A NEW INDEX OF THE BUSINESS CYCLE

FIRST VERSION: January 15, 2020

THIS VERSION: August 5, 2020

William Kinlaw, Mark Kritzman, and David Turkington

William Kinlaw is a senior managing director at State Street Associates in Cambridge, MA.

wbkinlaw@statestreet.com

Mark Kritzman is the CEO of Windham Capital Management in Boston, MA, and a senior lecturer at MIT Sloan School of Management in Cambridge, MA.

kritzman@mit.edu

David Turkington is a senior managing director at State Street Associates in Cambridge, MA.

dturkington@statestreet.com

Abstract

The authors introduce a new index of the business cycle that uses the Mahalanobis distance to measure the statistical similarity of current economic conditions to past episodes of recession and robust growth. Their approach has a key advantage compared to approaches that simply aggregate data, such as the Conference Board indexes, or approaches that rely on regression models. It considers the distribution of recession data separately from the distribution of growth data. This feature, along with the construction of the index as a relative probability, has the consequence of shifting the weights that are placed on the index inputs based on their prevailing values. In addition, their framework makes it possible to measure how the relative importance of the economic variables from which the index is constructed vary through time, which yields valuable insights about the dynamics of the business cycle.

A NEW INDEX OF THE BUSINESS CYCLE

1   Introduction

We introduce a new index of the business cycle that uses the Mahalanobis distance to measure the relative similarity of current economic conditions to past episodes of recession and robust growth. Conventional approaches to business cycle forecasting, such as the Conference board indexes, rely on simple aggregations of economic variables. Other approaches use regression models to forecast business cycle outcomes based on samples that mix recession data and growth data with all other data. Our approach distinguishes between standard deviations and correlations of economic variables that occur during recessions and those that occur during periods of growth, which allows the weights our approach places on inputs to vary with prevailing economic conditions. We believe this distinguishing feature explains why our index, which we call the KKT Index, identifies and predicts recessions more reliably than comparable indexes. Our methodology also enables us to measure the time varying importance of the variables we use to construct the index, which provides valuable intuition about the dynamics of the business cycle.

We proceed as follows. In Section 2, we describe the Mahalanobis distance as it was originally conceived by Mahalanobis to analyze human skulls, and we discuss more recent applications of this versatile statistic. We also include a technical overview of the Mahalanobis distance. In Section 3, we explain how we adapt the Mahalanobis distance to create an index of the business cycle. In Section 4, we describe in detail how our approach differs from other

approaches. We also describe how to measure the relative importance of the economic variables through time.   In Section 5 we offer empirical evidence of our index's effectiveness in assessing the state of the economy. We summarize our analysis in Section 6. We also include two appendices: one which describes the methodology used by the National Bureau of Economic Research (NBER) to measure the business cycle, and one which describes the methodology used by the Conference Board to construct their economic indexes.

2   The Mahalanobis Distance

The Mahalanobis distance was introduced originally in 1927 and modified in 1936 to analyze resemblances in human skulls among castes in India.[1] Mahalanobis compared a set of measurements for a chosen skull to the average of those measurements across skulls within a given caste. He also compared the co-occurrence of those measurements for a chosen skull to their covariation within the caste. He summarized these comparisons in a single number which he used to place a given skull in one caste versus another caste. The Mahalanobis distance has since been applied across many different fields. Chow, Jacquier, Kritzman, and Lowry (1999), for example, derived the Mahalanobis distance independently to measure turbulence in the financial markets. They compared a set of asset class returns for a given interval to their averages and covariances over a prior history to measure the statistical unusualness of that set of returns as an indication of financial turbulence. They reasoned that the more unusual were the returns the more likely it was that they were driven by disruptive events instead of noise, and therefore more characteristic of financial turbulence.

3

The Mahalanobis distance has also been applied in medicine to diagnose diseases. Su and Li (2002), for example, applied the Mahalanobis distance to diagnose liver diseases. Wang, Su, Chen, and Chen (2011) used the Mahalanobis distance to diagnose obstructive sleep apnea, and Nasief, Rosado-Mendez, Zagzebshi, and Hall (2019) used it to diagnose breast cancer. These are but a few of its applications to medicine.

The Mahalanobis distance has also been applied to detect anomalies in self-driving vehicles (Lin, Khalastchi and Kaminka, 2010), and to improve the forecast reliability of linear regression analysis (Czasonis, Kritzman, and Turkington, 2020). In this latter application the authors show that the prediction of a linear regression equation is mathematically equivalent to a weighted average of the past values of the independent variables in which the weights are the relevance of the observations as defined by the sum of two Mahalanobis distances. This equivalence allows a researcher to censor a sample to exclude insufficiently relevant observations in order to derive a more reliable forecast. This innovation has also been applied to improve the forecast of the stock-bond correlation (Czasonis, Kritzman, and Turkington, 2020). Before we proceed to our application of the Mahalanobis distance, it might be useful to provide a technical overview within the context of its original purpose.

The Mahalanobis distance, as originally conceived to measure the statistical similarity of human skulls, is given by Equation 1.

$$d = (x - \mu)\Sigma^{-1}(x - \mu)' \qquad (1)$$

4

In Equation 1, d equals the Mahalanobis distance, x equals a row vector of values for a set of dimensions used to characterize a skull, μ equals the average values from a chosen group of skulls, and Σ^{-1} equals the inverse of the covariance matrix of the group's dimensions, and $'$ denotes matrix transpose. The term $(x - \mu)$ captures how similar each dimension, by itself, is to the group's average values. By multiplying $(x - \mu)$ by the inverse of the covariance matrix, it captures how similar the co-occurrence of the dimensions is to their co-occurrence in the group. This multiplication also converts the variables into common units. This feature is not important in analyzing human skulls because all variables are measured in the same units, centimeters. When we apply this formula to create an index of the business cycle, however, this feature is very handy because some of the variables we use are measured as percentage changes whereas others are measured as levels.

The Mahalanobis distance is a powerful analytical tool because it summarizes information about multivariate covariation in a single number that has an intuitive interpretation. To develop an understanding of the Mahalanobis distance, it may be helpful to consider the following. If we apply Equation 1 to a single (scalar) variable instead of a vector, the equation becomes a ratio with the squared distance from the mean, $(x - \mu)^2$, in the numerator and the squared standard deviation, σ^2, in the denominator. Taken together, this is equivalent to the squared z-score representing how many standard deviations away from average a data point lies, where $z = \frac{(x-\mu)}{\sigma}$. Thus, we may interpret the Mahalanobis distance for multiple dimensions as a generalization of the z-score that accounts for the correlations among the variables.[2]

Exhibit 1 illustrates the features of the Mahalanobis distance in two dimensions using hypothetical data. Let us imagine the dots represent values of the two dimensions for various skulls. The cluster of observations on the left-hand side of the chart pertain to hypothetical skull measurements for a group where the attributes are perfectly uncorrelated and have equal standard deviations. When the correlation among the variables equals zero, the Mahalanobis distance of a given point reflects the average of its squared z-scores in the same fashion as the most common measurement of physical distance: the Euclidean distance. The two data points shown, A and B, have identical Euclidean distances from this group's average, and because the variables are uncorrelated these points also have identical Mahalanobis distances from the mean. Other points that fall on the same iso-distance curve, shown as a dotted line circle, also share the same distance value. The cluster of observations on the right-hand side of the chart are from a different sample with different characteristics; the attributes are positively correlated and have different standard deviations. Points C and D have identical Euclidean distances from their average for the group, but their Mahalanobis distances are not the same. Point C conforms to the typical correlation pattern of the data, so it is less unusual and less distant from the center of the distribution. Point D reflects an opposite alignment of the two variables, which is highly unusual. Thus, points C and D fall on different iso-distance curves. The iso-distance curves for the Mahalanobis distance are rotated ellipses rather than circles, owing to their non-zero correlation. When there are more than two variables, the iso-distance curve is an ellipsoid within a higher-dimensional space. This is difficult to visualize, but the same intuitive interpretation applies in any number of dimensions. If desired, one may compute the probability of observing a Mahalanobis distance greater than some chosen value from a

collection of multivariate normal data as the area under the right-tail of a chi-squared

distribution. This statistic is often used in outlier detection and statistical testing. For our

application, as we describe later, we are more interested in the relative likelihood that an

observation corresponds to one sub-sample or another.

Exhibit 1: Scatter Plot of two Hypothetical Skull Dimensions
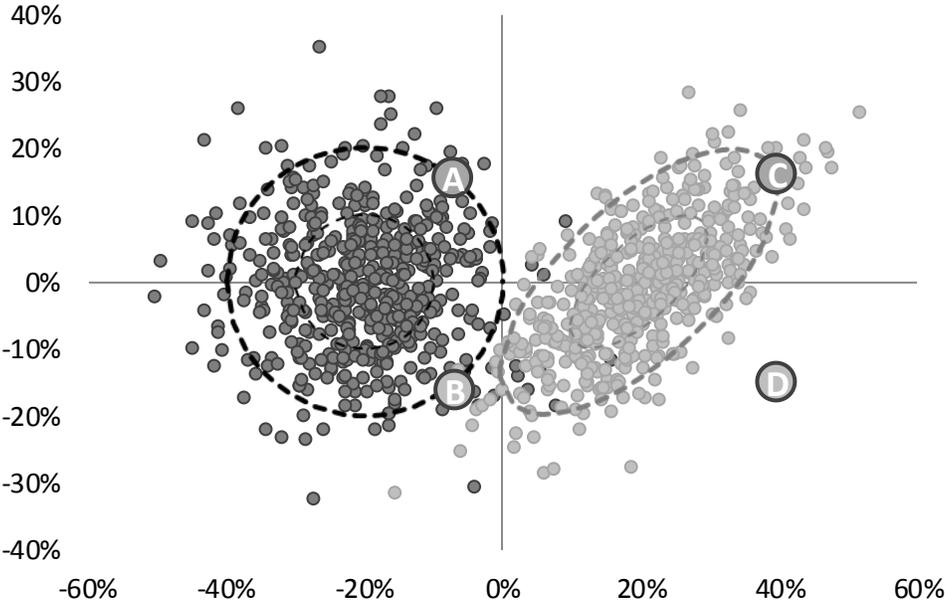
In summary, the Mahalanobis distance accounts for a couple of important features of

statistical similarity. It scales each value of the chosen observation by the variability of the

values in the group, which has the added benefit of converting all values into common units.

And it accounts for the co-occurrence of the values for a given observation.

The Mahalanobis distance is very general and may be computed for any data point so

long as the covariance matrix used in the calculation is valid. By valid, we mean that the

covariance matrix is positive semi-definite, which guarantees that it is possible for actual data to exhibit these correlation patterns. This is guaranteed to be the case if the data on which we estimate covariance contains more observations than variables. Nevertheless, we should also take care to select variables that are not highly redundant. If two variables have correlations very close to 1, the Mahalanobis distance will be highly sensitive to divergences and may occasionally emit misleading signals. We address these considerations in the design of our index by choosing variables that are conceptually distinct from one another, but which may behave differently during alternate phases of the business cycle.

3   An Index of the Business Cycle

We now apply the Mahalanobis distance to create an index of the business cycle which, as mentioned earlier, we call the KKT Index. In this application, the observations are the values of a set of economic variables, and the groups are the values of these variables during periods of recession or robust growth. We construct our index using the following economic variables for the United States starting in January 1916 and ending in June 2020:[3]

- Industrial Production (one-year percentage change, measured monthly)

- Nonfarm Payrolls (one-year percentage change, measured monthly)

- Return of the Stock Market (one-year return, measured monthly)

- Slope of the Yield Curve (10-year rate minus the Federal Funds Rate)

    We begin our index in January 1956, the earliest date for which we have vintage data to reflect the values for economic variables that were available at each point in time prior to any

revisions. For each month starting in January 1956, we only use data that were available at that point in time, including any revisions previously released. To calculate the index value, we proceed as follows. First, we identify two sub-samples of previous months: one corresponds to recession and the other corresponds to robust growth. We define recessions as those periods identified as recessions by the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER). We define robust growth as months in which the year-over-year percentage change in industrial production ranked above the 75th percentile relative to its values in the prior 10 years. This positive tail of the growth distribution contains roughly as many data points as the recession sub-sample. We focus on robust growth rather than growth of any size because it is important that the regimes be symmetrically opposite each other and sufficiently separated from each other.

Next, we measure the Mahalanobis distance of the current observations for the economic variables, x, relative to both sub-samples:

$$d_{rec}(x) = (x - \mu_{rec})\Sigma_{rec}^{-1}(x - \mu_{rec})' \tag{2}$$

$$d_{gr}(x) = (x - \mu_{gr})\Sigma_{gr}^{-1}(x - \mu_{gr})' \tag{3}$$

In Equations 2 and 3, d_{rec} is the Mahalanobis distance of x with respect to the recession sample, μ_{rec} is a vector of the average values of the variables in the recession sub-sample, Σ_{rec}^{-1} is the inverse of the covariance matrix of the variables in the recession sub-sample, and d_{gr},

μ_{gr}, and Σ_{gr}^{-1} are the corresponding values for the robust growth sub-sample.[4] Whereas

Mahalanobis sought to determine if a set of dimensions for a skull was more plausibly

associated with one caste versus another, we seek to determine if the values for a set of

economic variables are more closely associated with the values that prevailed during past

recessions or past periods of robust growth.

We convert both Mahalanobis distances, d, into likelihoods, ξ, using the multivariate

normal probability density function (PDF):

$$\xi_{rec}(d) = (\det(2\pi\Sigma_{rec}))^{-1/2}e^{-d_{rec}/2} \tag{4}$$

$$\xi_{gr}(d) = \left(\det(2\pi\Sigma_{gr})\right)^{-1/2}e^{-d_{gr}/2} \tag{5}$$

In Equations 4 and 5, det is the matrix determinant, and e is the base of the natural

logarithm. Next, we rescale the likelihood of recession by dividing it by the sum of the

recession and robust growth likelihoods. We interpret this rescaled likelihood of recession as a

probability.

$$p_{rec} = \frac{\xi_{rec}}{\xi_{rec}+\xi_{gr}} \tag{6}$$

We repeat this process for each month in our sample, ending in June 2020, to generate a monthly time series of values for the index.

4 Comparison to Other Approaches

The key distinction of the KKT Index from other approaches to analyzing the business cycle, such as regression models, is that the KKT Index considers the distribution of recession data separately from the distribution of growth data. This feature, along with the construction of the index as a relative probability, has the consequence of shifting the weights that are placed on the index inputs based on their prevailing values. This feature makes it possible to measure how the relative importance of the inputs vary through time, which yields valuable insights about the dynamics of the business cycle.

Here is how we measure the time varying importance of the inputs. We start with the notion that the KKT Index value is the rescaled probability of recession as a function of the vector of current economic variable values, x. Viewing this calculation as a nested set of functions, as illustrated in Equations 2 through 6, we use the chain rule of calculus to compute the derivative of the probability of recession with respect to x, which is given by Equation 7.

$$\frac{\partial p_{rec}}{\partial x} = p_{rec}p_{gr}\left(\Sigma_{gr}^{-1}(x - \mu_{gr})' - \Sigma_{rec}^{-1}(x - \mu_{rec})'\right) \qquad (7)$$

Equation 7 has an intuitive interpretation. The term $\Sigma_{gr}^{-1}(x - \mu_{gr})'$ shows how much the

Mahalanobis distance to the mean of the robust growth sample changes for a given change in

x. The recession probability is positively related to the distance from growth. The term

$\Sigma_{rec}^{-1}(x - \mu_{rec})'$ shows how much the Mahalanobis distance to the mean of the recession

sample changes for a given change in x. The recession probability is negatively related to the

distance from recession. The term $p_{rec}p_{gr}$ scales the overall sensitivity so that the rescaled

probability cannot fall below zero or rise above 100 percent. The overall sensitivity of the index

value is greatest when it is around 50 percent, and smallest when it is around zero or 100

percent. We also observe that as x approaches the average values of either recession or robust

growth (and the index value approaches 1 or 0 accordingly), the term corresponding to that

sub-sample approaches zero, and the index is increasingly driven by the distance of x to the

opposing sub-sample. This occurs because when the index is very close to the center of one of

the sub-samples, the information from that subsample is completely reflected in the index and

fluctuations are likely to represent noise. At this point, the distance to the opposing sub-

sample is more informative.

Next, we multiply each element of the derivative vector by the standard deviation for

that variable (measured over the full sample). The result represents the sensitivity of the

recession probability to a standardized shock in each variable. Finally, we rescale this vector, so

its absolute value sum equals one, and we interpret the result as a measure of the relative

importance of the variable.

$$Variable\ importance = \frac{\frac{\partial p_{rec}}{\partial x} \circ \sigma}{\left|\frac{\partial p_{rec}}{\partial x} \circ \sigma\right|} \tag{8}$$

In Equation 8, σ is a vector of the full-sample standard deviations of the variables, $\circ$

denotes the element-by-element product of two vectors that yields the same size vector result,

and $|\cdot|$ computes the absolute-value norm of a vector.

The time-varying importance of each variable in driving index changes is a key feature

that distinguishes our methodology from more conventional regression-based approaches. It

allows the index to treat data differently during different phases of the business cycle. This

stands in contrast to Probit and Logit models, which are two of the most common methods

used to forecast probabilities.

Both models form predictions by transforming the result of a linear model, $\hat{y}_{linear} = \beta x$, into a variable that falls between 0 and 1:

$$p_{probit}(\hat{y}_{linear}) = \Phi(\hat{y}_{linear}) \tag{9}$$

$$p_{logit}(\hat{y}_{linear}) = \frac{1}{(1+e^{-\hat{y}_{linear}})} \tag{10}$$

In Equation 9, $\Phi(\cdot)$ is the cumulative distribution function of a standardized normal

variable. Due to the nonlinearity of their final predictions, we cannot fit probit and logit models

algebraically; we must determine β by an iterative numerical process. The point we want to

stress is that relative variable importance is simply equal to β when applied to probit and logit

models. It does not vary with x, as it does in the KKT Index.[5] Traditional linear regression

models evaluate data points in the context of one overall distribution of data. Our

methodology evaluates data points in the context of two distinct distributions, where the differences in covariance between the distributions contributes information. Our approach is related to quadratic discriminant analysis, though the latter is typically used to assign data points to classes explicitly, as opposed to measuring their relative likelihood as an index. Following this discussion, we may view the KKT Index as an extension to, or variant of, regression-based approaches. We calibrate the index to identify the business cycle conditions that prevail in each month, which is akin to using a probit or logit model to forecast (or "now-cast" as it is sometimes called) whether the economy is currently in a recession. We find that the level and trend of such an index contains useful information for predicting future conditions, and it may be applied to predict outcomes over various future horizons, as desired. It is also possible to construct an alternative index using the KKT Index methodology, but with the explicit objective of predicting recession occurrence in a defined future period rather than in the contemporaneous month. This application may be a subject for future research.

5   Results

We now present evidence of the effectiveness of our index. Exhibit 2 presents a time series of the KKT Index of the Business Cycle (solid black line) from January 1956 to June 2020. This line measures how much more likely it is that the conditions at any point in time are associated with recession instead of with robust growth. The periods defined as recessions by the NBER are indicated by the shaded bars. For example, as of November 2019, the value of the KKT index was 76% (left axis), which means that considering either recession or robust growth, this set of

conditions is more closely associated with recession 76% of the time (and with robust growth

24% of the time). Put differently, the index suggested a recession was more than three times as

likely as robust growth. An index level of 76% does not necessarily mean that the economy is

currently in recession. Rather, we should interpret it as an indication of the potential for the

economy to enter recession in the foreseeable future. Given historical guidance, the index

should be close to 100% when a recession is imminent or underway. It spiked to 93% in March

2020 amidst the COVID-19 pandemic and subsequent economic crisis, and registers at 98% as

of June 2020.

The dashed line in Exhibit 2 shows the Conference Board's Coincident Economic Index.[6]

This time series begins in February 1979, and its values are indicated by the right axis. A value

of 0 indicates neutral economic conditions, whereas large negative values coincide with

recessions (we inverted the scale to coincide with the KKT Index). It is quite apparent that the

Coincident Index's values in 2020 lie far outside the range of its historical experience. The

Coincident Index has risen in tandem with the six NBER recessions since 1980, but it is

important to keep in mind that this result likely occurred because the Conference Board and

NBER follow nearly identical procedures to define and identify recessions. The NBER defines a

recession as "a period of falling economic activity spread across the economy, lasting more than

a few months, normally visible in real GDP, real income, employment, industrial production,

and wholesale-retail sales."[7] With the exception of GDP, these indicators match the four

variables that make up the Conference Board's Coincident Index. By using a different set of

variables that include market conditions in addition to economic growth and employment, the

KKT Index provides a complimentary and differentiated view on the business cycle.

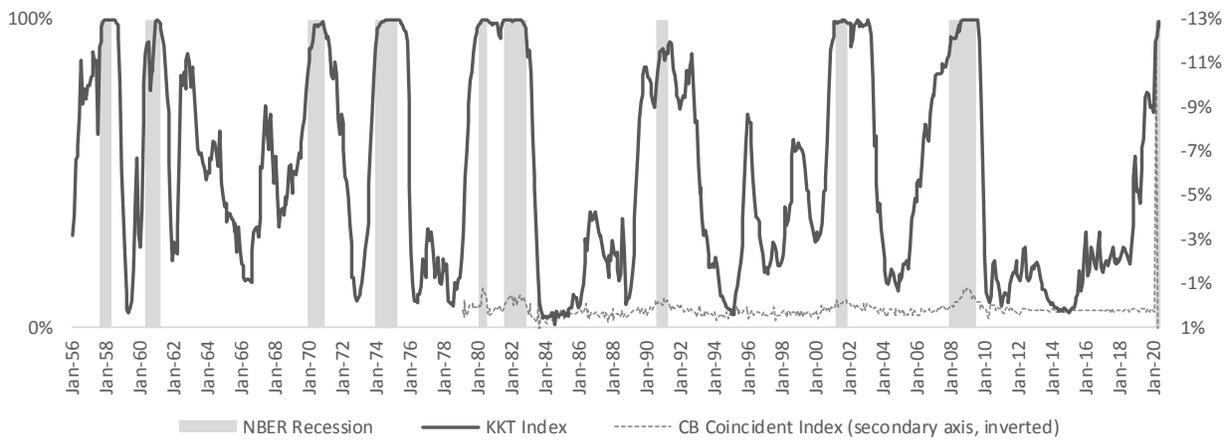Exhibit 2: KKT Index and Conference Board Coincident Index

Exhibit 3 shows the KKT Index alongside the Conference Board's Leading Economic

Index, which begins in January 1982.

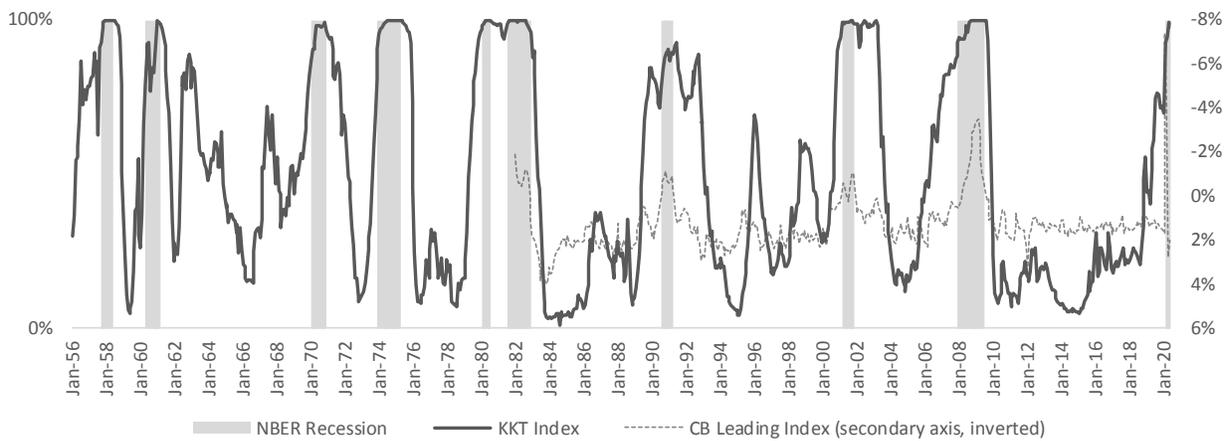Exhibit 3: KKT Index and Conference Board Leading Index

Exhibit 3 indicates that the major spikes in the Conference Board's Leading Economic Index tend to coincide with recessions rather than anticipate them. The KKT Index rises leading up to every recession so that the combination of its trajectory and level provides a reliable indicator of the likelihood recession. As of late 2019, the Conference Board's Leading Index remained flat while the KKT Index increased sharply.

Exhibit 4 compares the KKT Index to the yield curve, which we define as the 10-year rate minus the Federal Funds Rate. Many pundits believe that an inverted yield curve presages the onset of recession. Exhibit 4 tends to support this relationship, though the lead time is often quite long.

Exhibit 4: KKT Index and the Yield Curve
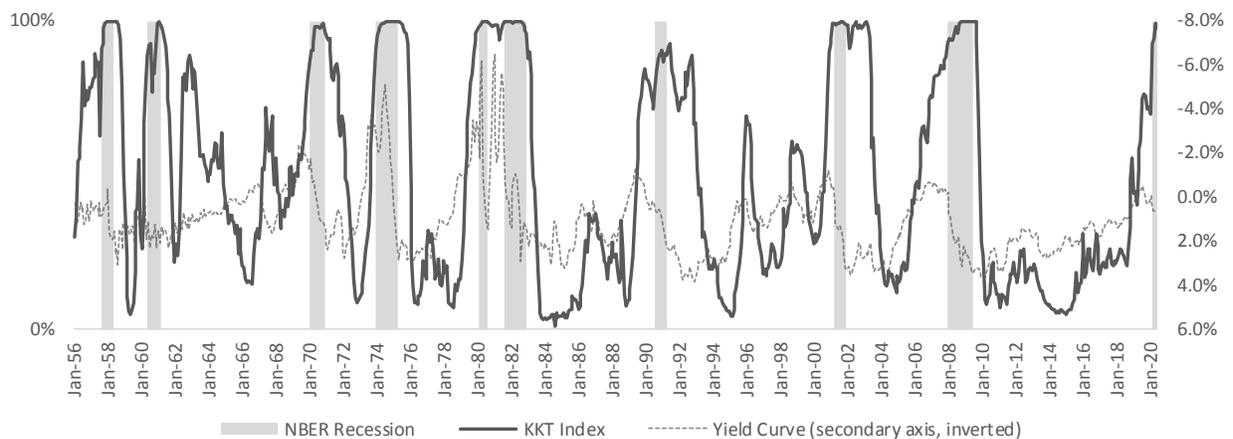
We next present an event study of the KKT Index of the Business Cycle. The shaded bar in Exhibit 5 represents the events which are either recessions or periods of robust growth that occurred since 1956. The width of the bar is not relevant. These events varied in duration. The

left side of the bar represents the beginning of the events while the right side represents the

end of the events, irrespective of their durations.

Exhibit 5: KKT Event Study
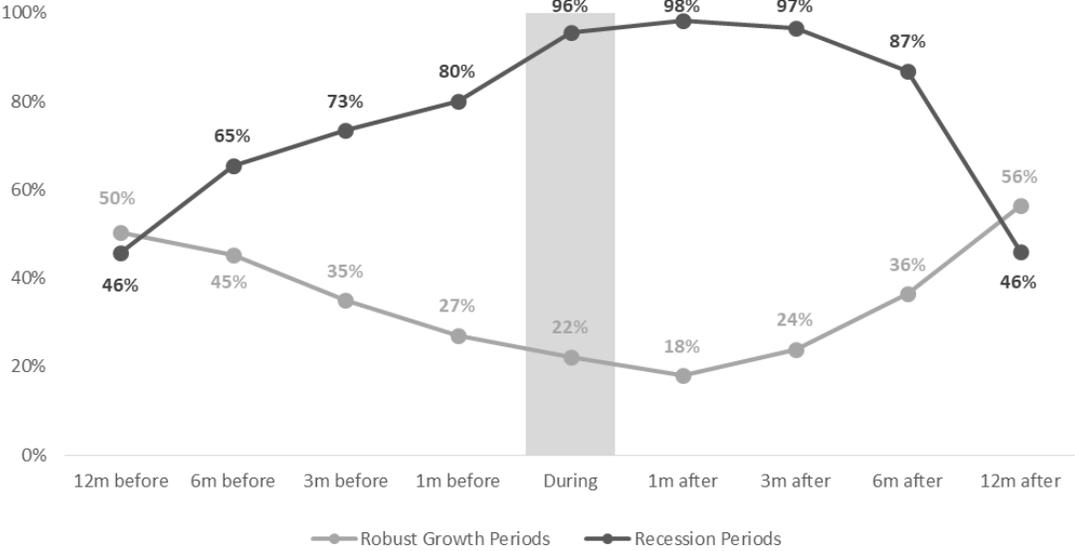
The dark line shows the level of the KKT Index leading up to, during, and following

recessions. The light line shows level of the index leading up to, during, and following periods

of robust growth. We define robust growth events based on annual changes in real GDP,

considering revisions.[8] Because our index is constructed as the relative likelihood of recessions,

we should expect it to be low during periods of robust growth, which it is.

Exhibit 5 reveals a stark separation in the level of the KKT Index depending on whether it

is measured around recession events or robust growth events. It also reveals that extreme

levels of the index tend to persist for a few months following the conclusion of the events,

which is a feature that businesspeople, policymakers, and investors should consider in their application of this index.

Exhibit 6 compares the level of the KKT Index to realizations of recessions within various time spans. We are interested in analyzing periods when the probability of recession is rising. Therefore, we require that the standardized shift of the index – defined as its current level minus its average over the past year, divided by its standard deviation over the past year – is greater than 1.

Exhibit 6: KKT Index and Recession Realizations

Rising and above threshold:	50%	60%	70%	80%	90%	Unconditional Frequency
This month	36%	43%	53%	62%	87%	14%
Next 1m	40%	48%	57%	66%	91%	14%
Next 3m	47%	55%	64%	73%	96%	16%
Next 6m	59%	66%	77%	82%	96%	20%
Next 12m	73%	78%	87%	89%	96%	28%
Next 18m	78%	82%	87%	89%	96%	34%

We report the unconditional frequency of recession for the various time spans. Note the row corresponding to the realization of recessions over the subsequent six months for various levels of the index. When the index exceeded 50%, 59% of the time a recession occurred within the next six months. When it exceeded 60%, a recession occurred 66% of the time within the next six months. When the index exceeded 70% the frequency of recessions was 77%. When it exceeded 80%, recessions occurred 82% of the time. And when it exceeded

90%, recessions followed 96% of the time. The correspondence between the index level and

the incidence of recessions is remarkably strong; in fact, the correlation exceeds 99%. To put

this in perspective, the unconditional likelihood of a recession within any six-month period is

only 20%.

We next present the same analysis for the yield curve. Specifically, we show the

incidence of recessions that occur over varying time spans once the yield curve becomes

inverted and its one-year standardized shift is below -1.[9] And again we show the unconditional

frequency of recession for these time spans.

Exhibit 7: Yield Curve and Recession Realizations

Falling and below threshold:	0%	Unconditional Frequency
This month	10%	14%
Next 1m	15%	14%
Next 3m	21%	16%
Next 6m	39%	20%
Next 12m	60%	28%
Next 18m	82%	34%

Exhibit 7 shows the yield curve to be a much less reliable indicator of subsequent

recessions than the KKT Index, especially for short horizons. It is only informative for a horizon

of 18 months, and even for that horizon, it is less reliable than the KKT Index.

Next, we present more intricate tests of the efficacy of the KKT Index. We apply a

framework known as "receiver operating characteristic" (ROC) curves. This approach evaluates

the tradeoff between erroneously predicting (false positives or type I errors) and failing to

predict (false negatives or type II errors) the occurrences of recession. It evaluates this tradeoff

for the full range of possible thresholds that may be set for an indicator, because effectiveness

may vary with the threshold, and it is not necessarily clear in advance what threshold to use.

By varying the threshold, an indicator's false positive rate may be improved but only at the

expense of its false negative rate, and vice versa. The most effective models will present more

favorable opportunities to jointly minimize both types of errors.

It is possible to apply many different methodologies and data inputs to the task of

predicting recessions. We do not attempt to evaluate our index directly against every other

proposed measure, as the permutations of methodologies and data are simply too numerous.[10]

Indeed, we welcome and encourage future research applying our methodology to other data

such as credit spreads and other asset prices.[11] Instead, we focus our attention on two specific

comparisons that are related to the benefits of our approach. First, we compare the KKT Index

to an index derived from a logit model using the same inputs.[12] This test reflects the advantage

of accounting for distinct covariances in the recession and growth sub-samples, versus

accounting only for the full-sample covariances of the variables. Second, we compare the KKT

Index to the Conference Board's Leading Economic Index, which is among the most widely-

followed recession indicators. This test reflects the advantage of combining data with the

Mahalanobis distance compared using a weighted average. It also reflects differences in input

variables. All tests are out-of-sample and pertain to the occurrence of recession in the 12
months following a signal.

Exhibit 8 shows both comparisons over the common historical sample available in each
case. The KKT Index achieves a tradeoff similar to logit for extreme thresholds (left panel), but
it achieves a superior tradeoff in the middle of the curve. Compared to the Conference Board's
Leading Economic Index since 1982, the KKT Index performs slightly worse when it is calibrated
to predict cautiously, but it performs better for moderate and aggressive prediction thresholds.
It is important to note that these tests may be overly simplistic because they only account for
the level of each index and not whether it is increasing or decreasing. Conditioning on the
index trend generally improves prediction accuracy, but it leads to more variants of these tests
than we have space to present in this paper.

Exhibit 8: False Positive and False Negative Rates for Recession Occurrence in Next 12 Months

KKT Index (dark) versus Logit (light) since 1956

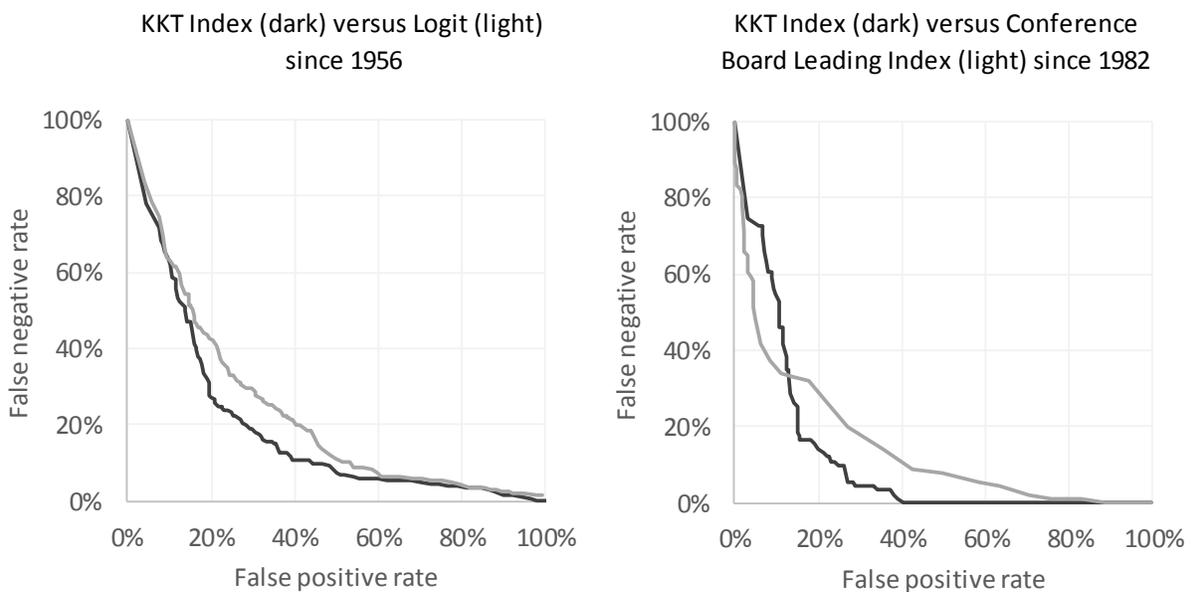KKT Index (dark) versus Conference Board Leading Index (light) since 1982

Exhibit 9 shows the relative importance of each variable through time. While all four variables play an important role overall, they are not all equally important at a given point in time. It appears that the yield curve and stock market are generally more important in the early stages of a recessionary episode, while industrial production and payrolls are more important in determining the depth of a recession and its subsequent recovery. We also observe interesting shifts across decades, as the yield curve and stock market have become relatively more influential after 2000, and industrial production has become less important. This methodology provides an intuitive yet statistically rigorous way to interpret the KKT Index. It may also be applied to evaluate the variables' relevance to current conditions.

Exhibit 9: Relative Importance of Variables Through Time
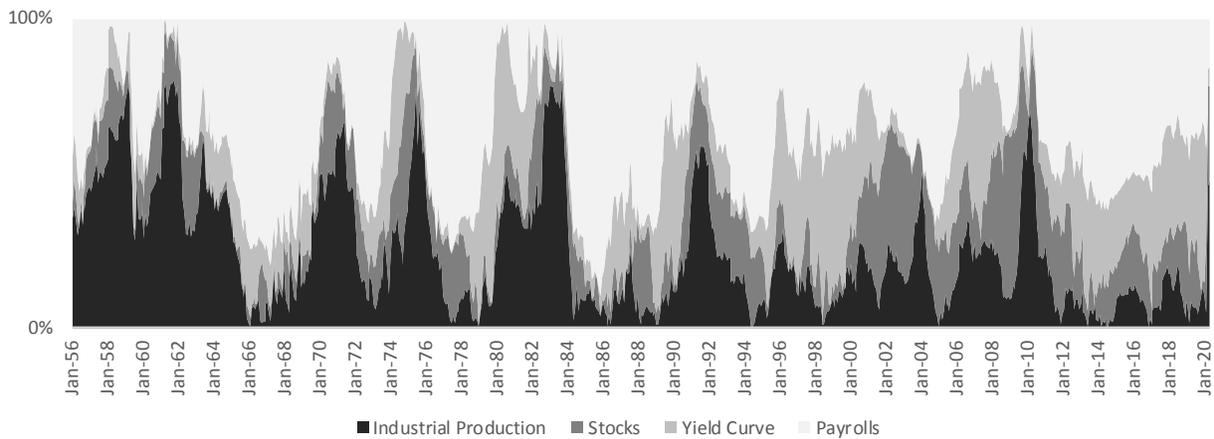
6   Conclusion

We apply the Mahalanobis distance to construct a new index of the business cycle. Specifically, we measure the statistical similarity of economic conditions each month to economic conditions that prevailed during prior periods of recession and robust growth. We then construct the index as the likelihood of recession relative to the likelihood of robust growth.

Unlike other approaches for forecasting the business cycle, which either rely on a simple aggregation of economic variables or use regression models, our approach distinguishes between the standard deviations and correlations of the economic variables that prevailed during recessions from those that prevailed during periods of growth. This separation allows the weights that our approach places on the inputs to vary with prevailing economic conditions. Our approach also allows us to measure changes in the relative importance of the economic variables, which yields valuable intuition about the dynamics of the business cycle.

We show that our index compares favorably to other commonly used measures of the business cycle, including the Conference Board Indexes, the yield curve, and results generated by logit models. Finally, we produce results showing the time varying importance of the economic variables that we use to construct our index, and we summarize key insights about the dynamics of the business cycle.

Appendix A   NBER Methodology

The National Bureau of Economic Research's (NBER's) Business Cycle Dating Committee was created in 1978 (though the NBER has been publishing business cycle dates since 1929). It consists of eight (or so) members. On its website and in the most recent statement following a recession, released in September 2010, the committee defines a recession as "a period of falling economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales."[13] (As a side note, it is important to acknowledge that the four indicators, other than GDP, map very closely to the four indicators used in the Conference Board's Coincident Economic Index, which we review in Appendix B.)

The September 2010 report also explains that: "The trough marks the end of the declining phase and the start of the rising phase of the business cycle. Economic activity is typically below normal in the early stages of an expansion, and it sometimes remains so well into the expansion."[14] The NBER's website (https://www.nber.org/cycles/recessions.html, accessed December 2019) provides further detail:

> The NBER's Business Cycle Dating Committee maintains a chronology of the U.S. business cycle. The chronology comprises alternating dates of peaks and troughs in economic activity. A recession is a period between a peak and a trough, and an expansion is a period between a trough and a peak. During a recession, a significant decline in economic activity spreads across the economy and can last from a few months to more than a year. Similarly, during an expansion, economic activity rises substantially, spreads across the economy, and usually lasts for several years.
>
> In both recessions and expansions, brief reversals in economic activity may occur – a recession may include a short period of expansion followed by further decline; an expansion may include a short period of contraction followed by further growth.

The Committee applies its judgment based on the above definitions of recessions and expansions and has no fixed rule to determine whether a contraction is only a short interruption of an expansion, or an expansion is only a short interruption of a contraction. The most recent example of such a judgment that was less than obvious was in 1980-1982, when the Committee determined that the contraction that began in 1981 was not a continuation of the one that began in 1980, but rather a separate full recession.

The Committee does not have a fixed definition of economic activity. It examines and compares the behavior of various measures of broad activity: real GDP measured on the product and income sides, economy-wide employment, and real income. The Committee also may consider indicators that do not cover the entire economy, such as real sales and the Federal Reserve's index of industrial production (IP). The Committee's use of these indicators in conjunction with the broad measures recognizes the issue of double-counting of sectors included in both those indicators and the broad measures. Still, a well-defined peak or trough in real sales or IP might help to determine the overall peak or trough dates, particularly if the economy-wide indicators are in conflict or do not have well-defined peaks or troughs.

National Bureau of Economic Research
(https://www.nber.org/cycles/recessions.html)

Appendix B   Conference Board Index Methodology

The Conference Board publishes monthly business cycle Indicators, including composite indexes

for leading, coincident, and lagging economic activity. Their Business Cycle Indicators

Handbook describes the design and historical evolution of the indexes as follows:

> … In 1961, under the direction of Julius Shiskin at the Bureau of the Census, the U.S. Government began publication of a monthly report, Business Cycle Developments (BCD). This work was undertaken in cooperation with the NBER and the President's Council of Economic Advisers and made extensive use of time-series charts of NBER indicators (80 U.S. series and indexes of industrial production for seven major trading partners). In 1968, the report was renamed Business Conditions Digest, and

in 1972 the indicators were shifted to another Commerce Department agency, the Bureau of Economic Analysis… (p. 9)

… In 1995, the BEA […] transferred its program of research and production of business cycle indicators to The Conference Board… (p. 10)

… Wesley C. Mitchell and Arthur F. Burns originated the indicator approach that made extensive use of business cycle indicators in the mid-1930s at the NBER… Over subsequent decades, the approach was developed and refined, mostly at the NBER under the leadership of Geoffrey H. Moore… (p. 13)

… Clearly, the peaks and troughs in the coincident index line up closely with the official peak and trough dates from the NBER. The largest deviation is the three months at the 1960 peak. Eight of the last 13 turning points match exactly, and all turning points in the coincident index correspond to either the beginning or end of a recession… (p. 14)

"Business Cycle Indicators Handbook." 2000. The Conference Board.

The Conference Board's Dec 19, 2019 release of U.S. Business Cycle Indicators explains that "The leading, coincident, and lagging economic indexes are essentially composite averages of several individual leading, coincident, or lagging indicators." [15] Each input is weighted by a standardization factor which essentially normalizes for the standard deviation of that series. The standardization factors are scaled to sum to 1. The Coincident Economic Index includes four variables: (1) employees on nonagricultural payrolls, (2) personal income less transfer payments, (3) industrial production, and (4) manufacturing and trade sales. The Leading Economic Index comprises ten different variables, including stock prices and the interest rate spread on 10-year Treasury bonds versus the Federal Funds Rate, both of which correspond to variables we use in our index.

Notes

References

Brereton, R. G. 2015. "The Mahalanobis distance and its relationship to principal component scores." Journal of Chemometrics, January 9, 2015. https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.2692

Chow, G., E. Jacquier, M. Kritzman, and K. Lowry. 1999. "Optimal Portfolios in Good Times and Bad." Financial Analysts Journal, vol. 55, no. 3: 65-71 (May/June).

Estrella, A. and G. A. Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." Journal of Finance vol. 46, no. 2: 555–76.

Estrella, A. and F. S. Mishkin. 1998. "Predicting U.S. Recessions: Financial Variables as Leading Indicators." Review of Economics and Statistics, vol. 80, no. 1: 45–61.

Gilchrist, S. and E. Zakrajsek. 2012. "Credit Spreads and Business Cycle Fluctuations," American Economic Review, vol. 102: 1692–720 (June).

Gilchrist, S., V. Yankov, and E. Zakrajšek. 2009. "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets." Journal of Monetary Economics, vol. 56, no. 4: 471–93.

Lin, R., E. Khalastchi, and G. Kaminka. 2010. "Detecting Anomalies in Unmanned Vehicles Using the Mahalanobis Distance." The Maverick Group, Computer Science Department, Bar-Ilan University, Ramat-Gan, Israel 52900.

Mahalanobis, P. C. 1927. "Analysis of Race-Mixture in Bengal." *Journal of the Asiatic Society of Bengal*, vol. 23: 301-333.

Mahalanobis, P. C. 1936. "On the Generalised Distance in Statistics." *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1: 49-55.

Nasief, H., I. Rosado-Mendez, J. Zagzebski, and T. Hall. 2019. "A Quantitative Ultrasound-Based Multi-Parameter Classifier for Breast Masses," Ultrasound in Medicine & Biology, vol. 45, no. 7: 1603-1616 (July).

Pike, T. and F. Vazquez-Grande. 2019. "Out-of-Sample Performance of Recession Probability Models." FEDS Notes No. 2019-12-13-1.

Stock, J. H., and M. W. Watson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." Journal of Economic Literature, vol. 41, no. 3: 788–829.

Su, C. and S. Li. "A MD Based Classifier for Diagnosing Diseases. 2002 "Journal of Chinese Institute of Industrial Engineers, vol. 19, no. 5: 41-47 (September).

Wang, P., C. Su, K. Chen, and N. Chen. 2011. "The Application of Rough Set and Mahalanobis Distance to Enhance the Quality of OSA Diagnosis." Expert Systems with Applications, vol. 38, no. 6: 7828-7836 (June).

---

[1] See Mahalanobis (1927) and Mahalanobis (1936).

[2] For further description and interpretation of the Mahalanobis distance including its relationship to principal components, see for example Brereton (2015).

[3] Industrial Production is the one-year percentage change in the monthly Industrial Production Index which is available from the Archival Federal Reserve Economic Data (ALFRED) repository maintained by the Federal Reserve Bank of Saint Louis starting in January 1920 (code: INDPRO). Prior to 1920 we use the Index of Industrial Production and Trade for the United States from ALFRED (code M1204BUSM363SNBR) as a proxy for Industrial Production. Nonfarm Payrolls is the rolling one-year percentage change in monthly Nonfarm Payrolls which is available from ALFRED starting in January 1940 (code: PAYEMS). Prior to 1940 we use the Index of Factory Payrolls for the state of New York from ALFRED (code M08G8AUS000NYM335NNBR) as a proxy for Nonfarm Payrolls. We scale down the standard deviation of the one-year changes in New York payrolls to adjust for the higher volatility of this series relative to nonfarm payrolls for the United States; the two series are 99% correlated during the period for which overlapping data is available. The return of the stock market is the rolling one-year price return of the S&P 500 Composite Index which is available monthly on Robert Shiller's website from 1871 to present and can also be replicated using data from Datastream. The slope of the yield curve is measured as the average difference between the ten-year U.S. treasury yield (code: USTRCN10) and the U.S. Federal Funds Rate (code: USFDFUND) over the preceding 12 months which is sourced from Datastream and captured on the 15th day of the prior month. Prior to November 1964, we use the difference between the ten-year U.S. Treasury yield and the annual risk-free rate, both available on Ken French's website, as a proxy for the yield curve. Prior to 1926 we use the difference between the ten-year U.S. Treasury yield (available on Robert Shiller's website) and the Federal Reserve Bank of New York discount rate (available on the NBER website) as a proxy for the yield curve. Finally, NBER recessions are identified by NBER and are available at a monthly frequency from ALFRED. For both Nonfarm Payrolls and Industrial Production data, we select all available vintages from ALFRED so we are able to identify the data that was actually available at each point in time, prior to any future revisions. Not all vintages are available prior to 1956, hence we treat the pre-1956 period as "in sample" to the extent that the data includes the benefit of subsequent revisions; our out-of-sample analysis begins in January 1956.

[4] We winsorize the vector of current observations, x, to prevent extreme outliers from distorting the results. Specifically, if any of the four variables are more than three growth standard deviations above the growth mean, we reset them to this level; if any of the variables are less than three recession standard deviations below the

recession mean, we reset them to this level. We do not make any adjustments to observations for the purposes of computing either covariance matrix.

[5] For example, the analog of Equation 7 for a logit (or logistic) regression model is $\frac{\partial p_{logit,rec}}{\partial x} = p_{rec}p_{gr}\beta$. If we rescale these vectors as per Equation 8 to derive the measure of relative variable importance, it will simply equal β, the fixed vector of variable coefficients.

[6] We obtain data for the Conference Board's Coincident and Leading Indexes from the Archival Federal Reserve Economic Data (ALFRED), including prior vintages of point-in-time data where available. We plot the index values from the earliest vintage for which each month's observation is available.

[7] From "The NBER's Recession Dating Procedure," January 7, 2008, https://www.nber.org/cycles/jan2003.html, accessed on July 24, 2020.

[8] In this event study, we use the NBER's designations of recessions. We define robust growth periods as quarters in which the trailing year's real GDP growth ranks near the top of its distribution over the prior ten years. We set a threshold of 85% to match roughly the historical frequency of recessions.

[9] Removing the requirement that the yield curve slope is falling (a standardized shift less than -1) does not materially change the results.

[10] For motivation and tests of other promising methods, see for example Pike and Vazquez-Grande (2019).

[11] For example, Estrella and Hardouvelis (1991) applies the term structure as a predictor of the real economy, Gilchrist and Zakrajsek (2012) and Gilchrist et al. (2009) analyze prediction with credit spreads, and Estrella and Mishkin (1998) and Stock and Watson (2003) evaluate the use of market-derived variables more generally.

[12] We compute the inputs in the exact same manner, including the winsorization procedure described above.

[13] "September 20, 2010 announcement of June 2009 business cycle trough/end of last recession," accessed December 2019 via www.nber.org/cycles/main.html.

[14] "September 20, 2010 announcement of June 2009 business cycle trough/end of last recession," accessed December 2019 via www.nber.org/cycles/main.html.

[15] Accessed December 2019 via www.conference-board.org.