The Yield Curve as a Leading Indicator - FAQ

What does the evidence say, in short?

The difference between long-term and short-term interest rates ("the slope of the yield curve" or "the term spread") has borne a consistent negative relationship with subsequent real economic activity in the United States, with a lead time of about four to six quarters. The measures of the yield curve most frequently employed are based on differences between interest rates on Treasury securities of contrasting maturities, for instance, ten years minus three months. The measures of real activity for which predictive power has been found include GNP and GDP growth, growth in consumption, investment and industrial production, and economic recessions as dated by the National Bureau of Economic Research (NBER). The specific accuracy of these predictions depends on the particular measures employed, as well as on the estimation and prediction periods. However, the results are generally statistically significant and compare favorably with other variables employed as leading indicators. For instance, models that predict real GDP growth or recessions tend to explain 30 percent or more of the variation in the measure of real activity. See Estrella and Hardouvelis (1991). The yield curve has predicted essentially every U.S. recession since 1950 with only one "false" signal, which preceded the credit crunch and slowdown in production in 1967. There is also evidence that the predictive relationships exist in other countries, notably Germany, Canada, and the United Kingdom. See Estrella and Mishkin (1997) and Bernard and Gerlach (1998).

How and when were the relationships first identified?

Analysis of the behavior of interest rates of different maturities over the business cycle goes back at least to Mitchell (1913). However, Kessel (1965) may have been the first to make specific reference, verbal if not quantitative, to the behavior of term spreads. He pointed out that various spreads between long- and short-term rates tend to be low at the start of recessions - the business cycle peak - and high as expansions get under way after a cyclical trough. Butler (1978) made a connection between the yield curve as a predictor of short-term interest rates and the implications of declining short-term rates for contemporaneous economic activity, foreshadowing some of the later logic. He correctly predicted that there would be no recession in 1979, though a year later the situation would have been quite different. The late 1980s saw a spurt in research on the yield curve as a leading indicator. Laurent (1988, 1989) used the term spread to predict GNP growth. Harvey (1988) developed a theoretical relationship between the real term spread and subsequent real consumption growth, and presented some confirming empirical evidence. Furlong (1989) noted some predictive power for recessions, but expressed skepticism about the yield curve's reliability as a leading indicator at that time. Estrella and Hardouvelis (1989, 1990, 1991) showed empirically that the yield curve may be used to predict real growth in consumption, investment, and aggregate GNP, as well as NBER-dated recessions. Stock and Watson (1989) developed a sophisticated model of coincident and leading indicators, with the term spread forming part of the latter. Chen (1991) looked at the predictive relationship from the point of view of financial market investors.

How long have these relationships existed, and do they still hold?

Kessel (1965) presents graphical evidence that shows that the term spread tends to be negative at cyclical peaks, using data that go back as far as 1858. He finds some evidence of the leading indicator properties of the spread in time periods from 1914 to 1933 and from 1954 to 1961. Bordo and Haubrich (2004) provide regression-based statistical evidence using U.S. data from 1875 to 1997 and Baltzer and Kling (2005) perform a similar exercise with German data from 1870 to 2003. Both papers conclude that predictability varies over time and that it seems to be related to monetary policy credibility. Most other studies use data since the 1950s. The most recent data suggest that some of the statistical models from which forecasts based on the term structure are constructed have undergone structural changes, particularly with regard to parameter values. The predictive power may still be there, but the values of the parameters that record the sensitivity to movements in the term spread may be somewhat changed from earlier periods. However, the

simple rule of thumb that the difference between ten-year and three-month Treasury rates turns negative in advance of recessions is still reliable, with negative values observed before both the 1990-1991 and 2001 recessions.

Are formal models needed to extract the information content in the yield curve, or are there also rules of thumb?

Much of the research on this topic has been based on formal statistical models, such as linear regressions and non-linear statistical equations. These models are useful in that they provide quantitative guidelines about the sensitivity of output growth to changes in the term spread and about the precise lead-lag relationships exhibited by the data series, as well as recession probabilities. Nevertheless, simple rules of thumb are available, such as the fact that yield curve inversions (negative term spreads) are followed by recessions. Estrella and Mishkin (1996b) and Estrella (2005b) show that this approach may be made a bit more precise, though still simple, by presenting the relationship in tabular or graphical form.

What statistical models have been formulated?

Much of the analysis in the literature has focused on "continuous variables," data series whose values can in principle be any real numbers. Examples of these series are growth rates in GNP, GDP, industrial production, consumption, and investment. To predict these series, analysts have generally relied on relatively standard regression equations, taking care to deal with some important statistical issues. The most common issue results from the presence of overlapping observations in many of the applications. For example, suppose that we are interested in forecasting GDP growth over the next year using quarterly datThen, two consecutive observations of the variable being forecasted correspond to time periods that have three quarters in common. Standard measures of statistical significance are in general inconsistent, and must be adjusted by using, for instance, the generalized method of moments estimators of Hansen (1982) or Newey and West (1987). When the objective is to predict recessions, the methodology used is typically a probit or logit equation, in which the forecasted variable only assumes the values one and zero (either the economy is in a recession or it is not). Note, however, that special techniques are also required in these models if the forecast horizon produces overlapping observations. See Estrella and Rodrigues (1998).

Which interest rates to use: Treasuries, fed funds, Eurodollar, swap, corporate?

Research on the United States business cycle has relied mostly on interest rates for U.S. Treasury securities. One reason is convenience: data for many maturities are available continuously from the 1950s to the present in a consistent format. Another reason is that the pricing of these securities is not subject to significant credit risk premiums that, at least in principle, may change with maturity and over time. For similar reasons, studies of other countries tend to use data on national government debt securities. Rates on coupon bonds and notes are most easily accessible, but researchers in many countries have also produced zero-coupon rates, which may be directly matched with the timing of the forecasts. Some analysts have also used, as short-term rates, the U.S. federal funds rate and rates closely controlled by central banks in other countries. These are useful for some purposes, but the control exercised by the central bank implies that these rates are not fully reflective of the expectations of financial market participants. The spread between ten-year Treasuries and fed funds is a very accurate predictor of U.S. recessions during some time periods, but less so in others. At certain times, concerns have been voiced that special factors, such as the federal budget surpluses of 1999-2000, might affect the supply or demand of Treasury securities and distort the signals incorporated in their prices. Attempts have been made to use other interest rates for which a yield curve may be constructed, like Eurodollar, swap or corporate rates. An important drawback to the use of such data, however, is their limited availability with regard to the length of historical time series, the number of points along the yield curve, or both.

What maturity combinations work best?

When the yield curve is used to predict inflation (e.g., as in Mishkin (1990a, 1990b), interest rate maturities are matched precisely with the forecast horizons for inflation. For instance, to predict the difference between inflation expected in the next five years and inflation expected in the next year, the difference between five-year and one-year Treasury yields is used. When forecasting real activity, in contrast, the best results are obtained empirically by taking the difference between two Treasury yields whose maturities are far apart. At the long end, the clear choice seems to be a ten-year rate, which is the longest maturity available in most countries on a consistent basis over a long sample period. At the short end, there is a wider variety of choices. An overnight rate, such as the fed funds rate, is close to the extreme of the maturity spectrum. However, its usefulness as an indicator of market expectations is confounded by its fairly direct control by the Federal Reserve. A common choice currently is the two-year Treasury rate, perhaps because of the liquidity of the associated instruments. Background research in connection with Estrella and Mishkin (1998) suggested that the three-month Treasury rate, when used in combination with the ten-year Treasury rate to predict U.S. recessions, provides a reasonable combination of accuracy and robustness over long time periods. In the end, most term spreads are highly correlated and provide similar information about the real economy, so the particular choices with regard to maturity amount mainly to fine tuning and not to reversal of results. The caveat is that a benchmark that works for one spread may not work for another. For instance, the ten-year minus two-year spread may invert earlier than the ten-year minus three-month spread, which tends to be larger.

Is it the level or the change in the spread that matters?

With many leading economic indicators, either individual variables or indices, analysts focus on the change or growth rate in the variable as a forecaster of future real economic conditions. In contrast, it is the level of the term spread - not the change - that helps forecast both recessions and changes in real economic activity. For recessions, it is clearly the level that matters. In a probit model of the probability of recession, a given change in the spread can have a very different impact, depending on the initial level of the spread. When the curve is very steep, say the spread is above 300 basis points, a change of 50 basis points in the spread hardly changes the probability of recession. However, if the spread starts out at 50 basis points, a decrease of that magnitude may raise the implied probability by 10 percentage points or more. Theoretical explanations of these empirical results are not easily formulated. A suggestive heuristic argument is that the term spread, being a difference between interest rates of different maturities, incorporates an element of expected changes in rates and is thus indicative of future changes in real activity. In 1996, the Conference Board added the yield curve spread to its index of leading indicators, focusing on monthly changes in the spread. Note, however, that it announced in June 2005 that it would adjust its procedures so as to focus on the level of the spread and not on the change.

Does it matter if changes are driven by the short or the long end?

The best forecast of future real activity is provided by the level of the term spread, not the change in the spread, nor even the source of the change in the spread. Thus, if a low or negative value of the spread is reached via an increase in the short-term rate or a decrease in the long-term rate, it is only the low level that matters. In the six months preceding the trough of each yield curve inversion in the United States since 1960, we see a decline in the ten-year Treasury rate in two of seven cases (before the 1990-1991 and 2001 recessions) and an increase in the other cases. The direction of the change in the ten-year rate at the time of the signal does not appear to be indicative of the strength or duration of the subsequent recession. It is clear, however, that each recessionary episode is preceded (with varying lead times) by a substantial increase in the short-term rate.

Is an inversion required for a signal?

Although economic theory suggests that the yield curve should help forecast real output, no theory establishes a clear connection specifically between yield curve inversions and recessions. However, since 1960, a yield curve inversion (as measured by the difference between ten-year and three-month Treasury rates) has preceded every recession on record. In fact, in terms of monthly averages, the ten-year rate was at least 12 basis points below the three-month rate before every recession in that period. In contrast, very low positive levels of the spread have been observed without a subsequent recession. Specifically, there were two episodes in the 1990s in which the term spread attained very low positive levels (42 and 12 basis points respectively), but did not invert. In both of those cases, economic activity continued unabated after the troughs or low points for the spread. Thus, using inversion as a benchmark, there were no "false positives" during the period. While inversions and recessions may not be inevitably connected by theory, they correspond to extreme values of the term spread and output growth, respectively, which are in fact theoretically linked.

Does the signal have to be persistent?

Daily or even intraday changes in the term spread can be substantial. For example, for the spread between ten-year and 3-month rates, one-day changes of over 25 basis points occur about 2½ percent of the time. In some cases, these changes may be driven by market expectations of economic fundamentals and consequently may be persistent. In many instances, though, high-frequency changes in the spread may result from temporary demand or supply imbalances in the markets for Treasury securities, which may be quickly reversed and thus may not be truly reflective of changes in expectations about real economic conditions. One way to distinguish between perceived changes in fundamentals and temporary market phenomena is to trace the persistence of yield curve signals. A signal that lasts only one day may be dismissed, but a signal that persists for a month or more should be looked at carefully. Statistically, these distinctions may be captured by using data averaged over one month or more, which is quite common in the literature, or by including lagged effects in the model, as in Chauvet and Potter (2005).

Is the evidence robust over time?

An informal way to assess the robustness of yield curve forecasts of real activity is to examine visually the ex post accuracy of the results. In some cases, as with the regularity that yield curve inversions precede recessions, the evidence is immediate and quite consistent, as in the United States since 1960. The only apparent miss was in 1967, when the economy experienced a "credit crunch" that the NBER did not classify as a recession, despite a marked decline in industrial production. In other cases, such as with forecasts of real GDP growth, it is not altogether obvious whether the forecasting performance is consistent in different subperiods. Thus, it is helpful to apply formal statistical tests of model stability to the forecasting equations. Stock and Watson (2003) find instability in a large proportion of the models that are frequently used to forecast output growth. Their results stress the need to look for changes in parameter values that may result, for instance, from changes in monetary policy regime. Estrella, Rodrigues and Schich (2003) find modest evidence of instability in models used to forecast growth in industrial production, which suggests a similar caveat. However, the latter find no evidence of instability in models used to forecast recessions in Germany and the United States, suggesting that models with qualitative dependent variables may be more robust to changes in policy or other economic conditions.

How do binary models that predict recessions compare with models that forecast continuous dependent variables (e.g., real GDP, industrial production growth)?

These two types of models may be compared in two dimensions: accuracy and robustness. There is evidence that the most accurate binary models perform about as well as the most accurate continuous models. For example, Estrella and Hardouvelis (1991) and Estrella and Mishkin (1997) find that the R-squared in models of cumulative GNP growth is similar in magnitude to the pseudo R-squared in probit models of recession. The sample period may influence the results

somewhat, particularly in the case of the continuous models. Stock and Watson (2003) find substantial evidence of instability over time in various models of output growth, and Estrella, Rodrigues and Schich (2003) find some instability in models of industrial production growth. However, the latter paper suggests that binary models of recessions are quite robust over time, both in Germany and in the United States.

How does the yield curve compare with other indicators?

In general, the yield curve tends to perform quite well in comparisons with other leading indicators, including the traditional leading indexes and their components, and other variables with potential predictive power. Indicators such as stock prices and interest rates may have similar performance to the yield curve at some horizons, but none seem to dominate the yield curve as a predictor. For instance, Dueker (1997) and Dotsey (1998) compare the yield curve with a few other variables as a leading indicator of recessions, and find generally supportive statistical evidence. Estrella and Mishkin (1998) compare the term structure as a predictor of recessions with a large number of alternative indicators, and find that it is among the best in tests of statistical significance, particularly at horizons of about one year. Stock and Watson (2003) examine a large number of competing indicators in forecasts of output growth and find that all of them fall short of ideal properties, but that within these limitations the term structure "comes closest" to achieving those goals.

How does the yield curve perform out of sample, and can it be supplemented with other indicators?

The yield curve performs quite well in out of sample tests of predictive accuracy, and it is not clear that, in general, supplementary information can improve its predictive performance. Estrella and Mishkin (1998) find that, at predictive horizons beyond one quarter, there is no match for the term structure as a predictor of recessions. Not only do a large number of other indicators fall short, but predictions deteriorate as those other indicators are added to the term spread in out of sample forecasting exercises. Stock and Watson (2003) similarly examine a large number of competing indicators, but focus on forecasts of output growth rather than recessions. They find that the term spread works best, but that it exhibits some instability. The nature of this instability seems sufficiently idiosyncratic that combining the term spread with some other indicators may improve performance in their equations.

How are predictions related to monetary policy?

Although there are different views of the instruments and channels of monetary policy, a tightening of monetary policy usually means a rise in short-term interest rates, typically intended in the end to lead to a reduction in inflationary pressures. When those pressures subside, it is expected that a policy easing will follow. Expected future short-term rates are important determinants of current long-term rates. Thus, long rates tend to respond to a monetary tightening by increasing, though given that a policy reversal is expected, they tend not to increase by as much as short-term rates. Thus, a simple explanation of the predictive power of the yield curve for future output growth is that a monetary tightening both slows down the economy and flattens (or even inverts) the yield curve. Monetary policy is therefore an important determinant of the predictive power of the yield curve. However, given that private-sector expectations are incorporated in interest rates, and given that those expectations are based on some concept of the structure of the economy, monetary policy is not likely to be the single determinant of the predictive power. Formal models with these properties are presented in Eijffinger, Schaling and Verhagen (2000), Hardouvelis and Malliaropulos (2004), and Estrella (2005a).

How are predictions related to market expectations of the economy?

One of the most pervasive theories of the determinants of the yield curve is the expectations hypothesis, which posits that long-term interest rates are averages of expected future short-term rates. In the "pure" version of the hypothesis, this is the whole story. That version has been repeatedly rejected in econometric tests [see Campbell (1995)], so it is likely that there are other

important determinants as well, such as risk or liquidity premiums. Nevertheless, it is difficult to dispute that interest rate expectations play an important role, and that those expectations are related to future real demand for credit and to future inflation. A rise in short-term interest rates induced by monetary policy may lead to a future slowdown in real economic activity and demand for credit, putting downward pressure on future real interest rates. At the same time, slowing activity may result in lower expected inflation. By the expectations hypothesis, these expected declines in future short-term rates would tend to reduce current long-term rates and flatten the yield curve. Clearly, this scenario is consistent with the observed correlation between the yield curve and recessions. A formal model with these properties is presented in Estrella (2005a), and alternative specifications are considered in Eijffinger, Schaling and Verhagen (2000), Rendu de Lint and Stolin (2003), Malliaropulos (2003) and Hardouvelis and Malliaropulos (2004).

Is there causality from economic activity to the yield curve?

Most of the literature in this field deals with the yield curve as a predictor of future activity, but in principle there could be influences in the opposite direction, from economic activity to the yield curve. In a dynamic theoretical model with rational expectations, such as Estrella (2005a), both directions play a role. The term spread contains expectations of future activity, and it is affected by current monetary policy, which is influenced in turn by current economic activity. Empirically, Estrella and Hardouvelis (1990) use U.S. data to examine the effect of monetary policy on the yield curve, and Estrella and Mishkin (1997) perform a similar analysis for a panel of European economies. Evans and Marshall (2001) find consistent evidence that monetary policy shocks affect the nominal yield curve. In the context of a vector autoregression, Diebold, Rudebusch and Aruoba (2004) find that the influence in the direction from activity to the term structure is even stronger than the predictive relationships (though Stock and Watson (2003) warn about overinterpretation of such results). In general, theory and evidence are both supportive of a bidirectional relationship.

Should we expect the predictive power of the term spread for real activity to persist?

Accumulated experience with the forecasting power of the yield curve suggests that it is much more than a passing phenomenon. Warnings of its actual or possible demise are often voiced, as in Butler (1978), Furlong (1989), Watson (1991) and - to some extent - Dotsey (1998), but the fact remains that recessions still seem to follow inversions quite inevitably, as recently as in 2000-2001. Like many empirical models, some formal predictive models that forecast output growth based on the term spread seem to have a structural break around 1979-1980. Stock and Watson (2003) find substantial evidence of a break for models that predict output growth and Estrella, Rodrigues and Schich (2003) find more modest evidence for models that predict industrial production. However, this evidence does not necessarily imply that the predictive power of the yield curve has disappeared altogether, only that the values of the parameters in the formal models may have changed. Models of a more qualitative nature, such as those that predict recessions, seem to be affected much less or not at all, as documented by Estrella, Rodrigues and Schich (2003). Theory suggests (e.g., Estrella (2005a)) that there is a persistent predictive relationship between term spreads and future real output, though the precise parameters may change over time. Since yield curve inversions and economic recessions correspond to extreme values of those variables, a connection between inversions and recessions may be systematically detectable even if parameters change over time within reasonable bounds. Thus, although yield curve inversions may not be followed by recessions as a matter of universal mathematical principle, they should definitely raise warning flags about future output growth.