reliable forecasts? To answer these questions, we need to compare their changing patterns across recessions and then develop formal statistical tests.

Figure 8: Change of yield spreads and curvature across recessions
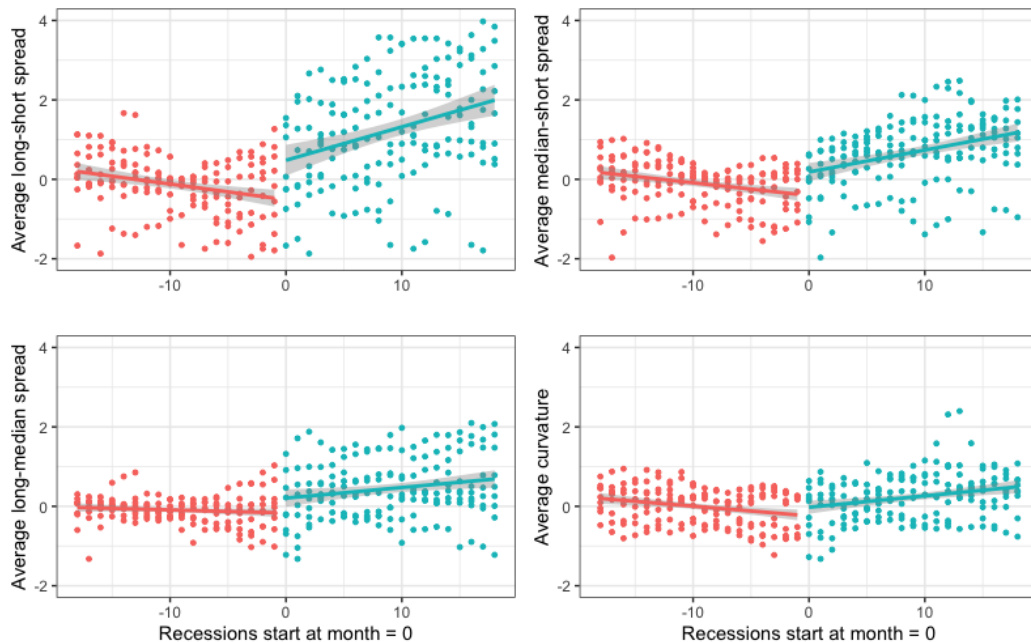
Following the earlier event study approach, I focus on the 18-month window asymmetric around the starting month of recessions and compare the patterns of the yield spreads across the board. Furthermore, I mimic the regression discontinuity design (RDD) and fit the data with a linear equation on both side of the threshold in order to detect whether the spreads drop significantly below zero prior to the recessions. No data adjustment is made for overlapping recession stages. The results are in Figure 8 with 95% confidence bands for the estimated equations. Several observations sharpen our cross-sectional comparison: 1) all yield spreads and curvature share a common trend—sliding down before recessions and rising up in and after recessions; 2) the long-short and median-short spreads show a statistically significant dip below zero when closing in to the border; 3) the long-short and median-short spreads show a statis-

tically significant regression discontinuity across the threshold; 4) the long-median spread and the curvature do not exhibit significantly different patterns across the border, which indicates a lack of variation and strength in forecasting recessions.

To estimate and forecast recession probabilities, a statistical model is required. For a binary recession variable, the most widely adopted model in the forecasting literature is the Probit model. Based on the Probit statistical model, macroeconomists apply the slope of the yield curve, or the yield spread, to predict recessions and many other economic activities. [19]

While it is straightforward to measure the yield curve slope as the difference between a long-term yield and a short-term yield, the question is which yield shall be chosen? In the literature, the choice of the long and short yields depends on data availability and measurement consistency (Estrella and Trubin, 2006). For instance, Estrella and Hardouvelis (1991) chose the 10-year Treasury bond rate and the 3-month Treasury bill rate. Benzoni and Chyruk (2018) employed the 10- minus 2-year Treasury spread. Another popular choice is the 10-year Treasury yield less the Federal funds rate. However, were it not for convenience, there is no reason to omit other maturity sectors of the yield curve. To fully explore the information content of the entire yield curve, the three average yield spreads and the shapes will all be used in forecasting recessions.

Unlike the numerical spread measures, the classified yield curve shape is a categorical variable and takes five discrete values. To distinguish their information content and generate meaningful forecast in regressions, separate binary variables are defined for distinct shapes as inputs in the Probit model, where each shape only takes zero or one.[20]

$$P(Recession_{t+h}|Spread_t) = P(Z \le \alpha + \beta * Spread_t) = \Phi(\alpha + \beta * Spread_t) \tag{1}$$

$$P(Recession_{t+h}|Shape_t) = P(Z \le \gamma_B B_t + \gamma_D D_t + \gamma_F F_t + \gamma_H H_t + \gamma_U U_t) \tag{2}$$

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-0.5u^2} du \tag{3}$$

The simple Probit model is our starting point and well-suited to compare the recession predictive content of the yield spreads and the shapes. Covering the full sample of April 1953 to

---

[19] Another binary response model in application is the Logit model with which Stock and Watson (1989), for example, have also documented predictive power of the yield curve slope. Both the Probit and Logit models generate similar estimation results. Furthermore, based on a stability test, Estrella et al. (2003) identified that the spread model that predicts recession is stable over time in countries like the United States and Germany.

[20] The advantage of setting up the model in this manner comes from the mutually exclusive content of each regressor in the time series data. Therefore, the estimation results based on this model would reflect the average predictive strength of a particular shape over the entire estimation sample, independent of other shapes. Note that no intercept shall be included to avoid regression multicollinearity.

December 2020, our first exercise provides coefficient estimates and model fits in forecasting recession one year ahead.

In Table 7, almost all coefficient estimates are statistically significant except for three of the minor shapes.[21] Based on the peusdo-R^2 measures, the models using the median-short spread and the shapes have the best fit, whereas the curvature shows little predictive power relative to a constant intercept model. Among the spreads, the sub-sample analysis shows that the long-short spread produces finer fitted model in the first half sample (1953-1986) and the median-short spread in the second (1987-2020) as well as the overall sample.

[Table 7 is about here]

Figure 9 visualizes the estimation and comparison results. Both the spread and shapes display maximum forecast power of recessions around four quarters ahead, whereas their forecasted recession probabilities are much lower in the most immediate and distant future. By construction, a decline in yield spread is associated with rising probability in future recessions. In contrast, the shapes do not generate a continuous range of recession probability forecasts as the spread does. Nonetheless, the four minor shapes on average produce recession forecasts at least twice as much as their benchmark upward-sloping shape whose forecasts are uniformly smaller than the unconditional recession probability 14.9% in the sample.

Figure 9: Recession probability forecasts: yield curve spread vs shapes (1953.04-2020.12)
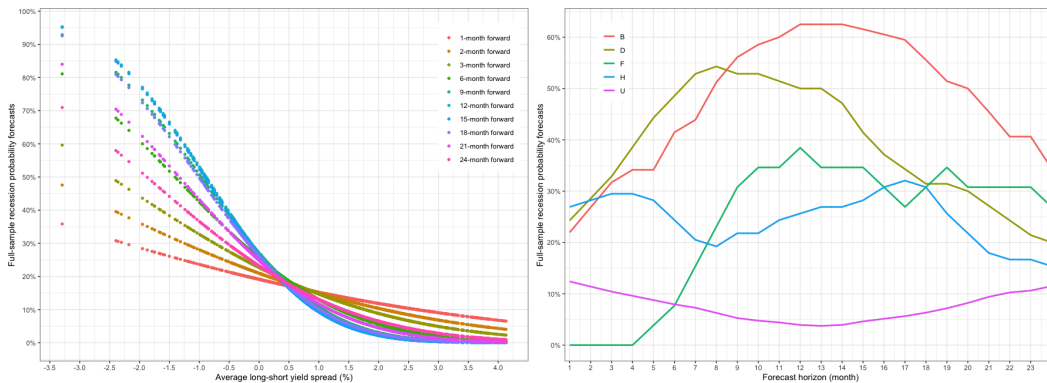
Figure 10 recovers the one-year-ahead full-sample estimates of recession probabilities. Out of eleven recessions, the median-short spread dominates the long-short and the median-short spreads in approaching true recession probabilities; the shape predictor outperforms the median-short spread in most recessions except the ones in the mid-70s and early-80s. Despite the binary

---

[21]Due to the binary nature of the variables as well as their less frequent occurrence, the estimates of the coefficients of the shape model exhibit much higher variation than the spread model.

nature of the shapes, the step-wise recession probabilities they generate in this exercise is on average more accurate than any other spreads in predicting recessions a year ahead.

Figure 10: Full-sample one-year-ahead recession probability forecasts (1953.04-2020.12)
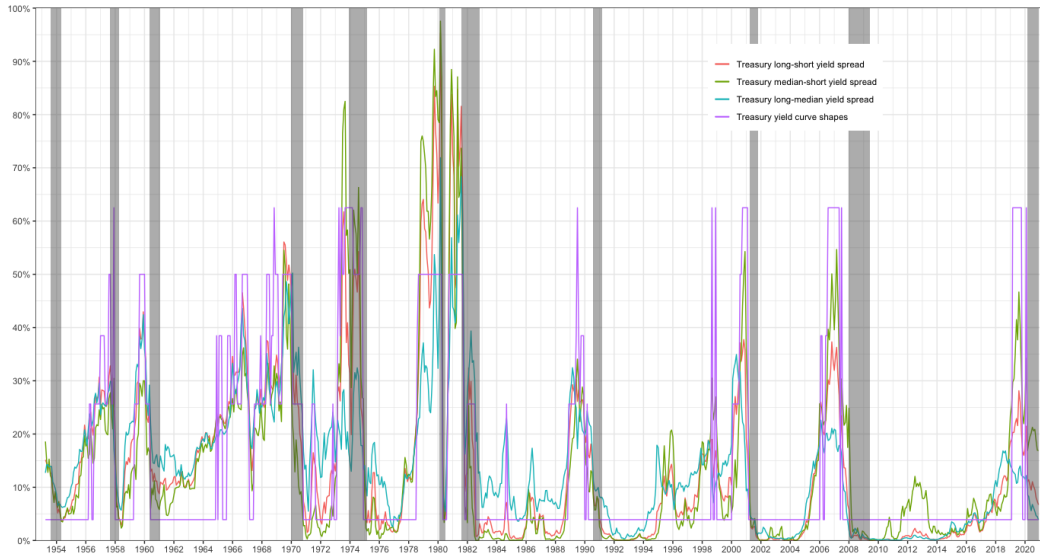
Though counter-intuitive and unexpected, these findings imply that the size of the negative spread alone might not necessarily be a good predictor of the forthcoming recessions; different choices of the yield spread perhaps weights more. In particular, the discovery of potentially more powerful predictors than the long-short spread is encouraging. It is time to evaluate their predictive power in formal metrics.

5.2   Forecast evaluation

To measure the predictive performance of any model or predictor, the mean absolute error (MAE) and the root mean square error ($RMS E$) are two most popular metrics, corresponding to $L1$ and $L2$ loss functions, respectively. While the MAE is more robust to outliers, the $RMS E$ penalizes more heavily those predictions far away from actual values. A more specialized measure adopted in evaluating probability forecast is the log probability score (LPS), which coordinates the in-sample estimation criterion with the out-of-sample loss function for the yield spread forecasts (Rudebusch and Williams, 2009).

For any forecasting horizon h and full sample size T, the calculation of prediction error is shown in equations below, where $\hat{P}_i$ is the forecasted recession probability at $i + h$ period ahead based on the all observations (used in the estimation sample) up to $T - h$ and I_i is the actual recession indicator taking a binary value of either zero or one.

$$MAE(\hat{P}_i) = \frac{1}{T-h} \sum_{i=1}^{T-h} | \hat{P}_i - I_i | \tag{4}$$

$$RMSE(\hat{P}_i) = \sqrt{\frac{1}{T-h} \sum_{i=1}^{T-h} (\hat{P}_i - I_i)^2} \tag{5}$$

$$LPS(\hat{P}_i) = \frac{-1}{T-h} \sum_{i=1}^{T-h} [(1 - I_i)ln(1 - \hat{P}_i) + I_i ln(\hat{P}_i)] \tag{6}$$

$$CER(\hat{P}_i) = E(\hat{I}_i) = \frac{1}{T-h} \sum_{i=1}^{T-h} |I_i - \hat{I}_i|, \text{ where } \hat{I}_i = \begin{cases} 1, & \text{if } \hat{P}_i > \text{cutoff} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

In practice, recession prediction is also a classification problem. A threshold or cutoff value, therefore, must be chosen to decide whether the probabilistic forecasts of a recession can be considered as a positive event. As the most common metric used to quantify a binary classifier, the classification error rate (CER) or misclassification rate is defined as percent of false signals out of total predictions. Similarly, if I_i is the recession indicator from the test data and $\hat{I}_i$ is the binary classifier of a recession event from the ith forecast using the training data in estimation, for any chosen cutoff value, $\hat{I}_i = 1$ if $\hat{P}_i$ >cutoff and $\hat{I}_i = 0$ otherwise. A simple choice of 50% for the threshold is common. Alternatively, a reasonable cutoff can be calculated from the unconditional recession probability in the training sample.

More recently, a good number of research applied receiver operating characteristics (ROC) and area under the curve (AUC) method to assessing recession classification accuracy (Berge and Jordà, 2011; Lahiri and Wang, 2013; Liu and Moench, 2016). This approach first computes two types of recession classification outcomes—false positive rate (TPR) and true positive rate (TPR)—for a full range of threshold, traces out the ROC curve by plotting TPR against FPR, and then summarizes the forecast performance in a single value by integrating over the ROC curve. As a comprehensive measure of probabilistic forecast error, a rule of thumb in model selection is to favor the one with the highest AUC.

It is worth noting that these measures of forecast errors incorporate different philosophy and standards in assessing the relative performance of competing models/predictors. Hence, multiple comparisons and statistical tests are necessary to further our analysis.

5.2.1 In-sample analysis

In our full-sample evaluation, Table 8 compares the different measures of forecast errors between the spread models and the shape model for horizons up to eight quarters ahead. First, all forecast errors reach their minima around four or five quarters ahead. Second, among the spreads, the median-short spread has the smallest errors in the four-quarters-ahead recession

forecasting by all measures. Last, the shapes fare slightly better than the median-short spread except for the *LPS* and *AUC* measures.

[Table 8 is about here]

When estimating the model for the subsample (1953-1986), the yield curve shapes outperform the spreads by three measures (*MAE*, *RMS E*, *LPS*) and the long-short spread slightly dominates the median-short by all measures around four or five quarters ahead. For the second subsample (1987-2020), the median-short spread is the champion by all measures and the shapes do slightly better than the long-short spread by *MAE*, *RMS E*, *CER*. Diebold-Mariano tests (1995, 1997) for the equality of predictive accuracy show that the relative forecast performance of the shapes, the long-median, and the long-short spreads are not statistically significant, but they are all better than the long-median spread. To save space, the results for D-M tests and sub-sample analysis are in the Appendix.

5.2.2   Out-of-sample analysis

In real-time forecasting, expanding- and rolling-window analyses are routinely applied to evaluate the out-of-sample performance among competing models. The procedure developed in this part shall carry more weight than the in-sample analysis.

The expanding window approach starts from a certain point in time and adds in one additional observation recursively in estimation and forecast. The choice of a starting point must at least include enough historical data to produce estimation results. In practice, no strict rule is imposed regarding the minimum portion of the sample in its first estimation. An alternative out-of-sample forecasting evaluation employs the rolling window analysis. The estimation begins with a fixed bandwidth and slides the window stepwise to the end of the sample. While the expanding window method explores all information available in the sample and weighs historical data more heavily regardless of their relevance to present situation, the rolling window omits all historical data and tracks the most ongoing development closely. A comparison between the two can also serve as a robustness check. To cover at least two recessions in estimation, our analysis selects an initial window of 20 years (1953-1972).[22]

Figure 11 delivers the comparisons for forecasting horizons from one-month to 30-month in the future. Overall, forecast errors fall initially but rise quickly beyond the 16-month horizon. Similar to the full-sample estimation results, all measures of forecast error indicate an optimal forecasting horizon in between four and five quarters ahead. Though statistically insignificant,

---

[22]Due to the announcement procedure of the NBER business cycle dating committee, U.S. recessions have always been identified at least three months later than its "book" value (what the public later know in hindsight). The forecasting procedure developed here is strictly speaking not based on the real-time information.

Figure 11: Expanding window out-of-sample forecast errors (1953-2020)
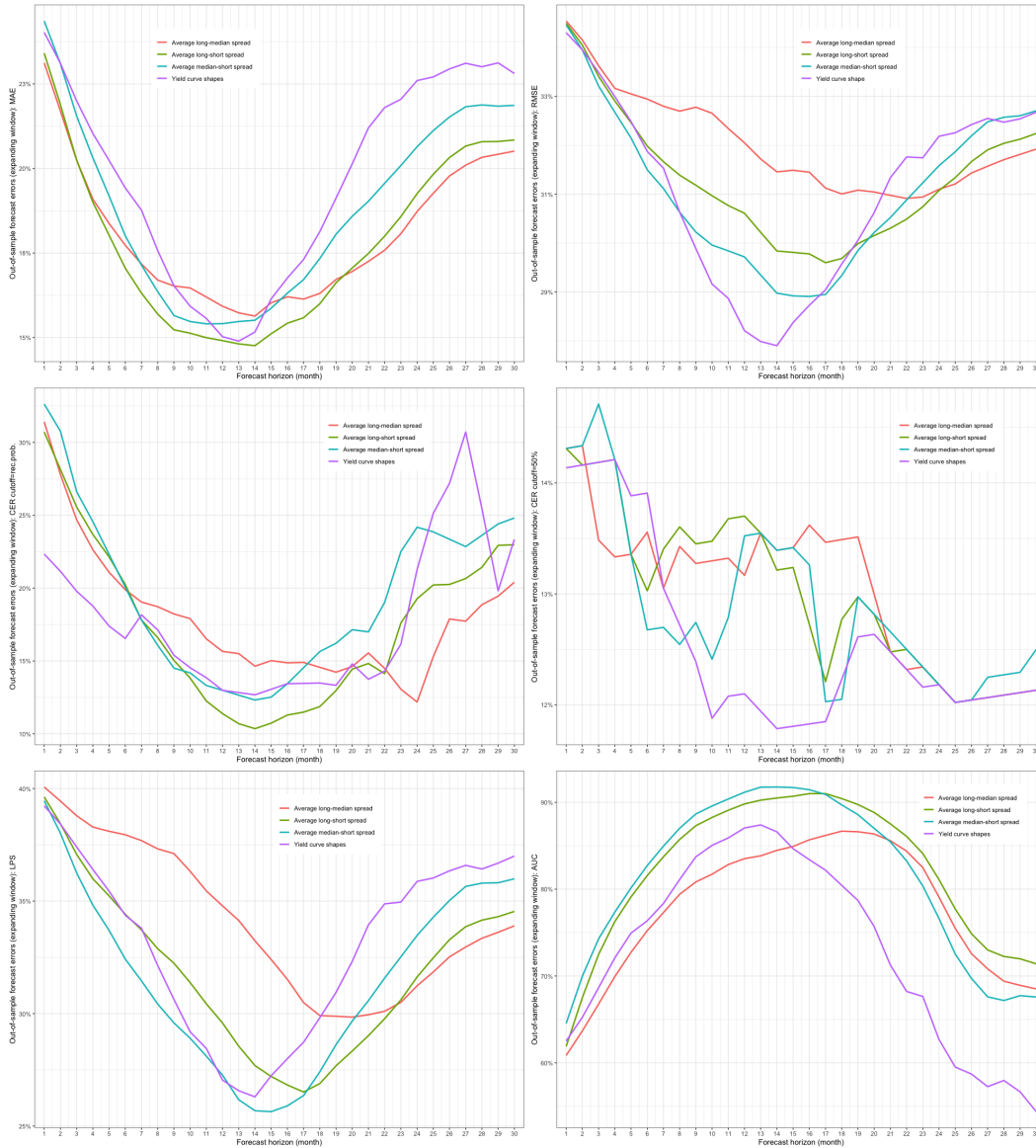
Top left panel: *MAE*–mean absolute error. Top right panel: *RMS E*–root mean square error. Middle left panel: *CER* cutoff=rec.prob–classification error rate for a probability threshold equal to 15%. Middle left panel: *CER* cutoff=50%–classification error rate for a probability threshold equal to 50%.Bottom left panel: *LPS*–log probability score. Bottom right panel: *AUC*-area under the curve. A window width of 20 years is initialized in expanding window analysis. Other choices of window width are shown in the Appendix.

the *MAE* slightly favors the long-short spread in the near- and median-term horizons. Nonetheless, the *RMS E*, *LPS*, and *AUC* all select the median-short spread as best predictor among
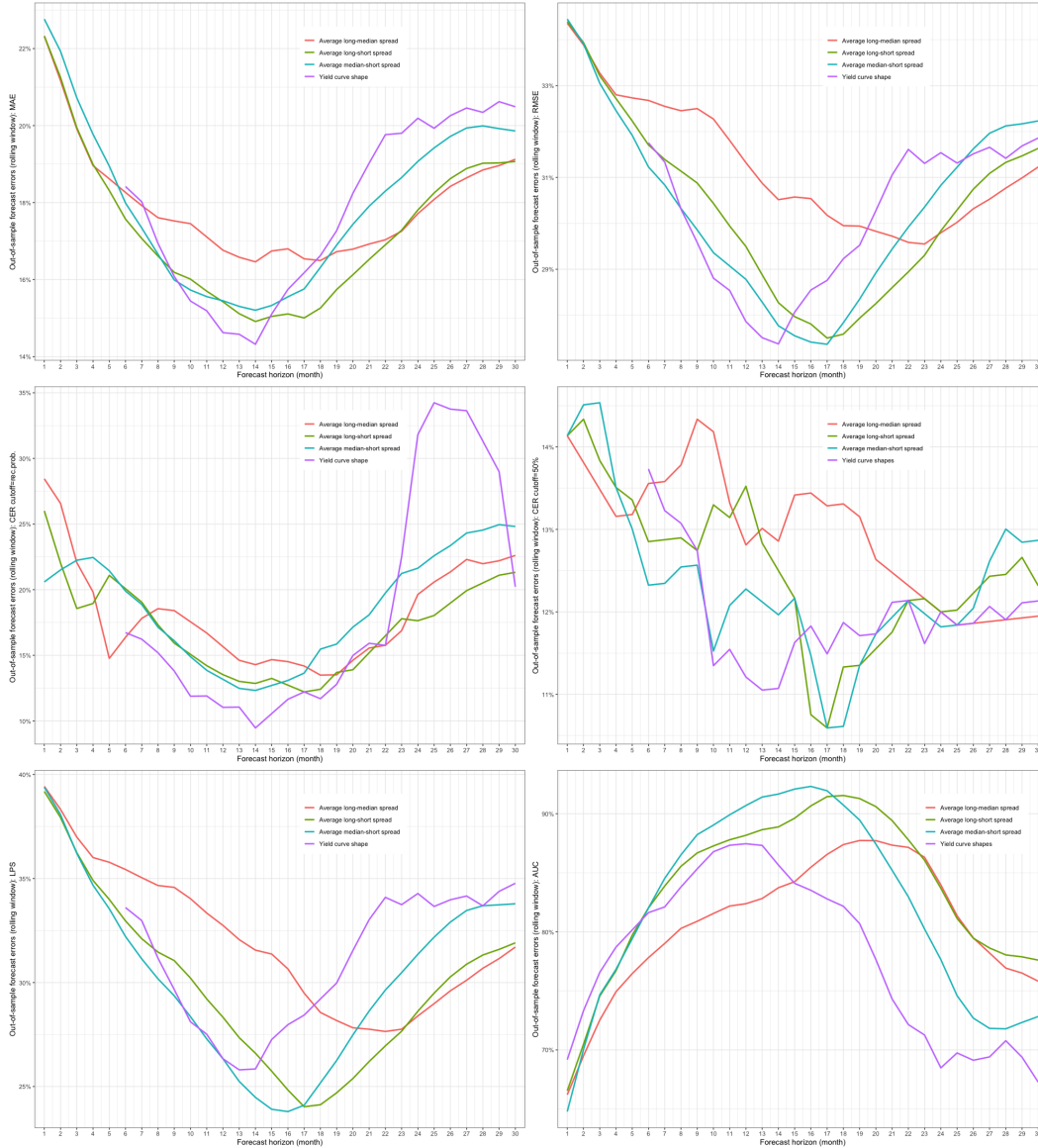
Figure 12: Rolling window out-of-sample forecast errors (1953-2020)

Top left panel: *MAE*–mean absolute error. Top right panel: *RMSE*–root mean square error. Middle left panel: *CER* (cutoff=rec.prob)–classification error rate for a probability threshold equal to 15%. Middle left panel: *CER* (cutoff=50%)–classification error rate for a probability threshold equal to 50%.Bottom left panel: *LPS*–log probability score. Bottom right panel: *AUC*-area under the curve. A fixed window width of 20 years is initialized in rolling window estimation and forecast. Other choices of window width are shown in the Appendix.

all spreads. The yield curve shapes also beat the long-short spread in terms of the *RMSE* and *LPS*, irrespective of window width. While the *CER* with a 15% cutoff probability prefers the

long-short spread, the one with 50% suggests the opposite. It is worth noting that the long-short and long-median spreads excel in long-horizon forecast (beyond 18 months). Other choices of window width further confirm most of these findings.

Figure 12 presents the results for the rolling window forecasts. The results are very similar to those of the expanding window analysis. By most measures except for the MAE, the median-short spread dominates the long-short spread for horizons up to 17 months, irrespective of the width of estimation window. Measured by the MAE, the predictive model based on the shapes improves a little and can even overtake all the spreads around one-year ahead forecast, though the results are not robust when the window size increases. Again, Diebold-Mariano tests (1995, 1997) for the equality of predictive accuracy show that the relative forecast performance of the shapes, the long-median, and the long-short spreads are not statistically significant, but they are all better than the long-median spread.

6 Concluding Remarks

In the year preceding the 2020 recession, the B-shape yield curve became the harbinger again and signaled nine times. In retrospect, it is intuitive, straightforward, and reliable to just look at the shapes. This paper emphasizes the significant role of tracking the yield curve shape in monitoring business cycles and predicting recessions.

Controlling for different choices of estimation procedure, window width, forecasting horizons, and forecast errors, the median-short spread is, on the whole, a more powerful predictor of recessions than the long-short spread, especially after the 1980s. Nevertheless, the performance of the yield curve shape is also remarkable by various measures of forecast error, though not statistically significant relative to the long-short spread. Even within a simple Probit model, the analysis developed here highlights the complexity and difficulty in real-time recession forecast evaluation. To address structural breaks and serial correlation for more robust results, more sophisticated model specification and estimation techniques, such as in Chauvet and Potter (2005), remain for future research.

To conclude, the shape of the Treasury yield curve is a concise and valuable indicator in studying the evolution of the term structure in general and predicting recessions in particular. While presenting a good number of new findings, this paper poses challenges and puzzles in search of hidden forces driving the evolution of the yield curve shape. Future research cries out for a unifying economic theory to explain the recession predictive power of the yield curve. Among the puzzles, the time-varying pattern of recession shape signals is perhaps most intriguing. Moreover, would the shape of the yield curve reflect underlying states of the economy? How do risk premia vary across the shapes?

The lack of data limits our shape identification of the Treasury yield curve for the first half of the 20 century. Future research can extend the data further back and seek more evidence on

the links between the yield curve shape and the recession. With respect to the exceptions (the wrong and missing signals), it is more crucial to examine the policy and institutional constraints than to test its predictability. In hindsight, the yield curve could make mistakes but policies also distorted the market signals. To investigate the causes shaping the yield curve over the business cycle, noises contained in the low-frequency macroeconomic data are notoriously difficult to filter out. A promising direction is to look into high frequency data and analyze the market microstructure. Ultimately, demand and supply of the Treasuries determine the equilibrium term structure, hence, the shape of the yield curve.

References

Benzoni, L., Chyruk, O., 2018. Why does the yield-curve slope predict recessions? *Chicago Fed Letters*, Federal Reserve Bank of Chicago. No. 404.

Campbell, J. Y., 1995. Some lessons from the yield curve. *Journal of Economic Perspectives*, 9(3), 129-152.

Chen, B., 2019. *Shapes and transitions of the interest rate term structure*. Doctoral Dissertation, The City University of New York, Graduate Center. New York.

Chauvet, M., Potter, S., 2005. Forecasting recessions using the yield curve. *Journal of Forecasting*, 24, 77-103.

Croushore, D., Marsten, K., 2016. Reassessing the relative power of the yield spread in forecasting recessions. *Journal of Applied Econometrics*, 31, 1183-1191.

Diebold, F. X., Mariano, R. S., 1995. Comparing forecast accuracy. *Journal of Business and Economic Statistics*, 13, 253-26.

Estrella, A., Hardouvelis, G. A., 1991. The term structure as a predictor of real economic activity. *Journal of Finance*, 46(2), 555-576.

Estrella, A., Mishkin. F. S., 1998. Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1), 45-61.

Estrella, A., Rodrigues A. P., S. Schich, 2003. How stable is the predictive power of the yield curve? Evidence from Germany and the United States. *Review of Economics and Statistics*, 85(3): 629-644.

Estrella, A., Trubin, M. R., 2006. The yield curve as a leading indicator: some practical issues. Federal Reserve Bank of New York *Current Issues in Economics and Finance*, 12(5).

Fabozzi, F. J., et al, 2019. *Foundations of global financial markets and institutions*. Fifth edition. MIT Press.

Gürkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54, 2291-2304.

Harvey, D., Leybourne, S., Newbold, P. , 1997. Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2), 281-291.

Kessel, R., ed. 1965. *The cyclical behavior of the term structure of interest rates*. National Bureau of Economic Research. New York.

Laurent, R. D., 1988. An interest rate-based indicator of monetary policy. *Federal Reserve Bank of Chicago Economic Perspectives*, 12(1), 3-14.

Litterman, R. B., Scheinkman. J., 1991. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1), 54-61.

Macaulay, F. R., 1938. *Some theoretical problems suggested by the movements of interest rates, bond Yields and stock prices in the United States since 1856*. National Bureau of Economic Research. New York.

Malkiel, B. G., 1966. *Term structure of interest Rates: expectations and behavior patterns.* Princeton University Press. New Jersey.

Mishkin, F. S., 2019. *The Economics of money, banking and financial markets.* 12th Edition. Pearson.

Rudebusch, G. D., Williams, J. C., 2009. Forecasting recessions: the puzzle of the enduring power of the yield curve. *Journal of Business and Economic Statistics*, 27(4), 492-503.

Wheelock, D. C., Wohar, M. E., 2009. Can term spread predict output growth and recessions? A survey of the literature. *Federal Reserve Bank of St. Louis Review*, 91(5), 419-440.

Lahiri, K., Wang, J. G., 2013. Evaluating probability forecasts for GDP declines using alternative methodologies. *International Journal of Forecasting*, 29(1), 175–190.

Berge, T. J., Jordà, 2011. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3-2, 246–277.

Liu, W., Moench, E., 2016. What predicts US recessions? *International Journal of Forecasting*, 32, 1138–1150

Table 7: Probit model recession forecast coefficient estimates and model fits (12-month ahead)

Panel A: Spread model $P(Recession_{t+12}|Spread_t) = P(Z \leq \alpha + \beta * Spread_t) = \Phi(\alpha + \beta * Spread_t)$

Predictor		1953.04-2020.12				1953.04-1986.12				1987.01-2020.12		
	α	β	MF R^2	ML R^2	α	β	MF R^2	ML R^2	α	β	MF R^2	ML R^2
$Spread_{l-s}$	-0.61*	-0.70*	0.25	0.19	-0.80*	-0.82*	0.26	0.21	-0.23	-0.80*	0.28	0.18
	(0.07)	(0.07)			(0.09)	(0.10)			(0.14)	(0.12)		
$Spread_{m-s}$	-0.62*	-1.32*	**0.29**	**0.21**	-0.74*	-0.19*	**0.26**	**0.21**	-0.25	-1.86*	**0.36**	**0.22**
	(0.07)	(0.12)			(0.09)	(0.15)			(0.13)	(0.25)		
$Spread_{l-m}$	-0.75*	-1.00*	0.15	0.12	-0.92*	-1.74*	0.24	0.20	-0.56*	0.90*	0.14	0.10
	(0.06)	(0.12)			(0.08)	(0.23)			(0.13)	(0.17)		
$Curvature$	-1.07*	-0.48*	0.04	0.03	-0.75*	1.03*	0.20	0.17	-1.29*	-0.40*	0.03	0.02
	(0.06)	(0.10)			(0.08)	(0.19)			(0.09)	(0.14)		
							0.10	0.09				

Panel B: Shape model $P(Recession_{t+12}|Shape_t) = P(Z \leq \gamma_B B_t + \gamma_D D_t + \gamma_F F_t + \gamma_H H_t + \gamma_U U_t)$

Sample	γ_B	γ_D	γ_F	γ_H	γ_U	MF R^2	ML R^2
1953.04-2020.12	0.31	0.00	-0.29	-0.65*	-1.76*	0.28	0.21
	(0.20)	(0.15)	(0.25)	(0.15)	(0.09)		
1953.04-1986.12	0.50	-0.04	-0.38	-0.78*	-1.88*	0.28	0.23
	(0.36)	(0.15)	(0.31)	(0.18)	(0.16)		
1987.01-2020.12	0.23	5.02	-0.14	-0.18	-1.68*	0.29	0.18
	(0.24)	(166.2)	(0.41)	(0.33)	(0.11)		

Note: Models are estimated via maximum likelihood method. The intercept term is excluded from the shape model to avoid multicolinearity problem. Coefficient estimates with an asterisk are significant at less than 0.1% . Standard errors are shown in parentheses. MF R^2 is the McFadden's pseudo R-squared and ML R^2 is the maximum likelihood pseudo R-squared. Both are a measure of goodness of fit for the Probit model regression. The pseudo R^2 is defined as $1 - ln\hat{L}(M_a)/ln\hat{L}(M_r)$, where M_a is the unrestricted model with intercept and slope predictors, M_r is the restricted model with only an intercept term, and $\hat{L}$ is the estimated likelihood.

Table 8: Full-sample evaluation of recession probability: forecast errors (1953.04-2020.12)

| | MAE | | | | RMSE | | | |
	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$
1Q	0.239	0.236	0.245	0.237	0.345	0.343	0.349	0.344
2Q	0.218	0.211	0.233	0.215	0.327	0.325	0.338	0.328
3Q	0.196	0.186	0.219	0.188	0.310	0.305	0.327	0.307
4Q	0.187	0.178	0.210	0.175	0.305	0.302	0.321	0.296
5Q	0.189	0.182	0.209	0.185	0.309	0.306	0.322	0.304
6Q	0.199	0.199	0.210	0.203	0.321	0.319	0.327	0.319
7Q	0.212	0.215	0.217	0.224	0.331	0.331	0.334	0.335
8Q	0.223	0.227	0.225	0.236	0.337	0.338	0.339	0.343

| | CER (cutoff=50%) | | | | CER (cutoff=15%) | | | |
	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$
1Q	0.148	0.147	0.149	0.149	0.42	0.37	0.49	0.24
2Q	0.143	0.144	0.147	0.147	0.33	0.29	0.41	0.21
3Q	0.139	0.131	0.138	0.133	0.27	0.24	0.36	0.20
4Q	0.135	0.131	0.137	0.129	0.24	0.22	0.33	0.18
5Q	0.133	0.133	0.139	0.128	0.24	0.24	0.31	0.20
6Q	0.140	0.143	0.147	0.135	0.27	0.27	0.30	0.22
7Q	0.139	0.141	0.145	0.140	0.30	0.31	0.31	0.26
8Q	0.141	0.139	0.144	0.141	0.35	0.35	0.34	0.29

| | LPS | | | | AUC | | | |
	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$	$Spread_{l-s}$	$Spread_{m-s}$	$Spread_{l-m}$	$Shapes$
1Q	0.392	0.387	0.405	0.390	0.679	0.698	0.635	0.661
2Q	0.357	0.345	0.383	0.360	0.773	0.792	0.710	0.721
3Q	0.319	0.303	0.359	0.317	0.839	0.864	0.763	0.795
4Q	0.303	0.290	0.344	0.292	0.866	0.887	0.791	0.831
5Q	0.303	0.296	0.339	0.308	0.857	0.868	0.797	0.803
6Q	0.318	0.324	0.337	0.340	0.829	0.819	0.798	0.744
7Q	0.342	0.353	0.347	0.377	0.783	0.761	0.775	0.655
8Q	0.364	0.375	0.363	0.396	0.722	0.691	0.727	0.588

Note: 1Q represents one-quarter ahead forecast, etc.. *MAE*–mean absolute error, *RMSE*–root mean square error, *CER*–classification error rate, *LPS*–log probability score, *AUC*–area under the curve. The cutoff value is the threshold used to classify probability forecasts of a recession as a positive event. The unconditional recession probability is 14.9% in the sample.