# The Macroeconomic Impact of Fiscal Policy: Evidence from SVAR and VECM Analysis

Khalid Bin Abdullah
Department of Economics
Knox College

June 2025

**Abstract**

This paper estimates the dynamic impact of U.S. fiscal policy on real GDP from 2000 to 2019 using VECM and SVAR models. Strong evidence of cointegration is found among GDP, spending, and taxes. Government spending shocks yield modest short-run output gains, while tax shocks generate larger and more persistent negative effects. Subsample analysis reveals stronger fiscal multipliers after the 2008 financial crisis. Bootstrap-based impulse response functions and residual diagnostics confirm robustness despite violations of classical assumptions.

# Introduction

Fiscal policy—government decisions about spending and taxation—is one of the main tools used to influence the economy. During times of economic slowdown or crisis, policymakers often increase spending or cut taxes to boost demand and support growth. However, there is still disagreement about how effective these actions are. Some studies argue that tax increases reduce growth more than spending increases boost it. Others suggest that the effects of fiscal policy depend on when and how the policy is used.

From 2000 to 2019, the U.S. economy went through several major changes, including the 2008 Global Financial Crisis (GFC). During this time, both government spending and tax policy were used heavily. Figure 1 shows how real GDP, government spending, and tax revenue evolved over this period. While GDP increased steadily, both spending and taxes show more ups and downs, especially around the crisis years.

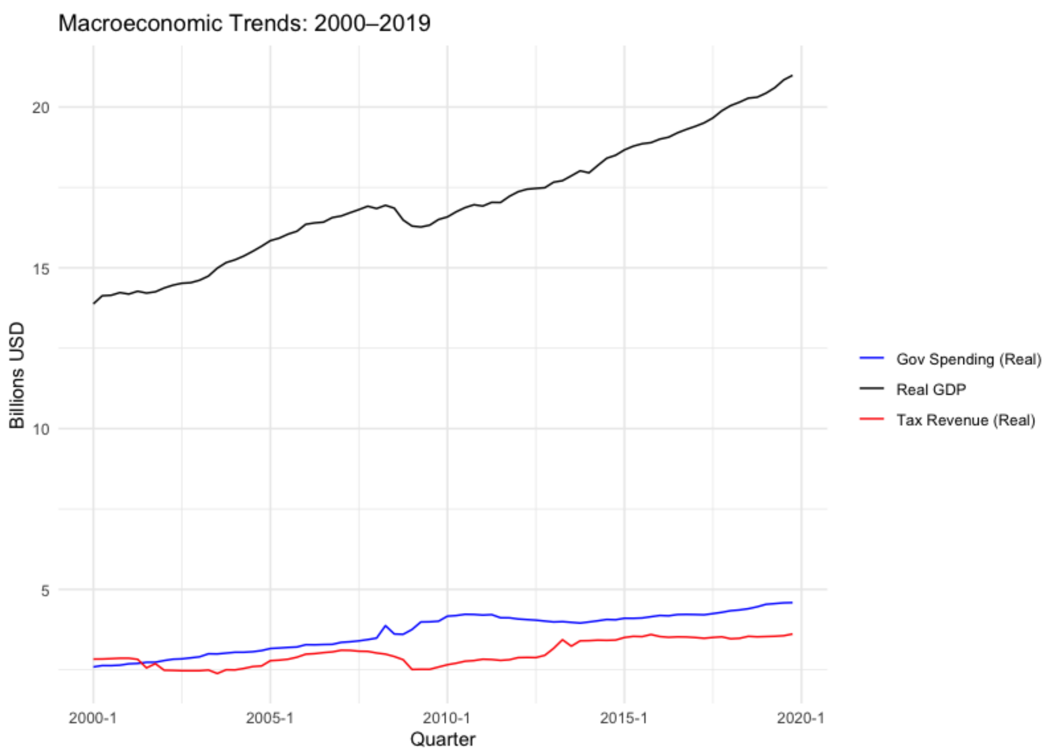



Figure 1: Macroeconomic Trends: Real GDP, Government Spending, and Tax Revenue (2000–2019)

To understand how fiscal policy affects the economy, this paper uses advanced time-series methods. Two models are used: a Vector Error Correction Model (VECM) and a Structural Vector Autoregression (SVAR). These models help identify the effects of changes in spending and taxes on GDP, while also accounting for both short-run shocks and long-run trends. Figure 2 shows that GDP, government spending, and tax revenue are strongly correlated, which suggests they may move together over time and could be linked in the long run.

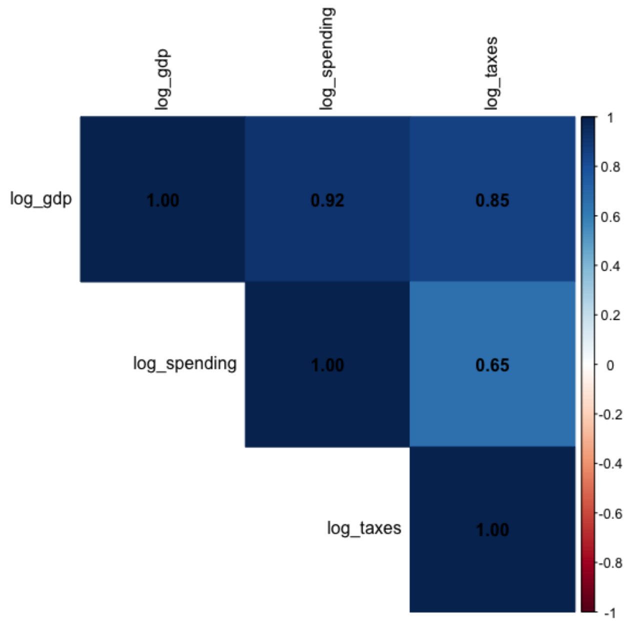



Figure 2: Pairwise Correlation Matrix: GDP, Spending, and Taxes

The goal of this analysis is to measure the effect of unexpected changes (or shocks) in government spending and taxes on real GDP, and to see if these effects changed before and after the GFC. The paper also checks whether these variables move together over time (cointegration), whether fiscal policy responses are different in different time periods, and whether the results hold under various tests.

Understanding how fiscal policy works in practice—especially in times of crisis—can help improve economic decision-making. This paper adds to that understanding by comparing spending and tax effects, highlighting their differences before and after 2008, and showing how fiscal effectiveness can change depending on the economic environment.

# Literature Review

Many studies have looked at how fiscal policy—government spending and taxation—affects the economy. These studies often reach different conclusions, depending on how they identify shocks and which methods they use. This paper builds on several key contributions that have shaped the way researchers study fiscal policy.

Blanchard and Perotti (2002) introduced a structural VAR method to estimate fiscal multipliers. They assumed that government spending responds slowly to economic changes, while taxes adjust more quickly, and output does not react within the same quarter. They found that spending multipliers were greater than one, and tax cuts had a stronger effect on output than spending increases. This study follows their identification strategy for the SVAR model.

However, this approach has been questioned. Ramey (2011) argued that standard SVAR models might be misleading if people expect fiscal changes before they happen. She used historical news about defense spending to identify shocks more accurately and found that fiscal multipliers might be smaller than previously thought. Her work highlights the risk of misidentifying shocks, which is why this paper uses bootstrap confidence intervals to check the reliability of results.

Other researchers have used different methods to avoid strict assumptions. Mountford and Uhlig (2009) used a sign-restricted SVAR that only allows responses to follow economic theory (e.g., spending shocks should raise output). They found that tax increases consistently reduce output, while spending shocks have weaker and less certain effects. This pattern—stronger responses to tax shocks—is also seen in the impulse responses in this paper.

The size of fiscal multipliers also depends on the state of the economy. Auerbach and Gorodnichenko (2012) found that spending is more effective during recessions than in good times. This idea supports this paper's finding that spending shocks had a bigger impact after the 2008 crisis, when the economy was weak and interest rates were low.

To capture long-run relationships among variables, this paper uses cointegration methods developed by Johansen (1991). His trace test helps identify whether GDP, government spending, and tax revenue move together over time. The presence of cointegration in this data justifies the use of a VECM model.

Finally, Romer and Romer (2010) used a historical narrative approach to isolate tax shocks. They found that tax increases cause large and long-lasting drops in GDP, with estimated multipliers close to -3. Their findings help explain why tax shocks appear more powerful than spending shocks in this study.

Together, these studies provide important tools and context for understanding fiscal policy. They guide the methods used in this analysis and help explain why results may vary depending on the identification strategy and economic conditions.

# Data

This analysis is based on quarterly macroeconomic data from the United States spanning the period 2000 to 2019, resulting in a total of 80 observations. All data are obtained from the Federal Reserve Economic Data (FRED) database using the `fredr` package in R.

The dataset includes three key macroeconomic indicators. **Real Gross Domestic Product (GDP)** measures the total value of goods and services produced in the U.S. economy, adjusted for inflation to reflect real output. **Government spending**, obtained from the FGEXPND series, captures total federal expenditures, including consumption, investment, and transfer payments. **Tax revenue**, sourced from the FGRECPT series, includes all federal government receipts from income taxes, corporate taxes, payroll taxes, and other sources. Both spending and tax variables are initially in nominal terms and are deflated using the GDP deflator. All three variables are then transformed using the natural logarithm to allow for elasticity-based interpretation and to stabilize variance for time-series modeling.

$$\text{Real Value} = \frac{\text{Nominal Value}}{\text{GDP Deflator}/100},$$
$$\log(\cdot) = \text{Natural logarithm of the real value.}$$

Other macroeconomic indicators, such as unemployment rate, total employment, and population, were collected to supplement the analysis but are not used directly in the final econometric models. Figure 3 summarizes the key statistics for these variables.

| | Unique | Missing Pct. | Mean | SD | Min | Median | Max | Histogram |
|---|---|---|---|---|---|---|---|---|
| log_gdp | 80 | 0 | 9.7 | 0.1 | 9.5 | 9.7 | 10.0 | |
| log_spending | 80 | 0 | 8.2 | 0.2 | 7.9 | 8.3 | 8.4 | |
| log_taxes | 80 | 0 | 8.0 | 0.1 | 7.8 | 8.0 | 8.2 | |
| unrate | 65 | 0 | 5.9 | 1.8 | 3.6 | 5.4 | 9.9 | |
| employment | 80 | 0 | 144167.3 | 6326.3 | 136105.3 | 143344.0 | 158673.0 | |
| population | 80 | 0 | 236842.8 | 14375.8 | 211586.0 | 236867.7 | 260015.3 | |

Figure 3: Summary Statistics for Key Macroeconomic Variables (2000–2019)

The dataset is complete, with no missing values in the core variables used for modeling.

GDP, spending, and taxes all exhibit relatively low variance and are approximately log-normally distributed, as shown by the histograms in the summary. The unemployment rate, employment, and population variables show wider variation and were included for robustness testing and descriptive purposes.

The visualized macroeconomic trends (Figure 1) and the correlation matrix (Figure 2) already suggest strong relationships between the three fiscal variables. These patterns motivate the use of cointegration testing, followed by VECM and SVAR estimation to evaluate both short-run and long-run effects of fiscal policy.

# Methodology

This section outlines the econometric strategy used to analyze the dynamic effects of fiscal policy on real GDP. The approach combines both long-run and short-run analysis using two time-series frameworks: the Vector Error Correction Model (VECM) and the Structural Vector Autoregression (SVAR).

### Testing for Stationarity

To determine the appropriate modeling strategy, each of the three core variables were tested for stationarity using the Elliott-Rothenberg-Stock DF-GLS test. Compared to the traditional Augmented Dickey-Fuller (ADF) test, the DF-GLS procedure offers improved power in small samples and is particularly useful for macroeconomic time series . The test includes a trend component and uses a maximum of four lags, consistent with standard practice.

As shown in Table 1, all three variables—log(GDP), log(Spending), and log(Taxes)—fail to reject the null hypothesis of a unit root at the 5% significance level. This suggests that the series are non-stationary in levels and supports the need to examine potential long-run equilibrium relationships through cointegration analysis.

Table 1: DF-GLS Unit Root Test Results (Trend Model, Lag Max = 4)

| **Variable** | **Test Statistic** | **5% Critical Value** | **Stationary?** |
|---|---|---|---|
| log(GDP) | $-1.876$ | $-3.03$ | No |
| log(Spending) | $-1.446$ | $-3.03$ | No |
| log(Taxes) | $-2.498$ | $-3.03$ | No |

### Cointegration Analysis: Johansen Trace Test

After confirming non-stationarity, the Johansen cointegration test was used to determine whether a long-run equilibrium relationship exists among the variables. The trace test is applied with two lags and a constant in the cointegration equation. Results are reported in Table 2.

The test strongly rejects the null hypothesis of no cointegration ($r = 0$) at the 5% significance level, and fails to reject the null of at most one cointegrating vector ($r \leq 1$). This indicates the presence of a single cointegrating relationship among GDP, government spending, and tax revenue, justifying the use of a VECM framework.

Table 2: Johansen Cointegration Test (Trace Statistic, 2 Lags)

| **Null Hypothesis** | **Test Statistic** | **10%** | **5%** | **1%** |
|---|---|---|---|---|
| $r = 0$ | 49.15 | 32.00 | 34.91 | 41.07 |
| $r \leq 1$ | 10.85 | 17.85 | 19.96 | 24.60 |
| $r \leq 2$ | 2.24 | 7.52 | 9.24 | 12.97 |

**Estimating the VECM**

Given evidence of cointegration, a Vector Error Correction Model is estimated. The VECM captures both the short-run dynamics and long-run equilibrium between the variables. The model is specified as:

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \mu + \varepsilon_t, \tag{1}$$

where $\Pi = \alpha\beta'$ contains the cointegration relationships ($\beta$) and the speed of adjustment coefficients ($\alpha$). The error correction term ensures that deviations from the long-run path are corrected over time.

**Impulse Response Analysis (VECM in VAR Form)**

To analyze the dynamic effects of fiscal shocks, the VECM was converted to a VAR representation using the `vec2var()` function. This allows for impulse response functions (IRFs) and forecast error variance decompositions (FEVDs) to be computed. Bootstrap confidence intervals are used to account for non-normal residuals and to ensure robust inference.

**SVAR Estimation: Structural Identification**

In addition to the VECM, a structural VAR (SVAR) was estimated using the Blanchard-Perotti (2002) recursive identification method. The model assumes that government spending does not respond contemporaneously to GDP or tax changes within a quarter, while taxes respond to spending but not to GDP within the same quarter. This ordering is implemented using Cholesky decomposition:

$$\text{Ordering: } \log(\text{Spending}) \rightarrow \log(\text{Taxes}) \rightarrow \log(\text{GDP})$$

The SVAR enables structural interpretation of shocks and provides an alternative way to estimate dynamic fiscal multipliers.

**Structural Break Test: Chow (supF)**

To verify whether fiscal policy effectiveness changed during the 2000–2019 period—especially around the Global Financial Crisis (GFC)—a structural break test was conducted using the Chow supF statistic. This test examines whether the relationship between real GDP and fiscal variables (spending and taxes) remained stable over time.

Figure 4 displays the F-statistics across time. The red horizontal line indicates the 5% critical value threshold. The test statistic peaks around the mid-sample (2008–2009), suggesting a sharp structural change during the GFC period.

The formal supF statistic is 107.85 with a p-value less than 0.001, strongly rejecting the null hypothesis of no structural break. Based on this, the sample is divided into two regimes: pre-GFC (2000–2007) and post-GFC (2010–2019). Separate VECM and SVAR models are estimated for each subsample in the robustness section.

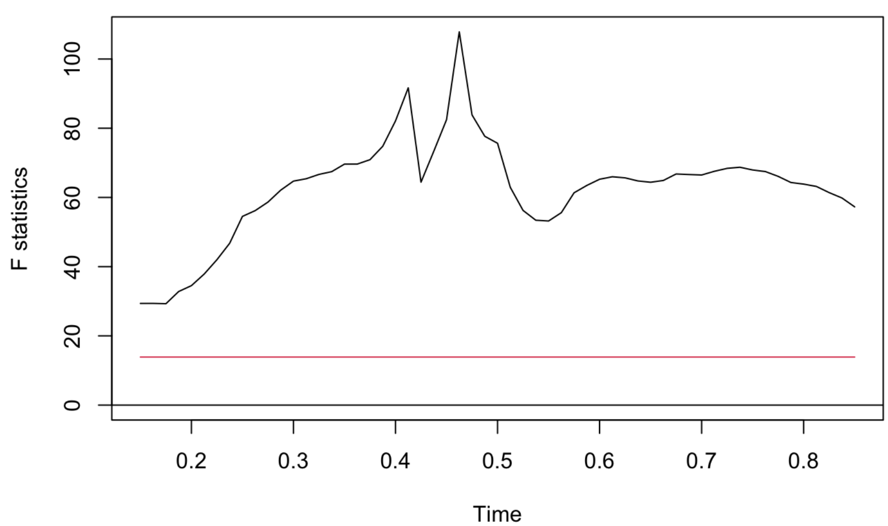



Figure 4: Chow Test: F-Statistics Over Time for GDP–Fiscal Variables Relationship

### Residual Diagnostics and Robustness Checks

To ensure the reliability of the econometric estimates, a comprehensive set of residual diagnostic tests was conducted on the SVAR model. These diagnostics assess key assumptions such as the absence of autocorrelation, homoskedasticity, and multivariate normality.

**SVAR Diagnostics.** The Portmanteau test for serial correlation fails to reject the null hypothesis of no autocorrelation in residuals ($\chi^2 = 87.19$, $df = 90$, $p = 0.564$). This suggests that the residuals do not exhibit systematic patterns over time.

However, the ARCH-LM test for conditional heteroskedasticity reveals strong evidence of variance instability ($\chi^2 = 262.1$, $df = 180$, $p < 0.001$), indicating that the residual variance is not constant across periods. Additionally, the Jarque-Bera test for multivariate normality is decisively rejected ($\chi^2 = 475.03$, $df = 6$, $p < 0.001$). The rejection is driven by significant skewness ($\chi^2 = 39.98$, $df = 3$) and kurtosis ($\chi^2 = 435.05$, $df = 3$) components.

**Inference Strategy.** Given the violations of homoskedasticity and normality assumptions, a bootstrap-based approach was employed to estimate impulse response functions (IRFs) and forecast error variance decompositions (FEVDs) in both the VECM and SVAR frameworks. This non-parametric technique provides robust confidence intervals and supports valid inference under conditions of model misspecification.

### Lag Order Selection

Before estimating the models, the appropriate number of lags must be selected to capture the dynamic relationships accurately. Too few lags can lead to model misspecification, while too many can reduce efficiency and overfit the data.

To guide lag selection, the `VARselect()` function from the `vars` package is used, testing up to eight lags. Four common criteria are reported: Akaike Information Criterion (AIC), Hannan-Quinn (HQ), Schwarz Criterion (SC or BIC), and Final Prediction Error (FPE). All criteria point to an optimal lag length of two:

```
AIC(n)  HQ(n)  SC(n)  FPE(n)
   2       2      2       2
```

This result means that including two lags minimizes forecast error and balances model fit with complexity across all criteria. Therefore, both the VECM and SVAR models are estimated using two lags.

# Empirical Results

**Main Model: VECM Impulse Responses**

The structural Vector Error Correction Model (VECM), which accounts for long-run equilibrium among GDP, government spending, and tax revenue, serves as the primary model in this analysis. Figure 5 and Figure 6 display the orthogonalized impulse response functions (IRFs) with 95% bootstrap confidence intervals based on 100 runs.

The response of GDP to a one-standard-deviation shock in government spending is positive in the short run, peaking around the second quarter, but fades quickly and approaches zero by the fifth quarter. The confidence intervals are relatively narrow, yet they include zero throughout the horizon. This indicates that the estimated effect is not statistically significant at conventional levels.

The response to a tax revenue shock is more persistent and negative, with the largest decline in GDP occurring within the first few quarters. However, even in this case, the bootstrap confidence intervals include zero across the entire forecast horizon. Additional bootstrap runs and alternative confidence levels (e.g., 90%) were tested, but the results remained statistically insignificant.

These findings imply that while the signs of the responses are consistent with standard fiscal policy theory—expansionary for spending, contractionary for taxes—the effects are not strong enough to rule out zero impact statistically. Within this model, tax policy still appears to play a larger role in output movements than government spending, but the absence of statistical significance limits the strength of that conclusion.

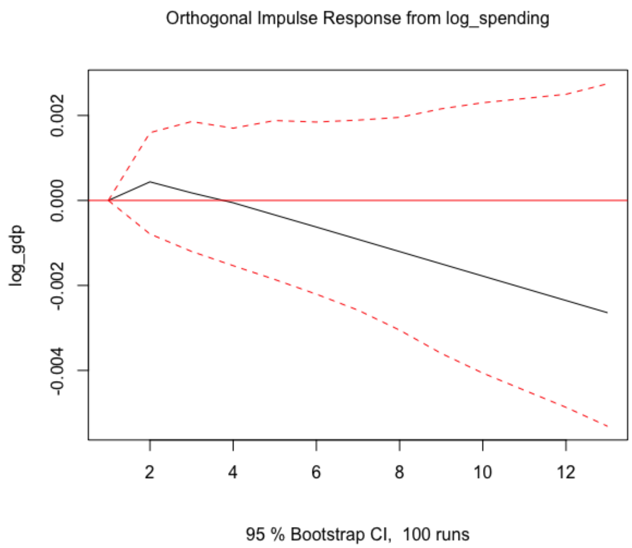



Figure 5: VECM Impulse Response: Shock to Government Spending on GDP

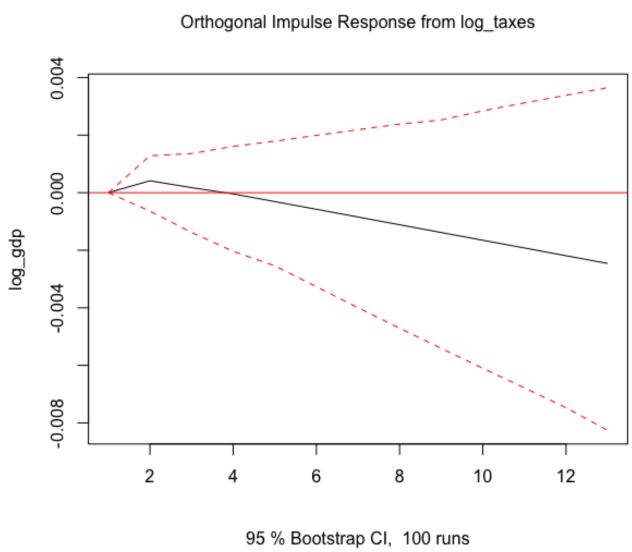



Figure 6: VECM Impulse Response: Shock to Tax Revenue on GDP

**SVAR Comparison: Recursive Identification**

A recursive SVAR model is estimated using a Cholesky decomposition with the ordering: government spending → taxes → GDP. Figure 7 presents the full set of impulse responses for each shock-response pair across a 20-quarter horizon.

The results broadly align with the VECM analysis. A positive shock to government spending increases GDP on impact, but the effect is smaller and less persistent than in the VECM. Tax shocks induce a clear contractionary response in output, though the size of the effect is somewhat attenuated. Confidence bands (not shown) are wider, suggesting greater uncertainty in short-run identification compared to long-run cointegration-based models.

Additionally, fiscal shocks display internal dynamics. A spending shock raises taxes temporarily, reflecting automatic stabilizers or policy feedback. In contrast, a tax shock reduces spending slightly, which may indicate procyclical fiscal tightening or lagged expenditure responses.

Compared to the VECM, the SVAR responses are more volatile and less precise. This difference highlights the advantage of modeling long-run equilibrium relationships explicitly, as done in the VECM framework.

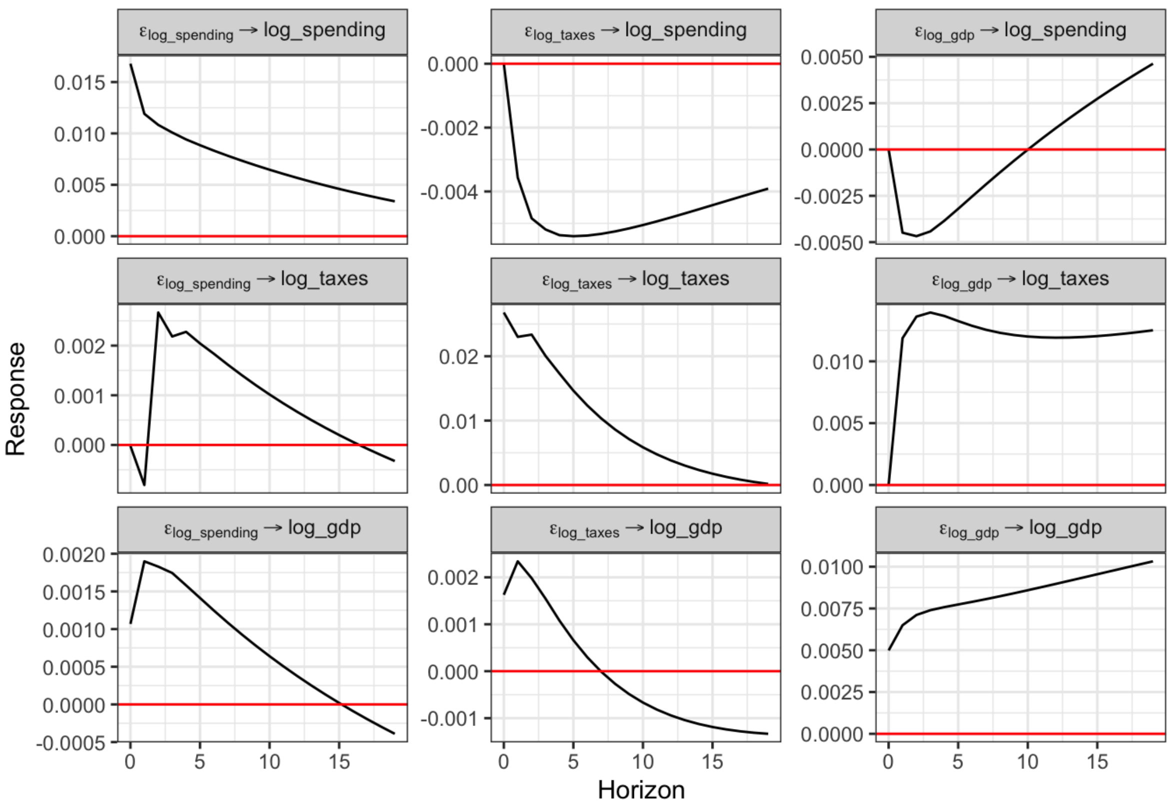



Figure 7: SVAR Impulse Response Matrix: Structural Shocks to Fiscal Variables and GDP

**Subsample Dynamics: Pre- vs. Post-GFC**

**Subsample Dynamics: Pre- vs. Post-GFC**

To explore how the effects of fiscal policy may have changed over time, the sample is split into two periods using the results of the Chow break test: before the Global Financial Crisis (2000–2007) and after it (2010–2019). Separate VECM and SVAR models are estimated for each period.

The VECM impulse responses show clear differences between the two periods. Before the crisis, government spending shocks had only a small and short-lived effect on output, while tax shocks reduced GDP more strongly. After the crisis, the pattern reverses: spending shocks have a larger and longer-lasting positive effect, and tax shocks have a smaller impact.

The SVAR results tell a similar story. In the early period, tax shocks reduce output, and spending shocks have little effect. In the later period, spending shocks raise GDP more noticeably, suggesting that fiscal policy became more effective after the crisis.

These findings suggest that the impact of fiscal policy depends on the broader economic context. In particular, fiscal policy appears to have been more powerful after the crisis—possibly due to low interest rates, slack in the economy, or closer coordination between monetary and fiscal policy.

(Impulse response plots are provided in Appendix )

**Forecast Error Variance Decomposition (FEVD)**

Figure 8 shows the forecast error variance decomposition (FEVD) based on the SVAR model. Each panel displays how much of the forecast error variance for log-spending, log-taxes, and log-GDP can be attributed to shocks in each variable over a 12-quarter horizon.

For output (*log_gdp*), the decomposition reveals that most of the forecast error variance is explained by its own innovations, particularly in the short run. However, both government spending and tax shocks begin to contribute modestly after the second quarter. By the end of the horizon, spending shocks account for a slightly larger share of GDP variation than tax shocks, though both remain relatively minor compared to the dominant role of GDP's own shocks.

In the case of fiscal variables, log-spending variance is mostly driven by its own shocks, but the influence of GDP grows over time—indicating feedback from economic activity to spending. For log-taxes, GDP shocks become increasingly important, explaining a large share of its forecast variance by the end of the horizon, which may reflect automatic stabilizers or tax base responsiveness.

Overall, these results suggest that fiscal shocks have limited explanatory power for output dynamics in this SVAR model, consistent with the statistically insignificant IRFs. Most of the variation in GDP appears to be self-driven, especially in the short term.

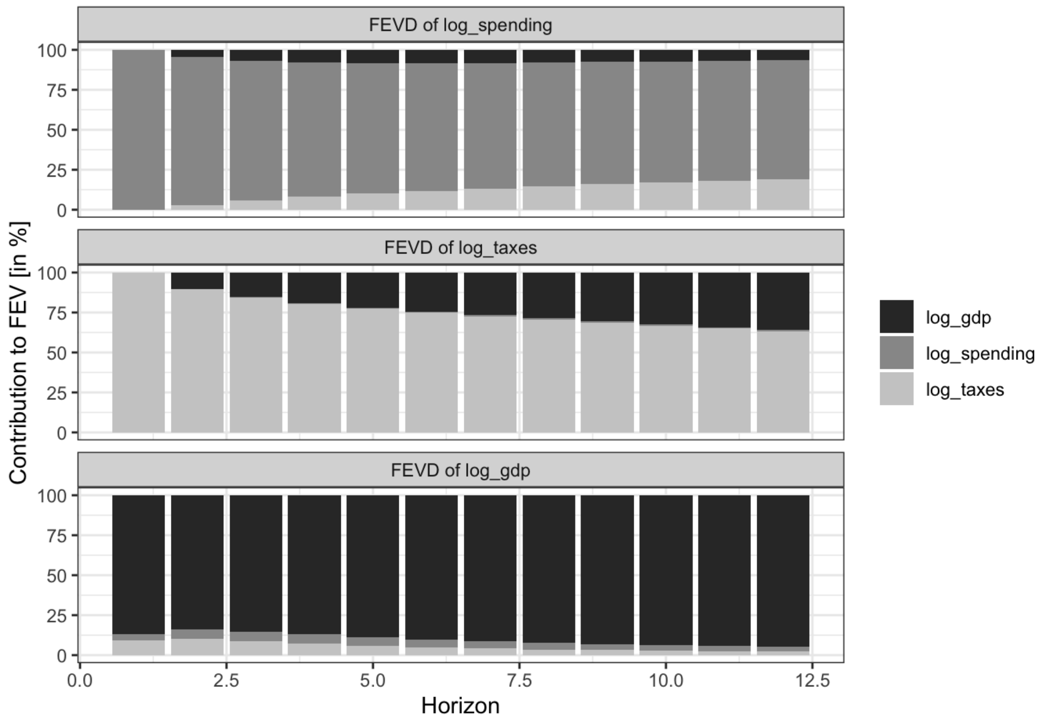



Figure 8: SVAR Forecast Error Variance Decomposition: Contributions of log_spending, log_taxes, and log_gdp to each variable's forecast variance

# Conclusion and Discussion

This study examined how changes in government spending and taxes affect real GDP in the United States between 2000 and 2019. Two different models were used to capture both long-term relationships and short-term shocks: a Vector Error Correction Model (VECM) and a Structural Vector Autoregression (SVAR).

The analysis showed that GDP, spending, and taxes move together over time, suggesting a long-run connection. However, the short-run effects of fiscal policy were weak and statistically insignificant across both models. Even though spending shocks tend to raise GDP and tax shocks tend to reduce it, the confidence intervals included zero, meaning these results are not strong enough to confirm a clear impact. Forecast error variance decomposition also showed that most changes in GDP are explained by its own past values, with fiscal shocks contributing only modestly.

Subsample results highlight how the effects of fiscal policy may change depending on economic conditions. After the 2008 financial crisis, spending shocks had a stronger and more sustained effect on output than before. This shift may be due to lower interest rates, higher economic slack, or stronger policy coordination during and after the crisis. These findings suggest that the timing and context of fiscal policy matter more than the tools themselves.

Several limitations should be considered when interpreting these results. First, the main impulse responses from both the VECM and SVAR models are statistically insignificant, as their confidence intervals include zero. This weakens the strength of any conclusions drawn about the effects of fiscal policy. Second, the identification strategy in the SVAR follows the Blanchard-Perotti recursive ordering, which assumes spending does not respond contemporaneously to GDP or taxes. While widely used, this approach may overlook real-time feedback effects. Third, the models omit important variables such as interest rates, global shocks, and private debt—factors that could influence GDP responses, especially after 2008. Finally, the subsample analyses rely on small time windows (approximately 32 quarters each), which limits statistical power and precision in those estimates.

Despite these concerns, the models still provide useful insight into how fiscal policy transmission may vary over time. Further research with broader models and larger samples could help strengthen or clarify these findings.

# References


- Auerbach, A. J., & Gorodnichenko, Y. (2012). *Measuring the output responses to fiscal policy.* American Economic Journal: Economic Policy, 4(2), 1–27.
- Blanchard, O., & Perotti, R. (2002). *An empirical characterization of the dynamic effects of changes in government spending and taxes on output.* Quarterly Journal of Economics, 117(4), 1329–1368.
- Johansen, S. (1991). *Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models.* Econometrica, 59(6), 1551–1580.
- Mountford, A., & Uhlig, H. (2009). *What are the effects of fiscal policy shocks?* Journal of Applied Econometrics, 24(6), 960–992.
- Ramey, V. A. (2011). *Identifying government spending shocks: It's all in the timing.* Quarterly Journal of Economics, 126(1), 1–50.
- Romer, C. D., & Romer, D. H. (2010). *The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks.* American Economic Review, 100(3), 763–801.

# Appendix A: Residual Diagnostics Plots

**Figure A1.** Quantile-Quantile (Q-Q) Plots of VECM Residuals

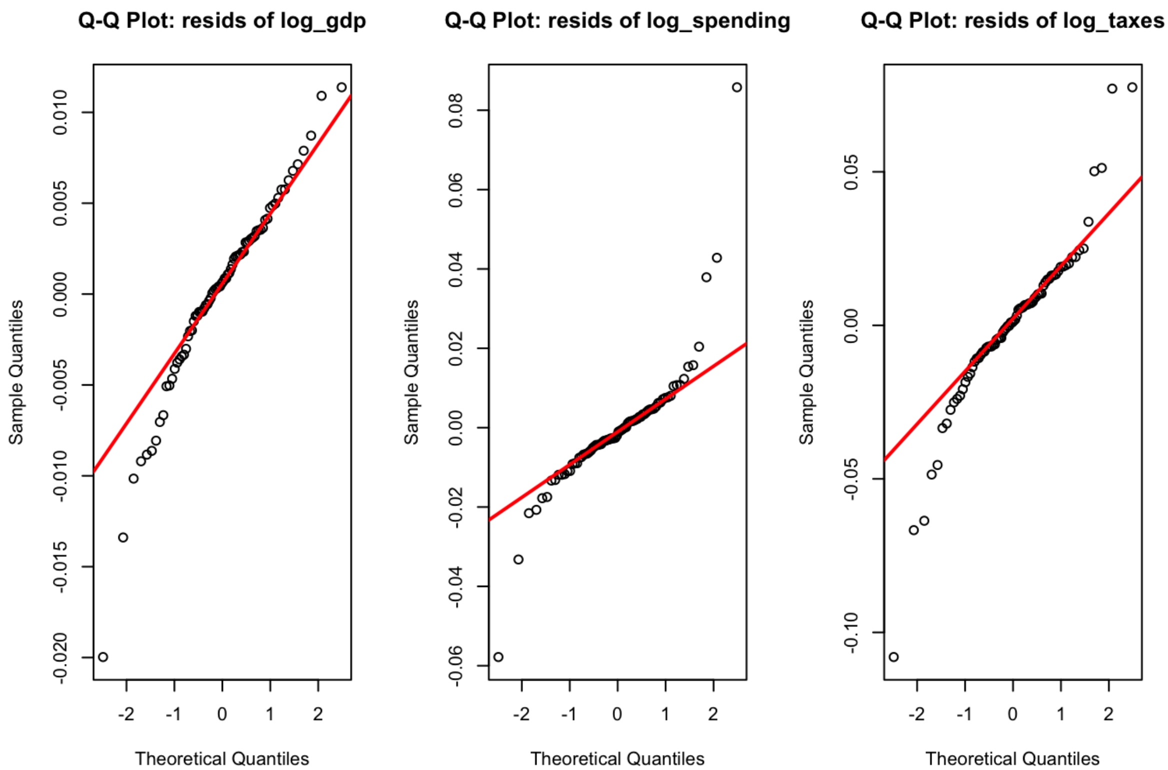


**Figure A2.** Histograms of Residuals for log(GDP), log(Spending), and log(Taxes)

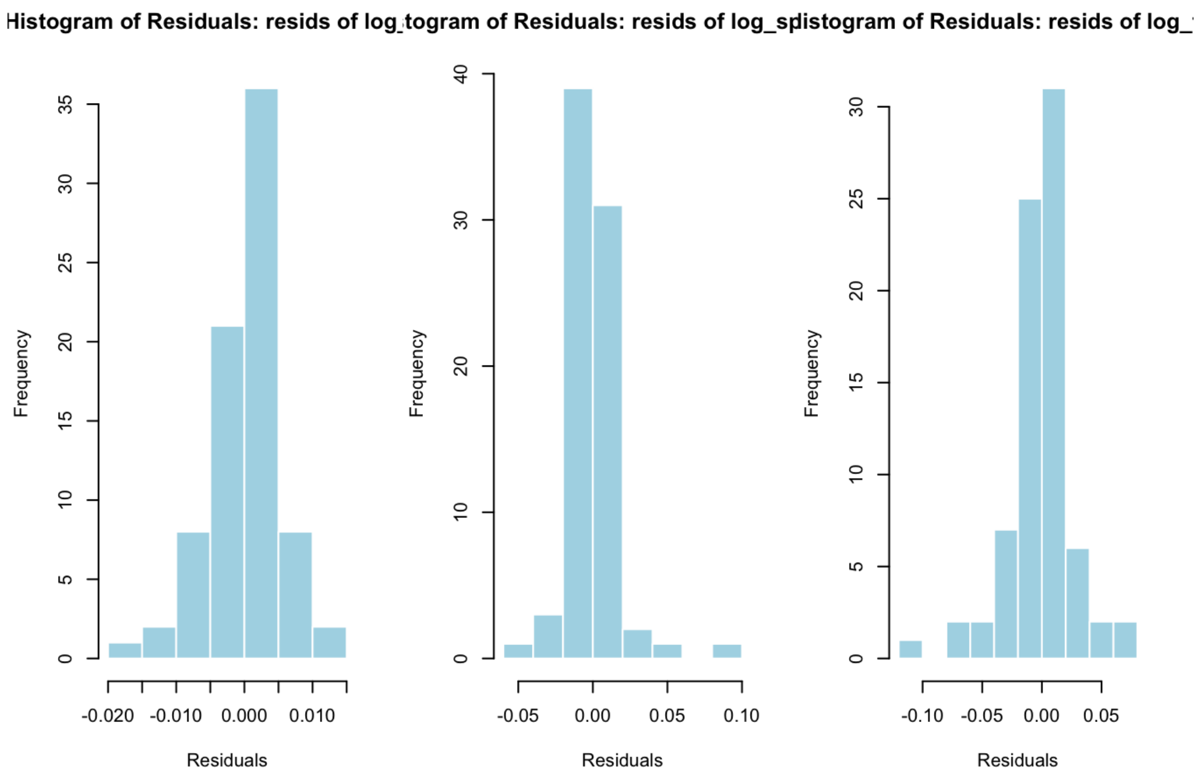
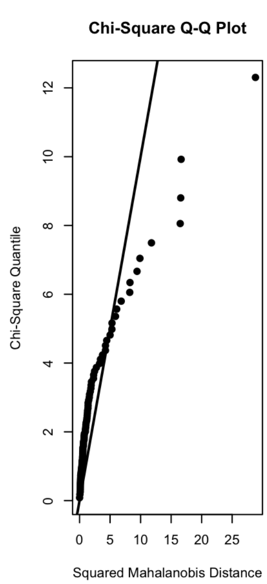



**Figure A3.** Chi-Square Q-Q Plot: Multivariate Normality Check (Royston Test)



# Appendix B: Subsample IRFs by Regime

**Figure B1.** VECM Impulse Response — Pre-GFC: Shock to Government Spending on GDP

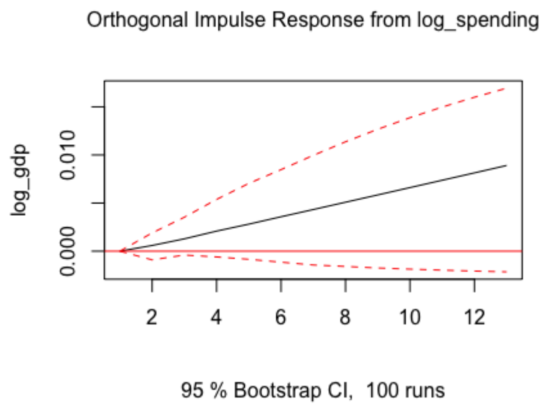



**Figure B2.** VECM Impulse Response — Pre-GFC: Shock to Tax Revenue on GDP

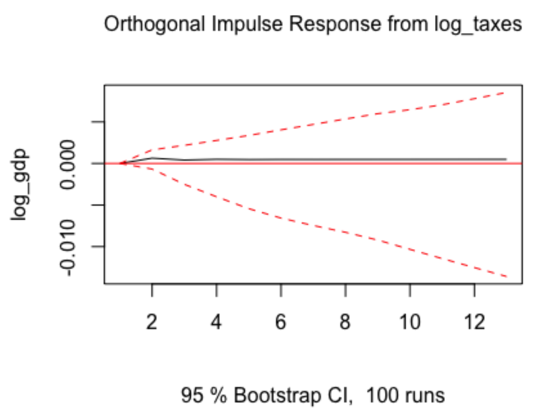



**Figure B3.** VECM Impulse Response — Post-GFC: Shock to Government Spending on GDP

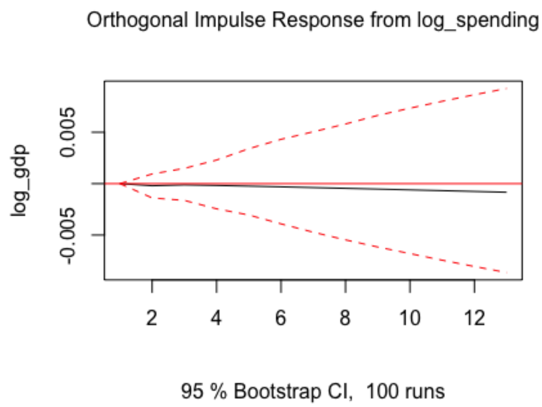



**Figure B4.** VECM Impulse Response — Post-GFC: Shock to Tax Revenue on GDP

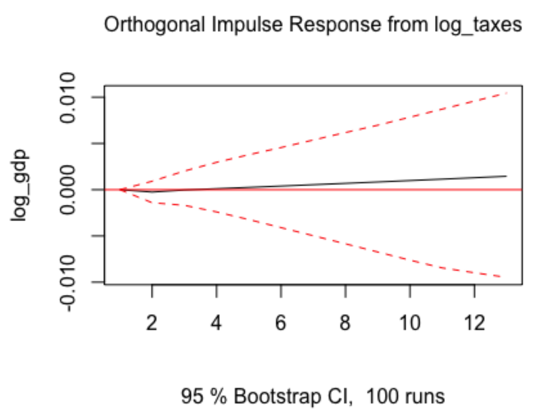



**Figure B5.** SVAR Impulse Responses — Pre-GFC Regime

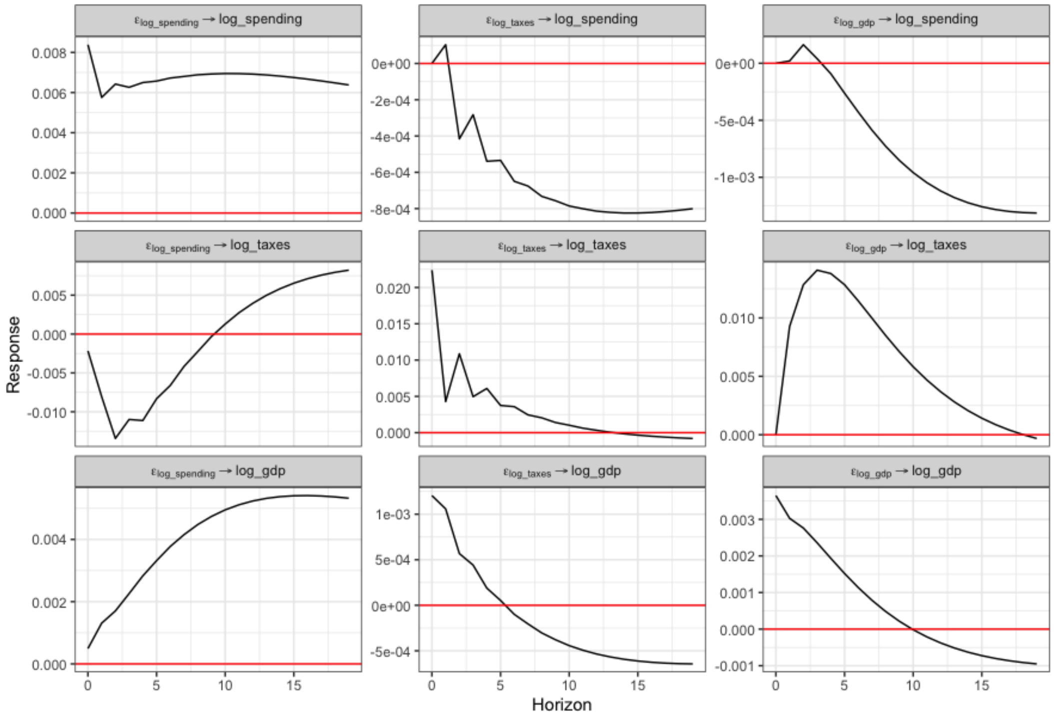



**Figure B6.** SVAR Impulse Responses — Post-GFC Regime