Testing the Power of Leading Indicators to Predict Business Cycle Phase Changes

By

Allan Layton
School of Economics and Finance
Queensland University of Technology
GPO Box 2434, Brisbane, Australia, 4001.
  Phone: +61 7 3864 2947
  Fax: +61 7 38644150
  Email a.layton@qut.edu.au

and

Daniel R. Smith
Faculty of Business Administration
Simon Fraser University
8888 University Drive
Burnaby BC V5A 2S6, Canada,
  Phone: +1 604 291 4675
  Fax: +1 604 291 4920
  Email: drsmith@sfu.ca
Testing the Power of Leading Indicators to Predict Business Cycle Phase Changes

BIOGRAPHY

Allan Layton

Professor Allan Layton, BEcon (Hons), MEcon, PhD (Qld), received his Doctorate from University of Queensland in 1982. Since then he has held academic posts at Latrobe University, Griffith University, Macquarie University, University of Queensland, and Queensland University of Technology in Australia and at Columbia University in the US. He is currently Professor of Economics and Head of School of Economics and Finance at Queensland University of Technology. His research and teaching interests span international business cycle analysis, financial and monetary economics, macroeconomic policy, provisioning of public infrastructure, and applied econometrics. He is the author of the 1994 book Modern Australian Macroeconomics, co-author of the book Economics for Today published in 2002, and has published widely in international academic journals. He has also acted as consultant to a number of private and public sector organisations in Australia and in the US.
Dr. Daniel Smith, B.Bus. (Hons), M.Bus., Ph.D., is an Assistant Professor of Finance in the Faculty of Business Administration, Simon Fraser University in Vancouver Canada. Daniel earned his Ph.D. in finance from the University of British Columbia in 2002 and has published in *Applied Economics*, the *Journal of Business and Economic Statistics*, and the *Journal of International Money and Finance*. Before joining SFU, Daniel taught at the Queensland University of Technology and UBC. Daniel’s research interests include asset pricing, portfolio selection, fixed income, econometrics and business cycle modeling.
Testing the Power of Leading Indicators to Predict Business Cycle Phase Changes

Abstract

In the business cycle literature researchers often want to determine the extent to which models of the business cycle reproduce broad characteristics of the real world business cycle they purport to represent. Of considerable interest is whether a model’s implied cycle chronology is consistent with the actual business cycle chronology. In the US, a very widely accepted business cycle chronology is that compiled by the National Bureau of Economic research (NBER) and the vast majority of US business cycle scholars have, for many years, proceeded to test their models for their consistency with the NBER dates. In doing this, one of the most prevalent metrics in use since its introduction into the business cycle literature by Diebold and Rudebusch (1989) is the so-called quadratic probability score, or QPS. However, an important limitation to the use of the QPS statistic is that its sampling distribution is unknown so that rigorous statistical inference is not feasible. We suggest circumventing this by bootstrapping the distribution. This analysis yields some interesting insights into the relationship between statistical measures of goodness of fit of a model and the ability of the model to predict some underlying set of regimes of interest. Furthermore, in modeling the business cycle, a popular approach in recent years has been to use some variant of the so-called Markov regime switching (MRS) model first introduced by Hamilton (1989) and we therefore use MRS models as the framework for the paper. Of course, the approach could be applied to any US business cycle model.

Keywords: Markov Regime Switching, Business Cycle, Quadratic Probability Score.
Testing the Power of Leading Indicators to Predict Business Cycle Phase Changes

1.0 Introduction

In the business cycle literature it is very common for researchers to be interested in determining the extent to which their models of the business cycle reproduce broad characteristics of the real world business cycle they purport to represent. Thus, analysts will compare the simulated amplitude of business cycle swings implied by their business cycle models with the actual business cycle amplitude. The average durations of recessions and expansions are also usually analysed for comparability with the actual business cycle and so on. Finally, of great interest is whether the model’s implied cycle chronology is consistent with the actual business cycle chronology, defined as the dates at which the economy moves from one phase into another (say expansion into recession (peak), or recession into expansion (trough)).

In the US, the most widely accepted business cycle chronology is that compiled by the National Bureau of Economic research (NBER). This is a monthly chronology which extends back to before WWI.¹ This chronology is provided in Table 1. The “actual” chronology of a country’s business cycle is not directly observable and so many other countries do not have such a widely accepted set of dates for the peaks and troughs in the cycle but, for the US, it is reasonable to regard the NBER chronology as, in effect, amounting to the actual US business cycle chronology. The vast majority of US business cycle scholars have, for many years, regarded it as such and have proceeded to test their models for their consistency with the NBER dates.

1 Insert Table 1 about Here

¹ For the purposes of this paper only the chronology dating from the 1950s will be used.
In doing this, one of the most prevalent metrics in use since its introduction by Diebold and Rudebusch (1989) is the so-called quadratic probability score, or QPS, defined in the next section. QPS is simply a mean square error measure where the “actuals” consist of zeros and ones and the “predictions” are model-generated probabilities varying between zero and one. In the case of the US business cycle, the NBER dates would define a series of zeros (for recessions, say) and ones (for expansions). If a model were able to perfectly predict the phase of the economy every period the model-generated probabilities would coincide perfectly with the zeros and ones deriving from the NBER dates and QPS would be zero. The higher the QPS, the worse the fit of the model to the NBER chronology. Researchers will therefore commonly calculate the QPS for their models or compare QPS statistics for different models and draw conclusions about the relative superiority of one model over another. Examples of researchers who have used QPS in the business cycle context include Diebold and Rudebusch (1989), Filardo (1994), Lahiri and Wang (1994), Layton (1997), Chauvet (1998) and others.

However, quite an important limitation to the use of the QPS statistic is that its sampling distribution is unknown so that rigorous statistical inference is not feasible. We suggest circumventing this by bootstrapping the distribution. This analysis yields some interesting and valuable insights about the relationship between statistical measures of the goodness of fit of a model and the ability of the model to predict some underlying set of regimes (and hence phase shift turning points).

Furthermore, in modeling the business cycle, a popular approach in recent years has been to use some variant of the so-called Markov regime switching (MRS) model first introduced by Hamilton (1989). The MRS model is non-linear and assumes a times series under study can be in one of a small number of discrete underlying (usually unknown latent) states. The probability rule governing the likelihood of different values being observed for the series is allowed to vary across states with the probability of the series transitioning from one state to another being governed by a set of Markov transition probability parameters (more in the next section).

2 Actually, the metric had its origin in weather forecasting, being introduced by Brier (1950).
The transition parameters can be assumed to be invariant through time (constant transition probability, or CTP, models), or, more interestingly, can be allowed potentially to vary through time and to depend on some set of underlying determinants (time varying transitional probability, or TVTP, models). Many papers in recent times have therefore compared CTP and TVTP models for some measure of the macro economy to determine whether a given set of determinants of interest could be shown to have statistically significant informational content in forecasting future business cycle phase shifts (i.e., business cycle turning points).

The TVTP models typically include one or more lagged leading economic indicators as the putative determinants of the transition probability parameters. In testing their informational value, in addition to the usual sorts of statistical fit measures and tests (for example, the value of the log likelihood – or R-Sq – and likelihood ratio tests), the QPS statistic is commonly calculated to determine the closeness of the model-generated probabilities to the NBER chronology. We argue that this is a very important diagnostic to be used in evaluating any empirical US business cycle model. We also argue that it is quite possible for a given model to produce a superior statistical fit but perform poorly in fitting the actual US business cycle chronology. In such cases the specification of the model in question should be reconsidered.

In sum, the purpose of this paper is to demonstrate how formal statistical inference can be carried out using QPS. In doing this we use MRS models as the framework. However, the approach could – and should – of course be applied to any US business cycle model. We also provide an actual example of a model incorporating a popular leading indicator which provides a reasonable statistical fit compared with another but which nonetheless performs quite considerably worse when it comes to replicating the US business cycle chronology. Finally, using the QPS metric, we also perform an out-of-sample forecasting analysis of a preferred model.
2.0 Developing the Basic Models

In this section we develop a univariate model to represent the leading indicator used in the analysis as well as alternative CTP and TVTP models for the business cycle measure we use in the study. Inter alia, these models are important as we use them extensively in the bootstrapping work we carry out for our estimated QPS statistic.

The basic data for our study comprise monthly observations on the experimental leading (XLI) and coincident (XCI) indexes of US economic activity developed and maintained by Stock and Watson. These data are described more fully in the Data Appendix.

2.1 Univariate Time Series Model of the Leading Index (XLI)

The first step is building a time-series model that is capable of explaining the important features of the data. We consider a range of univariate time-series models for the dynamics of XLI and settle on an AR (2) process with asymmetric GARCH effects and fat tails.

We employ two different approaches and find the same model. We first estimate the mean dynamics and then test the volatility specification. We consider a range of AR models using adjusted $R^2$ to compare them, which selects an AR (2) model. We then verify that the residuals from this AR (2) model are serially uncorrelated.\(^3\) To test for heteroskedasticity, and ARCH effects in particular, we estimate the Ljung-Box (LB) statistic for squared residuals using 6 lags which is 42.1654 and clearly significant.

We also find that we need to allow for AR and GARCH effects using a standard model selection approach. We consider all AR-GARCH models, with both symmetric and asymmetric volatility specifications and both normal and T (and estimated degrees of freedom) distributions out to 3 lags. The conditional mean is modeled as

\[^3\] The Ljung-Box portmanteau statistic with 6 lags is 9.1955 with a p-value of 0.1629.
\[ \mu_t = \phi_0 + \sum_{i=1}^{m} \phi_i y_{t-i} \]

and the conditional variance is modeled as

\[ h_t = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} + \sum_{i=1}^{r} \delta_i e_{t-i}^2 1(e_{t-1} < 0), \]

where \( e_t = y_t - \mu_t \), and the last term captures asymmetry in the conditional volatility specification. We estimate the unconditional volatility \( \sigma^2 \) which is related to \( \omega \) by the relationship \( \sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{q} \beta_i - 0.5 \sum_{i=1}^{r} \delta_i} \). We model the standardized residual

\[ z_t = \frac{e_t}{h_t^{0.5}} \]

as either a standard normal random variable or a standardized T random variable with \( v \) degrees of freedom (where we actually estimate \( v^{-1} \) as a parameter). We search over all combinations of models with \( m \leq 3 \) and \( p, q, r \leq 2 \). We select the “best” model to maximize the Bayesian information criterion (BIC).

The resulting model is a symmetric AR (2)-GARCH (1, 1) model with a standardized-T distribution. We report the parameter estimates and robust standard errors in Table 2.

**Insert Table 2 about Here**

The last two lines of Table 2 present the LB test for serial correlation in the standardized residuals \( z \) with 12 lags and the LB test for serial correlation is the squared standardized residuals \( z^2 \) with 12 lags. Neither diagnostic test is statistically significant indicating that the model does a good job at whitening the standardized residuals.

### 2.2 Developing Markov-Switching Models for the Coincident Index (XCI)

Consider a constant transition probability (CTP) Markov regime switching (MRS) model with two phases (states) and without any autoregressive dynamics. The MRS model is
characterized by a latent state variable $S_t = \{1, 2\}$ which describes the state of the economy and evolves as a first-order Markov chain with transition probabilities

\[

d(S_{t+1} \mid S_t = 1) = p,
d(S_{t+1} = 2 \mid S_t = 1) = 1 - p,
d(S_{t+1} = 1 \mid S_t = 2) = 1 - q,
d(S_{t+1} = 2 \mid S_t = 2) = q
\]

Although the state variable $S_t$ is unobservable, we do observe a variable $y_t$ whose distribution does depend on $S_t$. In particular we model the conditional distribution of $y_t$ as

\[
    f(y_t \mid S_t = i; \theta) = \frac{1}{\sqrt{2\pi}\sigma_i^2} \exp\left(\frac{-(y_t - \mu_i)^2}{2\sigma_i^2}\right)
\]

where $\theta = (p, q, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)'$ is the parameter vector. We estimate the parameters by maximum likelihood and infer the state of the economy using Hamilton’s (1989) recursive filter.

We do not include any further dynamics for the conditional mean and variance beyond changes in the conditional state probabilities because we are primarily focused on tracking changes in regime. Results for the CTP model are provided in Table 3 below.

**Insert Table 3 about Here**

We also estimate an extended model which allows the transition probabilities to potentially depend upon movements in the leading index (XLI). The specification of the transition probabilities in this time varying transition probability (TVTP) model is as follows:

\[
    p(S_{t+1} = 1 \mid S_t = 1, I_t) = p_{t+1} = (1 + \exp(-\beta_{01} - \beta_{11}X_t))^{-1}
\]

and

\[
    p(S_{t+1} = 2 \mid S_t = 2, I_t) = q_{t+1} = (1 + \exp(-\beta_{02} - \beta_{12}X_t))^{-1}.
\]
The logistic transformation ensures that the transition probabilities lie between zero and one and X denotes the XLI.

As far as ex ante prediction of the US business cycle phases is concerned, we measure the model’s ability to explain the regimes using the following QPS statistic:

\[
QPS = \frac{1}{T} \sum_{t=2}^{T} \left( \Pr(S_t = 1 \mid I_{t-1}) - \lambda_t \right)^2
\]

where: \( \Pr(S_t = 1 \mid I_{t-1}) \) is the ex ante forecast probability of a recession in the next period formed using information available last period which is denoted by \( I_{t-1} \), \( \lambda_t \) is a dummy variable taking the value one when the NBER chronology indicates that the US economy was in recession in month \( t \) (a recession is defined as any month between a peak (non-inclusive) and the subsequent trough (inclusive)). The QPS measure is basically an MSE measure where we are predicting a dummy variable using a conditional probability rather than a continuous random variable.

The parameter estimates from the TVTP model are also presented in Table 3. The likelihood ratio statistic for the null hypothesis of constant transition probabilities is 18.86. Under the null, this statistic is distributed as chi-squared with 2 degrees of freedom (if the conditional density is correctly measured) and so the calculated value of 18.86 equates to a p-value <0.0001.\(^4\)

At least as important as the LL measure of statistical improvement is the TVTP model’s performance in predicting turning points in the US business cycle (as defined by the NBER chronology), such performance being measured by the respective models’ QPS statistics. We report the QPS statistics in Table 3 and, as can be seen, there is an improvement (of about 24%) in QPS for the TVTP model over the CTP model, suggesting some support for the superiority of the model incorporating the information present in the XLI.

\(^4\) We also calculated the robust Wald test for this same null hypothesis. The value of this robust test statistic is 14.5708 which again is significant at better than the 0.0001 significance level.
However, as noted in the introduction, quite an important limitation to the use of the QPS statistic is that its sampling distribution is unknown so that rigorous statistical inference is not feasible. We suggest circumventing this by bootstrapping the distribution. Indeed, we find that this analysis yields some interesting insights into the relationship between statistical measures of the goodness of fit of an MRS model and the ability of the model to predict some underlying set of regimes (and hence turning points).

3.0 Bootstrapping the Distribution of the QPS Statistic with “Useless” Leading Indicators.

Our first experiment constructs the distribution of the QPS statistic when the leading indicator is “useless” and contains no information about the NBER official business cycle chronology. The objective of the experiment is to understand the behavior of QPS under the null hypothesis that XLI contains no information useful in dating the business cycle.

We proceed by simulating 1000 artificial XLI series all containing the same number of observations and having the same time-series dynamics as the actual XLI. In particular we simulate sample paths using the estimated parameters from the univariate AR (2)-GARCH-T time-series model we estimated for XLI in Section 2. Note that although the artificial leading indicators and XLI display the same time-series dynamics, the artificial data are randomly generated independent of the sample path of XCI and therefore have no systematic relationship to the US business cycle; hence the “useless” label. To reiterate, this experiment helps us gauge the likelihood that the improvement in the QPS statistic that we observe in the real data could have arisen purely by chance. For each artificial leading indicator we re-estimate the TVTP MRS model’s parameters (using the actual XCI data) and calculate the corresponding QPS and LL.  

When simulating each of the 1000 artificial series of the leading indicator we initialize the first two values of the series at their unconditional mean, simulate 1538 subsequent observations, and then drop the initial 1000 to avoid any contamination from the starting values selected.
It should be noted that, by virtue of the definition of the QPS statistic, this simulation experiment is designed to assess the distribution of the QPS statistic improvements under the null that the leading indicator is of no value in relation to the NBER business cycle chronology. It is of course possible that the unknown underlying latent states of CXI are devoid of business cycle economic content and therefore track the NBER recession dates very poorly. Thus, it should be emphasized that this current exercise relates to the model’s predictive abilities in relation to the NBER recession dates rather than the unknown underlying regimes of CXI itself. In another experiment below we address this issue.

Furthermore, the transition probabilities may not be constant but rather vary in response to some other leading indicator not under consideration. We cannot assess this hypothesis with the current experiment. The experiment is set up to test the statistical significance of the improvement in QPS using the specific indicator, XLI. If the improvement proves insignificant the issue of whether XCI’s transition probabilities are truly constant remains an open one. However, later, we conduct an experiment in which we simulate data that truly comes from a CTP MRS model.

Despite these two issues the question which this current experiment asks is nonetheless very useful and economically interesting. In recent years, many investigators have sought to investigate the ability of various leading indicators of interest to them to date the NBER chronology using an MRS model of some particular series or coincident index. Of critical interest then is conducting inference under the null that the selected leading indicator(s) contain no information to predict phase changes beyond the coincident indicator in question. For instance, in the current case, we care relatively little about whether the true underlying states of XCI come from a CTP model or whether some other leading indicator is, in some sense, better than XLI. What we are interested in stating is how likely the observed 24 percent improvement in the QPS statistic using XLI could have arisen simply by random chance. This question is directly addressed by the current experiment.
Of the 1000 simulated XLI series, only about half (445) resulted in the TVTP model’s QPS statistic being lower than the estimated CTP model for XCI ($QPS^{CTP}$ of 0.1104). Of these, no artificial XLI series actually produced a QPS lower than the 0.0844 obtained using the real XLI in the TVTP model; the smallest artificial $QPS^{TVTP}$ was 0.0866, still larger than that found using the real XLI. The empirical distribution of the QPS improvement is provided visually in Figure 1. As is clear from the figure, under the null hypothesis that XLI is without useful informational content for NBER phase changes, the probability of observing an improvement as large as the actual observed improvement of 24% is zero. In fact, the largest improvement we observed in our 1000 simulations is only 21.5%.

### 3.2 Bootstrapping the Distribution of QPS with Known States

In our previous analysis we used simulation to statistically evaluate the closeness of ex ante model generated forecasts of state probabilities – derived from a TVTP MRS model of XCI using XLI – to the official NBER business cycle chronology. An important aspect of that approach is that it uses the actual NBER chronology and the XCI. However, there are a couple of limitations. Firstly, although it is quite reasonable to believe that the NBER chronology and the XCI latent phases are closely linked, this may well not be the case (in which case the usefulness of XCI would presumably need to be seriously re-considered by the NBER). Hence, if this were the case, a TVTP model of XCI using XLI may well be highly statistically significant as far as modeling the latent states of XCI but may nonetheless perform quite poorly in terms of fitting the US business cycle chronology as defined by the NBER dates. A second limitation is the inability of the experiment to shed light on the behavior of the QPS statistic under the alternative hypothesis that the transition probabilities for XCI really do vary through time, but according to some other cause, other than movements in XLI.

To complement our previous analysis we conduct an experiment in which we jointly generate the states, coincident index and leading index. This allows us to analyze the
relationship between the forecast state probabilities and the known true states of the artificial data and investigate the properties of the empirical distribution of the QPS statistic under these circumstances. In all our simulations we again use 1000 replications.

### 3.2.1 Simulating CTP Models with known states

In these experiments we simulate a series $y_t$ which has the same dynamics as the CTP MRS model of XCI. We then estimate a CTP model for the simulated $y_t$ and calculate the QPS for the estimated model (Experiment 2). The key advantage of this is that, because $y_t$ is a generated series, we know its true states, and can use them in the calculation of QPS. We are then able to compare the improvement in both QPS and LL using a spurious TVTP model for $y_t$ incorporating some independently randomly generated useless leading indicator, $X_t$ (Experiment 3).

We proceed in a number of steps for each of the 1000 simulated data series $i$:

1. Simulate $T$ observations on the useless leading indicator $X^{(i)}_t$ using the estimated AR(2)-GARCH(1,1) (as in Experiment 1)

2. Simulate the state vector $S^{(i)}_t$ independently of $X$ using the CTP model transition probability estimates for the actual XCI series. We draw the first observation from the ergodic distribution for the state of XCI, and simulate transitions between states using draws from a uniform random number generator.\(^6\)

\(^6\) In particular, let $u^{(i)}_t$ be a draw from a continuous uniform random number defined on $[0,1]$. We set $S^{(i)}_1 = 1$ when $u^{(i)}_t \leq (1-q)/(2-p-q)$ and $S^{(i)}_1 = 2$ otherwise. We then recursively simulate future states: if $S^{(i)}_{t-1} = 1$ then $S^{(i)}_t = 1$ if $u^{(i)}_t \leq p$ otherwise set $S^{(i)}_t = 2$ and if $S^{(i)}_{t-1} = 2$ then set $S^{(i)}_t = 2$ if $u^{(i)}_t \leq q$ otherwise set $S^{(i)}_t = 1$. 

15
3. Given the sequence of states $S_t^{(i)}$ we simulate the artificial coincident index series by drawing a standard normal variate $z_t^{(i)}$ and using the states and the CTP model parameter estimates as:

$$y_t^{(i)} = \mu_0 + (S_t^{(i)} - 1)\mu_1 + z_t^{(i)} \sqrt{\sigma_0^2 + (S_t^{(i)} - 1)\sigma_1^2}.$$

For each simulated triple $y_t^{(i)}, X_t^{(i)}, S_t^{(i)}$, we analyze the model fit by estimating both a CTP model – the correct specification – and a TVTP model (the incorrect specification). The CTP model is estimated using only $y_t^{(i)}$, whereas we use both $y_t^{(i)}$ and $X_t^{(i)}$ when estimating the spurious TVTP model. We use $S_t^{(i)}$ only when calculating $QPS^{CTP}$ and $QPS^{TVTP}$ for each series $i$.

It will be recalled from Table 2 that the CTP model for the actual XCI produced a QPS statistic of 0.1104 when the NBER chronology was used in the calculation of QPS. A relevant question to ask is: How big is a QPS of .1104? The results of Experiment 2 shed some light on this. In Figure 2 we plot the empirical density of QPS when $y_t^{(i)}$ is generated by a CTP model, its states are known and used in calculating the QPS statistic, and when a CTP model is estimated using the generated data.

Also plotted as a vertical line in the figure is the QPS of .1104 computed from comparing the CTP MRS model of the real XCI with the official NBER chronology. It is rather evident that 0.1104 is in no way an outlier. The empirical 90 percent confidence interval is (0.0629, 0.1369). The value of .1104 is only slightly higher than both the mean (0.0998) and median value (0.0999). Furthermore, there is sufficient sampling variation to produce a cross-sample standard deviation of 0.0229, which means our observed value is within one standard deviation of the mean. These results may be interpreted as suggesting that, whatever are the latent states of the real XCI, the NBER chronology may not be too different from them.
In our third experiment we then estimated a spurious TVTP model for each $y_t^{(i)}$ using a generated useless leading indicator. In Figure 3 we present the resulting empirical distribution for QPS (using the known true states of $y_t^{(i)}$) calculated from these 1000 estimated models. It turns out that, when the data are truly from a CTP model, only 5 percent of spurious TVTP models estimated from the simulated series exhibited a QPS improvement greater than 7.1 percent. Again, the 24% improvement in evidence using the TVTP model using the real XCI and XLI – and the NBER business cycle chronology - would seem to suggest strongly that the business cycle forecasting improvement using XLI in the TVTP framework is highly significant. In fact, in this experiment where the true DGP is CTP, in only 2 of the 1000 estimated models were QPS improvements observed which were better than the actual observed 24%.

3.2.2 Simulating TVTP Models with known states

We are also interested in analyzing the power of QPS - and LL - to detect when TVTP is the true model for $y_t^{(i)}$. Does QPS contain informational content for identifying truly useful leading indicators? To that end we simulate leading indicators that have similar dynamic properties to XLI and use these to construct a set of states that follow a Markov chain with TVTPs which depend on the simulated leading indicator (Experiment 4). The simulation proceeds as above except that, in generating $y_t^{(i)}$, rather than use the means and variances from the CTP model for XCI, we use the equivalent TVTP estimates, and rather than use the CTP Markov chain to simulate the states for $y_t^{(i)}$, we simulate $S_t^{(i)}$ using the time-varying transition probability expressions below with parameter values used from Table 2.

$$p_{rs}^{(i)} = \left(1 + \exp(-\beta_{0s} - \beta_{1s} X_t^{(i)})\right)^{-1} \quad \text{and} \quad q_{rs}^{(i)} = \left(1 + \exp(-\beta_{0s} - \beta_{1s} X_t^{(i)})\right)^{1}$$

Again, it is critical that we know the underlying states for the purpose of calculating QPS, even though they are treated as unknown in the estimation stage. The resulting empirical distribution for QPS is provided in Figure 4.
Earlier it was found that, when the true model is CTP but a spurious TVTP model is used, 5% of computed QPS improvements were above 7.1%. In other words, the 5% critical value for rejecting a correct null of CTP was found to be 7.1%. What we see here is that, when we simulate data on \( y_t^{(i)} \) using a TVTP model and generate the leading indicator using the same dynamics as exhibited by XLI with the same transition parameters, 96.6 percent of the simulated series exhibited a QPS improvement - over the estimated CTP model - higher than the 5% “critical value” of 7.1% improvement! That is, when the true model generating \( y_t^{(i)} \) is TVTP, in only 3.4% of the time would we accept the incorrect null of a CTP model? Of course, it should be noted that this power depends critically on the parameter values we use in the simulations and which we have set equal to the actual estimated values found in our empirical study.

4.0 Statistical Fit or Economic Content?

We have argued that in business cycle modeling the economic content of a model - as represented here by the improved forecasting of business cycle dates - beyond the standard measures of statistical fit (as measured, for example, by LL) is of very great importance. This means that a model which tests as statistically significant may in fact turn out to be quite a poor model of the US business cycle chronology. In such cases we would argue that the model needs to be reconsidered despite any apparent statistical significance of its parameters.

In this section we seek to make this point more concretely by demonstrating that a time-series that is generally thought of as being a good predictor of growth-rates in the economy and a recession predictor, has a good statistical fit as a leading indicator but performs very poorly in terms of predicting the NBER dates.

We consider the term spread between 10 year treasury bonds and the 3 month Treasury bill (TS). We take all series from the FredII database at the St Louis Federal Reserve.\(^7\)

\(^7\) [http://research.stlouisfed.org/fred2/](http://research.stlouisfed.org/fred2/)
Harvey (1989) shows that the current value of the TS is an apparently very useful predictor of the growth in real per capita GDP over the subsequent four quarters. This result is rather robust. The empirical finance literature also typically includes the TS as a conditioning variable which purports to capture business cycle conditions and, in fact, XLI actually includes the spread of 10 year over 1 year Treasury bond yields which is actually quite similar to our TS variable.

The parameter estimates for the TVTP model using TS are presented in Table 4. The LL of -377 is actually remarkably close to the LL when using XLI. However the QPS statistic is 0.1130 which is substantially higher than the 0.0844 obtained using XLI. In fact is even worse than the QPS we found using the CTP model! Thus, TS seems to provide quite a good statistical fit to XCI but nonetheless it provides quite a poor fit to the NBER recessions. This was also actually quite common in the bootstrapping simulations carried out in the previous section.

Insert Table 4 about Here

5.0 Comparing the performance out-of-sample

We now turn our attention to comparing the ability of the models to forecast recessions out of sample. In particular, we compare the ability of the constant (CTP) and time-varying (TVTP) models to forecast a recession occurring some time in the next six months (a common objective in practice). We focus on two different metrics: 1) using the predictor model, the average estimated probability that a recession will soon occur in the period leading up to a recession actually occurring, and 2) the fraction of correct predictions of a recession event in all instances in which the model predicts a recession occurring.

The event of the economy being in recession in one or more of the next six months is the complement to the event of the economy being in expansion in all the next six months which is given by
The conditional probability of a recession in the next six months is expression particularly simple in the CTP model:

$$\xi_t(6) = 1 - p \Pr(S_t = 0 \mid I_t) - (1 - q_t) \Pr(S_t = 1 \mid I_t))(1 - q)p^5.$$.

This intuition also holds true for models that allow for time-varying transition probabilities. However, things are complicated somewhat because the future transition probabilities are unknown in which case the conditional probability of a recession in the next six months is

$$\xi_t(6) = 1 - p_t \Pr(S_t = 0 \mid I_t) - (1 - q_t) \Pr(S_t = 1 \mid I_t))\prod_{k=1}^{5} p_{t+k}$$

where $$p_{t+k} = (1 + \exp(-\beta_{01} - \beta_{02}X_{t+k}))^{-1}$$ which depends on future values of the state variable $$X_{t+k}$$ and which are unobservable. One simple approach to calculating this is to forecast values of the leading indicator and use these in the calculations

$$\xi_t(6) \approx 1 - p_t \Pr(S_t = 0 \mid I_t) - (1 - q_t) \Pr(S_t = 1 \mid I_t))\prod_{k=1}^{5} \hat{p}_{t+k}$$

where $$\hat{p}_{t+k} = (1 + \exp(-\beta_{01} - \beta_{02}\hat{X}_{t+k}))^{-1}$$ and $$\hat{X}_{t+k}$$ is the optimal forecast of $$X_{t+k}$$ conditional on information available at time $$t$$.

We forecast out-of-sample and re-estimate all model parameters every month. The forecast recession probabilities at date $$t$$ uses parameters estimated using data available at that date. We use an expanding window with at least 20 years or 240 observations of data, so our first forecast is at December 1978, so we have a total of 4 peaks (January 1980, July 1981, July 1990, and March 2001).

---

8This approximation is motivated by the duration calculation suggested by Filardo and Gordon (1998) but which omitted Jensen’s inequality. The nonlinearity of the logistic transformation and the stochastic nature of future variables of the leading indicators can be allowed for using Monte Carlo methods, using

$$\xi_t(6) = 1 - p_t \Pr(S_t = 0 \mid I_t) - (1 - q_t) \Pr(S_t = 1 \mid I_t))\frac{1}{N} \sum_{i=1}^{N} \prod_{k=1}^{5} p_{t+k}^{(i)}$$

where $$p_{t+k}^{(i)} = (1 + \exp(-\beta_{01} + \beta_{02}X_{t+k}^{(i)}))^{-1}$$. This produced results which were only very marginally different from those reported.
The first comparison is to look at the average forecast probability of a recession in all months when a recession does occur in at least one of the subsequent six months. The average forecast probability produced by the CTP model is 61.36% while the TVTP model has an average probability of 74.25%.

The second comparison is to find all instances where the model predicts that a recession is likely some time in the next six months (which we define as all instances where P>0.5) and compare the percentage of correct classifications. We report these results in Table 5. There were 54 months that preceded at least one recessionary month occurring in the following six months. The CTP model correctly predicted only 32 of these events and missed 22, whereas the TVTP model correctly predicted 40 of these and missed only 14. However, the TVTP does tend to falsely predict recessions slightly more frequently than the CTP model.

Insert Table 5 about Here

6.0 Conclusions

The use of the quadratic probability score to evaluate the extent to which a business cycle model replicates the official NBER business cycle chronology is now quite common in US empirical business cycle modeling. However, formal statistical inference using QPS is hampered because its sampling properties are unknown. We present an empirical method to assess the statistical significance of the QPS statistic in the context of Markov-regime switching models of the business cycle. The method involves bootstrapping the empirical distribution of the QPS statistic for the application at hand.

Using this method we find that the experimental leading index of Stock and Watson provides statistically significantly improved forecasts of the NBER business cycle chronology for the US over and above what can be provided by Stock and Watson’s experimental coincident index itself. We also find that, whatever are the unknown latent
states of the experimental coincident index, they would appear to be quite a “close” approximation to the official phases of the business cycle as determined by the NBER-determined business cycle chronology.

We also find that another commonly used leading indicator of the business cycle, viz., the interest rate term spread, whilst appearing to provide about as good a statistical fit to the experimental coincident index, nonetheless performed quite substantially worse in forecasting the NBER chronology than does the experimental leading index as measured by its QPS performance. We would argue therefore that, in determining the usefulness of a putative leading indicator for the US business cycle, the model’s QPS should certainly be calculated and tested for its statistical significance.

Finally, we have carried out an evaluation of the out-of-sample performance of the experimental leading index and found that it provides an improvement in forecasting business cycle recessions beyond that which is provided by the coincident index itself.
References


Table 1: Augmented NBER Chronology - [http://www.nber.org/cycles/](http://www.nber.org/cycles/).

<table>
<thead>
<tr>
<th>Trough</th>
<th>Peak</th>
<th>Recession Duration</th>
<th>Expansion Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 1949</td>
<td>July 1953</td>
<td>45 months</td>
<td></td>
</tr>
<tr>
<td>May 1954</td>
<td>August 1957</td>
<td>10 months</td>
<td>39 months</td>
</tr>
<tr>
<td>April 1958</td>
<td>April 1960</td>
<td>8 months</td>
<td>24 months</td>
</tr>
<tr>
<td>February 1961</td>
<td>December 1969</td>
<td>10 months</td>
<td>106 months</td>
</tr>
<tr>
<td>November 1970</td>
<td>November 1973</td>
<td>11 months</td>
<td>36 months</td>
</tr>
<tr>
<td>March 1975</td>
<td>January 1980</td>
<td>16 months</td>
<td>58 months</td>
</tr>
<tr>
<td>July 1980</td>
<td>July 1981</td>
<td>6 months</td>
<td>12 months</td>
</tr>
<tr>
<td>November 1982</td>
<td>July 1990</td>
<td>16 months</td>
<td>92 months</td>
</tr>
<tr>
<td>March 1991</td>
<td>March 2001</td>
<td>8 months</td>
<td>120 months</td>
</tr>
<tr>
<td>November 2001</td>
<td></td>
<td>8 months</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Parameter Estimates for AR (2)-GARCH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Robust SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>0.3393</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>1.1424</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.2452</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.2366</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0454</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9381</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.1086</td>
</tr>
<tr>
<td>LL</td>
<td>-809.8408</td>
</tr>
<tr>
<td>LB(z,12)</td>
<td>13.0500</td>
</tr>
<tr>
<td>LB($z^2$,12)</td>
<td>18.3271</td>
</tr>
</tbody>
</table>

Notes: LB (z, 12) is the Box-Ljung portmanteau test for autocorrelation in the standardized residual with 12 lags, and LB ($z^2$, 12) tests for serial correlation in the squared standardized residuals.
Table 3: Parameter Estimates of the CTP and TVTP MRS Models for XCI.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CTP Model</th>
<th>TVTP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Robust SE</td>
</tr>
<tr>
<td>p</td>
<td>0.9727</td>
<td>0.0104</td>
</tr>
<tr>
<td>q</td>
<td>0.9185</td>
<td>0.0310</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>0.3422</td>
<td>0.0279</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.1310</td>
<td>0.0964</td>
</tr>
<tr>
<td>$\sigma_0^2$</td>
<td>0.1459</td>
<td>0.0136</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.6705</td>
<td>0.1761</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>2.4261</td>
<td></td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>0.3193</td>
<td></td>
</tr>
<tr>
<td>$\beta_{01}$</td>
<td>2.8984</td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.4013</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>-385.3328</td>
<td>-375.9028</td>
</tr>
</tbody>
</table>

Notes: LL stands for log likelihood and QPS stands for quadratic probability score. QPS-Ex Ante is the QPS statistic using ex ante probabilities of the state in period t+1 made in period t, $p(S_{t+1} = 1 | I_t)$. The TVTP model’s QPS is 24% lower than that for the CTP model.

Table 4: Parameter Estimates of the TVTP MRS Model using the 10 year Term Spread.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Robust SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>0.3501</td>
<td>0.0246</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.1377</td>
<td>0.0891</td>
</tr>
<tr>
<td>$\sigma_0^2$</td>
<td>0.1466</td>
<td>0.0144</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.6354</td>
<td>0.1621</td>
</tr>
<tr>
<td>$\beta_{00}$</td>
<td>2.6337</td>
<td>0.4492</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>1.1866</td>
<td>0.4137</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
<td>3.3931</td>
<td>1.1280</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.5898</td>
<td>0.4697</td>
</tr>
<tr>
<td>LL</td>
<td>-377.8444</td>
<td></td>
</tr>
<tr>
<td>QPS</td>
<td>0.1130</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Out-of-sample recession predictions.

<table>
<thead>
<tr>
<th></th>
<th>Total Predicted Recessions</th>
<th>Correct Predictions</th>
<th>Wrong Predictions</th>
<th>Missed Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTP</td>
<td>61</td>
<td>32</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>TVTP</td>
<td>71</td>
<td>40</td>
<td>31</td>
<td>14</td>
</tr>
</tbody>
</table>
Figure 1: Plot of percentage improvement in QPS from TVTP models estimated for artificially generated series for XLI.

Notes: The figure provides a Kernel-based nonparametric estimate of the density of the percentage change in the QPS arising from using a TVTP model for XCI using an artificially generated leading indicator. One thousand artificial XLI series were generated, each with the same DGP as the actual XLI but which were generated without any relationship to XCI. In each case, a TVTP model was estimated for XCI using the artificial XLI and the resulting QPS compared with the QPS obtained from the original CTP model for XCI. Note that a negative number represents an improvement associated with a lower QPS statistic. The vertical line denotes the 24% improvement we observe after including the real XLI to model the transition probabilities.
Figure 2: Plot of density of CTP QPS statistic in artificially generated data using CTP model.

Notes: We report the Kernel-based nonparametric estimate of the density of the QPS measure derived from comparing the forecast state probability with the true latent state when coincident index data are artificially generated from a CTP model with parameter estimates from Table 1. The vertical line denotes the QPS value of .1104 which resulted from calculating QPS using the estimated state probabilities from the estimated CTP model for the actual XCI series and the NBER chronology.
Notes: We report the Kernel-based Nonparametric estimate of the density of the percentage change in the QPS measure from estimating a TVTP model using simulated data for the leading index, states and coincident index. The data are simulated using a DGP for the CI which is a CTP MRS model with parameters matching the CTP model parameter estimates in Table 1. Note that a negative number represents an improvement in QPS associated with the spurious TVTP model. The vertical line denotes the 24% improvement observed earlier after including the actual XLI to model the transition probabilities for XCI with QPS calculated using the NBER chronology.
Notes: We report the Kernel-based nonparametric estimate of the density of the percentage change in the QPS measure arising from estimating both a TVTP and a CTP model using simulated data for the leading index, states and coincident index. The data are simulated using a DGP for the CI which is a TVTP MRS model with parameters matching the TVTP model parameter estimates in Table 1. Note that a negative number represents an improvement in QPS associated with an estimated TVTP model as compared with an estimated CTP model for the same generated data (a TVTP model being the true model). The vertical line denotes the 24% improvement observed earlier after including the actual XLI in a TVTP model to model the transition probabilities for XCI, with QPS calculated using the NBER chronology.
Data Appendix

We use the experimental coincident and leading indicator indexes of Stock and Watson which may be downloaded from Watson’s website: http://www.wws.princeton.edu/~mwatson/.

Our sample period is from March 1959 through December 2003, given a sample size of 538 observations.

The Experimental Coincident Index

The Experimental Coincident Index is a weighted average of four broad monthly measures of U.S. economic activity:

1. Industrial Production
2. Real personal Income, total, less transfer payments
3. Real manufacturing and trade sales, total
4. Total employee-hours in nonagricultural establishments

The weighted average is computed using current and recent values of the growth rates of these four series. This weighted average, in growth rates, is then cumulated to create an index in levels. The index is constructed so that it equals 100 in July 1967. The average monthly rate of growth in the Experimental Coincident Index is 3.0% at an annual rate. Thus the Experimental Coincident Index has approximately the same trend growth rate as real GNP, which grew at an average annual rate of 3.1% from 1960 to 1988. The Experimental Coincident Index is approximately one and one half times more volatile than real GNP: the standard deviation of quarterly growth (at annual rates) in the Experimental Coincident Index is 5.5% while the corresponding standard deviation for real GNP is 3.9%.

The Experimental Leading Index

The Experimental Leading Index is a forecast of the growth of the Experimental Coincident Index over the next six months (that is, for the six months subsequent to the month for which the data is available). The forecast is stated in percentage terms on an annual basis. Thus, for example, the Experimental Leading Index for April represents a forecast of the percent growth in the Experimental Coincident Index between April and October, at annual rates.

The Experimental Leading Index is a weighted average of seven leading indicators. These series, with their abbreviations used in the table, are:
1. Housing BP: Housing authorizations (building permits) -- new private housing.

2. MD Unf Ord: Real Manufacturers’ unfilled orders: durable goods industries (smoothed).

3. Exchange Rates: Trade-weighted index of nominal exchange rates between the U.S. and the U.K., Germany, France, Italy, and Japan (smoothed).

4. Part Tim Wk: Number of people working part-time in nonagricultural industries because of slack work (smoothed).

5. 10Yr TBond Rate: The yield on a constant-maturity portfolio of 10-yr U.S. Treasury bonds (smoothed).

6. 3mtCP, 3mtTB Spr: The spread (difference) between the interest rate on 3-month commercial paper (financial) and the interest rate on 3 month U.S. Treasury bills.

7. 10yrTB, 1yrTB Spr: The spread (difference) between the yield on constant-maturity portfolio of 10-yr U.S. Treasury bonds and the yield on 1-year U.S. Treasury bonds.

The data are plotted in the following figure.

Figure A1. The first plot is of month-on-month growth rates in the Stock-Watson Experimental Coincident Index (XCI). The second plot is the previous month’s value of the experimental leading indictor (XLI). We superimpose a dummy variable for the NBER business cycle where zero denotes an expansionary month, while the recession months are some positive constant (chosen to fit the two time series).
LISTING OF DISCUSSION PAPERS - 2003

Valadkhani A, Demand for M2 in Developing Countries: An Empirical Panel Investigation, No 149, July 2003


Lahiri R, Tax Distortions in a Neoclassical Monetary Economy in the Presence of Administration Costs, No 151 September 2003

Layton A, & Smith D, Duration Dependence in the US Business Cycle, No 152, August 2003


Robinson M, The Output Concept and Public Sector Services, No 155, September 2003

Worthington A, Brown K, Crawford M, & Pickernell D, Socio-Economic and Demographic Determinants of Household Gambling in Australia, No 156, September 2003


Worthington A, & Higgs H, Weak-form Market Efficiency in European Emerging and Developed Stock Markets, No 159, September 2003


Worthington C, Debt as a Source of Financial Stress in Australian Households, No 164, November 2003


Lahiri R, Cooperation v/s Non-cooperation in R&D Competition with Spillovers, No 166, December 2003


LISTING OF DISCUSSION PAPERS - 2004


Drew M, Mallin M, Naughton T, Veeraraghavan M, Equity Premium: - Does it Exist? – Evidence from Germany and United Kingdom, No 170, January 2004


Drew M, Stanford J, Portability of Superannuation Balances, No 173, March 2004


Valadkhani A, Worthington A, Layton A, An Analysis of the Rising Cost of Education in Australia, No 175, April 2004


Guégan D, How Can We Define the Concept of Long Memory? - An Econometric Survey, No 178, April 2004


Lahiri R, On Skill Heterogeneity and Inflation, No 181, June 2004


Jeisman S, Exchange Market Pressure in Australia, No 183, June 2004


Worthington, A, The distribution of Financial Literacy in Australia, No 185, November 2004

Higgs H, Worthington A, Systematic Features of High-Frequency Volatility in Australian Electricity Markets: Intraday Patterns, Information Arrival and Calendar Effects, No 186, December 2004

LISTING OF DISCUSSION PAPERS - 2005

Behrens P.  Economic Development Strategies: Examples from Europe and Australia, No 187, January 2005


Zhang S.  Consumption Behaviour Under Institutional Transitions in China, No 189, March 2005

Collet J & Fadili J Simulation of Gegenbauer Processes Using Wavelet Packets - No 190, March 2005

Clements A & White S.  Non-Linear Filtering With State Dependant Transition Probabilities: A Threshold (Size Effect) SV Model.  No 191, April 2005

Clements A & White S.  Non-Linear Filtering For Stochastic Volatility Models with Heavy Tails and Leverage - No 192, April 2005

Lee B.  Significance of Employing a Multilateral Index Formula for Interstate Comparisons: A Case Study of the Australian Farm Sector - No 193, June 2005

Li S, Alfay E.  Evidence on the Arbitrage Efficiency of SPI Index Futures and Options Markets – No 194, June 2005


Hoffman M, Worthington A, & Higgs H.  Modelling Residential Water Demand with Fixed Volumetric Charging in a Large Urban Municipality: The Case of Brisbane, Australia – No 196, June 2005

Wilson, C.  Exposure to Pesticides, Ill-health and Averting Behaviour: Costs and Determining the Relationships – No 197, July 2005
Wilson C & Tidseel C. Globalisation, Concentration of Genetic Material and Their Implications for Sustainable Development – No 198, July 2005

Pollard G. The Cost Decision: A New Discount Approach for Net Cost Projects – No 199, July 2005
Testing the Power of Leading Indicators to Predict Business Cycle Phase Changes

Allan Layton
&
Daniel Smith

Discussion Paper No 200, August 2005

Series edited by
Dr Steven Li

School of Economics and Finance

ISSN 1324-5910