Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Expectations and Fiscal Policy: An Empirical Investigation

Roberto Perotti

Working Paper n. 429

This Version: June 17, 2011

Expectations and Fiscal Policy: An Empirical Investigation

Roberto Perotti[*]

First version: January 31, 2011
This version: June 17, 2011

Abstract

With fiscal foresight, the shocks identified by standard Vector Autoregression (SVAR) techniques can be non-fundamental for the variables of interest. In an important paper, Ramey (2011) uses direct measures of the private sector's forecast revisions of defense or federal spending to estimate the effects of government spending shocks in a VAR, obtaining the "expectations - augmented" VAR, or EVAR. The response of GDP to these shocks is smaller than 1, and consumption and the real wage fall: this is consistent with the neoclassical model, but the opposite of recent results from SVARs.

In this paper, I make three points. First, EVARs and SVARs give virtually the same results. Ramey reaches the opposite conclusion because she never estimates the two specifications on the same sample and with the same government spending variable.

Second, the evidence from EVARs is not robust. It is enough to dummy out just two quarters during WWII (when rationing was introduced) or during the Korean War (when new Fed regulation discouraging the purchase of durables was introduced) for the negative effects of defense spending shocks to disappear.

Third, the forecast revision of federal spending from the Survey of Professional Forecasters has high explanatory power for government spending, but for the "wrong" reason: the predictive power of expected government spending growth is extremely low, so that the forecast error is effectively actual spending growth less noise.

1

1 Introduction

A popular approach to estimating the effects of government spending shocks consists of regressing a government spending variable on past information, and tracing the dynamic effects of the residual of this regression on the variables of interest. This is the methodology embedded in the standard Vector Autoregression (SVAR) approach[1] of e.g. Blanchard and Perotti (2002). Contributions based on this methodology typically find that, on a post-Korean war sample, a surprise increase in total government spending on goods and services leads to positive responses of GDP, private consumption, and the real product wage in manufacturing or the business sector.

The focus on these three variables has two motivations. They are obviously of primary policy relevance, and they are also important in discriminating among alternative models and mechanisms of operation of fiscal policy. In a neoclassical model with lump-sum taxation, "throw - in - the - ocean" government spending, that has no productive or utility externality, affects the economy via a pure wealth effect: as government spending rises, the present value of taxes rises correspondingly; forward - looking individuals feel poorer, and reduce their demand for the consumption good and for leisure; hence, consumption falls, labor supply increases, and the real wage falls. Because investment also falls if government spending is sufficiently persistent, GDP is likely to increase less than the increase in government spending; in other words, the output multiplier is less than 1. In contrast, in a Keynesian model consumption and, in some versions, the real wage typically increase in response to a rise in government spending, and the output multiplier is typically larger than 1 (see e.g. Ravn, Schmitt-Grohé and Uribe 2006, Galí, López-Salido, and Vallés 2007, and Bilbiie 2011).

In an important recent paper, Ramey (2011) argues that in the SVAR approach the government spending shocks estimated by the econometrician are likely to be anticipated, and that this can lead to a spurious finding of a positive effect of these shocks on consumption and the real wage. Formally, with fiscal foresight the econometrician's information set is smaller than that of the private agents, so that the true fiscal policy shocks cannot be recovered from the estimated shocks: the MA representation is non-invertible, or non-fundamental for the variables used in the VAR.[2] The solution she proposes is to use direct measures of changes

---

[1]The acronym "SVAR" usually stands for "Structural VAR". But as described below, in the present context this approach has nothing structural as it is usually meant by this adjective in the VAR literature: it relies on a simple Choleski decomposition. When this approach is used to study the effects of tax shocks instead of spending shocks, identification is not obtained by a simple triangularization of the variance - covariance matrix of the residuals, hence the adjective "structural".

[2]See Lippi and Richlin (1994) or Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007) for general treatments of non - fundamentalness, and Leeper, Walker and Young (2008) and Forni and Gambetti (2011) for specific applications to fiscal foresight.

in the expectations of the present value of government spending: after simply appending it to a SVAR, one can then study the response to a shock to this variable. I will call this the "expectations - augmented" VAR, or EVAR.

Ramey (2011) focuses in particular on changes in the expectations of the present value of defense spending because this variable can be regarded as the quintessential "throw - in - the - ocean" government spending. She shows that in this incarnation of the EVAR, or "defense news VAR", the response of GDP to defense spending shocks is smaller than 1, the response of the various components of private consumption is negative or zero, and the real wage falls (although not in all samples). Barro and Redlick (2011) and Hall (2009) study the response to defense shocks in a SVAR with annual data, and reach a similar conclusion. In all these cases, the results are driven by WWII and the Korean War, which display by far by far the largest changes in defense spending. Ramey (2011) also uses forecast revisions of federal spending constructed from data of the Survey of Professional Forecasters (SPF); in this second incarnation of the EVAR, or SPF EVAR, she finds that over the last three decades private sector forecast revisions have even stronger negative effects on GDP and consumption.

In this paper, I make three points. First, and contrary to what claimed by Ramey (2011), EVARs and SVARs give virtually the same results. This is important because EVARs, and in particular defense news EVARs, are difficult to generalize to other countries and, within the US, other periods besides WWII and the Korean War. Many papers continue to use the SVAR approach despite the fact that Ramey's criticism is well founded in theory. Unless one shows that SVARs are still useful in practice, there is little basis to use them.

Ramey reaches the opposite conclusion - that EVARs and SVARs give radically different answers - because she never estimates the two specifications on the same sample and with the same government spending variable: she compares responses to total government spending shocks from a SVAR estimated over the post-WWII sample to responses to defense or federal spending shocks from EVARs estimated over several samples. I show that responses to a defense spending or federal spending shock from a SVAR are always virtually identical to responses from an EVAR with the same set of variables and estimated over the same sample.

Second, the evidence from EVARs - that the effects of spending shocks on consumption and the real wage are zero or negative - turns out not to be robust. In the defense news EVAR it is enough to dummy out just two quarters during WWII (when rationing was introduced) or two quarters during the Korean War (when new Fed regulation discouraging the purchase of durables was introduced) for the negative effects of defense spending forecast revisions on GDP and private consumption to disappear; in the SPF EVAR it is enough to eliminate the last year of the sample for the negative effects of federal spending forecast revisions to

become positive.

Third, I show that in the SPF EVAR the growth of federal spending predicted by professional forecasters has very little explanatory power for actual federal spending growth. Ramey (2011) constructs the forecast revision for federal spending as the difference between actual spending growth and predicted spending growth, and notices that this has indeed a strong predictive power for government spending. However, this is precisely because the predictive power of expected government spending growth is extremely low, so that the forecast error is almost equivalent to actual spending growth less some noise. This is the reason why SVARs and SPF EVARs give almost identical impulse responses. I also show that the SPF EVAR results are all due to that component of the constructed forecast error that is not in the information set of the professional forecasters, hence it is hard to attribute these effects to the wealth effect of the neoclassical model.

I conclude that the available evidence is that the output multiplier of government spending is likely to be in the neighborhood of 1. There is no robust evidence of a negative response of either GDP or private consumption or the real wage to government spending shocks; if anything, there is evidence of a positive (although not large) response of consumption and the real wage. A negative response of private consumption to defense spending shocks, and in some cases a very small or negative response of GDP itself, is instead perhaps the main conclusion of Barro and Redlick (2011), Hall (2009), and Ramey (2011).

The outline of the paper is as follows. Section 2 presents the SVAR and EVAR approaches. Section 3 presents the evidence from a defense news EVAR estimated over the longest sample, 1939:1-2008:4, and compares it to the evidence from a SVAR with the same variables and over the same sample. Section 4 presents the evidence on the sample that starts in 1947:1, and discusses the role of the Korean War. Section 5 discusses the construction of the revision of federal spending forecasts, using the Survey of Professional Forecasters, and the results from the corresponding EVAR. In section 6 I discuss issues of exogeneity of the shock. In section 7 I conclude by discussing how to reconcile the evidence presented so far with evidence from SVARs typically estimated in the literature, that consistently display positive effects on GDP and consumption.

2 SVARs and EVARs

2.1 A simple model

Consider a simple neoclassical growth model with log preferences, inelastic labor supply, and complete depreciation of capital. This is the same model as in Leeper, Walker and Young

(2008) and Mertens and Ravn (2010), except that I allow for government spending on goods and services G_t, financed by lump sum taxation T_t. These two papers focus on the issue of fundamentalness; here I exploit the extreme simplicity of the model to provide an economic intuition as to why neglecting fiscal foresight in a SVAR can lead to spurious findings of "neo-keynesian" results.

The competitive equilibrium is found by maximizing

$$\max E_t \sum_{i=0}^{\infty} \beta^i log C_{t+i} \tag{1}$$

s.t.

$$
\begin{align}
C_t + K_t + G_t &= Y_t \tag{2a} \\
Y_t &= Z_t K_{t-1}^{\alpha} \tag{2b}
\end{align}
$$

Z_t is a technology shock that is distributed lognormally with mean 0 and variance σ_z^2.

Letting small letters denote log deviations from the steady state, I postulate the following law of motion of government spending

$$g_t = a_{t/t-1} + \varepsilon_t \tag{3}$$

$a_{t/t-1}$ and ε_t are white noise shocks that are uncorrelated with each other and with z_t. $a_{t/t-1}$ represents news or announcements about government spending in period t, that become known in period $t-1$. Two properties of (3) are important. Because of decision and implementation lags, g_t does not respond contemporaneously to the other endogenous variables, like y_t; this helps identification in that it suggests a triangular structure, or Choleski decomposition, with g_t first. However, for the same reasons g_t is partly decided in advance, as evidenced by the presence of $a_{t/t-1}$; this complicates identification because typically the econometrician does not observe news about future government spending.

Obviously however government spending is not known in advance in its entirety; hence, I allow for a random term ε_t that generates a discrepancy between announcements and actual realizations.

After log-linearizing the Euler equation and the resource constraint, the solution for k_t is

$$k_t = \lambda_1 k_{t-1} + \delta z_t + \pi_0 a_{t/t-1} + \pi_1 a_{t+1/t} + \pi_0 \varepsilon_t \tag{4}$$

where

$$\pi_0 < 0, \quad \pi_1 = -\pi_0(1 - \phi) > 0, \quad \delta < 0 \tag{5}$$

$\lambda_1 < 1$ is the smaller root of the characteristic equation of the second-order difference equation describing the equilibrium behavior of k_t. ϕ is the inverse of the larger root. The exact expressions for these coefficients, with the solution of the model, are given in the Appendix.

Thus, a positive realization in t of the shock to future spending, $a_{t+1/t}$, causes an impact increase in k_t. c_t falls because of a negative wealth effect; since actual government spending in t does not move and k_{t-1} is given, from the resource constraint k_t must increase. Then k_{t+1} falls to make room for the actual increase in g_{t+1}; afterwards capital goes back to the steady state from below.

A positive realization of the shock to contemporaneous government spending, ε_t, causes instead an impact decline in k_t. Again because of the wealth effect, c_t falls, but for permanent income reasons less than one to one; given k_{t-1}, k_t must fall to make room for the increase in g_t.

2.2 EVARs

Suppose first the econometrician observes the government spending news $a_{t+1/t}$.[3] Then the econometrician can estimate the VAR

$$k_t = \lambda_1 k_{t-1} + \pi_0 a_{t/t-1} + \pi_1 a_{t+1/t} + \eta_{k,t}^E \qquad (6)$$

where

$$\eta_{k,t}^E = \delta z_t + \pi_0 \varepsilon_t \qquad (7)$$

I call this the "Expectations-Augmented" VAR, or EVAR, hence the superscript "E". Clearly, the econometrician can recover the coefficients of the impulse response to $a_{t+1/t}$ without bias. From (3), by regressing government spending g_t on the announcement $a_{t/t-1}$ the econometrician can also recover the contemporaneous shock to government spending ε_t; she could then insert it directly into (6), or regress the estimated residual $\widehat{\eta}_{k,t}^E$ on it, to get also the impulse response to ε_t. The MA representation corresponding to this system is fundamental: the econometrician can recover the original structural shocks ε_t and z_t using current and past values of the observable variables.

---

[3]To make the problem interesting, I will assume that the econometrician does not observe z_t.

2.3   SVARs

Now suppose the econometrician ignores the existence of fiscal foresight, and estimates the "Standard VAR", or SVAR:

$$k_t = \lambda_1 k_{t-1} + \mu_1 g_{t-1} + \eta_{k,t}^S \tag{8}$$

$$g_t = \rho_1 k_{t-1} + \rho_2 g_{t-1} + \eta_{g,t}^S \tag{9}$$

where the superscript "S" stands for "SVAR",

$$\eta_{k,t}^S = \delta z_t + \pi_0 a_{t/t-1} + \pi_1 a_{t+1/t} + \pi_0 \varepsilon_t \tag{10}$$

$$\eta_{g,t}^S = a_{t/t-1} + \varepsilon_t \tag{11}$$

and obviously $\mu_1 = \rho_1 = \rho_2 = 0$. The approach of Blanchard and Perotti (2002) consists of applying a Choleski decomposition where g_t comes first. In this case, it implies trivially a regression of the estimated residual $\widehat{\eta}_{k,t}^S$ on $\widehat{\eta}_{g,t}^S$. This would be appropriate if there were no fiscal foresight: the econometrician could estimate consistently the parameters of the impulse response function, and the impact effect of ε_t on k_t. Once again the estimated shock would be fundamental.

However, because of fiscal foresight $\eta_{k,t}^S$ is positively correlated with k_{t-1}, and specifically with $a_{t/t-1}$, hence the econometrician will not be able to recover the coefficients of the impulse response function.[4] Ignoring the difference between sample and population moments for simplicity, the Appendix shows that

$$\widehat{\pi}_0^S = \pi_0 \left[1 - \pi_1^2 \frac{var(a_{t/t-1})}{var(k_{t-1})} \frac{var(a_{t/t-1})}{var(a_{t/t-1} + \varepsilon_t)} \right] \tag{12}$$

If the term in brackets is less than 1 the estimate of π_0 is negative but smaller, in absolute value, than the true π_0.[5] Hence, the estimated impact response of k_t to a shock to contemporaneous government spending is biased towards 0. Intuitively, from (3) the econometrician estimates a shock to government spending only when the latter actually changes; but part of this change, $a_{t/t-1}$, is just the manifestation of an announcement that was made in $t-1$ and that had a positive effect on k_{t-1} and, through that, on k_t (see equation 4). Hence, the

---

[4]Note that the econometrician could estimate π_0 consistently by including g_t as a regressor in (8).

[5]A sufficient condition is that $\pi_1 = -\pi_0(1 - \phi)$ is less than 1 in absolute value, which is very likely for realistic values of the parameters $\widetilde{G}$, α and β, where $\widetilde{G}$ is the steady state share of government spending in GDP. For instance, with $\beta = .99$ and $\widetilde{G} = .5$ it is satisfied for $\alpha > .13$. The constraint is even easier to satisfy for smaller values of $\widetilde{G}$.

estimated effect of ε_t is less negative than the true one. [6]

The econometrician also concludes that a government spending shock has a negative (although estimated with a bias) impact on the capital stock, while in reality a pre-announced increase in government spending has a positive effect (i.e., π_1 is positive).

A similar argument shows that the Blanchard - Perotti approach estimates a smaller negative effect of ε_t on consumption than the true one. Like before, the econometrician is mixing the actual response to ε_t with the delayed response to $a_{t/t-1}$, which is less negative because of the positive effect of $a_{t/t-1}$ on k_{t-1}. In a model with elastic labor supply, it is easy to show that by the same argument the econometrician underestimates also the negative response of the real wage to ε_t.

In fact, the contributions that apply this approach, like Blanchard and Perotti (2002), Fatas and Mihov (2001), Galí, López-Salido, and Vallés (2007), Perotti (2007), Favero and Giavazzi (2009), and Auerbach and Gorodnichenko (2010), typically find a positive GDP response to a shock to total government spending, with multipliers that are slightly below or above 1, and a positive response of consumption; Monacelli, Perotti and Trigari (2010) also find a positive response of the real product wage.

2.4 Problems

This simple model provides the framework to discuss four problems with the SVAR approach that have been identified with the literature. The first two have to do with identification, the last two with the data.

1) It is clear that in (8) and (9) the econometrician cannot recover the true structural shocks. Without observing the news $a_{t+1/t}$, it is impossible to recover the shocks ε_t, z_t and $a_{t+1/t}$ separately. In this sense, the estimated shocks are trivially "non-fundamental". But the underlying problem is one of identification: in estimating a SVAR, the econometrician omits a variable, the news on future government spending. I call this the **"anticipation problem"** or **"non-fundamentalness problem"**.

2) The government spending shocks identified via a SVAR may not be really exogenous after all; there might be a term in k_t with a non-zero coefficient on the rights hand side of the government spending reaction function (3). Even if there were no fiscal foresight, this would make a Choleski decomposition invalid. I call this the **"exogeneity problem"**.

---

[6]The coefficient of k_{t-1} in (8) picks up some of the delayed effect of $a_{t/t-1}$ on k_t via the correlation between $a_{t/t-1}$ and k_{t-1}. This delayed effect is negative, hence the estimated residual $\widehat{\eta}_t^S$ is positively correlated with $a_{t/t-1}$, inducing a positive bias when regressed on g_t, which includes $a_{t/t-1}$. Exactly the same reasoning applies to the estimation of the consumption response.

3) There is just not enough variation in government spending other than defense to identify with confidence meaningful responses to government spending shocks. I call this the **"variance problem"**. A corollary of this is that one can only hope to learn from samples that include episodes of large military buildups, like WWII or the Korean war.

4) Finally, most of the items, other than defense, that make up government spending on goods and services (the variable typically used in a SVAR) either enter the utility function of the private sector, or have production externalities, or both. Thus, these items do not work via the wealth effect of pure "throw - in - the - ocean" government spending on goods and services, and as such are not useful to discriminate between alternative theories of the effects of government spending shocks. I call this the **"externality problem".**

2.5   Solutions

An EVAR obviously is designed to solve the first problem, by including observations on the present value of government spending forecasts directly in the specification. Perotti (2007) constructs a series of revisions of Congressional Budget Office forecasts of future government spending; the downside of this series is that it is only biannual, and started in 1984:1. Ramey (2011) uses two quarterly measures of forecast revisions. The first is an estimate of revisions to the present value of defense spending, from 1939:1 on, that she constructs from a reading of weekly magazines like *Business Week*; I will call this version of the EVAR the "defense news EVAR". The second is based on the median value of forecasts of federal and of state and local government spending, collected quarterly since 1981:3 in the Survey of Professional Forecasters; I will call this version of the EVAR the "SPF EVAR".

Besides the longer sample, one advantage of the defense news measure is that it can potentially address also the last three problems. In a sample that includes WWII and the Korean War there is a large variation in defense spending; this variation is plausibly exogenous; and defense spending is the quintessential "throw - in - the - ocean" government spending: resources taken from the private sector and providing no utility or productive externalities. Thus, I will start with the defense news VAR.

3   WWII

3.1   The defense news EVAR

I estimate the same specification of the defense news EVAR as in Ramey (2011), a VAR whose reduced form is

$$X_t = A(L)X_{t-1} + U_t \tag{13}$$

where X_t is a vector of variables that include, in addition to the defense news variable[7], the log of real per capita government spending on goods and services g_t, the log of real per capita GDP y_t, the three-month T-bill rate i_t, the Barro-Redlick average marginal income tax rate τ_t, and a sixth variable that rotates depending on the response one wants to study. These variables are: total hours, the real product wage in manufacturing, consumption of durables, of non durables, of services, and total investment.[8] Each equation includes four lags of these variables, a constant, and linear and quadratic time trends. The sample runs from 1939:1 to 2008:4.[9]

Columns 1 and 4 of Figure 1 display the median responses of the variables listed above to a shock to the defense news variable; thus, they reproduce the responses in Figure VIII of Ramey (2011), except that the responses of national income aggregates are expressed as percentage points of GDP instead of percentage changes.[10] The initial shock to defense news is normalized so that the maximum response of total government spending is one percentage point of GDP. Both 68 and 95 percent confidence bands are displayed.[11]

Total government spending peaks after one year; at about the same time, GDP increases by slightly less than 1 percent; durables and nondurables consumption fall by about .2 and .1 percentage points after one year; service consumption increases by about .1 percentage point of GDP after two years. Thus, the consumption responses are small, but they are significant, at peak or trough, at the 95 percent confidence level. Combining the responses of nondurables and services, there is no evidence of a negative response of nondurables and services: the claim that in this sample defense spending shocks lead to a decline in consumption therefore rests on the behavior of durables. The real product wage in manufacturing increases by .4

---

[7]This variable is constructed as each quarter's revision of the nominal value of future expected defense spending, divided by the previous period's nominal GDP.

[8]All variables were donwloaded from Valerie Ramey's website.

[9]The Barro-Redlick tax rate is available until 2006:4. It was extended by Valerie Ramey assuming that it "changed by the same percent in 2007 as my update of the Alexander-Seater (2009) series and (for want of more information) was constant through 2008." As we will see, extending the sample beyond 2006 is not innocuous for several results.

[10]These are obtained by multiplying the log responses by the average ratio of that variable to GDP

[11]These are computed based on 1000 bootstrap replications with replacement.

percent after 2 years, with the peak effect marginally insignificant. Investment falls.

3.2   Rationing

Defense spending variation in this sample is dominated by the two episodes of WWII and the Korean War. There is a long debate on whether WWII can provide any useful information on the effects of fiscal policy in "normal" times. Hall (2009) and Barro - Redlick (2011) argue that the combined effect on the multiplier of controls, rationing, the draft, and patriotism, is likely to be negative (Hall) or positive (Barro - Redlick). However, these authors also recognize that these are just conjectures based on intuition: we are in the dark on whether the existing estimates of fiscal multipliers during WWII provide an upper bound or a lower bound to the true response of GDP.

The response of consumption is complicated further by rationing and supply bottlenecks (see e.g. Higgs 1992). Table 1 displays the starting date for the main rationed items. The main durable goods were formally rationed from the first quarter of 1942. Supply bottlenecks, however, started before that, due to the enormous increase in military spending several quarters before Pearl Harbor: as Gordon and Krenn (2010) note, "pre-Pearl Harbor 'preparedness' emergency caused the share of government spending (including state, local, and Federal) to increase from 11.5 percent in 1940:Q2 to 25.6 percent in 1941:Q4, and [...] all of this increase took the form of federal government military expenditure" (p. 11). In addition, "[d]espite the high demand for automobiles, spending on consumer durables actually fell in the third and fourth quarters of 1941 due to capacity constraints" (p. 17).

Columns 2 and 5 of Figure 1 show that it is enough to dummy out 1941:4 and 1942:1 for a negative response of durables consumption in the defense news EVAR to disappear; in fact, now there is a positive and significant response on impact, which then falls back towards trend. The response of GDP is not affected, and remains positive.

Table 1 also shows that non-durable items were rationed at different times during the war, hence there is no specific quarter that can be naturally dummied out. In fact, the response of non-durables is not sensitive to the exclusion of a few quarters, and neither is the response of services.

3.3   The SVAR

Column 3 and 6 of Figure 1 display the responses from a SVAR. To ensure maximum comparability, the variables are the same as in the defense news EVAR, except that obviously

Table 1: **Main rationed items.**

Durables		Non-durables and semi-durables	
Tires	Jan. 1942 - Dec. 1945	Gasoline	May 1942 - Aug. 1945
Cars	Feb. 1942 - Oct. 1945	Sugar	May 1942 - 1947
Typewriters	Mar. 1942 - Apr. 1944	Fuel oil, kerosene	Oct. 1942 - Aug. 1945
Bicycles	Jul. 1942 - Sept. 1945	Rubber footwear	Oct. 1942 - Sep. 1945
Stoves	Dec. 1942 - Aug. 1945	Coffee	Nov. 1942 - Jul. 1943
		Shoes	Feb. 1943 - Oct. 1945
		Processed food	Mar. 1943 - Aug. 1945
		Meats, canned fish	Mar. 1943 - Nov. 1945
		Cheese, canned milk, fats	Mar. 1943 - Nov. 1945
		Solid fuels	Sep. 1943 - Aug. 1945
Source: "Rationed Goods in the USA During Second World War", http://www.ameshistoricalsociety.org/exhibits/rationitems.htm			

in the vector X_t the defense news variable is replaced by the log of real defense spending per capita. Defense spending is ordered first, and again the shock is normalized so that the peak effect on total government spending is 1 percentage point of GDP.

All responses are strikingly similar to those of the defense news EVAR in columns 1 and 4. Like the defense news EVAR, the SVAR shows a positive response of GDP, the real wage and service consumption, a negative response of durable consumption, and a flat response of nondurables. The sizes of the responses are almost identical to the defense news EVAR, except that the GDP response is smaller in the SVAR. Thus, there is no evidence in this sample that a SVAR leads to spuriously large positive effects of defense spending on GDP; the real wage, or consumption.

Ramey (2011) reaches the conclusion that SVARs and EVARs deliver very different results because she never estimates a SVAR and an EVAR with the same set of government spending variables and on the same sample. Her SVAR evidence comes from a specification where the government spending variable is total government spending, estimated over the post-WWII period; as several contributions (see those cited above) have shown, and as I show below, this specification does deliver large positive responses of GDP, private consumption and the real wage. This suggests that there may be important differences in the effects of federal (mostly defense) and of state and local spending, a point to which I will return later.[12]

---

[12]The SVAR is much more robust to the exclusion of a few quarters: in fact, when 1941:4 and 1942:1 are dummied out the responses (not shown) are virtually unchanged. The reason is that the defense news VAR

There are three important conclusions to be drawn from the evidence based on the full 1939:1-2008:4 sample: (i) defense news EVARs and SVARs give virtually the same answers; (ii) however, the defense news EVAR results on durable consumption depend crucially on two quarters associated with the introduction of rationing: when these two quarters are omitted, durable consumption, like service consumption, increases; (iii) the defense news EVAR is more sensitive to these few quarters than the SVAR.

I now show that the same conclusions can be drawn from a sample that includes only the Korean War.

4 Korea

4.1 The defense news EVAR

Skepticism about the information contained in WWII data may suggest using a post-WWII sample. Starting the sample in 1947:1 has the additional advantage that official quarterly national income data started to be collected on this date.

Columns 1 and 4 of Figure 2 show the impulse responses from a shock to the defense news variable; they reproduce the responses in Figure IX of Ramey (2011). The GDP response is less than 1, and about the same as in the sample starting 1939:1, although the standard errors are now much larger: even the peak is not significant at the 95 percent level (for this sample, Ramey 2011 displays impulse responses but not the standard errors). Following a positive impact response of about .25 percentage points of GDP, after one year consumption of durables falls below trend by about the same amount. Nondurables decline by about .1 percent of GDP, although the response is significant or close to significant only in the first four quarters; service consumption is flat, with very large standard errors. The real product wage has the opposite behavior than in the full sample: it now falls by a large amount, about 1.5 percent, then returns to trend after 6 quarters. Investment is flat, with rather large standard errors. Thus, among all the results presented in Ramey (2011), these are the most consistent with the neoclassical model.

4.2 Federal Reserve regulation at the outset of the Korean War

However, these "neoclassical" features - the low GDP effect, and the decline in consumption and in the real wage - rest crucially on two key quarters, 1950:3 and 1950:4. As always, one could argue that one should not discard any useful information contained even in a

---

has relatively few non-zero entries, and most are very small; hence, it is more sensitive to a few quarters.

few quarters; however, in those quarters the presence of two well-identified, exceptional concurrent factors makes it difficult to interpret the results as the effects of the defense news variable.

First, between July and October 1950, and again at the beginning of the first quarter of 1951, there were two waves of almost panic buying of durables, motivated by the memory of the WWII experience with rationing. These waves of buying were followed by large declines, as the public realized that the experience of WWII would not be repeated. For instance, the *Survey of Current Business*, February 1952, p. 5, writes: "Initially, the weakening of the basic civilian components of aggregate demand was largely the aftermath of the two waves of anticipatory spending which had followed the outbreak of hostilities in Korea. With the improvement of the military situation there, with many consumer needs unusually well satisfied by several months of heavy buying, with a reassuring volume of production dissipating the specter of serious shortages, and with the stabilization of prices removing another incentive to forward purchasing, consumers reduced their expenditure markedly in the second quarter [of 1951]".

Second, although there was no formal rationing, in the first two quarters of the Korean War important restrictions on the purchase of durables were introduced; both were motivated by developments preceding the war. On September 18, 1950, the Federal Reserve introduced Regulation W, setting higher downpayments than those prevailing in the market for the purchase of durable goods, and reducing the maturities of the loans; the rules were further tightened on October 16 1950.[13] The *Survey of Current Business*, November 1950, calculates that Regulation W might have decreased the purchase of durables by about $2.5 to $3 billion annually, or about 10 percent of total durable purchases and about 1 percent of 1950 GDP. Regulation W should be seen against a steady increase in installment credit at the end of the forties: by 1950, less than half of the durables purchased were paid cash; and in 1949 one every four new cars was purchased by households with less than $3,000 of income, against one in eight the year before.[14] In addition, Regulation X, also introduced in the fall of 1950, restricted the terms of mortgages; by mid-1951, it had caused a decline in homebuilding, which in turn was reflected in a decline in the purchases of durables and semi-durables like furniture and household equipment.[15]

Thus, in 1950:4 there were factors, unrelated to defense spending, that had a negative impact on the consumption of durables: the aftermaths of the first wave of panic buying, and the introduction of two Federal Reserve regulations. Columns 2 and 5 of Figure 2

---

[13] See the *Survey of Current Business*, November 1950, p. 11 for a detailed description.

[14] See the *Survey of Current Business*, November 1950, p. 12.

[15] See the *Survey of Current Business*, November 1951, p. 7.

show that when 1950:4 and 1951:1 are dummied out, the consumption of durables increases significantly, and never falls below trend. Both other types of consumption also increase significantly, after an initial decline, and so does investment. The response of GDP almost doubles and now exceeds 1; also in contrast to the regressions that includes 1950:4 and 1951:1, the real wage now increases. In addition, the standard errors bands tighten considerably: now all responses are significant or nearly significant at the 95 percent level.

Predictably, when also 1950:3 - a quarter with both a large increase in expectations and a wave of panic buying not offset yet by new regulation - is dummied out, the positive responses of consumption disappear; but the standard errors are now extremely large, and all responses are now entirely insignificant (results not shown).

Thus, when the three quarters between 1950:3 and 1951:1 are dummied out, there is no statistically significant information in the responses from a defense news EVAR that starts after WWII. This is not surprising: in 1950:3 and 1950:4 the expectation of the present value of future defense spending rose by 63 and 41 percent of GDP, respectively; the next two largest revisions during the Korean war were minuscule by comparison: -2.02 percent of GDP in 1953:1 and -3.06 percent in 1953:3.

4.3   The SVAR

Column 3 and 6 of Figure 2 display the responses from a SVAR over the same 1947:1 - 2008:4 sample. Although care should be exercised because the standard errors are large both here and in the defense news EVAR, it is noticeable that once again the SVAR responses are similar to the defense news EVAR responses in columns 1 and 4. The main differences are the smaller increase in GDP, the absence of the initial increase in the response of durables consumption, and a positive real wage response. Therefore, like in the WWII sample, a SVAR gives a very similar picture to that of the defense news EVAR, and if anything a more "conservative" picture in terms of responses of GDP and consumption relative to the full sample EVAR. Like before, a SVAR is not sensitive to the omission of a few critical quarters.

Ramey (2011) makes the point that the Ramey-Shapiro dummy helps predict the SVAR residuals, but not viceversa. Row 1 of Table 2 shows that indeed this is the case; however, once again all of the predictive power of the war dummy comes from Korea, and from the usual two quarters. In fact, if one omits the two quarters 1950:4 and 1951:1 (row 2), or if the sample starts in 1954:1 (row 3), then the military dummy variable no longer Granger causes the SVAR residual.

The main conclusions are the same as for the long sample: (i) the informational content of the Korean War rests on just two quarters: without these quarters, the response of

Table 2: **Granger causality.**

		F-stat	p-value
1	1947:1 - 2008:4	3.56	.008
2	excludes 1950:4 - 1951:1	1.59	.177
3	1954:1-2008:4	.85	.495

Regression of the residual of the defense spending equation from the 6-variable SVAR on 4 lags of the defense buildup dummy variable. The F-statistics refers to the exclusion of 4 lags of the defense buildup dummy variable.

consumption and of the real wage to defense shocks becomes positive, and the GDP response exceeds 1; but the interpretation of what happened in those quarters is clouded by a number of exceptional concurrent factors; (ii) the responses to defense spending shocks in a SVAR are virtually identical to those of a defense news EVAR; (iii) a SVAR is less sensitive to the exclusion of 1950:4 and 1951:1; but, with or without these quarters, a SVAR would not lead to the "incorrect" (from the point of view of the defense news EVAR) inference that in this sample private consumption increases in response to defense spending shocks.

5   The SPF EVAR

5.1   Constructing the forecast errors

After the Korean War there is not enough variation in the defense news variable to estimate responses to that variable with reasonable degrees of precision. However, starting in 1981:3, quarterly forecasts of real federal spending and of real state and local spending by professional forecasters, with horizons from 0 to 4 quarters ahead, have been assembled by the Survey of Professional Forecasters, and can be used to estimate an EVAR as in Ramey (2011): I will call this the SPF EVAR.

Let f_t be the log of federal government spending from the latest BEA National Income and Product Accounts. To construct the surprise in federal spending in t, it would be natural to define

$$f_t^u = f_t - f_{t/t-1}^e \tag{14}$$

where $f_{t/t-1}^e$ is the SPF expectation of federal spending. One problem is that in the SPF the base year changes several times during the sample, hence one would need vintage data from NIPA using the same base year as the SPF; but on those quarter t when the SPF base year

changes, there is no match between the SPF and NIPA data, and it is necessary to rebase the SPF expectations using the ratio of NIPA government spending in some quarter $t - i$ expressed in the two base years used in quarters t and $t - 1$.

An alternative, that does not require rebasing nor the use of vintage NIPA data, consists of using the forecast error of the rate of growth of f_t rather than its level:

$$\Delta f_t^u = \Delta f_t - \Delta f_{t/t-1}^e = (f_t - f_{t-1}) - (f_{t/t-1}^e - f_{t-1/t-1}^e) \tag{15}$$

This is the definition implemented by Ramey (2011). Note that if the ratio of two variables at different dates is preserved when changing the base year, and if variables dated t are in the information set of the private sector at t, then (15) would be identical to (14). But neither condition is true;[16] in fact, the correlation between f_t^u and Δf_t^u is only .41.

5.2   Results

To preserve comparison with Ramey (2011), I will use Δf_t^u. I estimate an SPF EVAR, i.e. an EVAR that has the same specification as the defense news EVAR in sections 3 and 4, except that Δf_t^u replaces the defense news variable itself. Also, since defense spending is largely concentrated in federal spending, I split total government spending into federal spending and state and local spending. Thus, I end up with a 7-variable VAR that includes the SPF forecast error of federal spending, the log of real per capita federal government spending on goods and services, the log of real per capita state and local government spending on goods and services, the log of real per capita GDP, the three-month T-bill rate, the Barro-Redlick average marginal income tax rate, and a seventh rotating variable. The sample runs from 1981:3 to 2008:4.[17]

Figure 3, columns 1 and 3, displays responses to a shock to Δf_t^u. To preserve comparability with the previous results, I now normalize all responses so that the peak (which usually occurs on impact) of the sum of the responses of federal and of state and local government spending is 1 percentage point of GDP. The response of federal government spending on impact reflects the shock to the forecast error almost one to one (see section 5.3 for an explanation). The pattern of responses is distinctly neoclassical. Output falls by more than 1.5 percent after 3 quarters, although the standard errors are large. The consumption of durables, non durables and services decreases by about .5 percentage points of GDP each; in each case the response is significant at the trough. The real product wage and investment

---

[16]In the survey, forecasters do not know yet f_{t-1} or f_t when forecasting f_{t+1}.

[17]This is the same VAR estimated by Ramey (2010) except that I have two government spending variables instead of total government spending.

also fall. The corresponding SVAR - i.e., based on the same list of variables except that obviously Δf_t^u is excluded - exhibits once again virtually the same responses to a federal spending shock, except that the standard errors are larger (columns 2 and 4).

Ramey (2011) concludes that these SPF EVAR results are further evidence of "rather contractionary effects" of government spending on GDP and consumption. However, these results are highly influenced by the last year of the sample. Figure 4 shows that, when 2008 is excluded, only the response of the real wage in the long run and the response of services on impact remain negative. GDP and hours increase by more than 1 percent, and the two other types of consumption are flat. The standard errors remain large, however the initial response of GDP is significant at the 95 percent level. The SVAR too exhibits the same pattern.

Why is 2008 so influential? At the end of that year, federal spending in real terms was up 9 percent relative to a year before; but GDP declined by almost 3 percent, and durable and nondurable goods consumption by 12 and 3 percent, respectively. Thus, that year saw federal spending and GDP and its components move by very large amounts in opposite directions. Moreover, the large increase in federal spending caught professional forecasters by surprise in each quarter: the anticipated component of Δf_t^u, $f_{t/t-1}^e - f_{t-1/t-1}^e$ in expression (15), is persistently negative.

So far I have estimated the response of output and private consumption to the surprise in the contemporaneous government spending. But the wealth effect in the neoclassical model depends on the revision in the whole present value of contemporaneous and future government spending. I approximate this present value by adding the forecast revisions 1 to 4 quarters ahead (the maximum horizon feasible with the SPF) to Δf_t^u, discounting with a factor of .99 per quarter:

$$E_t(PDV_t) = \Delta f_t^u + \sum_{j=1}^{4} .99^i (f_{t+j/t}^e - f_{t+j/t-1}^e) \qquad (16)$$

The results (not shown) are virtually identical to those displayed in Figures 3 and 4. The reason is the usual one: all the information content is in Δf_t^u; the revisions of forecasts of future spending have no explanatory power (see again the next section for an explanation).

The conclusions are once again similar to those of the previous sections: (i) the evidence of contractionary effects of federal government spending on GDP or consumption is not robust; the robust evidence points if anything to positive effects on GDP and consumption; (ii) the responses of a SVAR are once again similar (in fact, almost identical) to those of a SPF EVAR.

5.3 The predictive power of the SPF forecasts

Besides the instability and the large standard errors, there are two important reasons to be skeptical about the SPF results. The first concerns the predictive power of these expectations; the second concerns the mapping between the expectations as constructed and the actual expectations of the private sector. In this and the next subsections I discuss these issues in turn.

As Ramey (2011) notices, the forecast error Δf_t^u has an impressive explanatory power for federal (and total) government spending. Row 1 of Table 3 shows that a regression of Δf_t on lags 0 to 4 of Δf_t^u and on lags 1 to 4 of f_t, g_t, y_t, i_t, and τ_t gives an R^2 of .81; the marginal F-statistics[18] is 131.5. Ramey (2011) reports a marginal F-statistics of 201.9; if, like her, I omit lags 1 to 4 of the forecast error, I obtain a marginal F-statistics of 511.5 (row 2).[19]

These F-statistics are so large that they should be regarded with suspicion. By comparison, the defense news shock has a much lower forecasting power: on the sample that starts in 1955:1, the marginal F-statistics is 2.01; but even on the sample which exhibits the highest explanatory power, starting in 1947:1, the marginal F-statistics is 22.5 (see Ramey 2011, Table 3).

But what exactly does this large explanatory power of the forecast error tell us? It is important to note that the forecast error Δf_t^u is *constructed* as the actual change Δf_t less the predicted change $\Delta f_{t/t-1}^e$ (see equation 15). I now show that Δf_t^u has a high explanatory power for Δf_t for the "wrong" reason, i.e. because the predicted change has very little explanatory power, so that the variable Δf_t^u is dominated by the term Δf_t.

In fact, if the correlation between Δf_t^u and $\Delta f_{t/t-1}^e$ is close to 0, as it is the case in the data, a high R^2 in the regression of Δf_t on Δf_t^u implies a low R^2 in the regression of Δf_t on $\Delta f_{t/t-1}^e$, a high marginal F-statistics for Δf_t^u and a low F-statistics for $\Delta f_{t/t-1}^e$. Thus, the fact that Δf_t^u has good explanatory power for Δf_t simply means that $\Delta f_{t/t-1}^e$ is a poor predictor of Δf_t. In fact, row 3 of Table 3 shows that the R^2 in a regression of Δf_t on lags 0 to 4 of Δf_t^e is .23, with a marginal F-statistics of only 5.3.

The conclusion is that Δf_t^u has a large explanatory power for Δf_t because it is constructed as Δf_t less a variable - the forecast revision $\Delta f_{t/t-1}^e$ - that turns out to be largely noise.

---

[18] This is the test of the exclusion of lags 0 to 4 of Δf_t^u.

[19] There are three differences between the regressions underlying the statistics in Ramey's Table IV and those in Table 3 here: her dependent variable is total government spending; her sample starts in 1968:1 because for the first thirteen years she uses forecast errors for defense spending, from unpublished material of the SPF; and on the right hand side she has only lag 0 of the forecast error (this seems to contradict the explanatory notes to her Table IV, that indicate that lags 0 to 4 are included, but can be inferred from the

Table 3: **Explanatory power of forecasts and forecast errors for** Δf_t

	(1)	(2)	(3)	(4)	(5)
	indep. var.	R^2	F-stat	coeff. of lag 0	t-stat
1	Δf_t^u	.81	131.5	.99	23.51
2	Δf_t^u	.89	511.46	1.03	20.47
3	$\Delta f_{t/t-1}^e$	.21	5.30	.23	1.77

See the text for the definition of Δf_t^u and $\Delta f_{t/t-1}^e$.
Row 1: regression of Δf_t on lags 0 to 4 of Δf_t^u and on lags 1 to 4 of f_t, g_t, y_t, i_t, and τ_t.
Row 2: regression of Δf_t on lags 0 of Δf_t^u and on lags 1 to 4 of f_t, g_t, y_t, i_t, and τ_t.
Row 3: regression of Δf_t on lags 0 to 4 of $\Delta f_{t/t-1}^e$ and on lags 1 to 4 of f_t, g_t, y_t, i_t, and τ_t.
Column 2: R^2 of the regressions.
Column 3: marginal F-statistics for the exclusion of lags 0 to 4 of Δf_t^u (row 1), of lag 0 of Δf_t^u (row 2), and of lags 0 to 4 of $\Delta f_{t/t-1}^e$ (row 3).
Column 4: coefficient of lag 0 of Δf_t^u or $\Delta f_{t/t-1}^e$.
Column 5: its t-statistics.

5.4 The information set of the private sector

The definition of Δf_t^u given above assumes that f_t is in the information set of the private sector at time t. In reality, it is not: in the SPF survey, respondents do not know the value of f_t, nor that of f_{t-1}. Hence, one can decompose the forecast error Δf_t^u into:

$$\Delta f_t^u = \Delta f_t - \Delta f_{t/t-1}^e = \underbrace{(\Delta f_t - \Delta f_{t/t}^e)}_{\textbf{time t's surprise in } \Delta f_t} + \underbrace{(\Delta f_{t/t}^e - \Delta f_{t/t-1}^e)}_{\textbf{revision of expectation of } \Delta f_t} \tag{17}$$

where $\Delta f_{t/t}^e \equiv f_{t/t}^e - f_{t-1/t}^e$.[20] The first component on the right hand side of (17), $\Delta f_t - \Delta f_{t/t}^e$, is the realization of Δf_t in excess of its expectation in t; this is not in the information set of the private sector at time t. The logic of the neoclassical model is that an increase in "throw-in-the-ocean" government spending that is unknown to the private sector does not reduce the agent's perception of her human wealth and therefore does not have a wealth effect on her labor supply or consumption. The second component, $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$, is the

programs posted on her website).

[20]In the Survey of Professional Forecasters, respondents a time t do not know the exact value of the variable in t or $t-1$.

revision in the private sector's expectation of Δf_t; this is in the private sector's information set at time t. It is the correct measure of the expectation revision at time t if one wants to test the key mechanism at work in the neoclassical model, namely the wealth effect.

As Table 4 shows, the term $\Delta f_t - \Delta f_{t/t}^e$ has a large explanatory power for the actual change in federal government spending: instead, the term $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$ has an F-statistics of just 1.59. This is not surprising: like $\Delta f_{t/t-1}^e$, $\Delta f_{t/t}^e$ is largely noise, hence $\Delta f_t - \Delta f_{t/t}^e$ behaves very similarly to $\Delta f_t - \Delta f_{t/t-1}^e$; on the other hand, $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$ is the difference between two expectations, at t and $t-1$, both of which have very poor forecasting power for Δf_t.

I estimate two different SPF EVARs, with each of the two components of Δf_t^u in place of Δf_t^u itself. The sample stops in 2007:4. Figure 5 displays the results: columns 1 and 3 display the responses to the first component, $\Delta f_t - \Delta f_{t/t}^e$; columns 2 and 4 display the responses to the second component, $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$. The responses to the first component are virtually indistinguishable from those to Δf_t^u in Figure 3.

The second component has an enormous, and implausible, negative effect on GDP and hours, up to - 5 percent. However, and not surprisingly given the results of Table 4, the standard errors are extremely large. Note also that the response of federal government spending is now about half that to the first component (i.e., a much larger shock to the second component is required to achieve a maximum response of total government spending equal to 1 percentage point of GDP, the assumed normalization). This is another instance of the "bad instrument" problem: government spending forecasts convey little information on future government spending, and so does their revision $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$.

Thus, the evidence suggests that the component of the forecast error constructed by Ramey (2011) that is not in the information set of the private sector, $\Delta f_t - \Delta f_{t/t}^e$, drives the estimated effects of Δf_t^u; the remaining component, which can be related to the wealth effect of the neoclassical model, is essentially noise.

Table 4: **Decomposing the forecast error**

	(1)	(2)	(3)	(4)	(5)
	indep. var.	R^2	F-stat	coeff. of lag 0	t-stat
1	$\Delta f_t - \Delta f_{t/t}^e$	.76	32.70	.81	11.32
2	$\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$	.38	2.80	.66	2.76

The Table has the same structire as Table 3, except for the independent variables in the regressions, as listed in column (1). See the text for the definition of $\Delta f_t - \Delta f_{t/t}^e$ and $\Delta f_{t/t}^e - \Delta f_{t/t-1}^e$.

6 The exogeneity problems

Barro and Redlick (2011) argue that the government spending shock in a SVAR is likely to be endogenous, as a higher GDP leads to higher taxes and therefore to more government spending: this is one manifestation of the exogeneity problem described in section 2.4. However, Barro and Redlick (2011) use yearly data: their argument is unlikely to hold at the quarterly frequency. In fact, because of decision lags, contemporaneous discretionary government spending is unlikely to respond within a quarter to any news about the economy; and, unlike some welfare spending or tax revenues, it is difficult to think of automatic mechanisms linking government spending on goods and services to economic conditions.

But there is a different sense in which government spending on goods and services can be endogenous: suppose future GDP is expected to increase for exogenous reasons. Then future government spending might be expected to increase (high expected revenues allow higher expected government spending) or to decrease (if policymakers are expected to use future government spending in a countercyclical fashion).

This suggests controlling for the forecast revision of future GDP when estimating the response to a shock to the revision of the present value of government spending. Therefore, I add the present value of forecast revisions of GDP to the present value SPF EVAR introduced above (see 16). I then estimate responses to the component of the revision of the present value of government spending that is orthogonal to the forecast revision of the present value of GDP.[21] This makes virtually no difference to the results (not shown).

7 Conclusions: reconciling the evidence

Two conclusions can be drawn from the analysis so far. First, there is no robust evidence that truly unanticipated shocks to defense or federal spending in EVARs have negative effects on GDP, consumption or the real wage; in fact there is some evidence of positive, although not large, effects, which can be large in the case of the real wage. Second, EVARs and SVARs, once specified to make them comparable, give essentially the same answer.

---

[21]Formally, this is obtained by ordering the forecast revision of the present value of GDP first in the Choleski decomposition, the forecast revision of the present value of federal spending second, and by estimating responses to the federal spending shock thus orthogonalized.

The first conclusion appears to contradict a large number of SVAR studies, who find effects on GDP and consumption that are not just positive, but large. The second conclusion appears to contradict the notion that, if there is fiscal foresight, EVARs and SVARs estimate two different things, and the latter does not estimate anything structural.

How does one reconcile these seeming contradictions? Consider the first. There are three major differences with most existing SVAR studies (see e.g. Blanchard and Perotti 2002 or Galí et al 2007): (i) they typically start from 1954:1 or later, thus leaving out WWII and the Korean War; (ii) their government spending variable is total spending on goods and services, not just defense or federal spending; (iii) they display one standard error bands instead of two, thus often giving the impression of more significant responses.

Column 1 of Figure 6 displays estimates from a SVAR that uses total spending on goods and services, starting in 1954:1. Now there is consistent evidence of expansionary effects on GDP (whose peak effect is about 1), consumption (the figures displays also the effects on total consumption and consumption of nondurables and services, that are typically displayed in the existing literature), and the real wage; at peak, these effects are significant at the e68 percent but in general not at the 95 percent level, even though often with long delays of up to three years.

When I estimate a SVAR on the same sample starting in 1954:1, but with total government spending split into federal and state and local spending (columns 2 and 3 of Figure 6), it is clear that it is state and local spending that has positive effects on GDP, consumption, and the real wage: these effects are large and statistically significant.[22]

Of course impulse responses to a shock to state and local spending are subject to the usual criticism of fiscal foresight. The corresponding EVAR, based on SPF forecast errors of state and local government spending, can only be estimated starting in 1981:3. Δs_t^u and $\Delta s_{t/t-1}^e$ have the same properties as the corresponding variables for federal spending: their correlation is minimal, .04; and Δs_t^u has an enormous explanatory power for Δs_t, with an F-statistics 1214 (see Table 5). Like before, part of the explanation is again the very low explanatory power of $\Delta s_{t/t-1}^e$ itself (see row 3).

The standard errors of the resulting EVAR are extremely large, and nothing informative can be obtained from these regressions. The same applies to a SVAR estimated over the same sample that starts in 1981:3.

Thus, when estimated using the same specification and over the same sample, and whenever there is enough variation to provide meaningful estimates, SVARs and EVARs appear to give the same answer, a fact unnoticed by Ramey (2011) because she never estimates the same specification over the same sample. One possible conclusion from this observation is

---

[22]In each case, I order first the variable that is being shocked.

Table 5: **Explanatory power of forecasts and forecast errors for Δs_t**

	(1)	(2)	(3)	(4)	(5)
	indep. var.	R^2	F-stat	coeff. of lag 0	t-stat
1	Δs_t^u	.97	390.54	.96	39.57
2	Δs_t^u	.94	1214.47	1.01	35.98
3	$\Delta s_{t/t-1}^e$	.31	3.39	.19	.54

See the text for the definition of Δs_t^u and $\Delta s_{t/t-1}^e$.
Row 1: regression of Δs_t on lags 0 to 4 of Δs_t^u and on lags 1 to 4 of s_t, g_t, y_t, i_t, and τ_t.
Row 2: regression of Δs_t on lags 0 of Δs_t^u and on lags 1 to 4 of s_t, g_t, y_t, i_t, and τ_t.
Row 3: regression of Δs_t on lags 0 to 4 of $\Delta s_{t/t-1}^e$ and on lags 1 to 4 of s_t, g_t, y_t, i_t, and τ_t.
Column 2: R^2 of the regressions.
Column 3: marginal F-statistics for the exclusion of lags 0 to 4 of Δs_t^u (row 1), of lag 0 of Δs_t^u (row 2), and of lags 0 to 4 of $\Delta s_{t/t-1}^e$ (row 3).
Column 4: coefficient of lag 0 of Δs_t^u or $\Delta s_{t/t-1}^e$.
Column 5: its t-statistics.

that SVARs can be trusted even in those cases where they do not have an EVAR counterpart for lack of data. This includes other countries, that do not have enough variation in defense spending to construct a defense news variable, or responses to total or state and local government spending over the post-Korean war sample in the US, which show a large positive effect on GDP, consumption and the real wage. Others might still distrust the evidence from these SVARs.

Now consider the second seeming contradiction: an EVAR estimates a different object than a SVAR, yet they deliver almost identical impulse responses. From (12), the SVAR bias in estimating π_0 (and the equivalent for consumption) depends on the variance of $a_{t/t-1}$. If this variance is small, so is the bias in estimating the parameters of the impulse response. In the context of this specific model, if the parameters are estimated consistently, one should see a negative effect on output and consumption. However, the general point is that if the variance of the forecast revisions is small, then omitting the latter does not generate a large bias in a SVAR. This is consistent with several papers (see e.g. Forni, Gambetti and Sala 2011) showing that news shocks have small variance and have a limited role in explaining business cycle fluctuations.

In addition, note that if $\Delta f_{t/t-1}^e$ is essentially noise as we have seen, the forecast error Δf_t^u is just Δf_t plus noise, hence including $.\Delta f_t^u$ in the VAR is redundant and will generate

the same impulse responses as when Δf_t only is included.

Thus, the results of the present paper are consistent with those of Chahrour, Schmitt-Grohé and Uribe (2010), who generate the data from a DSGE model in which part of the shocks (to taxation) are anticipated, and show that a SVAR à la Blanchard and Perotti (2002) displays minimal bias. The results of this paper are also consistent with Forni and Gambetti (2011), who show that a SVAR that is affected by the non-fundamentalness problem delivers impulse responses to government spending shocks that are very similar to those of a large dynamic factor model that is immune to this problem.

Appendix

Consider the model

$$maxE_t \sum_{i=0}^{\infty} \beta^i logC_{t+i} \tag{18}$$

$$C_t + K_t + G_t = Y_t \tag{19}$$

$$Y_t = Z_t K_{t-1}^{\alpha} \tag{20}$$

$logZ_t$ is normal with mean 0 and variance σ_z^2. The Euler equation is

$$1 = E_t \left[\beta(1 + R_{t+1})C_t/C_{t+1} \right] \tag{21}$$

Assuming for simplicity $(1 + R_{ss}) = 1/\beta$, and using $K_{ss}^{1-\alpha} = \alpha\beta$ from the Euler equation, log linearization of the resource constraint gives

$$\alpha\beta k_t + \widetilde{G}g_t + (1 - \alpha\beta - \widetilde{G})c_t = z_t + \alpha k_{t-1} \tag{22}$$

where small letters denote log deviations from the non-stochastic steady state. Hence

$$c_t = \frac{1}{\xi} \left[-\alpha\beta k_t - \widetilde{G}g_t + z_t + \alpha k_{t-1} \right] \tag{23}$$

where

$$\xi = (1 - \alpha\beta - \widetilde{G}) \tag{24}$$

Log linearization of the rhs of the Euler equation gives

$$c_t - E_t c_{t+1} + E_t z_{t+1} + (\alpha - 1)k_t = 0 \tag{25}$$

Replacing for c_t and c_{t+1} from (23) in the log-linearized Euler equation gives

$$E_t k_{t+1} - (\alpha\beta)^{-1}(\alpha\beta + \alpha + \xi(1-\alpha))k_t + \beta^{-1}k_{t-1} = (\alpha\beta)^{-1}(1-\xi)E_t z_{t+1} - (\alpha\beta)^{-1}z_t \tag{26}$$
$$-(\alpha\beta)^{-1}\widetilde{G}E_t g_{t+1} + (\alpha\beta)^{-1}\widetilde{G}g_t$$

Replacing $E_t g_{t+1}$ with $a_{t+1/t}$ and $E_t z_{t+1} = 0$ gives

$$E_t k_{t+1} - \omega k_t + \beta^{-1}k_{t-1} = (\alpha\beta)^{-1}z_t - (\alpha\beta)^{-1}\widetilde{G}(a_{t+1/t} - a_{t/t-1}) + (\alpha\beta)^{-1}\widetilde{G}\varepsilon_t \tag{27}$$

$$\omega \equiv (\alpha\beta)^{-1}(\beta\alpha + \alpha + \xi(1-\alpha)) \tag{28}$$

Now consider the characteristic equation

$$\lambda^2 - \omega\lambda + \beta^{-1} = 0 \tag{29}$$

Let

$$\lambda = \frac{\omega \pm \sqrt{\omega^2 - 4\beta^{-1}}}{2} \tag{30}$$

A sufficient condition for $\lambda_1 < 1$ and $\lambda_2 > 1$ is that $\widetilde{G}$ should not be "too large". Under this condition, the solution for k_t is

$$k_t - \lambda_1 k_{t-1} = (\alpha\beta)^{-1}\phi \sum_{i=0}^{\infty} \phi^i E_t z_{t+i} - \frac{\widetilde{G}}{\alpha\beta}\left[\phi a_{t/t-1} + \phi^2 a_{t+1/t} - \phi a_{t+1/t} + \phi\varepsilon_t\right] \tag{31}$$

where $\phi = 1/\lambda_2$. Hence we get equation (4) in the text

$$k_t = \lambda_1 k_{t-1} + \delta z_t + \pi_0 a_{t/t-1} + \pi_1 a_{t+1/t} + \pi_0\varepsilon_t \tag{32}$$

where

$$\pi_0 = -\frac{\widetilde{G}}{\alpha\beta}\phi \ < 1 \qquad \pi_1 = -\pi_0(1-\phi); \qquad \delta = -\frac{\phi}{\alpha\beta(1-\phi)} \tag{33}$$

Now suppose one estimates

$$k_t = \lambda_1 k_{t-1} + \eta_{k,t}^S \tag{34}$$

where

$$\eta_{k,t}^S = \delta z_t + \pi_0 a_{t/t-1} + \pi_1 a_{t+1/t} + \pi_0\varepsilon_t \tag{35}$$

Note that k_{t-1} is correlated with $\eta_{k,t}^S$, and specifically with $a_{t/t-1}$, hence

$$\widehat{\lambda}_1 = \lambda_1 + \pi_0 \frac{cov(a_{t/t-1}, k_{t-1})}{var(k_{t-1})} \tag{36}$$

But since all the a's are independent of each other

$$\frac{cov(a_{t/t-1}, k_{t-1})}{var(k_{t-1})} = \pi_1 \frac{var(a_{t/t-1})}{var(k_{t-1})}; \tag{37}$$

and

$$\widehat{\eta}_{k,t}^{S} = k_t - \widehat{\lambda}_1 k_{t-1} \tag{38}$$

$$= (\lambda_1 - \widehat{\lambda}_1)k_{t-1} + \eta_t^k \tag{39}$$

$$= -(\pi_0 \pi_1)\frac{var(a_{t/t-1})}{var(k_{t-1})}k_{t-1} + \eta_{k,t}^{S} \tag{40}$$

Now suppose one applies the Blanchard - Perotti approach and regresses $\widehat{\eta}_{k,t}^{S}$ on $\widehat{\eta}_{g,t}^{S} = a_{t/t-1} + \varepsilon_t$. The coefficient is given by

$$\widehat{\pi}_0^{S} = \frac{cov(\widehat{\eta}_{k,t}^{S}, \widehat{\eta}_{g,t}^{S})}{var(\widehat{\eta}_{g,t}^{S})} \tag{41}$$

$$= -(\pi_0 \pi_1)\frac{var(a_{t/t-1})}{var(k_{t-1})}\frac{cov(k_{t-1}, a_{t/t-1})}{var(a_{t/t-1} + \varepsilon_t)} + \pi_0 \frac{var(a_{t/t-1} + \varepsilon_t)}{var(a_{t/t-1} + \varepsilon_t)} \tag{42}$$

$$= \pi_0 \left[1 - \pi_1^2 \frac{var(a_{t/t-1})}{var(k_{t-1})}\frac{var(a_{t/t-1})}{var(a_{t/t-1} + \varepsilon_t)}\right] \tag{43}$$

References

Auerbach, A. and Y. Gorodnichenko (2010): "Measuring the Output Responses to Fiscal Policy ", NBER Working Paper No. 16310, forthcoming, *Amercian Economic Journal, Macroeconomics*

Barro, R. and C. J. Redlick (2011): Macroeconomic effects from government purchases and taxes", *Quarterly Journal of Economics*

Bilbiie, F. (2011): "Non-Separable Preferences, Frisch Labor Supply and the Consumption Multiplier of Government Spending: One Solution to a Fiscal Policy Puzzle", forthcoming in *Journal of Money, Credit and Banking*

Blanchard, O. and R. Perotti (2002): "An Empirical Characterization of the DynamicEffects of Changes in Government Spending and Taxes on Output" *Quarterly Journal of Economics*, 117 (2002), 1329–1368.

Chahrour, R., S. Schmitt-Grohè and M. Uribe (2010): "A model-based evaluation of the debate on the size of the tax multiplier", NBER WP No. 16169, forthcoming, *American Economic Journal, Macroeconmics*

Fatas, A. and I. Mihov (2001): "The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence" CEPR Discussion Paper No. 2760

Favero, C. and F. Giavazzi (2007): "Debt and the Effects of Fiscal Policy", NBER Working Paper No 12822

Fernandez-Villaverde J., J.F. Rubio-Ramrez, T.J. Sargent and M.W. Watson (2007): "ABCs (and Ds) of Understanding VARs". *American Economic Review* 97(3):1021-1026.

Forni, M. and L. Gambetti (2011): "Fiscal foresight and the effects of government spending", mimeo, University of Modena and Reggio Emilia

Forni, M., L. Gambetti, and L. Sala (2011): "No news in business cycles", mimeo, Bocconi University

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 (2007), 227–270

Gordon, R. and R. Krenn (2010): "The end of the Great Depression 1939-41: policy contributions and fiscal multipliers", NBER WP No. 16380

Hall, R. (2009): "By How Much Does GDP Rise if the Government Buys More Output?", *Brookings Papers on Economic Activity*, Fall, pp. 183-231

Higgs, R. (1992): "Wartime Prosperity? A Reassessment of the U.S. Economy in the 1940s", Journal of Economic History, 52(1), 41-60

Leeper, E.M., Walker, T.B. and S.S. Yang (2008): "Fiscal Foresight: Analytics and Econometrics", NBER Working Paper No. 14028.

Lippi, M. and L. Reichlin (1994). "VAR analysis, non fundamental representation, Blaschke matrices", *Journal of Econometrics* 63, 307-325

Mertens, K. and M. O. Ravn (2010): "Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach" *Economic Journal*, vol. 120(544), pages 393-413, 05.

Monacelli, T., R. Perotti, and A. Trigari (2010): "Unemployment Fiscal Multipliers", *Journal of Monetary Economics*

Perotti, R. (2007): "In Search of the Transmission Mechanism of Fiscal Policy," *NBER Macroeconomics Annual*, 22, 169–226

Ramey, V. (2011): "Identifying government spending shocks: It's all in the timing", *Quarterly Journal of Economics*, 1-50

Ravn, M., S. Schmitt-Grohé and M. Uribe (2006): "Deep habits", *Review of Economic Studies*, vol. 73(1), pp. 195-218
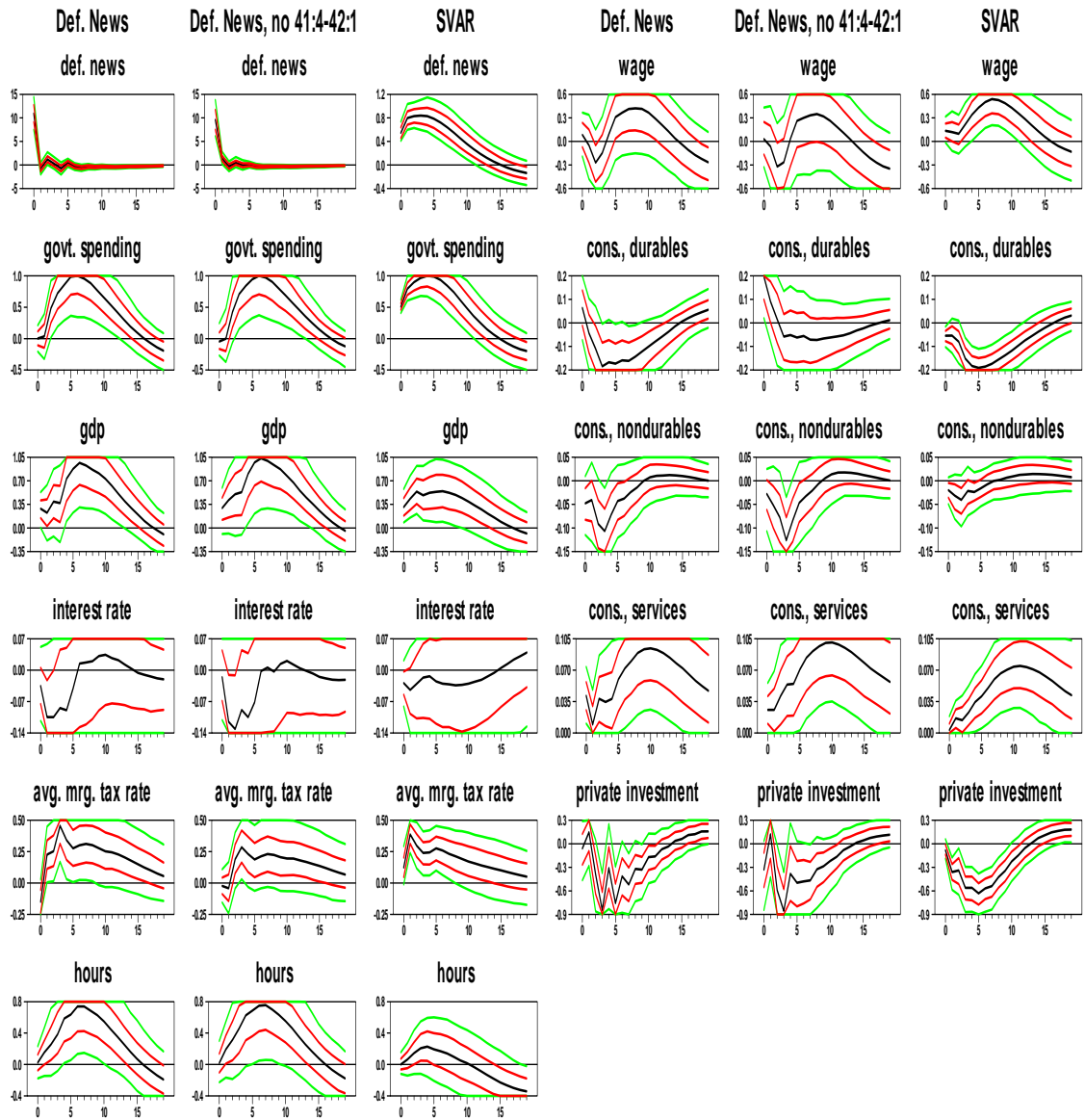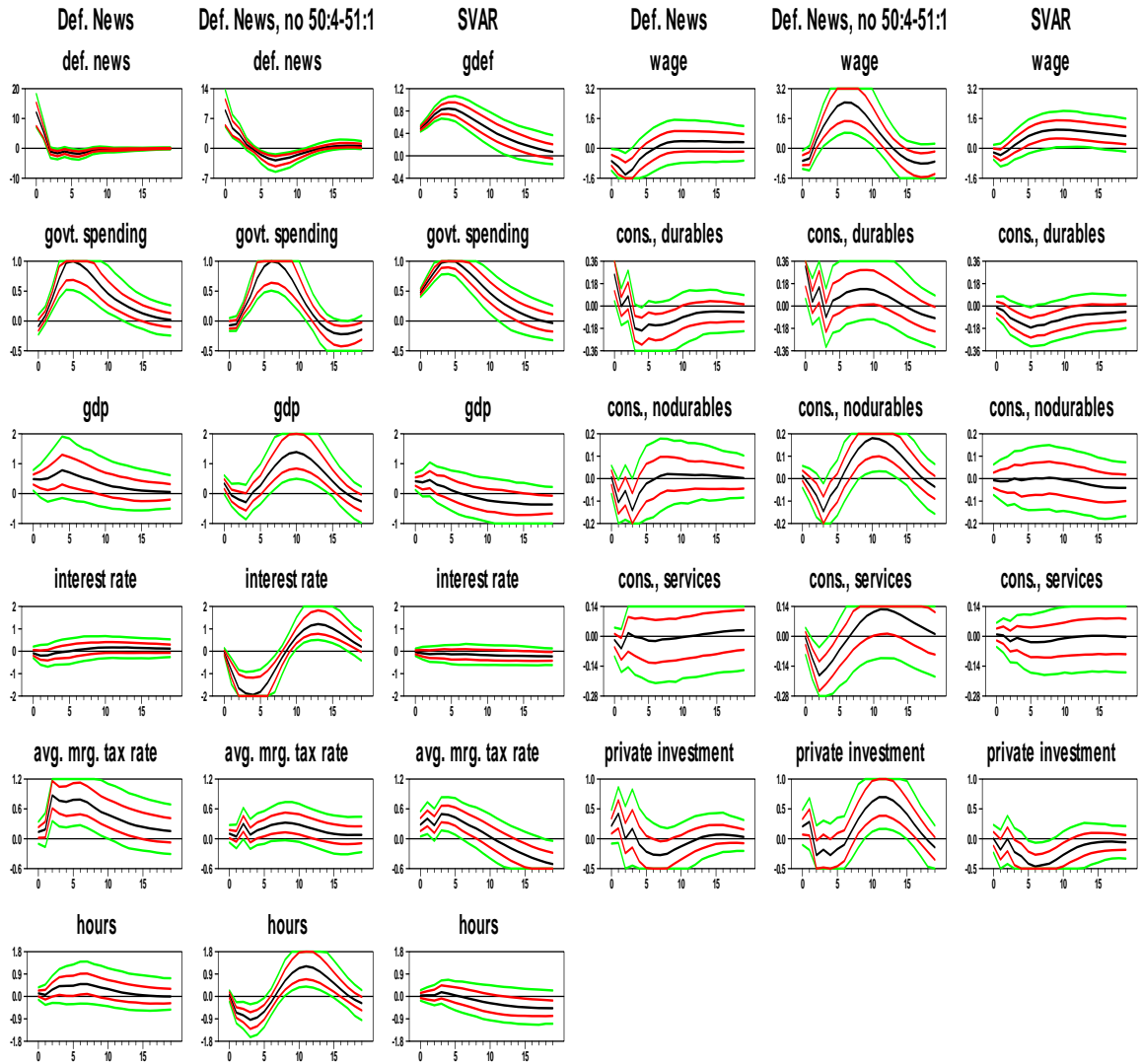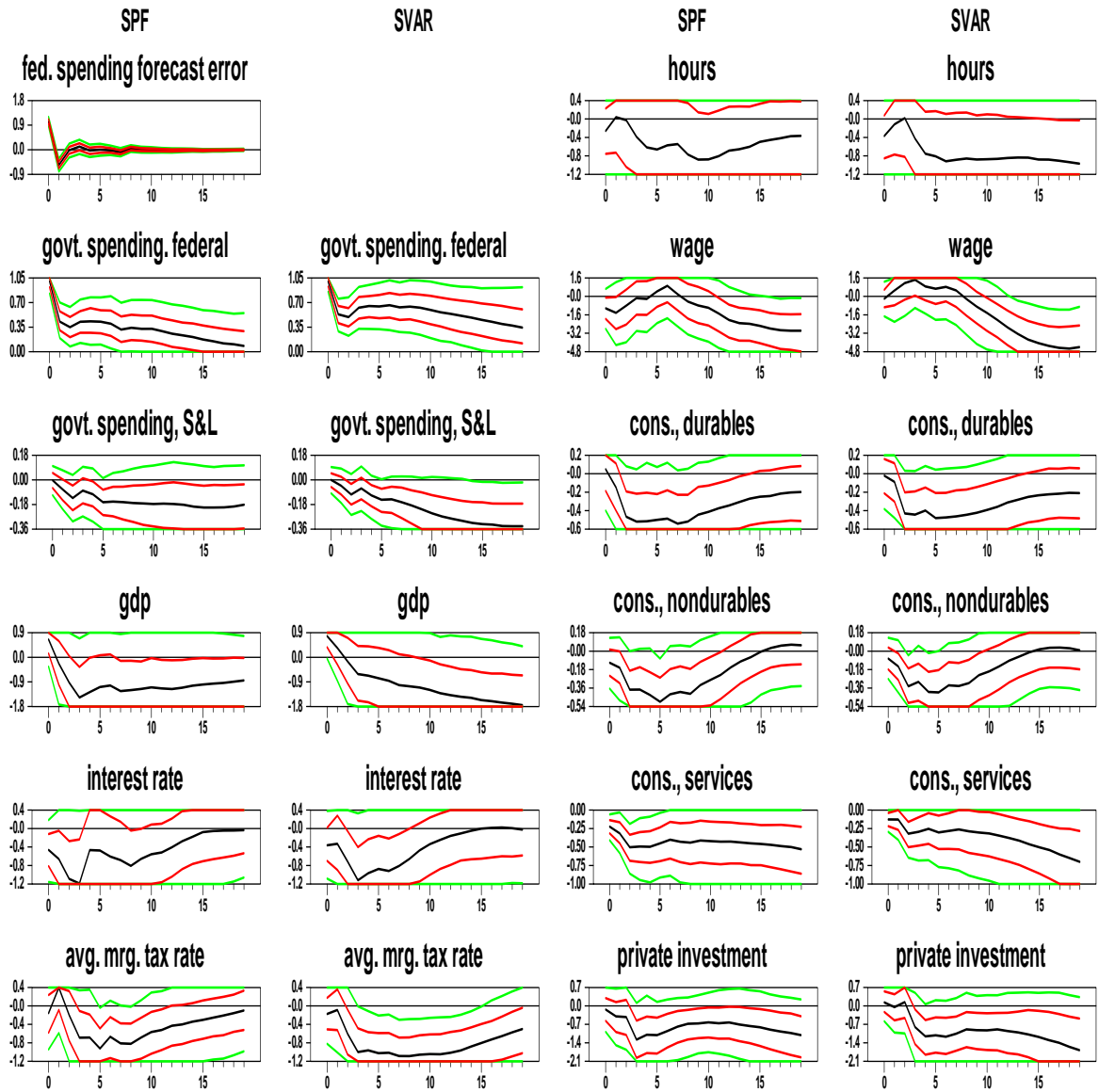
Figure 1: WWII

Figure 2: Korea

Figure 3: Survey of Professional Forecasters, federal spending shocks
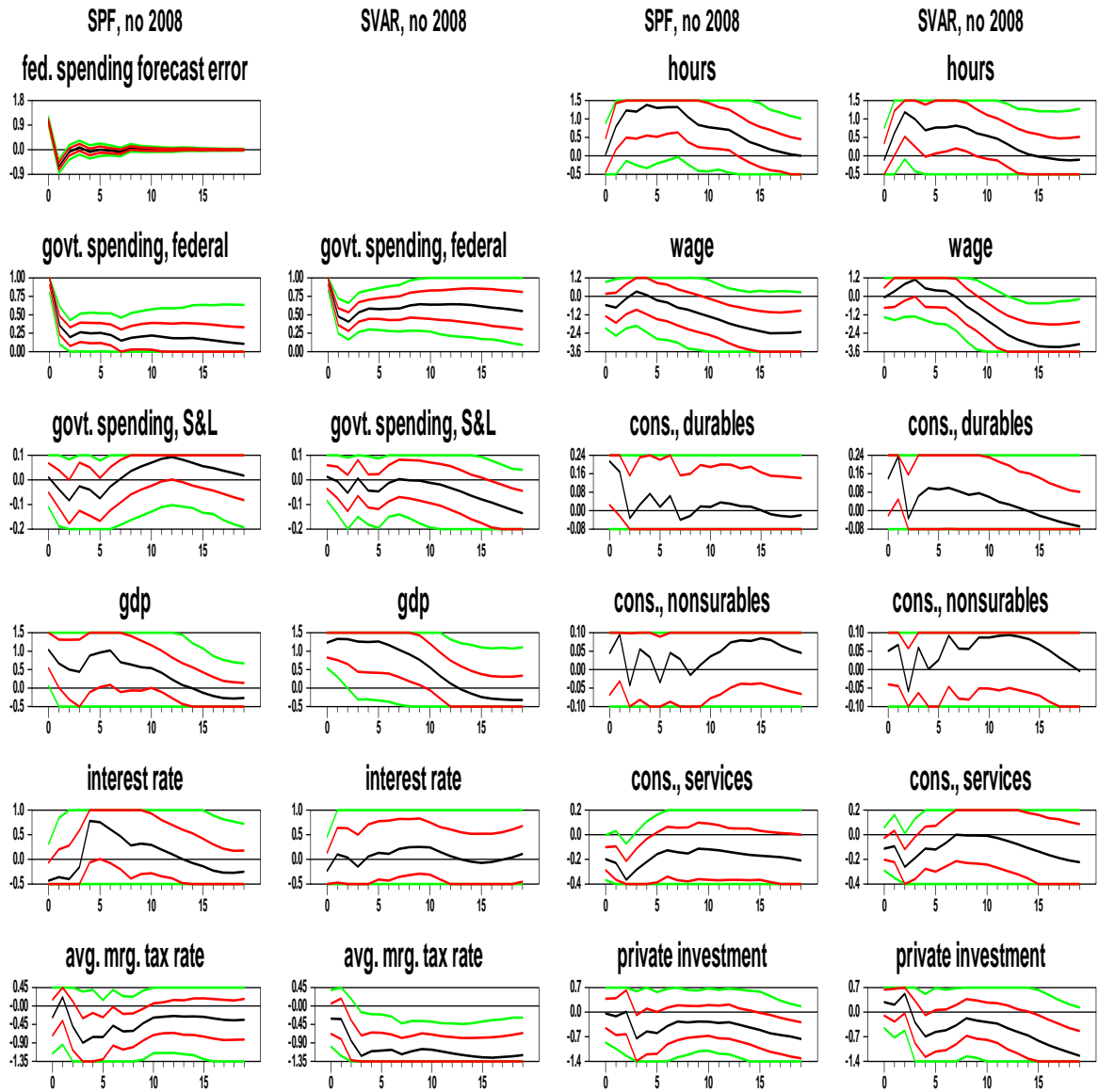
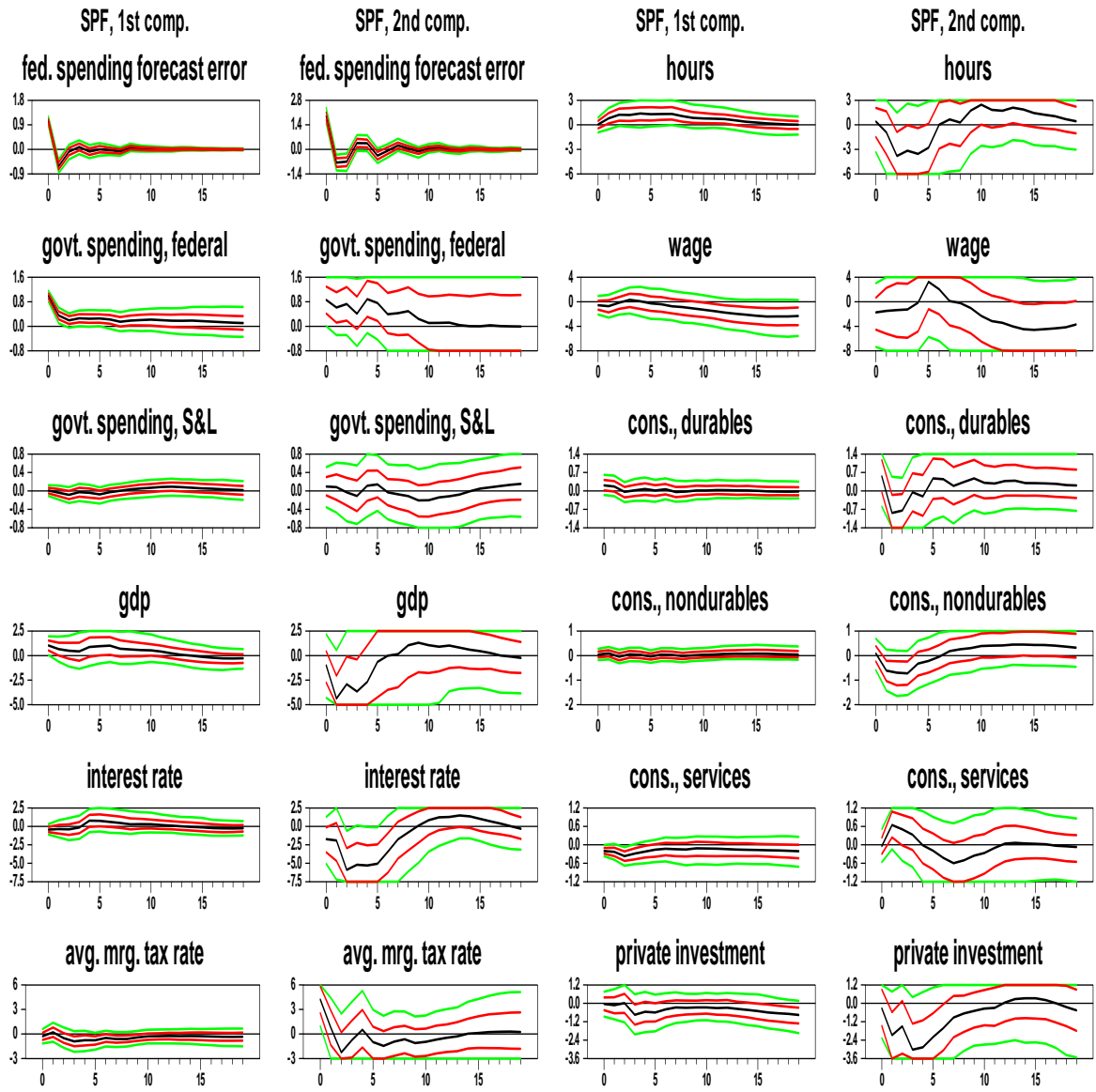Figure 4: Survey of Professional Forecasters, federal spending shocks, excluding 2008
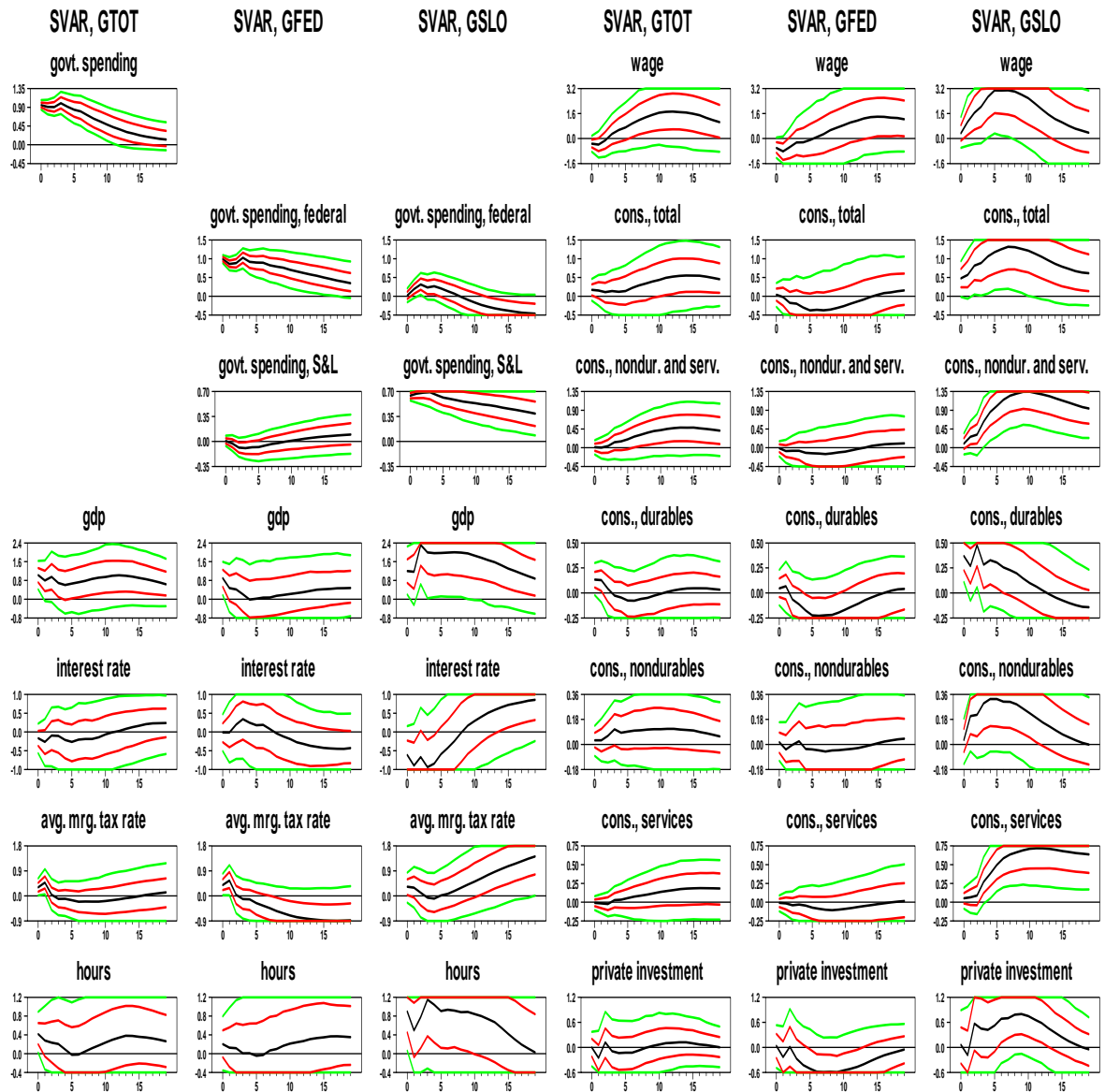
Figure 5: Survey of Professional Forecasters, decomposition of forecast error

Figure 6: SVARs, starting 1954:1