A Reply to Roberto Perotti's "Expectations and Fiscal Policy: An Empirical Investigation"

Valerie A. Ramey
University of California, San Diego
and NBER

June 30, 2011

Abstract

This brief note challenges the claims made by Roberto Perotti in his recent paper comparing structural VARs to expectational VARs.

I thank Neville Francis for helpful comments.

1 Introduction

Robustness is always a concern in empirical work, particularly in time series analysis where sample periods and seemingly small changes in specification can lead to different results. Thus, papers that analyze robustness are an important part of the literature. Perotti's (2011) recent paper analyzes how results differ across two key types of VARs and how inserting dummy variables for key quarters can sometimes change results. Based on presenting some 200 impulse response functions, Perotti (2011) makes several claims about the results in Structural VARs (SVARs) and what he calls "Expectational VARs" (EVARS). This brief note challenges those claims.

Perotti (2011) begins by claiming that SVARS and EVARS give the same results and that my previous work (Ramey (2011a)) never estimated both systems on the same sample with the same data. I will argue that both parts of this claim are incorrect. Second, the paper claims that including a few key dummy variables eliminates the negative effects of government spending. I will show that his paper uses these dummy variables in a manner that is inconsistent with his narrative, and that when they are used consistently, they do not differ from the full sample results. Third, the paper questions the construction of government spending shocks using professional forecast errors. Perotti conducts an interesting decomposition of the forecast errors. The theoretical analysis quite intriguing; however, inspection of the results from the two specifications he presents shows that there is no statistically significant difference.

2 SVARs vs. Defense News EVARS

In his latest contribution to the long-running debate on the methodology of analyzing government spending, Perotti (2011) claims that structural VARS and expectational VARs give the same results. Such a result is potentially helpful because it means that researchers can simply use standard times series techniques on familiar macro series rather than having to gather more data. Specifically, Perotti argues that SVARS with shocks to defense spending give the same results as my VARs using news, but both of those results differ from SVARs with shocks to total government spending. Referring to my recently published paper in the *QJE*, Perotti (2011) contends: "Ramey reaches the opposite conclusion ... because she never estimates the two specifications on the same sample and with the same government spending variable." (Perotti (2011), p. 4). I will now show that all parts of these claims are incorrect.

The entire Section III of my paper is dedicated to comparing SVARS and EVARs using the same data and the same sample. In that analysis, I used the original Ramey-Shapiro military date variable (Ramey and Shapiro (1998)) to maintain comparability with the literature. I did not conduct the same comparison with the richer news variable created later in the paper because I did not want to add to an already long paper. (The published paper in the *Quarterly Journal of Economics* is 50 pages.) I will

show these additional results below and they will reveal that Perotti's claims are not correct. It appears that Perotti's claims are based on the mistaken notion that direct comparison requires one to use orthogonalized shocks to defense spending rather than to total government spending. This is not correct. An SVAR can always be interpreted as an instrumental variables (IV) regression. Viewed in this context, the exercise I performed consists of comparing two instruments for total government spending, the first instrument being the VAR shock to total government spending using a Choleski decomposition and the second being the shock to the military date variable. Defense spending is not included in either VAR. In both VARs, I compare the response of all variables when the peak rise in total government spending has been normalized to the same number. Since there is no significant feedback of other variables to the news variable, in essence I am simply comparing the effects of the same size increase in total government spending on variables of interest, using two different instruments for the *same measure of government spending*.

Even if his econometric argument were correct, Perotti's claims that shocks to defense spending in an SVAR give different results than shocks to total government spending in an SVAR are simply not correct. Despite presenting 200 impulse response functions (surpassed only by the 500 impulse response functions presented in Perotti (2008)), Perotti never conducts a direct comparison of the two types of SVARs that he claims give different results.

I suspect that Alexander Pope's aphorism about words and leaves also applies to impulse response functions, so I am reluctant to present yet more impulse response functions. However, I do wish to set the record straight, so I will present just as many as I need to show that Perotti's claims are not borne out. Therefore, I will present the impulse response functions for the key variables of interest for three specifications, over three main samples. In contrast to Perotti, I will present the point estimates for all three specifications *on the same graph* so that comparison is facilitated. As discussed in Section V of my published paper, the standard VAR includes log per capita values of real government spending and GDP, the 3-month treasury bill rate, the average marginal tax rate of Barro and Redlick (2011), and another variable that is rotated in. Total hours is the variable used for the government spending and GDP results. In the other cases, the variable shown on the graph is the last variable. The standard SVAR orders total government spending first and uses a Choleski decomposition to identify shocks to it. The EVAR augments the system to include news about the expected discounted value of government spending and uses shocks to this variable, ordered first. Perotti's defense spending SVAR replaces my news variable with defense spending and uses a Choleski decomposition to identify shocks to defense, ordered first. Perotti claims that the results differ across the first and third specification, yet he only estimates the second and third specification.

In my published paper, I highlighted several key differences across the EVAR vs. SVAR specifications: (1) government spending rose much earlier in the SVAR specification; (2) the peak of GDP was much higher in the EVAR than in the SVAR, implying a

greater multiplier from the EVAR; (3) consumption components tended to fall in EVAR (though services rose for the period that included WWII), but rise in the SVAR; and (4) in post-WWII samples, real product wages decreased in the EVAR but increased in the SVAR. Figure 1 shows the results from the 1939-2008 sample, Figure 2 shows the results from 1947-2008, and Figure 3 shows the results from 1954-2008. All shocks have been normalized so that the increase in government spending at its peak is equivalent to one percent of GDP. The consumption variable responses have also been converted to percent of GDP.[1] The standard error bands are shown for the EVAR, and are based on bootstrap standard errors. I will now highlight seven results.

1. In 23 of 24 cases, the two SVAR results are numerically very close. Thus, whether the standard VAR shock is identified with respect to defense spending or to total government spending does not matter. (The only exception is nondurable consumption in the 1939-2008 sample.) Hence, Perotti's contention that using shocks to defense spending in an SVAR leads to fundamentally different results than using shocks to total government spending in an SVAR is simply incorrect. It appears that he came to this conclusion without ever comparing both sets of results.

2. All three samples continue to show that the response of government spending to a news shock is delayed, but is instantaneous with respect to an identified SVAR shock to either government spending or defense spending. These results are consistent with my news hypothesis and the simulations from a theoretical model, shown in the working paper version of my paper (Ramey (2009)).

3. In the 1939-2008 and the 1947-2008 samples, the results continue to show a much bigger response of GDP to government spending in the EVAR than in either SVAR specification. Again, these results are consistent with my news hypothesis and the simulations from the theoretical model. As discussed in my published paper, the response of GDP is estimated to be negative most of the time in the 1954-2008 sample, but the standard error bands are exceedingly wide.

4. Nondurables consumption falls in all three samples in the EVAR. It also falls for the two SVARs in the WWII sample, but rises slightly in the two post-war samples. All three specifications show a rise in services consumption when WWII is included and somewhat erratic behavior for 1947-2008. In 1954-2008, the EVAR shows a significant fall in services consumption, whereas the SVAR specifications show little change.

5. Durable consumption continues to show the spike upward followed by a decline

1. The conversion factors are based on government spending averaging 21 percent of GDP, nondurable consumption averaging 21 percent of GDP, services consumption averaging 34 percent, and durable expenditures averaging 9 percent. Note that the results shown in my published paper are percent changes in the variable itself, not as a percent of GDP.

in the news specifications that include the Korean War; the SVARS do not. As discussed in my paper, there was panic buying of durable goods as soon as North Korea invaded South Korea, *a feature of the Korean War period that is completely missed by both SVAR specifications*.

6. Real wages always rise in the SVAR specifications. In contrast, the news specification typically shows declines.

7. The standard errors are big on most responses. The same is true for the ones computed using SVARs. As Perotti's paper shows, the SVAR response of nondurable and services consumption in the post-WWII is never statistically different from zero. The literature is still recovering from the fact that for years many researchers claimed statistical significance based on one-standard deviation error bands.

To summarize, Perotti's claim that using SVAR shocks to defense spending rather than SVAR shocks to total government spending leads to different results is not true. On the other hand, the difference in point estimates between the EVAR and the SVAR specifications continues to hold. However, most responses are estimated imprecisely, so statistical significance is questionable.

3 Perotti's Dummy Variables

When discussing the presumed similarities of standard SVARS and EVARs, Perotti also argues that results change if he simply "dummies" out key quarters. He argues that several quarters during WWII should be dummied out because of rationing and that several quarters during the Korean War should be dummied out because of Regulation Q and Regulation W. Let's consider his arguments and methods in more detail.

The first point that should be noted is that a comparison of his selected quarters for dummying out line up with some of the largest news shocks in my data. Thus, significant information is being lost if these quarters are eliminated. Second, if one believes the structure of the economy is truly different in these quarters, then one should also account for this in the lag coefficients of the VAR. My understanding is that Perotti simply included dummy variables in the constant term in the quarters in question.

For WWII, Perotti advocates dummying out the fourth quarter of 1941 and the first quarter of 1942 because of rationing. However, his own Table 1 shows that rationing did not start for any goods until the first and second quarters of 1942. As it turns out, the fourth quarter of 1941 had no rationing, but because of Pearl Harbor, it contains the biggest news shock of the sample (as a percent of GDP) based on my narrative

method.[2] If one takes his rationing argument seriously, then the quarters that should be dummied out are 1942:1 and 1942:2, not 1941:4 and 1942:1.

Perotti also makes various arguments for why 1950:3, 1950:4, and 1951:1 should be dummied out. I am not persuaded that these arguments justify omitting quarters with so much information. Nevertheless, even if one accepts these arguments, one is struck by the fact that (1) Perotti does not use these dummies for his full 1939-2008 sample results; and (2) for some unknown reason, he only shows results with only 1950:4 and 1951:1 dummied out. Thus, there appear to be several inconsistencies in the way he translates his narrative arguments into his estimation.

What happens if Perotti's dummy variable method, whose econometric basis remains dubious, is at least applied more consistently? In particular, consider using the full sample from 1939 - 2008 and using dummies that match up with Perotti's rationing table and all three quarter for he gives arguments during the Korean War. Thus, instead of using Perotti's dummies for 1941:4 and 1942:2 in one sample and 1950:4 and 1951:1 in another sample, suppose that we instead use dummies for the two quarters where the most rationing took effect according to Perotti's Table 1, 1942:1 and 1942:2, as well as the three quarters that Perotti thinks are suspect during the Korea War: 1950:3, 1950:4 and 1951:1.

Figure 4 compares my original results that use no dummy variables to results with the five quarters dummied out. In every case, the responses using the dummy variables are within two standard deviations of the original responses. Thus, simply applying Perotti's dummy variable methods in a way that is consistent with his narrative reverses his conclusions.

4 Questions about Forecast Errors

The final part of his paper examines my use of the Survey of Professional Forecasters to construct forecast errors. Because the defense news variable did not pass my first-stage test for samples starting after 1954, I used forecasts from the Survey of Professional Forecasters to construct news shocks to analyze the post-Korean War period. Perotti makes two arguments against these results. One concerns the inclusion of 2008 and the other concerns the nature of the forecast errors. I will discuss each briefly.

On the first point. Perotti argues that dropping 2008 changes the results. It should be noted that Perotti chose to limit his Survey of Professional Forecasters sample to 1981 to 2008, whereas I used 1969 to 2008, so 2008 might be more influential for his sample.[3] Perotti argues that 2008 is "so influential" because real federal spending was up 9 percent during the year but GDP fell. To investigate this claim, Figure 5

2. Perotti argues that supply bottlenecks reduced consumption even before rationing took effect. However, he does not seem to realize that "supply bottlenecks" are a manifestation of "crowding out." According to the neoclassical view, increased government spending competes with private demand; bottlenecks that reduce the supply available for private consumption are part of this mechanism.

3. My published paper discusses in detail the motivation for using the entire sample back to 1969

shows quarterly growth rates of real federal spending, total government spending, and real GDP, as well as a scatter graph of GDP vs. real federal spending growth. Perotti argues that the last four observations are "influential" and should be dropped. Simple inspection of the graphs shows that the growth rates of the two government spending variables and GDP are not outliers for this sample. For example, federal spending rose significantly several times in the 1970s, 1980s, and after 9/11. The rise in total government spending during 2008 was not particularly noteworthy. Real GDP did fall substantially during 2008, but it also did so in the mid-1970s and the early 1980s. In his discussion, it appears that Perotti advocates dropping 2008 because federal spending rose while GDP fell, i.e., that the 2008 observations are in the southeast quadrant of the lower right graph. His suggestion is highly questionable since dropping observations because *GDP and government spending happen to be moving in opposite directions* will surely bias results toward finding a large positive effect of government spending on GDP.

On the second point, Perotti decomposes the forecast errors and argues that it is the wrong component of my constructed forecast error that is driving the result. This analysis is quite intriguing and raises some interesting issues about the use of professional forecasts to construct forecasts errors. I agree completely that it would be useful to gather the real time data in order to check robustness of forecasts errors for levels versus growth rates. However, Perotti does not gather the necessary data and instead does an econometric decomposition of the growth forecasts errors using already available data. Rather than discuss in detail his argument about the components, I will direct the reader to his Figure 5 impulse responses. Note that whether the first or second component is used, not one of the responses of GDP, the consumption components, real wages, or investment is statistically significant from zero. Thus, it is not clear to me whether it really matters in the data.

5 Conclusion

This short note has made several points. First, contrary to Perotti's assertion, in my published paper I directly compared the two ways of identifying shocks on the same data in the same sample. Second, I showed that his proposed "correct" comparison yields results little different from my comparison. Third, I showed that his finding that dummy variables change key results disappears once the dummy variables are timed to be consistent with his own narrative. Fourth, I commented briefly on his intriguing Survey of Professional Forecasters critique.

Perotti's paper makes multiple arguments about the robustness of EVARs. Unfortunately, the paper does not conduct a systematic investigation of the robustness of SVARS, even though the stated goal of the paper is to provide support for use of standard SVARs. While robustness exercises are an important part of empirical work, I suspect that gathering new data rather than estimating models on the existing data

might have higher expected returns for understanding how government spending affects the economy. My recent forum piece (Ramey (2011b)) discusses many papers that do just this.

References

Barro, Robert J. and Charles J. Redlick, 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126(1): 51–102.

Perotti, Roberto, 2008. "In Search of the Transmission Mechanism of Fiscal Policy." In *NBER Macroeconomics Annual 2007*, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, pp. 169–226. University of Chicago Press.

Perotti, Roberto, 2011. "Expectations and Fiscal Policy: An Empirical Investigation." Technical report, Bocconi.

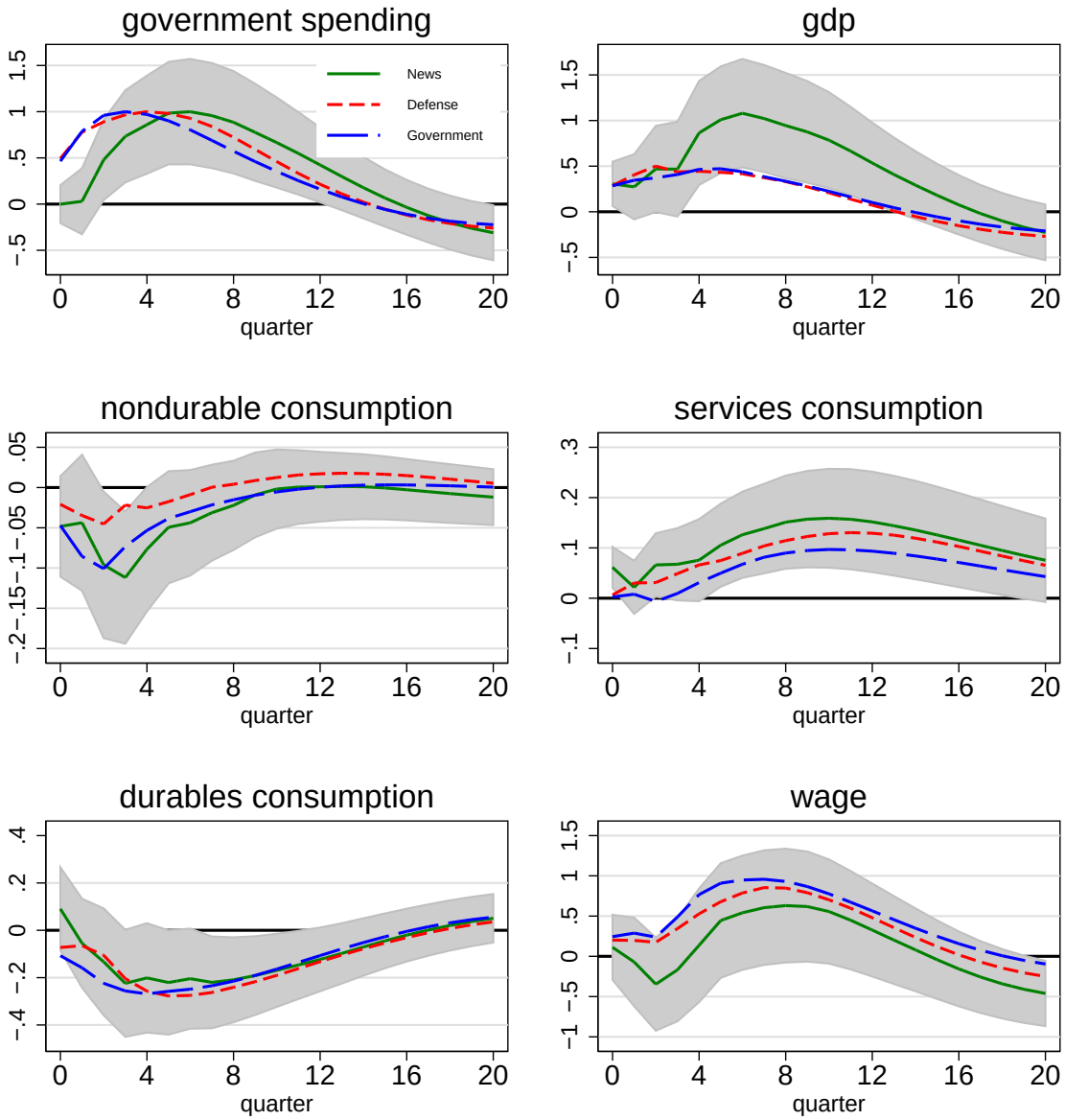
Ramey, Valerie A., 2009. "Identifying Government Spending Shocks: It's All in the Timing." Working paper 15464, National Bureau of Economic Research.

Ramey, Valerie A., 2011a. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economic* 126(1): 51–102.

Ramey, Valerie A., 2011b. "Can Government Purchases Stimulate the Economy?" Unpublished paper, University of California, San Diego.

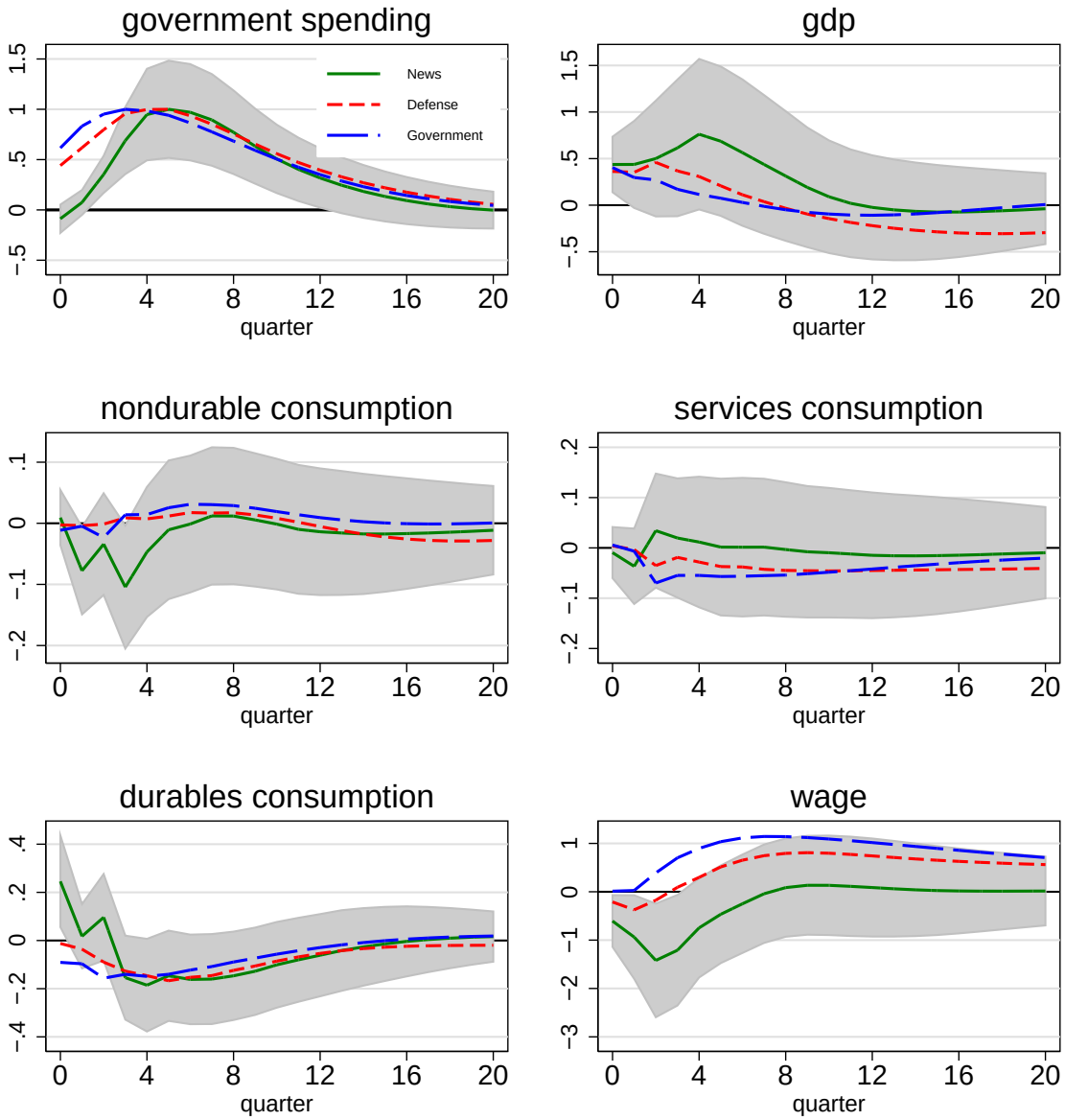
Ramey, Valerie A. and Matthew D. Shapiro, 1998. "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie-Rochester Conference Series on Public Policy* 48: 145–94.

Figure 1. Comparison of VAR Identification Schemes, 1939:1 - 2008:4



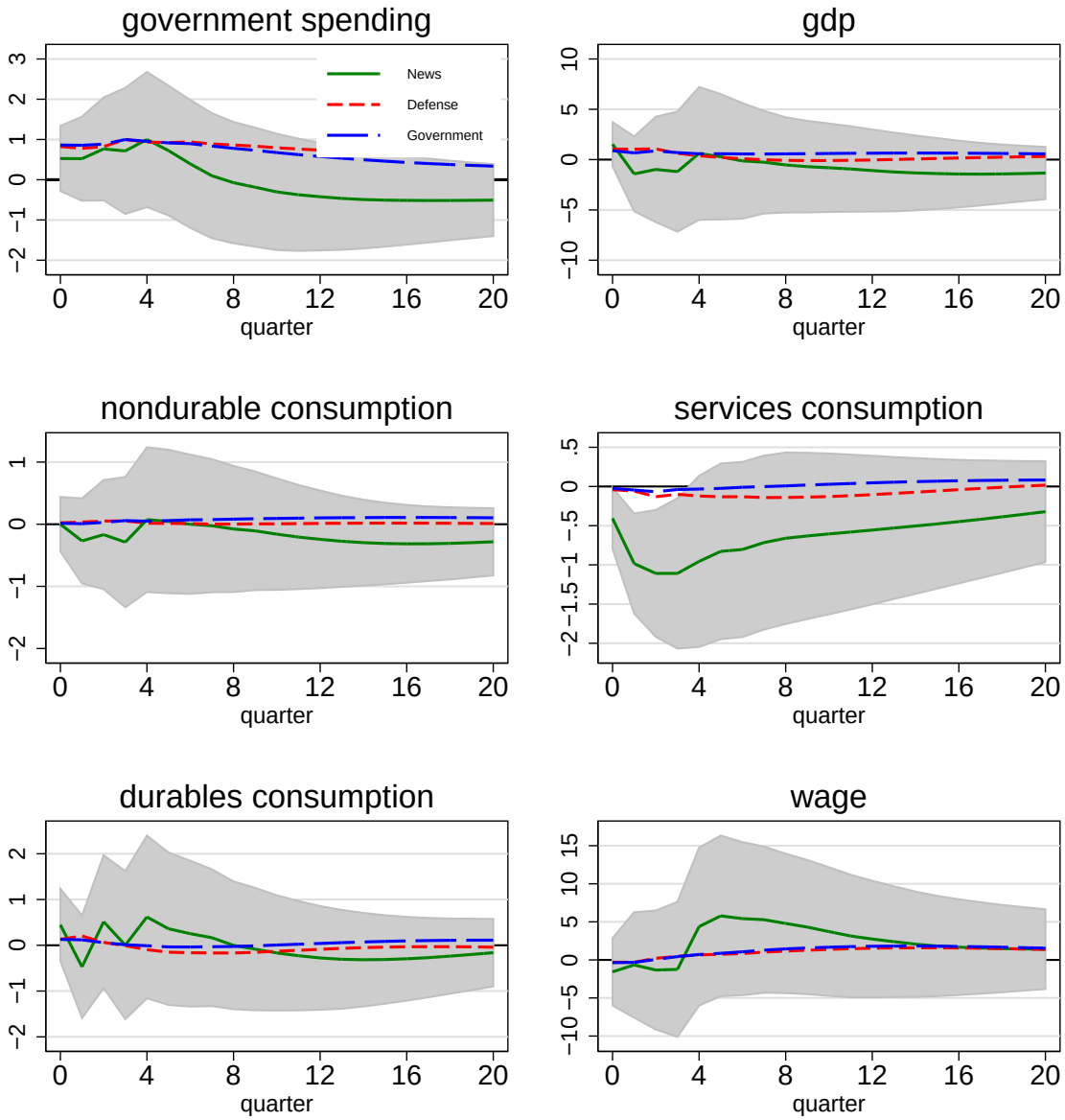
Notes: IRFs based on shocks to Ramey's news variable in an EVAR are denoted "News." IRFs based on SVAR shocks to defense spending are denoted "Defense" and to total government spending are denoted "Government." The standard error bands are 95 percent bands based on bootstrap standard errors in the "News" EVAR.

Figure 2. Comparison of VAR Identification Schemes, 1947:1 - 2008:4



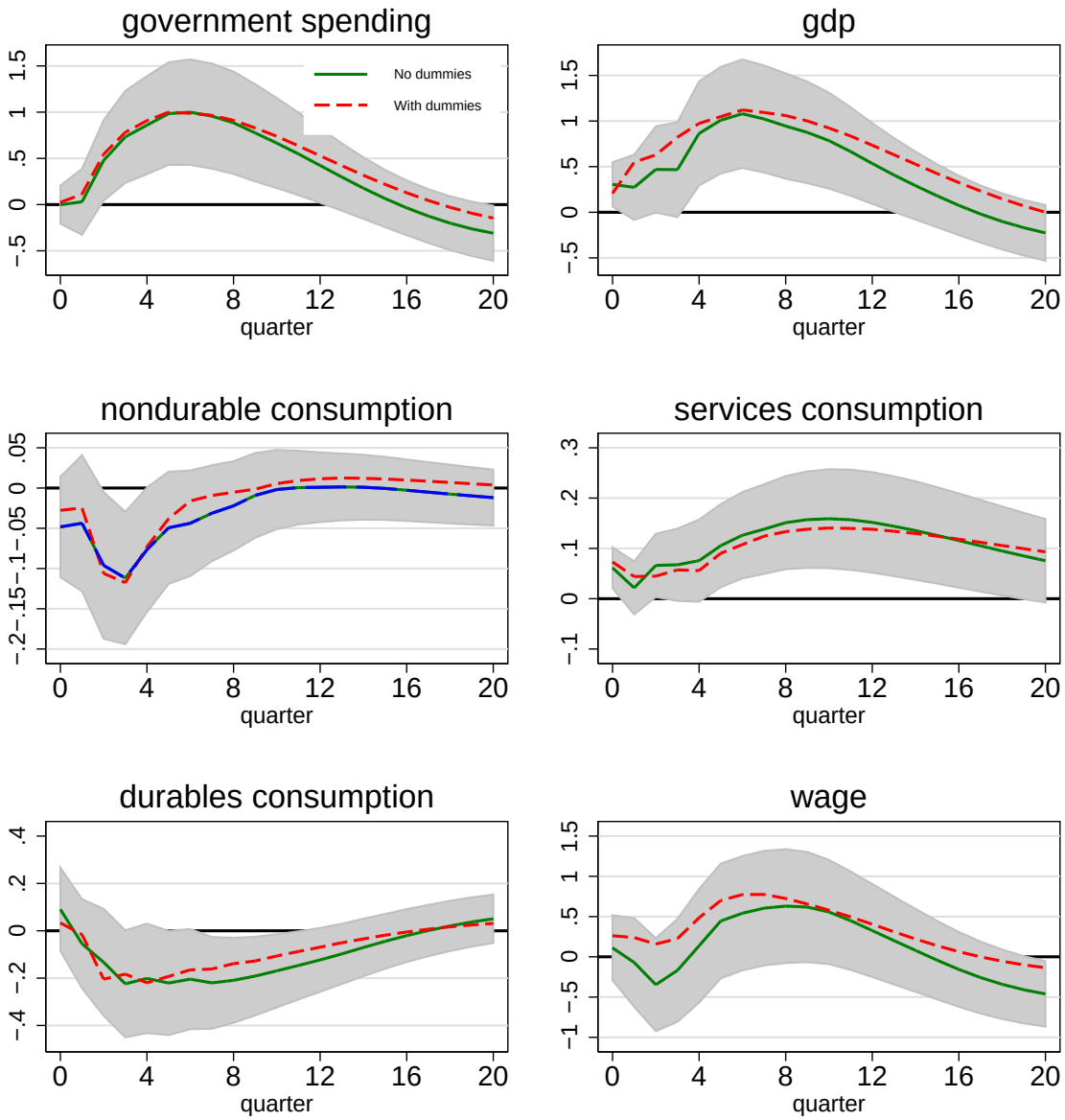
Notes: IRFs based on shocks to Ramey's news variable in an EVAR are denoted "News." IRFs based on SVAR shocks to defense spending are denoted "Defense" and to total government spending are denoted "Government." The standard error bands are 95 percent bands based on bootstrap standard errors in the "News" EVAR.

Figure 3. Comparison of VAR Identification Schemes, 1954:1 - 2008:4



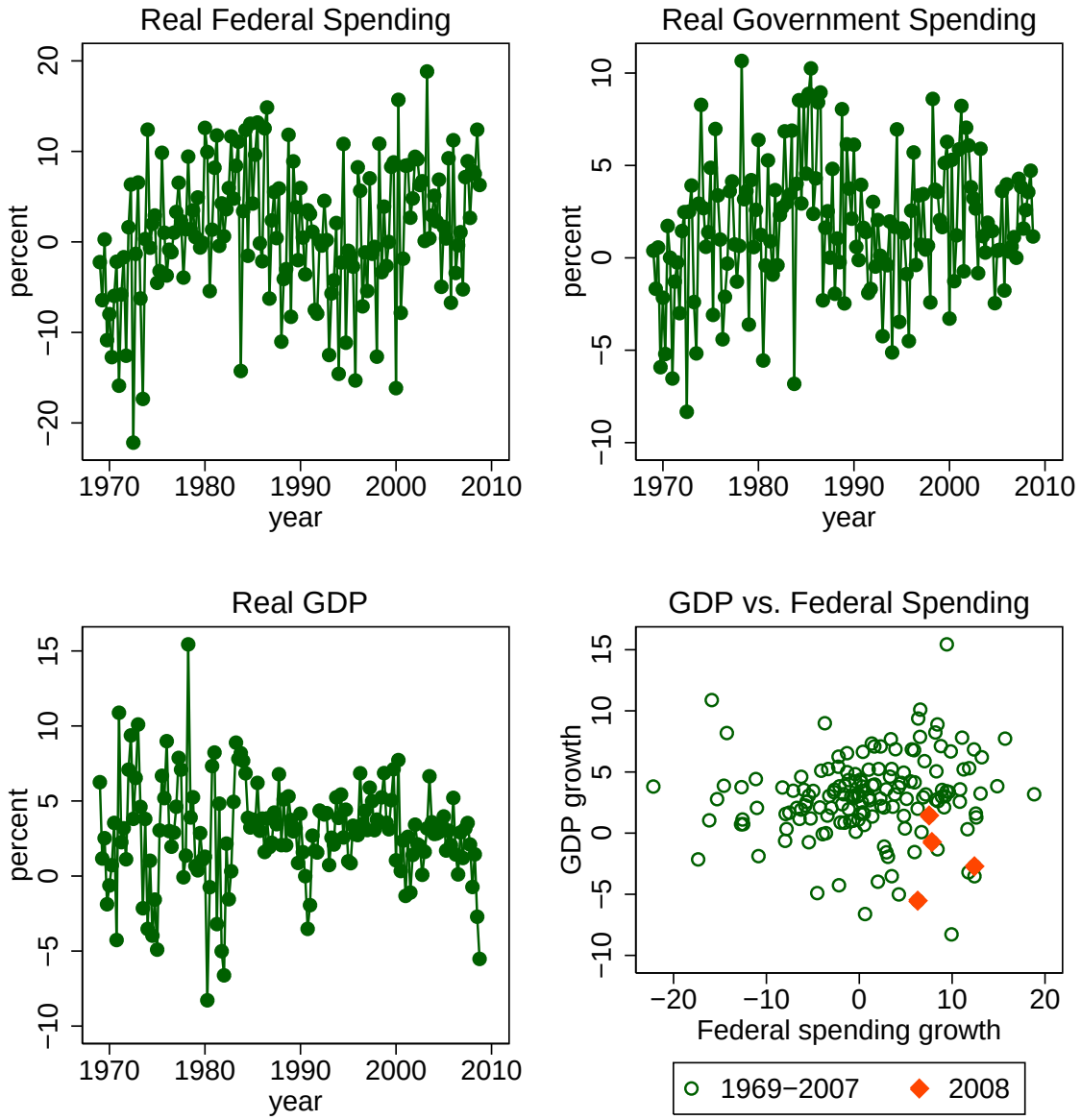
Notes: IRFs based on shocks to Ramey's news variable in an EVAR are denoted "News." IRFs based on SVAR shocks to defense spending are denoted "Defense" and to total government spending are denoted "Government." The standard error bands are 95 percent bands based on bootstrap standard errors in the "News" EVAR.

Figure 4. Comparison of News VARs with and without Dummy Variables, 1939:1 - 2008:4



Notes: All IRFs are based on shocks to Ramey's news variable. In the dummy variable specification, 1942:1, 1942:2, 1950:3, 1950:4, and 1951:1 are dummied out. The standard error bands are 95 percent bands based on bootstrap standard errors in the EVAR over the full sample.

Figure 5. Annualized Growth Rates, 1969 - 2008



Source: NIPA Data used by Ramey (2011).