Comparing Probability Forecasts in Markov Regime Switching Business Cycle Models

By

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

And

Allan Layton:
Faculty of Business
University of Southern Queensland
Toowoomba Qld 4350, Australia
Email:layton@usq.edu.au

January 27, 2007

Comparing Probability Forecasts in Markov Regime Switching Business Cycle Models

Abstract

We evaluate techniques for comparing the ability of Markov regime switching (MRS) models to fit underlying regimes of a series of interest. This is particularly important in the business cycle literature where one may be interested in determining whether using leading indicators to allow transition probabilities to vary improves the ability of MRS models to fit the NBER business cycle chronology. This is typically done using the quadratic probability score, or QPS (Diebold and Rudebusch (1989)). Although it is possible to statistically compare the QPS statistics for two MRS models using the Diebold and Mariano (1995) (DM) test statistic for comparing forecasts, we find using a Monte Carlo experiment that the DM statistic tends to under-reject (the null of "no difference in forecast accuracy") when comparing MRS models. This we believe is because of the strong non-normality of the forecast errors of such models. Furthermore, using simulation-based inference we demonstrate that leading indicators improve the fit of an MRS model of the US business cycle chronology by 24 percent, such improvement having a *p*-value of 0.001.

Keywords: Markov Regime Switching, Diebold and Mariano statistic, Quadratic Probability Score, Monte Carlo, Business Cycle,.

JEL Classification: C15, C53, E32

1.0 Introduction

The Markov Regime Switching (MRS) model has become an important tool for modeling many financial and economic time series. It has been particularly prominent in the business cycle literature. In this 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, including phase amplitudes and durations. Of particular interest in the context of MRS models is the degree to which a particular model can reproduce the actual phase change chronology of the business cycle. 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] 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.

By allowing the transition probabilities in MRS to vary with leading indicators we can assess their usefulness in predicting business cycle phase changes. The most important metric for assessing the ability of various MRS models to "fit" the business cycle is the so-called quadratic probability score (QPS) introduced by Diebold and Rudebusch (1989). The QPS statistic is simply a mean square error measure comparing the probability of a regime (the prediction) with an indicator variable for the regime. Smaller QPS statistics identify models that fit the NBER chronology better. The magnitude of the QPS statistic has been used in the literature to compare different regime-switching business cycle models. This metric can be used to determine, for example, whether using leading indicators to allow the transition probabilities to vary through time can improve

[1] For the purposes of this paper only the chronology dating from the 1950s will be used. The chronology is available from the NBER website.

their fit to the actual business cycle chronology. Examples of researchers who have used QPS in the business cycle context include Chen (2002), Diebold and Rudebusch (1989), Filardo (1994), Kholodilin and Yao (2005), Lahiri and Wang (1994), Layton (1998), Chauvet (1999) and others. Unfortunately the sampling distribution of the QPS statistic is unknown, and formal statistical comparisons of the fit of different model-generated chronologies to the actual NBER business cycle chronology are rare. The only attempts at a statistical comparison of models using the QPS statistic of which we are aware is that by Chen (2002) and Kholodilin and Yao (2005) who use the asymptotic Diebold-Mariano (1995) test.

The Diebold and Mariano (1995) statistic (DM statistic) is an important general-purpose tool for comparing two different competing sets of model-based forecasts. Diebold and Mariano (1995) propose a test statistic which involves calculating an asymptotic Z statistic using the variance of the loss-differential (in our context, the difference between the squared forecast errors). This can be applied in a straightforward fashion to the QPS statistic. In many applications the theoretical asymptotic standard normal distribution for the Z-statistic appears to be a very good approximation and the test performs satisfactorily. However, comparing QPS statistics in MRS models presents somewhat of a specific challenge to the DM approach because the distribution of the loss-differential is extremely non-normal. The forecast errors are usually very small while the system stays in a given regime but can be relatively large around the times of regime change, giving rise to an empirical distribution much more peaked than the normal distribution.

This strong non-normality is exacerbated when considering differences in the squared forecast errors. As a result the normal distribution is quite a poor approximation in even quite large samples. We demonstrate using a Monte Carlo experiment that in a wide range of models the DM statistic consistently under-rejects the null hypothesis even when the samples are as large as 500 or even 1000 observations. We therefore recommend caution when implementing the DM statistic to compare MRS models of the US business cycle.

Given the poor performance of the DM statistic in samples that are typical in many business cycle empirical analyses we turn to a purely empirical approach to formal statistical inference. We present an empirical application to modeling the business cycle using the experimental coincident index (XCI) of Stock and Watson. We find that allowing the transition probabilities to vary with their experimental leading index improves the model's statistical fit, as measured by log-likelihood, and its ability to fit the NBER chronology, as measured by its QPS statistic, by around 24 percent. This QPS improvement has an empirical *p*-value of 0.001 and is almost 1/20th of the *p*-value of 0.018 of the DM test statistic using the theoretical asymptotic distribution of the standard normal.

The paper proceeds as follows. In Section 2 we present the MRS model of the business cycle and present an application to the US business cycle. We discuss the Diebold-Mariano (1995) statistic in Section 3. In Section 4 we evaluate the finite-sample properties of the DM statistic using a Monte Carlo experiment. Section 5 assesses the ability of leading indicators to improve the fit of MRS models to the NBER chronology and we conclude in Section 6.

2.0 Regime Switching Models

The Markov Regime Switching (MRS) model of Hamilton (1989) has become a very important tool in understanding the dynamics of both economic and financial time series. The conditional distribution of a time series variable, y_t, depends on the value of a discrete state latent variable S_t. We will present our analysis using a very simple two-state model where $S_t = \{1,2\}$ and there are no autoregressive dynamics. The simplest MRS specification models the dynamics of S_t as a Markov process with constant transition probabilities:

$$\begin{aligned}
p(S_{t+1} = 1 \mid S_t = 1) &= p \\
p(S_{t+1} = 2 \mid S_t = 1) &= 1 - p \\
p(S_{t+1} = 1 \mid S_t = 2) &= 1 - q \\
p(S_{t+1} = 2 \mid S_t = 2) &= q
\end{aligned}$$

We model the conditional distribution of y_t given S_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 denote this model the constant transition probability (CTP) model. We estimate the parameters by maximum likelihood and infer the state of the economy using Hamilton's (1989) recursive filter. This filter produces a sequence of regime probabilities for each observation in the data.

We can extend the model to include time-varying transition probabilities (TVTP) which depend on some variable X_t. In MRS models of the business cycle it is common to define X_t using leading indicators, e.g., Diebold, Lee, and Weinbach (1994), Fillardo (1994) and Layton (1998), and, in a regime-switching model of interest rates, Gray (1996) defines X_t as the level of short-term interest rates. The transition probabilities are defined as:

$$p(S_{t+1} = 1 \mid S_t = 1, X_t) = p_{t+1} = \left(1 + \exp(-\beta_{01} - \beta_{11} X_t)\right)^{-1}$$

$$p(S_{t+1} = 2 \mid S_t = 2, X_t) = q_{t+1} = \left(1 + \exp(-\beta_{02} - \beta_{12} X_t)\right)^{-1}.$$

using the logistic transformation to ensure that the transition probabilities lie between zero and one.

2.1 The Quadratic Probability Score

It is common when modeling the US business cycle using MRS to compare the regime probabilities with the NBER business cycle chronology. This is done both informally by comparing graphs of a binary regime indicator variable with the regime probabilities, or more formally using the QPS statistic::

$$QPS = \frac{1}{T} \sum_{t=1}^{T} \left(\Pr(S_{t+1} = 1 \mid I_t) - \lambda_{t+1} \right)^2$$

where: $\Pr(S_{t+1} = 1 \mid I_t)$ is the ex ante forecast probability of a recession in the next period formed using information available up to the current period which is denoted by I_t, λ_{t+1}

is a dummy variable taking the value one when the NBER chronology indicates that the US economy was in recession in month *t+1* (a recession is defined as any month between a peak (non-inclusive) and the subsequent trough (inclusive)). The metric originated in weather forecasting (Brier, 1950) and was introduced to the MRS literature by Diebold and Rudebusch (1989). It has since been applied to business cycles by Filardo (1994), Lahiri and Wang (1994), Layton (1998), Layton and Smith (2000), Chauvet (1998) and others. It can be used to compare, for example, the performance of constant and varying transition probability models. In particular, a model with a small QPS statistic might be preferred to another model with a larger QPS statistic as it more closely reproduces the NBER chronology.

2.2 Regime Switching Models of the US Business Cycle

We estimate a two-regime CTP model for the experimental coincident (XCI) index developed by Stock and Watson (1989). We also use their experimental leading index (XLI) to estimate a TVTP model. Our sample period is from March 1959 through December 2003, giving a sample size of 538 observations. We plot the data in Figure 1. Their relationship to the underlying NBER chronology is clearly evident. The parameter estimates for both the CTP and TVTP models are reported in Table 1. Consider first the CTP results. Both regimes are quite persistent (as evidenced by the estimated transition parameters being close to one) and the average growth rate in the expansion state (state 0) is much higher than in the recession state. The variance of XCI is much higher when the economy is in recession. A very similar patter emerges from the TVTP model which appears to be a superior explanation of the XCI data as evidenced by a likelihood ratio statistic of 18.86 for the null hypothesis of constant transition probabilities.[2] The forecast and smoothed regime probabilities are presented in Figure 2 for the CTP model, and in Figure 3 for the TVTP model. Both models appear to do a pretty good job of fitting the NBER chronology.

[2] Under the null of no improvement in fit to the XCI data, this statistic is distributed as chi-squared with 2 degrees of freedom and so the calculated value of 18.86 equates to a p-value <0.0001.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.

Insert Table 1 about here

Insert Figures 1, 2, and 3 about here

We also calculate the QPS statistics for each model using the NBER official chronology as the reference of the actual state of the US economy. These dates may be downloaded from the NBER website http://www.nber.org/cycles/. The TVTP model produces a 24% improvement in QPS over the CTP model, suggesting that allowing the transition parameters to vary according to XLI has quite substantially improved the power of the model to predict the NBER-determined dates of the US business cycle phase changes. Of course, whether this improvement is a statistically significant one needs to be formally determined.[3]

3.0 Comparing Forecast Accuracy: The Diebold-Mariano (1995) Statistic

Suppose we are interested in comparing two forecasts of some random variable w_t, which are denoted by $\hat{w}_{i,t}$ for i=1,2. Diebold and Mariano (1995, DM hereafter) construct an asymptotic Z-statistic for comparing these two forecasts. Define the forecast errors as $e_{i,t} = w_t - \hat{w}_{i,t}$. The DM test evaluates the forecasts in terms of an arbitrary loss metric $g_{i,t}$ which we will define here as the squared error $g_{i,t} = e_{i,t}^2$. The test statistic is then given by

$$Z = \frac{T^{-1}\sum_{t=1}^{T}(g_{1t} - g_{2t})}{\sqrt{S/T}}$$

where S is an estimator of the variance of the sample mean loss differential $T^{-1}\sum_{t=1}^{T}(g_{1t} - g_{2t})$. DM suggest estimating S using the spectral density of the loss

[3] The model we use is very simple and does not allow for autoregressive dynamics in the conditional mean other than that which is generated by changes in regime. To explore this possibility we also fit both the CTP and TVTP models allowing for AR(1) dynamics. Consistent with results reported in Layton and Smith (2000) we find that adding AR(1) dynamics to the CTP model reduces the model fit as measured by QPS which increases to 0.1196. Interestingly, we find that the TVTP model with AR(1) dynamics actually improves on the basic TVTP model, producing a QPS statistic of only 0.0790. We thank an anonymous referee for this suggestion.

differential $g_{1t} - g_{2t}$ at zero frequency, similar to the estimators presented in Newey and West (1987) and Andrews (1991).[4] The numerator of the *Z* statistic is the difference between the mean-squared prediction errors from each forecast.

We can use the DM statistic to compare different MRS models in terms of their ability to reproduce the NBER chronology. In the context of MRS models w_t is defined as the business cycle indicator variable λ_t and the two forecasts are the regime probability predictions estimated from the two models. In this context the numerator of the *Z* statistic is nothing other than the difference between the QPS statistics from each model.

Under suitable regularity conditions we can rely on the law of large numbers to derive the asymptotic standard normal distribution for the *Z* statistic. For most applications the normal distribution will be appropriate in even quite modest sample sizes. However, the MRS model presents a particular challenge for the DM test. Virtually all MRS models are characterized by quite persistent regimes often with probabilities of staying in a given regime from one period to the next of around 0.9. While the economy is in that regime the regime probability forecast errors are usually quite small. However, at periods around phase changes at either a business cycle peak or trough, the forecast errors can become relatively large.

We illustrate the extreme non-normality with our MRS model of the business cycle presented in the previous section. The problem arises because of the discrete nature of λ_t (only taking the values of zero or one) and the fact that most of the calculated loss differentials (the ($e^2{}_{1t} - e_{2t}{}^2$)) are either very close to zero or quite large through the dataset. This is illustrated in Figure 4 where we plot the time-series of the forecast errors (e_{it}) for both the CTP and TVTP models, as well as the calculated loss differential for each observation in the dataset using the MRS models discussed in the previous section..

[4] Following Newey and West (1987) we use an estimator based on the Bartlett Kernel and calculate the optimal number of lags to include following Andrews (1991).

Insert Figure 4 about here

Of course, non-normality in the underlying data is not a problem as we only require that the sample mean of the loss differential converges to a normal distribution asymptotically. Even in the MRS model the *Z* statistic will converge to a standard normal distribution asymptotically. However, the pronounced non-normality in the loss-differential when comparing regime probability forecasts from constant and varying transition probability MRS models is such that the DM statistic converges to the normal distribution extremely slowly. In the next section we demonstrate in a Monte Carlo experiment that, for empirically relevant sample sizes (for monthly business cycle analyses), the asymptotic normal distribution is a poor approximation to the likely actual finite sample distribution of the calculated DM statistic.

4.0 A Monte Carlo Experiment

To understand how well the asymptotic distribution approximates the finite sample distribution in MRS models we undertake the following Monte Carlo experiment. For each of 1000 replications, indexed by *i*, we simulate a sequence of regimes ($S_t^{(i)}$), observations ($y_t^{(i)}$) and leading indicators ($X_t^{(i)}$) as follows:

1. Simulate T observations (T being 250, 500 and 1000) on a useless leading indicator, $X_t^{(i)}$, using a zero-mean AR(1) process with autocorrelation coefficient ρ={0.5, 0.9}.
2. Simulate the state vector $S_t^{(i)}$ independently of *X*.[5]

[5] Specifically, let $u_t^{(i)}$ be a draw from a continuous uniform random number defined on [0,1]. We set $S_1^{(i)} = 1$ when $u_t^{(i)} \leq (1-q)/(2-p-q)$ and $S_1^{(i)} = 2$ otherwise. We then recursively simulate future states: if $S_{t-1}^{(i)} = 1$ then $S_t^{(i)} = 1$ if $u_t^{(i)} \leq p$ otherwise set $S_t^{(i)} = 2$ and if $S_{t-1}^{(i)} = 2$ then set $S_t^{(i)} = 2$ if $u_t^{(i)} \leq q$ otherwise set $S_t^{(i)} = 1$. Values for *p* and *q* are selected as in Table 2.

3. Given the sequence of states $S_t^{(i)}$ we simulate the artificial coincident index series by drawing a standard normal variate $z_t^{(i)}$ as: $y_t^{(i)} = \mu_1 + (S_t^{(i)} - 1)(\mu_2 - \mu_1) + z_t^{(i)}\sqrt{\sigma_1^2 + (S_t^{(i)} - 1)(\sigma_2^2 - \sigma_1^2)}$. Parameter values are selected as in Table 2.

For each simulated triple, $y_t^{(i)}, X_t^{(i)}, S_t^{(i)}$, we then estimate both a CTP and a TVTP model. (Note that the CTP model is, of course, the correct 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 TVTP model. We then use the actual $S_t^{(i)}$ and the estimated model-generated regime probabilities to calculate QPS^{CTP} and QPS^{TVTP} for each simulation, *i*, and calculate the DM statistic for that simulation. This is repeated for each of the 1000 replicates. We report the rejection frequencies using the two-sided test of the null hypothesis of equal predictive accuracy at both the 5 and 1 percent nominal sizes in Table 2 for a range of ten different model parameterizations which are reported at the head of each column.

The most striking result to take from this table is the poor finite sample performance of the DM statistic. Using the appropriate critical value from the standard normal distribution, even when we have 500 observations (a sample size that is particularly relevant as we have 538 observations in our empirical model) and have a very persistent leading indicator (ρ=0.9) we only reject between 0.4 percent and 1.8 percent of samples when the nominal size is 5 percent. A similar pattern emerges even when using *T*=1000 observations where, for the same nominal size of 5 percent, the rejection frequency varies from 0.1 percent to 2.3 percent. In fact at the five percent level we typically reject in less than one percent of the samples though we sometimes reject around 2 to three percent of samples. The general lesson to take from this experiment is that the DM statistic rejects far less frequently than the nominal size would suggest. We therefore believe that one should be very cautious about accepting the null hypothesis of equal forecast accuracy that is indicated by a small calculated DM statistic relative to the standard normal distribution. Rather, we recommend evaluating the statistical significance of any QPS improvement using Monte Carlo methods.[6]

5.0 Do Leading Indicators Help Fit the Business Cycle?

A leading indicator such at XLI may be regarded as helping to explain the business cycle if the regime probability forecasts improve when the transition probabilities are allowed to vary according to the leading indicator. In this section we statistically test the QPS improvement of our TVTP MRS model of the US business cycle.

Earlier, we reported that we found that allowing the transition probabilities to vary with XLI improves the QPS statistic over the CTP model by a seemingly impressive 24%. Furthermore, the DM statistic is -2.34 which has a *p*-value of 1.86 percent and is statistically significant at the 5 percent but not the 1 percent level. Our simulation experiments above suggest that this *p*-value may be understated as we uncovered a tendency for the DM statistic to under-reject when applied to regime shifting models. For this reason we undertake two Monte Carlo experiments to compute the finite-sample size of this DM statistic. We also evaluate the power of the tests to identify leading indicators that help improve the ability of the models to capture the business cycle chronology.

5.1 Assessing QPS Improvements Using Simulations

In order to conduct the experiments the first step is to estimate a model for XLI using the actual data on XLI. The dynamics of XLI are modeled as an AR(2)-GARCH(1,1) model with Student T innovations. We settled on this model after considering a range of AR-GARCH models and as indicated by the Ljung-Box specification tests, it does an admirable job of whitening the residuals (see Table 3 for parameter estimates).

Insert Table 3 About Here

[6] A similar picture emerges when using a one-sided test, though the under-rejection is slightly less pronounced.

We now present two different approaches to evaluating the statistical significance or p-value by Monte Carlo methods. In our first experiment we generate 1000 artificial useless "leading indicators" which have the same time-series dynamics as XLI but contain no information about the business cycle. We simulate sample paths using the estimated parameters from the univariate AR(2)-GARCH-T time-series model reported in Table 3.[7] 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.

For each artificial leading indicator we re-estimate the TVTP MRS model's parameters (using the actual XCI data in each case) and calculate the corresponding QPS – using the NBER dates. Because we use the NBER dates in the calculation of QPS this simulation experiment will provide the empirical distribution of the QPS statistic improvements – in comparison to that obtained using the CTP model for XCI - under the null that the leading indicator is of no value in relation to the NBER business cycle chronology. 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.

The resulting empirical distribution of the QPS improvement is provided visually in the upper plot in Figure 5 (labeled "Actual NBER Dates"). It is clear from the figure that, 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 0.0261 (or 24%), which is indicated by the vertical line, is effectively zero with a *p*-value of 0.001. The evidence from the DM statistic is much less compelling as the *p*-value was only 1.8 percent.

Insert Figure 5 about here

[7] 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.

As an aside, we also calculated the DM statistic on each of these 1000 artificial samples. It will be recalled that the asymptotic distribution of the DM statistic under the null of no predictive improvement is the standard normal. In fact, the asymptotic standard normal distribution provides a very poor approximation to the actual finite sample distribution in this experiment. The Bera-Jarque normality test finds evidence against normality at the 0.0015 level. The DM test under-rejects quite severely as it rejects the null hypothesis at the 5 percent level in less than 1 percent of samples. And this is despite using quite a large sample of 538 observations.

We also compute the *p*-value using an experiment in which we simulate both the coincident and leading indicators. In this case for each of 1000 replications we generate the leading indicator $X_t^{(i)}$ as above using the parameter estimates in Table 3. We also generate the regimes $S_t^{(i)}$ using the same approach as in section 3 but using the parameter estimates from the CTP MRS model reported in Table 1.Finally, rather than using the actual XCI series, here we generate an artificial coincident index, $y_t^{(i)}$, using the relevant CTP model parameter estimates from Table 1. For each simulated triple $y_t^{(i)}, X_t^{(i)}, S_t^{(i)}$ we analyze the model fit by estimating both the CTP model - which is the correctly specified model - and a TVTP model which uses a meaningless leading indicator. 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. The QPS statistics for both the CTP and TVTP models and the DM *Z* statistic are computed using $S_t^{(i)}$ rather than the NBER dates (see bottom graph in Figure 5).

We find that the DM statistic in this experiment exhibits the same size distortion reported previously. In particular, at the 5 percent level, we reject the null hypothesis in only 0.87 percent of samples. The Bera-Jarque statistic for the distribution of the DM statistic across the simulations is 33.79 which is strongly significant with a *p*-value of only 4.6×10^{-8} indicating that the finite sample distribution (with nonetheless over 500 sample observations) of the DM statistic is very different from the asymptotic normal.

Furthermore, our simulations suggest that the calculated p-value of the DM statistic of -2.34 from the real data is actually 0.001 (only 1 of the 1000 simulations were larger) which is very different from the asymptotic p-value of 1.8 percent. This again highlights the size distortions in the DM statistic when applied to MRS business cycle models.

5.2 Power

To understand the ability to detect improvements in model fit when the data are actually generated by varying transition probability models we simulate the leading indicators $X_t^{(i)}$ as above and, using these, we generate the regimes $S_t^{(i)}$ and coincident index $y_t^{(i)}$ from a TVTP model using the parameters reported in Table 1. Using these simulated data we generate Monte Carlo critical values of 0.0115 (1 percent) and 0.0202 (5 percent) which allow us to reject the null hypothesis of no improvement in 92.9 and 97.0 percent of samples. However, we only reject in 83.8 and 91.6 percent of samples using respectively the 1 and 5 percent critical value from the asymptotic distribution of the DM statistic. Both types of tests based on the QPS statistic have good power to detect varying transition probabilities implied in our model of business cycle dynamics. Yet the simulation-based critical values have higher power because of the poor size of the DM statistic using the asymptotic standard normal distribution.

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. One important application of the QPS statistic is to measure the ability of leading indicators to improve the fit of regime switching business cycle models to the NBER chronology. The Diebold-Marino (DM) statistic is a powerful tool widely used to compare two different forecasts of some time-series variable. It is simple to calculate and does a satisfactory job in many empirical applications. However, we have demonstrated that the DM statistic is not well suited to evaluating changes in QPS statistics associated with competing business cycle models as the asymptotic normal distribution in this case is a very poor approximation to the true distribution. The DM statistic tends to under-reject even in samples of as many as 500 or 1000 observations. As an alternative, we present an empirical method to assess the statistical significance of the QPS statistic in the context of Markov-regime switching (MRS) models of the business cycle. The method involves constructing the distribution of the QPS statistic by Monte Carlo simulations 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 alone.

References

Andrews, D.W.K. (1991), "Heterokedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Brier G.W. (1950), "Verification of Forecasts Expressed in Terms of Probabilities", *Monthly Weather Review*, 78, 1-3

Campbell R. Harvey (1989), "Forecasting Economic Growth with the Bond and Stock Markets," *Financial Analysts Journal*, September/October: 38-45.

Chauvet, Marcelle (1999), "Stock Market Fluctuations and the Business Cycle", Working paper, UC Riverside.

Chen, Max (2002), "The Use of the Composite Leading Index for Forecasting Business Cycle Turning Points", *Annual Meeting of the American Statistical Association, Joint Statistical Meetings - Business & Economic Statistics Section*, 509-518.

Diebold, Francis X. and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X. and Rudebusch, R (1989), "Scoring the Leading Indicators", *Journal of Business*, 62, 369 - 402.

Diebold, F.X., Lee, J.-H. and Weinbach, G. (1994), "Regime Switching with Time-Varying Transition Probabilities," in C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*. (*Advanced Texts in Econometrics*, C.W.J. Granger and G. Mizon, eds.), 283-302. Oxford: Oxford University Press.

Filardo, Andrew J. (1994), "Business Cycle Phases and Their transitional Dynamics", *Journal of Business and Economic Statistics*, 12, 299-308.

Filardo, Andrew J and Stephen F. Gordon (1998), "Business Cycle Durations", *Journal of Econometrics*, 85, 99-123.

Gray, Stephen (1996), "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process", *Journal of Financial Economics*, 42, 27-62.

Hamilton, James (1989), "A New Approach to the Analysis of Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Kholodilin, Konstantin, A. and Vincent W. Yao (2005), "Measuring and Predicting Turning Points Using a Dynamic Bi-Factor Model," *International Journal of Forecasting*, 21, 525-537.

Lahiri, Kajal and Jiazhuo G. Wang (1994), "Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model", 13, *Journal of Forecasting*, 245-263.

Layton, Allan P. (1998), "Further on the Influence of Leading Indicators on the Probability of US Business Cycle Phase Shifts", *International Journal of Forecasting*, 14, 63-70.

Layton, Allan P. and Daniel R. Smith (2000), "A Further Note on the Three Phases of the US Business Cycle," *Applied Economics*, 32, 1133-1143.

Newey, W., and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Stock, James H.. and Mark W. Watson (1989), "New Indexes of Coincident and Leading Indicators," *NBER Macroeconomics Annual*, 351-393.

Table 1: Parameter Estimates of the CTP and TVTP MRS Models for XCI.

	CTP Model		TVTP Model	
Parameter	**Estimate**	**Robust SE**	**Estimate**	**Robust SE**
p	0.9727	0.0104		
q	0.9185	0.0310		
μ_1	0.3422	0.0279	0.3484	0.0312
μ_2	-0.1310	0.0964	-0.1433	0.0997
σ_1^2	0.1459	0.0136	0.1448	0.0158
σ_2^2	0.6705	0.1761	0.6515	0.1861
β_{01}			2.4261	0.4311
β_{11}			0.3193	0.1339
β_{02}			2.8984	0.6691
β_{12}			-0.4013	0.1641
LL	-385.3328		-375.9028	% Change:
QPS-Ex Ante	0.1104		0.0844	-24%

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 2: Rejection Frequencies From the Diebold-Mariano (1995) Statistic on MRS Models: Two Sided Test

Model:		1	2	3	4	5	6	7	8	9	10	11	12
p		0.8	0.8	0.9	0.9	0.95	0.95	0.9	0.9	0.95	0.95	0.9	0.9
q		0.8	0.8	0.9	0.9	0.95	0.95	0.9	0.9	0.95	0.95	0.95	0.95
μ_1		2	1	2	1	2	1	2	1	2	1	2	1
μ_2		-2	-1	-2	-1	-2	-1	-2	-1	-2	-1	-2	-1
σ_1^2		1	1	1	1	1	1	1	1	1	1	1	1
σ_2^2		1	1	1	1	1	1	2	2	2	2	1	1
T	A	**Panel A: ρ=0.5**											
250	0.05	0.002	0.055	0.001	0.021	0.002	0.016	0.004	0.033	0.005	0.025	0.000	0.006
	0.01	0.000	0.013	0.000	0.002	0.000	0.002	0.000	0.005	0.000	0.002	0.000	0.001
500	0.05	0.001	0.021	0.004	0.012	0.007	0.016	0.007	0.028	0.013	0.018	0.000	0.000
	0.01	0.000	0.003	0.000	0.001	0.000	0.002	0.000	0.002	0.001	0.000	0.000	0.000
1000	0.05	0.000	0.022	0.006	0.014	0.011	0.017	0.012	0.025	0.019	0.018	0.000	0.002
	0.01	0.000	0.003	0.000	0.000	0.000	0.002	0.001	0.002	0.001	0.001	0.000	0.000
		Panel B: ρ=0.9											
250	0.05	0.001	0.034	0.005	0.016	0.001	0.02	0.004	0.024	0.008	0.020	0.000	0.003
	0.01	0.000	0.008	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.001	0.000	0.002
500	0.05	0.004	0.017	0.004	0.017	0.009	0.012	0.006	0.018	0.016	0.017	0.000	0.002
	0.01	0.000	0.005	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.003	0.000	0.000
1000	0.05	0.001	0.023	0.004	0.012	0.012	0.012	0.007	0.018	0.015	0.017	0.000	0.000
	0.01	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.000	0.000

Table 3: Parameter Estimates for AR(2)-GARCH Model for XLI

	Parameter	Robust SE
ϕ_0	0.3393	0.0725
ϕ_1	1.1424	0.0477
ϕ_2	-0.2452	0.0500
σ^2	1.2366	0.3323
α	0.0454	0.0526
β	0.9381	0.0906
v	0.1086	0.0349
LL	-809.8408	
LB(z,12)	13.0500	0.3654
LB(z^2,12)	18.3271	0.1061

Notes: The conditional mean is modeled as $\mu_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2}$ and the conditional variance is modeled as $h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}$ where $e_t = y_t - \mu_t$, and the last term captures asymmetry in the conditional volatility specification. We report the unconditional volatility σ^2, which is given by $\sigma^2 = \omega/(1-\alpha-\beta)$, rather than ω directly. The standardized residual $z_t = e_t / h_t^{0.5}$ is a standardized T random variable with v^{-1} degrees of freedom. 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.

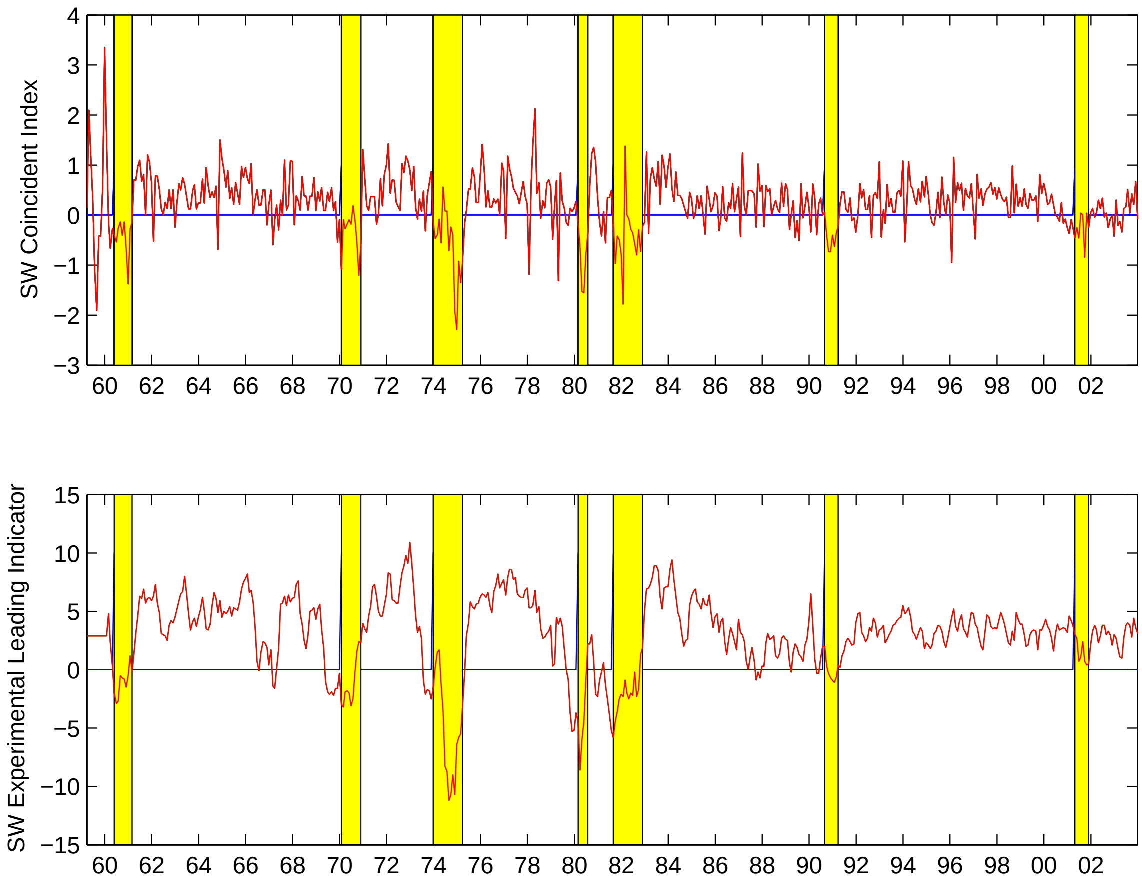

Figure 1: Plot of underlying data.
The first plot is of month-on-month growth rates in the Stock-Watson Experimental Coincident Index. The second plot is the previous month's value of the experimental leading indictor. The shaded areas denote recessions as identified by the NBER.

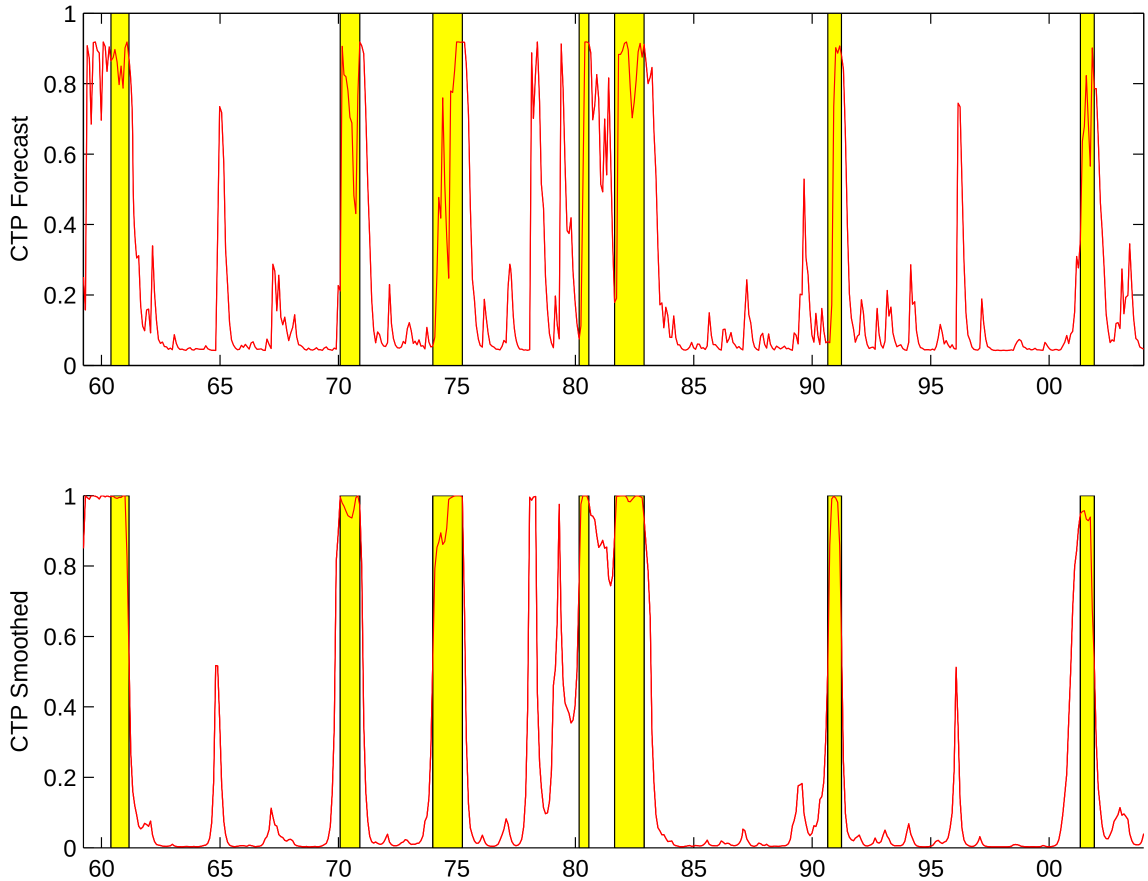

Figure 2: Plot of forecast and smoothed probabilities of a recession from MS model with CTP. The shaded areas denote recessions as identified by the NBER.

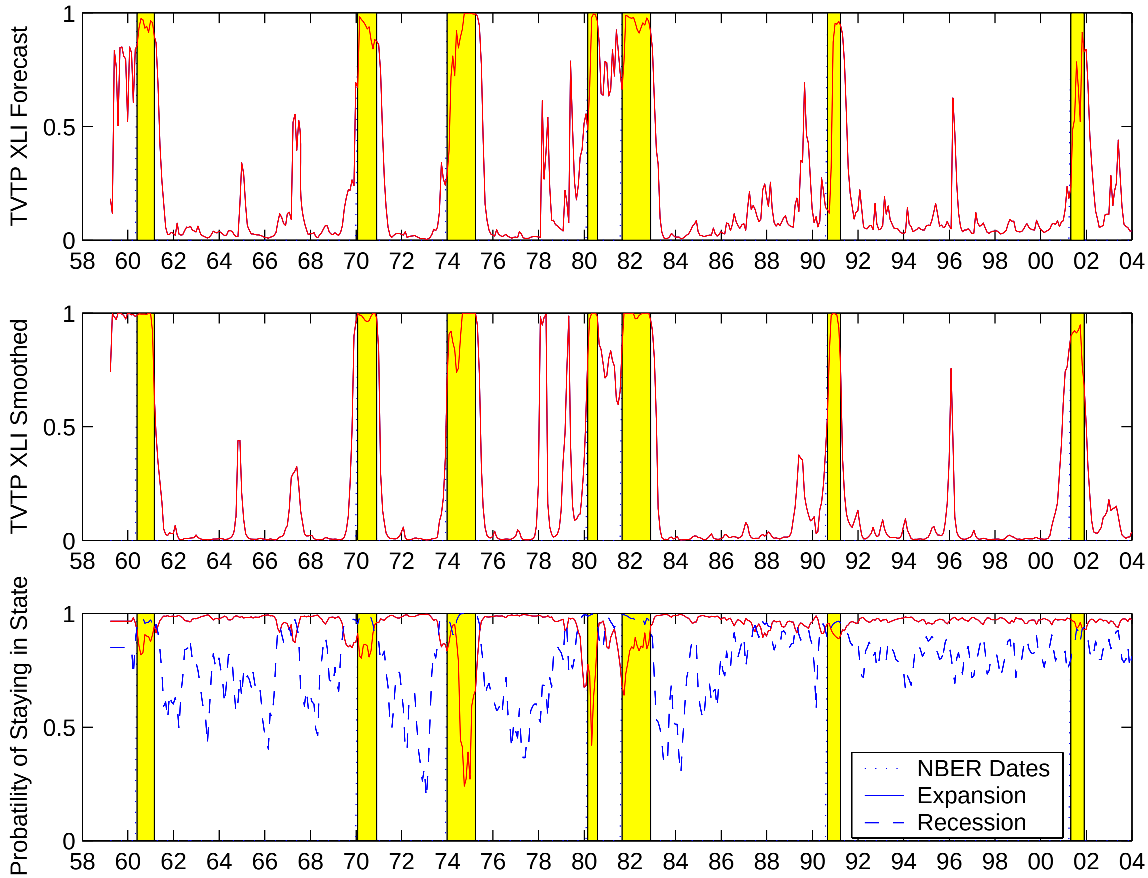

Figure 3: Plot of forecast and smoothed probabilities of recessions from MS model with TVTP using Stock and Watson's Experimental Leading Indicator. The shaded areas denote recessions as identified by the NBER.

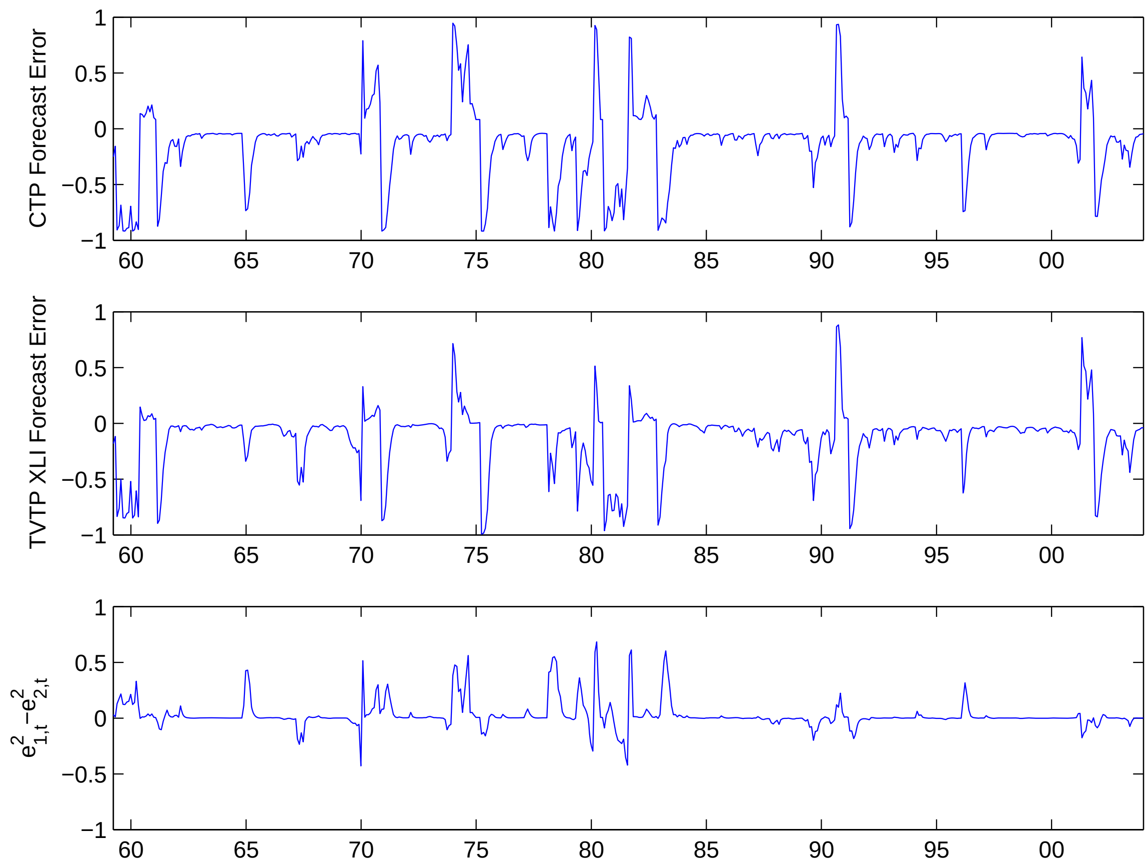

Figure 4: Time series plot of the forecast error (the difference between the recession dummy and the forecast probability of a recession next period) from a constant transition probability model (top panel), the forecast error from a time-varying transition probability model using the experimental leading indicator XLI (middle panel), and the difference between the two squared forecast errors (lower panel).

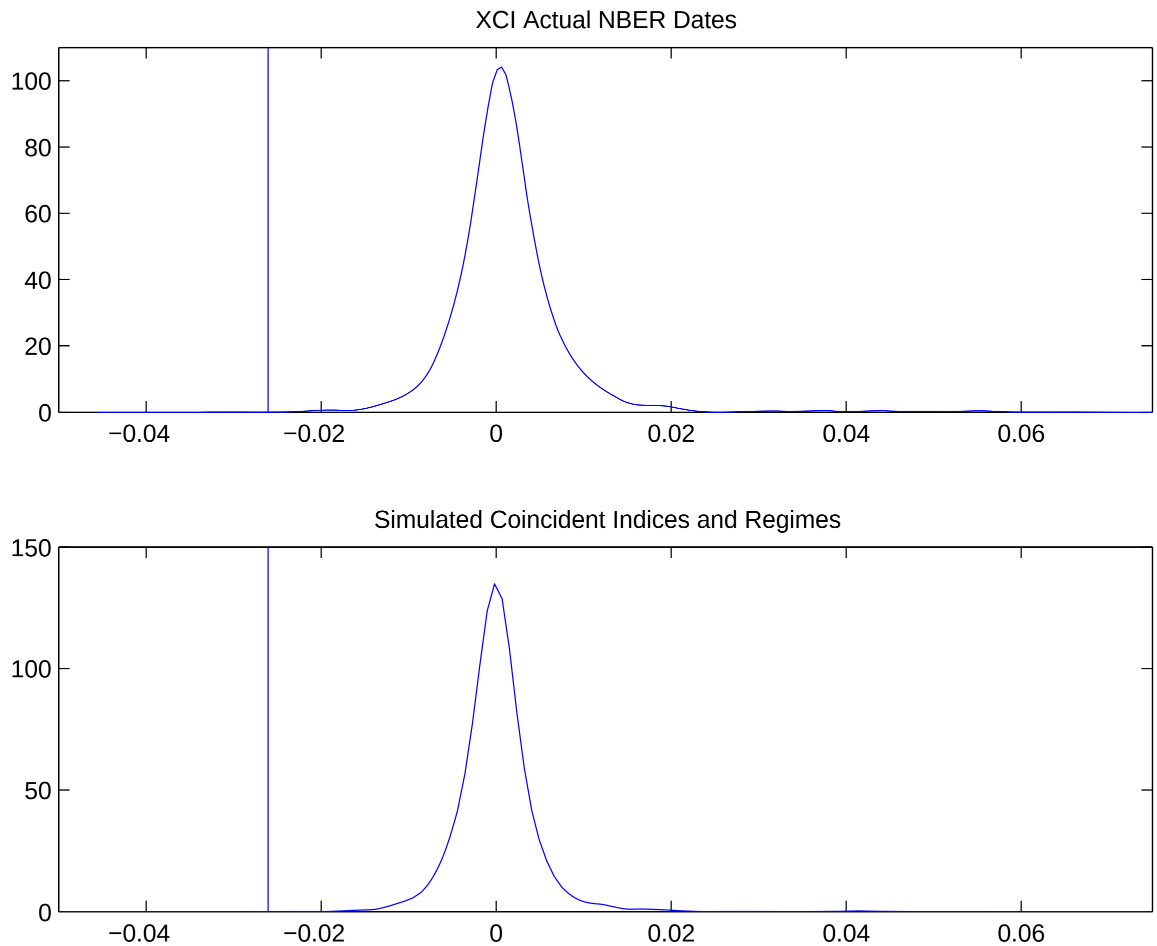

Figure 5: Plot of QPS improvement from TVTP when models estimated for artificially generated series for XLI.
The figure provides a Kernel-based nonparametric estimate of the density of the difference between the QPS from a CTP and TVTP MRS models where the leading indicators have no informational content as far as XCI is concerned. The upper plot uses XCI and actual NBER dates with artificial leading indicators. The lower panel uses artificial leading indicators, regimes and coincident indices. The vertical lines represent the actual improvement we found using XLI.