Reproducing Business Cycle Features: How Important Is Nonlinearity Versus Multivariate Information?

James Morley, Jeremy Piger, and Pao-Lin Tien*

First draft: July 17th, 2008

Abstract

Model evaluation has always been on the forefront of economic research. As modeling techniques advance over time, a wide variety of models have sprung up to satisfy the different needs of economists. It is therefore important to establish an efficient and reasonable approach to model comparison and evaluation, including when models are nonnested. In this paper, we consider the ability of time-series models to generate simulated data that display the same business cycle features which characterize U.S. GDP. Our analysis of multivariate linear models and univariate linear and nonlinear models allows us to investigate the extent to which multivariate information can account for the apparent univariate evidence of nonlinear dynamics in GDP. We find that certain nonlinear specifications yield an improvement over linear models in reproducing business cycle features, even when multivariate information is taken into account in some of the linear models.

JEL Classification:

* James Morley, Washington University in St. Louis; Jeremy Piger, University of Oregon; Pao-Lin Tien, Wesleyan University.

1. Introduction

Model evaluation has always been on the forefront of economics research. As modeling techniques advance over time, a wide variety of models have sprung up to satisfy the different needs of economists; from simple univariate and multivariate linear models to more complicated univariate and multivariate nonlinear models. It is therefore important to establish an efficient and reasonable approach to model comparison and evaluation. A recent paper by Morley and Piger (2006) takes on the issue of model comparison for U.S. real gross domestic product (GDP). The authors use a new business cycle dating algorithm that allows them to consistently evaluate a variety of univariate linear and nonlinear models in terms of their ability to produce simulated data that match the business cycle features exhibited by GDP. In this paper, we further the analysis in Morley and Piger (2006) by considering multivariate linear models. We want to investigate the extent to which multivariate information can account for the apparent univariate evidence of nonlinear dynamics in U.S. GDP demonstrated in Morley and Piger (2006).

This approach to model comparison is fairly new. The conventional method of conducting model evaluation is through hypothesis testing. However, when the models under consideration are non-nested – that is, one model is not simply a restricted version of another – straightforward comparison using hypothesis testing is often intractable. Employing out of sample forecasts to compare the models is another possibility, though the results are often very sensitive to the particular out of sample period used.[1] The business cycle features approach provides a useful alternative to these conventional methods. It can be viewed as related to a broader approach to model comparison known as "encompassing tests." Under this approach, there is no need to worry about whether the models under consideration are nested or not or what sample periods are being used, because competing time-series models are evaluated based on their ability to produce simulated data that can reproduce business cycle characteristics of actual U.S. GDP over any given sample period. The focus on business cycle features is very natural. Ever since Burns and Mitchell's (1947) extensive study of the cyclical behavior of economic activity, economists have sought to analyze economic fluctuations in terms of business cycle phases. This also provides a very intuitive way to assess the benefit of introducing nonlinearity into time-series models, as many of the nonlinearities explored for GDP have been motivated as related to the business cycle. One can also view this method of model evaluation as complementary to the more traditional methods. For example, if several non-nested models – such as an ARIMA model and a Markov-switching model – manage to pass the battery of conventional diagnostic tests and are equally favored, then these models' ability to produce simulated data that can match the business cycle features of GDP could help researchers make the difficult choice of which model to use based on an in-sample analysis.

[1] The Bayesian marginal likelihood approach to model comparison is also feasible, though this method may be computationally intensive at times, and more importantly, it is often quite sensitive to the specification of the priors. Furthermore, the Bayesian approach only provides us with a sense of the relative performance of different models and not an absolute measure of the ability of the models to explain the dynamics in the data. In contrast, the business cycle features encompassing test approach does provide an absolute measure.

Before we can present more detail about how to evaluate the performance of different models in reproducing business cycle features, we must first define what we mean by the business cycle. Note here that under this framework, "cycle" refers to the *classical* business cycle (or *reference cycle*) as described by Burns and Mitchell (1947) rather than the *cyclical component* of a series obtained after detrending the series. According to Burns and Mitchell (1947), the business cycle can be defined as a series of distinct phases in economic activity, with the phases corresponding to recession and expansion. The turning points of the phases are indicated as peaks and troughs. The general practice in the literature on business cycle features is to follow a model free algorithm that identifies peak and trough dates. Based on these dates, standard business cycle features, such as the average length of phases, cumulative growth during the phases, etc. are computed for the actual U.S. GDP data. Then, using the same algorithm, the corresponding set of business cycle features are computed for simulated data from a model in order to evaluate the ability of the model to reproduce features in the sample data.

Besides Morley and Piger (2006), a number of other papers in the literature have also employed this approach to assess the performance of different time-series models, including Hess and Iwata (1997), Harding and Pagan (2002), Galvão (2002), and Clements and Krolzig (2004) for U.S. data, and Demers and Macdonald (2006) for Canadian data. In the plethora of univariate and multivariate linear and nonlinear models that Hess and Iwata (1997), Harding and Pagan (2002), and Clements and Krolzig (2004) have considered, the simple linear ARIMA(1,1,0) or ARIMA(2,1,0) models always manage to reproduce business cycle features of actual real GDP just as well as, if not better than, their more complicated counterparts. Following the principle of parsimony, all three papers draw the conclusion that researchers should pick the simpler models over more complicated models, ceteris paribus. However, Galvão (2002), Morley and Piger (2006), and Demers and Macdonald (2006) find that while none of the models being considered dominates over all features, there are some features that nonlinear models are better at capturing than linear models. Hence there is added benefit and relevance for taking into account nonlinearity in time-series models.

Differences in the results reported could be due to the slight variations in the definition of the business cycle, the algorithm used to calculate the different phases of the cycle, or the set of business cycle features considered. Most papers follow the BBQ algorithm of Harding and Pagan (2002), which is a quarterly version of the BB algorithm for monthly data that Bry and Boschan (1971) developed. But in Morley and Piger (2006), the authors improved upon the BBQ algorithm by optimizing on the threshold values that indicate turning points. This modified algorithm MBBQ does a better job at matching National Bureau of Economic Research's (NBER) business cycle dates than BBQ when applied to U.S. real GDP. For that reason, we have chosen to follow in the footsteps of Morley and Piger (2006) and adopt the MBBQ algorithm for our analysis.

To keep a consistent framework for assessing whether the richer information contained in multivariate linear models can improve upon the deficiencies of univariate linear models and outperform the favored nonlinear model in Morley and Piger (2006),[2] we will be looking at the same set of business cycle features that they considered. The features include the mean and standard deviation of growth rates observed during Expansion and Recession phases, and the mean and standard deviation of the length of phases.[3] In addition, the Expansion phase is divided into a Recovery phase, defined as the first four quarters immediately following the Recession, and a Mature Expansion phase, defined as the remainder of the Expansion. Business cycle features for the Recovery and Mature Expansion phases are reported separately. Correlation between the cumulative growth observed during Recessions and that observed in the subsequent Recovery phase will also be considered.[4]

There are, of course, a wide range of multivariate linear models that are worthy of a detailed investigation. However, we have decided to focus our attention on just two widely used and very general classes of multivariate linear models: vector autoregression

[2] Only Clements and Krolzig (2004) have systematically compared univariate models against multivariate models, and they find that multivariate models do not do very well in terms of matching business cycle features of actual U.S. real GDP. Their business cycle dating algorithm is somewhat unconventional though, as they modified the BBQ algorithm such that it does not impose a minimum length requirement for business cycle phases, which may cause it to understate the duration of phases.

[3] Other papers in the business cycle features literature did not consider standard deviations, though we feel that they are an important feature to look at because they capture the substantial heterogeneity of business cycles.

[4] This feature is again not considered in other papers in the literature. But this is a business cycle feature that was central to Milton Friedman's (1964, 1993) analysis of the U.S. business cycle.

models (VAR) and vector error correction models (VECM). Specifically, in terms of VAR models, we consider the two variable model of Blanchard and Quah (1989) and the four variable model in Ahmed, Levin, and Wilson (2004). In terms of VECM models, we consider the three variable model in King, Plosser, Stock, and Watson (1991) that assumes two cointegrating relationships between consumption and output and investment and output. These three models are of particular interest to us because they are probably some of the most widely cited multivariate models in the economics literature, and are specifically designed to explain aggregate economic fluctuations.

Blanchard and Quah (1989) looked at the dynamic effects of aggregate demand and supply disturbances on the gross national product (GNP) by using GNP growth and unemployment rate in their VAR system. Ahmed, Levin, and Wilson (2004) investigated the source of the reduction in the volatility of GDP growth since 1984, and in their VAR system they included GDP growth, inflation, commodity price inflation and the federal funds rate. A very similar VAR model to that used in Ahmed, Levin, and Wilson (2004) was also implemented in Stock and Watson (2002) and Boivin and Giannoni (2006). The VECM in King et. al. (1991) is a classic model for looking at the importance of productivity shocks on economic fluctuations. The authors claim that their analysis applies to a wide class of real business cycle models and is superior to the bivariate VAR in Blanchard and Quah (1989). They included private GNP (y), consumption (c), and investment (i) in their system with ($c - y$) and ($i - y$) as the theoretical error-correction terms.[5]

In addition to investigating whether multivariate information matters in this context, an additional element we are adding to our analysis is the specification for the residuals while simulating data from the models of interest. Following the convention in the literature, we neglect parameter uncertainty in our simulations. Thus, the only source of variation across simulations arises from the residuals, which, in most of the literature, are assumed to be normally distributed. This parametric specification for the residuals might be improved upon by using a semi-parametric bootstrap approach – that is, shuffle the original residuals from the model estimation and then draw from this pool of residuals with replacement in order to construct the simulated series. This is a more general approach as we make no stringent assumption about the specific distribution of the residuals (residuals are non-parametric). If the true residuals are not normally distributed, this approach should improve the performance of the models in terms of the simulated data's ability to reproduce business cycle features.

Another important characteristic of the GDP data that previous research on the topic has largely ignored is the reduction in variance in the series starting around 1984. Economists have long noticed the reduction in volatility, though it is Kim and Nelson

[5] In the structural VAR literature, the type of identification method used is of vital importance. Blanchard and Quah (1989) implemented long-run restrictions while Ahmed, Levin, and Wilson (2004) used short-run restrictions. However, for the purpose of simulating data and calculating business cycle feature required here, identification of structural shocks is irrelevant. What matters is the variables included in each VAR or VECM model and the reduced form dynamics generated by the models.

(1999) and McConnell and Perez-Quiros (2000) who first provided rigorous empirical analysis of the phenomenon. Taking the reduction in volatility into account is especially important for nonlinear models such as the Kim, Morley, and Piger (2005) bounceback model (the favored model in Morley and Piger 2006) that we consider in this paper, as regime switching in mean can generate patterns that look similar to a structural break in volatility or other forms of heteroskedasticity that may be present in the sample data. As a robustness check on our results, we will compare the feature reproduction performance of the time-series models by allowing for a break in variance in GDP in 1984Q1.

The analysis in this paper shows that multivariate information does not appear to improve the performance of linear models over nonlinear models. All the models, be it univariate or multivariate, linear or nonlinear, are fairly well adapted at reproducing business cycle features such as the average growth rate of Recession and Mature Expansion phases, the average number of phases, and the average length and standard deviation of phases. However, the linear models have serious problems replicating the average growth rate of the Recovery phase, and the variation in the growth rate of both the Recession and Recovery phases. Most importantly, no linear model comes close to matching the strong negative correlation between the cumulative growth rate of the Recession phase and the Recovery phase of the cycle. This is where the nonlinear models triumph. Not only can the nonlinear models capture the higher than average growth rate during the Recovery phase, they are also able to generate a large enough negative correlation between the cumulative growth rates of the Recession and Recovery phases to match that exhibited by real GDP.

Results from the use of non-parametric versus parametric residual specifications are mixed. For most linear models, using non-parametric residuals helped their performance, but this is not the case for the nonlinear bounceback model or the linear VECM model. With non-parametric residuals, the linear models are better at replicating the standard deviation of growth rates of the Recovery or the Mature Expansion phases.

Finally, after taking into account the break in variance of U.S. real GDP in 1984Q1, all the multivariate linear models perform much better, especially in their ability to generate a strong enough negative correlation between the cumulative growth in the recession and recovery phases. This surprising improvement in the multivariate linear models prompted us to conduct a counterfactual experiment to see if the structural break alone drove the results rather than the dynamics generated by the multivariate linear models. The counterfactual analysis is based on the idea that if the multivariate linear models can in fact generate dynamics that capture this negative correlation between cumulative growth in recessions and recoveries, then just applying the estimated pre (or post) structural break date parameters of the multivariate linear models to simulate data for the whole sample period should still allow the models to generate a strong enough negative correlation between the growth rates in recessions and recoveries. Evidence shows that this is not the case, hence it is highly likely that the structural break is the main cause for the better performance of the multivariate linear models. Even if we take the results with break in variance at face value, however, the bounceback model without considering a break in variance still does a better job at matching business cycle features of real GDP than the multivariate linear models.

These results lead us to conclude that certain nonlinear specifications do yield improvement over linear models in reproducing business cycle features even when multivariate information is taken into account for the linear models.

The remainder of this paper will proceed as follows: Section 2 details the business cycle algorithm used to establish business cycle turning points in U.S. real GDP. The peak and trough dates calculated by the MBBQ algorithm will be compared to the dates from the BBQ algorithm as well as those established by the NBER. Section 3 defines the business cycle features we consider and documents these features for the U.S. real GDP. Section 4 specifies the time-series models under consideration and then evaluates the ability of the competing univariate and multivariate models to reproduce business cycle features exhibited by GDP. Finally, Section 5 concludes.

2. Business Cycle Dating Algorithm

Official business cycle dates – the peaks and troughs in the economy that define recessions and expansions – in the U.S. are determined by the NBER, a private, nonprofit, nonpartisan research organization founded in 1920. Within the NBER, the Business Cycle Dating Committee plays the key role in establishing business cycle dates. The committee reviews a variety of economic statistics and indicators of U.S. business conditions before deciding on the exact turning points in the economy. Given this set of official business cycle dates, it seems natural to use them as the benchmark for calculating business cycle features. However, the NBER chronology is only relevant for the actual U.S. GDP sample data, and not for the simulated data from the time-series models we are considering. Therefore, to establish turning points in the sample data and simulated data in a consistent fashion, we need to use a formal procedure capable of mimicking the NBER decision-making process as best we can.

The standard approach to establishing business cycle turning points in the literature is to use the Bry-Boschan Quarterly (BBQ) algorithm by Harding and Pagan (2002). The specifics of the algorithm can be summarized as follows:

Step 1: Using the log level of U.S. quarterly real GDP (y_t), establish candidate dates of peaks and troughs as local maxima and minima in the data such that a peak occurs at time t if:

$$y_{t-2} - y_t < 0; \quad y_{t-1} - y_t < 0; \quad y_{t+1} - y_t < 0; \quad y_{t+2} - y_t < 0,$$

and a trough occurs at time t if:

$$y_{t\text{-}2} - y_t > 0; \quad y_{t\text{-}1} - y_t > 0; \quad y_{t+1} - y_t > 0; \quad y_{t+2} - y_t > 0.$$

Step 2: Censor the turning points to ensure that peaks and troughs alternate. In the case of two consecutive peaks (troughs), eliminate the peak (trough) with the lower (higher) value of y_t.

Step 3: Censor the turning points to ensure that each business cycle phase (peak-to-trough and trough-to-peak) lasts a minimum of five quarters.

The peak and trough dates established by the NBER for the sample period 1948Q4 to 2007Q4,[6] along with the dates established by the BBQ algorithm applied to quarterly U.S. real GDP are reported in Table 1. The BBQ algorithm does an adequate job of matching the official NBER peak and trough dates. It identifies eight of the nine peaks and nine of the ten troughs reported by NBER. Only two of the peak dates differ from NBER dates by a quarter, and five of the trough dates differ by one to three quarters. It is interesting that all the errors made by the BBQ algorithm shift the turning points

[6] Even though U.S. real GDP data are available as early as 1947Q1, we choose to start our sample at 1948Q4. As a result, we have to ignore the first NBER peak date (1948Q4) in our evaluations of the BBQ and later MBBQ algorithms, as the earliest start date at which the algorithms can identify a turning point is 1949Q2. There are a couple reasons for shortening the sample period. First, this facilitates comparisons of our results with the results in Morley and Piger (2005), which used a sample starting from 1948Q4. Second, and most importantly, starting the sample at 1947Q1 seems to create problems for both the BBQ and MBBQ algorithms. If we start the sample at 1947Q1, we will have two consecutive quarters of decline followed by two consecutive quarters of increase in GDP right at the start of the sample. This not only causes the BBQ algorithm to pick up an extra trough date in 1947Q3, but it also throws off the precision of both dating algorithms in terms of their ability to produce trough dates that match those reported by the NBER. We believe that this is due to the interaction of the 1947Q1 observation with the minimum phase length and censoring requirements in Steps 2 and 3 of the algorithms. Hence, we believe that shortening the sample period by 7 quarters is worthwhile in order to make the algorithms more precise.

forward in time relative to the official NBER dates. This systematic error suggests that Step 1 of the BBQ algorithm may be modified to correct for it, which is what the MBBQ algorithm attempts to do. So, for MBBQ, step 1 of the algorithm can be stated as follows:

Step 1: Using the log level of U.S. quarterly real GDP (y_t), establish candidate dates of peaks and troughs as local maxima and minima in the data such that a peak occurs at time t if:

$y_{t-2} - y_t < \alpha_1$; $y_{t-1} - y_t < \alpha_1$; $y_{t+1} - y_t < \alpha_2$; $y_{t+2} - y_t < \alpha_2$,

and a trough occurs at time t if:

$y_{t-2} - y_t > \alpha_3$; $y_{t-1} - y_t > \alpha_3$; $y_{t+1} - y_t > \alpha_4$; $y_{t+2} - y_t > \alpha_4$.

This differs from BBQ in that the threshold parameters that signal turning points are allowed to deviate from 0. The thresholds are also allowed to vary from peak to trough and on different sides of the turning points. To determine the values of the α_i's, $i = 1, 2, 3, 4$, a grid search is conducted for values between -0.005 and 0.005, i.e. $\alpha_i \in (-0.005, 0.005)$. For each possible combination of the α_i's in the grid, a mean squared error (MSE) is calculated as:

$$MSE(\alpha_i) = \sqrt{\frac{\sum_{t=1}^{T} [MBBQ_t(\alpha_i) - NBER_t]^2}{T}},$$

where $NBER_t$ = 1 if quarter t is an NBER recession quarter and $NBER_t$ = 0 otherwise, while $MBBQ_t(\alpha_i)$ = 1 if quarter t is a recession quarter according to the MBBQ algorithm with threshold values α_i, and $MBBQ_t(\alpha_i)$ = 0 otherwise. The α_i's that minimizes MSE(α_i) are chosen to be the final threshold values for the algorithm. In the case of ties, α_i's that are closest to 0, as measured by $\sum_{i=1}^{4}|\alpha_i|$, are chosen.

The turning point dates established by the MBBQ algorithm are reported in Table 1 as well. Threshold values chosen for this sample period are: $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0.001$, $\alpha_4 = -0.002$. It is clear from Table 1 that the MBBQ algorithm offers substantial improvement over the BBQ algorithm, especially on the trough dates. It identifies the same number of peaks and troughs as the BBQ algorithm, though only two of the peak dates and two of the trough dates deviate by a quarter from the official NBER dates.

Observant readers will notice that both the BBQ and MBBQ algorithms miss the peak and trough dates identified by NBER in 2001. This was not the case in Morley and Piger (2006). Upon closer inspection of the data, we found that due to data revision in 2004, the U.S. real GDP output growth rate for 2001Q2 has changed from negative to positive.[7] As both dating algorithms require two quarters of decline or increase on both sides of turning points, this revision in GDP data implies that neither algorithm would be able to pick up any peaks or troughs in 2001. The data revision diminishes the ability of

[7] According to the St. Louis Fed Archival Federal Reserve Economic Data (ALFRED), U.S. real GDP (GDPC1) with a vintage date of June 25th 2004 still report a negative growth rate for 2001Q2, but in the next vintage (July 30th 2004) the same growth rate has been revised to a positive number.

the dating algorithms to mimic actual NBER chronology. However, given that both BBQ and MBBQ still do fairly well in picking out turning point dates that match up with the NBER dates prior to 2001, we believe that this problem is not serious enough for us to abandon the use of these algorithms.[8]

3. Business Cycle Features in U.S. Real GDP Data

The business cycle phases are defined as follows: (1) Recession – the quarter following a peak date to the subsequent trough date, (2) Expansion – the quarter following a trough date to the subsequent peak date, (3) Recovery – the first four quarters of the expansion phase, and (4) Mature Expansion – the remaining quarters of an Expansion phase following the Recovery phase.

Given this definition of phases, we consider the following business cycle features for any given realization of data:

- Number of business cycle peaks

[8] There is ample evidence that 2001 remains a recession phase despite the revision in GDP data. In the most recent memo released on January 7th, 2008 by the NBER Business Cycle Dating Committee, there is no mention of possibly revising the 2001 peak and trough dates. Also, even though the 2001 recession is no longer obvious from the level of the GDP series alone, it is still apparent in other series such as employment (total nonfarm payroll). In addition, nonlinear Markov-switching type models like the Kim et. al. (2005) bounceback model we consider here still identify 2001 as a recession episode with the updated GDP data.

- Average and standard deviation of Recession and Expansion phase lengths
- Average and standard deviation of annualized quarterly growth rates in Recession, Expansion, Recovery, and Mature Expansion phases
- Correlation between the cumulative decline during a Recession and the cumulative growth in the subsequent Recovery phase.

Table 2 presents the values of these business cycle features for quarterly U.S. real GDP data from 1948Q4 to 2007Q4 using turning points established by the NBER, the BBQ algorithm, and the MBBQ algorithm. As mentioned in the previous section, the dating algorithms failed to identify the 2001 peak and trough dates established by the NBER, so at first glance it would appear that the dating algorithms do a terrible job at replicating the NBER sample features. However, one should treat this more as an illustrative exercise to see which algorithm does a better job at matching the business cycle features exhibited by real GDP using the official NBER turning point dates. The results here corroborate with what we observe in Table 1, MBBQ does a superior job at matching the NBER sample feature values than BBQ because it replicates NBER turning points better. In all but four cases (average quarterly growth rates of the Expansion phase, average length of the Expansion phase, and the variation in the average length of Recession and Expansion phases) MBBQ produce feature values that are closer to the NBER sample features.

Typically, for the purpose of assessing the performance of the time-series models in replicating business cycle features, we compare the simulated features with the sample features produced by the dating algorithm (since the dating algorithm is used to produce turning points in the simulated data). However, due to complications with missing the 2001 peak and trough dates, we feel that it would be unfair to compare simulated features with the MBBQ sample features because the time-series models are designed to replicate behavior of actual GDP with NBER recessions and expansions, and there are some large differences between the sample features using NBER turning points and those produced by MBBQ.[9] Hence, we choose to compare simulated features with the NBER sample features instead.

Before concluding this section, there are a few things worth mentioning regarding the NBER sample features reported in Table 2. First, as one would expect, average quarterly growth rates differ quite a bit between the Recession and Expansion phases. Recessions are associated with negative growth rates, averaging around –1.9% per quarter, while Expansions are associated with positive growth rates close to 4.6% per quarter. When the Expansion phase is divided up into Recovery and Mature Expansion

[9] It is worth explaining the reasons for a few striking differences in the sample features of NBER compared with the BBQ or MBBQ algorithms. First of all, average quarterly growth rates of Recession phase using NBER dates are much smaller in absolute value than those produced by the algorithms because the algorithms happen to miss a few recession quarters with positive growth rates, and the 2001Q2 to 2001Q4 recession which the algorithms miss completely is mild compared to previous recessions. Second, the average quarterly growth rate for the Expansion phase is much higher using the algorithms than NBER dates because the 1991Q2 to 2001Q1 expansion that the algorithms failed to identify has much lower average growth rate than previous expansions. Finally, the algorithms produce average Expansion phase lengths that are much shorter than the NBER sample feature. Again, this is because the length of the 1991Q2 to 2001Q1 expansion (40 quarters) is not included in the calculation, which happens to be the longest of all expansions.

phases, it is striking to see that the average growth rate associated with the Recovery phase is almost twice as large as those reported for the Mature Expansion phase. Second, there is a large difference between the average length of the Recession and Expansion phases. Expansions appear to last nearly six times as long as Recessions. Third, the variability associated with the Recovery phase is much higher than for other phases in terms of the average quarterly growth rates. This high variability also applies to the average length of Expansion phase. Finally, there is strong negative correlation between the cumulative growth in a Recession phase and the cumulative growth in the subsequent Recovery phase. This corroborates the observation made in Friedman (1964, 1993).

4. Business Cycle Features in Simulated Data from Time-Series Models

4.1. Univariate Model Description

Two different univariate models are considered in this paper. First is the linear AR(2) model that has been found to do quite well in terms of matching features in the business cycle features literature, and is the preferred model in Clements and Krolzig (2004). Second is the Kim, Morley, and Piger (2005) bounceback model, a nonlinear model with Markov-switching parameters. This version of the bounceback model is termed BBV indicating that this particular specification will be able to depict V-shaped recessions.[10] The key difference between the bounceback model and the standard Hamilton (1989) 2-state Markov-switching model is that it would be able to capture a high-growth recovery phase in the first six quarters following the end of recessions. Furthermore, the strength of this high-growth recovery phase is related to the severity of the previous recession, as measured by its length up to six quarters. The BBV was the best performing time-series model in Morley and Piger (2006), beating even the 3-state Markov-switching model of Boldin (1996), which was also designed to capture high-growth recovery business cycle phases.

The specification and estimates of the two time series models for quarterly U.S. real GDP are presented in the appendix, section 2.6. The reported estimates are what we used to calibrate the data generating process in our Monte Carlo simulations that will be used for business cycle feature comparisons later on.

4.2. Multivariate Model Description

As mentioned in the introduction, we consider three different multivariate models. The two variable VAR model of Blanchard and Quah (1989) (B&Q), the four variable

[10] V-shaped recession refers to recessions exhibiting "sharpness," a term introduced by McQueen and Thorley (1993). A sharp series has the transition from contraction to expansion occurring more rapidly than the transition from expansion to contraction. This feature results in the level series being more rounded at peaks than at troughs.

VAR model in Ahmed, Levin, and Wilson (2004) (ALW), and the three variable VECM in King, Plosser, Stock, and Watson (1991) (KPSW). The specifications and estimates used for the Monte Carlo simulations of the multivariate linear models are presented in section 2.6, the appendix.[11] The motivations for considering these particular models have been explained in the introduction, so we will not elaborate further. However, we believe it is worthwhile to detail the main ideas behind the two classes of multivariate linear models we consider here.

Ever since Sims (1980) brought VAR into the spotlight in the economics literature, it has become a very popular econometric tool for economists to study the effect of different types of shocks on the economic variables of interest. The basic idea behind the VAR is that we assume all variables in the VAR system are endogenous, so each variable can be written as a linear function of its own lagged values and the lagged values of all the other variables in the system, plus an error term. Note that the errors of each linear function are assumed to be serially uncorrelated. There is much evidence showing VARs produce better forecasts than large structural models, though some have argued that the VAR is simply an overfit reduced-form version of a simultaneous equations model (see Hamilton 1994). Given the appealing forecast ability of VAR models and the simplicity

[11] Data used for estimation of the multivariate models vary from those used in the original paper on occasions. If the original model used an output variable that is not real GDP (for example, Blanchard and Quah 1989 used real gross national product), we replace that with real GDP in our estimation. As for the other variables used in the models, we try to stay as close to those used in the original study as possible. The estimation sample periods for the multivariate models all start later than 1948Q4 for a variety of reasons, sometimes it is due to data availability, sometimes it is because of the number of lags the estimation requires, and sometimes it is both. We try to use the longest possible sample here to produce the parameter estimates. Note that we continue to simulate data from 1948Q4 to 2007Q4 even though the estimated parameters are produced with varying sample lengths.

of its structure, we believe it is an important class of multivariate models to consider in our analysis here.

As for the VECM, one can view it as a multivariate version of the error correction model that Davidson, Hendry, Srba, and Yeo (1978) popularized. If the levels of the variables in a VAR system are cointegrated, then a VAR representation for the first differences of the variables may no longer be appropriate. Instead, a VECM representation should be used. Under the VECM specification, the first difference of the levels variables in the system can each be written as a function of lagged values of all the difference variables within the system, plus lagged values of the error correction terms and an error term. The error correction terms are the long-run relationships among the variables that arise from economic theory. For example, in King, Plosser, Stock and Watson (1991), the authors make use of the balanced growth theory[12] to motivate their use of the $(c - y)$ and $(i - y)$ error correction terms. These are called "error correction terms" because they reflect the current "error" in achieving long-run equilibrium. As an important extension of the VAR, this is a class of multivariate model that we should include in our analysis.

In the next subsection we use the estimated parameters reported in the appendix to simulate artificial real GDP series from 1948Q4 to 2007Q4, using the actual value of real

[12] Solow's (1970) balanced growth theory suggests that per capita consumption, investment, and output all grow at the same rate in steady state.

GDP in 1948Q4 as an initial value. For each model, we perform 10,000 simulations, saving the business cycle features for each simulation.

4.3. Business Cycle Features from Univariate Models

Table 3 presents the percentiles of sample values for business cycle features in terms of the simulated distributions of these features for the univariate models. These percentiles are based on 10,000 simulations, and they represent the proportion of simulated features that fall below the corresponding sample feature reported in Table 2 for actual real GDP using the NBER turning point dates. The percentiles provide us with a sense of how likely the univariate models could have produced a sample value for a particular business cycle feature as large or as small as that exhibited by the actual GDP data. Bold percentiles in the table indicate that the percentiles are less than 0.10 or greater than 0.90, implying that it was unlikely that the particular univariate time-series model could have simulated data that replicates the behavior of actual GDP for that particular feature. The numbers in parentheses in the table report the difference between a sample feature and the corresponding median simulated feature. This gives us a sense of whether a percentile is driven by closeness of the distribution in matching the sample feature or by a large dispersion of the simulated distribution.

Let us start with the AR(2) model with parametric residual draws (second column of Table 3). The model does a reasonably good job at replicating the features related to the number or length of phases. However, the large difference between the median value in the simulated data and the sample value for the length and standard deviation of the length of Expansion phase shows that there is a lot of dispersion in the simulated distribution. The AR(2) model also cannot reproduce the high Recovery growth rates exhibited by real GDP, and the standard deviation of quarterly growth rates for the phases are very far off from the sample data values as well. Finally, the AR(2) model does a very poor job at replicating the strong negative correlation between the cumulative growth rates of the Recession and Recovery phases exhibited by actual GDP. Using non-parametric residuals, the AR(2) model's performance improves somewhat. Looking down column one of Table 3, the model is now able to replicate the variability of the average growth rates of business cycle phases more satisfactorily, in particular the Recovery phase. However, even with non-parametric residuals, the AR(2) still produces a positive correlation between the cumulative growth rates of Recession and Recovery phases.

Turning our attention to the best performing model in Morley and Piger (2006), the bounceback BBV model, we can see that it clearly fares better than the AR(2) model. Column 4 of Table 3 shows that BBV with parametric residual specification can match all features reasonably well except for the standard deviation of quarterly growth rates of Recessions. It is especially remarkable that BBV can capture the high quarterly growth rate during the Recovery phase as well as the strong negative correlation between the cumulative growth rate in the Recession phase and the cumulative growth rate in the subsequent Recovery phase. The median simulated value for the average quarterly Recovery growth rate is close to 6% (7.1 % for the sample feature) while the median simulated value for the correlation feature is –0.44, close to the –0.66 reported in Table 2. Non-parametric residuals in this case do not lead to an improvement in the performance of the BBV model at all, creating percentiles in excess of 0.9 for the average quarterly growth rates of Recession and Expansion phases. However, they do allow the BBV to generate a slightly stronger negative correlation between the cumulative growth during Recession and Recovery phases (-0.49).[13]

The results reported here are consistent with the findings in Morley and Piger (2006) that the nonlinear bounceback model does a better job at capturing the important asymmetries in the business cycles than linear univariate models.

4.4. Business Cycle Features from Multivariate Models

[13] The weaker performance of the BBV model with bootstrapped residuals could be due to the problem of measuring residuals for such a model. In particular, residuals for Markov-switching models cannot be directly observed as they depend on the state (recession or expansion) and probability of switching or staying in that state. To get around this problem, we assume the state is observable by imposing the NBER peak and trough dates. Then, with the estimated model parameters, we calculate a set of residuals based on these states, allowing us to carry out the semi-parametric bootstrap procedure for the simulation exercise.

Table 4 reports percentiles of sample values for business cycle features in terms of the simulated distributions of these features for the multivariate models. A brief glance at the table would show that the three different multivariate models produce more or less the same results. All the models do well in terms of matching the number of peaks and the average length and variation of Recession and Expansion phases. However, as with the linear AR(2) model earlier, they fail completely in terms of being able to generate a high enough average quarterly growth rate for the Recovery phase or a strong enough negative correlation between the cumulative growth rate of Recession and the cumulative growth rate of Recovery phases. The ALW 4 variable VAR model even has trouble with the average quarterly growth rates in the Expansion phases. The multivariate models also cannot replicate the standard deviations associated with the quarterly growth rates of most of the business cycle phases.

Switching from parametric residual to non-parametric residuals improves the performance of all the models slightly. Mostly the improvement can be seen in being better able to match the standard deviation of quarterly growth rates of the business cycle phases. Consistent with the univariate findings, non-parametric residuals also help with generating a slightly stronger negative correlation between the cumulative growth rates of Recession and Recovery phases, though not strong enough to push the percentiles into an acceptable range. For the KPSW model, the non-parametric residuals actually worsen the performance of the model somewhat, by generating a median value of the average Expansion quarterly growth rate that is far too small relative to the NBER sample value.

Given the results reported in this table, one can conclude that multivariate information does not improve the performance of linear models. In the best case scenario, the B&Q model with non-parametric residuals replicate features about as well as the simple AR(2) with non-parametric residuals. This result is quite consistent with that reported in Clements and Krolzig (2004), who find multivariate models do no better, and often worse, than the univariate linear ARIMA models.

So far, we have shown that the bounceback BBV model is still the best performing model, supporting the results in Morley and Piger (2006). However, it is important to note that not all nonlinear time-series models do better in terms of business cycle feature reproduction when compared to linear models. For example, Morley and Piger (2006) found that the two-regime Markov-switching model of Hamilton (1989) performs about the same as the linear models. A key reason why the nonlinear BBV model does such a superior job in reproducing business cycle features is that there is a mechanism embedded in the model to capture high growth recoveries. This is what Galvão (2002) found as well when considering related models. Among the fifteen univariate nonlinear models she investigated, only two (a three-regime Markov-switching model and a state space model with Markov-switching in the transitory component) were able to account for the asymmetries in the shape of the U.S. business cycle, and these two models are characterized by mechanisms to capture high growth recoveries.

4.5. Business Cycle Features and the "Great Moderation"

As mentioned in the introduction, there is much evidence in the economics literature for the marked decline in the volatility of U.S. GDP growth since the mid 1980s, which is often labeled the "Great Moderation." The magnitude of the decline is striking. McConnell and Perez-Quiros (2000) show that the variance of output fluctuations since 1984 is only one fourth of the variance for the period ending 1983. There is much discussion as to the reason for the decline in volatility; some argue it is good monetary policy or better business practices, while others believe it is simply good luck (variance of exogenous shocks hitting the U.S. economy dropping sharply). Regardless of the reason, this is an important feature of the U.S. GDP data that should be taken into account in our simulations.

One major concern for not addressing this issue is that the linear models would be at a great disadvantage in our analysis because linear models cannot "automatically" pick up a reduction in variance while the bounceback model can potentially proxy for the structural break in variance or other forms of heteroskedasticity through its Markov-switching structure. So the superior performance of the bounceback model may be due to it capturing the break in variance rather than the asymmetries related to the business cycle. Therefore, to make sure that our results are robust, we consider a break in the variance of real GDP growth in 1984Q1 for all five time-series models presented earlier.

To implement the structural break, we only consider non-parametric residuals for all the models. This implies that the residuals or error terms for each of the time-series model are going to be drawn with replacement from two separate groups stemming from the original estimation residuals, pre-structural break (1948Q4 to 1984Q1) and post-structural break (1984Q2 to 2007Q4), depending on the quarter being simulated.

For the nonlinear bounceback model, in addition to looking at the specification with a structural break in variance in 1984Q1 (BBV1), we consider a specification in which we allow the shift parameter that represents deviation from long-run growth during recessions to change as well as the variance before and after 1984Q1 (BBV2).[14]

Table 5 reports the results in terms of the time-series models' ability at reproducing business cycle features while taking into account the Great Moderation. Looking at the univariate models first, one can see that the basic findings are very similar to those reported in Table 3. The AR(2) model fails to reproduce the exact same features as it did before taking the structural break into account (average quarterly growth rates of Recovery phase, standard deviation of quarterly growth rates of Recession and Mature expansion phases, and correlation between cumulative growth rates of Recession and Recovery phases). The one noticeable difference is that the median value of the 10,000

[14] This specification is motivated by one of the models (Model IV) in Kim and Nelson (1999), which the authors have found to be one of the two preferred models in their paper in terms of the model's ability at fitting real U.S. GDP data under the Bayesian marginal likelihood model selection procedure.

simulated series for the correlation feature is now negative (–0.07), which is somewhat more compatible with the sample feature than the small positive correlation (+0.07) it generated before taking the structural break into account. However, the correlation is still well below the –0.66 reported for the sample feature using NBER chronology.

As for the bounceback model, BBV1 does slightly worse than BBV with parametric residuals and slightly better than BBV with non-parametric residuals. Though compared to either of the models without structural break, BBV1 simulates a negative median value for the correlation between cumulative growth in the Recession phase and cumulative growth in the subsequent Recovery phase that is closer to the NBER sample feature. Results for the BBV2 specification show that it performs much more poorly than any of the other BBV specifications. It is hard to say what caused the deterioration of the performance of the BBV models after imposing the structural break. One possibility is that estimates of nonlinear dynamics are less precise because of the additional parameters related to structural change that have been added into the model.

The most interesting results in Table 5 are probably those related to the multivariate models. There are dramatic improvements in the performance of all the multivariate models, especially the KPSW VECM with non-parametric residuals. The models are now better at matching the variation in the quarterly growth rates of business cycle phases. But perhaps the most notable change is in the correlation feature. The multivariate models are now able to generate a more negative correlation between the cumulative growth rates of Recession and Recovery phases such that the proportion of simulated data below the corresponding NBER sample feature value is just slightly above 10%. This result is quite surprising given that none of the linear specification in Morley and Piger (2006) report a proportion higher than 10%. Even some nonlinear models in Morley and Piger (2006) report percentiles that are far less than 10%.

However, one should be cautious in interpreting this result as a validation for the success of multivariate linear models in capturing business cycle asymmetries exhibited by real GDP. First of all, the median correlations for the 10,000 simulations for all the multivariate linear models are still only mildly negative. B&Q reports the most negative median correlation at -0.24, which is less than that reported for either versions of the BBV (-0.53 for BBV1 and -0.39 for BBV2), and far less than the sample feature correlation of -0.66 calculated using the NBER chronology. Furthermore, the fact that the multivariate linear models cannot produce a strong enough negative correlation before taking into account the structural break in variance implies that there is something about the volatility reduction in 1984 that helped generate it, rather than something inherent in the dynamics of the linear models.

To test our hypothesis that the stronger negative correlation between the cumulative growth rates of the Recession and the Recovery phases is entirely driven by the one-time structural break in GDP variance, we conduct a simple counterfactual experiment detailed in the next subsection.

4.6. Counterfactual Experiment for Multivariate Linear Models

If there is something about the linear dynamics in the multivariate models that allow them to capture the strong negative correlation between growth in recessions and growth in recoveries exhibited by real GDP, it should be a recurring feature of the simulated data prior to the structural break date of 1984Q1 and after it as well. So consider this thought experiment: What would happen if the pre-1984Q1 parameters for the multivariate linear models applied for the whole sample period? Would this generate a strong enough negative correlation between the growth rates in recessions and recoveries? Similarly, what would happen if the post-1984Q1 parameters for the multivariate linear models applied for the whole sample period? Would there be a strong enough negative correlation this time?

These questions lead us to a simple counterfactual experiment where we estimate each of the multivariate models using pre-1984Q1 data and post-1984Q1 data separately. We then assume that the pre (post) break date parameters apply to the whole sample period and simulate corresponding counterfactual data to calculate the implied correlation between the cumulative growth rate of the Recession phase and the Recovery phase. We consider both parametric and non-parametric residual specifications, though the results are very similar. Table 6 details the outcome of the counterfactual experiment.

It is clear from the table that strong negative correlation between growth rates in recession and recovery phases are not a recurring trend using either pre or post break date parameters for any of the multivariate linear models. Under counterfactual 1 (pre-1984Q1 parameters), the median correlations for the simulations are only slightly negative or zero. With low corresponding percentiles, these results show that it is very unlikely that the sample value could have arisen from such models. Under counterfactual 2 (post-1984Q1 parameters), the median correlations for the simulations for all of the multivariate linear models are not even negative anymore. Though it is curious that the corresponding percentiles reported are all within the 0.1 to 0.9 boundary. This could be due to the poor fit of the models. Obviously, just using the pre (post) break date parameters to fit the whole sample period means that the simulated data will be much more (less) volatile than actual GDP between 1948Q1 and 2007Q4. With pre-break parameters, all the multivariate linear models over-predict the number of peaks (median of 11 to 14 peaks compared to 9 reported by NBER). In contrast, with post-break parameters, all the multivariate linear models far under-predict the number of peaks (median of 3 or 4 peaks only). The extreme deviation from actual GDP using post-1984Q1 parameters is the most likely explanation for the strange percentile numbers reported in Table 6 for counterfactual 2.

Through this counterfactual experiment, we have found some evidence that support our conjecture that the multivariate linear models with break in variance in

1984Q1 are not really capturing the negative correlation between the cumulative growth rates of the Recession and Recovery phases. And even if we take the results reported in Table 5 at face value, compared to the preferred model before imposing the structural break (BBV), the best performing models in Table 5 (KPSW and BBV1) still fare worse in terms of reproducing business cycle features. These results illustrate that, while a more general model will always fit the data better in sample, it does not necessary do better on other dimensions.[15]

5. Conclusion

In this paper, we implemented a novel approach to model comparison by assessing the ability of various time-series models to reproduce business cycle features exhibited by U.S. real GDP. Following Morley and Piger (2006), we use the most accurate possible business cycle dating algorithm to calculate business cycle turning points for the simulated data from each of the time-series models. The univariate linear and nonlinear models and the multivariate linear models we consider here allow us to answer the question of whether multivariate information can enrich the linear models such that they would succeed where univariate linear models have failed in terms of replication of certain business cycle features.

[15] This is analogous to the idea that more parsimonious models can forecast better out-of-sample, even if they fit worse in sample.

From the simulation exercises, a few important results emerge. First of all, the use of semi-parametric bootstrap approach to residual specification seems to benefit some models, particularly the linear models. At the same time, the fact that the linear models with non-parametric residuals fail to capture the strong negative correlation between the cumulative growth of the Recession phase and the cumulative growth of the Recovery phase while the BBV model with normal parametric residuals does suggests that the failure of the linear models is not due to the misspecification of the error terms. Perhaps the semi-parametric bootstrap procedure improved the performance of the linear models only because it allowed the linear models to better approximate the nonlinear models.

Secondly, the imposition of a structural break in the variance of real GDP growth in 1984Q1 had a noticeable impact on the performance of the multivariate linear models, enabling the VAR and VECM models to come closer to matching the BBV's ability to replicate most of the business cycle features considered here. However, our counterfactual experiment shows that this improvement may not be as impressive as it first appears.

Finally, the bounceback nonlinear model specification is by far the best performing time-series model among the ones we consider here. It can capture not only the usual features other papers in the literature report, such as the length and variation of business cycle phases or the average and standard deviation of quarterly growth rates of business cycle phases, but also important business cycle asymmetries that economists have observed in real GDP. Specifically, BBV does an excellent job at replicating the higher than average growth rates during the Recovery phase and the strong correlation between the severity of a recession and the strength of the subsequent recovery. This result is consistent with findings in Morley and Piger (2006) and corroborates the results in Galvão (2003) and Demers and Macdonald (2006). What this suggests is that the nonlinearity present in the U.S. business cycle is not something that linear models can pick up just by allowing for multivariate information. There is something fundamentally different about the dynamics of coming out of a recession that the linear models simply cannot replicate.

Appendix

Here we present the estimates for quarterly U.S. GDP for the five time-series models under consideration. The reported estimates are used to calibrate the data generating process used in our Monte Carlo simulations. The AR(2) and the Kim, et. al. (2005) bounceback model are univariate, while the Blanchard and Quah (1989) VAR, the Ahmed et. al. (2004) VAR, and the King et. al. (1991) VECM are multivariate. For the univariate models, Δy_t is defined as annualized growth rate of output to be compatible with the specification in Morley and Piger (2006). For the multivariate models, Δy_t is defined as natural log difference of output to be compatible with their original specifications.

The AR(2) model:

Estimation period 1948Q4 to 2007Q4.

$\Delta y_t = 0.0214 + 0.2976\Delta y_{t-1} + 0.0858\Delta y_{t-2} + \varepsilon_t$,

$\sigma_\varepsilon = 0.0383$.

The Kim, Morley, and Piger (2005) bounceback model (BBV):

Estimation period 1948Q4 to 2007Q4.

$\Delta y_t = 3.3521 - 4.4383 S_t + 1.3052(1 - S_t)\sum_{j=1}^{6} S_{t-j} + \varepsilon_t$,

$\sigma_\varepsilon = 3.1122$, $P(S_t = 1 | S_{t-1} = 1) = 0.7321$, $P(S_t = 0 | S_{t-1} = 0) = 0.9450$,

where $S_t = 1$ corresponds to recessions and $S_t = 0$ corresponds to expansions.

The Kim, Morley, and Piger (2005) bounceback model with break in variance (BBV1):

Estimation period 1948Q4 to 2007Q4.

$$\Delta y_t = 3.1464 - 4.4459 S_t + 1.5110(1 - S_t)\sum_{j=1}^{6} S_{t-j} + \varepsilon_t ,$$

$\sigma_\varepsilon = 4.0732$ for t = 1948Q4 to 1984Q1,

$\sigma_\varepsilon = 1.9881$ for t = 1984Q2 to 2007Q4,

$P(S_t = 1 | S_{t-1} = 1) = 0.7630$, $P(S_t = 0 | S_{t-1} = 0) = 0.9716$,

where $S_t = 1$ corresponds to recessions and $S_t = 0$ corresponds to expansions.

The Kim, Morley, and Piger (2005) bounceback model with break in variance and shift parameter (BBV2):

Estimation period 1948Q4 to 2007Q4.

For t = 1948Q4 to 1984Q1: $\Delta y_t = 3.4862 - 5.0392 S_t + 1.0566(1 - S_t)\sum_{j=1}^{6} S_{t-j} + \varepsilon_t$,

$\sigma_\varepsilon = 3.8646$.

For t = 1984Q2 to 2007Q4: $\Delta y_t = 3.4862 - 1.6000 S_t + 0.3355(1 - S_t)\sum_{j=1}^{6} S_{t-j} + \varepsilon_t$,

$\sigma_\varepsilon = 1.7767$.

For all t: $P(S_t = 1 \mid S_{t-1} = 1) = 0.7973$, $P(S_t = 0 \mid S_{t-1} = 0) = 0.9218$,

where $S_t = 1$ corresponds to recessions and $S_t = 0$ corresponds to expansions.

Blanchard & Quah (1989) 2 variable VAR model (B&Q):

Estimation period 1950Q1 to 2007Q4.

$$\begin{aligned}\Delta y_t = {} & 0.0022 + 0.1254\Delta y_{t-1} + 0.1682\Delta y_{t-2} + 0.0532\Delta y_{t-3} + 0.1426\Delta y_{t-4} + 0.06208\Delta y_{t-5} \\ & + 0.1596\Delta y_{t-6} - 0.0158\Delta y_{t-7} + 0.0231\Delta y_{t-8} - 0.7470u_{t-1} + 1.5542u_{t-2} - 0.5442u_{t-3} \\ & + 0.5880u_{t-4} - 0.8945u_{t-5} + 0.3827u_{t-6} - 0.2552u_{t-7} - 0.0012u_{t-8} + \varepsilon_t,\end{aligned}$$

$$\Sigma_\varepsilon = \begin{bmatrix} 0.0000762302 & -0.0000148006 \\ -0.0000148006 & 0.0000070473 \end{bmatrix},$$

where u_t denotes the civilian unemployment rate and the order of the variables in the VAR is $[\Delta y_t \;\; u_t]'$. The quarterly unemployment rate is the average of the monthly unemployment rate series.

Ahmed, Levin, and Wilson (2004) 4 variable VAR model (ALW):

Estimation period 1955Q3 to 2007Q4.

$$\begin{aligned}\Delta y_t = {} & 0.0076 + 0.2145\Delta y_{t-1} + 0.1660\Delta y_{t-2} + 0.0021\Delta y_{t-3} - 0.0328\Delta y_{t-4} + 0.0673\Delta cpi_{t-1} \\ & - 0.0316\Delta cpi_{t-2} + 0.0214\Delta cpi_{t-3} - 0.1648\Delta cpi_{t-4} - 0.0144\Delta ppi_{t-1} + 0.0263\Delta ppi_{t-2} \\ & - 0.0238\Delta ppi_{t-3} + 0.0139\Delta ppi_{t-4} + 0.0160 ffr_{t-1} - 0.2538 ffr_{t-2} + 0.1140 ffr_{t-3} + 0.0998 ffr_{t-4} \\ & + \varepsilon_t,\end{aligned}$$

$$\Sigma_\varepsilon = \begin{bmatrix} 0.000066 & -0.000001 & 0.000006 & 0.000023 \\ -0.000001 & 0.000021 & 0.000039 & 0.000012 \\ 0.000006 & 0.000039 & 0.000165 & 0.000039 \\ 0.000023 & 0.000012 & 0.000039 & 0.000120 \end{bmatrix},$$

where Δcpi_t denotes the consumer price inflation rate, Δppi_t is the inflation rate of the producer price index: all commodities, and ffr_t is the federal funds rate. The order of the variables in the VAR is $[\Delta y_t \;\; \Delta cpi_t \;\; \Delta ppi_t \;\; ffr_t]'$. The quarterly *cpi*, *ppi*, and *ffr* are all constructed by picking the end of quarter value of the equivalent monthly series.

King, Plosser, Stock, and Watson (1991) 3 variable VECM (KPSW):

Estimation period 1949Q2 to 2007Q4.

$$\begin{aligned} \Delta y_t = {} & 0.0008 + 0.0895\left(c_{t-1} - y_{t-1} + 0.4178\right) - 0.0265(i_{t-1} - y_{t-1} + 2.0545) + 0.1462\Delta y_{t-1} \\ & + 0.0526\Delta y_{t-2} + 0.0276\Delta y_{t-3} - 0.0636\Delta y_{t-4} + 0.1650\Delta y_{t-5} + 0.0752\Delta y_{t-6} - 0.0865\Delta y_{t-7} \\ & + 0.0057\Delta y_{t-8} + 0.2790\Delta c_{t-1} + 0.1360\Delta c_{t-2} - 0.0079\Delta c_{t-3} + 0.1432\Delta c_{t-4} - 0.1311\Delta c_{t-5} \\ & - 0.0009\Delta c_{t-6} + 0.1878\Delta c_{t-7} - 0.0724\Delta c_{t-8} + 0.0134\Delta i_{t-1} + 0.0202\Delta i_{t-2} - 0.0084\Delta i_{t-3} \\ & + 0.0138\Delta i_{t-4} - 0.0333\Delta i_{t-5} - 0.0026\Delta i_{t-6} + 0.0121\Delta i_{t-7} + 0.0100\Delta i_{t-8} + \varepsilon_t, \end{aligned}$$

$$\Sigma_\varepsilon = \begin{bmatrix} 0.000078 & 0.000039 & 0.000270 \\ 0.000039 & 0.000057 & 0.000053 \\ 0.000270 & 0.000053 & 0.001631 \end{bmatrix},$$

where c_t denotes real personal consumption expenditure and i_t is the real gross private domestic investment. The order of the variables in the VECM is $[y_t \;\; c_t \;\; i_t]'$ and the two cointegrating relationships based on the balance growth theory are $(c_t - y_t)$ and $(i_t - y_t)$.

References

Ahmed, Shaghil, Andrew Levin, and Beth Anne Wilson (2004): "Recent U.S. Macroeconomic Stability: Good Policies, Good Practices, or Good Luck?" *The Review of Economics and Statistics*, vol. 86, no. 3, pg. 824-832.

Blanchard, Olivier J., and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, vol. 79, no. 4, pg. 655-673.

Bry, Gerhard, and Charlotte Boschan (1971): Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, New York: National Bureau of Economic Research.

Boivin, Jean, and Marc P. Giannoni (2006): "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics*, vol. 88, no. 3, pg. 445-462.

Boldin, Michael D. (1996): "A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle," *Studies in Nonlinear Dynamics and Econometrics*, vol. 1, no. 1, pg. 35-46.

Burns, Arthur F., and Wesley C. Mitchell (1947): Measuring Business Cycles, New York: National Bureau of Economic Research.

Clements, Michael P., and Hans-Martin Krolzig (2004): "Can Regime-Switching Models Reproduce the Business Cycle Features of US Aggregate Consumption, Investment and Output?" *International Journal of Finance and Economics*, vol. 9, no. 1, pg. 1-14.

Davidson, James E. H., David F. Hendry, Frank Srba, and Stephen Yeo (1978): "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers Expenditure and Income in the United Kingdom," *Economic Journal*, vol. 88, no. 352, pg. 661-692.

Demers, Frédérick, and Ryan Macdonald (2007): "The Canadian Business Cycle: A Comparison of Models," Bank of Canada Working Paper 2007-38.

Friedman, Milton (1964): "Monetary Studies of the National Bureau," The National Bureau Enters Its 45th Year, 44th Annual Report, New York: National Bureau of Economic Research. [Reprinted in Friedman, Milton (1969): The Optimum Quantity of Money and Other Essays, Chicago: Aldine.]

Friedman, Milton (1993): "The "Plucking Model" of Business Fluctuations Revisited," *Economic Inquiry*, vol. 31, no. 2, pg. 171-177.

Hamilton, James D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, vol. 57, no. 2, pg. 357-384.

Hamilton, James D. (1994): Time Series Analysis, Princeton: Princeton University Press.

Harding, Don, and Adrian Pagan (2002): "Dissecting the Cycle: a Methodological Investigation," *Journal of Monetary Economics*, vol. 49, no. 2, pg. 365-381.

Hess, Gregory D., Shigeru Iwata (1997): "Measuring and Comparing Business-Cycle Features," *Journal of Business & Economic Statistics*, vol. 15, no. 4, pg. 432-444.

Galvão, Ana Beatriz C. (2002): "Can Non-linear Time Series Models Generate US Business Cycle Asymmetric Shape?" *Economics Letters*, vol. 77, no. 2, pg. 187-194.

Kim, Chang-Jin, James Morley, and Jeremy Piger (2005): "Nonlinearity and the Permanent Effects of Regressions," *Journal of Applied Econometrics*, vol. 20, no. 2, pg. 291-309.

Kim, Chang-Jin, and Charles R. Nelson (1999): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics*, vol. 81, no. 4, pg. 608-616.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, vol. 81, no. 4, pg. 819-840.

McConnell, Margaret M., and Gabriel Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review*, Vol. 90, no. 5, pg. 1464-1476.

McQueen, Grant, and Steven Thorley (1993): "Asymmetric Business Cycle Turning Points," *Journal of Monetary Economics*, vol. 31, no. 3, pg. 341-362.

Morley, James, and Jeremy Piger (2006): "The Importance of Nonlinearity in Reproducing Business Cycle Features," Nonlinear Time Series Analysis of Business Cycles, Contributions to Economic Analysis series, vol. 276. Amsterdam and San Diego: Elsevier, pg. 75-94.

Sims, Christopher A. (1980): "Macroeconomics and Reality," *Econometrica*, vol. 48, no. 1, pg. 1-48.

Solow, Robert M. (1970): Growth Theory: An Exposition, Clarendon Press.

Stock, James H., and Mark W. Watson (2002): "Has the Business Cycle Changed and Why?" Mark Gertler and Kenneth Rogoff (Eds.) NBER Macroeconomics Annual, New York: National Bureau of Economic Research.

TABLE 1

PEAK AND TROUGH DATES FROM NBER BUSINESS CYCLE DATING COMMITTEE AND THE BBQ AND MBBQ ALGORITHMS APPLIED TO U.S. REAL GDP (1948Q4 – 2007Q4)

Business Cycle Peaks			Business Cycle Troughs		
NBER	BBQ	MBBQ	NBER	BBQ	MBBQ
1948Q4	-	-	1949Q4	**1949Q2**	1949Q4
1953Q2	1953Q2	1953Q2	1954Q2	**1954Q1**	1954Q2
1957Q3	1957Q3	1957Q3	1958Q2	**1958Q1**	**1958Q1**
1960Q2	**1960Q1**	**1960Q1**	1961Q1	**1960Q4**	**1960Q4**
1969Q4	**1969Q3**	**1969Q3**	1970Q4	1970Q4	1970Q4
1973Q4	1973Q4	1973Q4	1975Q1	1975Q1	1975Q1
1980Q1	1980Q1	1980Q1	1980Q3	1980Q3	1980Q3
1981Q3	1981Q3	1981Q3	1982Q4	**1982Q1**	1982Q4
1990Q3	1990Q3	1990Q3	1991Q1	1991Q1	1991Q1
2001Q1	-	-	2001Q4	-	-

Note: Bold indicate that the identified turning points differ from the NBER dates. We ignore the first NBER peak date in our evaluation of the BBQ and MBBQ algorithm because given our sample period, the earliest date at which the algorithms can identify a turning point is 1949Q2.

TABLE 2

BUSINESS CYCLE FEATURES FOR U.S. REAL GDP (1948Q4 – 2007Q4)

	NBER	BBQ	MBBQ
Average quarterly growth rates			
Recession	–1.92	**–2.96**	–2.49
Expansion	4.59	4.78	**4.98**
Recovery	7.10	**5.52**	7.23
Mature expansion	3.94	**4.57**	4.29
Std. deviation of quarterly growth rates			
Recession	3.33	**3.10**	3.13
Expansion	3.54	**3.83**	3.75
Recovery	4.18	**4.75**	4.25
Mature expansion	3.05	**3.51**	3.31
Number of phases			
Number of peaks	9	8	8
Average length of phases			
Recession	3.44	**3.00**	3.50
Expansion	19.67	17.88	**17.13**
Std. deviation of length of phases			
Recession	1.13	1.31	**1.41**
Expansion	12.72	11.34	**10.88**
Correlation between growth rates			
Recession/Recovery	–0.66	**–0.36**	–0.68

Note: Because the earliest date at which the algorithms can identify a turning pint is 1949Q2, we ignore the first peak in 1948Q4 when calculating the sample features associated with the NBER dates. Bold indicates that the feature values produced by the algorithm is "further away" from the NBER sample feature values.

TABLE 3

PERCENTILES OF BUSINESS CYCLE FEATURES FOR UNIVARIATE MODELS

Features	AR(2) (non-parametric)	AR(2) (parametric)	BBV (non-parametric)	BBV (parametric)
Average quarterly growth rates				
Recession	0.71 (+0.26)	0.63 (+0.14)	**0.93 (+0.75)**	0.69 (+0.20)
Expansion	0.89 (+0.47)	0.80 (+0.30)	**0.90 (+0.43)**	0.89 (+0.40)
Recovery	**1.00 (+3.12)**	**1.00 (+2.94)**	0.83 (+0.95)	0.90 (+1.23)
Mature expansion	0.33 (–0.19)	0.18 (–0.37)	0.76 (+0.20)	0.66 (+0.11)
Standard deviation of quarterly growth rates				
Recession	**0.96 (+0.79)**	**0.99 (+1.06)**	0.86 (+0.53)	**0.98 (+1.00)**
Expansion	0.43 (–0.06)	0.46 (–0.02)	0.27 (–0.18)	0.46 (–0.03)
Recovery	0.87 (+0.99)	**0.97 (+0.95)**	0.56 (+0.09)	0.61 (+0.16)
Mature expansion	**0.04 (–0.57)**	**0.01 (–0.56)**	**0.09 (–0.40)**	0.13 (–0.28)
Number of phases				
Number of peaks	0.61 (+1)	0.40 (0)	0.55 (+1)	0.50 (0)
Average length of phases				
Recession	0.60 (+0.16)	0.60 (+0.17)	0.36 (–0.31)	0.49 (–0.01)
Expansion	0.24 (–4.76)	0.42 (–1.33)	0.31 (–3.21)	0.33 (–2.67)
Standard deviation of length of phases				
Recession	0.29 (–0.38)	0.27 (–0.42)	0.14 (–0.94)	0.20 (–0.70)
Expansion	0.16 (–6.36)	0.27 (–3.58)	0.21 (–4.98)	0.24 (–4.42)
Correlation between growth rates				
Recession/Recovery	**0.06 (–0.73)**	**0.04 (–0.73)**	0.29 (–0.17)	0.24 (–0.22)

Note: Percentiles are based on 10,000 simulations. They represent the proportion of simulated features that fall below the corresponding sample feature reported in Table 2 for actual real GDP using the NBER peak and trough dates. Bold indicates a percentile that is less than 0.1 or greater than 0.9, implying that it was unlikely that the particular time-series model could simulate data that replicates the behavior of actual GDP for that particular feature. The numbers in parentheses correspond to the difference between a sample feature and the corresponding median simulated feature.

TABLE 4

PERCENTILES OF BUSINESS CYCLE FEATURES FOR MULTIVARIATE MODELS

Features	B&Q (non-parametric)	B&Q (parametric)	ALW (non-parametric)
Average quarterly growth rates			
Recession	0.68 (+0.20)	0.65 (+0.15)	0.46 (–0.05)
Expansion	0.90 (+0.37)	0.78 (+0.23)	**0.99 (+0.78)**
Recovery	**1.00 (+2.41)**	**1.00 (+2.39)**	**1.00 (+3.15)**
Mature expansion	0.31 (–0.16)	0.17 (–0.31)	0.69 (+0.17)
Std. dev. of quarterly growth rates			
Recession	**1.00 (+1.08)**	**1.00 (+1.20)**	**0.97 (+1.13)**
Expansion	0.48 (–0.02)	0.28 (–0.13)	0.80 (+0.25)
Recovery	0.89 (+0.82)	**0.94 (+0.76)**	**0.94 (+1.20)**
Mature expansion	**0.04 (–0.51)**	**0.00 (–0.65)**	0.19 (–0.26)
Number of phases			
Number of peaks	0.40 (0)	0.24 (–1)	0.59 (+1)
Average length of phases			
Recession	0.80 (+0.44)	0.72 (+0.32)	0.77 (+0.44)
Expansion	0.36 (–1.96)	0.55 (+0.57)	0.24 (–4.76)
Std. dev. of length of phases			
Recession	0.48 (–0.02)	0.39 (–0.13)	0.45 (–0.08)
Expansion	0.30 (–3.17)	0.43 (–0.87)	0.19 (–6.13)
Correlation between growth rates			
Recession/Recovery	**0.08 (–0.52)**	**0.04 (–0.59)**	**0.07 (-0.65)**

Features	ALW (parametric)	KPSW (non-parametric)	KPSW (parametric)
Average quarterly growth rates			
Recession	0.42 (–0.07)	0.68 (+0.21)	0.71 (+0.21)
Expansion	**0.97 (+0.62)**	**0.95 (+0.58)**	0.84 (+0.34)
Recovery	**1.00 (+3.00)**	**1.00 (+2.77)**	**1.00 (+2.51)**
Mature expansion	0.51 (+0.01)	0.53 (+0.02)	0.32 (–0.19)
Std. dev. of quarterly growth rates			
Recession	**1.00 (+1.33)**	**0.95 (+0.88)**	**1.00 (+1.08)**
Expansion	0.73 (+0.13)	0.56 (+0.04)	0.23 (–0.17)
Recovery	**0.98 (+1.04)**	**0.93 (+0.92)**	**0.93 (+0.72)**
Mature expansion	**0.03 (–0.40)**	**0.05 (–0.47)**	**0.00 (–0.69)**
Number of phases			
Number of peaks	0.34 (0)	0.37 (0)	0.14 (–2)
Average length of phases			
Recession	0.72 (+0.33)	0.67 (+0.24)	0.60 (+0.14)
Expansion	0.46 (–0.56)	0.42 (–1.22)	0.70 (+2.48)
Std. dev. of length of phases			
Recession	0.31 (–0.19)	0.36 (–0.17)	0.27 (–0.28)
Expansion	0.29 (–2.71)	0.35 (–2.22)	0.55 (+0.64)
Correlation between growth rates			
Recession/Recovery	**0.04 (–0.67)**	**0.07 (–0.54)**	**0.04 (–0.58)**

Note: Percentiles are based on 10,000 simulations. They represent the proportion of simulated features that fall below the corresponding sample feature reported in Table 2 for actual real GDP using NBER peak and trough dates. Bold indicates a percentile that is less than 0.1 or greater than 0.9. The numbers in parentheses correspond to the difference between a sample feature and the corresponding median simulated feature.

TABLE 5

PERCENTILES OF BUSINESS CYCLE FEATURES FOR ALL MODELS WITH STRUCTURAL BREAK IN VARIANCE 1984Q1

Features	AR(2) (non-parametric)	BBV1 (non-parametric)	BBV2 (non-parametric)
Average quarterly growth rates			
Recession	0.79 (+0.44)	0.69 (+0.25)	0.88 (+0.50)
Expansion	0.70 (+0.27)	**0.96 (+0.65)**	**0.90 (+0.43)**
Recovery	**0.99 (+2.55)**	0.88 (+1.42)	**0.95 (+1.58)**
Mature expansion	0.27 (–0.31)	0.86 (+0.32)	0.69 (+0.15)
Std. deviation of quarterly growth rates			
Recession	**0.90 (+0.67)**	**0.95 (+0.94)**	0.92 (+0.62)
Expansion	0.25 (–0.28)	0.34 (–0.13)	0.34 (–0.14)
Recovery	0.69 (+0.43)	0.44 (–0.11)	0.54 (+0.07)
Mature expansion	**0.05 (–0.74)**	0.11 (–0.36)	0.14 (–0.37)
Number of phases			
Number of peaks	0.70 (+2)	0.72 (+2)	0.31 (–1)
Average length of phases			
Recession	0.56 (+0.11)	0.51 (+0.02)	**0.10 (–1.22)**
Expansion	0.36 (–2.56)	0.19 (–6.33)	0.61 (+1.67)
Std. deviation of length of phases			
Recession	0.28 (–0.46)	0.24 (–0.70)	**0.03 (–1.69)**
Expansion	0.23 (–6.17)	0.15 (–7.55)	0.36 (–2.00)
Correlation between growth rates			
Recession/Recovery	**0.10 (–0.59)**	0.35 (–0.13)	0.19 (–0.27)

Features	B&Q (non-parametric)	ALW (non-parametric)	KPSW (non-parametric)
Average quarterly growth rates			
Recession	0.71 (+0.26)	0.63 (+0.18)	0.76 (+0.37)
Expansion	0.81 (+0.30)	**0.95 (+0.65)**	0.80 (+0.35)
Recovery	**0.98 (+2.11)**	**1.00 (+2.78)**	**0.99 (+2.29)**
Mature expansion	0.29 (–0.18)	0.60 (+0.10)	0.35 (–0.16)
Std. deviation of quarterly growth rates			
Recession	**0.99 (+1.00)**	**0.92 (+0.87)**	0.89 (+0.73)
Expansion	0.31 (–0.16)	0.58 (+0.07)	0.38 (–0.10)
Recovery	0.72 (+0.43)	0.83 (+0.77)	0.76 (+0.48)
Mature expansion	**0.04 (–0.58)**	0.14 (–0.38)	**0.08 (–0.54)**
Number of phases			
Number of peaks	0.52 (+1)	0.62 (+1)	0.57 (+1)
Average length of phases			
Recession	0.77 (+0.44)	0.73 (+0.34)	0.67 (+0.24)
Expansion	0.34 (–2.33)	0.34 (–2.90)	0.37 (–2.22)
Std. deviation of length of phases			
Recession	0.46 (–0.04)	0.40 (–0.15)	0.37 (–0.17)
Expansion	0.23 (–4.56)	0.22 (–5.49)	0.25 (–4.95)
Correlation between growth rates			
Recession/Recovery	0.13 (–0.42)	**0.10 (–0.53)**	0.12 (–0.47)

Note: Percentiles are based on 10,000 simulations. They represent the proportion of simulated features that fall below the corresponding sample feature reported in Table 2 for actual real GDP using the MBBQ algorithm. Bold indicates a percentile that is less than 0.1 or greater than 0.9, implying that it was unlikely that the particular time-series model could simulate data that replicates the behavior of actual GDP for that particular feature. The numbers in parentheses correspond to the difference between a sample feature and the corresponding median simulated feature.

TABLE 6

COUNTERFACTUAL EXPERIMENT RESULT FOR MULTIVARIATE MODELS

Correlation between Cumulative Growth in Recession Phase and Cumulative Growth in Recovery Phase	Pre-structural break Parameters (Counterfactual 1)	Post-structural break Parameters (Counterfactual 2)
B&Q		
Non-parametric		
Median value	-0.14	0.00
Proportion below sample feature	**0.04**	0.34
Parametric		
Median value	-0.11	0.00
Proportion below sample feature	**0.03**	0.28
ALW		
Non-parametric		
Median value	0.00	0.00
Proportion below sample feature	**0.02**	0.32
Parametric		
Median value	0.00	0.04
Proportion below sample feature	**0.02**	0.27
KPSW		
Non-parametric		
Median value	-0.15	0.09
Proportion below sample feature	**0.05**	0.27
Parametric		
Median value	-0.12	0.14
Proportion below sample feature	**0.03**	0.18

Note: Proportion percentiles are based on 10,000 simulations. They represent the proportion of simulated correlation that fall below the corresponding sample correlation reported in Table 2 for actual real GDP using NBER peak and trough dates. Bold indicates a percentile that is less than 0.1 or greater than 0.9. The structural break date is 1984Q1.