Macroeconomic Dynamics (2023), **27**, pp. 141–170
doi:10.1017/S1365100521000341

ARTICLE

The economic effects of government spending: using expectations data to control for information

Matthew Hall[1] and Aditi Thapar[2,*]

[1]Boston Consulting Group and [2]University of Michigan, Ann Arbor, MI 48104, USA
*Corresponding author. Email: aditi@adititthapar.com

Abstract

We present a new methodology that uses professional forecasts to estimate the effects of fiscal policy. We use short-term forecasts to better identify exogenous shocks to government spending by controlling for anticipatory information already in the public domain. We use longer-term forecasts to net out expectations from the future path of other variables, which improves accuracy and efficiency by focusing on more precise measures of the impact of shocks. We show that this improves the statistical fit relative to both local projection methods and vector autoregression-based analyses that do not control for the entire future path of expectations.

Keywords: Fiscal policy, multipliers, impulse response, forecasts

JEL Classfications: E62, E30, H30

1. Introduction

Since the beginning of the Great Recession, and especially after short-term interest rates hit the zero lower bound in 2008, there has been a resurgence of research about the effectiveness of fiscal policy. One of the challenges in estimating the effects of government spending shocks is that fiscal policy takes a long time to implement. Bills are drafted and debated, then they must pass both the House and the Senate before they can be signed into law by the President. All this must take place before any actual spending occurs. For this reason, many changes in fiscal spending are anticipated. In this paper, we present a new methodology that uses the professional forecasts to control for anticipated policy. Our approach allows us to control for anticipated effects very efficiently and therefore generates impulse response functions that are much more statistically significant than standard alternatives.

The macroeconomic effects of government spending are highly contested on both theoretical and empirical grounds. Calibrated structural models tend to deliver results that depend on the set of assumptions on which they are built.[1] Models differ in the size of the impact of government spending on output and in the sign of the impact on consumption and wages. Traditional Keynesian models predict a larger impact on output and a positive impact on consumption, hours worked, and real wages. In contrast, neoclassical models imply that consumption and real wages fall in response to positive government spending shocks. Due to the decline in consumption, neoclassical models typically produce a smaller response of output than do Keynesian models.

Empirical work also delivers conflicting results. Estimates based on vector autoregressions (VARs) largely support the Keynesian view and tend to deliver relatively large effects of

© Cambridge University Press 2021. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

government spending on GDP. VARs, however, have been criticized for not adequately controlling for anticipated information. The information set of the VAR is typically limited to only a handful of contemporaneous variables. A far cry from the actual information available to agents. Failure to control for anticipated changes in government spending will imply that the government spending shocks identified in the VAR are not truly exogenous shocks, but rather a function of omitted variables.

In order to overcome these difficulties and identify government spending shocks that are truly exogenous, another branch of the literature considers shocks to military spending. Military spending arguably responds to foreign threats, so that changes in military spending are largely independent of current macroeconomic conditions. These shocks are often identified via narrative approaches that involve reading newspaper and magazine articles to find announcements of changes in military spending. In general, but not always, structural VARs tend to show larger multipliers, while the narrative-based analyses find evidence for more modest ones.[2] Ramey (2011*b*) attributes the differences in the results to VARs being unable to adequately control for anticipated future policy actions.

In this paper, we use professional forecasts to control for anticipated changes in fiscal policy. Using forecasts has two advantages. As already stated, short-term forecasts help to better identify exogenous shocks to government spending by controlling for anticipatory information already in the public domain. Longer-term forecasts are useful too. Longer-term forecasts help to control for confounding information and lead to more efficient estimates. We demonstrate the gains of our method relative to a standard structural VAR and a VAR that is augmented with expectations data. We also show the efficiency gains relative to standard local projection techniques that are gaining in popularity in the literature.

In the next section, Section 2, we discuss some recent papers that are relevant to our analysis. We then present the details of our approach in Section 3 and compare it to the VAR and local projection methods. We show that all three methodologies construct impulse response functions by projecting forecast errors at various horizons on identified shocks, which themselves are derived from one-step ahead forecast errors. The primary difference between the three approaches is the source of the forecasts that underlay the construction of the forecast errors. Our approach, which we call the professional forecaster model, uses professional forecasts to construct the forecast errors. The other approaches employ internally generated linear forecasts. The identification of shocks from forecast errors is essentially the same in all three approaches.

This characterization of the three models places the focus on the forecasts. The model with the best forecasts will perform best. Professional forecasters have an advantage in that they can make flexible use of all available information. Forecasters can alter their model as the economy changes. They can average the results of several models. They can use judgment. In contrast, VARs and local projection methods fit a single linear model to the entire data set. Where these methods have an advantage is that they can be estimated over the entire data set, including future observations and revised data that are unavailable to professional forecasters. We show that in spite of these advantages, using professional forecasts leads to greater statistical power in the estimates of the impact of government spending on economic activity.

Section 4 presents results using all three methodologies. For the professional forecast model, we construct forecast errors using the Federal Reserve's Greenbook forecasts. Regardless of the methodology, GDP, consumption, government spending, and real wages all rise in response to a rise in government spending. In all cases, prices fall. The results for GDP, consumption, and wages are stronger and have much greater statistical power (lower p-values) using professional forecasts than either the VAR or the local projection method. Beyond statistical significance, there are two principle differences between the professional forecaster model and the VAR. The first is that the professional forecaster model predicts that residential construction rises 1 year after a fiscal stimulus and that this effect is statistically significant at the 10% level. The VAR, on the other hand, predicts a large but statistically insignificant decline in residential construction. The second

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

is that the VAR predicts a smaller rise in government spending than do the estimates based on the Greenbook forecasts.

The decline in prices following an expansionary fiscal shock is counterintuitive but appears to be a common finding in the literature. Canova and Pappa (2007) and Mountford and Uhlig (2009) find similar results. Recently, D'Alessandro *et al.* (1993) and Jørgensen and Ravn (2020) have constructed models with endogenous skill accumulation or technological change to explain this phenomenon. These features effectively make the aggregate supply curve downward sloping, so that a fiscal stimulus increases productivity and reduces prices.

Section 5 contains various extensions and robustness checks. First, we consider multipliers rather than impulse responses. We calculate multipliers in two ways, cumulating the impulse responses as in Mountford and Uhlig (2009), and estimating the multipliers directly using using professional forecasts as controls as in Ramey and Zubairy (2018). In both cases, we find multipliers that rise from between one half and one on impact and rising to just under two after six quarters. These estimates are in line with the literature. Second, rather than constructing forecast errors, we allow the professional forecasts to enter the estimation as a dependent variable. This has little effect on the analysis. Third, we consider alternative professional forecasts. We consider the Survey of Professional Forecasters (SPF), the Blue Chip Economic Indicators, and the University of Michigan's Research Seminar in Quantitative Economics (RSQE) forecasts. The results derived from the Blue Chip look very similar to the results from the Greenbook. The RSQE forecasts imply a weaker response of output and consumption and much more crowding out of business investment. The SPF produce results that tend to lie between the other forecasts. These differences appear to be due to the differences in the implied government spending shocks rather than differences in the forecasts of the dependent variables. Finally, we apply our methodology to military spending. We use the Greenbook to construct unanticipated innovations in military spending and use these to instrument for government spending. The results are very similar to our baseline results. One of the criticisms of using innovations in defense spending to estimate the effects of fiscal policy is that military spending might not be representative of government spending in general. The similarity between our results for military spending and for government spending suggest that this might not be a problem.

Section 6 concludes.

2. Related Literature

VARs have been used extensively to estimate the effects of government spending since Blanchard and Perotti (2002).[3] The main advantage of VARs is their ability to capture complex patterns in the data with very few structural assumptions. The main disadvantage is that, given the small set of variables, the system may not entirely control for all of the information available to agents. The identified shocks may, therefore, not really be shocks at all but functions of omitted variables. For example, Ramey (2011*b*) finds that "shocks" to defense and nondefense spending are anticipated by professional forecasters a few quarters prior to the actual changes in spending.

The literature has approached the informational deficiencies of VARs in a number of ways. The narrative-based approach focuses on events, such as wars, that are likely to be exogenous to current economic conditions.[4] These studies tend to use local projection methods. The main advantage of this approach is the exogeneity of the shocks. The main disadvantage is that there tend to be few military spending shocks and that these shocks are associated with a certain type of government expenditure which may or may not be representative of overall spending.[5]

Another way to control for missing information in a VAR is to include expectations among the regressors. Both Ramey (2011*b*) and Auerbach and Gorodnichenko (2012) augment traditional VARs with one-period ahead forecasts of government spending. This improves the identification of fiscal shocks by netting out the expected change in fiscal policy.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

All of these approaches project some measure of future economic activity on some measure of innovations in fiscal policy. Our approach is to orthogonalize both the innovations in fiscal policy and the measure of future economic activity with respect to the expectations of professional forecasters. Netting expectations from innovations in fiscal policy improves the identification of fiscal shocks, as in the literature that augments a VAR with expectations. Netting expectations from the future path of other variables improves accuracy and efficiency by focusing on a more precise measure of the impact of the shock. We show that this improves the statistical fit of the model relative to both local projection methods and VAR-based analyses that do not control for the entire future path of expectations.

Two papers that are closely related to ours are Dominguez and Shapiro (2013) and Blanchard and Perotti (2002). Both to some extent control for the path of expectations. Dominguez and Shapiro (2013) use revisions of forecasts of GDP along with narrative evidence from news articles to analyze the effect of news on the forecasted path of recovery from the Great Recession. We differ in that we consider forecast errors rather than forecast revisions; we project forecast errors on identified shocks rather than news; and we consider a longer sample period than the Great Recession.[6]

Blanchard and Perotti (2002) regress 2-year cumulative output forecast errors on forecasts of fiscal consolidation in 2010 for 26 European countries to determine whether forecasters are systematically biased in their estimate of the fiscal multiplier. Their results suggest that actual fiscal multipliers were larger than forecasters assumed during the Great Recession, but there is no evidence of fiscal multipliers being over- or underestimated in the precrisis period from 1997 to 2008. In this paper, we consider the US economy rather than a panel of European countries, we estimate the effects of government spending rather than the structural fiscal balance, and we assume a longer sample than the Great Recession.

Thapar (2008) also uses professional forecasts and futures markets to estimate the impact of monetary policy. Her focus is on the price puzzle: the observation that the price level often rises in response to tight monetary policy. She finds that controlling for the information set of the Federal Reserve does not solve the price puzzle.

3. Econometric Framework

Despite a substantial literature studying the effects of fiscal policy on the economy, there are still two outstanding issues. The first issue involves obtaining exogenous changes to fiscal policy, which can be used to estimate the effects of policy changes on the economy. Obtaining these exogenous shocks has been a major focus of the literature. A second issue is the inefficiency and lack of statistical significance at conventional levels of estimates produced via VAR and the local projection approaches. This issue has not received much emphasis in the literature. As Ramey (2016) states, "*Because the Jordà method for calculating impulse response functions imposes fewer restrictions, the estimates are often less precisely estimated and are sometimes erratic.*"

We postulate that the lack of statistical significance in the estimates of the effects of changes in government spending arise from the inability to adequately control for the information set of agents. VAR-based studies are constrained to include only a small set of variables. Local projection estimations have slightly more flexibility in including a larger number of variables, since the degrees of freedom grow one for one with each included variable. However, even there it is difficult to include a large number of variables in the estimation, given data limitations.

We begin this section by discussing our econometric framework, which uses short- and long-term professional forecast data.[7] We then compare our methodology to the VAR and local projection approaches. We conclude by discussing the advantages and disadvantages of using the professional forecasts.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

3.1 Professional Forecasts Approach

Consider an $n \times 1$ vector of economic variables, X. All variables are in logs, so that differences reflect growth rates. Let X_t denote the vector at date t and let X_i denote the ith element of this vector. One of these elements is real government spending, G. For an arbitrary variable or vector V_t, let $v_{t,t+s}$ denote the growth rate in V between periods t and $t+s$. For example, $g_{t,t+5}$ is the growth rate in government spending between period t and period $t+5$.

The first step in estimating the effect of fiscal policy requires one to obtain unanticipated innovations in the policy variable. We use professional forecasts such as the Federal Reserve's Greenbook forecasts to calculate the one-quarter ahead forecast error on government spending FEG_t^{pf} as:

$$FEG_t^{pf} = g_{t-1,t} - g_{t-1,t}^{pf}. \tag{1}$$

Here, $g_{t-1,t}^{pf} = \mathbb{E}[g_{t-1,t}|\mathbb{I}_{t-1}^{pf}]$ is the professional forecast of the growth rate in government spending: the expectation of $g_{t-1,t}$ given the information set available to forecasters at time $t-1$, $\mathbb{I}_{t-1}^{pf}$.

As is well established by now, using professional forecasts does not by itself guarantee the exogeneity of these innovations. We have nothing new to say about this issue and our methodology is flexible with respect to the identification technique employed. In this paper, we follow Blanchard and Perotti (2002). We assume that government spending does not respond contemporaneously to other macroeconomic variables in the system and that government spending decisions are made prior to tax decisions, so that tax innovations also do not affect spending decisions contemporaneously. These assumptions imply that the true structural shocks to government spending, $shock_t^{pf}$, are exactly equal to the one-step ahead forecast errors:

$$shock_t^{pf} = FEG_t^{pf}. \tag{2}$$

The second step in estimating the effect of changes in government spending involves estimating the response of other variables of interest to a government spending shock. To do this, we first construct the vector of longer-term forecast errors for the vector X_t as:

$$u_{t-1,t+s}^{pf} = x_{t-1,t+s} - x_{t-1,t+s}^{pf}, \quad \text{for } s = 0, 1, 2, ..., H. \tag{3}$$

Here, $x_{t-1,t+s}^{pf} = \mathbb{E}[x_{t-1,t+s}|\mathbb{I}_{t-1}^{pf}]$ is the professional forecast of the growth rate in the vector X between $t-1$ and $t+s$. These longer-term forecast errors give the difference between how the economy actually evolved and how it was expected to evolve in the absence of any shocks. To estimate the effects of a change in government spending, we project these long-run forecast errors, $u_{t-1,t+s}^{pf}$, onto our structural shocks, $shock_t^{pf}$. We estimate

$$u_{t-1,t+s}^{pf} = c + \beta^s shock_t^{pf} + \theta_{t+s} \quad \text{for } s = 0, 1, 2, ..., H. \tag{4}$$

We estimate (4) separately via least squares for each of the $H+1$ horizons that forecasts exist and for each variable X_i in X.[8] Note that the error term θ_{t+s} includes the effects of fiscal shocks between periods $t+1$ to $t+s$ and non-fiscal shocks from period t to $t+s$. We therefore use a heteroskedasticity and autocorrelation consistent covariance matrix to estimate our standard errors. The elements of the vector β^s and $\beta^{x_i,s}$ are the coefficients of interest. $\beta^{x_i,s}$ represents the effect of a 1% point unanticipated increase in government spending, $shock_t^{pf}$, today on the growth rate in X_i between $t-1$ and $t+s$. Plotting the $\beta^{x_i,s}$ as a function of s gives the impulse response function for the variable X_i.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

3.2 Other Approaches for Estimating Impulse Response Functions

We compare our approach to two other leading methods used in the empirical literature on the estimation of the effects of fiscal policy. Section 3.2.1 compares our methodology to a structural VAR, and Section 3.2.2 compares our methodology to the local projection approach of Jorda (2005).

3.2.1 Vector autoregression and expectations-augmented VAR

Consider again the $n \times 1$ vector of economic variables, X_t. Recall that all variables are in logs. A structural VAR with p lags models the evolution of X_t as a linear function of its own lags and a vector of exogenous shocks e_t^{VAR}:

$$AX_t = \Gamma + B(L)X_{t-1} + e_t^{VAR}. \tag{5}$$

Here, A is an $n \times n$ matrix, Γ is an $n \times 1$ vector, $B(L)$ is a lag polynomial of order p, and e_t^{VAR} is an $n \times 1$ vector of structural disturbances. The reduced-form VAR associated with (5) is defined as:

$$X_t = A^{-1}\Gamma + A^{-1}B(L)X_{t-1} + A^{-1}e_t^{VAR}.$$

We can represent the vector of reduced-form errors, $u_t^{VAR} = A^{-1}e_t^{VAR}$, as the one-step ahead forecast errors generated by the linear VAR model:

$$u_t^{VAR} = X_t - \mathbb{E}[X_t|X_{t-1}, X_{t-2}, ..., X_{t-p}]. \tag{6}$$

Note that if the implied VAR forecasted growth rate $\mathbb{E}[g_{t,t+s}|X_{t-1}, X_{t-2}, ..., X_{t-p}]$ for government spending is equal to the professional forecast of government spending growth, $g_{t-1,t}^{pf}$, then the two methodologies produce the same reduced-form errors for government spending. In that case, the element of u_t^{VAR} associated with government spending is equal to FEG_t^{pf}.

The structural errors in the VAR are related to the reduced-form errors by the following relation:

$$e_t^{VAR} = Au_t^{VAR}. \tag{7}$$

Identifying assumptions are made on the matrix A. For example, a standard assumption in the VAR literature is to identify structural errors by assuming that A is lower triangular and that the structural shocks are independent. For example, Blanchard and Perotti (2002) assume that A is lower triangle and order G_t first. In this case, the element of e_t^{VAR} associated with government spending, call it $shock_t^{VAR}$, is equal to the element of u_t^{VAR} associated with government spending. If the element of u_t^{VAR} associated with government spending is equal to FEG_t^{pf}, then this identification scheme implies $shock_t^{VAR} = shock_t^{pf}$. The identification step is therefore the same in the VAR and the professional forecaster approach.

Given the structural shocks, we can obtain the impulse response functions. Note that we can represent the linear VAR with p lags as an MA (∞) process as given by:

$$X_t = A^{-1}\Gamma + \sum_{i=0}^{\infty} D_i e_{t-i}^{VAR},$$

where, for each i, D_i is an $n \times n$ matrix of coefficients. The s-step ahead forecast error from the VAR can be represented as:

$$X_{t+s} - \mathbb{E}[X_{t+s}|X_{t-1}, X_{t-2}, ..., X_{t-p}] = A^{-1}\Gamma + \sum_{i=0}^{s-1} D_i e_{t+s-i}^{VAR}. \tag{8}$$

Since the e_{t+s}^{VAR} are assumed to be independent, D_i can be interpreted as the effect of e_t^{VAR} on X_{t+i}. The column of D_i associated with the government spending shock, therefore, traces out

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

the impulse response functions. The regression in equation (8) is very similar to the regression in equation (4). Both project forecast errors on shocks. The difference lies in where the forecasts come from. The VAR uses an internally generated forecast. If $\mathbb{E}[x_{t,t+s}|X_{t-1}, X_{t-2}, ..., X_{t-p}] = \mathbb{E}[x_{t,t+s}|\mathbb{I}^{pf}_{t-1}]$ and the identification scheme is the same, then the shocks and the forecast errors are independent of the methodology used, whether professional forecasts or VAR. In this case, the two methodologies yield exactly the same results. $D^{x_i,g}_s$, the element of the matrix D_s associated with the dependent variable X_i and the shock to government spending, is equal to $\beta^{x_i,s}$.

One of the main criticisms of the structural VAR approach is that the VAR does not account for the fact that many fiscal policy changes are announced well before they are enacted. For example, Ramey (2011*b*) finds that changes in defense and nondefense spending are anticipated by private agents a few quarters prior to the actual changes in spending, implying that VAR shocks are not true innovations. To control for announcement effects, several authors augment VARs with expectational variables. A common practice is to add one-step ahead professional forecasts of government spending, $g^{pf}_{t-1,t}$, or the forecast error, FEG^{pf}_t, to the list of variables in the VAR and to order this additional variable first when identifying the structural shocks from the reduced-form errors. For example, $\hat{X}_t$ is taken to be $\{g^{pf}_{t-1,t}, X_t\}$ or $\{FEG^{pf}_{t-1}, X_t\}$. Note that in either case if the professional forecasts are equal to the standard VAR forecasts without the expectations included, $\mathbb{E}[g_{t-1,t}|X_{t-1}, \ldots X_{t-p}]$, then either $g^{pf}_{t-1,t}$ or FEG^{pf}_{t-1} is redundant and has no effect on the estimation.[9] If the $g^{pf}_{t-1,t}$ and the standard VAR forecasts are different, then this approach will derive fiscal shocks that are orthogonal to both $g^{pf}_{t-1,t}$ and $\{X_{t-1}, \ldots X_{t-p}\}$. There are two differences between this approach and our approach. The first is that this approach allows the $g^{pf}_{t-1,t}$ to enter the VAR flexibly. We investigate how this affects our results in Section 5.2 below. Second, this approach only adds the one-step ahead forecasts, $g^{pf}_{t-1,t}$, to the information set, whereas we control for longer-term forecasts of all variables, $x^{pf}_{t+s|t-1}$, when estimating the effect of government spending on X_{t+s}.

3.2.2 Local projection methods

Local projection methods have become increasing popular in the literature.[10] The local projection method begins with a set of identified structural shocks to government spending, $shock^{lp}_t$. Romer and Romer (2010) and Ramey (2016) use narrative methods to obtain a series of structural shocks, but any identification scheme will do. One then estimates impulse response functions via linear projection:

$$x_{t,t+s} = c + b^s shock^{lp}_t + \sum_{i=1}^{p} d^s_i Z_{t-i} + \eta_{t+s} \text{ for each } s = 0, 2, ..., H, \tag{9}$$

where c is a constant, Z_t is a vector of controls, d^s_i is a vector of coefficients (one vector for each lag i and each horizon s), and η_{t+s} is the error term. The control variables may include lags of the dependent variable. The regression (9) is run separately for each variable X_i in X and each horizon s.

Note that $\mathbb{E}[x_{t,t+s}|Z_{t-1}, \ldots Z_{t-n}] = c + \sum_{j=1}^{p} d^j \cdot Z^j_{t-i}$ so that (9) also projects forecast errors on the structural shocks. The specification of the local projection (9) is essentially the same as the VAR (8).[11] In both cases, the forecast is a linear function of the variables included in the regression.

Let $b^{x_i,s}$ denote the element of the vector b^s associated with the variable X_i. The $b^{x_i,s}$ trace out the responses of X_i to a shock in period t as a function of the horizon s . If model shocks are the same, $shock^{lp}_t = shock^{pf}_t$, and if the forecasts are the same, $\mathbb{E}[x_{t,t+s}|\mathbb{I}^{pf}_{t-1}] = \mathbb{E}[x_{t,t+s}|Z_{t-1}, \ldots Z_{t-n}]$,

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

the impulse response function derived from the local projection will be identical to that using professional forecasts: $b^{x_i,s} = \beta^{x_i,s}$.

One advantage of the local projection method is that it simplifies the inclusion of additional controls, Z^j without the need to estimate a large system of equations. Data and sample restrictions, however, typically restrict the number of control variables one can include.

3.2.3 Advantages and disadvantages of using professional forecasts

All three methodologies project $x_{t,t+s}$ on identified government spending shocks and controls. All three methodologies control for the expected future path of fiscal policy. There are two main differences between the professional forecaster methodology, on the one hand, and the VAR and the local projection method, on the other: the information set and the forecasting model.

The information set in the VAR is limited to X_{t-1} and its lags, whereas the information set in the local projection is limited to Z_{t-1} and its lags. There are good reasons to believe that the information set of professional forecasters is much larger than either of these. Presumably professional forecasters incorporate all relevant and available information into their forecasts. This includes not only variables beyond those included in the VAR or local projection, but also current information on the future evolution of these variables. Such information might be particularly important in the context of fiscal policy, as fiscal policy is subject to long and variable lags in implementation. Forecasters presumably are aware of legislation that is in progress, the results of elections that shift the balance of power in Congress and foreshadow future legislation, and news articles reporting the intentions of policy-makers. The Federal Reserve Board, in particular, has privileged access to information regarding the policy-making process. For all of these reasons, professional forecasts likely incorporate a far greater amount of information than do other methodologies. For this reason, professional forecasts are likely to produce more accurate and more efficient estimates of the impact of government spending shocks.

There is one important sense, however, in which professional forecasts might reflect less information than is included in the VAR. This is because the forecast $\mathbb{E}[x_{t,t+s}|\mathbb{I}^{pf}_{t-1}]$ is typically formulated and announced before the end of period $t-1$. Many economic variables that are included in standard VARs are not calculated and released until the period has ended. This is particularly true when data are revised over time, as is the case with real GDP where revisions can be as large as the period-to-period changes. Much of the information incorporated in the VAR in period t is therefore not processed and made available to agents in the economy, including professional forecasters, until much later, although the Federal Reserve likely has access to such information before others. For this reason, it remains possible that VARs and local projection methods use more information than do professional forecasts. It is therefore an empirical question which methodology is more efficient.

Even if professional forecasters use the same information set as the VAR, there is the question of how this information is processed to generate expectations. In principle, the professional forecaster has an advantage here. The VAR and the local projection are limited to using the same linear projection over the entire sample. Professional forecasters can use general methodologies which incorporate nonlinearities, model averaging, and judgment. They can change their model as the world changes. Professional forecasters should therefore be better able to react to changes in the economy, trend breaks, and regime shifts.[12] For this reason also, professional forecasts are likely to produce more accurate and more efficient estimates of the impact of government spending shocks.

That being said, it is likely that neither methodology utilizes the true model of the world. Misspecification is a problem for all methodologies. Downward-biased forecasts will lead to overestimation of positive shocks and responses and underestimation of negative shocks and responses with obvious implications for the estimation of multipliers. One way to think of this is that there are likely two sets of forecasts relevant for economic fluctuations. There is the forecast based

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

on the physical state of the world and the forecast in people's heads. The latter is the relevant forecast for intertemporal economic decisions such as investment, consumption, and saving. The VAR focuses on the forecast based on the physical world. It assumes that its linear forecast is the true forecast based on the state of the world and that this forecast is a sufficient statistic for the forecast in people's heads. The professional forecast methodology focuses on what people are thinking. It assumes that the forecast in people's heads is a sufficient statistic for the forecast based on the state of the world. Both are interesting cases worthy of study and we likely learn something from each. It is also certain that neither is exactly correct. It is an important topic for future research to incorporate both types of forecast into a single model and to investigate the implications of heterogeneous and inaccurate private sector forecasts for inference in imperfect statistical models.

4. Empirical Application

In this section, we compare the professional forecaster methodology to the VAR and local projection method. We first discuss the exact specification used in each approach, the data, and then the results.

4.1 Specifications

For the professional forecast model, we consider the vector $X_t^{pf} = \{G_t, Y_t, C_t, IBF_t, IRC_t, W_t, P_t\}$ where G_t represents real government spending, Y_t represents real GDP, C_t is real consumption expenditures, IBF_t represents real business fixed investment, IRC_t represents real investment in residential construction, W_t represents real wages, and P_t represents the GDP deflator. For each variable, we use data on expected growth rates to construct the forecast errors (3) and the shocks, $shock_t^{pf}$. We then estimate equation (4) separately via least squares for each of the $H+1$ horizons that forecasts exist and for each variable X. We construct Newey–West confidence intervals to correct for serial correlation. The time period for the estimation is quarterly 1978–2011. 1978 is the date at which the Federal Reserve began providing forecasts beyond four quarters.

For the standard VAR, $X_t^{VAR} = \{G_t, T_t, Y_t, C_t, IBF_t, IRC_t, W_t, P_t\}$ where in addition to the variables already introduced T_t represents tax receipts.[13] To identify shocks to government spending, government spending is ordered first in the VAR.[14] To increase power, when estimating the impulse response functions for G_t and Y_t, we consider a VAR in (8) that includes only $\{G_t, T_t, Y_t\}$ and when estimating the impulse response for any other variable X_i we consider a VAR that includes $\{G_t, T_t, Y_t, X_{it}\}$. The lag length is 4. The time period for the estimation of the VAR is 1968 to 2011.[15]

The expectation augmented VAR or E-VAR that we estimate is similar to the VAR with the addition of the lagged one-step ahead forecast error $g_{t-1,t}^{pf}$. When identifying the shocks, $g_{t-1,t}^{pf}$ is ordered first and G_t is ordered second, so that the identified shock $shock_t^{EVAR}$ is orthogonal to $g_{t-1,t}^{pf}$. As is the case with the VAR, when estimating the impulse response for G_t and Y_t we consider a VAR that includes only $\{g_{t-1,t}^{pf}, G_t, T_t, Y_t\}$ and when estimating the impulse response for any other variable X_i we consider a VAR that includes $\{g_{t-1,t}^{pf}, G_t, T_t, Y_t, X_{it}\}$. The E-VARs are estimated over the 1968–2011 sample period, based on availability of one-quarter ahead Greenbook forecasts.

The local projection model considers the same set of variables as our professional forecast model, $X^{lp} = \{G_t, Y_t, C_t, IBF_t, IRC_t, W_t, P_t\}$. To facilitate comparison, the shocks in the local projection model are taken to be same as those in the professional forecaster model, $shock_t^{lp} = shock_t^{pf}$. The regression (9) is run separately for each variable X_i and each horizon s. Following Ramey

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

(2016), the list of controls Z include a linear time trend, as well as two lags each of the shock, government spending, tax revenues, GDP, as well as the variable X_i (if X_i is any variable other than government spending or GDP). Like the professional forecast model, the local projection model is also estimated over 1978 to 2011. We present Newey–West confidence intervals to correct for serial correlation.

4.2 Data

To implement our empirical methodology, we need quarterly forecasts of government spending, GDP, and any additional variables that we want impulse response functions for. In this paper, we focus on the Federal Reserve Board's Greenbook forecasts. We use a number of data sources to obtain our baseline data set. We describe the data series, their sources, and their construction briefly here. Appendix C provides further detail.

To guide the Federal Open Markets Committee (FOMC) make informed decisions about monetary policy, the research staff at the Federal Reserve Board produces forecasts of the US economy. These forecasts are available to the FOMC 6 days prior to every scheduled meeting. Colloquially, the forecast report was called the "Greenbook" forecast report until 2010, at which point the name was changed to the "Tealbook" report. The Federal Reserve releases the Greenbook reports to the public, but with a 5-year lag. This constrains the end of our sample to be 2011. We begin our analysis in 1978 when the maximum forecast horizon was increased to seven quarters.

Some of the forecasts of the variables of interest are available in levels, others in growth rates, and a select few are in both levels and growth rates. Still others, like GDP and the GDP deflator, which had been available in both levels and growth rates, began to be published only in growth rate terms beginning in 2005. This implies that whether we decide to use data in levels or growth rates, we would need to extrapolate some data to ensure that we have a consistent time series for each variable. In this paper, we decided to use growth rates for all variables of interest.

We are interested in estimating the cumulative effects of government spending on variables of interest at different horizons. Since we are working with growth rates rather than levels, for the purpose of estimating impulse response functions, we calculate cumulative growth rates for each variable of interest X as, $x_{t,t+s} = \frac{X_{t+s}}{X_t} - 1$, for each date t and horizon $s = 1, 2, ..., H$ for which we have forecast data available. As discussed above, data limitations imply that $H = 7$.

In this paper, we use current available data and not real-time data for constructing all forecast errors. Appendix C.5 justifies our motivation for this choice. Briefly, since we are working with cumulative growth rates, estimating forecast errors from real-time data is not necessarily desirable. Cumulative growth rates are subject to revisions over progressively longer time horizons. In this case it is not clear, even in theory, which real-time data concept is appropriate to compare to a forecast. Any choice would be arbitrary and potentially vary according to the length of time over which the growth rate is calculated. In addition, since we are working with growth rates rather than levels, definition changes are less likely to have large effects on the growth rate of the variables of interest.

The Greenbook forecasts estimate the growth of future real government expenditures on consumption and investment, including federal, state, and local expenditures. We use this set of estimates as our measure of government spending. All other variables have standard definitions. All realized data have been downloaded from the Federal Reserve Bank of St. Louis' FRED data portal.

In Section 5.4, for the purpose of comparison, we also use quarterly forecasts from the University of Michigan's RSQE, the Blue Chip Economic Indicators, and the SPF conducted by the Federal Reserve Bank of Philadelphia. These alternative forecast data sets are described in Section C of the Appendix.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

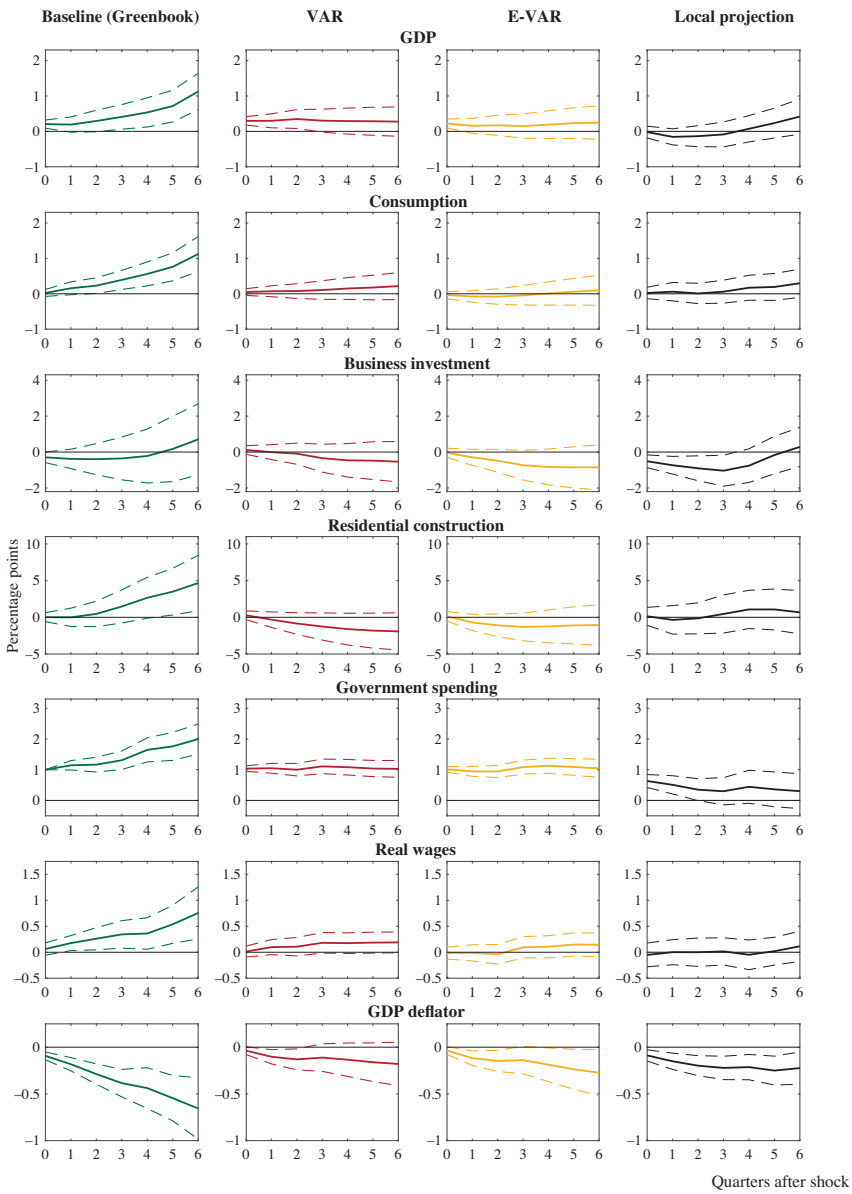

Figure 1. Baseline results: Impulse responses of a shock to government spending.
Each row in the figure graphically presents the impulse response to a 1% point increase in government spending on the variables listed in the row header. Each column represents the results from the different methodologies presented in Section 3. The baseline results in column 1 and the local projection are estimated over the 1978–2011 sample. The VAR and the E-VAR are estimated over the 1968–2011 sample. In all the figures, solid lines (—) represent the impulse response and dashed lines (– –) represent the 90% confidence intervals.

4.3 Estimating the Effects of Government Spending

Table 2 in the Appendix presents coefficient estimates and p-values for all variables estimated in this section. Figure 1 presents this information visually. The columns correspond to different estimation methods. The rows of the figure show the impulse responses of different dependent variables to government spending shocks. Solid lines in the figures represent these impulse responses. All of the impulse responses reported in this paper depict the effect on the variable of

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

interest in period $t + s$ of a 1% point shock to real government spending in period t. Dashed lines illustrate 90% confidence intervals.

The leftmost column presents our baseline results, estimated using equation (4) individually for each variable and each horizon s using Greenbook forecast errors over the 1978–2011 sample period. We find that real GDP, consumption, and wage all rise in response to an increase in government spending. For GDP, our results imply that a 1% point rise in government spending leads to a 0.2% point increase in GDP on impact. The effect rises to 1.1% points over seven quarters. The impact on GDP is significant at the 5% level at most horizons. The effect of the shock on consumption mirrors the effect on GDP. It is small and statistically insignificant for the first two quarters, indicating lags in the transmission from fiscal policy to the private sector. The effect on consumption increases to 1.1% point six quarters after the shock. The impact on consumption is significant at the 2% level after two quarters.

There is weak evidence that increases in government spending crowd out business fixed investment. We find that a 1% boost to government spending leads to a 0.3% point fall in business investment on impact. This result, however, is not statistically significant at standard levels (p = 0.104). On the other hand, there is somewhat stronger evidence that residential construction is boosted in response to a fiscal stimulus. The results are economically and statistically significant one and a half years after a shock to fiscal policy.

A particularly interesting finding is that prices fall in response to a shock to fiscal policy, regardless of the estimation methodology. We discuss this finding in further detail in Section 5.1.

The second column of Figure 1 shows the results from a standard VAR. Overall, the results from the VAR are qualitatively similar to our baseline results: GDP, consumption, and real wages all rise in response to a rise in government spending. Prices fall. The responses of these variables, however, are muted relative to the professional forecaster model and the estimates are not as statistically significant. The p-values for most variables are high. The effect on GDP is only significant for the first three quarters. The effects on consumption and real wages are not significant even at the 10% level. There are two qualitative differences between our baseline estimates and the VAR estimates. The first is that the professional forecaster model finds that residential construction begins to rise 6 months after a fiscal stimulus and that this effect is statistically significant at the 10% level after 1 year. The VAR, on the other hand, predicts a large but statistically insignificant decline in residential construction. The second is that the VAR predicts a smaller rise in government spending than do the estimates based on the Greenbook forecasts. In the professional forecaster model, government spending continues to rise following the shock. In the VAR, government spending is flat.

The third column of Figure 1 depicts the results for the E-VAR. These results are qualitatively similar to our baseline results and the VAR. The response of GDP and government spending rise and prices fall. The impact on GDP beyond the first quarter is no longer statistically significant even at the 10% level. Consumption and the real wage show little response but are imprecisely estimated. Business investment falls slightly more than in the VAR but is imprecisely estimated.

The last column represents the impulse responses based on the local projection method. In contrast to our baseline results and the VAR-based model, the local projection impulse response functions imply that GDP, consumption, and real wages are all largely unchanged in response to a shock in government spending. Additionally, the local projection approach implies a large and statistically significant decline in business fixed investment, which neither of the other approaches predict.

Figure 2 compares the point estimates and the statistical significance of the estimates. The left column depicts the point estimates for all four methodologies in a single graph for each of our seven variables. In the right column, we plot the p-values. The p-values are truncated and values greater than 0.25 are plotted as exactly equal to 0.25. This allows us to focus our attention on the statistically significant estimates. The solid green line represents our baseline results; the dashed red line is the VAR; the dash-dot yellow line is the E-VAR; and the black dotted line is the

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

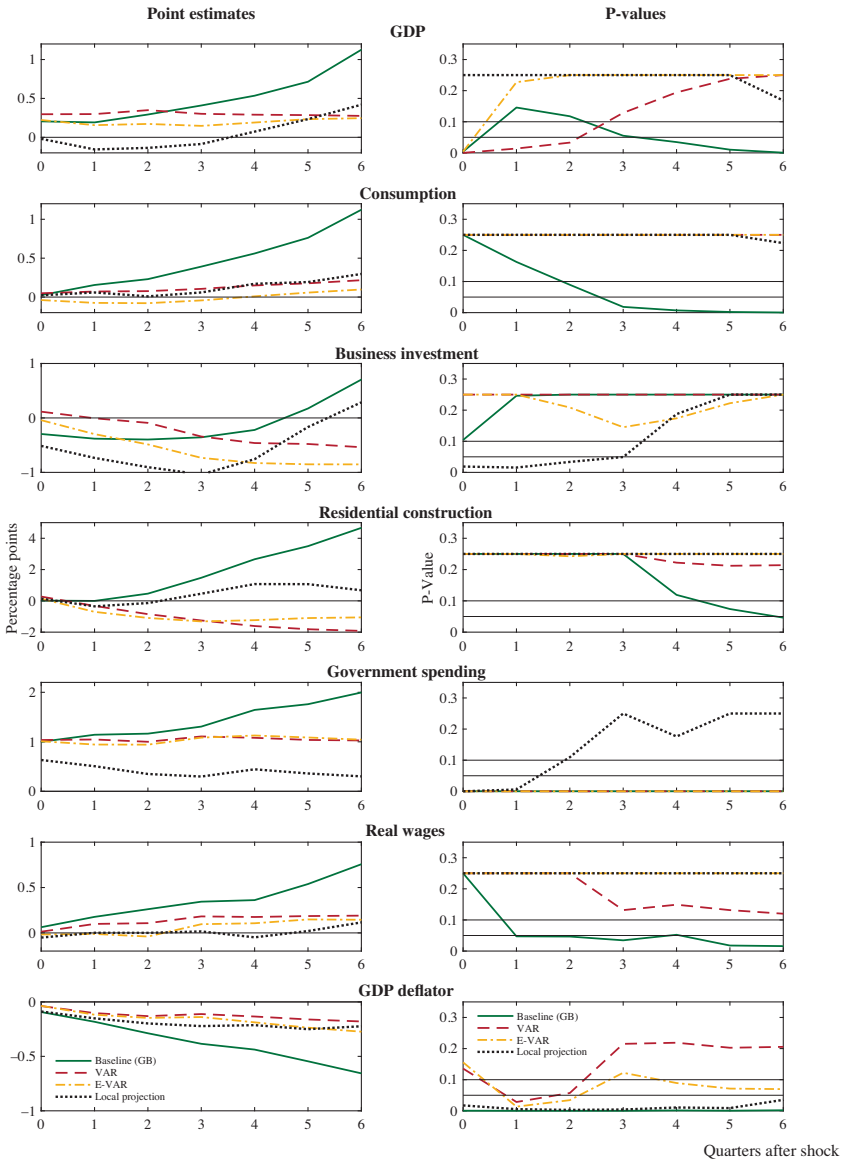

Figure 2. Baseline results: Highlighting the efficiency gains from using forecast data.
This figure is an alternative representation of the results in Figure 1 and Table 2. Each row in the figure graphically presents the results for a 1% point increase in government spending on the variables listed in the row header. The left column presents the point estimates from the different methodologies presented in Section 3. The right column presents the *p*-values associated with each point estimate in the left column. To visually focus on results that are statistically significant, all *p*-values greater than 0.25 are winsorized to 0.25. The horizontal thin black lines (–) represent significance at the 5% and 10% levels. The baseline (GB) and local projection results are estimated over the 1978–2011 sample. The VAR and the E-VAR are estimated over the 1968–2011 sample.

local projection estimation. The thin black horizontal lines in the right column, at 0.05 and 0.1, represent the conventional significance levels of 5% and 10%, respectively.

The main point we want to emphasize is the noticeably greater statistical significance of the estimates for the professional forecaster model. The professional forecaster model tends to produce stronger and more statistically significant estimated responses at all horizons, but especially

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

longer horizons. The exceptions are that the effect on GDP is more precisely estimated at short horizons with the VAR, and all other methodologies except the local projection method produce noisy estimates of business fixed investment. Longer horizons are where we would expect to see efficiency gains from using long-term forecasts. Professional forecasts appear to summarize relevant information about the future in a parsimonious manner and improve the statistical power of the estimates.

5. Discussion and Extensions

5.1 The "Fiscal Price Puzzle"

An interesting finding we highlight in this paper is the presence of a "Fiscal Price Puzzle." No matter which methodology we use, we find that prices decline following an expansionary government spending shock. Conventional wisdom would suggest the opposite. An increase in government spending should be inflationary. Similar counterintuitive responses of prices to fiscal policy have been reported by Canova and Pappa (2007) and Mountford and Uhlig (2009), but this phenomenon has been largely ignored in the literature.[16]

Recently, a handful of authors have constructed theoretical models consistent with this result. In essence, these papers argue that the aggregate supply curve may be downward sloping. D'Alessandro *et al.* (1993) add skill accumulation to a standard New Keynesian model. In their model, increases in government spending lead to increases in hours. The increase in hours increases the incentive to accumulate skills. The subsequent increase in total factor productivity lowers marginal costs and reduces inflation. Jørgensen and Ravn (2020) introduce variable technological utilization into a standard New Keynesian model. In their model, firms respond to an increase in government spending (and the resulting increase in aggregate demand) by adopting new technologies that raise productivity. Higher productivity leads to the decline in prices and inflation.

5.2 The Rational Expectations Assumption

An implicit assumption underlying our approach is that private sector forecasts are rational. In Appendix Table 3, we compare the accuracy of the Greenbook forecasts with the forecasts implicit in the VAR. The mean and mean squared errors are very similar across the two specifications. Only for real wages does the VAR significantly outperform the Greenbook forecasters. Overall, there is no evidence of bias of misspecification in the Greenbook forecasts.

Another way to evaluate the rational expectations assumption is to allow the Greenbook forecasts to enter flexibly in the estimation. Our baseline specification assumes that expectations feed one to one into future realizations. We can relax this assumption by moving the forecast to the right-hand side of the regression.

$$x_{t,t+s} = c + \gamma^s shock_t^{pf} + \mu^s \mathbb{E}[x_{t,t+s} | \mathbb{I}_{t-1}^{pf}] + \theta_{t+s} \text{ for } s = 0, 1, ..., H. \tag{10}$$

Here, γ^s is the coefficient of interest.[17] Regression (10) allows expectations to enter flexibly, as they do in the E-VAR.

Figure 3 presents the results. The first column reproduces our baseline estimation from Figure 1. The second column presents results from estimating equation (10). The results are almost identical. We can also test whether $\mu^{x_i,s}$ is equal to 1. For real wages, business investment and the first four quarters of government spending, we can reject that $\mu^{x_i,s}$ is equal to 1 at the 5% level (it is generally smaller than 1). For all other variables, and GDP in particular, we cannot reject the hypothesis that $\mu^{x_i,s} = 1$, at the 5% level for any horizon s.[18] Therefore, while in several cases, the restriction that $\mu^{x_i s}$ is equal to 1 is rejected, freeing up this coefficient has very little effect on estimates of the impulse responses.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

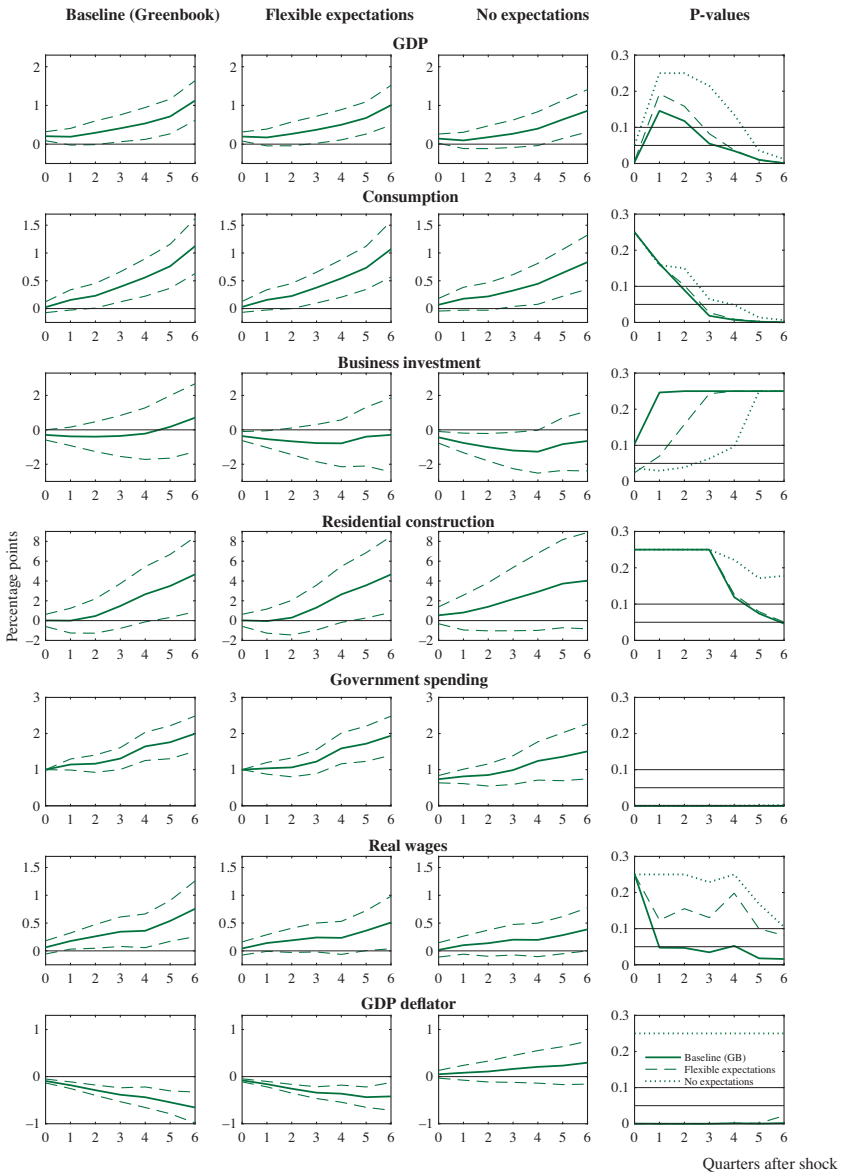

Figure 3. Rational expectations assumption.
Each row in the figure graphically presents the impulse response of a 1% point increase in government spending on the variables listed in the row header. Each column represents the results from different estimations. The first column presents our baseline results using Greenbook forecasts, estimated using equation (4). The second column presents results using Greenbook forecasts, estimated using equation (10). The last column presents results estimated by a modified version of equation (10), in which we omit expectation of the dependent variable from the estimation. We estimate, $x_{t,t+s} = c + \gamma^s shock_t^{pf} + \theta_{t+s}$ for $s = 0, 1, ..., H$. The estimation sample is 1978–2011. In all the figures, solid lines (—) represent the impulse response and dashed lines (– –) represent the 90% confidence intervals.

The third column of Figure 3 shows the results of estimating equation (10) without including $\mathbb{E}[x_{t,t+s}|\mathbb{I}_{t-1}^{pf}]$ at all. These estimates are similar to our baseline estimates but are not as statistically significant. The last column compares the p-values for the three regressions. For GDP, consumption, and real wages, the baseline regression produces results with the most statistical significance.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

For example, estimates are only significant on impact and after 1 year when expectations are not included. This illustrates the importance of controlling for long-run expectations.

5.3 Multipliers

The impulse response functions estimated in equation (4) in Section 3.1 estimate the elasticity of output with respect to government spending at various horizons. The multiplier differs in that it is not an elasticity but a derivative. It relates the dollar change in GDP to the dollar change in government spending.

There are many approaches to constructing multipliers. We consider two. The first, which is followed by Mountford and Uhlig (2009), is to calculate multipliers from our elasticity estimates. We consider the following:

$$\phi^s = \frac{\sum_{i=0}^{s} \beta^{y,i}}{\sum_{i=0}^{s} \beta^{g,i}} \times \frac{Y}{G}. \tag{11}$$

Here, ϕ^s is the multiplier over a horizon s. The numerator cumulates the estimated elasticities of GDP in reponse to a government spending shock. The denominator cumulates the estimated elasticities of government spending. The ratio gives the percent change in GDP between $t-1$ and $t+s$ per percent change in government spending. Y/G is the average ratio of GDP to government spending over the sample period. Multiplying by Y/G transforms these percent changes into level differences.[19]

An alternate approach, which is followed by Ramey and Zubairy (2018), is to estimate multipliers directly. Adapting this approach to our methodology, we estimate

$$\sum_{i=0}^{s} \left(\frac{Y_{t+i} - \mathbb{E}\left[Y_{t+i} \,\middle|\, \mathbb{I}_{t-1}^{pf} \right]}{\tilde{Y}_{t-1}} \right) = c_s + \phi^s \frac{\sum_{i=0}^{s} \left(G_{t+i} - \mathbb{E}\left[G_{t+i} \,\middle|\, \mathbb{I}_{t-1}^{pf} \right] \right)}{\tilde{Y}_{t-1}} + \theta_{t+s} \tag{12}$$

for $s = 1, 2, ..., H$ and $\tilde{Y}$ is either actual or potential GDP. ϕ^s is the multiplier. Here, the left-hand side is the cumulative change in the level of GDP relative to expectations. We regress this on the cumulative change in the level of government spending. We divide both sides by $\tilde{Y}_{t-1}$ in order to make the data stationary.

Table 1 presents multipliers using both methodologies. The first column shows multipliers constructed as in equation (11) using our baseline elasticities estimated using equation (4) and the Greenbook forecasts. The second and third columns show estimated multipliers. In the second column $\tilde{Y}$ is actual GDP, whereas in the third column $\tilde{Y}$ is potential GDP. The two methods yield very similar results. The multipliers are larger for the Mountford and Uhlig (2009) methodology at short horizons, and approximately equal for all methodologies at longer horizons. The multipliers using the Mountford and Uhlig (2009) methodology, begin slightly below 1 at short horizons and rise to slightly below 1.6 at longer horizons. The estimated multipliers begin at 0.6 at short horizons and rise to slightly above 1.6 at longer horizons. The advantage of estimating the multiplier is that it is easy to construct confidence intervals. Both estimated multipliers are statistically significant at conventional levels, especially at longer horizons, and statistically indistinguishable from the constructed multipliers.

The last column presents multipliers estimated using (11) for our baseline three-variable VAR that includes the log of real government spending, real government receipts, and real GDP. The VAR is estimated over the 1968–2011 period, which matches the VAR sample in our baseline estimation results in Figure 1. The multipliers implied by the VAR remain around 1.2–1.3 from impact to one and a half years later. The VAR multipliers are slightly higher than our baseline estimates at short horizons and slightly lower than our baseline estimates at longer horizons.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

Table 1. Fiscal multipliers

Horizon	Implied multiplier Baseline (Greenbook)	Direct estimation ($\tilde{Y}_{t-1}$ = Lagged GDP)	Direct estimation ($\tilde{Y}_{t-1}$ = Lagged Potential GDP)	Implied Multiplier (VAR)
0	0.9	0.6 (0.3) [0.07]	0.6 (0.3) [0.06]	1.2
1	0.8	0.7 (0.3) [0.06]	0.7 (0.3) [0.05]	1.2
2	0.9	0.7 (0.3) [0.04]	0.7 (0.3) [0.03]	1.3
3	1.1	0.8 (0.4) [0.02]	0.8 (0.4) [0.02]	1.3
4	1.2	1.1 (0.4) [0.00]	1.0 (0.4) [0.00]	1.2
5	1.3	1.4 (0.4) [0.00]	1.4 (0.4) [0.00]	1.2
6	1.6	1.6 (0.5) [0.00]	1.6 (0.5) [0.00]	1.2

Notes: Standard errors are in () and *p*-values are in []. The first three columns of the table show output multipliers for a $1 increase in government spending using Greenbook forecasts. The first column uses equation (11), while the second and third columns estimate the multiplier directly using equation (12). The last column shows the multipliers for a VAR implied by equation (11). The VAR is estimated over the 1968–2011 sample period and includes four lags each of the log of real government spending, real government receipts, and real GDP.

In the empirical fiscal policy literature, estimates of the government spending multiplier range anywhere from 0.5 to 2.0. The multipliers implied by the Greenbook forecasts are at the low end of this range at short horizons and at the high end of this range at longer horizons.

5.4 Other Private Forecasts

So far we have focused on the impact of government spending shocks as implied by the Federal Reserve's Greenbook forecasts. A natural question to ask is whether other professional forecasts would deliver similar results. In this section, we apply our estimation strategy to three alternative data sets with professional forecasts: The SPF, The Blue Chip Economic Indicators, and University of Michigan's RSQE forecasts. Here, we briefly describe each data set. For a detailed description, see Appendix Section C.

The SPF is produced by the Federal Reserve Bank of Philadelphia. It is a survey of professional forecasters. The number of forecasters has varied over time and ranges from 26 at the beginning of our sample to 43 at the end. It presents consensus forecasts for all of our variables, except real wages, for horizons up to 1 year. The Blue Chip Economic Indicators are considered to be the leading private sector forecasts. They contain consensus forecasts of 30–40 professional forecasters for up to seven quarters ahead. Unfortunately, the Blue Chip forecasts are only available for GDP, consumption, and the GDP deflator. They do not include forecasts of any fiscal variables. We therefore need to find another source for the government spending shocks. For comparability to earlier results, we use the shocks derived from the Federal Reserve's Greenbook forecasts when we present results for the Blue Chip forecasts. Like the Greenbook forecasts, the University of Michigan's RSQE forecasts are based on a large-scale structural model of the US economy and rely heavily on judgment in producing the final estimates. Tables 3 and 4 compare the performance of the different forecasts at different horizons. The mean errors and the root mean squared errors (RMSEs) are similar across forecasters.

Figure 4 compares the the results using each forecast data set.[20] The sample period is 1983–2011. This is the period for which all data sets overlap. The left column plots the point estimates,

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

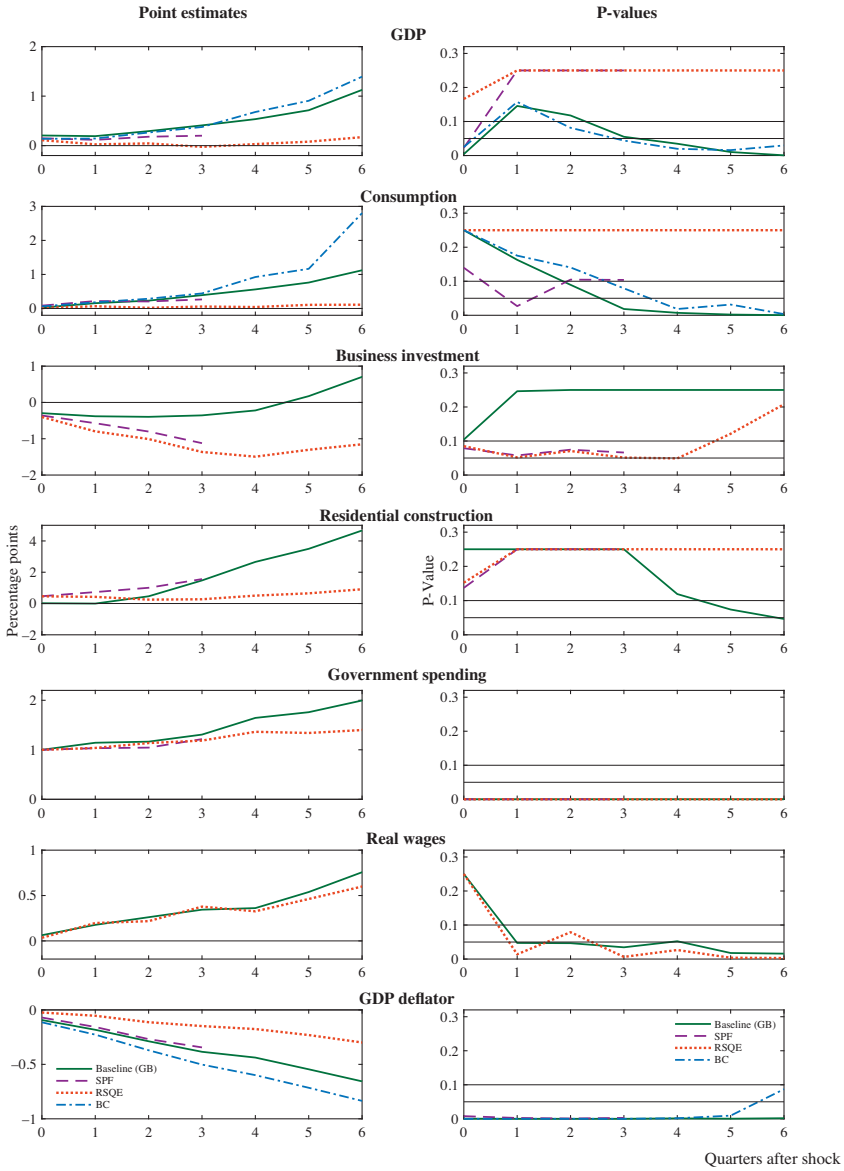

Figure 4. Comparing different forecasters—Greenbook, SPF, RSQE, and blue chip forecasts
Each row in the figure graphically presents the results for a 1% point increase in government spending on the variables listed in the row header. The left column presents the point estimates using different forecast data sets, estimated using our baseline estimation methodology presented in equation (4) in Section 3. The right column presents the *p*-values associated with each point estimate in the left column. To visually focus on results that are statistically significant, all *p*-values greater than 0.25 are winsorized to 0.25. The horizontal thin black lines (–) represent significance at the 5% and 10% levels. All results are estimated over the 1983–2011 sample, which is when data are available for all the different forecast data sets. The Blue Chip results use the Greenbook shocks, since the Blue Chip forecasts do not collect data on government spending.

while the right column plots the *p*-values. In the right column, *p*-values greater than 0.25 are plotted as exactly equal to 0.25. This allows us to focus attention on the statistically significant estimates. The solid green line represents our baseline results; the dotted red line is the RSQE; the dashed purple line is SPF; and the dash-dot blue line is the Blue Chip. The thin black horizontal

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

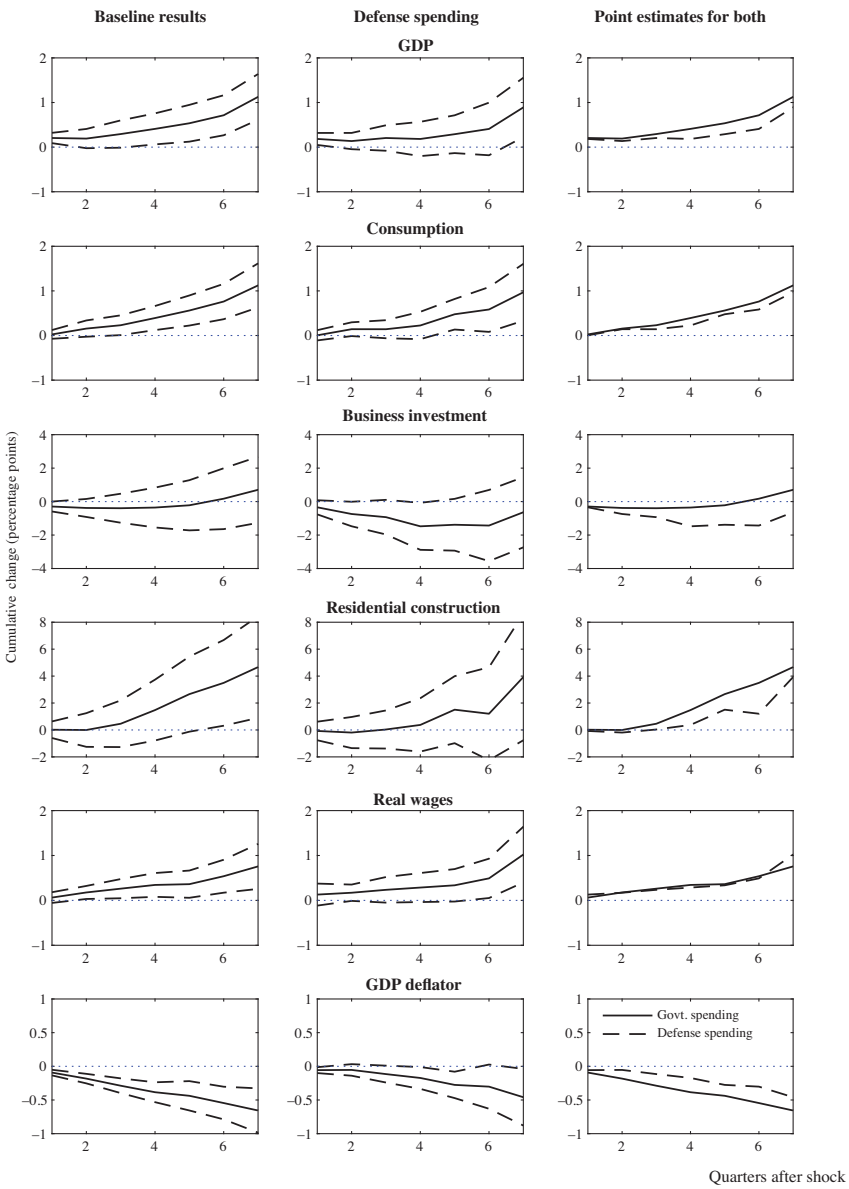

Figure 5. Baseline results versus defense spending shocks (1983–2011).
Each row in the figure graphically presents the impulse response to a 1% point increase in government spending on the variables listed in the row headers. The first column presents our baseline estimation using equation (4) estimated using Greenbook forecasts of total government spending. The second column presents results estimated using equation (4), with one difference. We use one-step ahead Greenbook forecast errors of defense spending to instrument for the total government spending shock ($shock_t^{pf}$). The equations are estimated over 1983–2011, which is the sample period for which we have forecasts of defense spending. In the first two figures, solid lines (—) represent the impulse response and dashed lines (– –) represent the 90% confidence intervals. The last column plots the point estimates from the first two columns together for each variable.

lines in the right column, at 0.05 and 0.1, represent the conventional significance levels of 5% and 10%, respectively.

Looking at the left column, we see that responses implied by the Greenbook forecasts and the Blue Chip forecasts produce very similar results, whereas the RSQE forecasts imply a weaker

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

response of output and consumption and much more crowding out of business investment.[21] The SPF forecasts produce results that tend to lie between the other forecasts. The different results appear to be mostly due to differences in the implied shocks and not differences in forecast errors of the dependent variable. Figure 6 in the Appendix shows the result of projecting forecast errors for GDP and business investment constructed from different private sector forecasts on the shocks derived from the Greenbook forecasts, as well as projecting Greenbook forecast errors on shocks derived from different private sector forecasts. We see there that the differences in Figure 4 between the RSQE-based estimates, on the one hand, and the Greenbook- and Blue Chip-based estimates, on the other, are mostly due to differences in the shocks. For example, when one projects the RSQE forecast errors for investment on the Greenbook shocks, the impulse response looks similar to projecting the Greenbook forecast errors on the Greenbook shocks. Looking at the right column, we see that all four sets of forecasts yield estimates with similar p-values.

Given that the differences in the results across forecaster data sets appear to be mostly due to the differences in the shocks, a likely candidate is that the Greenbook forecasts contain more short-term information than do the other forecasts aggregate more information.

5.5 Defense Shocks

Much of the literature on the effect of fiscal policy focuses on the impact of changes in defense spending for GDP, as defense spending is arguably orthogonal to current macroeconomic conditions.[22] In this section, we apply our methodology to defense spending and compare the results to our baseline results.

To apply our methodology to defense spending, we first construct one-step ahead errors in the growth rate of defense spending, FED_t^{pf}, using estimates of defense spending from the Greenbook. FED_t^{pf} is the difference between the realized growth in defense spending in period t and the Greenbook forecast from period $t-1$. We then use FED_t^{pf} to instrument for $shock_t^{pf}$ in equation (4).[23] Essentially, the new fiscal shock is a rescaling of FED_t^{pf}. The sample period is 1983–2011, since Greenbook forecasts for defense spending are available beginning in 1982.

Figure 5 presents the results. The first column reproduces our baseline estimation. The second column presents estimates using the forecast errors in defense spending as instruments. For ease of comparison, the last column plots the point estimates for both together in one graph. The results from the two methodologies are qualitatively similar. The impact of government spending shocks are slightly muted and less significant using defense spending as an instrument. The contraction in business investment is slightly stronger and more significant using defense spending as an instrument.

One criticism of using innovations in defense spending as an instrument for government spending shocks is that it is not clear that the impact of military spending is representative of the impact of government spending in general. Military spending targets specific industries and specific income groups. The impact of military spending is not felt evenly throughout the nation. That being said, the similarity between our results for military spending and for government spending overall indicate that these differences do not have a large effect on estimates of the impact of fiscal policy.

6. Conclusion

In this paper, we present a new methodology that uses professional forecasts to estimate the effects of fiscal policy on the economy. We use short-term forecasts to estimate a new measure of shocks to government spending. We show how using longer horizon forecasts to control for anticipated information can improve the statistical fit and improve efficiency of the estimates. We

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

also compare our professional forecaster approach to the VAR and local projection methodologies and show that all three approaches effectively project forecast errors of variables onto structural shocks.

Using the professional forecaster approach, we find that there is a statistically significant increase in GDP, consumption, and wages in response to an unanticipated increase in government spending. The inflation rate decreases in response to a fiscal expansion (the Fiscal Price Puzzle), which is a common finding in all three methodologies. The effects on business fixed investment are not statistically significant. Broadly speaking, our results are similar to the VAR-based results.

One of the advantages of our methodology is that it is easily adaptable to estimate state-dependent effects of various policy changes such as oil price shocks, or exchange rate shocks on the economy.

Acknowledgements. We would like to thank Charles Brown, Susan Collins, Kathryn Dominguez, Gabriel Ehrlich, George Fulton, Chris House, John Leahy, and seminar participants for helpful comments. Special thanks to George Mytilineos for outstanding research assistance. Thapar gratefully acknowledges the support of the National Science Foundation for their support under grant SES-1919362 and the Michigan Institution for Teaching and Research (MITRE) at the University of Michigan. All errors are our own.

Notes

1 For example, Cogan et al. (2007) employ a DSGE model with sticky prices and wages and estimate a fiscal multiplier that is significantly less than unity. In contrast, Gali et al. (2007) modify a New Keynesian model with a large proportion of rule-of-thumb consumers and find a multiplier of two. Sims and Wolff (2018*b*) consider the output and welfare effects of government spending shocks in a medium-scale DSGE model and find an output multiplier of just over 1.

2 See Blanchard and Perotti (2002), Perotti (2007), and Gordon and Krenn (2010) for examples of VAR approaches yielding a multiplier greater than 1, and Mountford and Uhlig (2009) for an example of a VAR approach with a multiplier below 1. Among narrative-based approaches, including use of war dates, see Ramey and Shapiro (1998), Eichenbaum and Fisher (2005), and Ramey (2011*a*) for estimates of a government spending multiplier less than 1.

3 Other prominent examples of VAR-based analyses of the effects of fiscal policy include Perotti (2007), Mountford and Uhlig (2009), and Auerbach and Gorodnichenko (2012).

4 See, for example, Ramey and Shapiro (1998), Barro and Redlick (2011), Ramey (2011*b*), and Ramey and Zubairy (2018). Favero and Giavazzi (2012) use the Ramey (2011*b*) war dates in a VAR. Nakamura and Steinsson (2014), who use heterogeneity in the response of states in the USA to national military spending buildups to identify shocks to fiscal policy. Ramey (2016) provides an excellent overview of the literature.

5 Canova and Pappa (2007) increase the number of observations by analyzing innovations in military procurement and the stock prices of defense contractors. Nakamura and Steinsson (2014) increase the number of observations by considering US states.

6 Note that under the assumption of rational expectations, the distinction between forecast errors and forecast revisions should not matter.

7 Thapar (2008) uses financial market information to estimate the effects of monetary policy.

8 A limitation of the professional forecaster approach is that it can only be applied to variables and horizons for which forecasts exist.

9 Note that these papers do not impose rational expectations, that is, they do not impose that the expectations of fiscal policy included in the VAR are equal to the expectations of fiscal policy generated by the VAR:

$$g^{pf}_{t-1,t} \neq \mathbb{E}[g_{t-1,t}|X_{t-1}, \ldots X_{t-p}].$$

10 Ramey (2016) applies the Jordà methodology to estimating government spending multipliers, while Romer and Romer (2010) apply it to estimating the effects of tax changes.

11 If the list of variables are the same and the identification scheme is the same, then the local projection method and the VAR are equivalent. Differences between the two methodologies tend to be more a mater of specification and identification than inference. See Plagborg-Møller and Wolf (2020) for a detailed discussion.

12 In current work, Thapar (2019) extends the methodology developed in this paper to study state dependence in the effects of fiscal policy. Also see Sheremirov and Spirovska (2019), Sims and Wolff (2018*b*), and Biolsi (2017).

13 Taxes are usually included in VAR studies of fiscal policy. We did not include taxes in the professional forecaster methodology, because there we only needed to include the dependent variables that we study.

14 Blanchard and Perotti (2002) also investigate ordering government spending after taxes. They find that their results are unaffected by this change. We therefore only present one set of VAR results, with government spending ordered first. We also

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

investigated orthogonalizing $shock_t^{pf}$ with respect to taxes. This had no qualitative or quantitative effect on the results with the professional forecaster methodology.

15 Estimating the VAR over the same sample period as the professional forecaster model yields very noisy results. Incorporating a longer sample allows the VAR time to form more accurate forecasts and places it on a more even footing with the professional forecaster model.

16 See Jørgensen and Ravn (2020) for a summary of the empirical estimates of the effects of government spending shocks on prices.

17 In estimating the equation specification above, one could in principle include additional control variables, or even additional forecasts as controls, to improve the accuracy of the estimates. We do not do so here, since our baseline specification does not include any additional controls.

18 For GDP, $\mu^{x_i,s}$ is significantly different from 1 at the 10% level for two horizons, although for one of these two horizons the p-value is just slightly below 10%.

19 Mountford and Uhlig (2009) calculate "present value" multipliers by discounting both $\beta^{y,i}$ and $\beta^{g,i}$ by the real interest rate over horizon i. This has very little effect on the calculation of ϕ^s.

20 We do not include results for real wages in this figure since neither the SPF nor the Blue Chip include forecasts for wages. Both RSQE and Greenbook predict that real wages increase in response to an increase in government spending.

21 Recall, the Blue Chip is using the Greenbook shocks.

22 See Ramey and Shapiro (1998).

23 We do not report the results of the first-stage of the two-stage-least-squares estimation. In all cases, the first-stage F statistic is over 50, which implies that we can reject the null that defense spending is a weak instrument for total government spending.

References

Auerbach, A. and Y. Gorodnichenko (2012) Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy* 4(2), 1–27.

Bachmann, R. and E. Sims (2012) Confidence and the transmission of government spending shocks. *Journal of Monetary Economics* 59(3), 235–249.

Barro, R. J. and C. J. Redlick (2011) Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics* 126, 51–102.

Biolsi, C. (2017) Nonlinear effects of fiscal policy over the business cycle. *Journal of Economic Dynamics & Control* 78, 54–87.

Blanchard, O. and R. Perotti (2002) An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics* 117(4), 1329–1368.

Blanchard, O. and D. Leigh (2013) Growth forecast errors and fiscal multipliers. *American Economic Review* 103(3), 117–120.

Canova, F. and E. Pappa (2007) Price differentials in monetary unions: The role of fiscal shocks. *Economic Journal* 117, 713–737.

Cogan, J. F., T. J. Cwik, J. B. Taylor, and V. Wieland (2010) New Keynesian versus old Keynesian government spending multipliers. *Journal of Economic Dynamics and Control* 34, 281–295.

Croushore, D. (1993) Introducing: The survey of professional forecasters. *Federal Reserve Bank of Philadelphia. Business Review,* November/December, 3–15.

d'Alessandro, A., G. Fella, and L. Melosi (2019) Fiscal stimulus with learning-by-doing. *International Economic Review* 60(3), 1413–1432.

Dominguez, K. M. E. and M. D. Shapiro (2013) Forecasting the recovery from the great recession: Is this time different? *American Economic Review* 103(3), 147–152.

Eichenbaum, Martin, and Jonas D.M. Fisher (2005) Fiscal Policy in the Aftermath of 9/11. *Journal of Money, Credit, and Banking* 37(1), 1–22.

Favero, C. and F. Giavazzi (2012) Measuring tax multipliers: The narrative method in fiscal VARs. *American Economic Journal: Economic Policy* 4(2), 69–94.

Galí, J., J. D. López-Salido, and J. Vallés (2007) Understanding the effects of government spending on consumption. *Journal of the European Economic Association* 5, 227–270.

Giorno, C., P. Richardson, D. Roseveare, and P. van den Noord (1995) Estimating potential output, output gaps and structural budget balances. *OECD Economics Department Working Papers* No. 152.

Gordon, R. J. and R. Krenn (2010) The end of the great depression 1939–41: Policy contributions and fiscal multipliers. *NBER Working Papers*. No. 16380.

Jordá, O. (2005) Estimation and inference of impulse responses by local projections. *American Economic Review* 95(1), 161–182.

Jørgensen, P. L. and S. H. Ravn (2020) The inflation response to government spending shocks: A fiscal price puzzle? *Working paper*.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

Kilian, L. and J. Kim (2011) How reliable are local projections estimators of impulse responses? *Review of Economics and Statistics* 93(4), 1460–1466.
Kilian, L. and B. Hicks (2013) Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting* 32, 385–394.
Mountford, A. and H. Uhlig (2009) What are the effects of fiscal policy shocks? *Journal of Applied Econometrics* 24, 960–992.
Nakamura, E. and J. Steinsson (2014) Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review* 104(3), 753–792.
Plagborg-Møller, M. and C. Wolf (2020) Local projections and VARs estimate the same impulse responses. *Econometrica*, forthcoming.
Perotti, R. (2011) Expectations and fiscal policy: An empirical investigation. *IGIER Working Paper 429*.
Perotti, R. (2007) In search of the transmission mechanism of fiscal policy. *NBER Macroeconomics Annual* 22, 169–249.
Ramey, V. A. (2016) Macroeconomic shocks and their propagation. *NBER Working Paper Series*, No. 21978.
Ramey, V. A. (2011a) Can government purchases stimulate the economy? *Journal of Economic Literature* 49(3), 673–685.
Ramey, V. A. (2011b) Identifying government spending shocks: It's all in the timing. *Quarterly Journal of Economics* 126(1), 1–50.
Ramey, V. A. and M. D. Shapiro (1998) Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy* 48, 145–194.
Ramey, V. A. and S. Zubairy (2018) Government spending multipliers in good times and in bad: Evidence from 20th century historical data. *Journal of Political Economy* 126(2), 850–901.
Romer, C. D. and D. H. Romer (2010) The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review* 100(3), 763–801.
Sheremirov, V. and S. Spirovska (2019) Fiscal multipliers in advanced and developing countries: Evidence from military spending. *Working paper*.
Sims, E. and J. Wolff (2018a) The state-dependent effects of tax shocks *European Economic Review* 107, 57–85.
Sims, E. and J. Wolff (2018b) The output and welfare effects of government spending shocks over the business cycle. *International Economic Review* 59(3), 1403–1435.
Thapar, A. (2008) Using private forecasts to estimate the effects of monetary policy. *Journal of Monetary Economics* 55, 806–824.
Thapar, A. (2019) The ineffectiveness of fiscal policy in recessions: Size does matter. *Working paper*.

Appendix

Appendix A. Baseline Coefficient Estimates

Table 2 presents the results from Figure 1 in tabular form. Each column presents the estimates for a different outcome variable, as specified in the column headers. The rows are grouped by methodology. The first block presents estimates using Greenbook forecasts and estimated using regressions in equation (4). The second and third blocks presents estimates based on the VAR and E-VAR presented in equation (8). The last block presents results estimated using the local projection approach from equation (9).

For GDP, consumption, real wages, and the GDP deflator, the p-values are much smaller using our baseline methodology. Our methodology implies results that are statistically significant at the 5% level at most horizons. For the local projection methodology, on the other hand, results are typically not statistically significant even at the 20% level. Our results imply that the effect of a government spending shock on investment are small and statistically insignificant, whereas the linear projection method implies large and statistically significant decreases in business investment. For prices, both methods predict a strong and statistically significant decline in inflation after an increase in government spending. By controlling for information in an efficient and parsimonious way, our approach leads to impulse responses that have much greater statistical significance than the standard local projection and VAR methodologies.

Appendix B. Alternative Forecasts: Are the Shocks or Forecasts Driving the Differences?

A natural question is whether the differences in the results between different data sets of professional forecasters are caused by differences in the shocks or differences in the the forecast errors projected onto the shocks to derive the impulse response functions. We explore this question in Figure 6, which plots impulse response functions estimated using different combinations of forecasts errors and government spending shocks. We focus on the response of GDP and business investment as these are were the main differences lie.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

Table 2. Baseline results using greenbook forecasts versus alternative estimation methods

Model	Horizon	GDP		Consumption		Business Investment		Residential Construction		Government Spending		Real Wages		GDP Deflator	
		coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value
Baseline (Greenbook)	0	0.20	**0.00**	0.02	0.68	-0.30	0.10	0.02	0.96	1.00	**0.00**	0.06	0.39	-0.09	**0.00**
	1	0.19	0.15	0.16	0.16	-0.38	0.25	0.00	1.00	1.14	**0.00**	0.18	**0.05**	-0.18	**0.00**
	2	0.29	0.12	0.23	**0.09**	-0.40	0.45	0.46	0.66	1.16	**0.00**	0.26	**0.05**	-0.29	**0.00**
	3	0.41	**0.06**	0.39	**0.02**	-0.36	0.62	1.48	0.28	1.31	**0.00**	0.34	**0.03**	-0.38	**0.00**
	4	0.54	**0.03**	0.56	**0.01**	-0.22	0.81	2.66	0.12	1.64	**0.00**	0.36	**0.05**	-0.44	**0.00**
	5	0.71	**0.01**	0.76	**0.00**	0.17	0.88	3.50	**0.07**	1.76	**0.00**	0.54	**0.02**	-0.55	**0.00**
	6	1.13	**0.00**	1.12	**0.00**	0.70	0.56	4.67	**0.05**	2.00	**0.00**	0.76	**0.02**	-0.66	**0.00**
VAR	0	0.30	**0.00**	0.05	0.40	0.12	0.43	0.27	0.45	1.04	**0.00**	0.01	0.84	-0.04	0.14
	1	0.30	**0.01**	0.07	0.45	-0.01	0.98	-0.32	0.62	1.05	**0.00**	0.10	0.26	-0.10	**0.03**
	2	0.35	**0.03**	0.08	0.55	-0.09	0.80	-0.85	0.35	1.00	**0.00**	0.11	0.32	-0.13	**0.06**
	3	0.30	0.13	0.11	0.51	-0.34	0.47	-1.25	0.27	1.11	**0.00**	0.18	0.13	-0.11	0.22
	4	0.29	0.19	0.15	0.43	-0.46	0.42	-1.61	0.22	1.08	**0.00**	0.18	0.15	-0.13	0.22
	5	0.29	0.24	0.18	0.40	-0.48	0.45	-1.81	0.21	1.04	**0.00**	0.19	0.13	-0.16	0.20
	6	0.28	0.28	0.22	0.35	-0.54	0.43	-1.92	0.21	1.03	**0.00**	0.19	0.12	-0.18	0.21
E-VAR	0	0.22	**0.00**	-0.04	0.53	-0.04	0.79	0.14	0.73	1.01	**0.00**	-0.02	0.80	-0.04	0.16
	1	0.16	0.23	-0.07	0.45	-0.30	0.28	-0.70	0.30	0.94	**0.00**	-0.01	0.90	-0.12	**0.01**
	2	0.17	0.32	-0.08	0.56	-0.49	0.21	-1.09	0.24	0.94	**0.00**	-0.04	0.73	-0.15	**0.03**
	3	0.15	0.48	-0.04	0.80	-0.73	0.15	-1.31	0.26	1.08	**0.00**	0.10	0.44	-0.14	0.12
	4	0.19	0.43	0.01	0.96	-0.83	0.17	-1.23	0.37	1.13	**0.00**	0.11	0.41	-0.19	**0.09**
	5	0.23	0.38	0.06	0.80	-0.85	0.22	-1.10	0.48	1.09	**0.00**	0.15	0.27	-0.24	**0.07**
	6	0.25	0.39	0.10	0.70	-0.85	0.26	-1.06	0.53	1.04	**0.00**	0.14	0.30	-0.27	**0.07**
Local Projection	0	-0.02	0.85	0.02	0.82	-0.51	**0.02**	0.14	0.85	0.63	**0.00**	-0.05	0.71	-0.09	**0.02**
	1	-0.16	0.26	0.06	0.71	-0.73	**0.02**	-0.35	0.77	0.51	**0.01**	0.00	0.99	-0.15	**0.01**
	2	-0.14	0.45	0.01	0.96	-0.90	**0.03**	-0.14	0.91	0.35	0.11	0.00	1.00	-0.20	**0.00**
	3	-0.09	0.69	0.06	0.76	-1.04	**0.05**	0.46	0.77	0.30	0.27	0.02	0.92	-0.22	**0.00**
	4	0.07	0.75	0.17	0.43	-0.76	0.19	1.07	0.50	0.44	0.18	-0.05	0.78	-0.21	**0.01**
	5	0.23	0.36	0.19	0.40	-0.17	0.79	1.07	0.53	0.36	0.30	0.02	0.90	-0.25	**0.01**
	6	0.42	0.17	0.30	0.22	0.29	0.67	0.68	0.71	0.30	0.38	0.12	0.52	-0.22	**0.04**

Notes: The table shows the response of the variables a 1% point increase in government spending using the methodologies described in Section 3.

The solid green lines in both graphs represent the impulse response functions, estimated using the Greenbook forecasts and Greenbook shocks to government spending. The additional lines in the top two panels of the graph depict impulse response functions estimated using forecasts from the SPF, RSQE, and Blue Chip data sets along with the Greenbook shocks to government spending. The figures in the bottom two panels represent impulse response functions estimated using Greenbook forecasts along with shocks to government spending from the SPF and RSQE forecasts. The impulse response functions in the top panels are more similar than the ones in the bottom panel indicating that the major source of difference is the shocks.

Appendix C. Data Appendix

C.1 Federal Reserve's Greenbook Forecasts

The Greenbook is the colloquial name given to the official report titled "Current Economic and Financial Conditions—Summary and Outlook" that is produced by the research staff at the Board of Governors of the Federal Reserve System.[24] This report is prepared for the FOMC prior to every scheduled FOMC meeting. It includes the forecasts of the US economy and is available to the FOMC members 6 days prior to every scheduled meeting. Before 1981, there was a regularly scheduled meeting every month. Since 1981, there have been eight scheduled meetings every year, approximately once every 6 weeks.

The Greenbooks are publicly available, but with a 5-year lag, which constrains our sample period to end in 2011. Although Greenbook forecasts are available from 1966 onward, the maximum forecast horizon was increased from five quarters ahead to seven quarters ahead in 1979. Eight-quarter ahead forecasts are available beginning in 1988. In this paper, we use data beginning in 1978.[25]

C.2 Blue Chip Economic Indicators

The Blue Chip Economic Indicators is a private survey of forecasters. Forecasts have been published every month since 1982. Over its history, the survey has maintained a relatively large and stable response rate of 30–50 forecasters. The consensus Blue

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

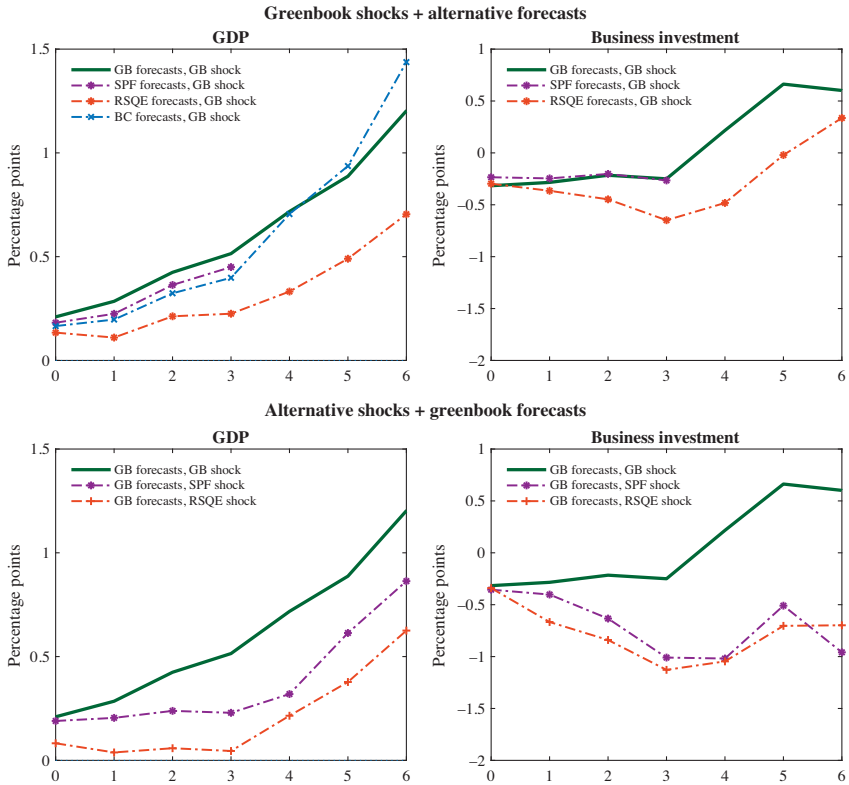

Figure 6. Are the shocks or forecasts driving the results: Impulse responses of GDP.
The figures show the impulse response functions of GDP and business investment using different combinations of forecasts and shock sources. The impulse responses are estimated using our baseline estimation equation (4). The top row uses Greenbook shocks to government spending with forecasts of GDP and investment from the Greenbooks, SPF, RSQE, and the Blue Chip data sets. The bottom row uses Greenbook forecasts of GDP and business investment with government spending shocks from Greenbook, SPF, and RSQE forecasts.

Chip forecasts are widely considered to be the best private sector forecasts, on par with the Greenbook forecasts in terms of forecast accuracy. In Section D, we compare the forecast performance of the various forecasts used in this paper and show that the Blue Chip forecasts are in fact on par with the Greenbook forecasts.

A major drawback of the Blue Chip forecasts is that they forecast a limited set of variables, and in particular do not ask respondents any questions about fiscal policy.[26] We can use the forecasts to obtain the impulse responses, but not to obtain shocks to fiscal policy.[27] In addition, they began forecasting consumption only in 1992 and they do not include forecasts of any investment category. While this restricts the usefulness of the Blue Chip forecasts, they still provide us with a valuable check on the accuracy of our results.

C.3 SPF Forecasts

The Survey of Professional Forecasters (SPF) is conducted once every quarter by the Federal Reserve Bank of Philadelphia.[28] We use SPF forecast data for the 1983–2011 period. The SPF data set contains mean forecasts of approximately 25–45 anonymous private sector forecasters. Anonymity implies that concerns over reputation and a desire for publicity do not bias forecasts.

The data set includes forecasts of quarterly real GDP, real consumption, real business fixed investment, real investment in residential construction, and the GDP deflator. Information on wages was not collected as part of the survey. The survey is conducted in the middle month of each quarter, at which time the survey respondents report their forecasts for each of the next four quarters and for the subsequent calendar year. The fact that the longer-term forecasts are on a calendar-year basis leads to an asymmetry between the short-term and the longer-term forecasts. This gives rise to a complex set of correlations in the data. We therefore only use the one-to-four-quarters ahead forecasts in our analysis.

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

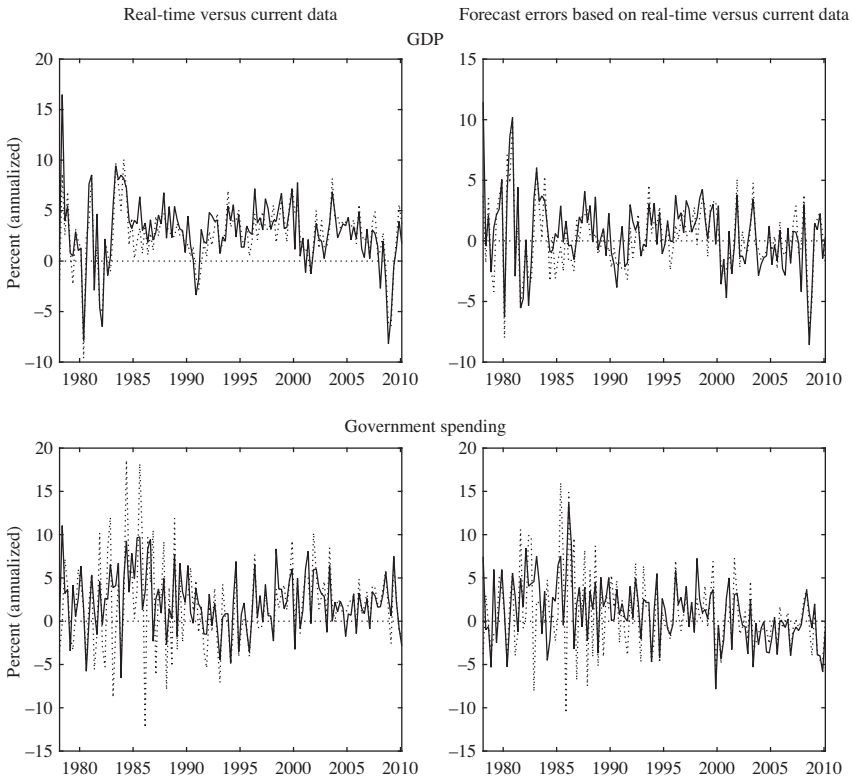

Figure 7. Growth rates of real-time versus current data and forecast error of growth rates.
Notes: Real-time refers to the data as it existed two quarters after the end of the reference quarter. Current refers to data as it currently exists. In the forecast errors column, real-time refers to errors constructed using the real-time data for actuals and current refers to current data being used for actuals. Positive shocks refer to spending being unexpectedly high.

C.4 RSQE Forecasts

The University of Michigan's RSQE, the original quarterly macroeconomic forecaster, began forecasting the US economy in 1952. They produce forecasts at regular intervals that have shifted slightly over time. Unfortunately, RSQE forecasts from 1952 to 1982 were destroyed in a fire. As with the Greenbook forecasts, the RSQE forecasts are judgment forecasts that are based on a large-scale structural model of the US economy.

We use RSQE's one through seven-quarter ahead forecasts of our variables of interest for the 1983–2011 sample period.[29] While their current practice is to publish detailed write-ups of their forecasts in March, May, September, and November, in the past their August set of projections were their featured (and sometimes only) third-quarter forecast. We draw one forecast from each quarter between 1983 and 2011, according to the following conventions: we use the March, May, and November forecasts throughout our analysis, and we use the August forecast from 1983–2007, and the September forecast from 2008 to 2011.

C.5 Real-Time or Current Data

In the use of forecast data, a primary issue in defining forecast errors is whether to use the data as is currently available (current data) or to use the data that were available around the time the forecasts were made (real-time data).[30] Many studies using forecast data tend to use real-time data to construct forecast errors; even though more recent vintages of historical data tend to reflect our improved understanding of the true value of the variable over time. These studies prefer to match forecasts with the contemporaneous understanding of the underlying data. The main argument for basing forecast errors on real-time data is that calculating forecast errors based on the currently available data might artificially inflate the forecast errors and incorrectly attribute definition changes as shocks.

In this paper, we base forecast errors on current data for three reasons. First, since we are working with cumulative growth rates, estimating forecast errors from real-time data is not necessarily desirable. Cumulative growth rates are subject

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

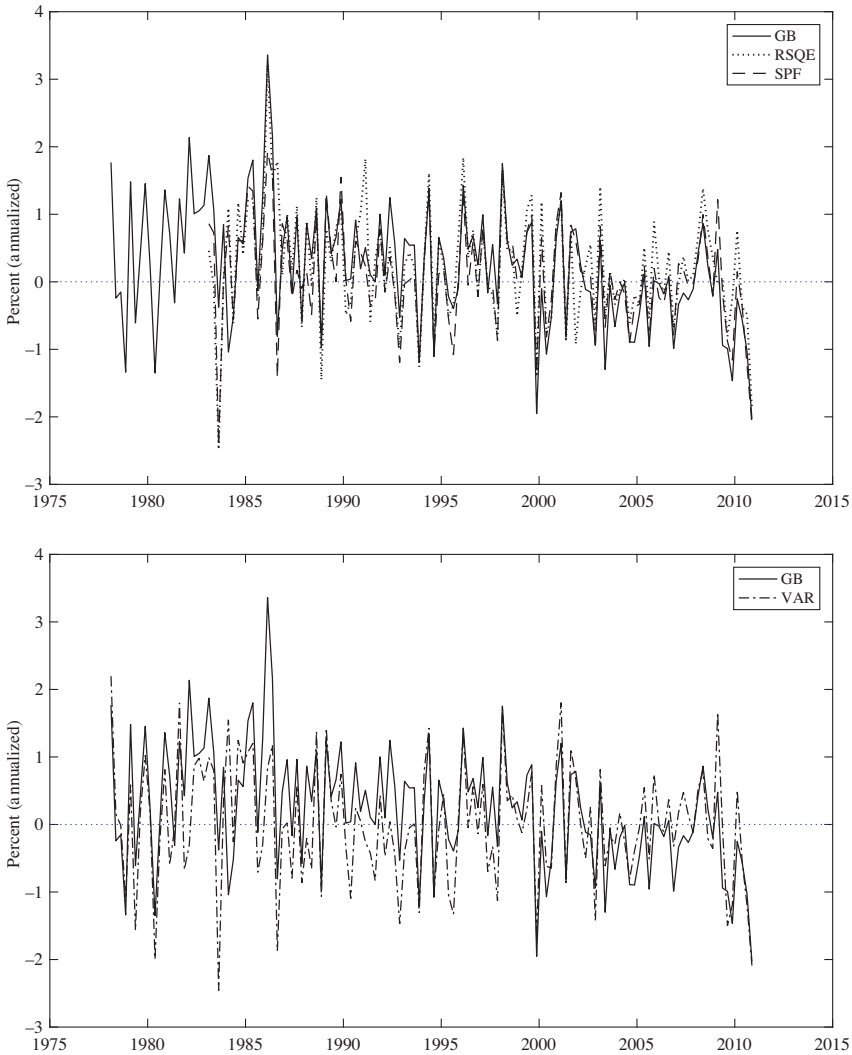

Figure 8. Structural errors.
Notes: The figure shows the structural shocks for government spending obtained from Greenbook, RSQE, and SPF forecast data sets, and from a VAR with G, T, Y using current data, and estimated only once over the entire sample. Positive shocks refer to spending being unexpectedly high.

to revisions over progressively longer time horizons than single-quarter rates. In addition, if an annual or comprehensive revision occurs between the forecast and the realization of the variable of interest, then not only the level but also the definition of variables is subject to change. In this case, it is not clear, even in theory, which real-time data concept is appropriate to compare to a forecast. Linked quarterly growth rates would reflect a blend of revisions and conceptual frameworks, while any other choice would be arbitrary and potentially vary according to the length of time over which the growth rate is calculated. Second, more recent vintages of historical data tend to reflect our improved understanding of the true growth rates over time which is what forecasters are likely trying to predict. Finally, since we are working with growth rates rather than levels, definition changes are less likely to have large effects on the growth rate of the variables of interest.

Figure 7 compares the real-time and current data for the growth rate of the variable indicated. The column on the left compares growth rates of the real-time data to those from current data for our two primary variables: GDP and government spending. The column on the right compares Greenbook forecast errors of the growth rate of the variable of interest, derived from real-time data versus using current data. An additional issue with using real-time data as the proxy for actual data in constructing forecast errors can be observed in the middle panel of the left column of the figure. Government spending

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

Table 3. Forecast error statistics: Mean

	FORECAST SOURCE	QUARTERS AFTER SHOCK						
		0	1	2	3	4	5	6
GDP	Greenbook	0.1	0.1	0.1	0.1	0.0	0.0	-0.1
	Blue Chip	0.1	0.1	0.1	0.1	0.2	0.2	0.3
	SPF	0.1	0.1	0.1	0.0	-	-	-
	RSQE	0.1	0.1	0.0	0.0	0.0	-0.1	-0.2
	VAR	0.0	0.1	0.1	0.1	0.1	0.1	0.1
Consumption	Greenbook	0.1	0.2	0.4	0.5	0.5	0.5	0.5
	Blue Chip	0.1	0.2	0.3	0.3	0.4	0.5	0.6
	SPF	0.2	0.3	0.4	0.6	-	-	-
	RSQE	0.1	0.2	0.3	0.4	0.5	0.6	0.6
	VAR	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Business Investment	Greenbook	0.1	0.1	-0.1	-0.3	-0.8	-1.1	-1.6
	Blue Chip	-	-	-	-	-	-	-
	SPF	0.0	-0.2	-0.3	-0.5	-	-	-
	RSQE	-0.1	-0.2	-0.5	-0.9	-1.4	-2.0	-2.7
	VAR	0.1	0.2	0.3	0.4	0.5	0.5	0.6
Residential Construction	Greenbook	0.5	0.6	0.4	0.1	-0.4	-0.9	-1.6
	Blue Chip	-	-	-	-	-	-	-
	SPF	0.1	0.0	0.0	0.0	-	-	-
	RSQE	0.1	-0.4	-1.3	-2.5	-3.6	-4.8	-5.7
	VAR	0.0	0.2	0.4	0.5	0.7	0.9	1.1
Government Spending	Greenbook	0.2	0.3	0.4	0.5	0.6	0.8	0.8
	Blue Chip	-	-	-	-	-	-	-
	SPF	0.1	0.2	0.3	0.4	-	-	-
	RSQE	0.2	0.5	0.8	1.0	1.3	1.6	1.8
	VAR	0.0	0.1	0.1	0.1	0.2	0.2	0.2
Real Wages	Greenbook	0.0	0.0	0.0	-0.1	-0.1	-0.2	-0.3
	Blue Chip	-	-	-	-	-	-	-
	SPF	-	-	-	-	-	-	-
	RSQE	0.0	0.0	0.1	0.1	0.2	0.2	0.2
	VAR	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
GDP Deflator	Greenbook	0.0	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2
	Blue Chip	-0.1	-0.2	-0.3	-0.4	-0.7	-0.8	-1.0
	SPF	-0.1	-0.2	-0.3	-0.4			
	RSQE	0.0	-0.1	-0.2	-0.4	-0.5	-0.7	-0.8
	VAR	0.0	0.0	0.0	0.1	0.1	0.1	0.1

Notes: The table shows the root mean squared errors for Greenbook forecasts (GB) and from two structural VARs. VAR (real-time) is based on data as it existed at the time it was published, while VAR (current) is based on currently available data. In both VARs, the baseline specification includes government spending, receipts, GDP, and the deflator. Additional variables are added one at a time. For government receipts, we report only the one-quarter ahead forecasts, which are the only horizon used in the paper.

tends to be revised significantly in later years. The period between 1981 and 1988 appears to be responsible for most of the subsequent revisions, and the current data are much less volatile than in the initial telling. GDP growth rates, however, are more similar between the real-time series and current data.

One of the most important steps in our analysis in this paper is estimating shocks to government spending. If we construct forecast errors from real-time data, then the revisions to government spending in the pre-1990 period are likely to be attributed as shocks. To avoid this misspecification of shocks as well as due to the complications with using real-time data when working with cumulative growth rates, in this paper, we use current data to construct forecast errors.

C.6 Comparison of Structural Errors

Figure 8 compares the structural errors for government spending from four different forecast sources: Greenbook, RSQE, SPF, and a VAR. The top panel compares structural errors from the three professional forecasters data sets, while the bottom

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

Table 4. Forecast error statistics: Root mean squared error

	FORECAST SOURCE	QUARTERS AFTER SHOCK						
		0	1	2	3	4	5	6
GDP	Greenbook	0.7	1.2	1.6	2.0	2.3	2.6	2.9
	Blue Chip	0.6	1.0	1.4	1.8	2.2	2.5	2.9
	SPF	0.7	1.2	1.6	2.0	-	-	-
	RSQE	0.6	1.0	1.5	1.9	2.3	2.6	3.0
	VAR	0.7	1.2	1.6	2.0	2.4	2.7	3.0
Consumption	Greenbook	0.6	0.9	1.3	1.7	2.0	2.4	2.7
	Blue Chip	0.4	0.8	1.2	1.6	2.1	2.5	3.1
	SPF	0.5	0.9	1.3	1.7	-	-	-
	RSQE	0.5	0.8	1.2	1.6	2.0	2.4	2.9
	VAR	0.6	0.9	1.1	1.5	1.8	2.1	2.4
Business Investment	Greenbook	1.8	3.3	4.7	6.2	7.7	9.0	9.8
	Blue Chip	-	-	-	-	-	-	-
	SPF	1.7	3.1	4.6	6.0	-	-	-
	RSQE	1.8	3.2	4.5	5.9	7.3	8.7	10.2
	VAR	1.5	2.4	3.4	4.5	5.7	6.7	7.6
Residential Construction	Greenbook	3.1	5.6	7.8	10.2	12.5	13.5	14.8
	Blue Chip	-	-	-	-	-	-	-
	SPF	3.4	6.2	8.9	11.2	-	-	-
	RSQE	3.3	6.0	8.7	11.4	14.0	16.5	19.5
	VAR	3.5	6.3	8.8	11.2	13.1	14.7	15.9
Government Spending	Greenbook	0.9	1.4	1.7	1.9	2.3	2.6	3.0
	Blue Chip	-	-	-	-	-	-	-
	SPF	0.8	1.1	1.4	1.7	-	-	-
	RSQE	0.9	1.3	1.7	2.0	2.4	2.8	3.1
	VAR	0.8	1.1	1.3	1.6	1.9	2.2	2.5
Real Wages	Greenbook	0.8	1.1	1.4	1.7	2.1	2.4	2.8
	Blue Chip	-	-	-	-	-	-	-
	SPF	-	-	-	-	-	-	-
	RSQE	0.8	1.1	1.4	1.6	1.9	2.0	2.3
	VAR	0.7	0.9	1.1	1.3	1.4	1.6	1.7
GDP Deflator	Greenbook	0.2	0.5	0.7	0.9	1.2	1.5	1.7
	Blue Chip	0.2	0.5	0.7	1.0	1.4	1.7	2.1
	SPF	0.2	0.5	0.8	1.1	-	-	-
	RSQE	0.3	0.4	0.6	0.9	1.1	1.4	1.7
	VAR	0.2	0.4	0.6	0.8	1.0	1.3	1.5

Notes: The table shows the mean absolute errors for Greenbook forecasts (GB) and from two structural VARs. VAR (real-time) is based on data as it existed at the time it was published, while VAR (current) is based on currently available data. In both VARs, the baseline specification includes government spending, receipts, GDP, and the deflator. Additional variables are added one at a time. For government receipts, we report only the one-quarter ahead forecasts, which are the only horizon used in the paper.

panel compares the Greenbook shocks to those obtained from a standard structural VAR with G, T, Y estimated using current data and estimated only once over the entire sample.

Notice, in both panels, there are minor differences in the shocks estimated from the different data sets; however, there are no obvious glaring differences between shocks. The pairwise correlation between each set of series lies between 0.7 and 0.9.

Appendix D. Forecast Evaluation

Tables 3 and 4 compare the means and RMSEs of the professional forecast data sets and for forecasts from a VAR. The Greenbook and VAR statistics are estimated over the 1978–2011 sample. The Blue Chip, SPF, and RSQE forecasts are estimated over the 1983–2011 sample, since the Blue Chip and RSQE data begin in 1983. We present statistics of the cumulative forecast errors, which are used in the estimation in the paper. The professional forecasters errors in this paper are computed using current (fully revised) data and not real-time data. The VAR in this section includes four lags of the logs of real GDP,

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press

government spending, government receipts, and one additional variable when we forecast inflation, consumption, and the two investment components. The VAR is estimated over the 1968–2011 sample.

Table 3 compares the average forecast error across the different data sets, which indicates the extent of bias in the forecasts. The VAR tends to have lower mean forecast error for all variables relative to the Greenbook forecasts, except for residential construction. However, the mean forecast error is not statistically different from zero for any of the forecasts and forecast horizons that we consider here.

Table 4 presents the RMSE, which is used to measure the accuracy of the forecasts. For all variables, the RMSE increases as the forecast horizon increases. We find this for the professional forecasters and the VAR. The different professional forecast data sets have similar statistical properties for all the variables we consider. For most variables, the Greenbook forecasts and the VAR have similar forecasting performance, with two exceptions. For business investment and for real wages, the VAR outperforms the professional forecasters. The VAR has an advantage over professional forecasts: the VAR is over-fitting since it uses out-of-sample data for estimation, and it has access to a consistent, revised data set.

Cite this article: Hall M and Thapar A (2023). "The economic effects of government spending: using expectations data to control for information." *Macroeconomic Dynamics* **27**, 141–170. https://doi.org/10.1017/S1365100521000341

https://doi.org/10.1017/S1365100521000341 Published online by Cambridge University Press