The Asymmetric Business Cycle*

James Morley
Washington University in St. Louis

Jeremy Piger
University of Oregon

February 24, 2009

ABSTRACT: The "business cycle" is a fundamental, yet elusive concept in macroeconomics. In this paper, we consider the problem of measuring the business cycle. First, we argue for the 'output-gap' view that the business cycle corresponds to transitory deviations in economic activity from a permanent or "trend" level. Then, we investigate the extent to which a general model-based approach to estimating trend and cycle for the United States in the postwar era produces measures of the business cycle that depend on models versus the data. We find strong empirical support for a nonlinear time series model that implies a highly asymmetric business cycle that is large and negative in recessions, but small and close to zero in expansions. Based on the principle of forecast combination, we use Bayesian model averaging to construct a model-free measure of the business cycle that also turns out to be highly asymmetric. This model-free measure of the business cycle is closely related to other measures of economic slack and implies a convex short-run aggregate supply curve. The asymmetric business cycle also potentially reconciles two long-standing, but competing theories about the main source of macroeconomic fluctuations.

Keywords: Business Cycle; Output Gap; Trend/Cycle Decomposition; Markov Switching; Asymmetry; Bayesian Model-Averaging; Short-Run Aggregate Supply
JEL Classification: E32, C22

---

* **Morley:** Department of Economics, Box 1208, Eliot Hall 205, Washington University, One Brookings Drive, St. Louis, MO 63130-4899, (morley@wustl.edu); **Piger:** Department of Economics, 1285 University of Oregon, Eugene, OR 97403-1285 (jpiger@uoregon.edu). We thank Todd Clark, Neal Ghosh, Jim Nason, Tara Sinclair, participants at the 2008 SNDE meetings and the 2008 SEA meetings, and seminar participants at the Kansas City Fed and George Washington University for helpful comments and suggestions. We also thank Jim Hamilton for providing some code. The usual disclaimers apply. Morley acknowledges support from the Weidenbaum Center on the Economy, Government, and Public Policy. Much of the work on this paper was done while Morley was a Research Fellow at the Federal Reserve Bank of St. Louis. The views expressed in this paper should not be interpreted as those of the Weidenbaum Center, the Federal Reserve Bank of St. Louis, or the Federal Reserve System. The most recent version of this paper, updated data for some of the business cycle measures, and a technical note on the computation of these measures can be found under working papers at http://artsci.wustl.edu/~morley.

Introduction

The "business cycle" is a broad term that connotes the inherent fluctuations in economic activity. Research on the measurement of business cycles has a long tradition in macroeconomics, with an early example provided by Wesley Mitchell (1927), the founder of the National Bureau of Economic Research (NBER). An integral part of business cycle measurement is its definition. Mitchell and the NBER defined the business cycle in terms of the alternation between periods of expansion and recession in the level of economic activity (which can be denoted the 'alternating-phases' definition). Other definitions exist, although they are often related in some way to the NBER definition. One popular alternative definition is that the business cycle represents transitory fluctuations in economic activity around a permanent or "trend" level (which can be denoted the 'output-gap' definition). This definition is associated with work on the U.S. business cycle by Beveridge and Nelson (1981), who propose a general approach to measuring the business cycle based on long-horizon forecasts produced by a time-series forecasting model.

In this paper, we revisit the problem of measuring the business cycle. We begin by arguing for consideration of the 'output-gap' definition of the business cycle. We then discuss how to conduct trend/cycle decomposition based on long-horizon forecasts for linear and nonlinear time series models, including how to implement an approach developed in Morley and Piger (2008) for empirically-relevant regime-switching processes.

When we apply model-based trend/cycle decomposition to U.S. real GDP, we find that the estimated cycle is highly dependent on model specification, with the key distinction being between linear models that imply symmetric gaps from trend and nonlinear regime-switching models that imply asymmetric gaps from trend. In order to discriminate between the different measures of the business cycle, we consider traditional model selection criteria, formal hypothesis testing, and Bayesian model comparison. The empirical evidence strongly favours nonlinear regime-switching model specifications and an asymmetric business cycle.

The Bayesian approach to model comparison allows us to construct a “model-free” measure of the business cycle based on the principle of forecast combination (i.e., a combined forecast can be superior to all of the individual forecasts that go into its construction), with weights for different time-series models based on “partial” marginal likelihoods that are robust to initial priors on model parameters. The resulting model-free business cycle is highly asymmetric, with small positive values in expansions and large negative values during NBER-dated recessions. Notably, this asymmetry and the corresponding negative mean of the business cycle imply much larger welfare costs to the business cycle than suggested by the Lucas (1987, 2003) calculation that assumes the business cycle is symmetric with a mean of zero.

Beyond a link between the ‘output-gap’ notion of the business cycle and the NBER’s ‘alternating-phases’ notion, we also find that the model-free business cycle is closely related to other measures of macroeconomic “slack” such as the unemployment rate and capacity utilization, even though the cycle is estimated via univariate analysis. In addition, we find that the asymmetric business cycle implies a convex short-run aggregate supply curve, suggesting that it takes a large negative cycle to bring about a significant disinflation, while a small positive cycle can lead to rising inflation. Finally, the asymmetric business cycle explains and potentially reconciles two widely-held, but very different theories about the main sources of macroeconomic fluctuations, with permanent productivity shocks able to explain the behaviour of output in expansions, while large, infrequent, and negative transitory demand shocks can explain the declines in economic activity in recessions. Taking these results together, we argue that the model-free business cycle captures a meaningful macroeconomic phenomenon and sheds more light on the nature of fluctuations in aggregate economic activity than simply looking at either the level or the growth rates of U.S. real GDP.

1. Definitions of the Business Cycle

In macroeconomics, fluctuations in economic activity are typically classified into three main categories: long-run growth, the business cycle, and seasonal patterns. These different sources of fluctuations may in fact be related to each other, but it can be useful to make some distinction between them. In this paper, we follow standard practice by considering seasonally-adjusted data. This implicitly treats the seasonal patterns as independent or, at least, not marginally relevant for making inferences about long-run growth or business cycles, although we note the existence of an interesting literature on the influence of seasonal fluctuations on business cycles (see, for example, Wen, 2002).

So, what is the "business cycle" as distinct from long-run growth? One notion that has been put forth by the NBER is that the business cycle corresponds to an alternation between relatively persistent phases of expansion and recession in economic activity that occur despite the positive *average* growth of economic activity in most industrialized countries. We refer to this notion of the business cycle as the 'alternating-phases' definition. One problem with this notion is that it is far from universal. Some countries have experienced many consecutive years of positive growth in the level of economic activity and thus have no business cycles in the strict NBER sense (e.g., Japan in the early postwar period).

A more general notion of the business cycle is that it corresponds to all short-run fluctuations in economic activity (again, beyond seasonal movements), without a distinction made between whether they correspond to an increase or outright decline in activity. The problem with this definition is that it merely labels the analysis of higher-frequency variation in economic activity as "business cycle analysis", without saying whether there is anything meaningful about the business cycle as a macroeconomic phenomenon. For example, this notion begs the question of why any attention is paid to whether the NBER deems there to be a recession or not.

A third notion of the business cycle is that it represents the transitory fluctuations of the economy around a long-run or “trend” level. In this paper, we argue that this ‘output-gap’ definition provides the most useful notion of the business cycle.[1] It implies a construct—the transitory component of real economic activity—that can be measured for any economy and is potentially useful for forecasting, policymaking, and theory. We emphasize that there is nothing about this notion of the business cycle that implies it is independent of long-run growth. Transitory fluctuations could be due to the same factors that drive long-run growth or they could be due to independent factors. It is ultimately an empirical question how important these different underlying factors are. Indeed, it is an empirical question as to whether the transitory component of economic activity is relevant for anything. Later in this paper, we provide some analysis that suggests it is. Furthermore, we find in measuring the transitory component for the U.S. economy that it displays large negative movements during NBER-dated recessions, suggesting a direct link between the ‘alternating-phases’ and ‘output-gap’ definitions of the business cycle.

Before discussing methods of measuring the transitory component of economic activity in the next section, it is worth providing a more formal discussion of the ‘output-gap’ definition

[1] It is possible to make a further distinction between transitory movements at different frequencies. This is the approach taken when a spectral filter is applied to a time series with the goal of isolating fluctuations at, say, the 1-5 year horizon. However, it is important to note that these spectral filters are based on the assumption that the time series being analyzed follows a stationary process. When applied to integrated processes, part of the filter (i.e., a differencing operator) is used up transforming the series to something that could be thought of as stationary, leaving the rest of the filter to amplify fluctuations in the transformed series at different frequencies than originally intended (see Cogley and Nason, 1995, and Murray, 2003, on this point). While we acknowledge that the isolation of transitory movements at different frequencies is an interesting issue in business cycle analysis, we consider the initial isolation of transitory movements for measures of real economic activity as the more important and challenging task for macroeconomists given the basic premise that not all fluctuations are transitory.

considered in this paper. First, following much of the literature, we use natural logarithms of U.S. quarterly real GDP, denoted y_t, as a measure of overall economic activity. We acknowledge that this measure has its limitations and does not always match up with the NBER's implicit measure of economic activity. However, it does a reasonable job on this front (see, for example, Harding and Pagan, 2002, and the subsequent literature on business cycle dating with real GDP). Then, given y_t, the 'output-gap' definition of the business cycle is the notion that economic activity can be meaningfully decomposed into a trend and a cycle as follows:

$$y_t = \tau_t + c_t , \tag{1}$$

$$\tau_t = \tau_{t-1} + \eta_t^* , \tag{2}$$

$$c_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j}^* , \tag{3}$$

where $\psi_0 = 1$, $\eta_t^* = \mu + \eta_t$, and $\omega_t^* = \varpi + \omega_t$, with η_t and ω_t following martingale difference sequences. The trend, τ_t, is the permanent component of y_t in the sense that the effects of the realized trend innovations, η_t^*, on the level of the time series are not expected to be reversed. By contrast, the cycle, c_t, is the transitory component of y_t in the sense that the Wold coefficients, ψ_j, are assumed to be absolutely summable such that the realized cycle innovations, ω_t^*, have finite memory. The parameter μ allows for non-zero drift in the trend, while the parameter ϖ allows for a non-zero mean in the cycle, although the mean of the cycle is not identified from the behaviour of the time series alone, as different values for ϖ all imply the same reduced-form dynamics for Δy_t, with the standard identification assumption being that $\varpi = 0$.

Whether the permanent and transitory components in (1) are meaningful macroeconomic phenomena is ultimately an empirical question, although it is clear that the trend should embody the steady-state effects of the factors that drive long-run growth in economic activity.[2] Such factors might also have transitory effects, so we do not want to assume the permanent and transitory shocks are uncorrelated (see Morley, Nelson, and Zivot, 2003, on this point).

In this paper, we consider the setting where some of the parameters describing the process in (1)-(3) can be regime-switching, as discussed in Morley and Piger (2008), and where some of the parameters can undergo structural breaks. In addition to finite-order unobserved-components (UC) models of the process in (1)-(3), we also consider processes for which there is no finite-order autoregressive moving-average (ARMA) representation of the Wold form in (3). Specifically, we conduct trend/cycle decomposition based on reduced-form forecasting models that capture the autocovariance structure for a general process as in (1)-(3), regardless of whether the process has a finite-order UC representation. Given a forecasting model that captures the autocovariance structure of a process as in (1)-(3), the methods we employ provide optimal estimates (in a minimum mean-squared-error sense) of trend and cycle.

2. Trend/Cycle Decomposition Based on Time Series Models

2.1 The Beveridge-Nelson Decomposition

There are many different approaches to trend/cycle decomposition. In terms of the 'output-gap' definition of the business cycle as transitory deviations around a trend, a particularly useful and general approach is the Beveridge-Nelson (BN) decomposition. The BN measure of trend is

[2] To be clear with our terminology, by "steady state" we have in mind the level to which the process would gravitate in the absence of future permanent or transitory innovations.

$$\hat{\tau}_t^{BN} \equiv \lim_{j\to\infty}\left\{E^M\left[y_{t+j}\middle|\Omega_t\right] - j\cdot E^M\left[\Delta y_t\right]\right\}, \tag{4}$$

where $E^M[\cdot]$ is the expectations operator with respect to a forecasting model and Ω_t is the set of relevant and available information observed up to time *t*. In words, the BN trend is the long-horizon conditional forecast of the time series minus any deterministic drift. The intuition for the BN measure of trend is that, as the forecasting horizon extends to infinity, a long-horizon forecast of a time series will no longer be influenced by the transitory component that exists at time *t*, and, therefore, will only reflect the trend component.

Both the conditional and unconditional expectations in (4) should be straightforward to calculate (either analytically or by simulation) given a forecasting model. Morley (2002) and Clarida and Taylor (2003) provide discussion and examples, while Appendix A provides the relevant formulas for the class of forecasting models to which the BN decomposition will be applied in this paper. Meanwhile, the BN trend provides an optimal estimate of the underlying trend of an integrated process in the following circumstances: First, the time series under analysis conforms to the trend/cycle process in (1)-(3), with constant drift, μ, constant Wold coefficients, ψ_j, and the mean of the cycle innovations equal to zero, $\overline{\omega}=0$. Second, the forecasting model captures the autocovariance structure of the process such that $E^M\left[y_{t+j}\middle|\Omega_t\right] = E\left[y_{t+j}\middle|\Omega_t\right]$.

This second requirement highlights the fact that accurate measurement of the transitory component requires an accurate forecasting model. This is important because it justifies our choice to not limit our consideration only to finite-order UC models in order to capture the process in (1)-(3). Such models represent a mere subset of all possible time series models and can place binding restrictions on the autocovariance structure of a given time series process. By considering a broader set of models, we aim to get $E^M\left[y_{t+j}\middle|\Omega_t\right]$ as close as possible to $E\left[y_{t+j}\middle|\Omega_t\right]$.

2.2 The Regime-Dependent Steady-State Decomposition

In considering a broader set of time series models, it should be noted that the BN trend does not generally provide an optimal or even unbiased estimate of trend when the process in (1)-(3) has regime-switching parameters, even if the forecasting model is correctly specified such that $E^M\left[y_{t+j}\middle|\Omega_t\right]=E\left[y_{t+j}\middle|\Omega_t\right]$. Morley and Piger (2008) propose a regime-dependent steady-state (RDSS) approach that generalizes the BN decomposition in order to provide optimal estimates in the setting where the underlying trend and/or cycle can be regime switching. The RDSS approach involves constructing long-horizon forecasts conditional on sequences of regimes and then marginalizing over the distribution of the unknown regimes.

The RDSS measure of trend is

$$\hat{\tau}_t^{RDSS} \equiv \sum_{\tilde{S}_t}\left\{\hat{\tau}_t^{RDSS}\left(\tilde{S}_t\right)\cdot p^M\left(\tilde{S}_t\middle|\Omega_t\right)\right\}, \tag{5}$$

$$\hat{\tau}_t^{RDSS}\left(\tilde{S}_t\right)\equiv \lim_{j\to\infty}\left\{E^M\left[y_{t+j}\middle|\left\{S_{t+k}=i^*\right\}_{k=1}^{j},\tilde{S}_t,\Omega_t\right]-j\cdot E^M\left[\Delta y_t\middle|\left\{S_t=i^*\right\}_{-\infty}^{\infty}\right]\right\}, \tag{6}$$

where $\tilde{S}_t=\{S_t,\ldots,S_{t-m}\}'$ is a vector of relevant current and past regimes for forecasting a time series, $p^M(\cdot)$ is the probability distribution with respect to the forecasting model, S_t is an unobserved Markov state variable that takes on *N* discrete values according to a fixed transition matrix, and i^* is the "normal" regime in which the mean of the transitory component is assumed to be zero. For a given forecasting model, the probability weights in (5), $p^M\left(\tilde{S}_t\middle|\Omega_t\right)$, can be obtained from the recursive filter given in Hamilton (1989). Appendix A provides the relevant formulas for constructing the expectations in (6) for the regime-switching models considered in this paper.

Morley and Piger (2008) show that the RDSS decomposition provides an optimal estimate of the underlying trend of an integrated process with regime-switching parameters when $E^M\left[y_{t+j}|\Omega_t\right]=E\left[y_{t+j}|\Omega_t\right]$. Meanwhile, the approach is general in the sense that it simplifies to the BN decomposition in the absence of regime switching.

2.3 UC Models

A direct way to conduct trend/cycle decomposition is to consider a finite-order parametric specification for the Wold form of the transitory component in (3). For example, a standard assumption is a finite-order stationary AR process:

$$\phi(L)c_t=\omega^*_{t-j}, \tag{3’}$$

where, again, the standard identification assumption for the mean of the cycle is that $\overline{\omega}=0$. If the shocks to the trend and the cycle in (2) and (3’) are assumed to be Gaussian (i.e., $(\eta_t,\omega_t)'\sim N(0,\Sigma_{\eta\omega})$), the Kalman filter can be employed to make optimal inferences about the trend and cycle. As discussed in Morley, Nelson, and Zivot (2003), the inferences based on the Kalman filter will be the same as those based on the BN decomposition given equivalent models of the autocovariance structure of a time series.

It is possible to extend linear UC models to allow for regime-switching parameters. For example, Lam (1990) considers the case where the drift parameter is regime switching (i.e., $\mu=\mu(S_t)$ in (2)). Kim and Nelson (1999a) consider the case where the mean of the innovations to the cycle is regime switching (i.e., $\overline{\omega}=\overline{\omega}(S_t)$ in (3’)). As with the RDSS decomposition, it is necessary for identification to assume a "normal" regime i^* in which the cycle is mean zero. Therefore, unlike the BN decomposition, there is no implicit assumption that the cycle is unconditionally mean zero. Indeed, the RDSS decomposition and inference based on optimal filtering will be the same given equivalent models of the autocovariance structure of the time series, while the BN decomposition will, in general, be different and not optimal.

3. Model-Based Measures of the U.S. Business Cycle

In this section, we present a range of model-based measures of the post-war U.S. business cycle. To keep the scope of our analysis manageable, we consider only univariate models. While this might seem to preclude useful multivariate information in terms of forecasting, we find that the set of models considered here covers the full range of possibilities in terms of the degree of predictability of U.S. real GDP. Also, in practice, trend/cycle decomposition is usually presented as a prior step to cross-series analysis. In particular, we are often interested in whether and how trend and cycle components of one time series are related to the trend and cycle components of many other series. As a general method, then, it is particularly useful if trend/cycle decomposition can be applied first at a univariate level and then the resulting measures of trend and cycle are considered in different multivariate settings. For instance, this is the approach taken in studies that use the Hodrick-Prescott filter or a bandpass filter and could help explain their popularity despite their tendency to produce spurious cycles (see Cogley and Nason, 1995, and Murray, 2003).

In terms of linear models, we consider AR models for the first differences of y_t:

$$\phi(L)(\Delta y_t - \mu) = e_t, \tag{7}$$

where $\phi(L)$ is a p^{th} order lag polynomial with roots outside the unit circle. We consider versions of the AR models with Gaussian errors (i.e., $e_t \sim N(0,\sigma_e^2)$) or Student *t* errors (i.e., $e_t \sim t(\nu,0,\sigma_e^2)$) and a set of values for the lag order *p* ranging from 0 to 12. Beyond the AR models, we also consider three UC models. The first model (UC-HP) is due to Harvey and Jaeger (1993) and corresponds to the Hodrick-Prescott filter with a smoothing parameter of 1600.[3] The second model (UC-0) has a standard UC specification as in (1), (2), and (3'), with an independent AR(2) cycle (i.e., $\eta_t \sim N(0,\sigma_\eta^2)$, $\omega_t \sim N(0,\sigma_\omega^2)$, and $E[\eta_t \omega_t] = 0$). The third model (UC-UR) has the same structure as the second model except that, following Morley, Nelson, and Zivot (2003), it allows for correlation between permanent and transitory movements by assuming a general variance-covariance matrix $\Sigma_{\eta\omega}$ for the shocks. Because the Kalman filter assumes Gaussian shocks, we do not consider Student *t* errors for the UC models.

In terms of nonlinear forecasting models, we consider Hamilton's (1989) model and different versions of Kim, Morley, and Piger's (2005) bounceback model. If the Hamilton model can be said to correspond to "L"-shaped recessions, with the economy growing from a permanently lower level following the end of recessions, the bounceback models allow for post-recession high-growth recoveries, with the three cases of "U"-shaped recessions, "V"-shaped recessions, and recoveries that are proportional to the "Depth" of the preceding recession. Each of these nonlinear models can be expressed as an AR model with a regime-switching mean that potentially depends on the current and *m* lagged states:

$$\phi(L)(\Delta y_t - \mu_t) = e_t \tag{8}$$

[3] Harvey and Jaeger's (1993) UC model assumes a random walk with drift for the permanent component, where the drift itself follows a random walk, plus noise for the transitory component (i.e., an AR(0) cycle). The variance of the shock to the permanent component is assumed to be zero, while the variance of the shock to the drift is assumed to be 1/1600 times as large as the variance of the noise, which is freely estimated.

$$\mu_t = \mu(S_t, ..., S_{t-m}), \tag{9}$$

where $S_t = \{0,1\}$ is a Markov state variable with fixed continuation probabilities $\Pr[S_t = 0 \mid S_{t-1} = 0] = p_{00}$ and $\Pr[S_t = 1 \mid S_{t-1} = 1] = p_{11}$. The Hamilton and bounceback models differ by their specifications for the time-varying mean:

1. Hamilton (H)

$$\mu_t = \gamma_0 + \gamma_1 S_t, \tag{10}$$

2. "U"-Shaped Recessions (BBU)

$$\mu_t = \gamma_0 + \gamma_1 S_t + \lambda \sum_{j=1}^{m} \gamma_1 S_{t-j}, \tag{11}$$

3. "V"-Shaped Recessions (BBV)

$$\mu_t = \gamma_0 + \gamma_1 S_t + (1 - S_t)\lambda \sum_{j=1}^{m} \gamma_1 S_{t-j}, \tag{12}$$

4. Recovery based on "Depth" (BBD)

$$\mu_t = \gamma_0 + \gamma_1 S_t + \lambda \sum_{j=1}^{m} (\gamma_1 + \Delta y_{t-j}) S_{t-j}, \tag{13}$$

The state $S_t = 1$ is identified as a low-growth regime by assuming $\gamma_1 < 0$. Following Kim, Morley, and Piger (2005), we assume $m = 6$ for the bounceback models, which allows recoveries to persist for up to six quarters following the end of a recession. Again, for these nonlinear AR models, we consider both cases of Gaussian errors and Student *t* errors.

In terms of nonlinear UC models, we consider Kim and Nelson's (1999a) version of Milton Friedman's "plucking" model (UC-FP-0) and a version due to Sinclair (2007) that allows for correlation between permanent and transitory shocks (UC-FP-UR). These models augment the linear UC-0 and UC-UR models described earlier by allowing for a regime-switching mean of the cyclical component in (3'):

$$\overline{\omega} = \tau S_t , \tag{14}$$

where S_t is defined as for the nonlinear AR models and the state $S_t = 1$ is identified by assuming $\tau < 0$. As in the linear case, we only consider Gaussian shocks for the nonlinear UC models.

While there are many other nonlinear models, these cover the range of possibilities in terms of whether recessions are permanent or transitory. In particular, the Hamilton model assumes the effects of regime switches into recessions are completely permanent, the plucking model assumes they are completely transitory, and the bounceback models allow for both possibilities and everything in between.

A substantial literature has documented evidence of structural breaks in the parameters of time-series models for post-war U.S. real GDP growth. In particular, the evidence for a structural break in volatility parameters sometime during 1984 (i.e., the so-called "Great Moderation") is as close to incontrovertible as it gets in time series analysis of macroeconomic data, and several studies have pointed out the importance of accounting for these volatility changes when estimating regime-switching models (e.g. Kim and Nelson, 1999b; McConnell and Perez-Quiros, 2000). Although less overwhelming than the evidence for the Great Moderation, there is also some evidence for a reduction in mean growth rates in the early 1970s (i.e., the so-called "productivity slowdown") that has been considered in a number of studies (e.g., Perron, 1989; Zivot and Andrews, 1992; Bai, Lumsdaine and Stock, 1998). In a recent paper, Perron and Wada (2005) argue that controlling for the productivity slowdown is crucially important for U.S. business cycle measurement. In particular, they show that measures of the business cycle for different UC models are less sensitive to the particular model used once a one-time break in the long-run average growth rate of U.S. real GDP is allowed for in 1973. Thus, for all the models that we consider, we allow for a break in long-run growth in the first quarter of 1973, and a break in conditional variance in the second quarter of 1984.[4] We will discuss the implications of allowing for these structural breaks where it is relevant for our inferences about the business cycle.

For trend/cycle decomposition given the linear forecasting models, we use the BN decomposition or, in the case of the UC models, the Kalman filter. Note, again, that the filtered inferences from the Kalman filter are equivalent to the BN decomposition using the corresponding reduced-form forecasting models of the UC models. For trend/cycle decomposition given the nonlinear forecasting models, we use the RDSS decomposition or, in the case of the nonlinear UC models, the Kim (1994) filter, which combines the Kalman filter with Hamilton's (1989) filter for Markov-switching models.[5] For the nonlinear models, we

---

[4] In practice, to keep the addition of parameters across models similar, each structural break is accommodated by a single-parameter scalar shift. In particular, for capturing the Great Moderation in the linear AR models, the conditional variance of the error term is allowed to change. For the nonlinear AR models, both the conditional variance and the difference in growth rates across regimes change (see Kim and Nelson, 1999b), but the change in conditional variance is assumed to be proportional to the squared change in the difference in growth rates. Finally, for both the linear and nonlinear UC models, we allow for a scalar shift in the level of the variance-covariance matrix of permanent and transitory shocks. For capturing the productivity slowdown in the linear AR models, the mean growth rate parameter is allowed to change. For the nonlinear AR models, the structural break is accommodated by a change in the mean growth rate of the high-growth (expansion) regime. For all UC models other than the UC-HP model, the drift parameter of the trend component is allowed to change. For the UC-HP model, no additional change in drift is allowed for, as the drift parameter for this model follows a random walk process and is thus already time varying. More complicated patterns of structural change for the nonlinear and UC models produce only small improvements in fit and have little effect on inferences about the business cycle.

[5] The Kim filter only approximates conditional expectations, but it is relatively straightforward to implement and has been found to produce very similar inferences to optimal methods for nonlinear UC models (see Kim, 1994). Meanwhile, even though the RDSS decomposition involves exact conditional expectations, it is based on the assumption, discussed in Morley and Piger (2008), that permanent and transitory innovations only depend on current or lagged regimes. This assumption, which is explicitly made for the nonlinear UC models, implicitly holds for the BBU and BBV models, meaning that the specification of a "normal" regime only affects the level of the cycle, not its magnitude or shape. However, for the BBD model, the transitory effects of past shocks can depend on future regimes because the model has implicit regime-switching autoregressive coefficients. In this case, the level and magnitude of the cycle can, in principle, be affected by the particular assumed sequence of future regimes, although the general shape will be robust. We have considered an extended version of the RDSS decomposition that allows for more complicated patterns for future regimes. Given the same "normal" regime at long horizons, we have found very similar measures of the cycle, including in terms of magnitude, to what is produced by the application of (6) to the BBD model. Thus, for simplicity of presentation, we consider the basic RDSS decomposition.

assume $i^* = 0$, corresponding to a cycle that is mean zero in expansions. We discuss this assumption in more detail in Section 6.

The raw data are seasonally-adjusted quarterly U.S. real GDP for the sample period of 1947:Q1 to 2006:Q4 and are taken from the St. Louis Fed (FRED) database. We conduct maximum likelihood estimation (MLE) for all of the models. To facilitate with model comparison, we need to ensure that the adjusted sample period is equivalent for all of the models. Complicating matters is the practical difficulty of conducting exact MLE for the nonlinear AR models. Our solution to this problem is to backcast a suitable number of observations prior to 1947:Q1 based on the long-run average growth rate. We then conduct conditional MLE for all of the models based on the same adjusted sample of 1947:Q2-2006:Q4 for 100 times the first differences of the natural logs of real GDP. For the UC models, we can incorporate the conditioning information provided by backcast observations by starting with a highly diffuse prior on an initial level of the trend component and using backcast observations as a training sample prior to evaluation of the likelihood. However, we found that whether we assumed up to 12 lags of growth in the training sample (equivalent to conditional MLE for the AR(12) models and implicitly nesting the relevant information for all of the lower-order AR models) or no training sample had little impact on the UC model estimates and likelihood values. Indeed, the likelihood values were slightly higher when we did not use a training sample based on backcast observations. Thus, for the UC models, we assume a diffuse prior for the level of the trend component in 1947:Q1 and evaluate the likelihood for 1947:Q2-2006:Q4.

For traditional model comparison, we use the Akaike and Schwarz Information Criteria (AIC and BIC). Later in the paper, we consider formal Bayesian model comparison based on marginal likelihoods for the different models under consideration. Model choices based on AIC and BIC are asymptotically equivalent to choices based on marginal likelihoods in the respective cases that prior information is as precise as the likelihood and prior information is relatively diffuse. Thus, AIC and BIC provide useful benchmarks with which to compare our marginal likelihood results. Meanwhile, in terms of information theory, Smith, Naik, and Tsai (2006) show that AIC places too small of a penalty on Markov-switching parameters from the perspective of minimizing the Kullback-Leibler divergence from the true model when the true model involves Markov switching, while Psaradakis and Spagnolo (2003) find that BIC tends to underestimate the true number of Markov-switching regimes. Thus, we also use Smith, Naik, and Tsai's "Markov switching criterion" (MSC) that is designed to minimize the Kullback-Leibler divergence in the general setting of regression models with Markov-switching parameters. Appendix B provides details on how we have calculated these information criteria.

Table 1 reports AIC, MSC, and BIC results for the various linear models, while Figure 1 reports the corresponding cycle measures. Beginning with the results for the AR models, AIC and MSC both pick the AR(4) model with Gaussian errors, while BIC picks the AR(2) model with Gaussian errors. Looking at Figure 1, we can see that the AR(2) and AR(4) cycles are small, noisy, and typically positive during NBER recessions. The reason for this counterintuitive result is that both models imply positive serial correlation at short horizons. Specifically, when output falls in a recession, there is a prediction of further declines (or at least below-average growth) in the short run, suggesting that output is above its long-run (i.e., trend) level. By contrast, the AR(12) model produces a more traditional-looking cycle that typically turns negative during NBER-dated recessions. In this case, the model implies negative serial correlation at longer horizons. Thus, when output falls in a recession, there is a prediction of compensating above-average growth at some point in the future, suggesting that output is below its long-run level. The very different cycle for the AR(12) model is notable in part because, even though the model is heavily discounted by BIC, it has a sizable improvement in likelihood over even the AR(8) model and is reasonably close to the low-order AR models when considering either AIC or MSC.

In terms of the UC models, it is interesting to note how similar the HP cycle looks to the AR(12) cycle discussed above. However, there are very different scales for these two cycles, with the HP cycle being much larger in amplitude. Furthermore, the UC-HP model has an extremely poor fit as judged by any of the information criteria, suggesting that the autocovariance structure implied by the UC-HP model is strongly at odds with the data. In terms of model comparison, BIC favours the UC-0 model, while AIC and MSC slightly favour the UC-UR model. Interestingly, from Figure 1, both the UC-0 and UC-UR cycles are large and persistent and have a similar pattern to the AR(12) cycle.

Comparing across all the linear models, the model selection criteria produce a mixed signal about the nature of the cycle. In particular, the most preferred model as judged by AIC or MSC is the UC-UR model, which, from Figure 1, produces a large, traditional business cycle that implies an important role for transitory fluctuations. The most preferred model as judged by the BIC is the AR(2) model with Gaussian errors. In contrast to the UC-UR model, the AR(2) produces a small, non-traditional business cycle, implying that most short-run fluctuations in output are permanent.

Before turning to the nonlinear models, it is worth comparing the results obtained for the linear models to the conclusions of Perron and Wada (2005), who argue that the sensitivity of model-based measures of the business cycle is due to a failure to account for structural breaks. In particular, they show that the cycles implied by different methods and specifications of linear models look more similar once a one-time break in the long-run growth rate of U.S. real GDP is allowed for in 1973. While the Perron and Wada finding is corroborated by the similarity in the implied cycles generated by the UC-0 and UC-UR models, the implied cycles for the low-order

AR models are small and with counterintuitive sign. Thus, contrary to Perron and Wada's claim, allowing for structural breaks does not resolve the sensitivity of business cycle measures to model specification.[6]

Table 2 reports AIC, MSC, and BIC results for the various nonlinear models, while Figure 2 reports the corresponding cycle measures.[7] Beginning with the nonlinear AR models, the results in Table 2 suggest that the bounceback model, in all of its versions, is preferred to the corresponding Hamilton model. Evidently, the idea that the regime switches correspond to only permanent movements in the level of output is not supported by the data. The nonlinear UC models fare somewhat better than the Hamilton model, at least according to AIC and MSC. However, all versions of the bounceback model are preferred to the nonlinear UC models regardless of the criteria considered.

Comparing across all the nonlinear models, each of the three criteria choose the "Depth" version of the bounceback model with no linear dynamics and with Gaussian errors. However, from the perspective of measuring the business cycle, all versions of the bounceback models yield similar results. In particular, from Figure 2, the different bounceback models all imply cycles that are highly asymmetric with a similar shape and amplitude. In particular, the cycles are close to zero during expansions and negative during recessions. This pattern suggests a direct

---

[6] Indeed, the robustness of the implied cycle for the UC-UR could also be questioned. In unreported results we have found a second peak of the likelihood function, only slightly lower than the global maximum, which implies a business cycle that looks very similar to that implied by the low-order AR models or the UC-UR model without structural breaks. Meanwhile, the implied cycles for all of the AR models are highly robust to structural break assumptions, suggesting that the sensitivity of UC model inferences could be due to weak identification of such models. In particular, it can be statistically difficult to distinguish between a large and volatile trend component and a large and highly persistent cycle component. As a consequence, there tends to be a pile-up problem in the form of UC model estimates implying a large and persistent cycle, especially in small samples, even when the true data generating process involves a volatile trend (see Nelson, 1988, and Ma and Nelson, 2008). Allowing for a one-time structural break in the long-run growth rate of U.S. real GDP shifts the estimates from a volatile trend to a large and persistent cycle, although, consistent with the notion that weak identification is behind the result, this finding is overturned when considering multivariate information (see Sinclair, forthcoming).

[7] We omit the cycle for the Hamilton model with linear AR(2) dynamics from Figure 2 because it is linear and similar to the cycles for the low-order linear AR models. Also, just as we omitted the linear AR(0) model from Figure 1 because there is no cycle for such a model, we omit the Hamilton model with an AR(0) assumption from Figure 2 because there is no cycle.

link between the NBER definition of the business cycle and the transitory component of U.S. real GDP. Specifically, in terms of recessions, the NBER appears to be identifying periods when there are large and negative transitory movements in real economic activity. We discuss the economic implications of these asymmetric measures of the business cycle in greater detail in Section 6. The implied cycles for the nonlinear UC models look more linear than those for the bounceback models, although they still have a negative mean.[8] However, again, these nonlinear UC models do not fare well under any of the model selection criteria.

Finally, we turn to the comparison between linear and nonlinear models. In doing so, we are particularly interested in the MSC results, which are designed to mimic AIC within each class of models, but to allow for comparison across models with different numbers of Markov-switching regimes, including models with one regime. According to the MSC results, the best model is the "Depth" version of the bounceback model with no linear dynamics. Indeed, it is the top overall choice based on AIC and BIC too. Among the linear models, only the AR(1) and AR(2) models come close, and only in terms of BIC. Notably, the findings in favour of the bounceback model are robust to allowing Student *t* errors, implying that it is the ability of the model to capture nonlinear dynamics, rather than fat tails in the unconditional distribution of output growth, that explains its empirical success.[9]

Based on traditional model comparison, then, the question of what the business cycle looks like can be narrowed somewhat. In particular, all three information criteria give the edge to a nonlinear model that implies an asymmetric business cycle, with the cycle close to zero during

---

[8] Again, consistent with a weak identification problem for UC models, we have found a second peak in the likelihood for the plucking models that implies a business cycle that looks a lot like that for the bounceback models and for the plucking models without structural breaks (or only one or the other break). Meanwhile, the cycle measures for the bounceback models are quite robust to different assumptions about structural breaks, although the cycles for models without structural breaks are somewhat larger in amplitude prior to 1984 and smaller afterwards.

[9] We found more support for Student *t* errors when considering models without structural breaks. However, this directly suggests that it is really the structural breaks (especially the Great Moderation) that matter for output growth, rather than fat tails in the error distribution.

expansions and negative during recessions. Based on the BIC only, a linear model that implies a small, nontraditional, cycle is competitive with the nonlinear models. Meanwhile, the model selection criteria appear to rule out linear models that imply large, symmetric cycles. At the same time, one should be cautious about relying solely on traditional model selection criteria to compare between linear and nonlinear models. Thus, we turn to more formal tests of nonlinearity and model comparison in the next two sections.

4. Tests of Nonlinearity

As discussed in the previous section, model comparison based on traditional information criteria supports a nonlinear model with no linear dynamics, namely the BBD-AR0 version of the bounceback model. However, as is evident in Tables 1 and 2, when comparing a more general version of the bounceback model that incorporates AR(2) dynamics to its nested linear AR(2) counterpart, BIC favours the linear model (although AIC and MSC continue to favour the nonlinear model). Thus, despite the evident importance of linearity versus nonlinearity in terms of implications for a model-based business cycle, it is not clear which case would be supported by formal statistical tests of nested models. In this section, we take up formal testing of nonlinearity within the context of a few of the key models considered in the previous section. In particular, we consider a null hypothesis of a linear AR(2) model and compare it to the nonlinear alternatives of the Hamilton model and the bounceback models.

Testing for nonlinearity of the Markov-switching form is difficult due to the presence of unidentified nuisance parameters under the null hypothesis of linearity and the singularity of the information matrix at the null. There have been different proposed tests to address this nonstandard testing environment, most notably by Hansen (1992) and Garcia (1998). Recently, Carrasco, Hu, and Ploberger (2007) (CHP hereafter) developed a relatively straightforward information-matrix-based test that is optimal for local alternatives to linearity and only requires estimation under the null hypothesis of linearity. While estimation under the null might only seem like a small advantage given that we have already estimated the alternative models in the previous section, it is helpful because the test requires parametric bootstrap experiments to assess statistical significance. The bootstrap experiments are made easier by only having to estimate models under the null of linearity given data generated under the null. On the other hand, there are some limitations in terms of what alternatives can be considered with the CHP test. Thus, at the end of this section, we also discuss results for parametric bootstrap experiments to assess the significance of a likelihood ratio (LR) test for nonlinearity.[10] In this case, the bootstrap experiments require estimation under the alternative. We address difficulties in estimating under the alternative by considering a grid of possible values for the continuation probabilities (see Kim, Morley, and Piger, 2005) and by considering a large number of starting values for MLE.

The CHP test can be applied to a broad set of random coefficient models, the most prominent of which are models with Markov-switching parameters. In addition to presenting the general test, CHP discuss how to implement the test in the specific case where parameters depend only on the current realization of a two-state Markov-switching process (also see Hamilton, 2005, for an accessible discussion of how to implement the test in this case). In Appendix C, we provide details on how to implement the test for a broader range of Markov-switching models, including the Hamilton and bounceback models considered in this paper, for which parameters can depend on current and lagged values of the state variable.

For our tests, we consider a linear AR(2) model as the null hypothesis. As in the previous section, we assume Gaussian or Student *t* errors and we allow for structural breaks in mean and variance. In terms of alternatives, we consider Hamilton's (1989) model, which implies "L"-

[10] Di Sanzo (forthcoming) uses Monte Carlo analysis to investigate the small sample properties of some statistical tests for Markov switching, including the CHP test and a parametric bootstrap LR test. He finds that both tests have good size properties, but that the LR test has higher power for the specific data generating processes considered in his study.

shaped recessions, and the “U”-shape-recession version of the bounceback model (BBU). Note that we do not consider the “V”-shape-recession version of the bounceback model (BBV) because the CHP test requires that the regime-switching parameters are a linear function of the state variables, while the BBV model has the mean depend in part on the product of the current and lagged states. The “Depth” version of the bounceback model (BBD) is also difficult to cast into the CHP framework due to the interaction between the lagged states and lagged growth rates. Thus, we consider a bootstrap LR test instead of the CHP test for the BBD alternative.

Table 3 presents the results for the tests of linearity with an AR(2) model as the null. For the Hamilton model as an alternative, we are unable to reject linearity using the CHP test. However, for the BBU model as an alternative, the bootstrap *p*-value of the CHP test is 0.03, meaning that we can reject the null of linearity at the standard 5% level. Likewise, the bootstrap *p*-value is 0.03 for the LR test with the BBD model as an alternative. These results are robust to consideration of models with Student *t* errors. Thus, using formal hypothesis tests, there is strong evidence for nonlinearity given nonlinear alternatives that allow for high-growth recoveries following the end of recessions.

However, while we are able to reject linearity for specific nonlinear alternatives, it must be acknowledged that the more alternatives we consider, the more we are faced with the potential of a size distortion in our overall test of linearity. Thus, in terms of measuring the business cycle, we are left with the slightly unsatisfactory situation that our inferences depend crucially on close to knife-edge test results about whether a linear model or nonlinear model provides a better description of the autocovariance structure of U.S. real GDP growth. Given this sensitivity of model-based measure of the business cycle and a fundamental uncertainty about the underlying role of nonlinearity in output growth dynamics, an obvious response is to model this uncertainty and use it to put weights on different measures of the business cycle in order to construct a "model-free" measure. This is the approach that we take in the next section.

5. A Model-Free Measure

In this section, we take a Bayesian approach to inference. This approach allows us to put probability weights on different models and, therefore, different measures of the business cycle. We can use these weights for formal model comparison and to construct a model-free measure that averages across the implied cycles for different models. In addition, the Bayesian approach also produces posterior distributions for the business cycle measures that provide us with a sense of the importance of parameter uncertainty for our inferences.

It is worth mentioning that, even without taking a Bayesian viewpoint, constructing a model-free measure of the business cycle has some justification. In particular, the BN and RDSS decomposition methods that we use to estimate the business cycle involve constructing long-horizon conditional forecasts based on time series models. It has long been understood in the forecasting literature that combined forecasts can outperform individual forecasts (see, for example, Bates and Granger, 1969). Thus, combining model-based business cycle estimates could produce a measure with a lower mean-squared-error than any of the individual estimates. Of course, the principle of combining forecasts does not answer the question of exactly how to combine forecasts. In this paper, we choose to use Bayesian model probabilities to construct weights on different forecasts.

For Bayesian inference, we employ Markov-chain Monte Carlo (MCMC) methods for posterior simulation and for calculation of marginal likelihoods for each model under consideration. Posterior simulation is done via a single-block Metropolis-Hastings (MH) algorithm with a random-walk chain. For each model, we generate 20,000 draws from the posterior after a burn-in of 1,000 draws. To calculate the marginal likelihood for a given model, we use the Chib and Jeliazkov (2001) method based on the "Bayes identity" and the MH output. We have verified that results are very similar across different runs of the MH sampler with different starting values.

Bayesian inference requires the specification of a prior distribution for the parameters of each model under consideration. Unfortunately, even if posterior inferences about model parameters are robust to different priors, posterior inferences about models themselves are less so. The reason is that marginal likelihoods for models, which are a crucial input for the construction of posterior model probabilities, are constructed by integrating out parameters from the likelihood over the prior:

$$f(y|M_i) = \int f(y|\theta_i, M_i)\, p(\theta_i|M_i)\, d\theta_i \ , \qquad (15)$$

where M_i is an indicator for the model, $f(y|M_i)$ is the model's "marginal likelihood", θ_i is the set of parameters for the model, $f(y|\theta_i, M_i)$ is the model's likelihood function, and $p(\theta_i|M_i)$ is the prior distribution for the model's parameters. Conceptually, the marginal likelihood evaluates the fit of a model by considering the predictive density of the data based on the model and the prior. This is a perfectly sensible way to evaluate a model and is analogous to an out-of-sample forecast comparison in the classical setting. However, it makes it clear why the prior on parameters is important, even asymptotically. A high prior weight on a parameter value that produces a poor forecast of the data will result in a model having a relatively low marginal likelihood. The practical consequence of this importance of priors for Bayesian model comparison is that models with uninformative priors and many parameters are almost always dominated by models with uninformative priors and fewer parameters. This favoritism of parsimony is fine if it is indeed a researcher's prior. However, if it is merely a shortcut to avoid the difficult task of prior elicitation in a time series context, then it is a problem because it makes model comparison based on marginal likelihoods meaningless.

In this paper, we strive to make our priors similarly informative across models. Table 4 presents the details of the prior distributions. Our benchmark priors (denoted as “Prior 1”) target a similar unconditional mean and variance for output growth across models. These priors are not particularly informative in the sense that posterior inferences about parameters are similar to maximum likelihood estimates. However, for robustness, we also consider highly uninformative priors (denoted as “Prior 2”) that correspond to substantially more uncertainty about the unconditional mean and variance of output growth, with the degree of uncertainty increasing with the complexity of the model.

In order to resolve any conflicting results driven by the different priors, we consider “partial” marginal likelihoods (see Lempers, 1971) in addition to the standard marginal likelihood given in (15). The partial marginal likelihood involves calculation of the marginal likelihoods for only a portion of the sample after an initial training sample. The training sample is used to determine posterior distributions that are used as priors for the calculation of the marginal likelihoods for the latter portion of the sample. Specifically, the sample is partitioned into an initial portion for prior determination and a latter portion for model evaluation:

$$f(y_{\kappa+1},..,y_T | y_1,..,y_\kappa, M_i) = \int f(y_{\kappa+1},..,y_T | y_1,..,y_\kappa, \theta_i, M_i)\, p(\theta_i | y_1,..,y_\kappa, M_i)\, d\theta_i\,, \quad (16)$$

where $0 < \kappa < T$. In (16), $p(\theta_i | y_1,..,y_\kappa, M_i)$ represents the posterior distribution for θ_i conditional on the training sample, $y_1,..,y_\kappa$, and will itself depend on the initial prior distribution, $p(\theta_i | M_i)$. To the extent that the training sample posterior is dominated by the likelihood, the partial marginal likelihood should be relatively robust to the initial parameter prior distribution. This approach is directly analogous to constructing a pseudo out-of-sample forecast, where a portion of the total sample is used for parameter estimation and the remaining sample is used for model evaluation. As such, the approach could be sensitive to the particular out-of-sample period chosen. Furthermore, there is little guidance for choosing how to partition the sample into training sample and evaluation sample. If the goal is to evaluate models, then it makes sense to have as long an evaluation sample as possible. If the goal is to minimize the impact of initial priors, then it makes sense to have as long a training sample as possible. We consider using the data prior to 1970 as a training sample and the sample period of 1970-2006 for model evaluation. We find that this training sample is long enough to eliminate most of the influence of initial priors, yet the evaluation sample is long enough to include a number of business cycle episodes that are important for distinguishing linear and nonlinear models.

The partial marginal likelihood approach is appealing as a means of making model comparison reasonably "objective" in a Bayesian setting. It certainly makes prior determination more transparent and allows the researcher to avoid even implicit "empirical Bayes" in which priors are based on the same sample data used to evaluate models. It also makes initial prior specification less crucial, which is useful given that prior elicitation can be extremely challenging for complicated models.[11]

Tables 5 and 6 report the log marginal likelihoods and log partial marginal likelihoods for the various linear and nonlinear models under consideration. Starting with the log marginal likelihoods and considering both sets of priors, the ranking of the models corresponds closely to

[11] In models with structural breaks, the partial marginal likelihood could still be strongly influenced by the initial priors on post-break parameters if the breaks occur after the end of the training sample. Fortunately, as structural breaks are common across the set of models we consider, this is not a particular problem in our case. Of course, this would be more problematic in comparing models with and without structural breaks. However, in unreported results, we find the evidence for structural breaks is so strong that even given very uninformative priors on post-break parameters, partial marginal likelihoods strongly favour the models that include structural breaks in all cases. As the models without structural breaks receive essentially zero weight in posterior probability calculations, we have chosen to focus only on the models that include structural breaks.

the BIC results in Tables 1 and 2, which is not particularly surprising given that a BIC statistic can be thought of as an approximation of the marginal likelihood given relatively uninformative priors. As with the BIC results, the low-order AR models are preferred to the higher-order AR models and UC models, while the BBD version of the bounceback model is preferred over the other nonlinear specifications. In terms of comparing across linear and nonlinear models, the rankings generally correspond to the BIC results for the benchmark priors (Prior 1), with the BBD-AR0 model being the most preferred model overall. However, for the highly uninformative priors (Prior 2), the more parsimonious linear AR(2) model comes out ahead.

Focusing on the benchmark priors, Table 7 reports parameter estimates for the linear AR(2) model and the nonlinear BBD-AR0 model, both with Gaussian errors. The similarity between the posterior inferences and MLE makes it clear that the benchmark priors are not strongly informative. Figures 3 and 4 provide a visual summary of the estimation results by plotting posterior (solid lines) and prior (dashed lines) densities for the parameters of the AR(2) and BBD-AR0 models, respectively. The posteriors are always dominated by the likelihood, although the priors for the parameters related to the difference in growth rates and the continuation probabilities for the nonlinear model are set so as to place no weight on the region of the parameter space that corresponds to a linear model (i.e., if there is no difference in growth rates between regimes or if one of the regimes persists for the entire sample period). Finally, Figure 5 displays the posterior median and 95% equal-tailed credibility bands for the cycles implied by these two models. Again, the similarity of the shapes to the corresponding cycles in Figures 1 and 2 reveals how much more important the likelihood is than the priors in determining the posterior inferences. Also, it is notable how precise the inferences about the cycle are, especially for the nonlinear model. Indeed, for the bounceback model, the posterior probability that the cycle goes negative during NBER recessions is always greater than 95%.

While the posterior inferences about the business cycle largely reflect the likelihood and not the prior, we cannot escape the fact that the overall ranking of linear and nonlinear models depends on priors. At the same time, posterior uncertainty about which model is best suggests a role for model-averaging in order to construct a model-free measure of the business cycle. To do so, it is necessary to determine the posterior probabilities for the different models under consideration. The two elements of these calculations are prior model probabilities and the marginal likelihoods reported in Tables 5 and 6:

$$p(M_i|y) \propto f(y|M_i)\,p(M_i), \tag{17}$$

where $p(M_i|y)$ and $p(M_i)$ are the posterior and prior model probabilities, respectively. For prior model probabilities, we assign equal weights for the two classes of linear and nonlinear models. Within each class of models, we assign equal weights for each specification. For specifications with different error distributional assumptions, we assign equal weights for each distributional assumption.

Figure 6 displays Bayesian model-averaged measures of the business cycle using model probabilities for the two sets of initial priors. Considering the top panel, it is easy to see the role of priors in determining the implied business cycle. The solid line corresponds to the business cycle measure for the benchmark priors. In this case, the cycle is large and asymmetric reflecting the large weight put on the bounceback model. Meanwhile, the dashed line corresponds to the business cycle measure given highly uninformative priors. In this case, the cycle is small and symmetric, reflecting the large weight put on the AR(2) model. Despite the very different measures of the business cycle for the two priors, the bottom panel reveals a way to address the sensitivity of inferences to priors. In particular, the bottom panel presents business cycle measures for the two initial priors, but allowing for a training sample of the data prior to 1970 before evaluating partial marginal likelihoods for the models. The similarity in inferences for the two initial priors reflects the fact that the posterior inferences for the various model parameters are dominated by the likelihood by 1970 and, therefore, depend little on the initial priors. Meanwhile, the implied cycles based on the partial marginal likelihoods correspond more closely to the cycle based on the benchmark prior and the full-sample marginal likelihoods, although its amplitude is somewhat smaller.

The partial marginal likelihoods for the sample of 1970-2006 are reported in Tables 5 and 6. Again, reflecting a diminishing importance of initial priors, the partial marginal likelihoods based on both Prior 1 and Prior 2 give the highest probability to the "Depth" version of the bounceback model with AR(0) dynamics and Student *t* errors.[12] It is also notable that the partial marginal likelihood results give sizeable posterior weight to a number of different models, including linear models, suggesting more benefit of model-averaging than for the full-sample marginal likelihoods, where almost all of the weight went to one model or another (i.e., the "Depth" version of the bounceback model for the benchmark priors and the AR(2) model for the highly uninformative priors).

Because the implied cycles are robust to initial priors when using the training sample, we use the measure based on the benchmark priors and the partial marginal likelihoods calculated for 1970-2006 as our "model-free" measure of the business cycle (i.e. the solid line in the bottom panel of Figure 6). For ease of presentation, this model-free measure is displayed by itself in Figure 7. It is a truly "model-free" measure in the sense that the weighting that underlies it

[12] The preference for models with Student *t* errors reflects the fact that priors based on data up to 1970 imply a large error variance with high precision, while the evaluation sample from 1970 contains data with both high volatility in the 1970s and data with low volatility from the mid-1980s and on. Even though all of the models allow for a structural break in variance in 1984, the prior on the post-break variance is quite diffuse. Thus, the models with Student *t* errors are more successful at predicting the heterogeneity in volatility over the full post-1970 data than models with Gaussian errors.

reflects an averaging across a number of models, with the weights being appropriate to integrate out model uncertainty.[13] The model-free cycle is noticeably asymmetric and negative during NBER-dated recessions. It has a large negative mean, but unlike the model-based measures for the bounceback models, the cycle does sometimes venture into positive territory (e.g., in the late 1960s and late-1990s). With this model-free measure in hand, we turn next to the question of whether it is a spurious artifact of an admittedly complicated filtering of U.S. real GDP data, or whether it captures the behaviour of a meaningful macroeconomic phenomenon that is sometimes referred to as "the business cycle".

6. Economic Relevance

The picture of the asymmetric business cycle in Figure 7 is striking in terms of its negative mean and correspondence to NBER-dated recessions. However, it is an open question as to whether it has any economic relevance. An obvious potential relevance is the implication that, if gaps from trend are negative on average, then stabilization policy, if successful, could raise the average level of output (see DeLong and Summers, 1988, and Yellen and Akerlof, 2006, on this point). In particular, the estimated means of the model-free cycle and the cycle for the preferred bounceback model (BBD-AR0) based on MLE are -0.40 and -0.78, respectively, implying that the potential welfare benefits of stabilization are at least an order of magnitude higher than suggested by Lucas (1987, 2003) based on a symmetric cycle.[14] The first issue, then,

[13] Technically, the model-free cycle is based on the weighted-average of the posterior means of the cycles for the different models. Thus, it is the overall posterior mean of the cycle given the prior weights on the different models. Put another way, the model uncertainty is integrated out according to the posterior probability weights for each model based on the partial marginal likelihoods.

[14] We report both the mean of the model-free cycle and the mean of the cycle for the bounceback model because the model-free cycle may be better at capturing the shape of the business cycle than its magnitude. In particular, to the extent that the true mean of the cycle is negative, the equal prior weight on linear models and nonlinear models implies a large prior weight on a zero-mean cycle that shrinks the model-free measure of the business cycle towards zero. It might be thought that the large prior weight on the nonlinear models pushes the mean of the model-free measure below zero. However, it should be noted that the nonlinear models nest the possibility of an unconditionally mean-zero cycle, even when priors constrain the growth rate to be lower in the "recession" regime. For example, the Hamilton model always produces an unconditionally mean zero cycle. Meanwhile, the prior of the bounceback

is whether the negative mean of the cycle is meaningful or if it is merely an arbitrary consequence of the assumption that the mean is zero in the expansion regime for the nonlinear AR and UC models.

To address the issue of whether we can meaningfully identify the mean of the cycle, rather than just normalize it, we start with Figure 8, which displays the estimated cycle for the preferred bounceback model. The cycle and the change in the cycle are presented for both possible assumptions about which regime has a mean-zero cycle, that is for $i^* = 0$ and $i^* = 1$. From the figure, it is clear that the shapes of the cycles for both assumptions are very similar, but the levels are very different.[15] Meanwhile, except for scale, the *changes* in the cycle are essentially the same across the two assumptions. Importantly, the changes in the cycle are zero at the same points of time in both cases. However, from a conceptual point of view, it only makes sense that the change in the cycle would be zero when output is at its "steady-state" level. Thus, the assumption of a mean-zero cycle in expansions is the only assumption that corresponds to our definition of the cycle as measuring transitory deviations from a steady-state trend.

Another method of evaluating the relevance of the negative mean for the cycle is to consider whether the sign of the cycle is correlated with the sign of other variables that should behave differently depending on whether the economy is above or below trend, such as the change in inflation or the change in the unemployment rate. Of course, it is not possible to use this approach to evaluate a cycle, such as the $i^* = 0$ cycle in Figure 8, that is always negative. However, we can use this approach to evaluate our "model-free" measure of the cycle in Figure 7 that is sometimes positive and sometimes negative. For inflation, we use the percentage change

parameter is centered at zero, implying that the prior mean of the cycle for the bounceback models is also zero. The prior mean of the cycle for the plucking model is negative, but this model receives relatively small prior weight and virtually no posterior weight in the construction of the model-free business cycle.

[15] The structural break in the growth difference parameter shifts the level of the cycle in the case that $i^*=1$, but has no impact of the level of the cycle in the case that $i^*=0$. Also, as discussed in footnote 5, the magnitude of the cycle is affected by the specification of i^*.

in core CPI. The raw data for core CPI and unemployment rate are taken from the FRED database. The series for core CPI and the unemployment rate cover the sample periods of 1958:Q1 to 2006:Q4 and 1948:Q1 to 2006:Q4, respectively. The correlation between the sign of the model-free cycle and the sign of the subsequent four-quarter change in inflation is 0.30, while the correlation with the sign of the contemporaneous four-quarter change in the unemployment rate is -0.28. Thus, these results support the identification assumptions that underlie the negative mean of the model-free cycle. Interestingly, the equivalent correlations are -0.12 and 0.38 for the AR(2) measure of the cycle. That is, inflation actually rises and unemployment rate is falling when the AR(2) cycle is negative and vice versa.

Beyond the issue of the mean of the cycle, a more direct indication of the relevance of a measure of the business cycle is its more general co-movement with other macroeconomic variables. Two variables that are often thought to correspond closely to the business cycle are the level of the unemployment rate and the level of capacity utilization. The raw data for capacity utilization are also taken from the FRED database and cover the sample period of 1967:Q1 to 2006:Q4. While the unemployment rate and capacity utilization are thought to reflect the business cycle, there appear to be some permanent movements in these series that complicate statistical analysis (i.e., we would need to consider trend/cycle decomposition for these variables as well if we wanted to conduct formal correlation analysis between the cycles of the different series). Figure 9 plots the model-free business cycle against the unemployment rate. While the long-run variation in the unemployment is fairly easy to see, there is also a clearly visible relationship between the model-free business cycle and variation in the unemployment rate. Likewise, while there is some apparent long-run variation in capacity utilization in Figure 10, there is also a remarkably strong relationship with the model-free business cycle. These strong relationships are notable because the model-free business cycle is based purely on univariate information in U.S real GDP. Meanwhile, the model-free business cycle has the advantage over the unemployment rate or capacity utilization of capturing transitory fluctuations in real economic activity, while abstracting from all long-run variation.

Given an asymmetric business cycle, there is an immediate question of what it implies for the behaviour of inflation. Again considering core CPI inflation, we find that the model-free business cycle is consistent with a convex short-run aggregate supply curve. Figure 11 displays the subsequent four-quarter change in inflation against the model-free business cycle in the first panel and against the four-quarter growth in money supply when the model-free cycle is positive in the second panel. Money supply growth is measured using M2 data, which are taken from the FRED database and cover the sample period of 1958:Q1 to 2006:Q4. The convex shape of the short-run aggregate supply curve is fairly apparent in the first panel, with large disinflations occurring primarily when the business cycle is large and negative and large increases in inflation occurring when the business cycle is close to zero or somewhat positive. While the first panel makes it clear that a large negative cycle can bring down inflation, it is not as clear why a positive cycle sometimes leads to higher inflation and sometimes does not. The second panel addresses this to some extent by showing that large increases in inflation are often associated with large increases in money supply. Of course, the wide distribution of inflation responses in both panels of Figure 11 given the business cycle and money supply growth directly implies that other factors, such as inflation expectations, would be necessary to help explain the movements (or lack thereof) in inflation.

In terms of long-run variation in output, Figure 12 plots U.S. real GDP against the model-free trend and against the trend for the preferred bounceback model based on MLE.[16]

[16] Again, given the large prior weight on linear models, the model-free measure is likely to be better at capturing the shape of the business cycle than its magnitude. Therefore, it may somewhat overstate the variation in trend during

Corresponding to the asymmetric behaviour of the model-free cycle, real GDP is close to trend during expansions and drops below trend in recessions. Thus, both measures of trend are consistent with the idea that movements in output are largely permanent in expansions, as though they are driven by productivity shocks. Meanwhile, there is little sign of "technological regress" in recessions because these events correspond to transitory movements in output below a trend level that rarely undergoes an outright decline itself, especially given the trend based on the bounceback model. Thus, the asymmetry of the business cycle has the potential to resolve a major debate in macroeconomics over the relative importance of productivity and demand shocks in driving fluctuations in economic activity. The volatile measures of trend suggest that permanent productivity shocks are more important, except in recessions, which appear to correspond to large negative transitory demand shocks. Thus, these results are consistent with Temin's (2008) argument that economists should move beyond the standard real business cycle model to explain periods of recession or "depression".

Overall, the results in the section suggest that, beyond the obvious link to NBER-dated recessions, the model-free business cycle captures a meaningful macroeconomic phenomenon. Specifically, the 'alternating-phases' and 'output-gap' notions of the business cycle appear to be related, with a measure based on the 'output-gap' notion having a strong relationship to traditional measures of economic "slack" such as the unemployment rate and capacity utilization. Meanwhile, the apparent asymmetry of the business cycle has the potential to explain the perseverance of two very different theories of the source of macroeconomic fluctuations, with permanent productivity shocks driving the economy in expansions and transitory demand shocks pulling output below trend during recessions.

---

recessions. Thus, to get a better sense of implied variation in trend if the test results for nonlinearity are taken at face value, we also report the trend for the bounceback model.

7. Conclusion

While model-based measures of the business cycle for the postwar U.S. economy are highly dependent on the particular time series model used to capture output growth dynamics, we have found strong support for a nonlinear model that captures a high-growth recovery following the end of recessions and implies a highly asymmetric cycle that is large and negative in NBER-dated recessions and close to zero in expansions. The nonlinear model is supported by traditional model selection criteria, formal statistical testing for nonlinearity, and Bayesian model comparison. However, we argue that there is enough uncertainty about the most appropriate model of output growth dynamics and that the measure of the business cycle is sensitive enough to model specification that it is helpful to average across model-based measures of the cycle using Bayesian model probabilities in order to construct a "model-free" measure of the business cycle. We find that this model-free business cycle is strongly asymmetric despite equal prior weights on linear and nonlinear models. Also, we find that the model-free cycle is closely related to other measures of economic slack, implies a convex short-run aggregate supply curve, and helps reconcile two long-standing, but competing theories about the main source of macroeconomic fluctuations. Based on our analysis, we conclude that the model-free business cycle captures a meaningful macroeconomic phenomenon and provides useful insights about aggregate economic activity above and beyond what can be gleaned by only looking at the level or growth rates of U.S. real GDP.

References

Bai, J., Lumsdaine, R.L., and J.H. Stock, 1998, Testing for and dating common breaks in multivariate time series, *Review of Economic Studies* 65, 395-432.

Bates, J.M. and C.W.J. Granger, 1969, The combination of forecasts, *Operations Research Quarterly (OR)* 20, 451-468.

Beveridge, S. and C.R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle, *Journal of Monetary Economics* 7, 151-174

Carrasco, M., Hu, L., and W. Ploberger, 2007, Optimal test for Markov switching, Working Paper.

Chib, S. and I. Jeliazkov, 2001, Marginal Likelihood from the Metropolis-Hastings Output, *Journal of the American Statistical Association* 96, 270-281.

Clarida, R. H. and M.P. Taylor, 2003, Nonlinear Permanent-Temporary Decompositions in Macroeconomics and Finance, The Economic Journal 113, C125-C139.

Cogley, T. and J.M. Nason, 1995, Effects of the Hodrick-Prescott filter on trend and difference stationary time series: Implications for business cycle research, Journal of Economic Dynamics and Control 19, 253-278.

DeLong, B. and L. Summers, 1988, How does macroeconomic policy affect output? *Brookings Papers on Economic Activity* 2, 433-480.

Di Sanzo, S., forthcoming, Testing for linearity in Markov switching models: A bootstrap approach, *Statistical Methods and Applications*.

Garcia, R., 1998, Asymptotic null distribution of the likelihood ratio test in Markov switching models, *International Economic Review* 39, 763-788.

Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357-384.

Hamilton, J.D., 2005, What's real about the business cycle? *Federal Reserve Bank of St. Louis Review* 87, 435-452.

Harvey, A.C. and A. Jaeger, 1993, Detrending, stylized facts and the business cycle, *Journal of Applied Econometrics* 8, 231-247

Hansen, B.E., 1992, The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, *Journal of Applied Econometrics* 7, S61-S82.

Harding, D. and A.R. Pagan, 2002, Dissecting the cycle: A methodological investigation, *Journal of Monetary Economics* 49, 365-381.

Kim, C.-J., 1994, Dynamic linear models with Markov switching, *Journal of Econometrics* 60, 1-22.

Kim, C.-J., J. Morley, and J. Piger, 2005, Nonlinearity and the permanent effects of recessions, *Journal of Applied Econometrics* 20, 291-309.

Kim, C.-J. and C.R. Nelson, 1999a, Friedman's plucking model of business fluctuations: Tests and estimates of permanent and transitory components, *Journal of Money, Credit and Banking* 31, 317-34.

Kim, C.-J. and C.R. Nelson, 1999b, Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economic and Statistics* 81, 608-616.

Lam, P.-S., 1990, The Hamilton model with a general autoregressive component: Estimation and comparison with other models of economic time series, *Journal of Monetary Economics* 26, 409-432.

Lempers, F.B., 1971, *Posterior Probabilities of Alternative Linear Models* (Rotterdam University Press).

Lucas, R.E., 1987, *Models of Business Cycles* (Basil Blackwell, Oxford).

Lucas, R.E., 2003, Macroeconomic priorities, *American Economic Review* 93, 1-14.

Ma, J. and C.R. Nelson, 2008, Valid inference for a class of models where standard inference performs poorly; Including nonlinear regression, ARMA, GARCH, and unobserved components, Working Paper.

McConnell, M.M. and G. Perez-Quiros, 2000, Output fluctuations in the United States: What has changed since the early 1980s? *American Economic Review* 90, 1464-1476.

Mitchell, W.A., 1927, *Business Cycles: The Problem and Its Setting* (NBER, New York).

Morley, J., 2002, A state-space approach to calculating the Beveridge-Nelson decomposition, *Economics Letters* 75, 123-127.

Morley, J.C., Nelson, C.R. and E. Zivot, 2003, Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *Review of Economics and Statistics* 85, 235-243.

Morley, J. and J. Piger, 2008, Trend/cycle decomposition of regime-switching processes, *Journal of Econometrics* 146, 220-226.

Murray, C.J., 2003, Cyclical properties of Baxter-King filtered time series, *Review of Economics and Statistics* 85, 472-476.

Nelson, C.R., 1988, Spurious trend and cycle in the state space decomposition of a time series with a unit root, *Journal of Economic Dynamics & Control* 12, 475-488.

Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361-1401.

Perron, P. and T. Wada, 2005, Let's take a break: Trends and cycles in U.S. real GDP, Working Paper.

Psaradakis, Z. and N. Spagnolo, 2003, On the determination of the number of regimes in Markov-switching autoregressive models, *Journal of Time Series Analysis* 24, 237-252.

Sinclair, T.M., 2007, Asymmetry in the business cycle: Revisiting the Friedman plucking model, Working Paper.

Sinclair, T.M., forthcoming, The relationships between permanent and transitory movements in U.S. output and the unemployment rate, *Journal of Money, Credit, and Banking*.

Smith, A., P.A. Naik, and C.-L. Tsai, 2006, Markov-switching model selection using Kullback-Leibler divergence, *Journal of Econometrics* 134, 553-577.

Temin, P., 2008, Real business cycle views of the Great Depression and recent events: A review of Timothy J. Kehoe and Edward C. Prescott's Great Depressions of the Twentieth Century, *Journal of Economic Literature* 46, 669-684.

Wen, Y., 2002, The business cycle effects of Christmas, *Journal of Monetary Economics* 49, 1289-1314.

Yellen, J.L. and G.A. Akerlof, 2006, Stabilization policy: A reconsideration, *Economic Inquiry* 44, 1-22.

Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the great crash, the oil price shock, and the unit root hypothesis, *Journal of Business and Economic Statistics* 10, 251-270.

Appendix A

This appendix provides formulas for the BN measure of trend implied by an autoregressive model, as well as for the RDSS measure of trend implied by the Hamilton and bounceback regime switching models.

Given a linear AR model of Δy_t, such as was given in (7), the BN trend for y_t can be easily calculated analytically using the state-space method in Morley (2002) as

$$\hat{\tau}_t^{BN} = y_t + HF(I-F)^{-1}(\Delta\tilde{y}_t - \tilde{\mu}), \tag{A.1}$$

where

$$H = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix},\ F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\Delta\tilde{y}_t = (\Delta y_t, \ldots, \Delta y_{t-p+1})', \text{ and } \tilde{\mu} = (\mu, \ldots, \mu)'.$$

Then, for a nonlinear AR model, such as was given in (8)-(9), the RDSS trend can be calculated as:

$$\hat{\tau}_t^{RDSS}(\tilde{S}_t) = y_t + HF(I-F)^{-1}(\Delta\tilde{y}_t - \tilde{\mu}_t) + \sum_{j=1}^{\infty}\left(E\left[\mu_{t+j}\left|\{S_{t+k} = i^*\}_{k=1}^{j}, \tilde{S}_t, \Omega_t\right.\right] - \overline{\mu}_{i^*}\right), \tag{A.2}$$

where H, F, and $\Delta\tilde{y}_t$ are the same as in (A.1), $\tilde{\mu}_t = (\mu_{S_t}, \ldots, \mu_{S_{t-m+1}})'$, and $\overline{\mu}_{i^*} = \mu(i^*, \ldots, i^*)$. The first two terms on the right hand side of (A.2) are analogous to the calculation of the BN estimate of trend in (A.1), with the second term corresponding to forecastable momentum due to linear dynamics. The third term arises due to nonlinear dynamics and accounts for forecastable momentum implied by any difference between the future time-varying mean and the regime-dependent average growth rate in regime i^*. When $i^* = 0$ for the models considered in this paper, the summation in (A.2) can be truncated at $j = m$ and calculated analytically. In general, the infinite sum can be calculated via simulation. Once $\hat{\tau}_t^{RDSS}\left(\tilde{S}_t\right)$ is calculated, $\tilde{S}_t$ can then be integrated out as in (5) using the probability weights, $p^M\left(\tilde{S}_t \middle| \Omega_t\right)$, to arrive at $\hat{\tau}_t^{RDSS}$. These probability weights can be obtained from the recursive filter given in Hamilton (1989).

Appendix B

This appendix provides details on how the information criteria considered in this paper are calculated. Note that because the versions of AIC, BIC, and MSC considered here place positive weight on model fit and negative weight on the number of model parameters, we seek to maximize these criteria.

For a given model, we calculate AIC and BIC as follows:

$$AIC = l(\theta) - k, \tag{B.1}$$

$$BIC = l(\theta) - \frac{\ln(T)}{2}k, \tag{B.2}$$

where $l(\theta)$ is the log-likelihood given parameter values θ, k is the number of parameters, and T is the sample size. The formula for MSC is more complicated and is given as follows:

$$MSC = l(\theta) - k_f \frac{T}{T - k_f - 1} - \frac{1}{2}\sum_{i=1}^{N} \frac{\hat{T}_i(\hat{T}_i + Nk_s)}{\max\left\{\hat{T}_i - Nk_s - 2,\ 1\right\}} + \frac{1}{2}T, \tag{B.3}$$

where N is the number of regimes, k_s is the number of Markov-switching parameters, k_f is the number of fixed parameters other than the continuation probabilities and a variance term (i.e., $k_f = k - N(N-1) - Nk_s - 1$), and $\hat{T}_i = \sum_{t=1}^{T} \Pr\left[S_t = i \middle| \Omega_T\right]$. The formula in (B.3) is modified somewhat from Smith, Naik, and Tsai (2006) to be on the same scale as AIC and BIC and to allow for k_f parameters to be fixed when $N > 1$. For the Markov-switching models considered in this paper, $N = 2$ and $k_s = 1$. In terms of interpreting the results for MSC, it should be noted that the formula in (B.3) is asymptotically equivalent to AIC when $N = 1$ and it incorporates a penalty for possible regime switching in the variance term when $N > 1$, although we always assume fixed variances in our Markov-switching models. See Smith, Naik, and Tsai (2006) for more details on the derivation of MSC and its relationship to other model selection criteria.

Appendix C

This appendix provides some details on how to implement the CHP test developed by Carrasco, Hu, and Ploberger (2007) in the setting where the regime-switching alternative allows parameters to depend on both current and lagged values of the state variable. The CHP test considers the null hypothesis of constant parameters against an alternative hypothesis of switching parameters. Let θ_t denote the potentially time-varying parameters, which are a subset of all model parameters Θ. The null hypothesis is that $H_0 : \theta_t = \theta_0$. The alternative hypothesis is that $H_1 : \theta_t = \theta_0 + \theta_t^*$, where $\theta_t^* = H\tilde{\xi}_t$, which is the product of a matrix of possible changes in parameter values given switching regimes and $\tilde{\xi}_t$ is a vector of current and lagging zero-mean Markov state variables that determine the prevailing regime for the parameters. The state vector evolves according to $\tilde{\xi}_t = F\tilde{\xi}_{t-1} + w_t$, where the vector w_t follows a martingale difference sequence (i.e., $E\left[\tilde{\xi}_{t-1}{}' w_t\right] = 0$), with $E\left[w_t w_t{}'\right] = Q$. Lagged state variables are incorporated into $\tilde{\xi}_t$ using identities, implying zero elements in w_t.

The general form of the CHP test statistic is given as follows:

$$TS_T(\beta) = \Gamma_T - \frac{1}{2T}\hat{\varepsilon}'\hat{\varepsilon}, \tag{C.1}$$

where

$$\Gamma_T = \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\gamma_t(\beta), \tag{C.2}$$

$$\gamma_t(\beta) = tr\left(\left(l_{t,\theta}^{(2)} - l_{t,\theta}^{(1)}l_{t,\theta}^{(1)\prime}\right)E\left[\theta_t^*\theta_t^{*\prime}\right]\right) + 2\sum_{s<t} tr\left(l_{t,\theta}^{(1)}l_{s,\theta}^{(1)\prime}E\left[\theta_t^*\theta_s^{*\prime}\right]\right), \tag{C.3}$$

$$l_{t,\theta}^{(1)} = \frac{\partial \ln f(y_t|\tilde{y}_{t-1},\Theta)}{\partial\theta},\ l_{t,\theta}^{(2)} = \frac{\partial^2 \ln f(y_t|\tilde{y}_{t-1},\Theta)}{\partial\theta\partial\theta'}, \tag{C.4}$$

and $\hat{\varepsilon}$ is the vector of residuals from an OLS regression of $\frac{1}{2}\gamma_t(\beta)$ on $l_{t,\Theta}^{(1)}$ (i.e, the scores with respect to all of the parameters under the null, not just those that are hypothesized to switch under the alternative), with β denoting a vector of all of the nuisance parameters in H and F that are not identified under the null. Because of the presence of nuisance parameters, the test is based on the supremum test statistic for a set of considered values of the nuisance parameters (i.e., a "sup" test statistic $\sup TS = \sup_{\beta\in\bar{\mathrm{B}}} TS_T(\beta)$, where $\bar{\mathrm{B}}$ is a compact subset of all possible values of the nuisance parameters B). Note that for the sup statistic, the scale of the nuisance parameters in H is not identified because it cancels out in the first-order condition with respect to that scale parameter. Thus, the value of an arbitrary nuisance parameter that is assumed not to take on the value of zero under the alternative can be normalized to 1 and the test statistic can be constructed as

$$\sup TS = \sup_{\beta'\in\bar{\mathrm{B}}'} \frac{1}{2}\left(\max\left(0, \frac{\Gamma_T}{\sqrt{\hat{\varepsilon}'\hat{\varepsilon}}}\right)\right)^2, \tag{C.5}$$

where β' is a vector of the remaining nuisance parameters. Also, note that the expectations terms in (C.3) can be solved as $E\left[\theta_t^*\theta_t^{*\prime}\right] = HVar\left(\tilde{\xi}_t\right)H'$, where $vec(Var\left(\tilde{\xi}_t\right)) = (I - F \otimes F)^{-1} vec(Q)$, and $E\left[\theta_t^*\theta_s^{*\prime}\right] = HF^{t-s}Var\left(\tilde{\xi}_t\right)H'$.

The asymptotic distribution of the CHP test depends on nuisance parameters. As a result, CHP rely on parametric bootstrap experiments to calculate the critical values. These experiments involve simulating *B* bootstrap samples based on the estimated null model and calculating the test statistic for each of these simulated samples. Then, the percentage of simulated test statistics that are larger than the sample statistic determines the bootstrap *p*-value for the test, while the bootstrap critical value for a test with nominal size α can be found by sorting the bootstrap test statistics from smallest to largest and finding the $(1-\alpha)B$ test statistic or the next largest if $(1-\alpha)B$ is not an integer.

A simple example helps illustrate the CHP test. Consider the null of an AR(0) model $\Delta y_t = \mu + e_t$, $e_t \sim N(0,\sigma_e^2)$ against the alternative of a two-state Markov-switching mean $\Delta y_t = \mu_t + e_t$, where $\mu_t = \gamma_0 + \gamma_1 S_t$ and $S_t = \{0,1\}$ is a two-state Markov-switching state variable with fixed continuation probabilities $\Pr[S_t = 0 \mid S_{t-1} = 0] = p_{00}$ and $\Pr[S_t = 1 \mid S_{t-1} = 1] = p_{11}$. Then, letting $\bar{\pi} \equiv E[S_t] = (1 - p_{00})/(2 - p_{11} - p_{00})$, $\mu_t = \mu + \gamma_1\xi_t$, where $\mu = \gamma_0 - \gamma_1\bar{\pi}$ and $\xi_t \equiv S_t - \bar{\pi}$. Thus, in terms of the general CHP test, $H = \gamma_1$, $\tilde{\xi}_t = \xi_t$, and $F = \rho$, where $\rho = p_{00} + p_{11} - 1$. Note that it is necessary to normalize the variance of the unobserved state variable in order to identify the magnitude of γ_1. We do this by setting $Q = 1$. Then, in constructing the test statistic, we set $\gamma_1 = 1$ and find the largest test statistic for $\rho \in (0.02, 0.98)$. In practice, we only consider positive values for ρ because alternatives with persistent regimes are what we are interested in for U.S. real GDP. Specifically, we are not interested in considering regime-switching models in order to capture outliers. Instead, we consider these models in order to capture persistent business cycle phases. This restriction can lower the value of the test statistic and, in some cases, it does. Of course, the same restriction is imposed when calculating the bootstrap distribution of the test statistic, so it can also lower the critical value of the test. Also, it is worth mentioning that we want to avoid the case where $\rho = 0$ because given the assumption of a linear Gaussian model under the null hypothesis, the second derivatives in (C.3) are equal to the negative of the outer-product of the scores (i.e., $l_{t,\theta}^{(2)} = -l_{t,\theta}^{(1)} l_{t,\theta}^{(1)\prime}$), meaning that $\gamma_t(\beta) = 0$ and, therefore, making it impossible to run the OLS regression to find $\hat{\varepsilon}$. As discussed in CHP, the test has no power in this case. On the other hand, the test has nontrivial power in other cases, including, in this setting, when $\rho \neq 0$.[17]

[17] The formal condition for nontrivial power is far more complicated and discussed in detail in CHP.

Table 1
Information Criteria for Linear Models

Model	Log Likelihood	# of Parameters	Akaike Information Criterion	Markov Switching Criterion	Schwarz Information Criterion
AR(0)	-297.50	4	-301.50	-301.56	-308.46
AR(1)	-287.00	5	-292.00	-292.10	-300.70
AR(2)	-283.97	6	-289.97	-290.11	**-300.40**
AR(4)	-281.52	8	-289.52	-289.77	-303.42
AR(8)	-280.92	12	-292.92	-293.51	-313.78
AR(12)	-273.66	16	-289.66	-290.74	-317.47
AR(0)-t	-296.64	5	-301.64	-301.74	-310.33
AR(1)-t	-286.29	6	-292.29	-292.43	-302.72
AR(2)-t	-283.51	7	-290.51	-290.70	-302.68
AR(4)-t	-280.94	9	-289.94	-290.26	-305.58
AR(8)-t	-280.37	13	-293.37	-294.28	-315.96
AR(12)-t	-273.62	17	-290.62	-291.86	-320.17
UC-HP	-505.15	2	-507.15	-507.17	-510.63
UC-0	-282.00	7	-289.00	-289.19	-301.17
UC-UR	-280.70	8	**-288.70**	**-288.95**	-302.60

Maximum likelihood estimation is based on the conditional likelihood for 1947:Q2-2006:Q4. Observations prior to 1947:Q2 are backcast based on the mean growth rate. The AIC, MSC and BIC are formulated such that the highest value (in bold) represents the preferred model.

Table 2
Information Criteria for Nonlinear Models

Model	Log Likelihood	# of Parameters	Akaike Information Criterion	Markov Switching Criterion	Schwarz Information Criterion
H-AR0	-286.13	7	-293.13	-294.42	-305.29
H-AR2	-282.71	9	-291.71	-293.18	-307.35
BBU-AR0	-280.34	8	-288.34	-289.69	-302.25
BBU-AR2	-279.77	10	-289.77	-291.20	-307.15
BBV-AR0	-281.24	8	-289.24	-290.56	-303.15
BBV-AR2	-279.33	10	-289.33	-290.73	-306.71
BBD-AR0	-277.95	8	**-285.95**	**-287.32**	**-299.86**
BBD-AR2	-277.14	10	-287.14	-288.58	-304.53
H-AR0-t	-285.83	8	-293.83	-295.14	-307.74
H-AR2-t	-282.27	10	-292.27	-293.77	-309.65
BBU-AR0-t	-280.19	9	-289.19	-290.57	-304.84
BBU-AR2-t	-279.62	11	-290.62	-292.10	-309.75
BBV-AR0-t	-280.88	9	-289.88	-291.22	-305.53
BBV-AR2-t	-278.97	11	-289.97	-291.41	-309.09
BBD-AR0-t	-277.59	9	-286.59	-287.97	-302.23
BBD-AR2-t	-276.53	11	-287.53	-288.99	-306.65
UC-FP-0	-281.38	10	-291.38	-292.23	-308.76
UC-FP-UR	-280.09	11	-291.09	-291.98	-310.21

Maximum likelihood estimation is based on the conditional likelihood for 1947:Q2-2006:Q4. Observations prior to 1947:Q2 are backcast based on the mean growth rate. The AIC, MSC and BIC are formulated such that the highest value (in bold) represents the preferred model.

Table 3
Tests of Nonlinearity

Null		Alternatives		
		L-shape (Hamilton)	U-shape (BBU)	Depth (BBD)
AR(2)	Test Statistic	0.13	4.28	13.32
	(*p*-value)	(0.52)	(0.03)	(0.03)
	95% crit. val.	1.40	3.97	12.20
AR(2)-t	Test Statistic	0.53	4.72	13.40
	(*p*-value)	(0.19)	(0.03)	(0.03)
	95% crit. val.	1.57	3.96	12.11

The test statistics for the L-shaped and U-shaped recession alternatives are based on Carrasco, Hu, and Ploberger (2007). The test statistics for the Depth-based recovery alternative are likelihood ratio statistics based on estimation using a grid for the continuation probabilities. All *p*-values and critical values are based on parametric bootstrap experiments with 499 simulations.

Table 4
Prior Distributions

Parameters	Description	Models	Prior 1	Prior 2
γ_0	Growth in High State	H, BB	$N(2,3^2)$	$N(2.5,10^2)$
$-\gamma_1$	Impact of low state	H, BB	$Gamma(\frac{100}{2},\frac{50}{2})$	$Gamma(\frac{15}{2},\frac{5}{2})$
λ	Bounceback coefficient	BB	$N(0,1)$	$N(0,10^2)$
$-\tau$	Mean of transitory innovations in low State	UC-FP-0, UC-FP-UR	$Gamma(\frac{100}{2},\frac{50}{2})$	$Gamma(\frac{15}{2},\frac{5}{2})$
μ	Unconditional mean growth	All except UC-HP, H and BB	$N(1,3^2)$	$N(1,10^2)$
ϕ_j	AR parameter at lag *j*	All except UC-HP	$TN(0,(2/j)^2)_{[\lvert z\rvert>1,\phi(z)=0]}$	-
$p_{00},\ p_{11}$	Markov state continuation probabilities	H, BB, UC-FP-0, UC-FP-UR	$Beta(4,16)$	-
ν	Degree of freedom for Student *t* errors	All except UC models	$Gamma(\frac{1}{2},\frac{0.1}{2})$	-
α_μ	Scale factor for structural break in drift	All except UC-HP	$Gamma(\frac{1}{2},\frac{1}{2})$	-
α_σ	Scale factor for structural break in variance	All	$Gamma(\frac{1}{2},\frac{1}{2})$	-
$\frac{1}{\sigma_e},\ \frac{1}{\sigma_\eta},\ \frac{1}{\sigma_\omega}$	Precision for independent shocks	All except UC-UR and UC-FP-UR	$Gamma(\frac{5}{2},\frac{2}{2})$	-
$\Sigma_{\eta\omega}^{-1}$	Precision for correlated shocks	UC-UR and UC-FP-UR	$Wishart(5,2\times I_2)$	-

Prior 1 is more informative than Prior 2. Distributions for Prior 2 are only reported when different than for Prior 1

Table 5
Bayesian Model Comparison for Linear Models

	Marginal Likelihood		Partial Marginal Likelihood	
Model	Prior 1	Prior 2	Prior 1	Prior 2
AR(0)	-309.65	-310.89	-158.65	-158.60
AR(1)	-301.07	-302.29	-155.65	-155.68
AR(2)	**-300.08**	**-301.28**	**-152.84**	**-152.80**
AR(4)	-301.74	-302.84	-155.07	-155.31
AR(8)	-306.92	-307.02	-157.25	-156.95
AR(12)	-304.68	-305.55	**-152.77**	-153.00
AR(0)-t	-309.68	-310.89	-158.19	-158.34
AR(1)-t	-301.22	-302.52	-156.28	-155.93
AR(2)-t	-300.53	-301.60	-153.27	-153.60
AR(4)-t	-302.14	-303.09	-155.25	-155.46
AR(8)-t	-306.81	-308.67	-156.73	-157.42
AR(12)-t	-305.69	-307.51	-161.76	-155.68
UC-HP	-510.84	-519.85	-294.09	-294.08
UC-0	-303.00	-304.23	-154.36	-154.36
UC-UR	-306.91	-308.43	-155.09	-153.64

Marginal likelihood calculations are based on the Chib and Jeliazkov (2001) method. “Marginal Likelihood” refers to the log marginal likelihood computed over the full sample (1947:Q2-2006:Q4), while “Partial Marginal Likelihood” refers to the log marginal likelihood computed over the subsample 1970:Q1-2006:Q4, using the data from 1947:Q2-1969:Q4 as a training sample. Partial marginal likelihoods are constructed using a prior that corresponds to the posterior for the training sample. “Prior” refers to initial priors before any training sample, with Prior 1 being more informative than Prior 2. The preferred models are in bold (models with log marginal likelihoods within 0.1 of each other are deemed to tie).

Table 6
Bayesian Model Comparison for Nonlinear Models

	Marginal Likelihood		Partial Marginal Likelihood	
Model	Prior 1	Prior 2	Prior 1	Prior 2
H-AR0	-302.09	-303.90	-153.13	-153.03
H-AR2	-304.30	-304.94	-154.15	-153.40
BBU-AR0	-299.73	-304.81	-154.17	-153.80
BBU-AR2	-303.24	-307.66	-154.23	-153.98
BBV-AR0	-301.93	-305.84	-155.59	-155.25
BBV-AR2	-304.85	-307.68	-160.56	-159.09
BBD-AR0	**-297.67**	**-303.01**	-153.02	-153.30
BBD-AR2	-300.97	-305.92	-153.58	-154.29
H-AR0-t	-302.67	-304.42	-154.38	-152.60
H-AR2-t	-304.72	-305.67	-154.83	-153.43
BBU-AR0-t	-300.41	-305.43	-153.70	-153.68
BBU-AR2-t	-303.75	-307.48	-153.87	-153.16
BBV-AR0-t	-302.90	-306.38	-156.85	-154.06
BBV-AR2-t	-305.75	-305.70	-157.63	-154.76
BBD-AR0-t	-298.16	-303.31	**-152.57**	**-152.46**
BBD-AR2-t	-301.99	-306.02	-153.37	-153.77
UC-FP-0	-310.93	-311.05	-162.01	-160.42
UC-FP-UR	-310.82	-307.20	-160.00	-158.85

Marginal likelihood calculations are based on the Chib and Jeliazkov (2001) method. “Marginal Likelihood” refers to the log marginal likelihood computed over the full sample (1947:Q2-2006:Q4), while “Partial Marginal Likelihood” refers to the log marginal likelihood computed over the subsample 1970:Q1-2006:Q4, using the data from 1947:Q2-1969:Q4 as a training sample. Partial marginal likelihoods are constructed using a prior that corresponds to the posterior for the training sample. “Prior” refers to initial priors before any training sample, with Prior 1 being more informative than Prior 2. The preferred models are in bold (models with log marginal likelihoods within 0.1 of each other are deemed to tie).

Table 7
Parameter Estimates for Preferred Linear and Nonlinear Models

Parameter	MLE	Posterior Mode	Posterior Mean	Posterior Median
		AR(2) Model		
$\mu_{pre1973}$	0.98	0.98	0.97	0.97
	(0.18)	(0.18)	(0.19)	[0.62, 1.36]
$\mu_{post1973}$	0.74	0.74	0.73	0.73
	(0.08)	(0.08)	(0.08)	[0.57, 0.90]
ϕ_1	0.24	0.24	0.25	0.25
	(0.06)	(0.06)	(0.06)	[0.12, 0.37]
ϕ_2	0.16	0.16	0.17	0.17
	(0.06)	(0.06)	(0.07)	[0.03, 0.30]
$\sigma^2_{e,pre1984}$	1.24	1.23	1.29	1.28
	(0.15)	(0.14)	(0.15)	[1.03, 1.64]
$\sigma^2_{e,post1984}$	0.20	0.20	0.21	0.21
	(0.03)	(0.03)	(0.03)	[0.16, 0.29]
		BBD-AR0 Model		
$\gamma_{0,pre1973}$	1.07	1.09	1.11	1.10
	(0.12)	(0.12)	(0.13)	[0.87, 1.38]
$\gamma_{0,post1973}$	0.86	0.86	0.86	0.86
	(0.05)	(0.05)	(0.05)	[0.76, 0.96]
$\gamma_{1,pre1984}$	-1.90	-1.90	-1.88	-1.87
	(0.20)	(0.16)	(0.16)	[-2.21, -1.56]
$\gamma_{1,post1984}$	-0.84	-0.83	-0.83	-0.82
	(0.10)	(0.09)	(0.09)	[-1.01, -0.65]
$\sigma^2_{e,pre1984}$	0.78	0.77	0.83	0.82
	(0.10)	(0.10)	(0.11)	[0.64, 1.06]
$\sigma^2_{e,post1984}$	0.15	0.15	0.16	0.16
	(0.02)	(0.02)	(0.02)	[0.12, 0.22]
λ	-0.10	-0.10	-0.10	-0.10
	(0.02)	(0.02)	(0.02)	[-0.15, -0.06]
p_{00}	0.94	0.93	0.92	0.92
	(0.02)	(0.02)	(0.02)	[0.87, 0.96]
p_{11}	0.79	0.80	0.79	0.79
	(0.07)	(0.05)	(0.05)	[0.67, 0.88]

Standard errors based on the negative inverse Hessian are given in parentheses beneath the MLEs and posterior modes. Posterior standard deviations are given in parentheses beneath the posterior means. 95% equal-tailed credibility bands are given in square brackets beneath the posterior median. Bayesian inferences are based on Prior 1. For the BBD model, the scale of the structural change is constrained to be the same for the growth rate differential and the standard deviation of the error term.

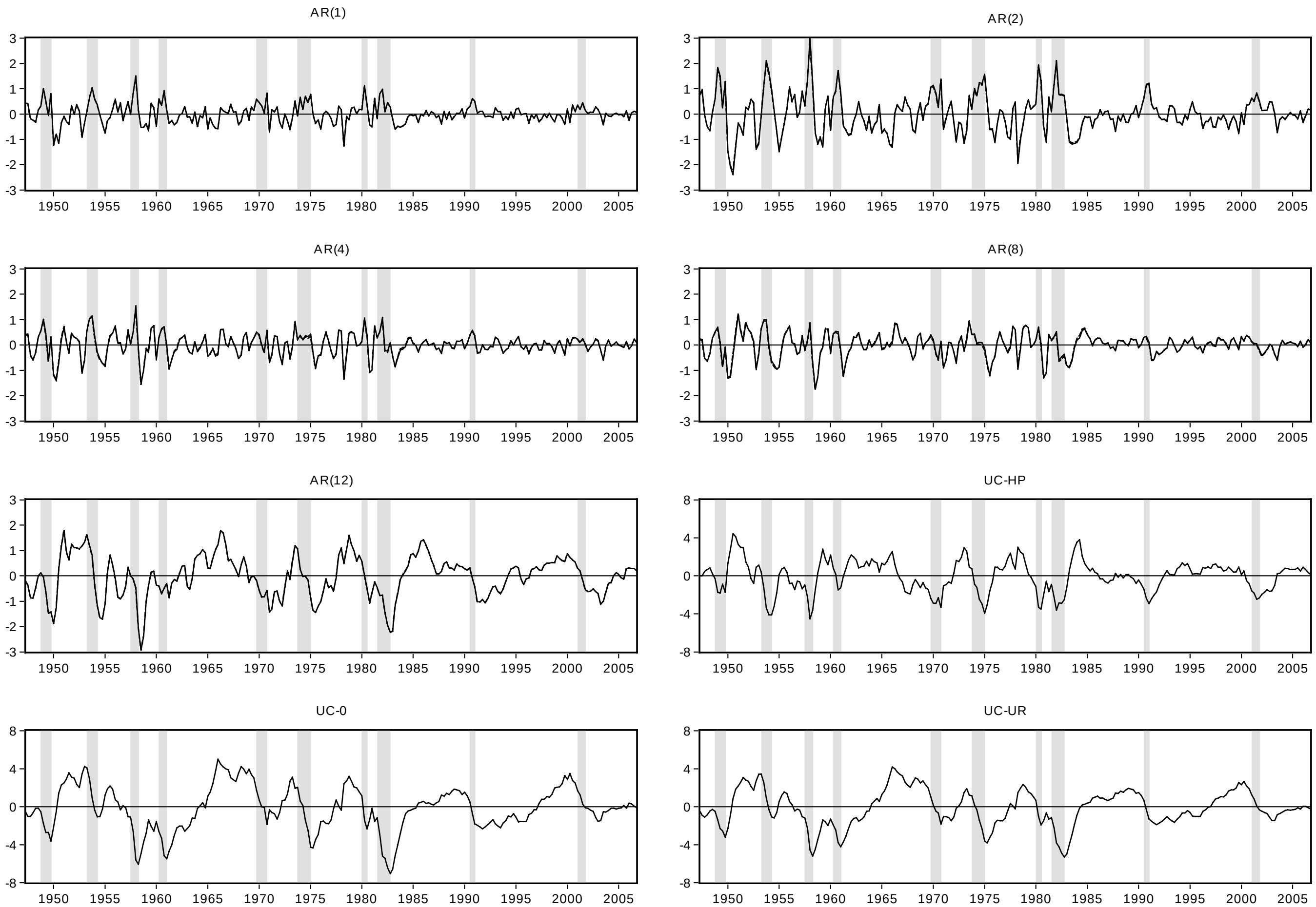

Fig. 1
Measures of the U.S. Business Cycle Based on Linear Models (NBER recessions shaded)

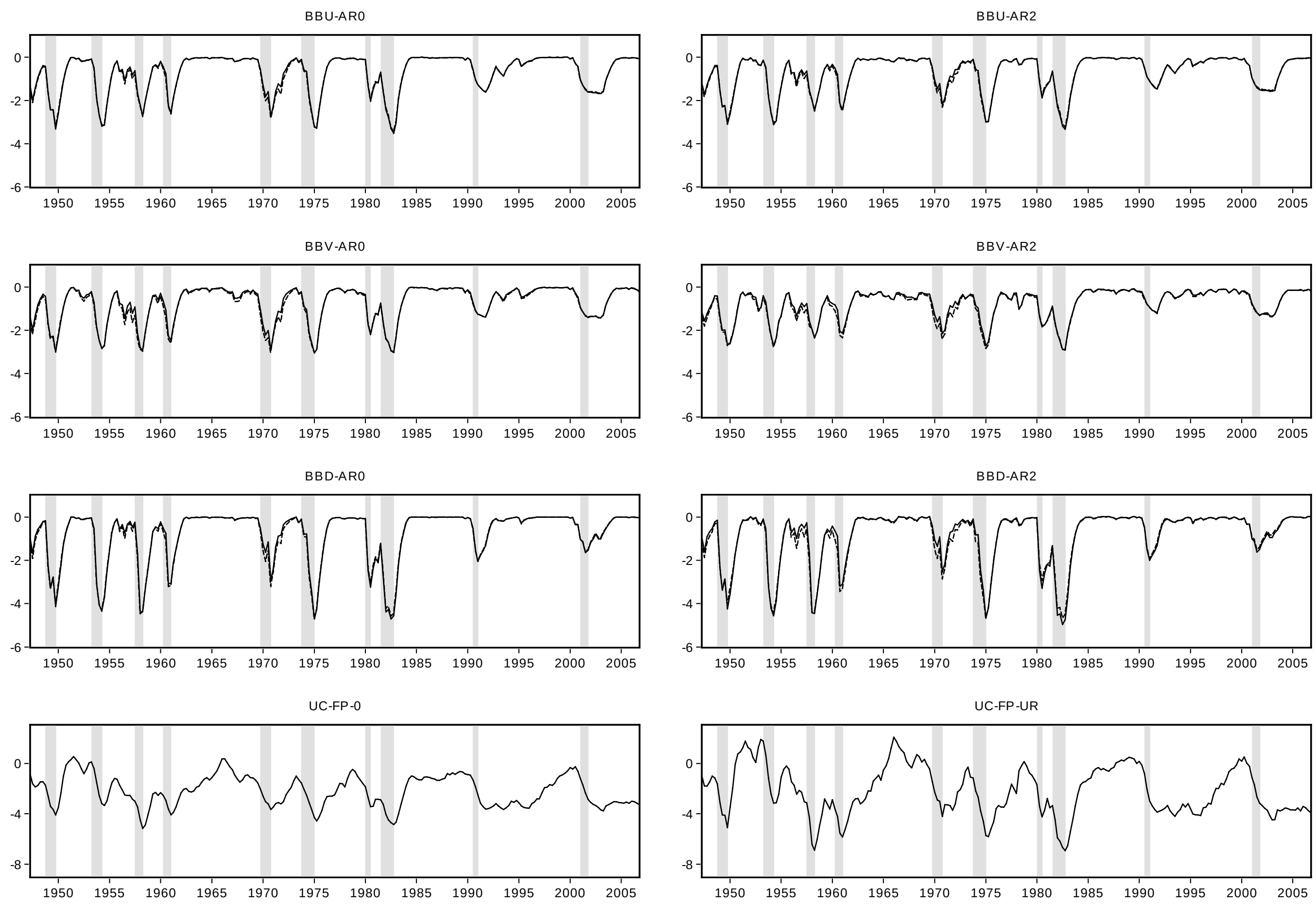

Fig. 2
Measures of the U.S. Business Cycle Based on Nonlinear Models (NBER recessions shaded)

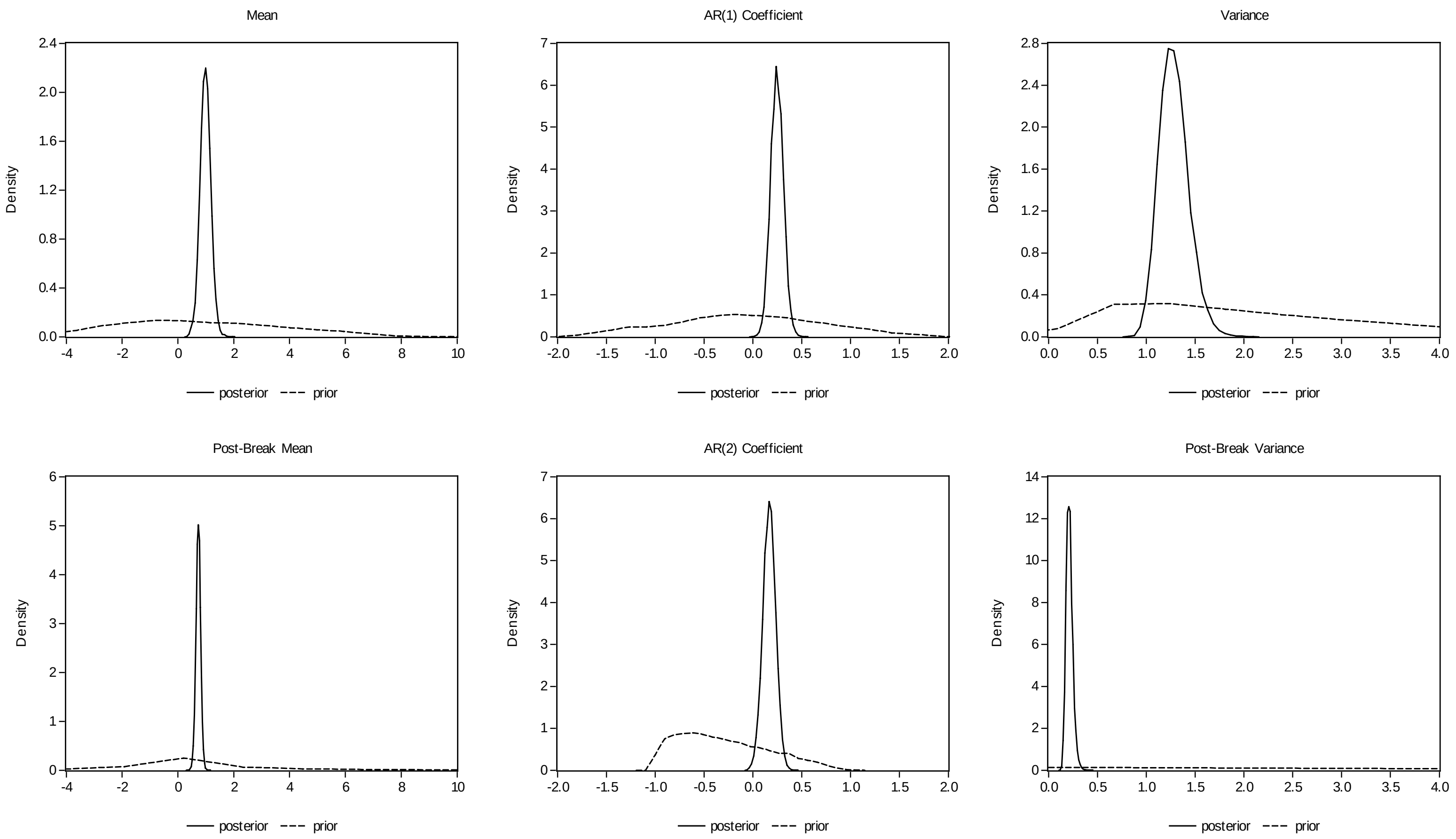

Fig. 3
Posterior and Prior Densities for AR(2) Model Parameters

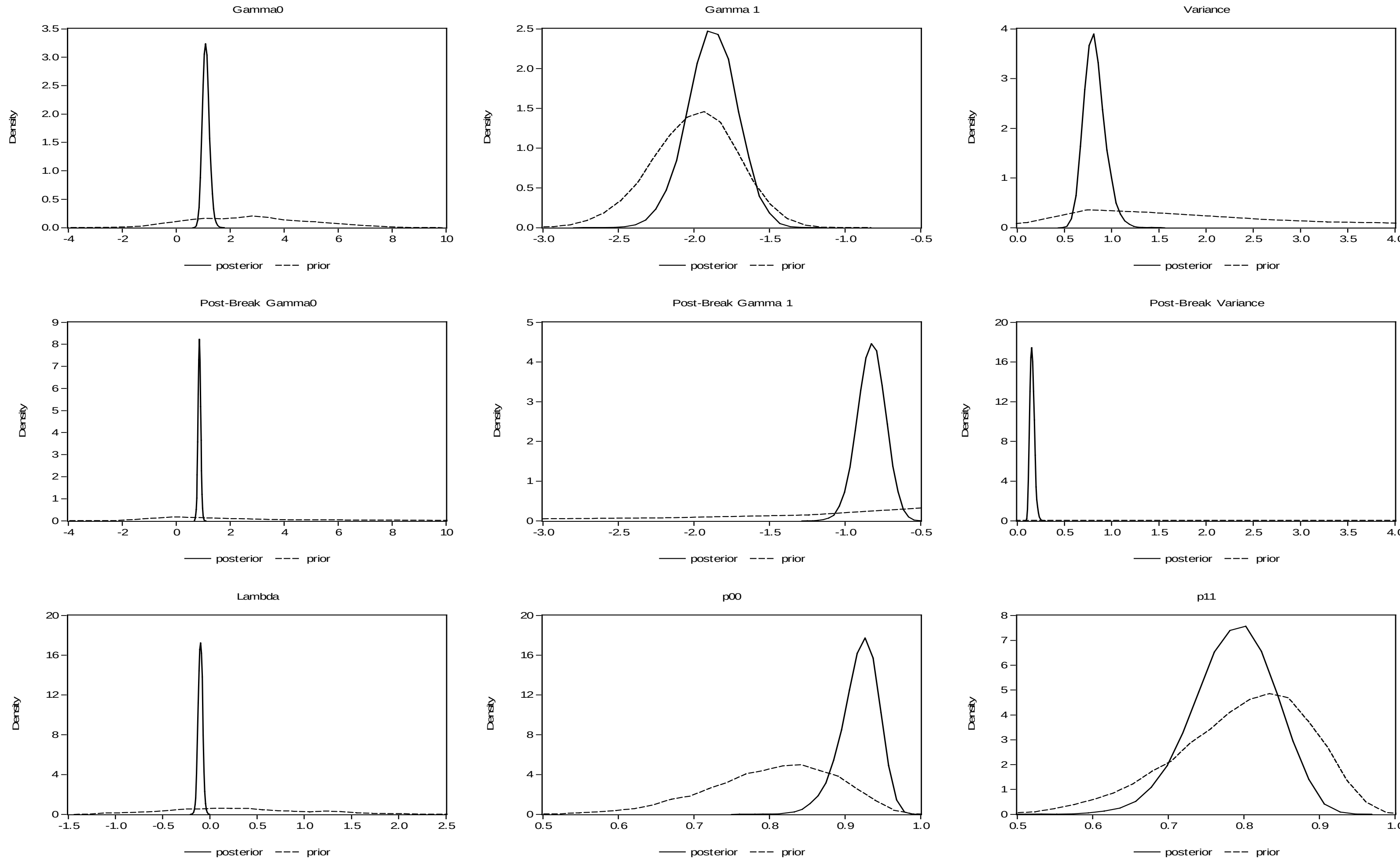

Fig. 4
Posterior and Prior Densities for Bounceback Model Parameters

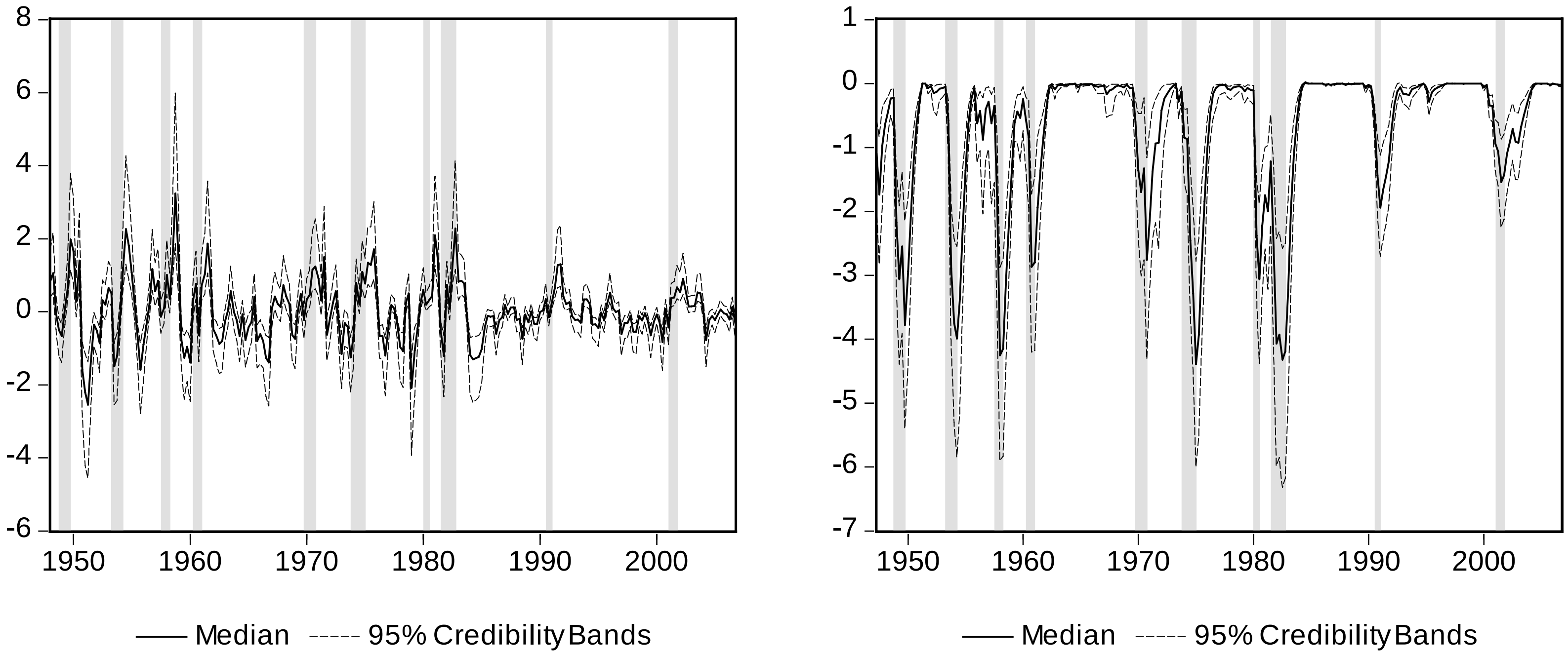

Fig. 5
Posterior Inferences for AR(2) and Bounceback Model Measures of the U.S. Business Cycle (NBER recessions shaded)

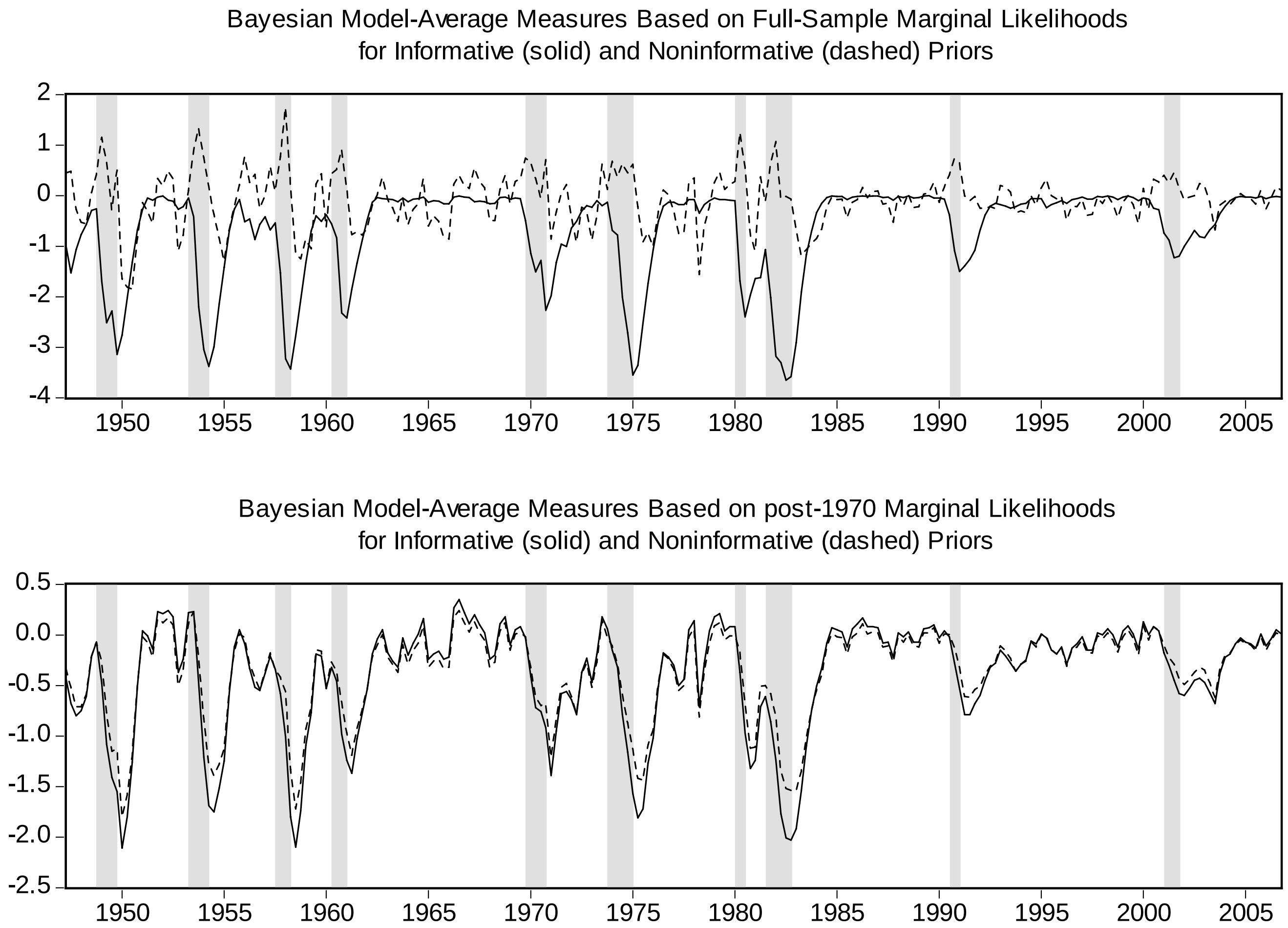

Fig. 6
Bayesian Model-Averaged Measures of the U.S. Business Cycle (NBER recessions shaded)

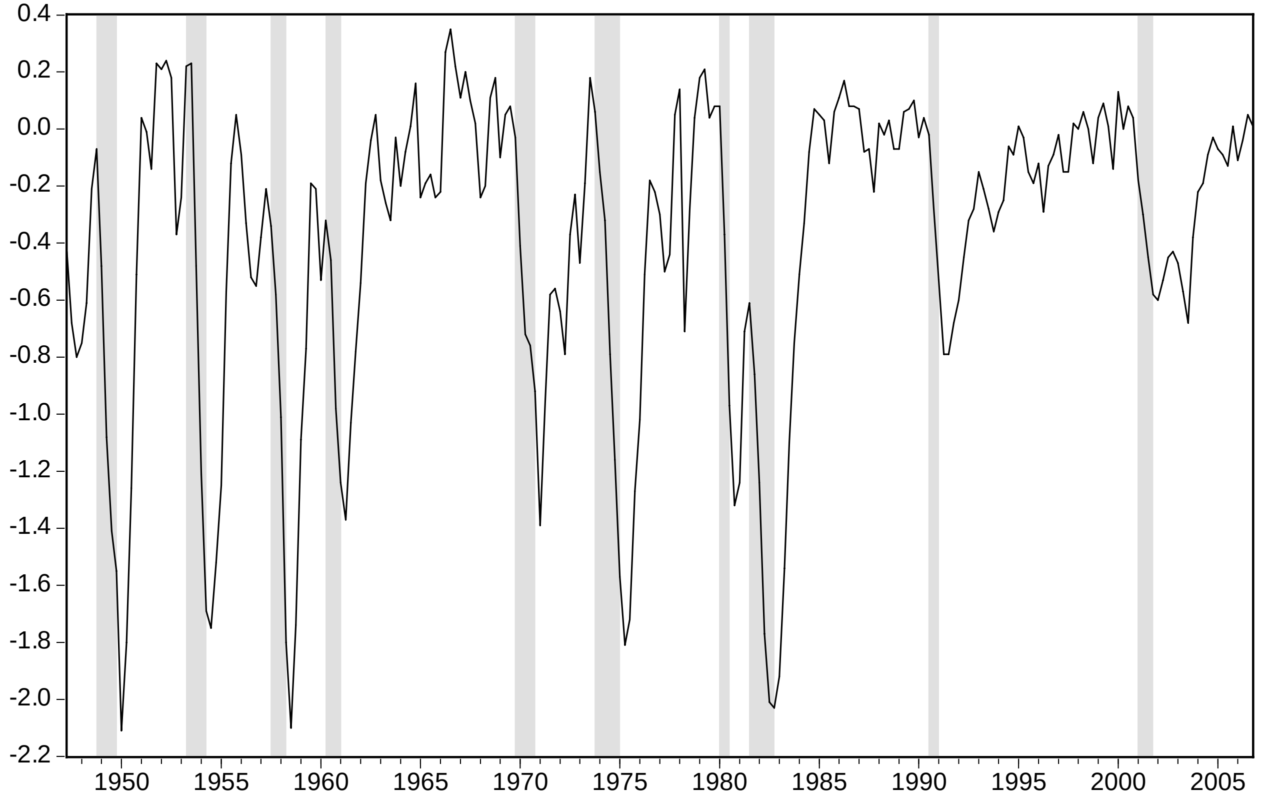

Fig. 7
The “Model-Free” Measure of the U.S. Business Cycle (NBER recessions shaded)

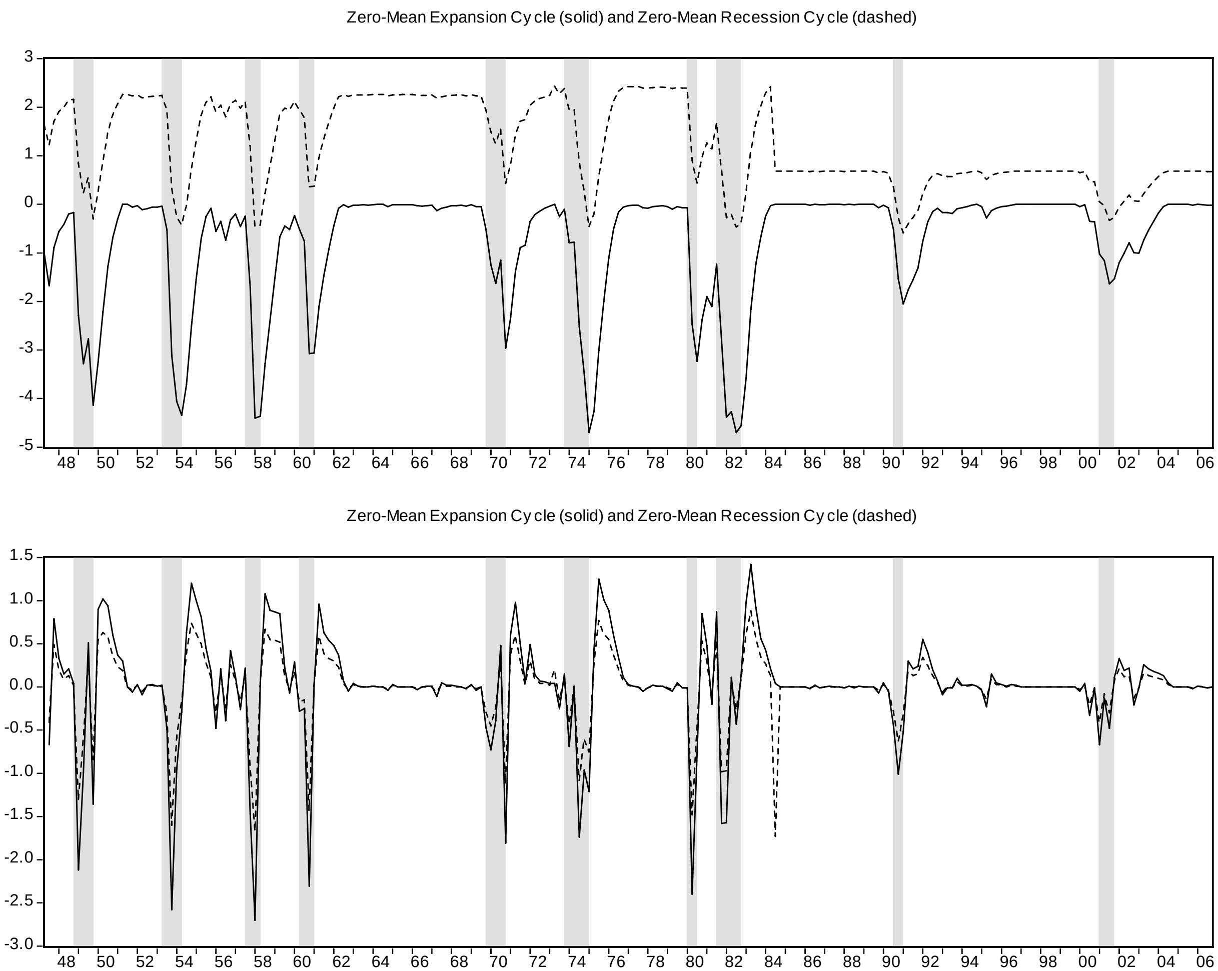

Fig. 8
Different Measures of the Level and Changes of the U.S. Business Cycle Based on the Bounceback Model (NBER recessions shaded)

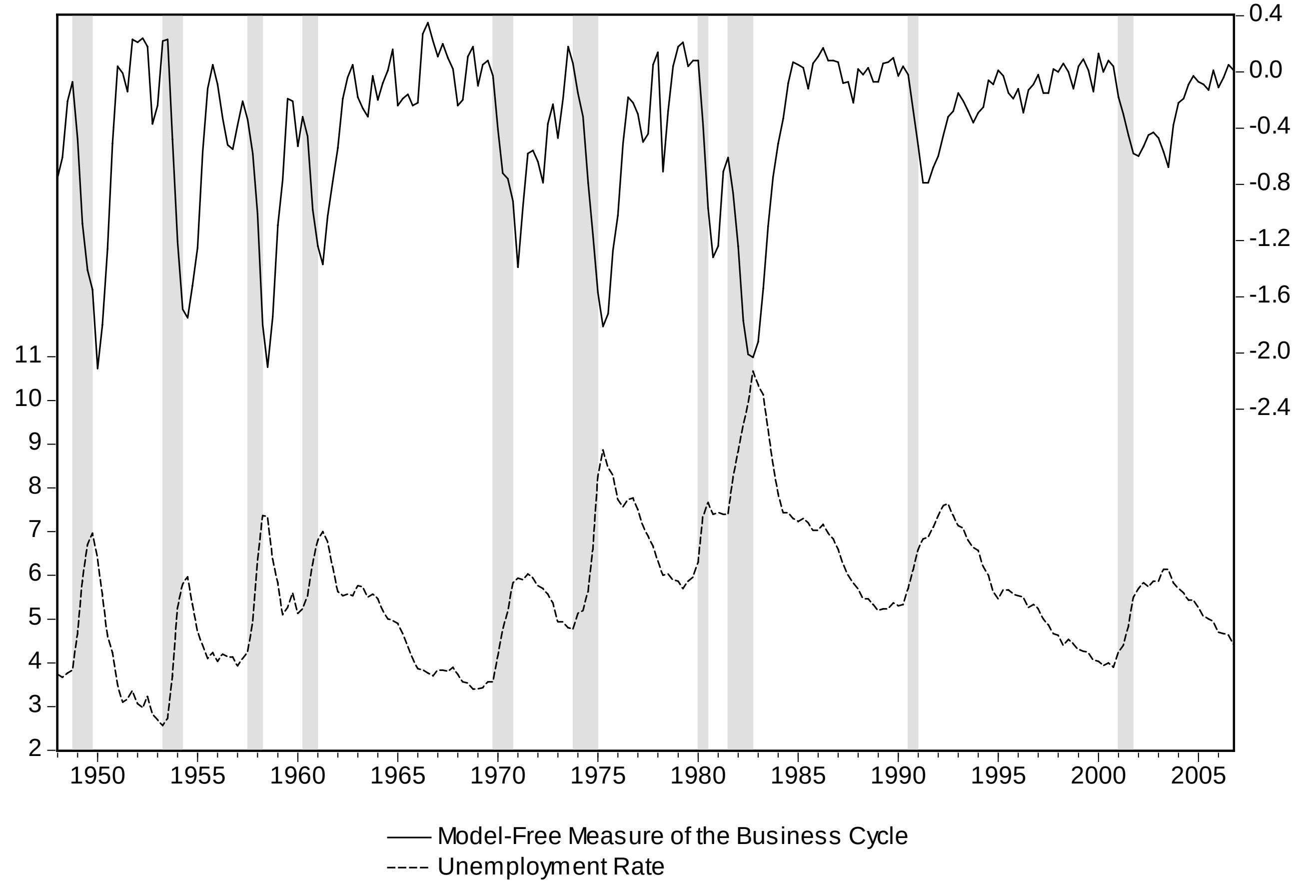

Fig. 9
The Model-Free Business Cycle and the Unemployment Rate (NBER recessions shaded)

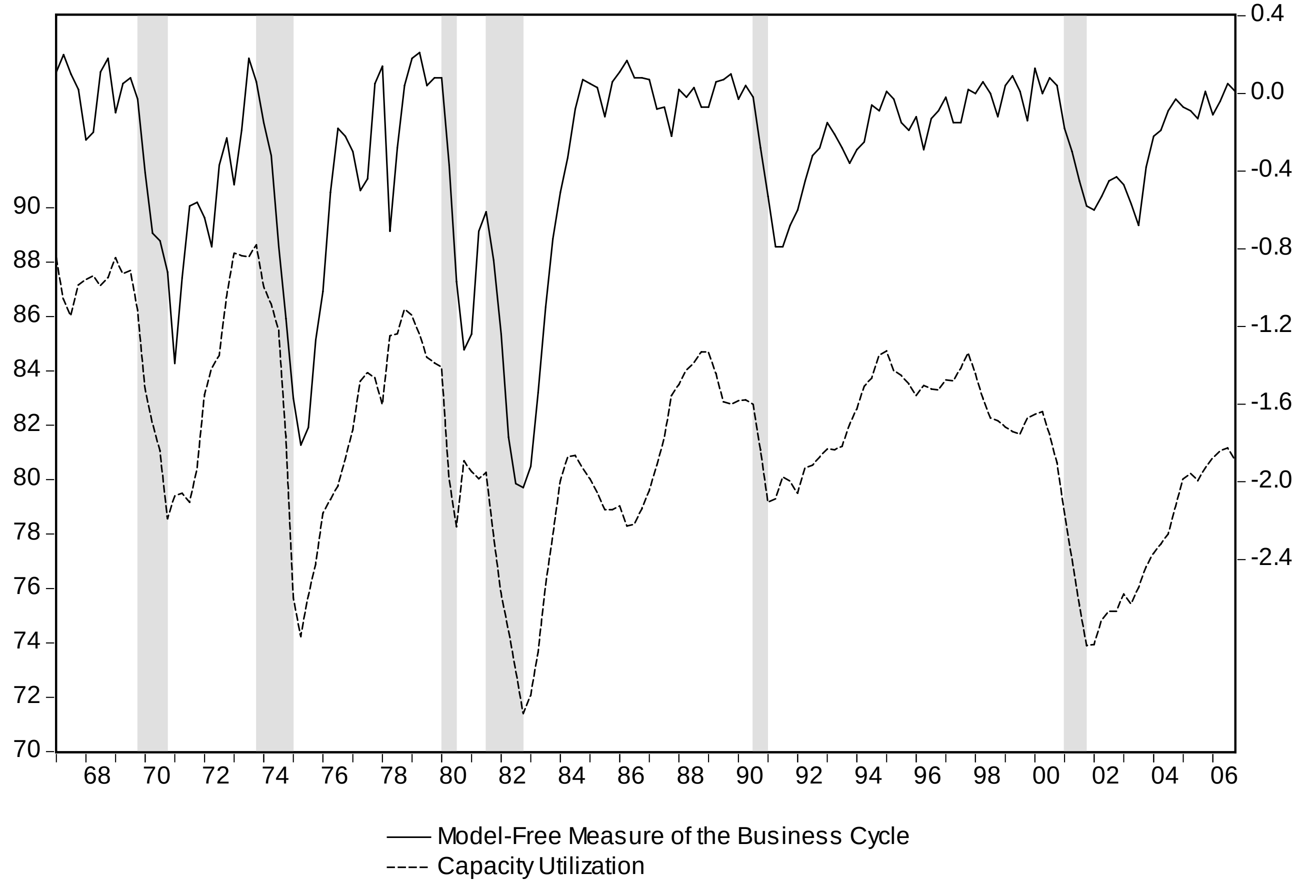

Fig. 10
The Model-Free Business Cycle and Capacity Utilization (NBER recessions shaded)

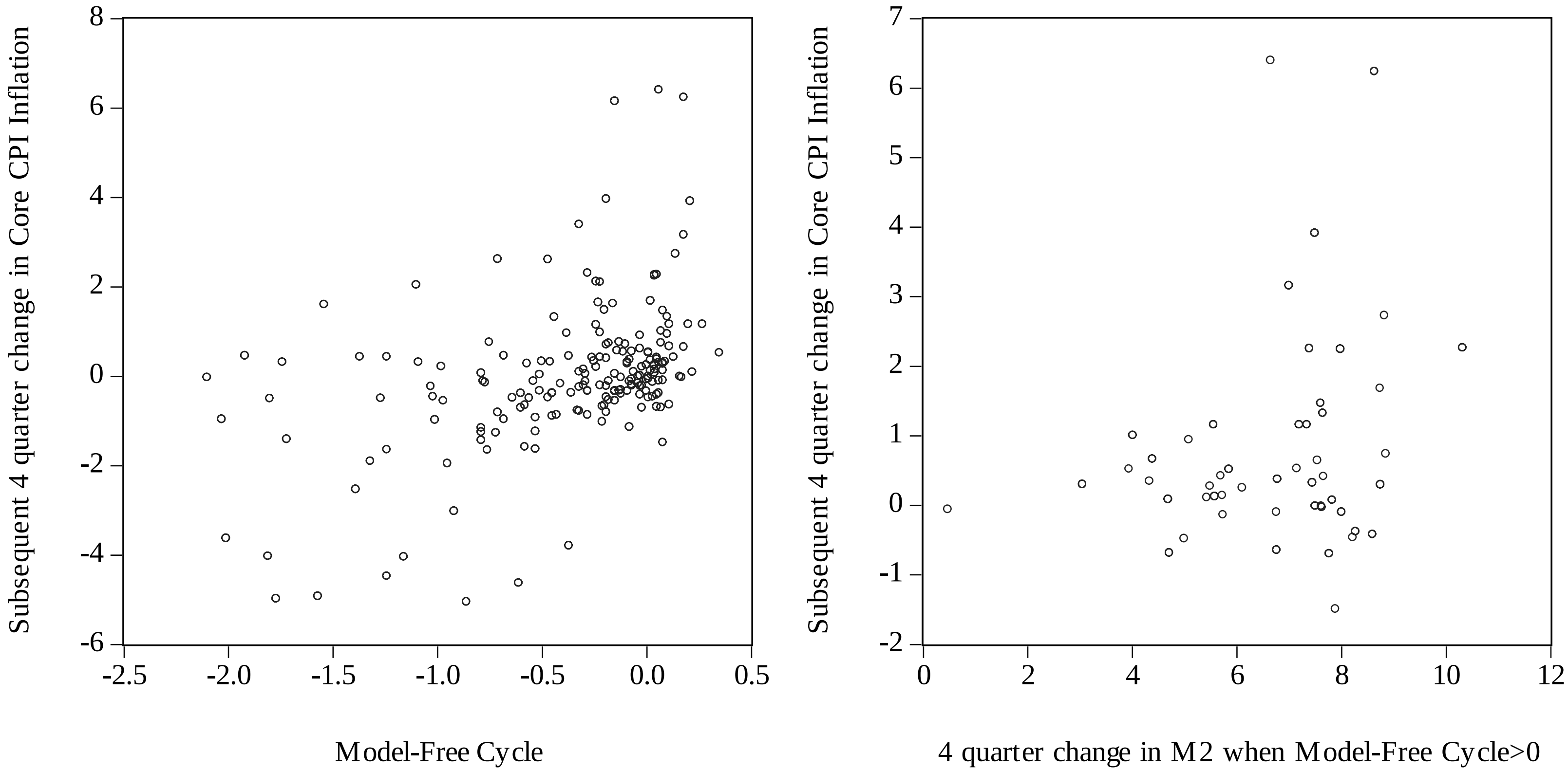

Fig. 11
Inflation Versus the Model-Free Business Cycle and Money Growth

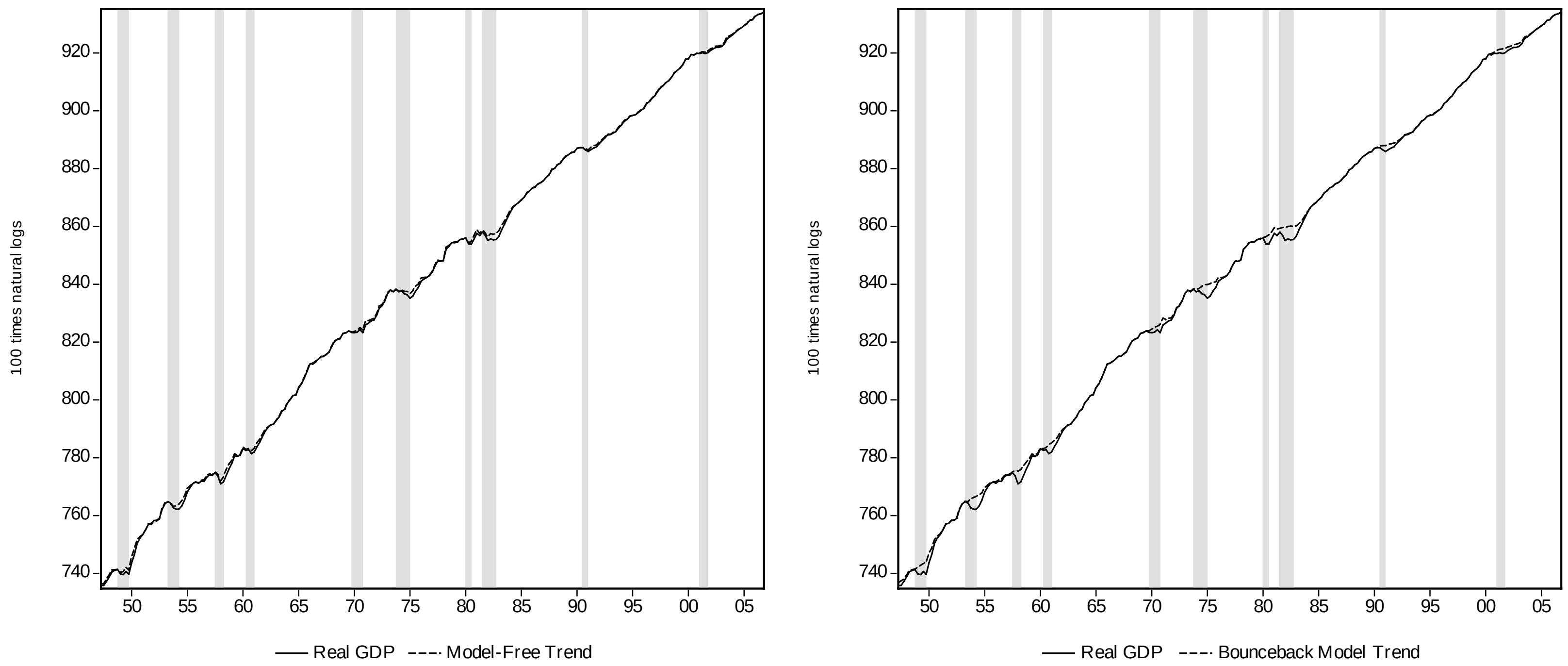

Fig. 12
Model-Free and Bounceback Model Measures of Trend (NBER recessions shaded)