Xiaoming Cai, Wouter J. Den Haan and Jonathan Pinder

Predictable recoveries

Article (Accepted version)
(Refereed)

Original citation: Cai, Xiaoming, Den Haan, Wouter J. and Pinder, Jonathan (2016) *Predictable recoveries.* Economica, 83 (330). pp. 307-337. ISSN 0013-0427
DOI: 10.1111/ecca.12185

Predictable Recoveries

Xiaoming Cai, Wouter J. Den Haan, and Jonathan Pinder*

January 22, 2016

Abstract

A random walk with drift is a good univariate representation of US GDP. This paper shows, however, that US economic downturns have been associated with predictable short-term recoveries and with changes in long-term GDP forecasts that are substantially smaller than the initial drop. To detect these predictable changes, it is important to use a *multivariate* time series model. We discuss reasons why univariate representations can miss key characteristics of the underlying variable such as predictability, especially during recessions.

Key Words: forecasting, unit root, business cycles

JEL Classification: C53,E32,E37

*Cai: Tongji University. E-mail: xmingcai@gmail.com. Den Haan: London School of Economics, CFM, and CEPR. E-mail: wjdenhaan@gmail.com. Pinder: London School of Economics and CFM. E-mail: j.m.pinder@lse.ac.uk. We would like to thank Bas Jacobs, Sweder van Wijnbergen and two anonymous referees for useful comments.

1 Introduction

Accurate forecasts of future economic growth are very valuable, for example, because they are needed for policymakers to decide on the appropriate stance of monetary and fiscal policy. Good forecasts are also important for the private sector, for example, for investment decisions or purchases of durable consumption goods. For these reasons, it is important that such forecasts are done with utmost care; forecasts that are too pessimistic or too buoyant could induce the wrong decisions and be quite harmful. Understanding what lies ahead is especially important during recessions, which explains the strong interest to understand what the short-term and long-term consequences of economic downturns are for future output levels.

Campbell and Mankiw (1987) argued that:

> "The data suggest that an unexpected change in real GDP of 1 percent should change one's forecast by over 1 percent over a long horizon."

Thus, shocks to GNP are permanent. Moreover, it implies that reductions in real activity are associated – if anything – with predictable deteriorations, not predictable recoveries. More recently, this quote was repeated on Mankiw's blog.[1] Campbell and Mankiw (1987) base their conclusion on estimated univariate ARMA models, that is,[2]

$$\phi(L)\,\Delta y_t = a_0 + \theta(L)\,e_t, \tag{1}$$

where y_t is the log of real GDP and e_t is a serially uncorrelated shock. In this class of time-series models, there is only one type of shock, that is, the response of output to realizations of e_t is always the same, independent of why there is a shock to output.

The contribution of this paper is twofold. First, we document that the claim made in Campbell and Mankiw (1987) is not very accurate. Using a simple *multivariate* time series model, we show that US recessions were often (but not always) followed by *predictable* recoveries.[3] Consistent with the results in Campbell and Mankiw (1987), these recoveries were not predicted by univariate time-series models.

The second contribution of this paper is to put forward reasons why univariate time-series models for GDP may lead to inaccurate forecasts. Key in our arguments is that GDP is an *aggregate* of other random variables.

The first reason is that a univariate representation does not have the flexibility to incorporate shocks with different persistence levels. A striking illustration is given in Blanchard *et al.* (2013). They construct an example in which the *correct* univariate specification of a stochastic variable that is the sum of an integrated variable *with* predictable changes and a stationary variable, also *with* predictable changes, is a random walk. That is, using only information about the aggregate variable, the correct univariate representation indicates that all changes are permanent, even though both innovations of the underlying system imply predictable further changes. We derive a more general version of this result.

The key lesson is the following. Macroeconomic aggregates are likely to be the sum of stationary and non-stationary variables. A correct univariate representation of such a variable must indicate that it is non-stationary, which means that the impact of the shock of the univariate representation necessarily has a permanent impact. We show that similar distortions occur when a random variable is the sum of two stationary variables with different persistence levels.

The second reason that univariate models may prove problematic is that the true ARMA representation of an aggregate variable may be more complex than the most complex ARMA process of each of its component series. This argument, pointed out by Granger and Morris (1976) and Granger (1980), means that with a finite data sample it might be difficult to identify the correct ARMA specification. This means that univariate time series models for aggregate variables may generate misleading forecasts. In this paper, we analyze how the under-parameterization of a univariate time series model can lead to biased forecasts.

We compare predictions of the univariate representation with those based on a VAR of GDP's expenditure components. It strengthens our argument that even such a simple multivariate time series model generates quite different forecasts during recessions. This finding is consistent with results from the forecasting literature that richer models can outperform univarate time series models.[4] Nevertheless, univariate time-series models have a long history and remain important. Nelson (1972) documents that large-scale macroeconometric models with many equations do not outperform forcasts made by simple ARIMA models. Similarly, Edge and Gurkaynak (2010) and Edge *et al.* (2010) show that forecasts made by DSGE models can be worse than a simple forecast of constant output growth.[5]

In section 2, we provide some theoretical background and discuss reasons why univariate representations may overestimate the long-run impact of economic downturns. In section 3, we illustrate some key time-series properties of US GDP. In section 4, we compare the precision of forecasts made by univariate and multivariate time-series models. In section 5, we document what this meant for forecasts made during US post-war recessions. In section 6, we show that multivariate representations also have advantages for predicting UK GDP, but for quite different reasons than the ones outlined above. The last section concludes.

2 Econometrics of univariate time-series models

In section 2.1, we illustrate why *univariate* time-series representations can give misleading predictions *even* if they are correctly specified. In particular, it is possible that the variable of interest, y_t, is a random walk and (i) it is not necessarily true that all changes in this variable have a permanent effect and (ii) the model's predictions made during recessions systematically overpredict the persistence of the downturn. In section 2.2, we give reasons why it may be difficult to get a correctly specified univariate representation for aggregate variables.

2.1 Univariate representation: Missing information and bias

Consider the following data generating process (*dgp*) for y_t:[6]

$$\begin{gathered} y_t \equiv x_t + z_t, \\ (1-\rho L)\, x_t = e_{x,t}, \\ (1-\rho L)(1-\rho_z L)\, z_t = e_{z,t}, \\ \mathbb{E}_t\left[e_{x,t+1}\right] = \mathbb{E}_t\left[e_{z,t+1}\right] = \mathbb{E}_t\left[e_{x,t+1}e_{z,t+1}\right] = 0,\ \mathbb{E}_t\left[e_{x,t+1}^2\right] = \sigma_x^2, \mathbb{E}_t\left[e_{z,t+1}^2\right] = \sigma_z^2, \end{gathered} \tag{2}$$

where $\mathbb{E}_t\left[\cdot\right]$ denotes the expectation conditional on current and lagged values of x_t and z_t. The persistence of the effects of $e_{x,t}$ on x_t is determined by the value of ρ and the persistence of the effects of $e_{z,t}$ on z_t is controlled by both ρ and ρ_z. We assume that

$$-1 < \rho < 1, \tag{3}$$

$$-1 < \rho_z \leq 1, \tag{4}$$

$$\frac{\rho_z}{\rho} > 1. \tag{5}$$

We define $e_{y,t}$ such that the following holds:[7]

$$(1-\rho_z L)\, y_t = e_{y,t}, \tag{6}$$

The unconditional autocovariance of $e_{y,t}$ and $e_{y,t-j}$, $\mathbb{E}\left[e_{y,t}e_{y,t-j}\right]$, is given by

$$\mathbb{E}\left[e_{y,t}e_{y,t-j}\right] = \frac{\rho^j}{1-\rho^2}\sigma_z^2 + \left(\left(\rho-\rho_z\right)\rho^{j-1} + \frac{\left(\rho-\rho_z\right)\rho^j}{1-\rho^2}\right)\sigma_x^2. \tag{7}$$

This implies that the autocovariances of $e_{y,t}$ are equal to zero if the following equation holds:[8]

$$\sigma_z^2 = \frac{\left(\rho_z-\rho\right)\left(1-\rho_z\rho\right)}{\rho}\sigma_x^2. \tag{8}$$

If this equation is satisfied, then $e_{y,t}$ is serially uncorrelated, and the correct univariate time-series specification of y_t is an $AR(1)$ with coefficient ρ_z.

In this univariate representation for y_t, there is only one shock, $e_{y,t}$, and the persistence of the effects of this shock is solely determined by ρ_z. Thus, the value of ρ does not matter at all! This is remarkable given that ρ affects the persistence of both fundamental shocks, $e_{x,t}$ and $e_{z,t}$.

To understand why the univariate representation misses key aspects of the underlying system, consider the case considered in Blanchard *et al.* (2013) when $\rho_z = 1$. The univariate representation is then given by

$$y_t = y_{t-1} + e_{y,t}. \tag{9}$$

That is, Δy_t is white noise and y_t is a random walk. Although y_t is a random walk, almost all changes in y_t imply predictable further changes according to the underlying multivariate *dgp*.[9] In particular, if $\Delta y_t < 0$ because $e_{x,t} < 0$, then there is a *predictable* recovery in y_t, since $x_t = \rho x_{t-1} + e_{x,t}$ and $0 < \rho < 1$. If $\Delta y_t < 0$ because $e_{z,t} < 0$, then there is a *predictable* further deterioration, since $\Delta z_t = \rho \Delta z_{t-1} + e_{z,t}$ and $\rho > 0$. If one only observes that $\Delta y_t < 0$, then one has to weigh the two possible cases and in this example the two opposing effects exactly offset each other, leading the forecaster to predict that the level of output will remain the same.

Although the implications are most striking when $\rho_z = 1$, which is the case considered in Blanchard *et al.* (2013), the analysis presented here makes clear that the univariate representation of y_t does not incorporate the role of ρ for *any* value of ρ_z such that $-1 < \rho_z \leqslant 1$.

The *dgp* considered in this section is special because the forecastability that is present in the different components cancels out and disappears in the univariate representation. It is true more generally, however, that important information is lost in the univariate representation of the sum of variables.

Is the predicted long-run impact correct on average? The previous discussion showed that the univariate representation given in equation (6) clearly misses some aspects of the underlying data generating process. Next, we turn to the question whether the univariate representation generates (long-term) predictions that are *on average* correct.

To simplify the discussion, we focus on a particular version of the *dgp* given in equation (2). We assume that $\rho_z = 1$ and equation (8) is satisfied, so that the univariate representation of y_t is a random walk. Moreover, we set $\sigma_x = \sigma_z = \sigma$, which implies that $\rho = 0.381966$ according to equation (8). Finally, we assume that $e_{x,t}$ and $e_{z,t}$ can take only two values, namely $-\sigma$ and $+\sigma$, both with equal probability. Note that the value of y_t remains unchanged if $e_{x,t}$ and $e_{z,t}$ have the opposite sign.

Although y_t has a random-walk representation, it systematically overpredicts the long-term consequences when output falls, i.e., during recessions, and it systematically underpredicts long-term consequences when output increases.

Before showing this, we first consider the case when output remains the same, which happens if $e_{x,t}$ and $e_{z,t}$ have the opposite sign. The (long-run) predictions based on the random-walk specification remain the same, since y_t remains the same. However, the true long-run predictions are affected as follows:

$$\begin{aligned}
&\lim_{\tau \longrightarrow \infty} \mathbb{E}_t\left[y_{t+\tau}\right] - y_t = +\sigma/\left(1-\rho\right) \text{ if } e_{z,t} = +\sigma \text{ and } e_{x,t} = -\sigma \text{ and} \\
&\lim_{\tau \longrightarrow \infty} \mathbb{E}_t\left[y_{t+\tau}\right] - y_t = -\sigma/\left(1-\rho\right) \text{ if } e_{z,t} = -\sigma \text{ and } e_{x,t} = +\sigma.
\end{aligned} \tag{10}$$

Thus, when y_t remains the same, then one fails to recognize that the long-run value of y_t has gone up half of the time and fails to recognize that this long-run value has gone down the other half of the time. However, the forecasts are not systematically wrong.

Now consider the case in which output drops, which happens when $e_{x,t} = e_{z,t} = -\sigma$. The drop in output is equal to $-\sigma_x - \sigma_z = -2\sigma$. The random-walk specification implies that the long-run impact is identical to the short-term impact, that is,

$$\lim_{\tau \longrightarrow \infty} \widehat{\mathbb{E}}_t \left[y^f_{t,t+\tau} \right] - y_t = -2\sigma, \tag{11}$$

where $\widehat{\mathbb{E}}_t [\cdot]$ is the expectation according to the (correct) univariate representation. The *true* long-run impact of the shock, however, is equal to

$$\lim_{\tau \longrightarrow \infty} \mathbb{E}_t \left[y_{t+\tau} \right] - y_t = -\sigma/(1-\rho) = -1.618\sigma. \tag{12}$$

That is, in a recession, the univariate representation systematically overpredicts the long-run negative impact of the economic downturn. Similarly, the univariate representation systematically overpredicts the long-run positive impact of an increase in y_t. So the predictions are not biased, but one clearly is too pessimistic during recessions and too optimistic during booms if one would make predictions based on the random-walk specification.

In this stylized example in which $e_{x,t}$ and $e_{z,t}$ can take only two values, one could drastically improve on the predictions of the univariate representation even if one could not observe x_t or z_t, but knows the true *dgp*. The reason is that a drop in y_t implies that $e_{x,t}$ and $e_{z,t}$ are both negative and an increase implies that both shocks are positive. The idea that the *magnitude* of the unexpected change in y_t has information about the importance of $e_{x,t}$ and $e_{z,t}$ is also true for more general specifications of $e_{x,t}$ and $e_{z,t}$, as long as one has information about the distribution of the two shocks. If one observes a very large drop in y_t, then it is typically the case that it is more likely that $e_{x,t}$ and $e_{z,t}$ are both negative than that $e_{x,t}$ is positive and $e_{z,t}$ is so negative it more than offsets the positive value of $e_{x,t}$ or vice versa. That is, the larger the economic downturn the larger the probability that a certain fraction of this downturn is driven by the transitory shock, that is, the larger the probability that a fraction of the drop in real activity will be reversed.

2.2 Aggregated variables and correctly specifying their *dgp*s

Aggregating ARMA processes. In this section, we highlight another problem with working with aggregated variables. We illustrate that the correct ARMA representation of an aggregate variable may very well be more complex than the most complex ARMA process for each of the component series. Formally, if x_t is an $ARMA(p_x, q_x)$ and z_t is an $ARMA(p_z, q_z)$, then $y_t \equiv x_t + z_t$ is an $ARMA(p, q)$ and p and q satisfy the following condition:[10]

$$p \leq p_x + p_z \text{ and } q \leq \max\{q_x + p_z, q_z + p_x\}. \tag{13}$$

These conditions give *upper* bounds for the $ARMA$ representation of the sum, y_t. Thus, the $ARMA$ representation of y_t is not necessarily of a higher order than those of x_t and z_t. In fact, in section 2.1 we gave an example in which an $AR(1)$ variable and an $AR(2)$ variable add up to an $AR(1)$ variable.[11] But that example relies on specific parameter restrictions. In practice, one should not rule out the possibility that the univariate representation of a sum of several random variables could be quite complex. In fact, Granger (1980) argues that an aggregate of many components—as is the case for typical macroeconomic variables—may exhibit long memory.[12]

One might think that the solution to this dilemma is to use more complex $ARMA$ processes for aggregate variables. The problem is that the model has to be estimated with a finite amount of data, consequently the values of p and q cannot be too high. But if the values of p and/or q are too low, then the *dgp* could be misspecified.[13]

Simple example. We will now give a simple example, in which the predictions of a univariate time-series model for an aggregated variable are quite bad if that time-series model is *not* more complex than the most complex time-series representation of the components.

Consider the following *dgp*:

$$\begin{aligned} y_t &\equiv x_t + z_t, \\ x_t &= \rho_x x_{t-1} + e_{x,t}, \\ z_t &= e_{z,t}, \\ \mathbb{E}_t\left[e_{x,t+1}\right] &= \mathbb{E}_t\left[e_{z,t+1}\right] = 0, \\ \mathbb{E}_t\left[e_{x,t+1}^2\right] &= \sigma_x^2, \\ \mathbb{E}_t\left[e_{z,t+1}^2\right] &= \sigma_z^2, \end{aligned} \tag{14}$$

with $-1 < \rho_x < 1$. Thus, y_t is the sum of two stationary random variables, an AR(1) and white noise. Equation (14) implies that

$$(1 - \rho_x L)\, y_t = e_{x,t} + (1 - \rho_x L)\, e_{z,t}. \tag{15}$$

The first-order autocorrelation of the term on the right-hand side is not equal to zero unless $\rho_x = 0$, but higher-order autocorrelation coefficients of this term are equal to zero. Consequently, y_t is an $ARMA\,(1,1)$. That is, there is a value for θ such that the following is the correct univariate time-series representation of y_t:

$$(1 - \rho_x L)\, y_t = (1 + \theta L)\, e_{y,t}, \tag{16}$$

where $e_{y,t}$ is serially uncorrelated. The value of θ is given by the following expression:[14]

$$\theta = \frac{\rho_x\left(-\mathbb{E}\left[e_{x,t}e_{z,t}\right] - \mathbb{E}\left[e_{z,t}^2\right]\right)}{\mathbb{E}\left[e_{y,t}^2\right]}. \tag{17}$$

The most complex component of y_t is x_t, which is an AR(1). So suppose that y_t is also modelled as an AR(1). That is,

$$y_t = \widetilde{\rho}_y y_{t-1} + \widetilde{e}_{y,t}. \tag{18}$$

If we abstract from sampling uncertainty, we can pin down the value of $\widetilde{\rho}_y$ using population moments:

$$\widetilde{\rho}_y = \frac{\mathbb{E}\left[y_t y_{t-1}\right]}{\mathbb{E}\left[y_t^2\right]} = \frac{\left(\rho_x + \theta\right)\left(1 + \rho_x \theta\right)}{\left(1 - \rho_x^2\right) + \left(\rho_x + \theta\right)^2}. \tag{19}$$

We are interested in whether this AR(1) specification would tend to over- or underestimate the long term effects of shocks by comparing $|\widetilde{\rho}_y|$ with $|\rho_x|$. If $|\widetilde{\rho}_y| > |\rho_x|$, then the AR(1) specification would tend to overstate the true degree of persistence. It is straightforward to show that $|\widetilde{\rho}_y| > |\rho_x|$ if and only if $\theta\rho_x > 0$, that is, if ρ_x and θ have the same sign.[15] Equation (17) implies that this happens if

$$-\mathbb{E}\left[e_{x,t}e_{z,t}\right] - \mathbb{E}\left[e_{z,t}^2\right] > 0. \tag{20}$$

This condition is satisfied if the covariance of $e_{x,t}$ and $e_{z,t}$ is sufficiently negative. Similarly, $|\widetilde{\rho}_y| < |\rho_x|$ if and only if ρ_x and θ have the opposite sign, which happens if

$$-\mathbb{E}\left[e_{x,t}e_{z,t}\right] - \mathbb{E}\left[e_{z,t}^2\right] < 0. \tag{21}$$

This condition would be satisfied if the two shocks are positively correlated.

To shed some light on the possible consequences of using an $AR(1)$ as the law of motion for y_t, we consider the case when the two shocks have the following very simple relationship:

$$e_{z,t} = \alpha e_{x,t}. \tag{22}$$

Since $e_{x,t}$ and $e_{z,t}$ are perfectly correlated, there is only one type of shock and there is a univariate time-series specification of y_t that completely captures the dynamics of y_t. Now we investigate what the consequences of misspecifying the $ARMA(1,1)$ process as an $AR(1)$—as an $AR(1)$ is the most complex of the individual underlying time series processes.

Figure 1 plots $\widetilde{\rho}_y$, i.e., the value of the coefficient of the $AR(1)$ representation of y_t, as a function of the true dominant root in the *dgp* of y_t, i.e., ρ_x. The top panel considers the case when the two shocks are negatively correlated ($\alpha < 0$). In this case, $\widetilde{\rho}_y$ is greater than ρ_x and so the $AR(1)$ process overstates the true amount of persistence. Conversely, if the shocks are positively correlated $\widetilde{\rho}_y$ is less than ρ_x, as shown in the lower panel.

[figure 1 around here]

These two panels document that long-term persistence is increased substantially for lower values of ρ_x when α is negative and that long-term persistence is decreased substantially for higher values of ρ_x when α is positive.

Figure 2 displays IRFs for three sets of parameter values. Each panel plots the true response of y_t to a one-time shock in $e_{x,t}$ and the response according to the $AR(1)$ specification for y_t. These three panels clearly document that misspecifying the aggregate variable y_t as an $AR(1)$—the correct specification of the most complex of the underlying processes—can give inaccurate impulse responses at both short and long horizons. The $AR(1)$ representation of y_t overestimates the long-term consequences of the shock when $e_{x,t}$ and $e_{z,t}$ are negatively correlated and underestimates them when the two shocks are positively correlated. The bottom two panels document that these bad long-term predictions only become apparent at forecast horizons of over 30 periods. At forecast horizons shorter than 30 periods, the $AR(1)$ representation of y_t *over*estimates the consequences of the crisis by a large margin when the shocks are *positively* correlated and vice versa. For example, when the shocks are negatively correlated, then the $AR(1)$ representation predicts that the initial reduction will be followed by an immediate but gradual recovery. By contrast, the true response is a further deterioration of almost the same magnitude followed by a somewhat faster recovery.

[figure 2 around here]

In this section, we focused on a case in which the most complex time-series specification of a component is an $AR(1)$, that is, a relatively simple process. Although the correct time-series specification of the aggregate is more complex, namely an $ARMA(1,1)$, it has only two parameters and one should be able to estimate this more complex time-series model with data sets of typical length. One can also improve on the $AR(1)$ specification by using higher-order AR processes, although these would—like the $AR(1)$—not be correct either, unless the number of lags is high enough to result in a sufficiently accurate approximation. However, the option to estimate a more complex representation may not always be feasible. If the two components are, for example, both an $AR(4)$, one would have to estimate an $ARMA(8,4)$, and if y_t is the sum of three $AR(4)$ processes, then one would have to estimate an $ARMA(12,8)$ to make sure that the univariate representation is not misspecified. In the next section, we document that a better strategy might be to estimate separate time-series models for the components and then explicitly aggregate the forecasts of the components to obtain forecasts for the aggregated variables.

3 Time series properties of US GDP

In this section, we discuss the relevance of the analyis in the last section by comparing an estimated univariate representation of US GDP with the representation that is implied by an estimated multivariate representation of its spending components.

3.1 Empirical specifications

The specification of the multivariate model is given by the following VAR:

$$\ln(s_t) = \sum_{j=1}^{p} B_j \ln(s_{t-j}) + e_{s,t}, \tag{23}$$

where s_t is a 5×1 vector containing the expenditure components, consumption, c_t; investment, i_t; government expenditures, g_t; exports, x_t; and imports. m_t. The forecast for

$y_{t+\tau}$ follows directly from

$$y_{t+\tau} \equiv e^{\ln(c_{t+\tau})} + e^{\ln(i_{t+\tau})} + e^{\ln(g_{t+\tau})} + e^{\ln(x_{t+\tau})} - e^{\ln(m_{t+\tau})}. \tag{24}$$

The estimated univariate representation for aggregate output is given by:[16]

$$\ln(y_t) = \sum_{j=1}^{p} a_j \ln(y_{t-j}) + e_t. \tag{25}$$

The time series for y_t itself is also constructed using equation (24) so that we are comparing like with like exactly. The key feature of the univariate time-series model is that there is only one type of shock. If output turns out to be lower than expected, i.e., $e_t < 0$, then the predicted effect on future values of y_t will always have the same pattern with the magnitude proportional to the value of e_t.

Both time-series processes are estimated with ordinary least squares (OLS). Given that the variables could very well be integrated, it is important to add enough lags to ensure that the shocks are stationary and spurious regression results are avoided. If the time series are known to be integrated, then efficiency gains are possible by imposing this. Additional restrictions can be imposed if the series are cointegrated. If these restrictions are correct, but are not imposed, then the estimated parameter values will converge towards the true parameter values at rate T, that is, there is superconsistency. If the restrictions are not correct and are nevertheless imposed, then the system is misspecified and the estimated system will not converge towards the true system. Because of superconsistency, we prefer not to impose these types of restrictions on the system.

3.2 Impulse response functions

The response of a negative one-standard-deviation shock to e_t on (the log of) US GDP, i.e., the impulse response function (IRF), is displayed in figure 3.[17] Even though the specification in equation (25) does not impose a unit root and contains a quadratic deterministic trend, the estimated specification documents that the response to the shock e_t is very persistent. It is exactly this type of result that underlies the argument of Campbell and Mankiw (1987) that one should expect economic downturns to have permanent effects.

[figure 3 around here]

If output is generated by the multivariate model, i.e., according to equations (23) and (24), then there are five reduced-form shocks that result in a drop in output. Consequently, there are five impulse response functions (IRFs), that is, five different ways in which output could respond. There are fierce debates in the economic literature on how to interpret shocks, but the interpretation of the shocks is not important for the point we want to make, that is, a model used to forecast GDP should allow for different forecasting patterns. For convenience, we will label the reduced-form shocks according to the dependent variable of the equation. For example, we will refer to $e_{c,t}$ as the consumption shock, but this is just a label and not meant to hint at a structural interpretation. The five IRFs are plotted in figure 4. The figure makes clear that according to the multivariate model there are shocks that have an extremely persistent impact on output. The figure also makes clear, however, that there are shocks that have a transitory impact on output.

[figure 4 around here]

3.3 Relevance of the theoretical arguments for modelling US GDP

The IRFs displayed in figure 4 indicate that several of the issues raised in section 2 could be relevant for forecasting US GDP using a univariate representation. The IRFs indicate that some events have long lasting consequences and others do not. For example, the "consumption shock" has a very persistent effect, but the "investment shock" and the "export shock" do not. This means that the analysis of section 2.1 is relevant. That is, since some components of US GDP are not stationary, the univariate representation will imply that *all* shocks to GDP will have a long-lasting effect.

With a finite sample, it is more difficult to determine whether the relatively parsimonious representation of GDP used here is the correct univariate representation. But the results of section 2.2 may give some guidance on potential problems. We find that the innovations of the components of GDP are positively correlated. As documented in figure 4, GDP consists of very persistent and not so persistent components. This resembles the example displayed in the bottom panel of figure 2. In this example, the univariate representation of the aggregate random variable overestimates the impact of shocks for a long period (up to 30 quarters), but underestimates the very long consequences.

4 Forecasting US GDP with univariate and multivariate models

We use the univariate and the multivariate time-series models to forecast future GDP levels. Forecasts are out-of-sample forecasts, because forecasts made at t^* only use data up to date t^*.[18] We use the latest vintage of data for each forecast.

The left panel of figure 5 plots the average forecast error at different forecast horizons according to the univariate and the multivariate time-series models. The figure shows that the predictive power of the univariate model is just as good as that of the multivariate model in terms of average forecast errors. This does, of course, not imply that there are no multivariate models that outperform a univariate model. In fact, Stock and Watson (2002) document that a forecasting model that uses indexes based on the principal components of many economic variables outperforms autoregressive univariate for most (but not all) variables. Nevertheless, the result is somewhat surprising. After all, the IRFs of the expenditure components indicate that GDP has components characterized by different persistence levels and the theoretical analysis indicated that there should be advantages in constructing forecasts of the aggregate by combining the separate forecasts of the components.

But average forecast errors may obscure some interesting patterns. In particular, the multivariate model turns out to do substantially better in forecasting at longer forecast horizons during recessions. The right panel of figure 5 shows forecast errors averaged across the six US recessions starting with the 1973-75 recession. NBER dates are used to determine whether a quarter falls in a recssion. The figure shows that the multivariate model generates much better forecasts at higher forecasting horizons.

[figure 5 around here]

Since *average* forecasting errors of the two types of models are similar, there must be periods when the univariate time-series model generates better forecasts. Interestingly, that happens during "ordinary" times, when the economy is neither doing very well nor very poorly, but continues to grow at a steady pace. The estimated multivariate models have fewer degrees of freedom and this seems to come at a cost during stable periods when simple forecasting rules suffice.

For the UK, the two time-series model generate forecast errors of similar magnitude even during economic downturns. The multivariate time-series model does generate more accurate forecasts, however, at the troughs of recessions. Below, we will discuss in more detail in which way UK recessions differ from US recessions.

5 Predictable US recoveries

In this section, we discuss in more detail the differences in forecasts of the univariate and the multivariate times-series model made at the trough of recessions.

Explaining the figures. Figures 6, 7, and 8 show the results for US recessions. The vertical lines in each figure indicate the forecasting point. The thick solid line plots the actual data. Each figure also plots the predicted growth path according to the two time-series models and a deterministic time trend.[19]

1973-75 US recession. The top panel of figure 6 displays the results for the 1973-75 recession.[20] Forecasts are made at the trough of the recession, 1975Q1. Forecasts from the univariate one-type-shock model indicate that output losses will be very persistent. Instead, there is a rapid recovery back to the long-term trend. Given that there are at times persistent changes in GDP, the univariate representation will always reflect this persistence to some extent.[21] By contrast, the forecast based on the multivariate model captures the fast recovery of GDP after the trough of the recession. In addition to the predicted short-term increase in growth rates, the multivariate model also captures the subsequent return to normal growth rates. Not surprisingly, the path forecasted in 1973Q2 does not predict the recessions of the early eighties.

The exercise discussed here should not be considered as a horse race of two forecasting models. What the results show is that (i) some economic downturns are followed by faster than normal growth and seem to have little or no permanent effects and (ii) this type of pattern is unlikely to be predicted by univariate representations, whereas multivariate VARs do have the flexibility to capture this.

[figure 6 around here]

1980 US recession. The bottom panel of figure 6 displays results for the first recession of the early eighties. Forecasts are made at the trough, 1980Q3. Both models predict that the shortfall of GDP relative to its trend value observed in 1980Q3 will remain of roughly the same magnitude up till 1984. This means that both models miss the short-lived pickup in growth rates just after 1980Q3 and both miss the second recession in the early eighties. In 1984, the economy has recovered from the second recession, although GDP is still below its trend value, and GDP is in fact close to the levels predicted by both models using data up to 1980Q3.

The two 1980Q3 forecasts diverge in their predictions for the post-1984 period. The 1980Q3 forecast according to the univariate representation predicts that the gap between

GDP and its (ex-post) trend value will not become smaller. By contrast, the 1980Q3 forecast based on the multivariate model indicates that the gap will become smaller, which is indeed what happened. In 1986, GDP was back to its trend value, which is in line with the 1980Q3 prediction according to the multivariate model.

The recovery predicted by the multivariate model in 1980Q3 is quite different from the recovery predicted in 1973Q2. Whereas, the multivariate model predicts a quick return at the trough of the seventies recession, it predicts a much more gradual return at the trough of the first early eighties recession.

1981-82 US recession. The top panel of figure 7 reports the results for the forecasting exercise when forecasts are made at the end of the second early-eighties recession, 1982Q4. From this point onwards, the US economy recovers remarkably quickly. Whereas the economy is almost 9% below its (ex-post) trend level at the end of 1982, this gap is only 2.5% at the end of 1984 and only 1% at the end of 1985. The multivariate model captures this remarkable recovery very well. It does not capture, however, the fact that in subsequent years the gap gets even smaller. The univariate representation completely misses the recovery and predicts, again, that ground lost during the recession is permanent.

Both the behavior of GDP during this recession and the fact that the remarkable recovery can be predicted by a simple time-series model strongly suggest that it is not always the case that an unexpected change in real output of x percent should lead to a change of the long-term forecast of x percent.

Although our multivariate model is a simple VAR, with five variables and four lags, it allows for a rich set of dynamics. It is, therefore, not always easy to understand what features of the data lead to particular predictions. For this particular period, it is possible to point at the reason why the model predicts a sharp recovery. The period just before 1982Q4 is characterized by sharp drops in investment and exports. As documented in figure 4, these correspond to *temporary* reductions in GDP. Consequently, the multivariate model predicts that these negative influences will disappear quickly. During 1982, both consumption and government expenditures have started to grow already, which according to figure 4 correspond to permanent positive changes in GDP. This is consistent with the predicted persistence of the recovery.

[figure 7 around here]

1990-91 US recession. The bottom panel of figure 7 displays the results for the recession of the early 1990s. The results differ from those reported above for previous recessions in that now both models predict a permanent loss in GDP. Although the loss in actual GDP is indeed very persistent and GDP does not get back to its trend level until 1997, the actual loss is not permanent.

2001 US recession. The results for the early naughties recession are displayed in the top panel of figure 8. During this recession, there is not a sharp contraction in output. It is better characterized by a period of near zero growth rates. The recovery is also very gradual. The multivariate model is wrong in predicting a short-term pick up in growth rates, but is correct in its longer-term forecast that the loss in GDP is not permanent. The univariate representation predicts again that there will be no recovery, not in the short term, which in this case is indeed what happened, and also not in the long term, which is not what happened.

[figure 8 around here]

US financial crisis, 2008-2009 The bottom panel of figure 8 plots the results for the forecasts made in 2009Q2, when the sharp fall in GDP had come to a halt.[22] Similar to forecasts made in previous recessions, the multivariate model again predicts that part of the loss in output relative to trend will be recovered in a couple years. Different from forecasts made in previous recession is that the univariate now also predicts a recovery. In fact, at this point in time, the univariate model predicts stronger long-term growth than the multivariate model. Unfortunately, forecasts of both models were too optimistic.

Starting in 2012, the multivariate model starts to predict the future reasonably well. In particular, it correctly predicts that output loss relative to trend will not be reversed.[23] The univariate representation remains more optimistic than the multivariate model until the end of the sample, sometimes marginally more optimistic, but typically substantially more optimistic. Using data up to the end of our sample, the univariate model predicts that output in 2025 will be 1% below its extrapolated trend value whereas the multivariate model predicts that the gap will be 4.5%.[24]

Why are forecasts made with a univariate model too pessimistic? In section 2, we gave two reasons why univariate representations could be too pessimistic regarding the long-term impact of negative shocks. The common element in both reasons is that it is difficult for a univariate representation to generate the best possible forecast when the variable of interest is a sum of variables with different persistence.

The first reason focused on the case where the shocks affecting the aggregate where different shocks. Even the correct univariate representation has only one shock and would never be able to capture that there are actually multiple shocks that affect the aggregate for different lengths of time. The second reason focused on the case where the components are driven by the same shock, but the estimated univariate model is not complex enough.

Figure 4 showed that US GDP does consist of components with different degrees of persistence. Moreover, shocks to these components are clearly correlated. Nevertheless, we doubt that that the reason the univariate model generates different forecasts is that it is not complex enough. Our results are robust to alternative specifications and resemble those found in the literature for a variety of univariate representations. It seems more plausible to us that US GDP is affected by different types of events which affect the US economy for different durations. Univariate representations would not be able to capture this.

6 Predictable UK recoveries?

UK recessions before the financial crisis. Post-war UK recessions are not as interesting as US recessions. Instead of sharp contractions, like those observed for the US, UK recessions were typically prolonged periods of low growth rates. Similarly, recoveries were very gradual. Although the multivariate model has better long-term predictions than the univariate representation in all but one of the recessions that occurred before the financial crisis, the predictions of the two models are roughly similar. Moreover, forecasted paths are close to straight lines, which is not surprising given the shallow aspect of economic downturns in the UK. The exception to these observations is the financial crisis, which will be discussed next.

UK financial crisis, 2008-2010. Figures 9 and 10 plot the realizations of UK GDP together with forecasts made by the two models at four different forecasting points. First consider the two panels of figure 9, which plot the results when forecasts are made at the middle of the period with large negative growth rates, 2008Q4, and at the end of this period, 2009Q2.

[figure 9 around here]

In the middle of the period when GDP dropped sharply, the univariate representation predicts an immediate and sustained return to positive growth rates. It is even somewhat more optimistic than the prediction of a random walk model with drift in that it predicts that GDP will grow faster than its trend in the next couple years, that is, it predicts that part of the reduction of the pre-crisis positive gap between GDP and its trend value will be recovered. By contrast, the multivariate model predicts that GDP will grow at rates that are somewhat lower than the trend growth rate, which is closer to the observed outcomes, although also too optimistic. In 2009Q2, the univariate representation still predicts that GDP will end up substantially above its trend value. The multivariate model forecasts that growth rates would be around zero for several quarters followed by a very gradual recovery. These forecasts are slightly below the actual outcomes.

[figure 10 around here]

The two panels of figure 10 plot the results when forecasts are made in 2009Q3 and 2010Q1. Both of these quarters are in the period when the UK economy had just started its recovery. For both forecasting points, the univariate representation's predictions indicate that the economy will start growing at rates slightly higher than those observed in the past so that it still predicts that part of the losses will be recovered. By contrast, the multivariate model—using data up to 2009Q3—predicts that there first will be a period with low growth rates, which eventually is followed by a period of faster growth rates. This is indeed what happened, although the predictions are a little bit too pessimistic. Half a year later, in 2010Q1, the forecasts of the multivariate model have improved somewhat and do a good job in predicting the subsequent development of UK GDP.

We do not want to argue that the multivariate model is a remarkably good forecasting model. Neither model does very well in predicting subsequent output growth during this period, although it is worth noting that the multivariate model realizes quickly that output losses will be very persistent. The point that we want to make is that multivariate models have the flexibility to predict different types of forecasting patterns. By contrast, univariate representations are quite restrictive and may miss both predictable recoveries and—as is shown here—a predictable deterioration during a downturn. The main reason why the univariate representation is restrictive is that it has only one type of shock. Since the GDP data used to estimate the univariate representation contains a persistent component, changes in GDP will *always* lead to changes in the long-term forecasts of the univariate model. Although, univariate forecasts always have a permanent component, we allow for the possibility that short-term forecasts are different from long-term forecasts, since our empirical univariate representation has four lags. But all of our estimated univariate representations imply predictions that are quite close to those of a random walk with drift.

7 Concluding comments

Macroeconomic forecasts are made with simple univariate models, for example, Campbell and Mankiw (1987),[25] as well as with advanced multivariate models, for example, Stock and Watson (2002).

In this paper, we reviewed reasons why univariate representations of a sum of random variables could miss key predictable aspects of this random variables. In fact, even if a random variable is a random walk, then that does not mean that there are no forecastable changes. In particular, if an aggregate consists of stationary and non-stationary variables, then the univariate representation will indicate that all shocks have permanent consequences even though that is, of course, not the case for shocks to the stationary components. Moreover, the correct specification of an aggregate of random variables could be quite complex. We argued that it might be better to estimate time-series models for the components and obtain forecasts for the aggregate by explicitly aggregating the forecasts of the components.

Despite the empirical observation that US GDP consists of very persistent and less persistent variables, the univariate and multivariate time-series model have similar forecasting performance in terms of average forecast errors. Such a finding may explain why forecasts based on univariate models are still taken seriously.

However, our simple multivariate time-series model clearly outperforms the univariate model, when it is used to forecast future GDP during recessions. Whereas the univariate model typically predicts that recessions have large and negative consequences, the multivariate model often correctly predicts that this is not the case. In some cases, for example, when the drop in GDP is mainly due to drops in components with less persistence such as investment and exports, it was possible to understand why the multivariate model performed better than the univariate model. In other cases it is not. Nevertheless, the sharply better performance of our simple multivariate model during recessions and the theoretical discussion indicate that one should be careful making forecasts with univariate time-series models.

One point that we do not address is the correct level of (dis)aggregation. Consumption is the sum of non-durable and durable consumption and both are sums of individual expenditures. So further disaggregation may lead to further improvements. It is not clear, however, whether one should disaggregate to the lowest possible level, since sampling variation typically increases when one considers disaggregated variables.

A Data sources

US data. Data are downloaded from the web site of the Federal Reserve Bank of St. Louis. They are (i) Consumption: real personal consumption expenditures; (FRED code: PCECC96); (ii) Investment: real gross private domestic investment (GPDIC1); (iii) Government expenditures: real government consumption expenditures & gross investment (GCEC1); (iv) Exports: real exports of goods & services (EXPGSC1); and (v) Imports: real imports of goods & services (IMPGSC1). All time series are seasonally adjusted quarterly data measured in billions of chained 2009 dollars. The data were last updated May 29, 2015.

The GDP data used is the sum of the consumption, investment, government expenditures, and exports minus imports. Adding up these real time series generates a time series that is extremely close, but not exactly identical to the actual GDP data. Our approach ensures that the components used in the multivariate model add up *exactly* to the data used in the univariate model. This way, we avoid clutter in the paper by describing small differences in the GDP data used in the two types of time-series models.

UK data. Data are from the Office of National Statistics. They are (i) household final consumption expenditures (ONS code: ABJR) plus final consumption expenditure of non-profit institutions serving households (HAYO); (ii) total gross fixed capital formation (NPQT); (iii) general government: Final consumption expenditures (NMRY); (iv) balance of payments: Trade in goods and services: Total exports (IKBK); (v) Balance of payments: Imports: Trade in Goods and services (YBIM). All data are seasonally adjusted quarterly data and the base period is 2011. The GDP data used is the sum of these five components. Investment in inventories are excluded, since they contain some very volatile high frequency movements.

B Robustness

Figures 11 through 16 display the results for several robustness exercises. Figure 11 documents that our result that multivariate time-series models generate more accurate long-term forecasts than univariate models is also true when no deterministic trend term is included, when only a linear trend term is included, and when the number of lags are chosen by AIC. Figures 12 through 16 illustrate that even the actual forecasts are very similar when the number of lags are chosen with AIC.[26] At the earlier forecasting dates, there is a bit of variation in the number of lags chosen by AIC, especially for the univariate specification. After this, the number of lags chosen for the univariate specification is three, which is one less than our benchmark number. For the multivariate specification, the number of lags remains two for a while and then jumps to five lags, one more than our benchmark number.

Notes

[1]See: `http://gregmankiw.blogspot.com/2009/03/team-obama-on-unit-root-hypothesis.html`.

[2]They allow for the possibility that $\theta(L)$ has a root equal to 1, which would imply that y_t is stationary around a deterministic time trend.

[3]We also compare univariate and multivariate time-series models to predict UK recoveries. Whereas several US recessions were followed by remarkable recoveries, economic recoveries in the UK were much more gradual and the predictions of the two types of models are similar. However, the multivariate model does outperform the univariate model during the great recession. In particular, the multivariate model correctly predicts a further deterioration in the initial phase of the economic downturn and correctly predicts its long-lasting impact.

[4]Fair and Schiller (1990) also show that GDP forecasts based on the sum of forecasts of GDP's components help improve forecasts when compared with univariate forecasts. They use univariate representations of the components, which makes it possible to disaggregate at a higher level. Stock and Watson (2002) generate forecasts using a small number of indexes that are based on the principal components of a large set of economic variables. We refer the reader to Chauvet and Potter (2013) for a recent survey of the forecasting literature.

[5]By contrast, Smets and Wouters (2007) show that their DSGE model performs better in forecasting than a Bayesian VAR.

[6]This time-series specification is a generalization of the one studied in Blanchard *et al.* (2013).

[7]It is always true that

$$(1-\rho_z L)(1-\rho L)\,y_t = (1-\rho_z L)\,e_{x,t} + e_{z,t}.$$

Thus, an equivalent definition of $e_{y,t}$ would be the following:

$$(1-\rho L)\,e_{y,t} = (1-\rho_z L)\,e_{x,t} + e_{z,t}.$$

These two equations are helpful in deriving the formulas in this section.

[8]$\sigma_z > 0$, since we assumed that $\rho_z/\rho > 1$.

[9]In the (very) special case that $(1-\rho)x_t$ happens to be equal to $\rho\Delta z_t$, then $\mathbb{E}[y_{t+k}] = y_t$ for $k \geq 1$.

[10]See Granger and Morris (1976).

[11]In theory it is, of course, even possible that the sum of random variables is not random.

[12]One aspect that seems to be ignored in the econometrics literature is that the *dgp*s of the individual components may be "aligned" to the same factors, which could mean that the time-series representations of the components are similar, making it less likely that the aggregate has a much more complex representation

than its components. For example, if markets are complete, then market prices will align agents' marginal rates of substitution—and, thus, their consumption growth processes—even if agents face very different income processes.

[13]The misspecification is likely to be worse than indicated in this section. Typically, log-linear processes are more suitable than linear processes. But if $y_t \equiv x_t + z_t$ and x_t and z_t are log-linear processes, then neither y_t nor $\ln(y_t)$ is a linear process and the convention of modelling $\ln(y_t)$ as a linear process is, thus, not correct. In fact, the effects of shocks on y_t would be time-varying. These issues are further discussed in Den Haan *et al.* (2011).

[14]Since $e_{y,t}$ is white noise, it must be true that

$$\mathbb{E}\left[(1+\theta L)\, e_{y,t} \times (1+\theta L)\, e_{y,t-1}\right] = \theta \mathbb{E}\left[e_{y,t}^2\right].$$

It is also true that

$$\mathbb{E}\left[(1+\theta L)\, e_{y,t} \times (1+\theta L)\, e_{y,t-1}\right] = \rho_x \left(-\mathbb{E}\left[e_{x,t}e_{z,t}\right] - \mathbb{E}\left[e_{z,t}^2\right]\right),$$

since $(1+\theta L)\, e_{y,t} = e_{x,t} + (1-\rho_x L)\, e_{z,t}$ and both $e_{x,t}$ and $e_{z,t}$ are white noise. Combining both equations gives the expression for θ.

[15]Equation (19) implies that $|\widetilde{\rho}_y| > |\rho_x|$ if

$$\begin{aligned} &\frac{\left(1-\rho_x^2\right)}{\left(1-\rho_x^2\right)+(\rho_x+\theta)^2}\theta > 0 \quad \text{when } \rho_x > 0, \\ &\frac{\left(1-\rho_x^2\right)}{\left(1-\rho_x^2\right)+(\rho_x+\theta)^2}\theta < 0 \quad \text{when } \rho_x < 0. \end{aligned} \tag{26}$$

Consequently, $|\widetilde{\rho}_y| > |\rho_x|$ if and only if $\theta\rho_x > 0$, that is, if ρ_x and θ have the same sign.

[16]We follow common practice and use four lags, unless stated otherwise. In appendix B, we show that the results are similar when the number of lags is chosen by AIC, although the associated long-term forecasts are somewhat less precise. Results not reported here indicate that long-term forecasts are substantially less precise if the Bayesian Information Criterion (BIC) is used. All models in this paper also include a constant and a linear-quadratic deterministic trend. Appendix B also shows that key results are very similar if no trend is included and when only a linear trend is included. Campbell and Mankiw (1987) also consider $ARMA$ representations, but the results are similar to those obtained with AR represenations. The only exception is when third-order MA components are included, but the authors point out that the implied impulse response functions of this specification are estimated very imprecisely .

[17]See Appendix A for further details on data sources. Whereas the forecasting exercise discussed in

the next section is based on real-time data, the results in this subsection are based on the full sample of quarterly US data from 1947Q1 to 2015Q1. The results are very similar if the sample ends in 2006Q4 and the financial crisis is, thus, excluded, except that the IRF of the "import" shock is then less persistent.

[18]Strictly speaking, this is *pseudo* out-of-sample forecasting, since future data is available at each forecasting point. We estimate specifications with two lags if they have fewer than 135 observations and four lags otherwise. The exact cutoff point does not matter, but it is important to only use only two lags at the early dates of our forecasting exercise, because the specifcations with four lags generate strange forecasts, which is likely to be due to the low number of degrees of freedom. Note that four lags means estimating 23 coefficients per equation.

[19]The time trend shown in the figures is a linear trend estimated on the full sample of GDP and is included as a point of reference. The linear-quadratic trends included in the univariate and multivariate models are estimated up until t^*.

[20]Because we focus on out-of-sample forecasts, we have only 109 quarterly observations for forecasts at the trough of this recession, which leaves few degrees of freedom when the VAR is estimated with the default specification, that is, four lags for each of the five variables and a quadratic deterministic trend. By using a VAR with only two lags for this recession, we avoid the strong sensitivity of forecasts when the forecasting date shifts slightly.

[21]However, since we use an $AR(4)$ to describe real output, our model does allow for a further predictable deterioration and/or for the possibility that (a large) part of the initial drop can be expected to be reversed.

[22]At the beginning of the financial crisis, both time-series models wrongly predict that a substantial part of the losses will be recaptured quickly. These results are not displayed in the graphs.

[23]These results are not displayed in the figures.

[24]The economy was substantially above its trend value before the crisis, which means that these long-term predictions imply larger losses relative to the hypothetical case when there would have been no financial crisis and subsequent average real output growth would have been equal to the trend growth rate.

[25]More recently, Edge and Gurkaynak (2010) and Edge *et al.* (2010), show that the forecasting performance of estimated DSGE models can be worse than a simple forecast of a constant output growth.

[26]Although not shown, the same is true for different trend specifications.

References

Blanchard, O. J., L'Huillier, J.-P. and Lorenzoni, G. (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review*, **103**, 3045–3070.

Campbell, J. Y. and Mankiw, N. G. (1987). Are output fluctuations transitory. *Quarterly Journal of Economics*, **102**, 857–880.

Chauvet, M. and Potter, S. (2013). Forecasting output. In A. Timmerman and G. Elliott (eds.), *Handbook of Economic Forecasting*, Amsterdam: North-Holland, pp. 1–56.

Den Haan, W. J., Sumner, S. W. and Yamashiro, G. (2011). Bank loan components and the time-varying effects of monetary policy shocks. *Economica*, **78**, 593,617.

Edge, R. M. and Gurkaynak, R. S. (2010). How useful are estimated dsge model forecasts for central bankers. *Brookings Papers on Economic Activity*, **Fall**, 209–259.

—, Kiley, M. T. and Laforte, J.-P. (2010). A comparison of forecasts performance between federal reserve staff forecasts, simple reduced-form models, and a dsge model. *Journal of Applied Econometrics*, **25**, 720–754.

Fair, R. C. and Schiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, **80**, 375–389.

Granger, C. and Morris, M. (1976). Time series modelling and interpretation. *Journal of the Royal Statistics Society A*, **139**, 246–257.

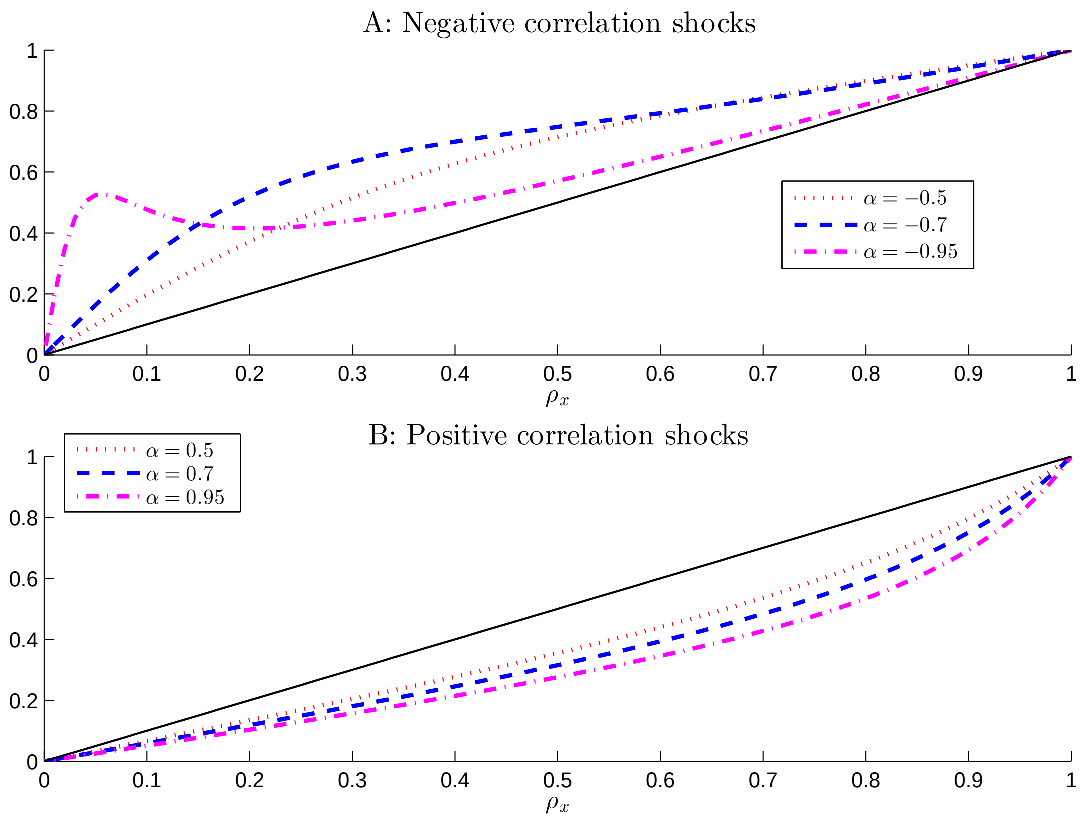
Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227–238.

Nelson, C. R. (1972). The prediction performance of the frb-mit-penn model of the u.s. economy. *American Economic Review*, **62**, 902–917.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, **97**, 586–606.

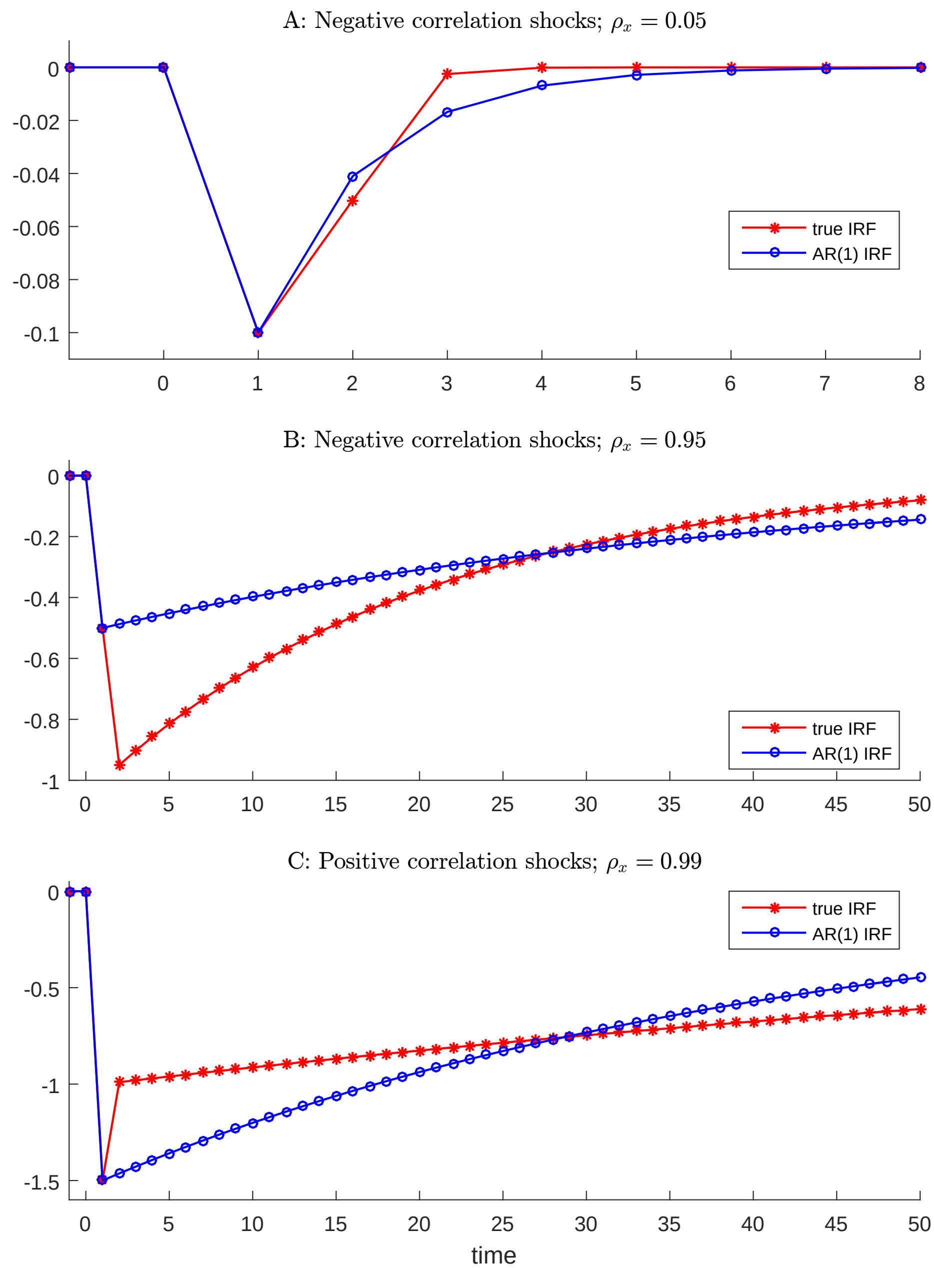
Stock, J. R. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, **20**, 147–162.

Figure 1: $AR(1)$ coefficient of $y_t = x_t + z_t$ according to incorrect univariate representation



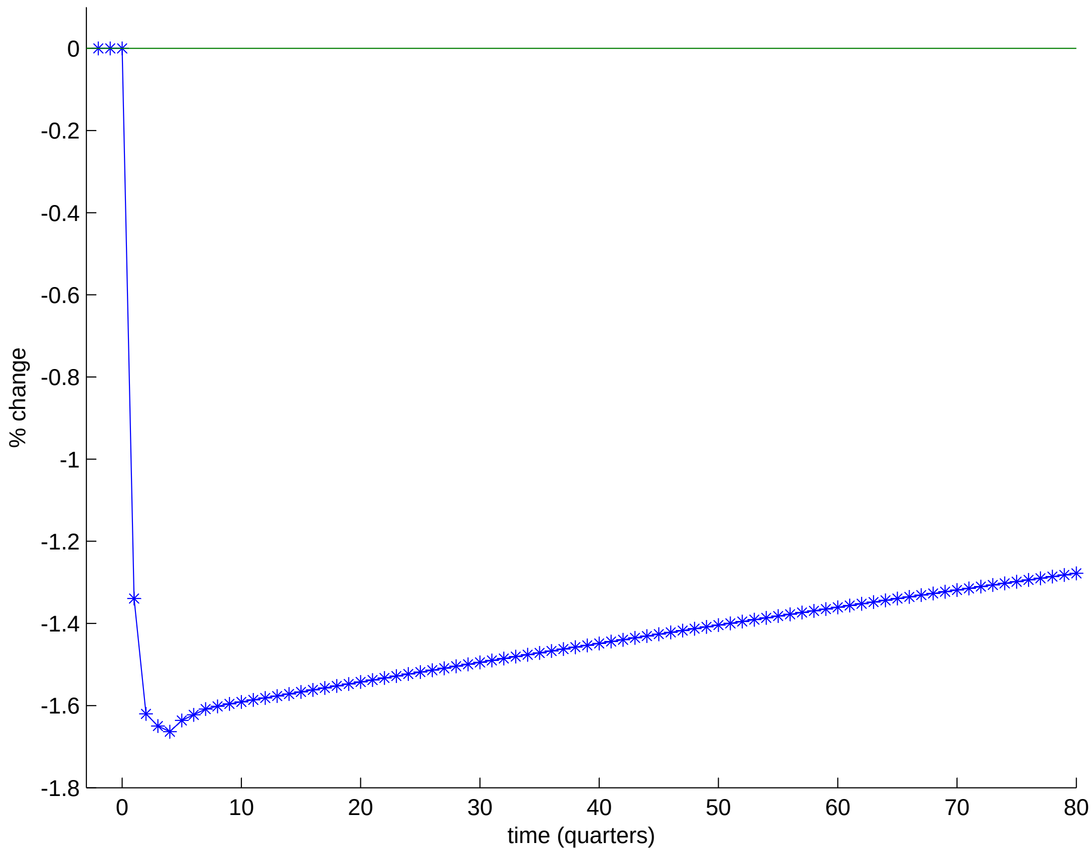
Notes: The graph displays the root of the $AR(1)$ representation of $y_t = x_t + z_t$ as a function of the AR root in the true time-series representation of y_t when $e_{z,t} = \alpha e_{x,t}$. The solid line is the 45°line.

Figure 2: IRFs of $y_t = x_t + z_t$ according to correct and incorrect univariate representation



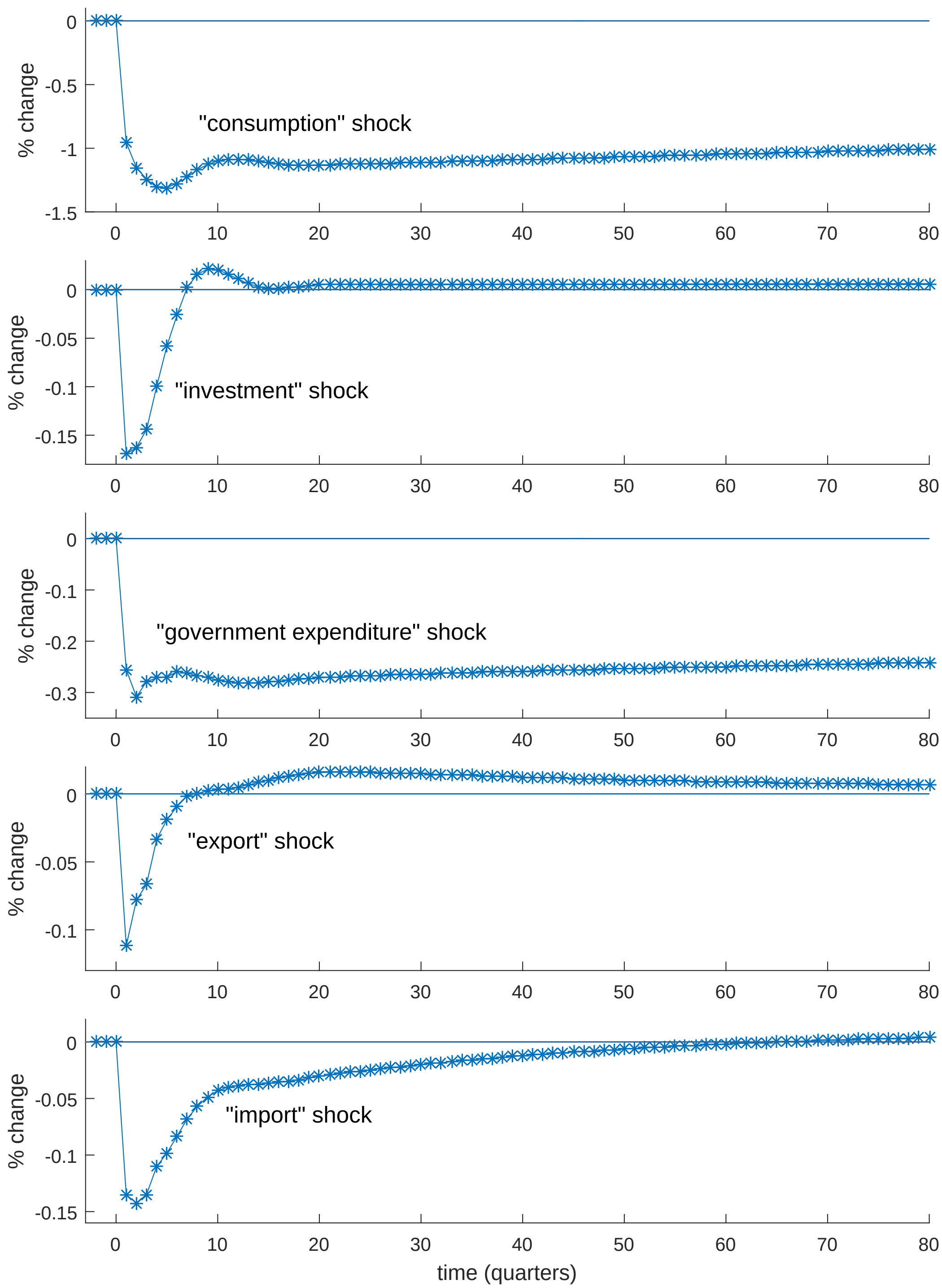
Notes: The graph plots the true responses of $y_t = x_t + z_t$ to a one-time shock in $e_{x,t}$ and the response according to the $AR(1)$ representation, which is the time-series representation of the most complex of the y_t components. In panel A, $e_{z,t} = -0.9e_{x,t}$; in panel B, $e_{z,t} = -0.5e_{x,t}$; and in panel C, $e_{z,t} = 0.9e_{x,t}$.

Figure 3: Effect of the shock in univariate representation on US GDP



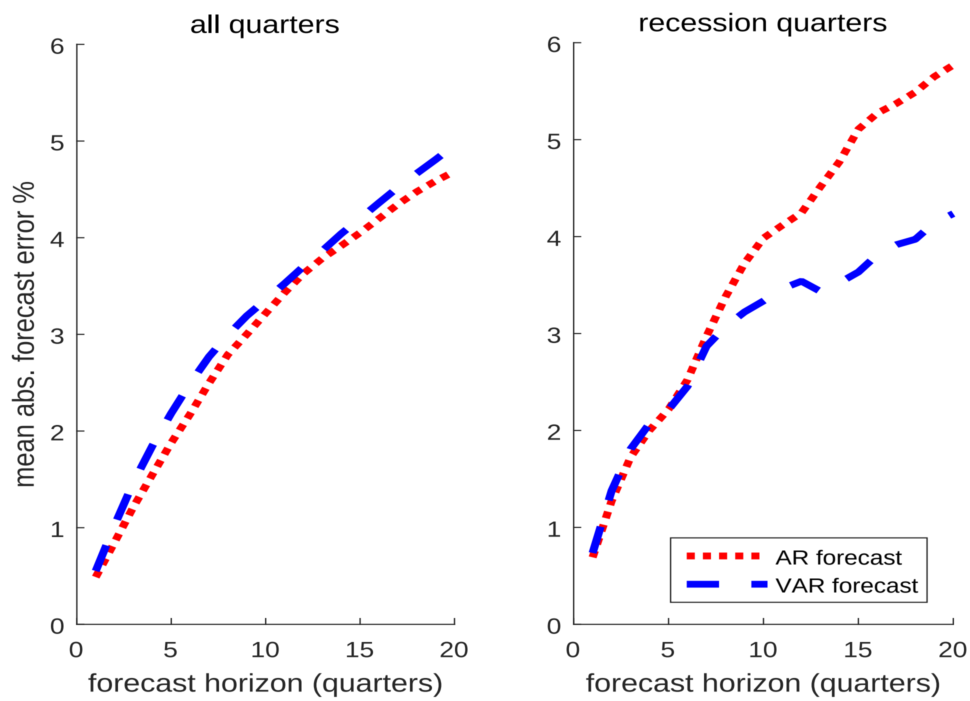
Notes:The graph plots the response of output following a one-standard-deviation negative shock according to the univariate, one-type-shock, model.

Figure 4: Effect of reduced-form VAR shocks on US GDP



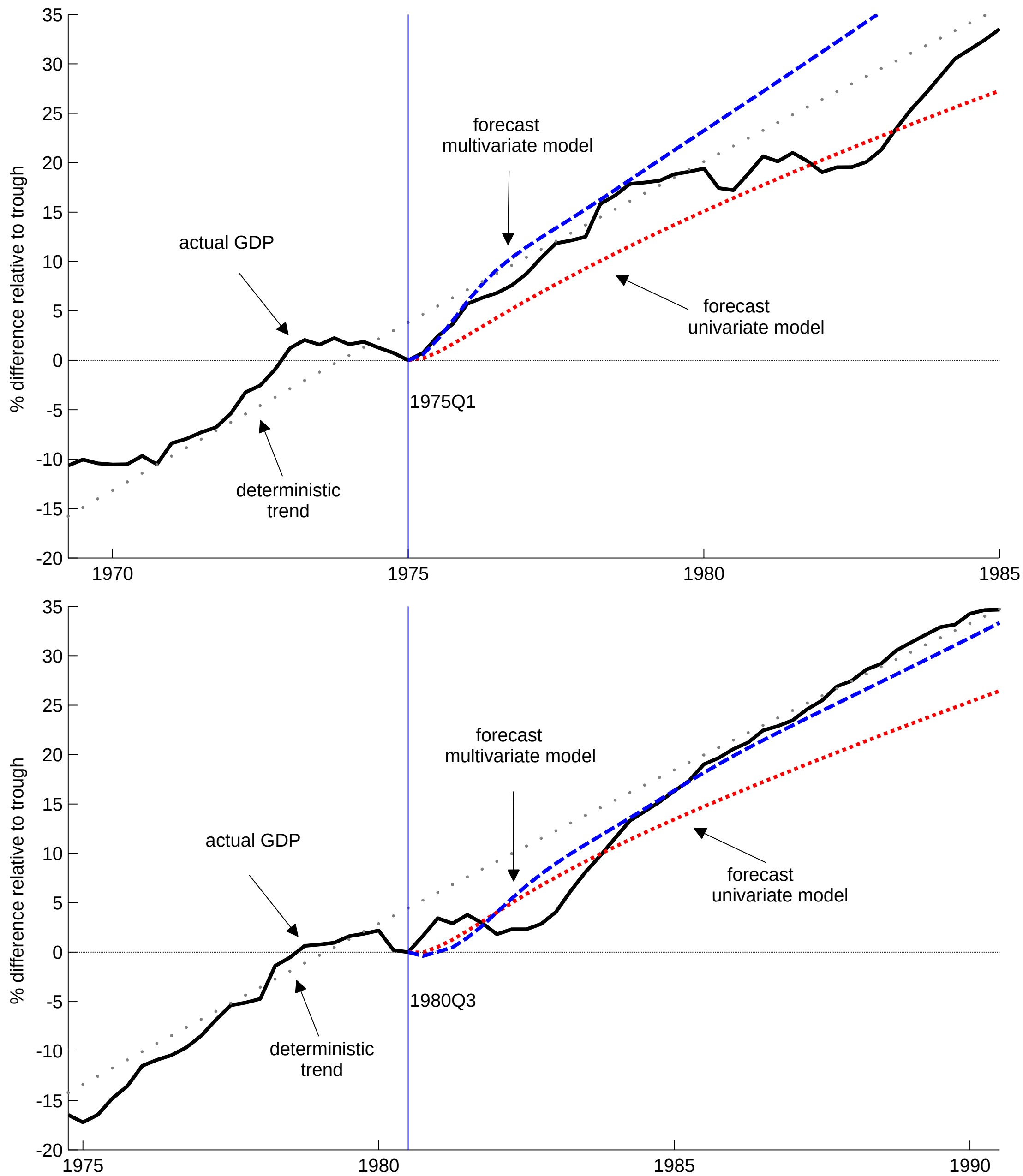
Notes: The graphs plots the predicted responses of output following a one-standard-deviation shock in the indicated reduced-form VAR shock that leads to a reduction in GDP.

Figure 5: Average forecast errors - US



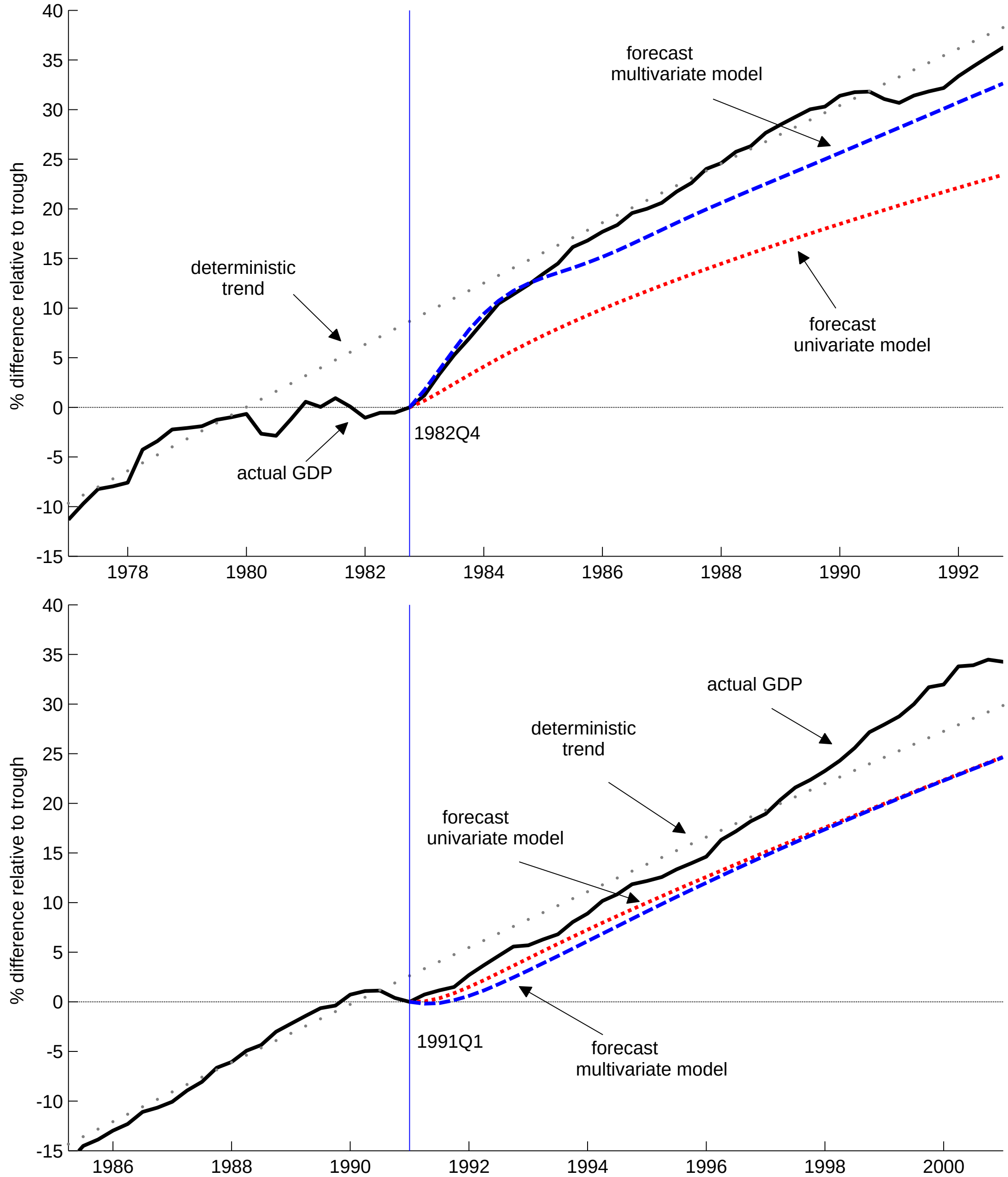
Notes: These graphs plot the average forecast errors of the indicated time-series model. NBER recessions dates are used to identify whether a quarter is a "recession quarter".

Figure 6: The 1973-75 and the 1980 US recessions



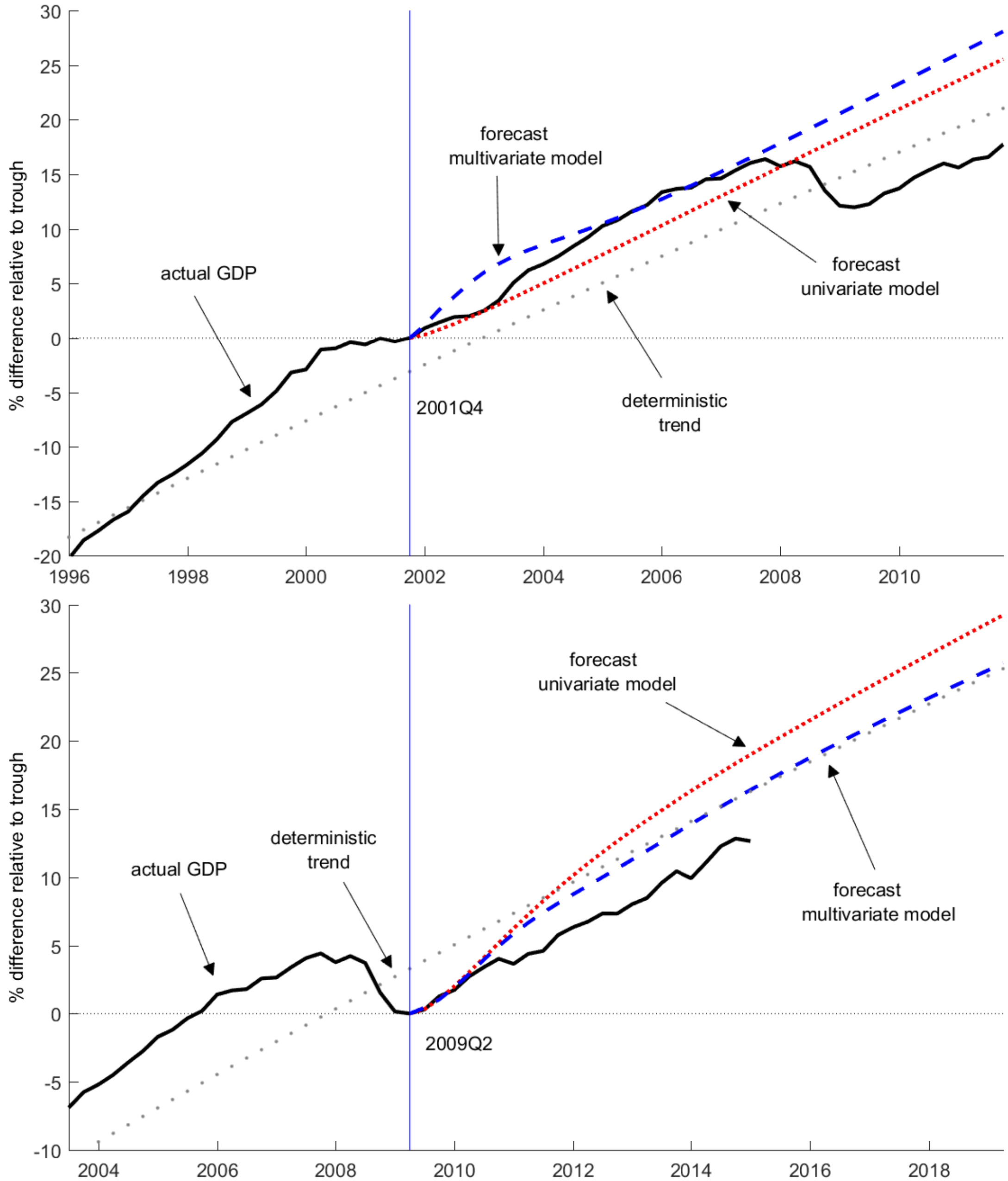
Notes: This figure plots the two forecasted time paths for US GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line.

Figure 7: The 1981-82 and the 1990-91 US recessions



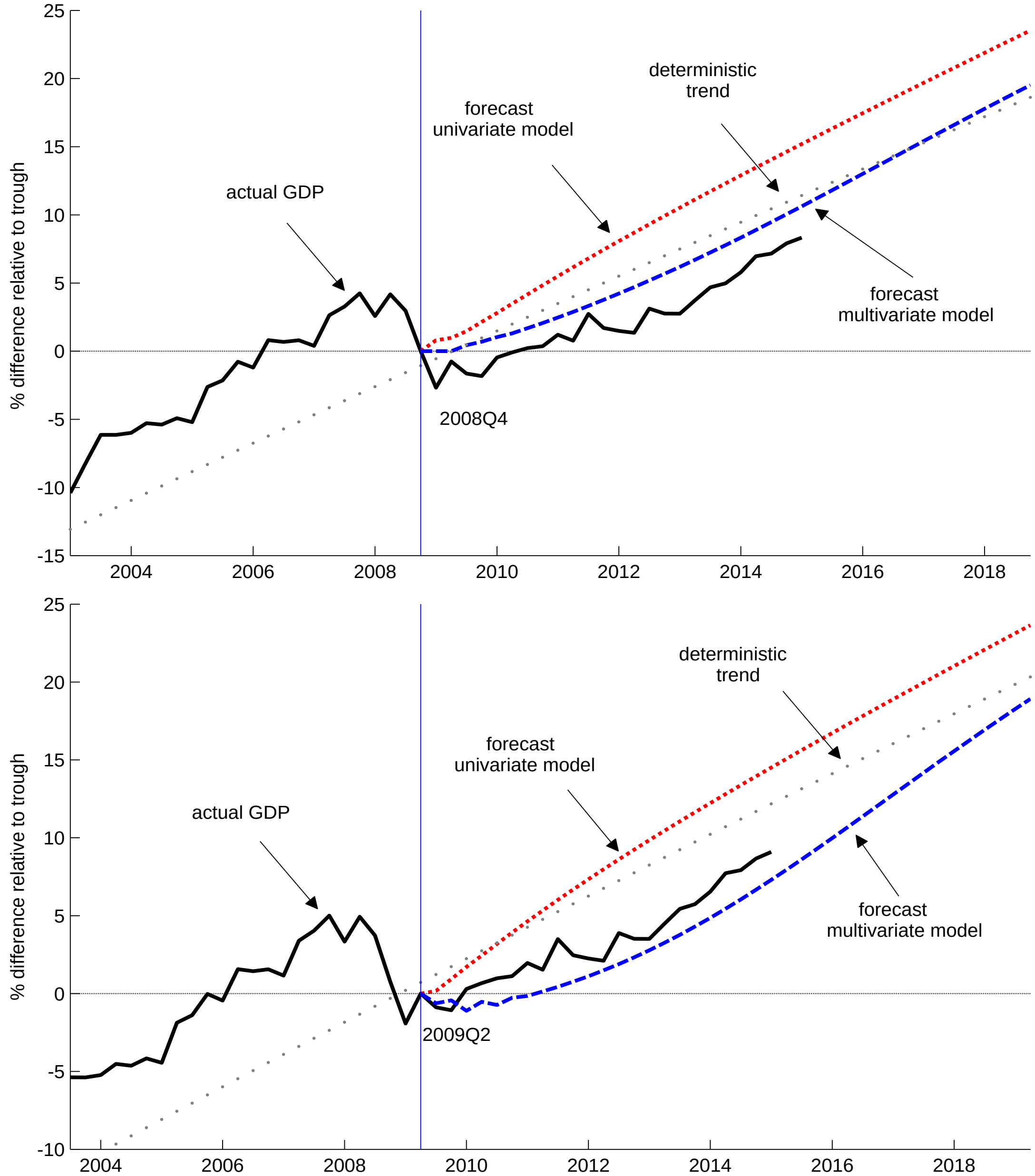
Notes: This figure plots the two forecasted time paths for US GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line.

Figure 8: The 2001 and great US recession



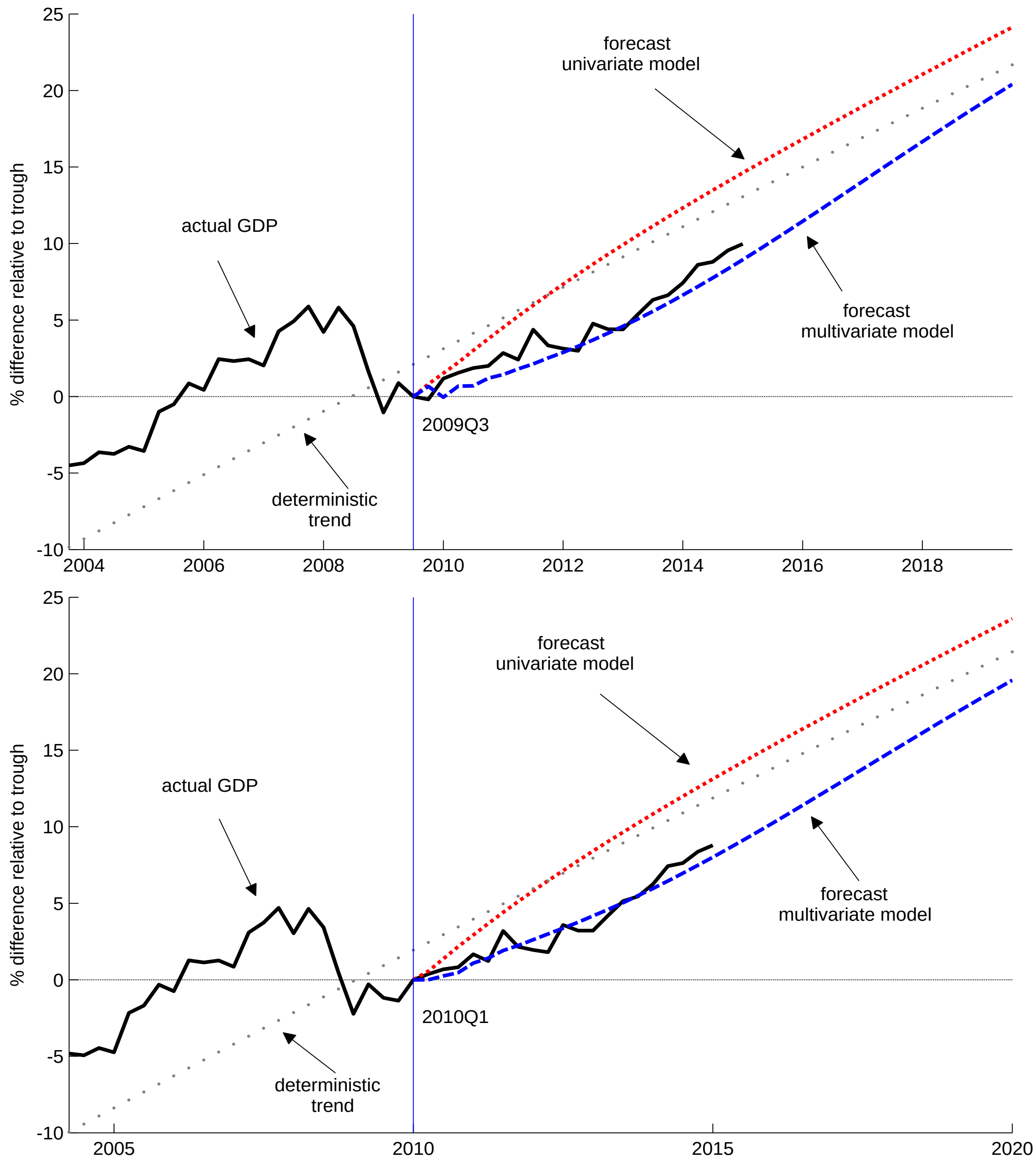
Notes: This figure plots the two forecasted time paths for US GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line.

Figure 9: The start and trough of the great UK recession



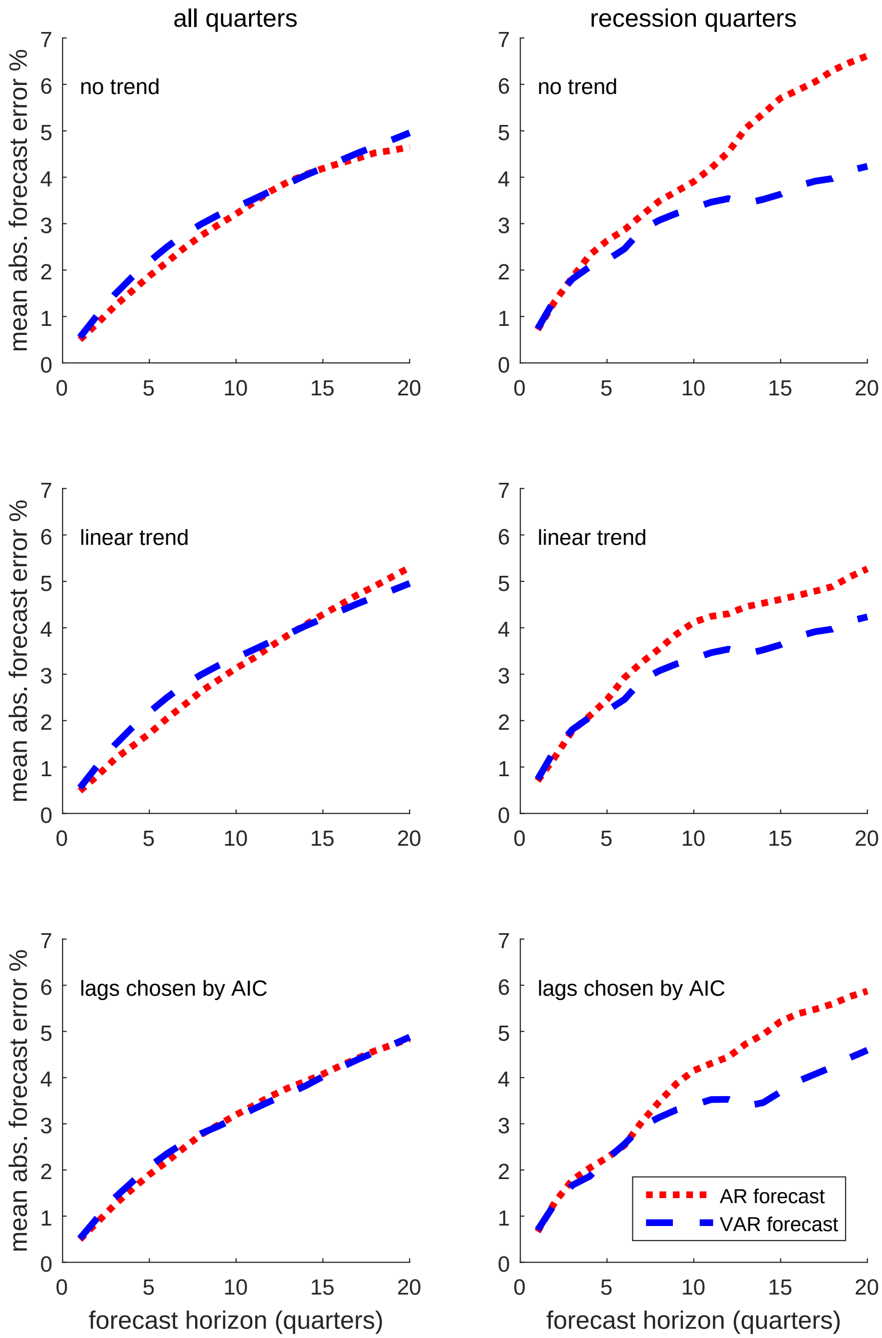
Notes: This figure plots the two forecasted time paths for UK GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line.

Figure 10: The initial recovery of the great UK recession



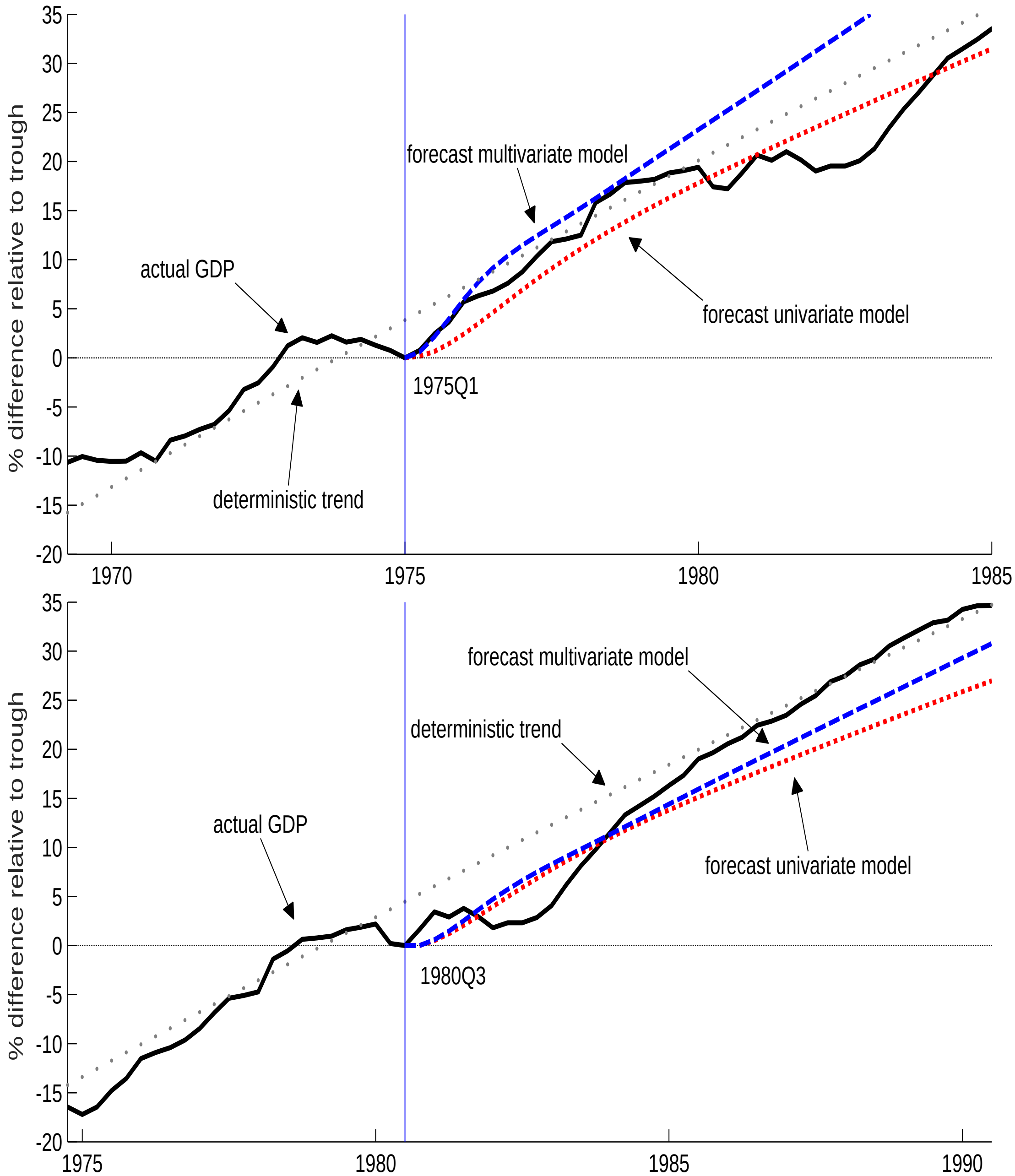
Notes: This figure plots the two forecasted time paths for UK GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line.

Figure 11: Average forecast errors - US - robustness



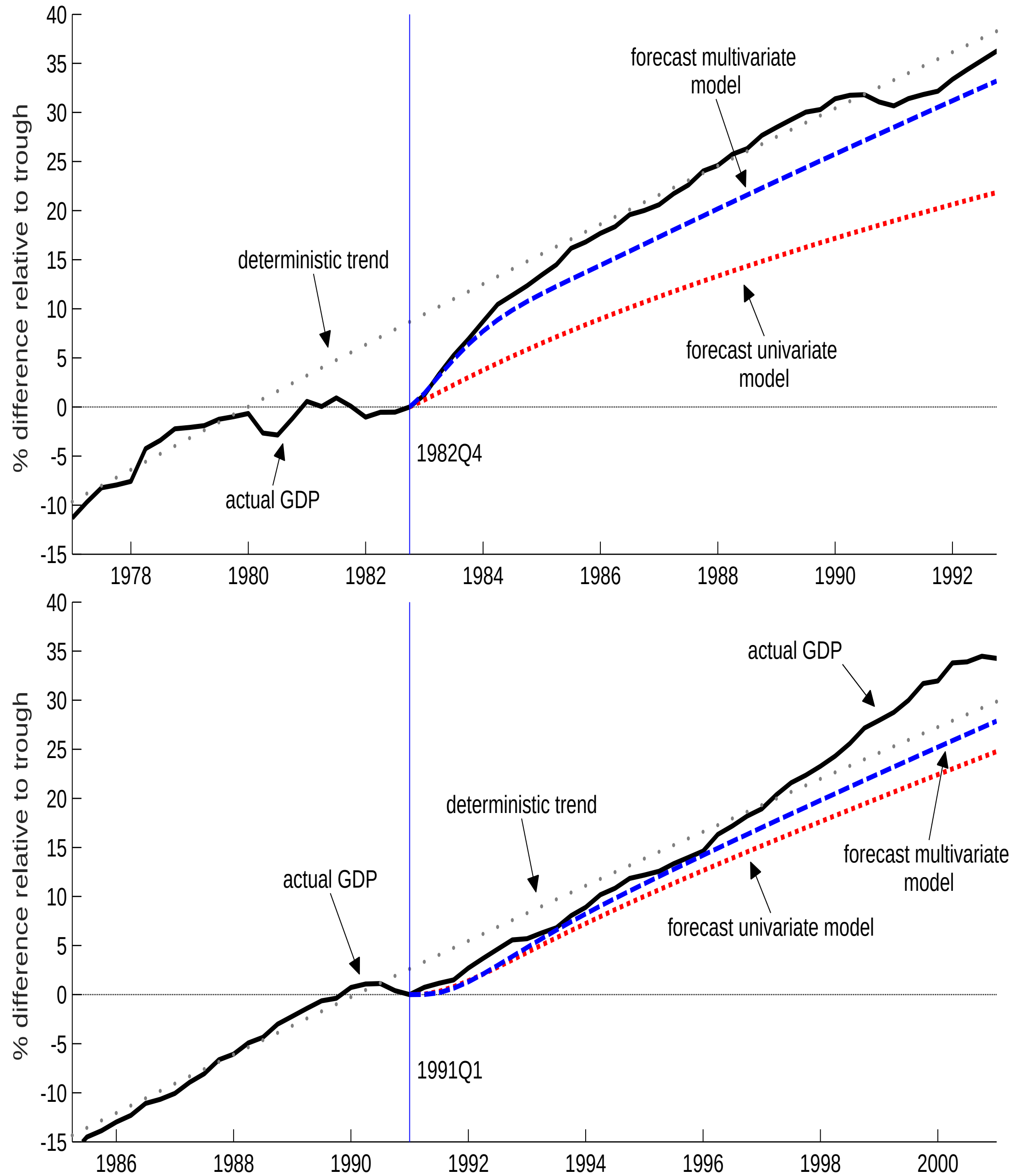
Notes: These graphs plot the average forecast errors of the indicated time-series model. NBER recessions dates are used to identify whether a quarter is a "recession quarter".

Figure 12: The 1973-75 and the 1980 US recession - AIC



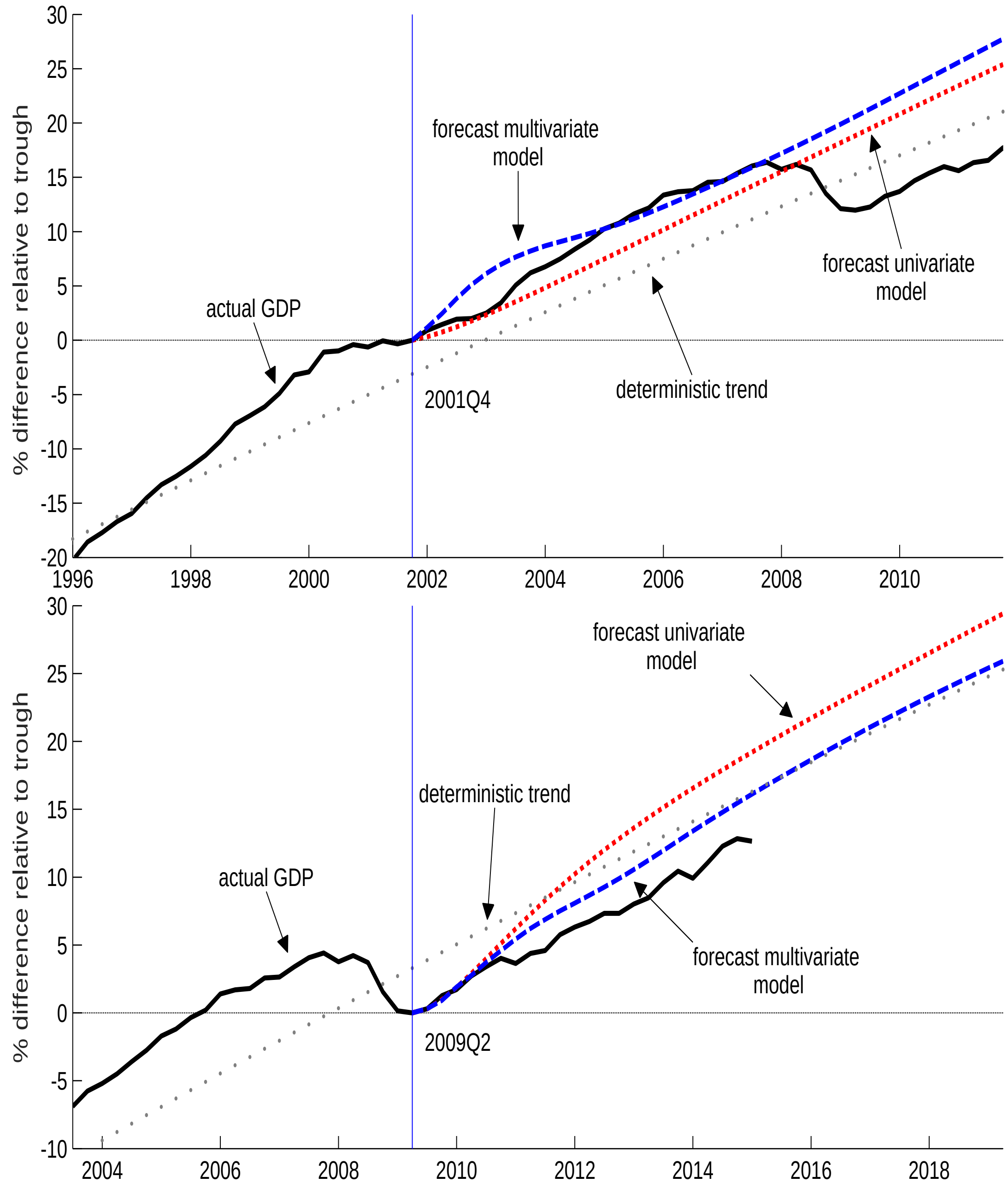
Notes: This figure plots the two forecasted time paths for UK GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line. Number of lags chosen with AIC.

Figure 13: The 1981-82 and the 1990-91 US recession - AIC



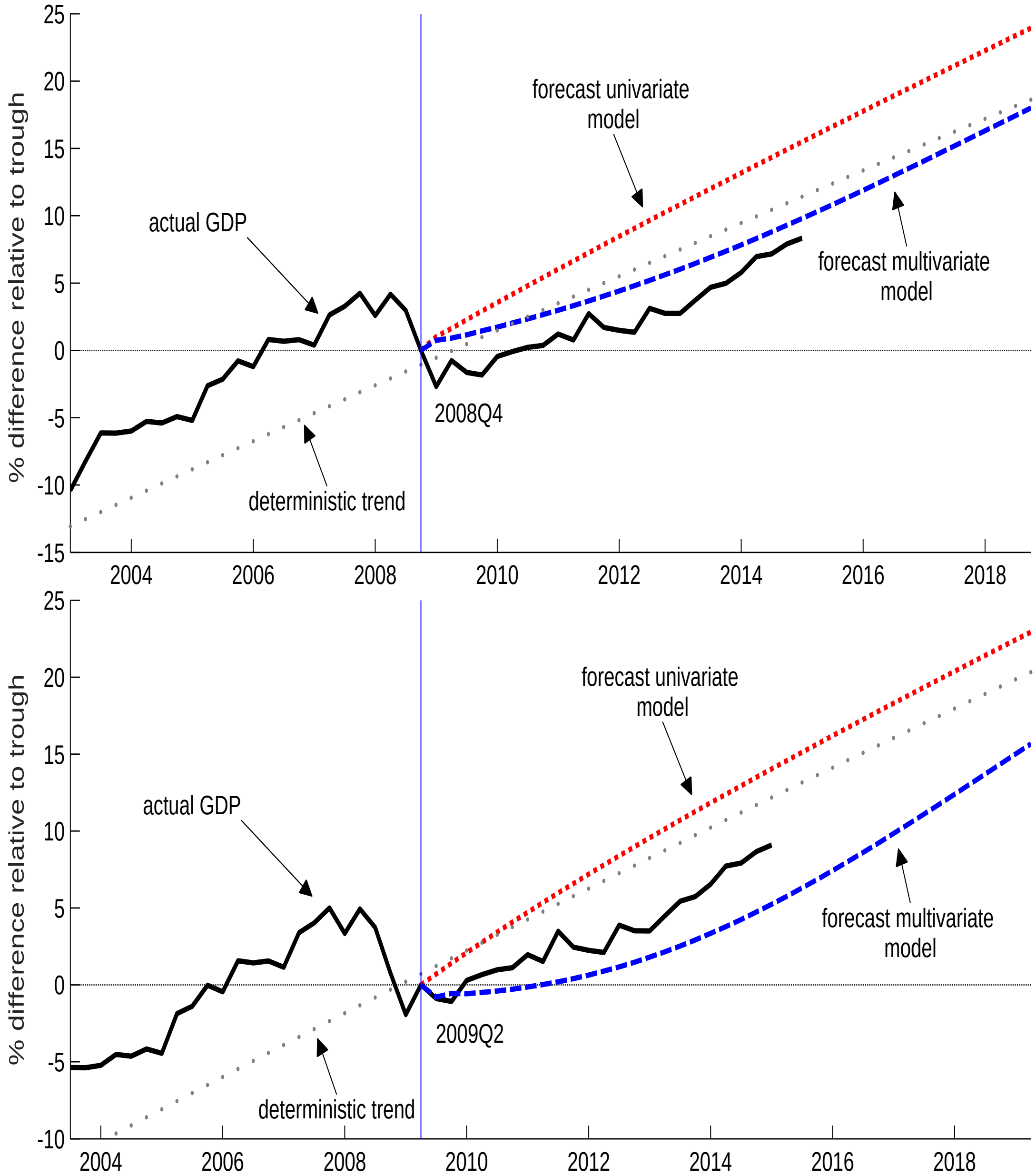
Notes: This figure plots the two forecasted time paths for US GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line. Number of lags chosen with AIC.

Figure 14: The 2001 and great US recession - AIC



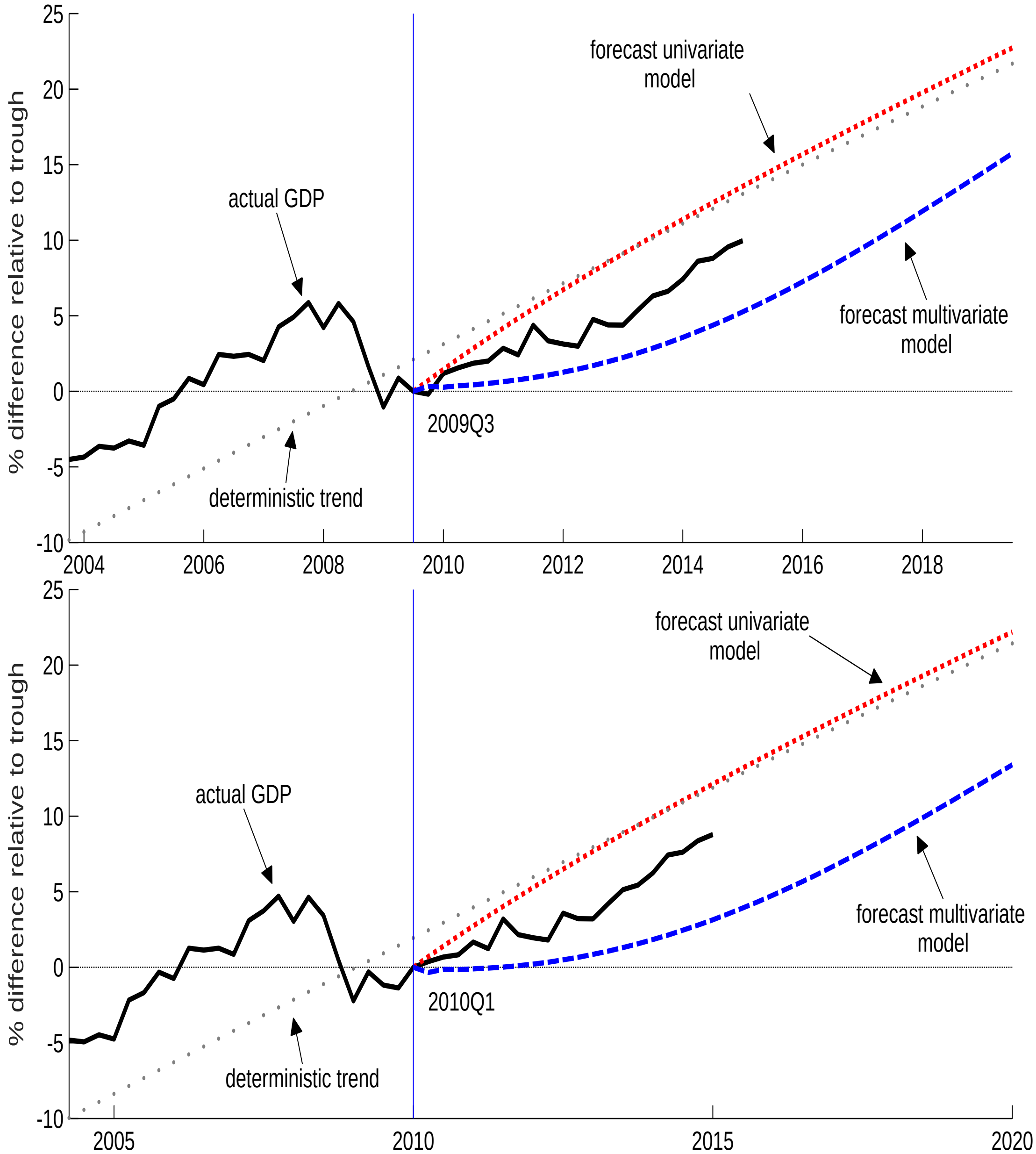
Notes: This figure plots the two forecasted time paths for US GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line. Number of lags chosen with AIC.

Figure 15: The start and trough of the great UK recession - AIC



Notes: This figure plots the two forecasted time paths for UK GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line. Number of lags chosen with AIC.

Figure 16: The initial recovery of the great UK recession - AIC



Notes: This figure plots the two forecasted time paths for UK GDP together with the realized values and a deterministic time trend. All four variables are relative to the value of GDP at the forecasting date, which is indicated by the vertical line. Number of lags chosen with AIC.