



Mankiw vs. DeLong and Krugman on the CEA's Real GDP Forecasts in Early 2009: What Might a Time Series Econometrician Have Said?

David O. Cushman¹

[LINK TO ABSTRACT](#)

Introduction

At the end of February 2009, the Council of Economic Advisers (CEA) of the new Obama administration forecasted a strong rebound in the U.S. economy from the recession. The CEA predicted that, after a further dip in 2009, real GDP would recover strongly, growing at annual rates of over 4% in 2011 and 2012 and achieving cumulative growth of 15.6% by 2013 compared to 2008. The CEA based its forecast on the newly decided size of the fiscal stimulus and on the “key fact...that recessions are followed by rebounds” and “deeper recessions are typically followed by more rapid growth.”²

A few days later, Greg Mankiw expressed doubts in a blog entry (Mankiw 2009b). He suggested that the administration was “premissing its forecast on the economy being *trend stationary*.” If so, shocks have only temporary effects on real GDP. After a large negative shock and its resulting recession, real GDP would

1. Westminster College, New Wilmington, PA 16172.

2. The CEA gave its rebound discussion and forecasts for 2009 and 2010 in “Economic Projections and the Budget Outlook” (CEA 2009). The full set of administration forecasts through 2019 is in Office of Management and Budget (2009), Table S-8, issued at the same time and cited in CEA (2009).

rebound to its unaltered long-run growth path, growing at a higher-than-normal rate to get there. Mankiw quoted the abstract of a paper he coauthored with John Campbell (Campbell and Mankiw 1987a), in which they noted they were “skeptical of this implication” (that shocks do not affect real GDP’s long-run growth path) for post-war U.S. real GDP. They argued instead for the unit root hypothesis. “It contrasts starkly with the trend-stationary hypothesis,” wrote Mankiw in his blog (2009b). If real GDP contains a unit root, shocks tend to permanently change real GDP’s growth path. In fact, Campbell and Mankiw (1987a) had concluded that a 1% negative shock to real GDP would lead to a permanent reduction in the growth path of *even more than* 1%. After a recession, one would therefore expect no rebound in real GDP.

Brad DeLong, in his own blog (DeLong 2009), immediately retorted that one needed to distinguish between permanent and transitory effects.

A fall in production that does not also change the unemployment rate will in all likelihood be permanent. A fall in production that is accompanied by a big rise in the unemployment rate will in all likelihood be reversed. You have to do a bivariate analysis—to look at two variables, output and unemployment.

The 2008 decline in real GDP was, as all know, accompanied by an increase in the unemployment rate, and so there would be a rebound in real GDP, just as predicted by the CEA. “And that is certainly the way to bet,” concluded DeLong.

Paul Krugman quickly took up the argument in his blog (Krugman 2009). In an entry provocatively titled “Roots of evil (wonkish),” Krugman wrote:

I always thought the unit root thing involved a bit of deliberate obtuseness—it involved pretending that you didn’t know the difference between, say, low GDP growth due to a productivity slowdown like the one that happened from 1973 to 1995, on one side, and low GDP growth due to a severe recession. For one thing is very clear: variables that measure the use of resources, like unemployment or capacity utilization, do NOT have unit roots: when unemployment is high, it tends to fall. And together with Okun’s law, this says that yes, it is right to expect high growth in future if the economy is depressed now.

Finally, Mankiw (2009c) wrote:

Paul Krugman suggests that my skepticism about the administration’s growth forecast over the next few years is somehow “evil.” Well, Paul, if you are so confident in this forecast, would you like to place a wager

on it and take advantage of my wickedness? Team Obama says that real GDP in 2013 will be 15.6 percent above real GDP in 2008. (That number comes from compounding their predicted growth rates for these five years.) So, Paul, are you willing to wager that the economy will meet or exceed this benchmark? I am not much of a gambler, but that is a bet I would be happy to take the other side of (even as I hope to lose, for the sake of the economy).

Krugman made no response to this that I know of, but given DeLong's certainly-the-way-to-bet statement above, Mankiw's bet may have been directed to the wrong person.³

Now, this blog exchange was more interesting than most because it involved three of what may be the four most popular economics blogs (Davis et al. 2011). Moreover, at the time I was immediately sympathetic to Mankiw. I liked Mankiw based on his style of presenting ideas, on having met him, and on having had a few, brief (and pleasant) email exchanges (there have been more since then). And I am also basically a conservative, market-favoring economist, just as Mankiw seems to be. In contrast, I am not a fan of shoot-from-the-hip analysis accompanied by incomplete evidence and, yet, extreme self-confidence; in my opinion, such often appears in DeLong's and Krugman's blogs. Nor do I like their frequent name calling, condescension, and snark. Thus, I was interested in trying to prove them wrong.⁴ Yet, their point about accounting for permanent versus transitory shocks in forecasting was a good one.⁵

Therefore, I wondered what a careful, non-DeLong/Krugman forecasting analysis would indicate. Would it confirm that the CEA forecast was an obvious one, as DeLong and Krugman expressed in their quick retorts? Or would Mankiw's bet look pretty good?

Other forecasts were then available. An Office of Management and Budget (OMB) report (2009) referenced by the CEA, which Mankiw likely looked at to

3. Although Mankiw's central source of rebound skepticism certainly seemed to be the unit-root idea, his initial blog on this (Mankiw 2009b) presented two additional reasons for doubt. For the interested reader I elaborate and clarify them in the Appendix.

4. In the interest of fuller disclosure, I'll reveal that more recently, in September 2011, I received (in contrast to pleasant Mankiw emails) an unpleasant email from DeLong. It was in response to a comment I had attempted to post on his blog. He wrote me, "Shame on you for trying to confuse the issue." He also did not allow the comment to be posted. As my project was already well underway, the event was clearly not an initial motivator, but may have served to spur me on. In case the reader wants to assess whether I was "confusing the issue," I provide details in the Appendix. The episode also provides a case study of the sorts of things I don't like in DeLong's and Krugman's blogs.

5. In his "Wanna Bet" blog, Mankiw (2009c) pointed out that Campbell and he (1987b) had, in fact, investigated transitory versus permanent shocks in GDP. The paper's abstract concludes, "We find no evidence for the view that business cycle fluctuations are more quickly trend-reverting."

get the CEA growth forecast through 2013, presented not only the CEA forecasts but also the forecast of the Congressional Budget Office (CBO) from January and the Blue Chip consensus forecast from February.⁶ In contrast to the CEA forecast of 15.6% growth from 2008 to 2013, the CBO forecast was for 12.4% and the Blue Chip consensus for just 9.1%. A February revision by the CBO (Elmendorf 2009), which took account of the stimulus bills then being considered by Congress, implied slightly higher growth of 12.7% by 2013.⁷ Since Mankiw's bet was that he would lose only if growth met or exceeded 15.6%, it seems he was not being all that bold (in line with being "not much of a gambler"!), independent of his unit root story.

In any event, I had no idea (nor would any outsider, I presume) how the CBO and Blue Chip forecasts were constructed. Therefore, it was unclear to me to what extent the CBO and Blue Chip forecasts reflected the issues that Mankiw, DeLong, and Krugman deemed important.

I decided to perform my own analysis. It would be based on forecasting techniques that seemed to me to be standard in the sense of being found in many standard econometric and time series textbooks and used often in the journal literature. I would also add a few refinements from the recent journal literature. I wanted the analysis to be reasonably thorough and as unbiased as I could manage.

I had other projects to complete, however, and two years passed before I returned to the idea. By then it was beginning to appear that Mankiw would have been on the way toward winning his bet. Real GDP for 2010 (as of the February 25, 2011, Bureau of Economic Analysis (BEA) release) was a mere 0.1% higher than in 2008, not 2.0% higher as forecasted by the CEA. And at the time of final revisions to this paper (August 2012), a Mankiw win seems even more likely, with real GDP for 2011 (as of the July 27, 2012, BEA release) only 1.0% higher than in 2008 instead of the CEA's forecasted 6.0%. But such developments shed no light on the outcome of the project I imagined in 2009.

6. The Blue Chip forecasts for 2009 and 2010 in OMB (2009) are from February 2009, and the subsequent years from October 2008. The Blue Chip consensus is the average of forecasts by approximately 50 private forecasters of a number of macroeconomic variables. The monthly issues of Blue Chip Economic Indicators present annual forecasts for the next two years and the March and October issues additionally present longer forecasts. See Aspen Publishers' website ([link](#)).

7. I doubt that Mankiw, DeLong, or Krugman would have taken into account the revised CBO figures as far as growth through 2013 is concerned, because the figures were not presented in a way easy to compare with 2008's calendar-year value. The revised CBO figures were presented as revisions to fourth-quarter levels relative to the CBO's January "baseline" values. Thus, computing the calendar-year values requires some interpolations and other assumptions. Details are in the Appendix. As noted in the previous footnote, the February 2009 Blue Chip forecast for 2013 was actually from October 2008, and so it's not clear whether it accounted for the possibility of stimulus.

In spring 2011, I decided to go ahead with the project. But I realized that the forecasting needed to proceed as if it were still March 2009, because it would not be fair to utilize data, information, or techniques only more recently available. I must also admit that developments since March 2009 have surely affected my motivation to go through with the investigation, but I have striven to maintain the same goal of an unbiased econometric analysis that I had then.

I began by imagining a hypothetical time series econometrician who would apply ARIMA (autoregressive integrated moving-average) models (Box and Jenkins 1970) to first differences of real GDP, as done by Campbell and Mankiw (1987a). ARIMA models are sometimes referred to as ARIMA(p, d, q) models, where the autoregressive lag order is p , the order of differencing is d , and the moving average lag order is q . ARIMA(p, d, q) models have been very popular for forecasting and so I decided to refer to my hypothetical econometrician as PDQ (not to be confused with the infamous classical composer P. D. Q. Bach). PDQ is envisioned as male and I will sometimes refer to PDQ with masculine pronouns (he, him).⁸ The popularity of ARIMAs for forecasting is attested to by their appearance in this capacity in many well-known textbooks that PDQ would know of (such as Pindyck and Rubinfeld 1997, Enders 2004, Tsay 2005, and Diebold 2008). Graham Elliott and Allan Timmerman (2008, 23) discuss the ARIMA model's "historical domination" in forecasting. Moreover, PDQ has seen ARIMA models applied in the literature to analyze the time series properties of real GDP. Prominent examples are Campbell and Mankiw (1987a), alluded to in Mankiw's blog post, and James Morley, Charles Nelson, and Eric Zivot (2003).

But as I proceeded, I realized that ARIMA modeling did not obviously address DeLong and Krugman's point about using the unemployment rate to help make the forecast. To do so in an obvious way, I felt PDQ should do what DeLong (2009) had suggested: employ a bivariate approach. DeLong's blog post (2009) included a graph of historical real GDP growth rates plotted against earlier unemployment rates. A regression line was included. It showed high unemployment being followed by higher than normal economic growth. But DeLong gave no other statistical results for the regression. Mankiw (2009c) questioned the statistical significance of the line in what he called a "cloud of points." Thus, I wanted PDQ to follow the bivariate suggestion, but with something more credible and *a lot* more thorough.

The obvious answer to me was to add a bivariate VAR (vector autoregression) approach to the project. Like ARIMAs, VARs are popular for forecasting. For example, in his well-known textbook, William Greene (2003, 587) writes that for "forecasting macroeconomic activity...researchers have found that

8. PDQ can also devote full time to the project, and can thus finish it much faster than I have been able to.

simple, small-scale VARs without a possibly flawed theoretical foundation have proved as good or better than large-scale structural equation systems.” Similarly, in their abstract Todd Clark and Michael McCracken (2006) write, “Small-scale VARs are widely used in macroeconomics for forecasting U.S. output, prices, and interest rates.” The paper goes on to reference empirical papers that have done so. Forecasting with VARs is discussed in a number of widely used texts (e.g., Enders 2004, Lütkepohl 2005, Stock and Watson 2007, and Diebold 2008).

PDQ’s main points and findings

I am now going to summarize PDQ’s key points, procedures, and forecasts. To document PDQ’s attempt to be thorough, I give a more detailed, blow-by-blow account, with graphs of the forecasts, in the subsequent sections of the paper. Still more details are found in the occasionally referenced Appendix. Here’s the summary:

1. Real GDP can have a unit root and there can nevertheless be rebounds without necessarily implying a trend-stationary process, contrary to what some might infer from Mankiw’s blog entries.
2. Though univariate, ARIMAs allow for both permanent and transitory effects. In contrast to DeLong’s (2009) assertion, you do not necessarily need a bivariate approach. However, *if* unemployment rate fluctuations capture most of the transitory effects, a bivariate output-unemployment VAR may be better.
3. Because underlying shocks are usually unobserved and output and unemployment are simultaneously determined, neither an ARIMA nor a VAR model can fully identify transitory and permanent shocks. This counters Krugman’s (2009) claim that it is easy to know whether the source of a GDP slowdown is transitory or permanent.
4. Both the ARIMA and the VAR approaches require lag order choices. But rather than pick one set of lag orders for each estimation approach, PDQ combines the forecasts of many lag order models using a recent approach called model averaging.⁹ PDQ tries both AIC (Akaike) and BIC (Schwarz-Bayesian) model weights in the averaging.
5. PDQ conducts structural stability tests. They suggest that forecast model estimation should start in 1986:3, rather than earlier, to lessen possible misspecification bias in the forecasts.

9. The Blue Chip consensus is a form of model averaging; it is private forecaster averaging.

6. PDQ computes model-averaged forecasts from the ARIMAs and VARs, reported later in the paper. He goes on to compute overall AIC and BIC average forecasts for his ARIMA and VAR forecasts. The resulting forecasted 2008-2013 growth rates are:
 - ARIMA/VAR (AIC): 13.8%
 - ARIMA/VAR (BIC): 11.5%
7. The VAR model does not dominate the weights, contradicting DeLong's (2009) assertion that "you have to...look at two variables, output and unemployment."
8. PDQ's overall forecasts for 2008-2013 growth straddle the CBO's 12.4% and revised 12.7% forecasts. Like the January CBO and Blue Chip consensus forecasts, they are well under 15.6%. Therefore, PDQ thinks *Mankiw would probably win his bet*.
9. PDQ also calculates confidence bands for his forecasts, unlike the CEA, CBO, or Blue Chip consensus.¹⁰ Plus-minus one standard deviation bands (containing roughly 68% of the probability) for PDQ's overall forecasts for 2013 are:
 - ARIMA/VAR (AIC): 10.6% to 17.0%
 - ARIMA/VAR (BIC): 7.8% to 15.3%
10. Using his overall, ARIMA/VAR model-averaged forecast standard errors, PDQ computes Mankiw's probability of losing. It is only 14% (BIC weights) to 28% (AIC weights). Accordingly, the extreme confidence exuded by DeLong and Krugman in the CEA forecast is not warranted.
11. The CEA forecast is very similar to several variations of a trend-stationary forecast, as Mankiw speculated.
12. Does PDQ forecast any rebound at all? With respect to the eventual long-run equilibrium net of trend, none of PDQ's model-averaged forecasts indicate a rebound from 2008's annual figure, and they indicate at best a trivial one from 2008:4.
13. Things that PDQ could not know: Through 2011, the Blue Chip forecast of real GDP is most accurate, followed closely by the January 2009 CBO forecast. But given the most recent value of real GDP, for 2012:2, the best forecasts for the next few years will likely turn out to be the Blue Chip consensus and PDQ's ARIMA forecasts. However,

10. In the case of the Blue Chip consensus, one could examine the range of individual forecasts, but, while useful, this would not have any known relation to probabilities as does a confidence interval. The revised, February CBO forecasts consist of "high" and "low" values, but there is no explanation except that the range "encompasses a majority of economists' views." The 12.4% and 12.7% values are midpoints.

the CBO's forecasted long-run growth rate of only 2.2% by 2018 is already starting to look plausible.

Foundations for two standard forecasting models

The basic idea of the forecasts to be computed by PDQ is to extrapolate from past movements of key variables to predict future movements of real GDP. The details of government and private response to downturns are not modeled. Instead, for example, if governments have successfully responded to recessions in the past, then low GDP generally leads to stronger recoveries than otherwise, and the model will forecast this when GDP is low.

The ARIMA and VAR that PDQ will use can be derived from a basic state-space model:

$$y_t = \mu_{t-1} + \varepsilon_t ; \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \quad (1),$$

$$\mu_t = \mu_{t-1} + a + \eta_t ; \quad \eta_t \sim NID(0, \sigma_\eta^2) \quad (2).$$

The log of real GDP, y , is determined by a permanent component (μ) and a transitory component (ε). The permanent component is determined by a constant trend (a) and permanent shocks (η). Permanent shocks to capacity or labor force participation (shocks to the long-run growth path) are given by η , and temporary shocks to capacity utilization or employment by ε . The shocks probably cannot be directly observed. But note that the overall shock to y_t consists of a combination of permanent and temporary shocks, and the temporary ones are by definition reversed. Thus, observed y_{t+1} will sometimes show “rebounds,” the frequency and size of which will depend on the relative sizes of the temporary and permanent shocks and their correlation. Regardless, y_t has a unit root. Permanent shocks occur every time period.

Ruey Tsay (2005) and Rob Hyndman et al. (2008) show how the equations (1) and (2) have an ARIMA(0,1,1) as a reduced form:

$$\Delta y_t = a + v_t - \theta_1 v_{t-1} \quad (3)$$

The value of θ_1 has a relationship to the unobserved shocks in (1) and (2) that depends on certain assumptions. One common assumption (e.g., found in Tsay 2005) is that the permanent and transitory shocks are uncorrelated. In this case, θ_1 is related (nonlinearly) to the signal-to-noise ratio, the relative variances of the

permanent and transitory shocks. The lower is the signal-to-noise ratio, the higher is θ_1 . Another common assumption (promoted by Hyndman et al. 2008) is that the shocks are perfectly correlated: $\eta_t = d\varepsilon_t$. In this case, $\theta_1 = 1 - d$ and $v_t = \varepsilon_t$. In both models, the higher is θ_1 , the more important are transitory shocks relative to permanent ones. Various serial correlation processes for the shocks, and other extensions, lead to more general ARIMA(p,d,q) models:

$$\Delta^d y_t = \sum_{k=1}^p \varphi_k \Delta^d y_{t-k} + a + v_t - \sum_{k=1}^q \theta_k v_{t-k} \tag{4}$$

where p and q are the autoregressive and moving average lag orders, and d is the order of differencing to achieve stationarity.

The discussion above indicates that, despite utilizing a univariate model, ARIMA forecasting takes into account both permanent and transitory effects and thus addresses, to a certain extent, DeLong and Krugman’s observations that these effects need to be distinguished in forecasting. The distinction (or identification) in the ARIMA is, however, not perfect. The ARIMA forecasting approach cannot identify what portion of any observed shock is permanent or temporary (unless the Hyndman et al. (2008) assumption of perfect correlation is correct). Instead, ARIMA estimation supposes that a constant fraction of a recently observed shock will be reversed. The fraction is based on the estimated average tendency of observed shocks to be transitory. In the basic ARIMA(0,1,1) case, this tendency is related to the θ_1 estimate, and in more complicated ARIMAs to all the φ ’s and θ ’s.

PDQ thus pursues the VAR approach to more directly address DeLong and Krugman’s criticism. If one is willing to associate unemployment with one or both shocks in the state-space model, then equations (1) and (2) can be transformed into a VAR that can be used for forecasting.

PDQ supposes that, as suggested by DeLong (2009), the transitory shock is related to unemployment: $\varepsilon_t = -b_1(un_t - \bar{un}) + \tau_t$, where un is the unemployment rate, \bar{un} its mean, and τ an independent transitory effect not captured by the unemployment rate. He substitutes this into equation (1). PDQ also allows for the possibility that unemployment fluctuations could be the source of some of the permanent effects: $\eta_{t-1} = -c_1(un_{t-1} - \bar{un}) + \psi_{t-1}$. For example, the skills of the unemployed may deteriorate so that they are less likely to be rehired (Pissarides 1992).

After the substitutions, PDQ transforms equations (1) and (2) into a first difference equation:

$$\Delta y_t = a_1 - b_1 u_n t + (b_1 - c_1) u_n t_{-1} + \tau_t - \tau_{t-1} + \psi_{t-1} \quad (5)$$

where $a_1 = a + c_1 \bar{m}$. Finally, PDQ supposes that fluctuations in u_n are related to Δy_t and an underlying transitory shock ω :

$$u_n t = a_2 - b_2 \Delta y_t + \omega_t \quad (6)$$

PDQ notes that Δy and u_n are now simultaneously determined. Equation (5) says that more employment increases the supply of output and equation (6) says that more output increases the demand for labor. Equations (5) and (6) constitute a first-order VAR but cannot be used for forecasting because of the presence of unlagged variables. To get rid of them, PDQ solves the system for the reduced form VAR:

$$\Delta y_t = [(a_1 - b_1 a_2) + (b_1 - c_1) u_n t_{-1} + \tau_t - \tau_{t-1} + \psi_{t-1} - b_1 \omega_t] / (1 - b_1 b_2) \quad (7),$$

$$u_n t = [(-b_2 a_1 + a_2) - b_2 (b_1 - c_1) u_n t_{-1} - b_2 (\tau_t - \tau_{t-1} + \psi_{t-1}) + \omega_t] / (1 - b_1 b_2) \quad (8).$$

Neither of the reduced form VAR equations has any explicit lags of Δy . Moreover, the portion of the transitory error ε that is not captured by u_n , which is τ , leads to the composite error terms being equivalent to a moving average process, just as the transitory error in equation (1) leads to a moving average process in (3). In estimation of the VAR, the moving average can be approximated with a sufficient number of lags of Δy and u_n . But if u_n captures most of the transitory effects, the moving average part will be small and additional lag terms trivial in importance. However, serial correlation in ψ and ω will also introduce additional lag terms.¹¹ The simultaneous determination of Δy and u_n means that it will not be possible to identify or separate out the effects of the underlying transitory and permanent shocks. Contrary to the implications of DeLong and Krugman, observed values of the unemployment rate do not necessarily measure transitory shocks.

PDQ believes that the relative merits of the ARIMA versus the VAR can be summarized as follows. If the unemployment rate in the VAR captures a high

11. The moving average aspect of equations (7) and (8) could be addressed by estimating the system as a VARMA (vector autoregressive moving average) model, just as the first-difference version of equations (1) and (2) can be estimated by an ARMA. However, PDQ has read of the difficulties in the identification and estimation of VARMA's discussed in Lütkepohl's (2005) time series text, where four chapters are devoted to the model. Perhaps because of the difficulties, the JMulti computer program ([link](#)), which is directly based on Lütkepohl's work, does not include VARMA estimation.

portion of the transitory shocks assumed unobservable in the state-space/ARIMA approach, then the VAR forecasts should be better because the VAR model will have less noise. For example, the transitory shock ω_t will be captured to a significant extent in the un_t value that will be used to forecast y_{t+1} using equation (7). But if *non-unemployment* transitory shocks are important, then the VAR will have a significant moving average error and contain more noise than the ARIMA unless its lag order is rather long (which then reduces statistical efficiency). In contrast, the ARIMA does not need long lag orders to handle an unspecified transitory shock.

Data and preliminary analysis

Data set

PDQ realizes that he should use the same data set as was available to the CEA. The CEA (2009) stated that “[t]he Administration’s economic assumptions were largely completed in early January and finalized on February 3rd.” Therefore, the CEA presumably had access to the Bureau of Economic Analysis (BEA) release of January 30, 2009, containing data through 2008:4.¹² The bloggers, meanwhile, could have looked at the February 27th release, but they would likely agree with PDQ that assessment of the CEA forecast should be based on the data available to the CEA at the time and not subsequent revisions. For unemployment the CEA would have had access to the January 9th release from the Bureau of Labor Statistics (BLS), and so unemployment rates reported as of this date are what PDQ uses.¹³

PDQ decides he needs to address three issues before actually computing forecasts. (1) What specific ARIMA and VAR lag specifications should he use? (2) His estimation period will end in 2008:4, but when should it begin? The available quarterly data starts soon after World War II, but the literature contains well-known evidence of structural changes since then, which, if included in the forecasting model estimation period, could bias the forecasts. (3) The ARIMA,

12. The archive of releases is available at the BEA’s website ([link](#)). The specific real GDP series is billions of chained (2000) dollars, seasonally adjusted at annual rates. The values are transformed to logs for estimation and forecasting.

13. The unemployment rate is the seasonally adjusted rate for age 16 and over, series LNS14000000. The current full data set is found at the BLS’s website ([link](#)), but it reflects revisions unavailable to PDQ. However, each January the BLS generates revised values for the most recent five years. I thank Karen Kosanovich of the BLS for providing me the data released in January 2009. After substituting the January 2009 data for the more recently released data for 2004–2008, we have the complete unemployment series as of January 2009.

equation (4), and the VAR, equations (7) and (8), contain Δy and u and thus assume that y has a unit root and is first-difference mean stationary, and that u is mean stationary. Is this reasonable?

Lag order selection

Many methods exist for choosing a forecasting model when many lag orders and variables are possible. For univariate models, Tsay (2005) and Hyndman et al. (2008) emphasize information criteria such as the AIC and BIC criteria. Hyndman et al. (2008) seem to slightly favor the AIC. Diebold (2008) favors the BIC. For forecasting with VARs, Helmut Lütkepohl (2005) also suggests using information criteria. He seems to have no clear favorite. PDQ decides to use both the AIC and BIC.

PDQ is, however, impressed with the relatively recent approach of using not just the forecasts of the top model from a given criteria, but weighted averages of the forecasts from many models. The weights or probabilities are computed from the models' AIC or BIC values (see Koop and Potter 2003, Hansen 2007, and Wright 2008). Bruce Hansen (2007) favors AIC over BIC based weights.¹⁴ The Appendix gives the formulas.

PDQ uses the same ARIMAs as in Campbell and Mankiw (1987a), 16 models with lag orders p and q of 0 to 3. In the weighting, PDQ assumes equal priors, a typical approach. For the VAR models, PDQ increases the maximum lag order to 4. Because equations (7) and (8) clearly include the possibility of different lag orders of variables within *and* between equations, PDQ allows Δy and u in each equation to each have different lag orders, ranging from 0 to 4 and thus yielding a total of 625 VAR models. Once again, PDQ assumes equal priors.¹⁵

When to start the estimation period

The full quarterly U.S. real GDP set from the BEA begins in 1947:1, and many papers analyzing the time-series properties of U.S. real GDP have used this starting date. The BLS quarterly unemployment data starts at almost the same point, 1948:1. Therefore, PDQ defines his full data set as starting in 1948:1, with first differences starting in 1948:2. However, PDQ recollects the famous paper by

14. However, Hansen (2007) is specifically promoting the use of the less well-known Mallows criteria over either the AIC or BIC. But this is in a single equation environment. The Mallows approach has not been extended to VARs as far as I know.

15. In contrast to the assumption of equal priors, some lag orders seem less plausible than others to PDQ, but he suspects that may be because he is already familiar with the data. The equal-prior assumption avoids possible bias from a data-based prior.

Pierre Perron (1989) about the effect of the 1970s oil price shock on GDP growth, and he has recently read the working-paper version of Perron and Tatsuma Wada (2009), which updates the argument that there was a change in real GDP's trend growth rate at the beginning of 1973. PDQ also recalls the "Great Moderation," the apparently increased stability of the U.S. economy starting in the early to mid 1980s (e.g., Kim and Nelson 1999, Stock and Watson 2002). PDQ is concerned that the implied heteroskedasticity in real GDP will bias forecast standard errors, and he wonders if there was also change in the growth rate or in the short-run dynamic parameters at that point.

PDQ examines the stability issue in two ways. First, he applies a breakpoint test to all the ARIMA models and to the 15 highest AIC-weighted and 10 highest BIC-weighted VAR models. (The test, which has to be bootstrapped, is too time-intensive to apply to all 625 VARs; the included VARs cover 74% of the AIC weight and 96% of the BIC weight.) Second, PDQ examines some key parameter estimates for a range of different estimation periods. If the estimates change a lot, forecasts based on the longer estimation periods will likely be unreliable.¹⁶

The breakpoint test is adapted from a quasi-likelihood ratio test discussed by James Stock and Mark Watson (2007, 567-570). The date of the breakpoint is assumed unknown, and the test examines all possibilities within the middle 70% of the dates. If the test rejects homogeneity, it also provides an estimate of the breakpoint date. The key parameter estimates that PDQ examines are the trend rate of real GDP growth and the infinite-horizon impulse responses of real GDP to various shocks. These are not individual parameter estimates but functions of individual parameter estimates that are of particular relevance for forecasting. For example, in the ARIMA model, trend growth is $a / (1 - \sum \phi_i)$ and the infinite-horizon impulse response to a shock is $(1 - \sum \theta_i) / (1 - \sum \phi_i)$. For both the ARIMAs and the VARs, PDQ computes the trend and impulse response estimates in a recursive manner. He starts with a short estimation period, 1994:1–2008:4, and then moves the start date backwards quarter by quarter until he reaches the longest period of 1948:2–2008:4. If the models are stable, the estimated trends and impulse responses should not change too much in terms of economic importance.

Details of the procedures and outcomes are given in the Appendix. Here I just summarize PDQ's findings and conclusion. For the ARIMAs, the breakpoint tests do not reject the no-break null hypothesis, but trend growth gets substantially higher as the estimation starting point goes back in time, particularly as it moves back into the 1960s. In this way does Perron's trend growth change manifest itself.

16. All procedures in the paper, except one noted in the Appendix, were coded and computed in TSP 5.1. The Appendix contains a link to a web page with the data, TSP files, and Excel files used to get the results in this paper.

In addition, the impulse response values show a great deal of instability in economic magnitude as the estimation start date moves backwards through the early 1980s. Turning to the VARs, the no-breakpoint hypothesis is strongly rejected, with two break dates emerging as most likely: 1973:2 and 1986:3. Trend growth shows the same pattern as in the ARIMAs, and the infinite-horizon impulse responses of y to Δy shocks and to u shocks change suddenly and substantially as the estimation starting date moves to points before 1985.¹⁷ Based on the various results, PDQ concludes that the estimation period for his forecasting equations should start in 1986:3. It's a shame that earlier data containing additional recessions cannot be used, but forecasts from longer data sets are likely to be biased.

Real GDP and the unemployment rate: Unit root or stationary processes?

The literature seems to lean in favor of a unit root in real GDP (e.g., Murray and Nelson 2000, Shelley and Wallace 2011). Regarding the unemployment rate, on first pass it seems that it would be stationary, being bounded by zero and 100 percent, and in theory tending to return to the natural rate. But the natural rate may change and the bounds are not very restrictive. Moreover, the empirical evidence is not clear. For example, Mehmet Caner and Bruce Hansen (2001) strongly reject a unit root in the unemployment rate but conclude there are two “regimes,” which consist of two different autoregressive processes that alternate irregularly over time. This is discussed in more detail in the Appendix, as are PDQ's own unit root tests on real GDP and the unemployment rate. PDQ's decision is to stay with Δy and u in the ARIMAs and VARs.

Forecasts

Preliminaries: Linear trends, other forecasts, and post-2008 real GDP values

Mankiw thought that the CEA forecasts looked as if the CEA expected a return to a deterministic linear trend. PDQ therefore fits a linear trend over

17. Because the simultaneity in his structural VAR equations (5) and (6) indicates that structural shocks will not be identifiable, PDQ uses the generalized impulse response procedure of Pesaran and Shin (1998). The procedure applies to the two reduced form VAR equations (with constants omitted) two pairs of reduced form shocks that reflect the correlation in the estimated residuals. See the Appendix.

the 1986:3–2008:4 period to see how the various forecasts relate to it.¹⁸ To help interpret relative movements in the upcoming graphs, Figures 1 to 3, PDQ normalizes all series using the extrapolation of the 1986-2008 trend. Therefore, in the graphs all values are log differences from the extrapolated 1986-2008 trend line. For a second benchmark, PDQ includes a trend of the same slope starting from the average real GDP value for 2007, the year of the most recent peak in the business cycle. In addition to PDQ’s forecasts, the graphs include four other forecasts, which are annual: the CEA forecast, the January and February CBO forecasts, and the February Blue Chip forecast.

The graphs also show the actual post-2008 real GDP values from the July 2012 BEA release. PDQ can’t know these, of course. They are there for us to see how good the forecasts have turned out to be so far. However, acquiring appropriate values for the comparison is not as straightforward as one might imagine. Later in 2009 and in the two following years, the BEA substantially revised the real GDP series. The revisions changed the base year, but much more important, they substantially increased the reported decline of real GDP in 2008 compared with 2007. How one links the currently reported data with the data used to make the forecasts significantly affects how the actual post-2008 values compare with the forecasted values. I use two approaches.

In the first approach, the quarterly log differences starting in 2009:1 of data reported in February 2012 are added one by one to the log of 2008:4 real GDP to create the post-estimation-period realized quarterly log values. This approach does not penalize the forecasters for not knowing that 2008’s quarterly real GDP growth rates would all subsequently be revised downward. But this approach does not correctly depict the negative *annual* growth rate from 2008 to 2009. My second approach does do so by applying the annual 2008-2009 growth rates reported in 2012 to the annual real GDP log value for 2008. The second approach generates annual values. These are the values that probably would be used by the bloggers to settle the bet, had it been accepted.

PDQ’s forecasts

PDQ is ready to compute his own forecasts of real GDP. The estimation period is 1986:3–2008:4 and the forecasts are dynamic. The AIC- and BIC-weighted ARIMA forecasts are in Figure 1. The AIC- and BIC-weighted VAR forecasts are in Figure 2.¹⁹ Points labeled “AIC trend value” and “BIC trend value”

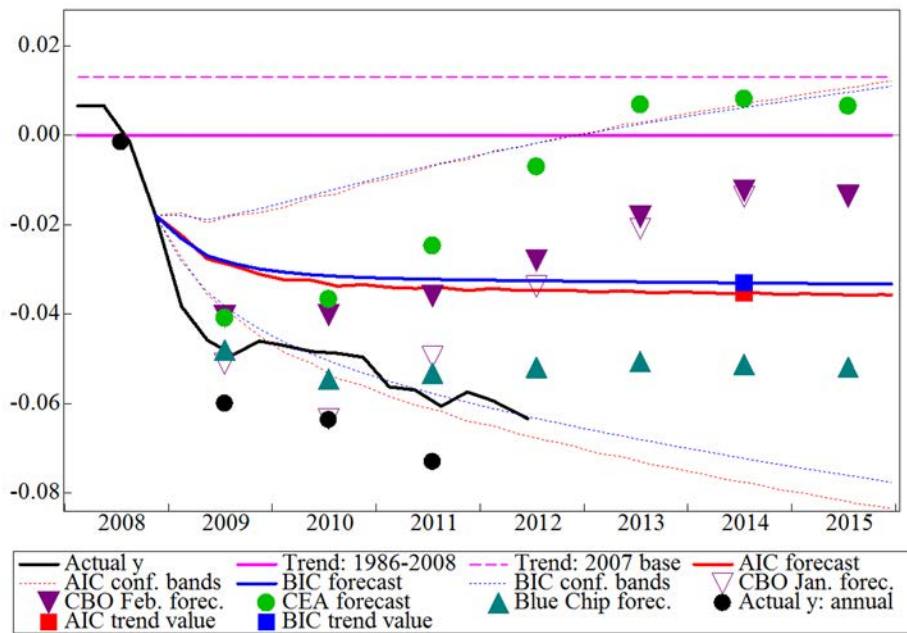
18. Specifically, he estimates an AR(3) model (as chosen by the AIC and BIC) with a constant and a linear trend.

19. A very small number of the VAR models were non-stationary (explosive) and were not used.

for the year 2014 are also plotted; these will be discussed later. The graphs also include AIC- and BIC-weighted plus-minus one standard error confidence bands for PDQ’s quarterly forecasts.²⁰ For more precise assessments, Table 1 gives the numerical values for some of the points in the graphs.

PDQ first notes that by 2013 the CEA forecast exceeds the extrapolation of his 1986-2008 linear trend, and it comes close in 2014 to the higher trend line extrapolated from 2007. This supports Mankiw’s statement that the CEA forecast was, in effect, assuming a return to a deterministic linear trend.

Figure 1. ARIMA forecasts, other forecasts, and actual values

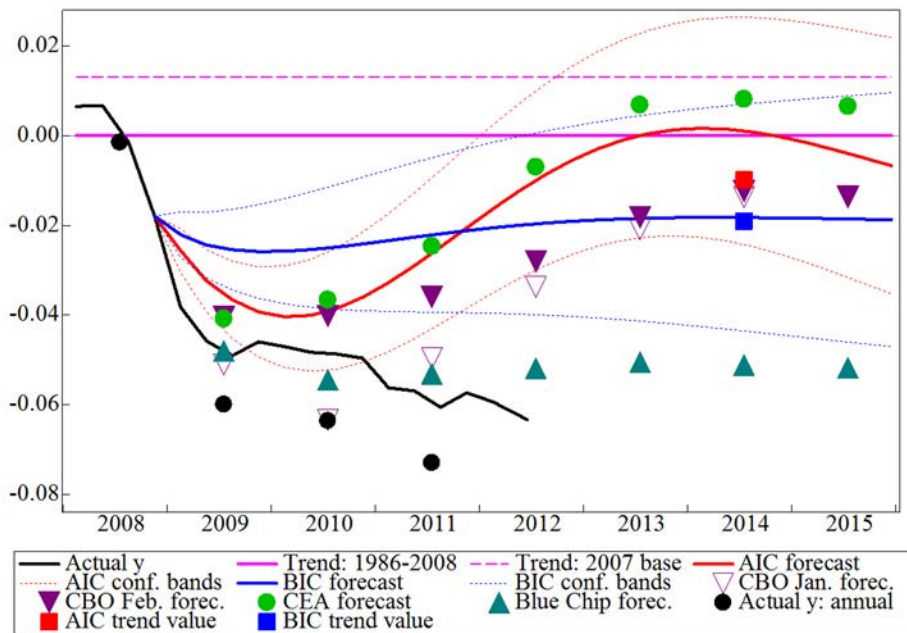


PDQ considers whether he thinks Mankiw would win his bet. None of PDQ’s weighted ARIMA or VAR forecasts meet or exceed the CEA forecast for 2013. In fact, although this isn’t shown in the graph, not a single one of PDQ’s individual ARIMA or VAR forecasts meets or exceeds the CEA forecast. However, the width of the weighted confidence bands suggests to PDQ that a Mankiw victory is not certain. Using the forecast standard error values and assuming normality

20. BIC averaging of forecast standard errors is discussed by Koop and Potter (2003). The accuracy of the confidence bands and the ability to make probability statements using them depends on the estimates being based on serially independent, homoskedastic, and normal residuals. PDQ conducts tests of these assumptions. PDQ finds only a few problems, which he deems of relatively little importance. Details are in the Appendix.

(see the Appendix), the probability that Mankiw loses the bet is 13% for the two ARIMAs and the VAR with BIC weighting and 37% for the VAR with AIC weighting.

Figure 2. VAR forecasts, other forecasts, and actual values



The VAR forecasts, particularly with AIC weighting, are closer to the CEA forecast than are the ARIMA forecasts. PDQ wonders which to believe. He suspects that DeLong and Krugman would favor the VAR results, because the VARs directly address their point that the high unemployment rate of 2008 should be taken into account in the forecasts. But given his own *a priori* uncertainty about the better approach, PDQ would prefer to settle this based on the data and some resulting AIC and BIC weights. But he is not aware of anyone discussing how to get such weights for combining single-equation ARIMA forecasts with two-equation VAR forecasts.²¹ Thus he improvises an approach that uses the VAR Δy equations alone to get weights for the VAR forecasts in order to combine them with the single-equation ARIMA forecasts. The procedure is described in the Appendix and combines the 16 ARIMAs with 21 of the original 625 VARs. Since the fundamental issue is whether *un* matters for forecasting, and because PDQ is quite uncertain about this, he gives equal prior probabilities to the set of models with *un* and to the

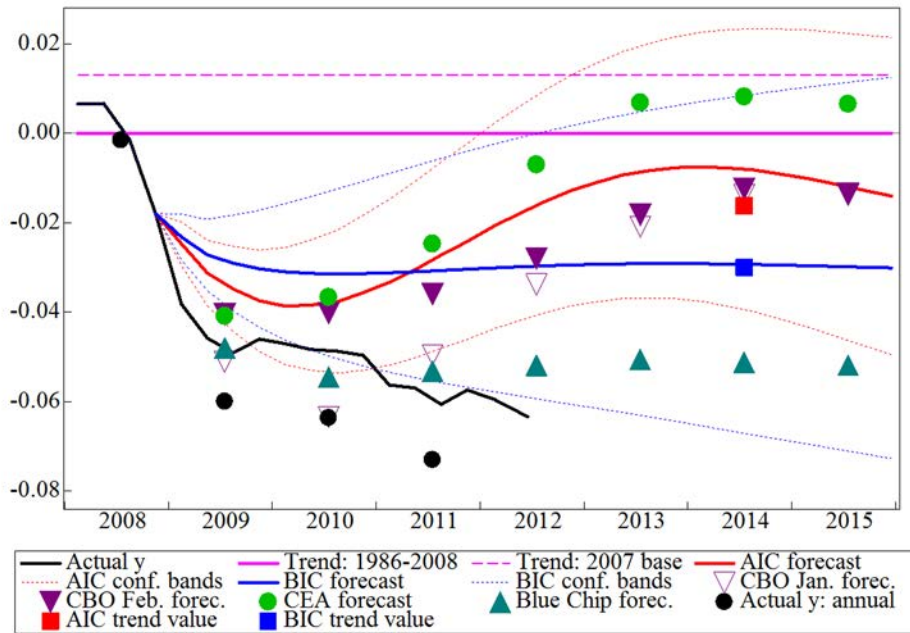
21. This is based on an email from Bruce Hansen dated January 23, 2012, to PDQ's alter ego, the author. Such a technique would have been no more likely to exist three years earlier when PDQ was working.

set without um (the “without um ” models being the ARIMA models and one of the VARs).²²

The results are in Figure 3. Unsurprisingly, the forecasts lie in between those of Figures 1 and 2, lying 1.6 and 3.6% below the CEA’s for 2013. The probability that Mankiw loses his bet according to the overall forecast can also be computed: combined AIC, 28%; combined BIC, 14%.

A key observation is that neither the VARs nor the ARIMAs clearly dominate in the weighting. The AIC weights favor the VAR models while the BIC weights favor the ARIMAs, but these “preferences” are not overwhelming. Thus, the weights do not support DeLong and Krugman’s notion that unemployment rates would be necessarily useful in forecasting real GDP.

Figure 3. Combined ARIMA/VAR forecasts, other forecasts, and actual values



22. The 17 equations without um (the 16 ARIMAs plus the AR(4) Δy equation) thus each get slightly more weight each (18.5/17) than each of the 20 equations with um (18.5/20).

TABLE 1. Various forecasts relative to the extrapolated 1986-2008 linear trend

	2013	2014
Admin (CEA) forecast (Figs. 1-3)	0.007	0.008
Linear trend: 2007 base (Figs. 1-3)	0.013	0.013
Linear trend: 1986-2008 (Figs. 1-3)	0.000	0.000
Linear trend: 2008 base (Figs. 1-3)		-0.002
Linear trend: 2008:4 base (Figs. 1-3)		-0.018
ARIMA AIC forecast (Fig. 1)	-0.035	
ARIMA BIC forecast (Fig. 1)	-0.033	
ARIMA AIC trend (Fig. 1)		-0.035
ARIMA BIC trend (Fig. 1)		-0.033
VAR AIC forecast (Fig. 2)	-0.001	
VAR BIC forecast (Fig. 2)	-0.019	
VAR AIC trend (Fig. 2)		-0.010
VAR BIC trend (Fig. 2)		-0.019
Combined AIC forecast (Fig. 3)	-0.009	
Combined BIC forecast (Fig. 3)	-0.029	
Combined AIC trend (Fig. 3)		-0.016
Combined BIC trend (Fig. 3)		-0.030

Any rebound at all?

PDQ's point forecasts are that real GDP will not recover to the levels forecasted by the CEA for Mankiw's bet year of 2013, but this does not necessarily tell us about rebounds because the forecasted adjustments are not yet complete. To judge rebounds, we need to define a starting point and then compute whether the forecasted long-run equilibrium is higher than that, adjusted for the forecasted trend growth. For example, some of PDQ's forecasts for the long-run recovery are actually more pessimistic than his 2013 values because in those cases 2013 is near the peak of an oscillation in the recovery path.

What starting point should PDQ apply? The CEA's February discussion was all in terms of annual growth rates, and 2008 was the base year for the CEA's first annual growth forecast, so 2008's annual real GDP would certainly seem reasonable as the rebound starting point. A variation would be to use the fourth quarter of 2008. Because this value is less than the annual value, it is a less stringent criterion for declaring a rebound. Finally, one could decide that if real GDP is forecasted to eventually recover at least somewhat from whatever low is ultimately reached after 2008, then a rebound is forecasted. Then all we need is some amount of forecasted

recovery from the troughs that all the forecasts show in 2009 or 2010. But the CEA was forecasting that a rebound would occur relative to 2008, so PDQ sticks to that.

PDQ then computes forecasted long-run trend values using the AIC and BIC weights. The values are plotted for the year 2014 in the Figures 1 to 3 to allow easy comparison with the CEA forecasted long-run trend value, which the annual CEA forecast reaches the same year. The long-run values can then be graphically compared with the two 2008 starting values by imagining a horizontal line drawn from the long-run value in 2014 back to the values in 2008. Or they can be compared numerically by applying a few computations to results in Table 1.

These comparisons assume that the various models' forecasted long-run growth rates are the same as the linear-trend annual growth rate (2.728%) used to normalize the values in the graphs. PDQ's ARIMA and VAR growth rates are not exactly equal to that, but they are close enough (2.697% to 2.736%) that applying the model-specific growth rates makes no discernible difference to the computed rebound sizes forecasted by PDQ's models. These growth rates are essentially the same as the Blue Chip long-run growth rate of 2.7%.

PDQ finds no rebounds relative to the annual 2008 value in the weighted forecasts of *any* of his models. The only weighted forecast rebounds he finds are relative to the 2008:4 starting point using the VAR and the combined ARIMA/VAR forecasts with AIC weighting. The rebounds are rather small, amounting to 0.8% (VAR) or 0.2% (combined ARIMA and VAR).²³ Note that real GDP's move to the high point of the AIC-weighted VAR forecast path near the end of 2013 is not a rebound in PDQ's view because it is temporary.

In contrast to PDQ's small to nonexistent forecasted rebounds, the CEA forecasted rebound is 2.6%, net of the same 2.728% long-run growth rate used for the other rebound computations. And if the CEA's own forecasted long-run growth rate of 2.6% is used instead for computing the CEA net forecasted rebound, it becomes 3.5%. To put it another way, the CEA forecast for 2013 rebounds almost exactly to a trend line extrapolated from 2007's real GDP using the CEA's growth rate of 2.6% (again consistent with Mankiw's trend reversion point).

Tyler Cowen has recently asked, "When was it obvious our recovery would be so slow?" (Cowen 2012). This was in response to several recent discussions involving the idea that rebounds tend to be stronger after deep recessions than after shallow ones (reminiscent of the 2009 CEA statement). PDQ's prediction in early 2009 of little to no rebound suggests an answer to Cowen's question, but the recent

23. This does not mean that PDQ's models predict stronger rebounds can never happen. Some impulse response analysis available from the author upon request shows the possibility of larger partial rebounds than in the present case.

discussion more specifically regards recovery from the now-known official ending point of the recession, June 2009.²⁴ We therefore should look at PDQ's forecasted recoveries from his post-2008 predicted troughs. His BIC-weighted forecasts show little to no recovery. Thus, in early 2009 it would have been predicted by (if not completely obvious to) a believer in PDQ's BIC-weighted results that the recovery would be very slow. A believer in PDQ's AIC-weighted results would have found slow recovery less obvious. Nevertheless, the AIC believer would not have predicted a recovery strong enough to get back to the 1986-2008 trend line.

How are all the forecasts doing?

I now turn to something 2009's PDQ cannot know: how all the forecasts are doing so far. The Blue Chip forecast was the most pessimistic and is the most accurate through 2011, marginally beating the January CBO forecast using the mean squared error criterion ($MSE = 0.000207$ versus 0.000209). The February CBO forecast, revised to account for the stimulus, does worse ($MSE = 0.0026$). The CEA and all PDQ forecasts but one have MSEs in the range of 0.0034 to 0.0039 , with PDQ's VAR-BIC trailing at 0.0053 . Basically, PDQ's forecasts are the least successful of the forecasts at catching the depth of the recession in 2009.²⁵

Looking beyond 2011, if real GDP does not soon experience what would be a remarkable burst of growth, then the relative performance of PDQ's ARIMA forecasts will rise, and PDQ's VAR forecasts will prove inferior (and DeLong and Krugman's insistence that the unemployment rate is essential to the forecasting will, again, be unsupported). Finally, the CBO's forecast of a much-lower growth rate of 2.2% later in the decade (given in OMB 2009) is starting to look very good.

24. Cowen cites Cochrane (2012b), who cites Taylor (2012), who cites Bordo and Haubrich (2012). Cochrane (2012b) draws various linear trends for real GDP (see also Cochrane 2012a) and thus apparently believes in linear trend stationarity and the implied full recoveries doubted by Mankiw (2009b). Cochrane (2012b) also points out that each year since 2009 the Administration forecasts have been more optimistic than the Blue Chip consensus forecasts. Bordo and Haubrich (2012) analyze U.S. recessions going back to the 1880s and find the pattern of stronger recoveries after deeper recession. But, as they mention, they do not address whether the recoveries are ever strong enough to get back to any former trend.

25. It's not clear whether the February Blue Chip forecast for 2009 and 2010 accounted for the Obama stimulus. It was, however, more *pessimistic* than the Blue Chip January forecast. The Blue Chip accuracy so far does not prove that the Blue Chip consensus would always be more accurate than other forecasts. To do so would require a study of its own.

Final remarks

In 2009, using some fairly standard techniques with a few refinements from the recent literature, my hypothetical time series econometrician, PDQ, confirms Mankiw's skepticism about the CEA's early 2009 optimistic forecasts of real GDP. Mankiw would likely win his bet, with a probability of 72% to 86% according to PDQ's overall estimates. In this instance, at least, Mankiw's intuition appears to PDQ to be superior to DeLong and Krugman's. Perhaps Mankiw's doubts were driven not only by his unit root point but also by his well-known skepticism regarding the Obama stimulus plan (e.g., Mankiw 2008, 2009a). And, although they didn't provide probabilities, the CBO and Blue Chip consensus forecasts, readily available at the time, also suggested Mankiw would win. It is therefore quite unclear to PDQ how DeLong and Krugman could have been so dismissive of Mankiw's skepticism and so certain of the CEA's forecasted rebound.

Appendix

Data and code for replication

Data and computer code to aid those interested in replication are available via econjwatch.org ([link](#)). To run all code as is, the reader would need Excel, TSP 5.1 (somewhat earlier versions should also work), and R. The reader can check the CBO data interpolations and recreate the breakpoint tests, weighted forecasts, all graphs, heteroskedasticity tests, normality tests, and unit root tests.

Mankiw's other two points

In addition to his primary point about unit roots and the permanency of shocks, Mankiw (2009b) raised two other points about the CEA (2009) rebound discussion. The CEA presented data for and a graph of the regression of the real GDP “rebound” (measured as the subsequent two-year growth rate) on the percentage peak-to-trough real GDP decline for eight U.S. post-war recessions. The regression line shows higher rebounds associated with deeper troughs, with a t -statistic of 1.97.

Mankiw's (2009b) first non-unit-root point involved the interplay between a sample selection issue and the possibility of heteroskedasticity. His write-up was somewhat ambiguous, but he kindly clarified matters for me by email. As in his blog, let G = real GDP growth and V its variance. In essence, the CEA regressed $G(t)$ on $G(t-1)$. But the CEA only used time periods consisting of recession followed by recovery (negative $G(t-1)$ followed by positive $G(t)$). Suppose there is positive ARCH(1).²⁶ Then big $V(t-1)$ means that somewhat big $V(t)$ is likely. Furthermore, an observed $G(t-1)$ that is far from its mean is more likely to come from a distribution with big $V(t-1)$ than otherwise. Thus, the more negative is $G(t-1)$, the more positive $G(t)$ is likely to be, biasing the CEA regression in the direction of showing big recessions followed by big recoveries.

Mankiw's second non-unit-root point was that the CEA regression data omitted the 1980 recession, but he did not elaborate much on the problem. I do so here. The CEA regression also omitted the 1949 recession. The CEA (2009) gave the following rationale for omitting these data points:

26. The assumption that the heteroskedasticity was ARCH was not explicit in the blog.

The 1949 recession is excluded because it was followed by the outbreak of the Korean War, resulting in exceptionally rapid growth. The 1980 recession is excluded because it was followed by another recession, resulting in unusually low growth.

In the case of the 1980 omission noted by Mankiw, it would seem that the CEA dropped the data point *because* it clearly contradicted the CEA's hypothesis. But the 1949 omission involves the same problem, because not only was that recession followed by strong growth, it was a shallow recession. If one reruns the regression with either or both of these two recessions included, the statistical significance of the rebound is completely eliminated (*t*-values fall to 1.06 or less). One might add that a regression based on only eight observations is not very believable.

Lead-up to and background for the DeLong email

In a September 26, 2011, *Wall Street Journal* essay, Harold Cole and Lee Ohanian wrote:

The Federal Reserve Board's Index of Industrial [P]roduction rose nearly 50% between the Depression's trough of July 1932 and June 1933. This was a period of significant deflation. Inflation began after June 1933, following the demise of the gold standard. Despite higher aggregate demand, industrial production was roughly flat over the following year.

A day later, Krugman (2011) posted a graph of industrial production and the producer price index (PPI) over 1929-1936 that showed generally positive co-movements including a net rise in both from July 1932 to June 1933. He wrote:

You might think that this looks pretty straightforward: output shrank when prices were falling, grew when they were rising, which is what a demand-side story would predict. But Cole and Ohanian focus on the month-to-month wiggles in 1932-33—conveniently omitting wiggles that went in an inconvenient direction—to claim that demand had nothing to do with it. This goes beyond holding views I disagree with (as does much of what happens in this debate). This is a deliberate attempt to fool readers, demonstrating that there is no good faith here.

On his blog, Stephen Williamson (2011) reacted:

Hal and Lee are two thoughtful and careful economists. I don't agree with everything they have ever said, but to call them liars is appalling.

DeLong (2011) joined in:

Williamson should be much more unhappy at Cole and Ohanian's claim that July 1932-June 1933 was "a period of significant deflation." The PPI in July 1932 is 11.1. The PPI in June 1933 is 11.2. Cole and Ohanian may be the only people who have ever managed to call a period during which the price level rose as "one of significant deflation".

On September 28, after reading these exchanges, I wondered how Cole and Ohanian could make such a mistake (if not attempting to "fool readers") and looked up the price data for myself. I found that, while the overall PPI did indeed have a net rise over the period as DeLong pointed out (about 2%), the CPI had a net fall of about 7%. I presumed that Cole and Ohanian based their "deflation" claim on the CPI rather than on the PPI.²⁷ My attempted comment on DeLong's blog was something to the effect that it was misleading to mention only that the PPI had risen a bit and so DeLong ought to have also mentioned that the CPI had fallen a lot. Although it was certainly a critical comment, I do not think I was obnoxious in expressing it, but since he did not publish my comment and I had not thought to keep a copy of what I filled into the comment form, I cannot prove it. Anyway, his emailed response was, "It's not a period of 'significant deflation' if one of your two price indexes is not falling. Shame on you for trying to confuse the issue." "Shame on me?" I wrote back. "I wouldn't have thought bringing a bit more information to bear was confusing the issue."

The revised CBO forecasts

The revised CBO forecasts (Elmendorf 2009) were given as percentage revisions to the level of real GDP in the fourth quarters of 2009-2019. The CEA, January CBO, and Blue Chip forecasts, and the Mankiw bet, were, however, all in terms of calendar years. To compare its own forecasts with the revised CBO forecasts, the CEA said that it "interpolated the impacts by quarter" (CEA 2009, footnote 3) to convert to calendar year terms, but the CEA did not precisely describe the procedure, and the CEA only applied it to 2009 and 2010. Therefore, I conducted my own conversion.

27. To recreate what would have been found in September, 2011, when the blogs and my attempted comment were written, I used the September 2011 vintage data from ALFRED.

Stimulus spending began in early March, and I assume that the initial date of its effect on GDP and thus the CBO's revision is one month later.²⁸ I assume that calendar-year real GDP values are centered at the beginning of July and fourth-quarter values in the middle of November. I divide the year into 36 periods where the 19th is deemed to be the beginning of July and the 32nd the middle of November. I then place the January CBO calendar-year values in period 19 and interpolate to get the remaining 35 values (except that actual third-quarter values and the CBO's fourth-quarter estimates as of January 2009 are used for the second half of 2008). In the interpolations, I also impose some smoothing of the January CBO growth rates from year to year. Next, I place in each year's 32nd period the midpoint of the low and high percentage increases to fourth-quarter levels in the CBO's February revision, and I interpolate to get the remaining 35 percentage increases. The resulting 36 percentage increases for each year are applied to the interpolated January CBO values, and the values for each year's 19th period are used to get the February CBO revision of annual real GDP growth rates on a calendar-year basis. My procedure gives growth rates for 2009 and 2010 that are similar to the CEA's estimates, and it also generates growth rates for the remaining years.²⁹

AIC and BIC weighting

Weighting a set of single equations is described in Hansen (2007). The formulas here are generalized to allow for weighting multi-equation models and to allow for unequal priors. Define the AIC criterion for model m with k constants and slopes and a sample size of n as $AIC_m = n \ln \left| \hat{\Sigma}_m \right| + 2k$ where $\hat{\Sigma}_m$ is the estimated error variance (single equation) or covariance matrix (multi-equation). Define the BIC criterion for model m with k constants and slopes as $BIC_m = n \ln \left| \hat{\Sigma}_m \right| + \ln(n)k$. Then, letting IC be in turn either AIC or BIC, and pr the prior probability, the weight for model m in a set of M models is

$$w_m = pr_m \cdot \exp\left(-\frac{1}{2}IC_m\right) \left/ \sum_{j=1}^M pr_j \cdot \exp\left(-\frac{1}{2}IC_j\right) \right.$$

28. See the "Stimulus Speed Chart" from ProPublica (Larson, Flavelle, and Knutson 2010).

29. The reader may wonder why I don't just use the CEA version of the CBO revision for 2009 and 2010. I do not because I compute that the 2009-2010 CEA growth rate for the revised CBO values is slightly too high to be consistent with Elmendorf (2009).

In computation, if the information criteria values are sufficiently large in absolute value, as they turn out to be in PDQ's application, they need to be normalized by subtracting some typical value from each one before applying the formula, or numerical errors occur.

Breakpoint tests

Stock and Watson (2007) present the QLR test as based on an F test. One computes F statistics for the null that all constant and slope parameters are constant against the alternative that at least one changes for all break dates in the middle 70% of the overall estimation period. The largest resulting F value is then compared to a nonstandard sampling distribution. Instead of an F test, PDQ computes a likelihood ratio test (for the ARIMAs) and a Wald test (for the VARs). Likelihood ratio and Wald tests are asymptotically equivalent (Greene 2003, 484), but in the case of the VARs, PDQ wants to test stability for a number of subsets of parameters because the overall null of stability was rejected. It is easier to do this with Wald tests because the model only needs to be estimated once instead of multiple times as with the likelihood ratio version. However, the subset results are not very revealing and are not reported here.

Because of the almost certain presence of heteroskedasticity in the post-War data, PDQ applies a recursive wild bootstrap approach to get p-values instead of comparing the likelihood ratio and Wald statistics with tabled distributions. Silvia Gonçalves and Lutz Kilian (2004) discuss the validity of the recursive wild bootstrap for autoregressive models. Assume autoregressive order r and let

$Y_{t-1} = (y_{t-1}, \dots, y_{t-r})'$. Estimate regression model $y_t = Y_{t-1}'\hat{\varphi} + \hat{\varepsilon}_t$ (with constants and trends included as desired). The estimated coefficients and residuals comprise the data generating process (DGP), with which one builds up simulated

data sets using $y_t^* = Y_{t-1}'\hat{\varphi} + \hat{\varepsilon}_t\eta_t$ where η_t is i.i.d.(0,1). PDQ follows Herman Bierens's EasyReg econometrics program procedure by generating η_t as a standard normal variable.³⁰ One then applies the desired statistical test to the simulated data sets to generate the sampling distribution. PDQ extends the procedure to the VARs and ARIMAs. The lag orders of the DGP are thus the same as in the model being tested, and the residuals are normal with a heteroskedastic component

measured by $\hat{\varepsilon}_t$. Also following Bierens, PDQ uses the first few actual data values for the initial lag values in the recursive process.

30. The current version is Bierens (2011), but the wild bootstrap procedure has been in the program since well before PDQ's working period in 2009.

For the two-equation VARs, PDQ needs a pair of residuals for each period t in the place of the single $\hat{\varepsilon}_t$ in the univariate case. The pair needs to reflect the heteroskedasticity in each equation *and* the covariance of the errors between equations. PDQ first computes moving three-period covariance matrices using the estimated residuals and then randomly selects a pair of values from the bivariate normal distribution with covariance matrix centered on period t .

Thus far, the creation of bootstrapped data sets has been described. PDQ generates 5,000 of them for each of the 16 ARIMAs and each of the 15 AIC-weighted and 10 BIC-weighted VARs (two of these are the same, so there are 23 tested VARs in all). For each model, then, PDQ has an actual breakpoint statistic for each break date and 5,000 simulated breakpoint statistics for each date. He uses the actual value and set of simulated values for each date to compute a bootstrapped p-value for that date. The date of the most significant p-value is the Stock-Watson estimate of the break date, if there is one. But to determine if there is one, an overall level of significance must be generated. Suppose the lowest p-value across all the dates is 0.03. If the null of stability is true, it is much more likely than 0.03 to get a p-value of this seemingly low value because the test has been run 167 times to cover the middle 70% of dates. Thus, we want an overall p-value that expresses how unlikely 0.03 really is. Now, each breakpoint has 5,000 simulated test statistics that can be converted into 5,000 p-values between 0 and 1. As a result, each of the 5,000 replications has a set of 167 p-values. PDQ's overall p-value is the fraction of replications with at least one p-value less than or equal to 0.03.³¹

As noted in the main text, none of the ARIMA tests were close to significance and so they are not reported here. Tables 2 and 3 give the VAR breakpoint test results for the highest weighted VAR models. Figures 4 and 5 graph the individual breakpoint p-values for the highest weighted AIC and BIC models in order to show how the results seem to suggest two breakpoints.

The main text mentions PDQ's alternative evidence of structural instability in the form of recursive estimates of trends and infinite horizon impulse responses. Figures 6 and 7 give the AIC- and BIC-weighted ones for the ARIMAs. The shocks for impulse responses are contractionary and the responses are cumulated and thus for y , not Δy . The dates on the horizontal axis refer to the beginning of the estimation period. Figures 8 to 10 give the AIC- and BIC-weighted trends and Pesaran-Shin generalized impulse responses in y for the VARs. (Because the um equations in the VARs are stable, the infinite horizon responses in um to shocks are zero.)

31. In this, PDQ is applying a procedure not specifically seen in the literature as of 2009 (so far as I know), so in this respect he seems to be violating my rule that he use only techniques available then. I therefore posit a creative burst on his part. Or perhaps he saw Cushman (2008) or a working paper version of Cushman and Michael (2011), in which similar approaches are developed.

In Figure 10, the ultimate response of y to a generalized shock in u for the estimation period commencing 1986:3 is negative. This means that the point estimate for the effect on y is for an incomplete rebound if any. However, this does not really tell us about the response to a transitory shock, because it is not possible to identify the transitory shock.

TABLE 2. Breakpoint test results for the top AIC-weighted VAR models

Lag orders	Model weight	Overall p-value	Break point	Break point 2
2, 3, 3, 3	0.171	0.019	86:3	75:4
2, 3, 2, 3	0.075	0.043	86:3	75:4
2, 3, 3, 4	0.057	0.000	86:3	75:4
2, 4, 3, 4	0.055	0.024	86:3–86:4	75:4
3, 3, 2, 3	0.049	0.017	86:3	75:4
0, 4, 3, 4	0.046	0.004	86:2–86:3	73:3, 75:4
1, 3, 3, 3	0.045	0.005	85:2–86:4	75:4
0, 3, 3, 3	0.043	0.007	85:3–87:1	75:4
2, 4, 2, 3	0.037	0.066	86:3	75:4
2, 4, 3, 3	0.034	0.032	86:3	75:4
2, 3, 4, 3	0.032	0.034	86:3	75:4
1, 4, 3, 4	0.030	0.011	86:1–86:3	75:4
3, 3, 3, 3	0.025	0.021	86:3	75:4
0, 4, 2, 3	0.020	0.030	86:3	75:4
1, 3, 2, 3	0.020	0.025	86:3	75:4

Note: The lag order column gives the Δy equation lag orders for Δy and u , then the u equation lag orders for Δy and u .

Figure 4. VAR breakpoint p-values: AIC

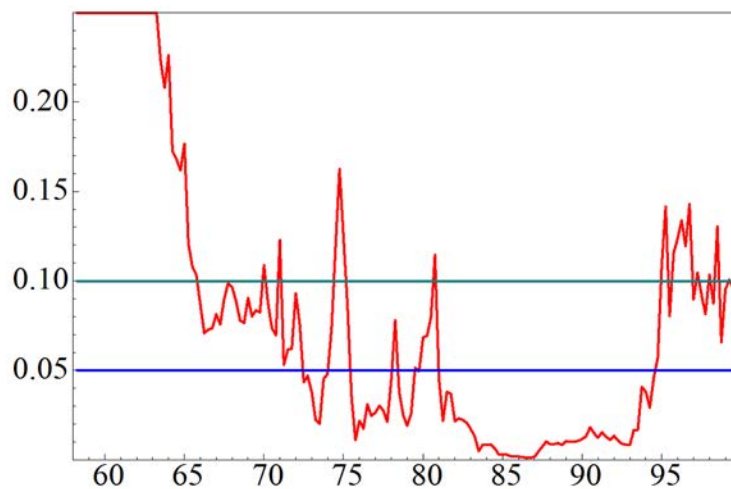


TABLE 3. Breakpoint test results for the top BIC-weighted VAR models

Lag orders	Model weight	Overall p-value	Break point	Break point 2
0, 2, 2, 2	0.377	0.010	73:2–73:3	86:3
0, 3, 2, 3	0.245	0.024	86:3	73:3
0, 2, 1, 2	0.166	0.008	73:2–73:3	86:3
1, 2, 2, 2	0.050	0.027	73:2	86:3
0, 3, 1, 2	0.042	0.018	73:2	86:2
0, 2, 2, 3	0.029	0.008	75:4	86:3
1, 2, 1, 2	0.022	0.053	73:3	86:3
0, 3, 3, 3	0.017	0.007	85:3–87:1	75:4
0, 4, 2, 3	0.008	0.030	86:3	75:4
0, 3, 2, 2	0.008	0.021	73:3	86:3

Note: Lag orders are denoted as in Table 2.

Figure 5. VAR breakpoint p-values: BIC

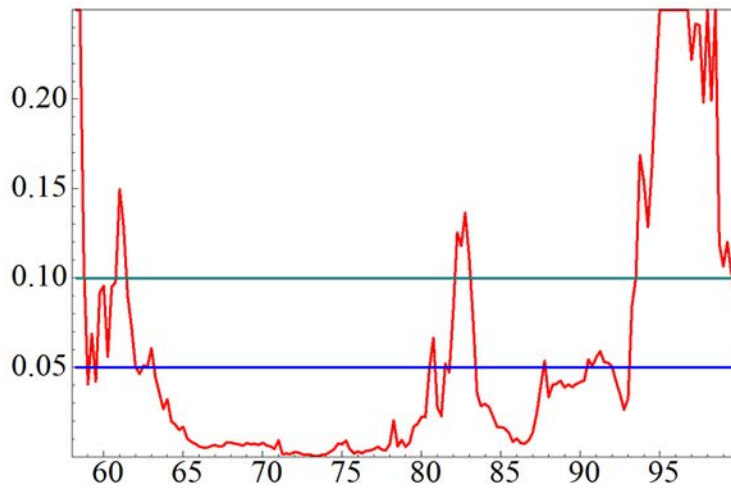


Figure 6. Recursive ARIMA trends

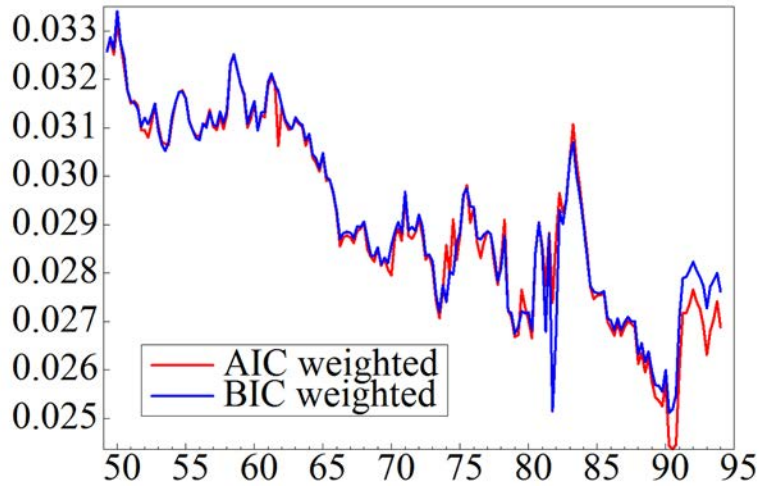


Figure 7. Recursive ARIMA impulse responses

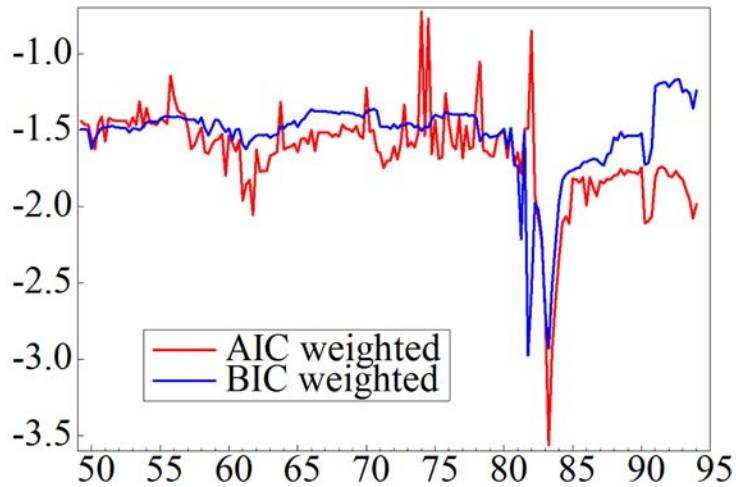


Figure 8. Recursive VAR trends

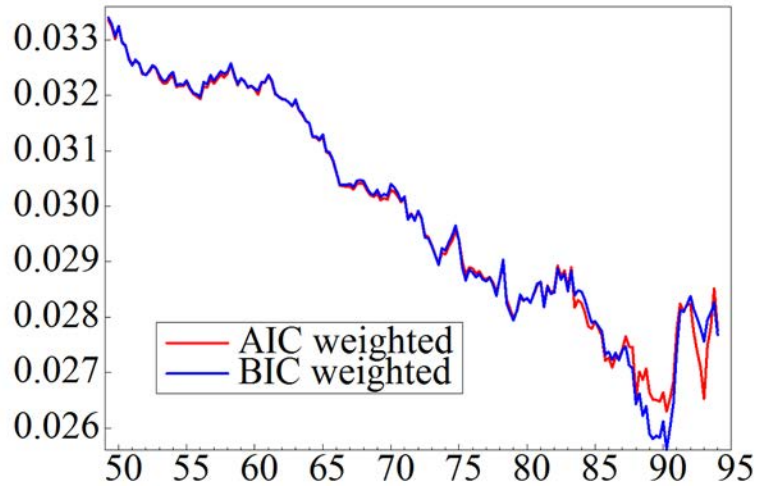


Figure 9. Recursive VAR impulse responses to Δy

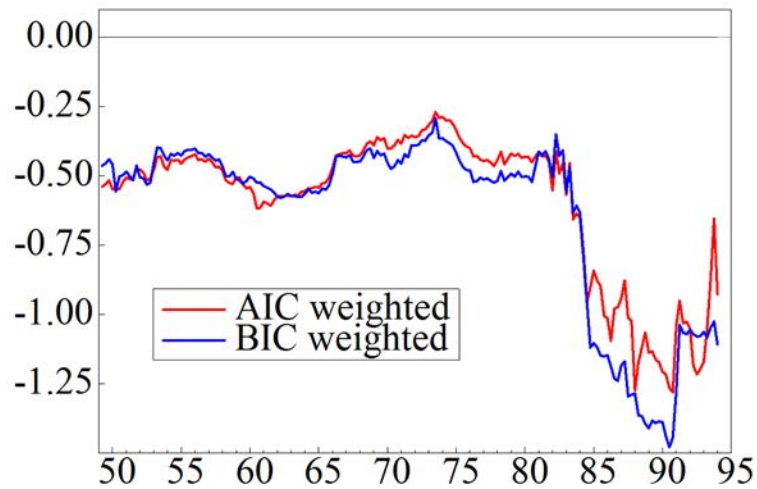
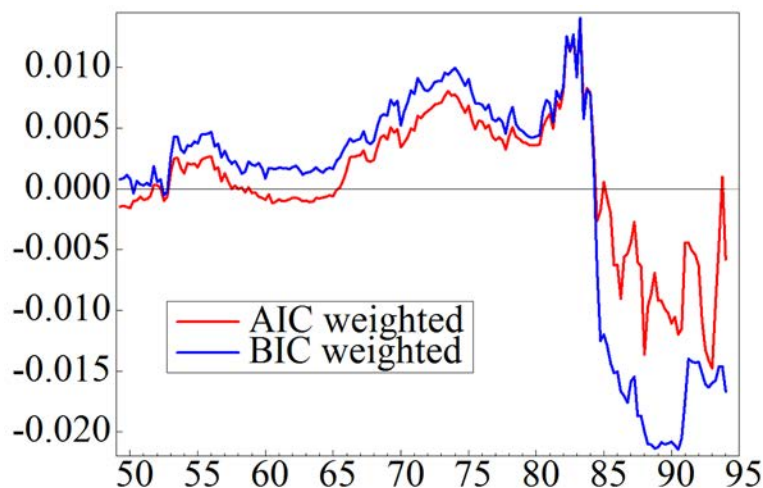


Figure 10. Recursive VAR impulse responses to um 

The Pesaran-Shin generalized shock responded to here is a pair of reduced form shocks, one for each equation, whose relationship to each other is based on the correlation of the reduced form VAR residuals. For an um shock, the shock to the um equation is 1.0 and the simultaneous shock to the Δy equation is the slope coefficient in the regression of the Δy equation residuals on the um equation residuals. For a Δy shock, the shock to the Δy equation is 1.0 and the simultaneous shock to the um equation is the slope coefficient in the regression of the um equation residuals on the Δy equation residuals.

Unit root tests

For results specific to his data set, PDQ applies the DF-GLS-MAIC test of the unit root null (Elliott, Rothenberg, and Stock 1996, Ng and Perron 2001) and the modified KPSS test of the mean or trend stationarity null (Harris, Leybourne, and McCabe 2007). The modification of the KPSS test is to filter out a near unit root (under the null) to reduce size distortion. A wild bootstrap (as in the breakpoint tests, but univariate) is used to get p-values. The DGP lag orders are determined by the MAIC lag choice of the DF-GLS-MAIC test. When bootstrapping the DF-GLS-MAIC test, in each replication the test chooses the lag order, which may therefore be different than that of the DGP (leading to “exact” p-values, as in Murray and Nelson 2000). The modified KPSS test uses a sample-size determined lag. PDQ uses $lag = \text{int}(12(T/100)^{0.25})$, a typical choice. To filter out the presumed near unit root, he uses a rho value of 0.90.

For y , the unit root null is not rejected (p-value = 0.795), confirming the Murray-Nelson (2000) finding. However, the stationarity null for y is *also* not rejected (around a linear trend, p-value = 0.366). For Δy the mean stationarity null is not rejected (p-value = 0.490) and the unit-root no-drift null *is* rejected (p-value = 0.017). Probably, then, Δy meets the assumptions of PDQ's ARIMAs and VARs. However, if y itself is actually trend stationary, then both the ARIMA and VAR models above would contain a moving average unit root ($\theta_1 = 1$, $c_1 = 0$, $\psi_{t-1} = 0$ in equations (3), (7), and (8)). The ARIMA procedure PDQ is going to use can handle this.³² The VAR, on the other hand, would need a long lag order to approximate the MA unit root process.

For u , mean stationarity is not rejected (p-value = 0.448), but neither is the unit root null (p-value = 0.237). Contrary to Krugman (2009), it is therefore not “very clear” that the unemployment rate has no unit root. For further evidence, PDQ looks at the weighted symmetric tau unit root test (Pantula, Gonzalez-Farias, and Fuller 1994) as implemented in TSP 5.1. The lag choice is from an AIC-plus-two rule. The test generates a weak unit root rejection (p-value = 0.069). Next, in light of the Caner and Hansen (2001) conclusion of two regimes for the autoregressive process (a threshold autoregression, TAR) for the adult male unemployment rate over the 1956-1999 period, PDQ applies their test to his 1986-2008 data (in monthly form), using code on Bruce Hansen's web page.³³ The possible second regime in Caner and Hansen's application, which PDQ follows exactly, is triggered by recent large increases (beyond some threshold) in the unemployment rate. Both the homogeneous-regime and unit-root nulls are rejected with (coincidentally equal) bootstrapped p-values of 0.012. If there are, in fact, two autoregressive regimes, PDQ's VARs may be misspecified. However, the two-regime result may not hold in the two-equation VAR. Furthermore, Caner and Hansen (2001), who were focusing on the econometric procedure, provide no economic rationale as to why the short-run dynamics and mean of unemployment might change when it has been rising a lot recently, regardless of its current level or other factors. Overall, PDQ concludes that assuming a homogeneous u process remains reasonable.

32. PDQ uses the exact maximum likelihood procedure of Mélard (1984) in the econometrics program

TSP 5.1. This allows $\hat{\theta} = 1$ and gets correct standard errors, unlike the traditional conditional likelihood approach in such a case. Furthermore, a long lag order is not necessary, contrary to the case of a purely autoregressive model.

33. Gauss, Matlab, and R code are provided by Hansen ([link](#)).

Serial correlation, heteroskedasticity, and normality tests

PDQ applies, where lag orders permit, Q(4), Q(6), and Q(8) Ljung-Box tests for serial correlation to the 16 ARIMA models. He applies AR(4) Wald tests to the 15 highest AIC-weighted VAR models and the 10 highest BIC-weighted models.³⁴ Q statistics are part of the standard ARIMA output in TSP 5.1. For the VAR models, both univariate and multivariate Wald tests are computed. The two univariate AR(4) tests are constructed by regressing each equation's residuals on its own one-to-four lagged residuals and a constant. The multivariate AR(4) test is the joint Wald test of significance of all non-constant coefficients in the two-equation system of the Δy and un equation residuals regressed on four lags of the residuals from both equations (therefore 16 jointly tested coefficients). The only evidence of serial correlation is in two BIC-weighted VARs where the Δy equation has no right-hand-side variables at all (other than a constant). The two VAR models' combined weight is only 15%. Because of the statistical insignificance of most of these results, I don't give them in a table, but they are available.

PDQ applies ARCH(4) tests for heteroskedasticity and Jarque-Bera (JB) tests for normality. The procedure is the same as for the univariate and multivariate AR serial correlation tests, except that the residuals are squared. P-value results are in Tables 4 to 6.

TABLE 4. ARIMA heteroskedasticity and normality test p-values

AR, MA orders	ARCH(4)	JB test
0, 0	0.534	0.133
0, 1	0.793	0.485
0, 2	0.309	0.520
0, 3	0.355	0.525
1, 0	0.770	0.657
1, 1	0.854	0.674
1, 2	0.860	0.709
1, 3	0.671	0.651
2, 0	0.677	0.685
2, 1	0.429	0.660
2, 2	0.592	0.706
2, 3	0.468	0.684
3, 0	0.562	0.668
3, 1	0.496	0.680
3, 2	0.549	0.696
3, 3	0.692	0.479

34. The top-weighted models are different from the ones for the breakpoint tests because the estimation period is different.

The ARIMAs have no significant heteroskedasticity or non-normality. Some of the VARs do show heteroskedasticity, but it appears to be in the u_t , not the Δy equation. Therefore, forecast confidence bands for Δy are likely to be tainted less than otherwise. The VAR normality tests raise no concerns. Moreover, they call into question the heteroskedasticity rejections, because heteroskedasticity will tend to be interpreted by the JB test (which assumes homoskedasticity) as a violation of normality. Similar lag structures explain many of the similarities among the results.

TABLE 5. AIC-weighted VAR heteroskedasticity and normality test p-values

Lag orders	Model weight	ARCH Δy	ARCH u_t	ARCH multi	JB Δy	JB u_t
1, 3, 1, 3	0.122	0.884	0.313	0.156	0.782	0.290
1, 3, 2, 3	0.077	0.884	0.205	0.172	0.782	0.238
0, 3, 1, 3	0.058	0.906	0.356	0.140	0.883	0.254
2, 3, 2, 3	0.049	0.825	0.161	0.085	0.849	0.220
0, 3, 2, 3	0.037	0.906	0.216	0.131	0.883	0.210
3, 3, 2, 3	0.034	0.680	0.161	0.175	0.738	0.220
2, 0, 4, 2	0.026	0.684	0.011	0.000	0.723	0.201
2, 0, 2, 3	0.026	0.684	0.042	0.001	0.723	0.158
1, 3, 1, 4	0.024	0.884	0.234	0.107	0.782	0.289
1, 3, 2, 4	0.024	0.884	0.104	0.093	0.782	0.239
2, 0, 2, 2	0.022	0.684	0.023	0.001	0.723	0.205
2, 3, 1, 3	0.022	0.869	0.313	0.110	0.808	0.290
1, 4, 1, 3	0.021	0.899	0.313	0.171	0.752	0.290
1, 3, 4, 3	0.020	0.884	0.105	0.163	0.782	0.255
1, 4, 2, 3	0.016	0.904	0.200	0.197	0.741	0.236

Note: Lag orders are denoted as in Table 2.

TABLE 6. BIC-weighted VAR heteroskedasticity and normality test p-values

Lag orders	Model weight	ARCH Δy	ARCH u_t	ARCH multi	JB Δy	JB u_t
1, 0, 1, 2	0.242	0.769	0.011	0.000	0.693	0.238
2, 0, 2, 2	0.122	0.684	0.023	0.001	0.723	0.205
0, 2, 0, 2	0.119	0.992	0.168	0.215	0.928	0.155
0, 0, 0, 2	0.107	0.534	0.039	0.003	0.160	0.072
1, 0, 1, 3	0.073	0.754	0.036	0.000	0.796	0.200
0, 2, 1, 2	0.049	0.992	0.062	0.015	0.928	0.176
0, 0, 1, 2	0.044	0.534	0.007	0.000	0.160	0.104
1, 0, 2, 2	0.040	0.769	0.008	0.000	0.693	0.187
0, 3, 1, 3	0.026	0.906	0.356	0.140	0.883	0.254
2, 0, 1, 2	0.016	0.709	0.017	0.000	0.682	0.231

Note: Lag orders are denoted as in Table 2.

Overall weighting of the ARIMA and VAR models

PDQ computes the 16 ARIMA AIC and BIC values as before, but for the VAR models he computes AIC and BIC values not for the 625 VARs, but for the 25 distinct Δy AR equations that appear in the VARs. But four of them, the ones with zero to three Δy lags and no un lags, are the same as the four ARIMA equations with no moving average terms, so the duplicate AR equations are dropped.³⁵ PDQ thus gets AIC and BIC weights for 37 equations. The 16 ARIMA forecasts are, of course, the same as before. To get the VAR forecasts, PDQ adds to each Δy equation the un equation that gives the highest VAR model probability among the 25 VARs with the given Δy specification.

References

- Bierens, Herman J.** 2011. EasyReg International. Department of Economics, Pennsylvania State University (State College, Pa.). [Link](#)
- Bordo, Michael D., and Joseph G. Haubrich.** 2012. Deep Recessions, Fast Recoveries, and Financial Crises: Evidence from the American Record. *NBER Working Paper* No. 18194. National Bureau of Economic Research (Cambridge, Mass.).
- Box, George, and Gwilym Jenkins.** 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Campbell, John Y., and N. Gregory Mankiw.** 1987a. Are Output Fluctuations Transitory? *Quarterly Journal of Economics* 102(4): 857-880.
- Campbell, John Y., and N. Gregory Mankiw.** 1987b. Permanent and Transitory Components in Macroeconomic Fluctuations. *American Economic Review* 77(2): 111-117.
- Caner, Mehmet, and Bruce E. Hansen.** 2001. Threshold Autoregression with a Unit Root. *Econometrica* 69(6): 1555-1596.
- Clark, Todd E., and Michael W. McCracken.** 2006. Forecasting with Small Macroeconomic VARs in the Presence of Instabilities. *Finance and Economics Discussion Series* 2007-41. Federal Reserve Board (Washington, D.C.).
- Cochrane, John.** 2012a. Just How Bad Is the Economy? *The Grumpy Economist*, July 31. [Link](#)

35. Because ARIMA estimation treats the lag values for the initial estimation periods differently than OLS estimation does, the overlapping models' estimates are not actually identical, but the differences are quantitatively trivial.

- Cochrane, John.** 2012b. Inevitable Slow Recoveries? *The Grumpy Economist*, August 16. [Link](#)
- Cole, Harold L., and Lee E. Ohanian.** 2011. Stimulus and the Depression: The Untold Story. *Wall Street Journal*, September 26. [Link](#)
- Council of Economic Advisers (CEA).** 2009. Economic Projections and the Budget Outlook. *Council of Economic Advisers Fact Sheets and Reports*, February 28. [Link](#)
- Cowen, Tyler.** 2012. When Was It Obvious Our Recovery Would Be So Slow? (Questions That Are Rarely Asked). *Marginal Revolution*, August 17. [Link](#)
- Cushman, David O.** 2008. Real Exchange Rates May Have Nonlinear Trends. *International Journal of Finance and Economics* 13(2): 158-173.
- Cushman, David O., and Nils Michael.** 2011. Nonlinear Trends in Real Exchange Rates: A Panel Unit Root Test Approach. *Journal of International Money and Finance* 30(8): 1619-1637.
- Davis, William L., Bob G. Figgins, David Hedengren, and Daniel B. Klein.** 2011. Economics Professors' Favorite Economic Thinkers, Journals, and Blogs (along with Party and Policy Views). *Econ Journal Watch* 8(2): 126-146. [Link](#)
- DeLong, J. Bradford.** 2009. Permanent and Transitory Components of Real GDP. *Brad DeLong's Semi-Daily Journal*, March 3. [Link](#)
- DeLong, J. Bradford.** 2011. Yes, Recovery in the Great Depression Started with Roosevelt: Stephen Williamson Takes to His Fainting Couch Department. *Brad DeLong's Semi-Daily Journal*, September 27. [Link](#)
- Diebold, Francis X.** 2008. *Elements of Forecasting*. 4th ed. Mason, Ohio: South-Western.
- Elliott, Graham, and Allan Timmerman.** 2008. Economic Forecasting. *Journal of Economic Literature* 46(1): 3-56.
- Elliott, Graham, Thomas A. Rothenberg, and James H. Stock.** 1996. Efficient Tests for an Autoregressive Unit Root. *Econometrica* 64(4): 813-836.
- Elmendorf, Douglas W.** 2009. Letter to the Honorable Judd Gregg. February 11. [Link](#)
- Enders, Walter.** 2004. *Applied Econometric Time Series*. 2nd ed. Hoboken, N.J.: John Wiley and Sons.
- Gonçalves, Silvia, and Lutz Kilian.** 2004. Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics* 123(1): 89-120.
- Greene, William H.** 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, N.J.: Prentice Hall.
- Hansen, Bruce E.** 2007. Least Squares Model Averaging. *Econometrica* 75(4): 1175-1189.

- Harris, David, Stephen Leybourne, and Brendan McCabe.** 2007. Modified KPSS Tests for Near Integration. *Econometric Theory* 23(2): 355-363.
- Hyndman, Rob J., Anne B. Koehler, J. Keith Ord, and Ralph D. Snyder.** 2008. *Forecasting with Exponential Smoothing*. Berlin: Springer.
- Kim, Chang-jin, and Charles R. Nelson.** 1999. Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle. *Review of Economics and Statistics* 81(4): 608-616.
- Koop, Gary, and Simon Potter.** 2003. Forecasting in Large Macroeconomic Panels using Bayesian Model Averaging. *Federal Reserve Bank of New York Staff Reports* No. 163. Federal Reserve Bank of New York.
- Krugman, Paul.** 2009. Roots of Evil (Wonkish). *The Conscience of a Liberal, New York Times*, March 3. [Link](#)
- Krugman, Paul.** 2011. Bad Faith Economic History. *The Conscience of a Liberal, New York Times*, September 27. [Link](#)
- Larson, Jeff, Christopher Flavelle, and Ryan Knutson.** 2010. Stimulus Speed Chart. *ProPublica*, July 30. [Link](#)
- Lütkepohl, Helmut.** 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- Mankiw, N. Gregory.** 2008. Stimulus Spending Skeptics. *Greg Mankiw's Blog*, December 18. [Link](#)
- Mankiw, N. Gregory.** 2009a. Is Government Spending Too Easy an Answer? *New York Times*, January 10. [Link](#)
- Mankiw, N. Gregory.** 2009b. Team Obama on the Unit Root Hypothesis. *Greg Mankiw's Blog*, March 3. [Link](#)
- Mankiw, N. Gregory.** 2009c. Wanna Bet Some of That Nobel Money? *Greg Mankiw's Blog*, March 4. [Link](#)
- Mélard, G.** 1984. Algorithm AS 197: A Fast Algorithm for the Exact Likelihood of Autoregressive-Moving Average Models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 33(1): 104-114.
- Morley, James C., Charles R. Nelson, and Eric Zivot.** 2003. Why Are the Beveridge-Nelson and Unobserved Components Decompositions of GDP So Different? *Review of Economics and Statistics* 85(2): 235-243.
- Murray, Christian J., and Charles R. Nelson.** 2000. The Uncertain Trend in U.S. GDP. *Journal of Monetary Economics* 46(1): 79-95.
- Ng, Serena, and Pierre Perron.** 2001. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica* 69(6): 1519-1554.
- Office of Management and Budget (OMB).** 2009. *A New Era of Responsibility: Renewing America's Promise*. February 26. Washington, D.C.: U.S. Government Printing Office. [Link](#)

- Pantula, Sastry G., Graciela Gonzalez-Farias, and Wayne A. Fuller.** 1994. A Comparison of Unit-Root Test Criteria. *Journal of Business and Economic Statistics* 12(4): 449-459.
- Perron, Pierre.** 1989. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica* 57(6): 1361-1401.
- Perron, Pierre, and Tatsuma Wada.** 2009. Let's Take a Break: Trends and Cycles in US Real GDP. *Journal of Monetary Economics* 56(6): 749-765
- Pesaran, H. Hashem, and Yongcheol Shin.** 1998. Generalized Impulse Response Analysis in Linear Multivariate Models. *Economics Letters* 58(1): 17-29.
- Pindyck, Robert S., and Daniel L. Rubinfeld.** 1997. *Econometric Models and Economic Forecasts*. 4th ed. New York: McGraw-Hill.
- Pissarides, Christopher A.** 1992. Loss of Skill During Unemployment and the Persistence of Employment Shocks. *Quarterly Journal of Economics* 107(4): 1371-1391.
- Shelley, Gary L., and Frederick H. Wallace.** 2011. Further Evidence Regarding Nonlinear Trend Reversion of Real GDP and the CPI. *Economics Letters* 112(1): 56-59.
- Stock, James H., and Mark W. Watson.** 2002. Has the Business Cycle Changed and Why? *NBER Macroeconomics Annual* 17: 159-230.
- Stock, James H., and Mark W. Watson.** 2007. *Introduction to Econometrics*. 2nd ed. Boston: Pearson Education.
- Taylor, John B.** 2012. Debate and Evidence on the Weak Recovery. *Economics One*, May 2. [Link](#)
- Tsay, Ruey S.** 2005. *Analysis of Financial Time Series*. 2nd ed. Hoboken, N.J.: John Wiley & Sons.
- Williamson, Stephen.** 2011. Hal Cole and Lee Ohanian Are Bad Guys. *Stephen Williamson: New Monetarist Economics*, September 27. [Link](#)
- Wright, Jonathan H.** 2008. Bayesian Model Averaging and Exchange Rate Forecasts. *Journal of Econometrics* 146(2): 329-341.

About the Author



David O. Cushman is Captain William McKee Professor of Economics and Business at Westminster College (Pennsylvania), and currently visiting professor (nonresident) at Oxford Brookes University. He received a Ph.D. from Vanderbilt University. Cushman previously held regular positions at the University of New Orleans, the University of South Florida, and the University of Saskatchewan, where he is Professor Emeritus of Economics. He has also held visiting positions at the University of North Carolina and Vanderbilt University, and been visiting scholar at Tilburg University and the European University Institute. His most highly cited papers have appeared in the *Journal of Monetary Economics*, *Review of Economics and Statistics*, *Journal of International Economics*, *Review of World Economics*, and *Journal of International Money and Finance*. Cushman also performs on the oboe and earlier in life played professionally with the Daytona Beach Municipal Band, Florida Symphony, and Nashville Symphony. His email is cushmado@westminster.edu.

[Go to Archive of Economics in Practice section](#)
[Go to September 2012 issue](#)



Discuss this article at Journaltalk:
<http://journaltalk.net/articles/5777>