

Kopecky, Joseph et al.

**Working Paper**

## Robustness Report on "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom", by James Cloyne (2013)

I4R Discussion Paper Series, No. 132

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Kopecky, Joseph et al. (2024) : Robustness Report on "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom", by James Cloyne (2013), I4R Discussion Paper Series, No. 132, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/298013>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



No. 132

I4R DISCUSSION PAPER SERIES

# **Robustness Report on “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom”, by James Cloyne (2013)**

Joseph Kopecky

Abel Brodeur

Lester Lusher

Douglas Campbell

Magnus Johannesson

Nikita Tsoy

**June 2024**

## I4R DISCUSSION PAPER SERIES

I4R DP No. 132

# **Robustness Report on “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom”, by James Cloyne (2013)**

**Joseph Kopecky<sup>1</sup>, Douglas Campbell<sup>2</sup>, Abel Brodeur<sup>3</sup>,  
Magnus Johannesson<sup>4</sup>, Lester Lusher<sup>5</sup>, Nikita Tsoy<sup>6</sup>**

*<sup>1</sup>Trinity College Dublin/Ireland*

*<sup>2</sup>New Economic School, Moscow/Russia*

*<sup>3</sup>University of Ottawa/Canada and IZA – Institute of Labor Economics, Bonn/Germany*

*<sup>4</sup>Stockholm School of Economics, Stockholm/Sweden*

*<sup>5</sup>University of Pittsburgh/USA*

*<sup>6</sup>Institute for Computer Science, Artificial Intelligence and Technology (INSAIT), and Sofia University, Sofia/Bulgaria*

JUNE 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

### Editors

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

E-Mail: [joerg.peters@rwi-essen.de](mailto:joerg.peters@rwi-essen.de)  
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3  
45128 Essen/Germany

[www.i4replication.org](http://www.i4replication.org)

# Robustness Report on “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom”, by James Cloyne (2013).\*

Joseph Kopecky<sup>†</sup>    Douglas Campbell<sup>‡</sup>    Abel Brodeur<sup>§</sup>  
Magnus Johannesson<sup>¶</sup>    Lester Lusher<sup>||</sup>    Nikita Tsoy<sup>\*\*</sup>

May 12, 2024

## Abstract

Cloyne (2013) constructs a novel dataset documenting fiscal tax shocks in the United Kingdom using the narrative approach developed by Romer and Romer (2010), and estimates the impact of tax changes on GDP. He finds that a tax cut of one percent of GDP causes a 0.6 percent increase in output in the initial quarter of the policy, rising to a peak of 2.5 percent over three years. We first reproduce all of the VAR tables and figures in the original paper, and then test for robustness through a number of changes to the baseline regression model, particularly: changes in lag structure, changes in the control set, alternative estimation procedures, and excluding influential observations. In 60% of robustness the impact effect is significant at the 95% level, with a mean estimated coefficient of 0.63, while in 70% of robustness tests the peak response remains significant at the 95% level, with a mean peak response of 2.27.

**JEL Classification:** E23, E32, E62, H20, H61

**Keywords:** Fiscal policy, taxation, narrative shocks, macroeconomics

---

\*This robustness report was written as part of a meta-robustness project entitled “The robustness reproducibility of the American Economic Review”. See Campbell, Brodeur, Dreber, Johannesson, Kopecky, Lusher and Tsoy (2024) for more information.

<sup>†</sup>Corresponding author. Department of Economics, Trinity College Dublin, E-mail: [jkopecky@tcd.ie](mailto:jkopecky@tcd.ie)

<sup>‡</sup>New Economic School, E-mail: [dolcampb@gmail.com](mailto:dolcampb@gmail.com)

<sup>§</sup>University of Ottawa and IZA. E-mail: [abrodeur@uottawa.ca](mailto:abrodeur@uottawa.ca)

<sup>¶</sup>Department of Economics, Stockholm School of Economics, E-mail: [magnus.johannesson@hhs.se](mailto:magnus.johannesson@hhs.se)

<sup>||</sup>Department of Economics, University of Pittsburgh, E-mail: [lesterlusher@pitt.edu](mailto:lesterlusher@pitt.edu)

<sup>\*\*</sup>INSAIT, Sofia University, E-mail: [nikita.tsoy@insait.ai](mailto:nikita.tsoy@insait.ai)

# 1 Introduction

Cloyne (2013) shows that changes in taxes have both contemporaneous output effects (on a scale smaller than the shock itself) and large effects at longer (roughly 2.5 year) horizons. He identifies tax shocks using the narrative fiscal approach developed in Romer and Romer (2010), compiling a novel dataset of exogenous tax changes within the UK for the period from 1947 to 2009. Cloyne (2013) tests the impact of these narrative shocks vector autoregression (VAR) series, primarily focusing on their effects on output, but also on imports and exports, real wages, and employment. Ultimately he provides evidence of a very similar size of output responses to tax shocks to those found by Romer and Romer (2010), who construct a similar narrative shock using data from the United States. Cloyne (2013) draws a direct comparison to the scale of the peak effect found in both papers, which is roughly 2.5 percent at 10-11 quarters after the shock takes place.

We first computationally reproduce all impulse response figures and regression results in Cloyne (2013) using the data and code provided by the authors on the AER website. We are able to perfectly replicate his results, though when reporting p-values have some trouble initially finding exactly those reported in the text as they are not included in the replication package.<sup>1</sup> We then run a number of robustness checks on what we view as the key identified result, the impulse response of output to the fiscal shock. These include altering the lag structure of the VAR for both the endogenous variables and the exogenous policy shock, the inclusion of additional endogenous controls to the estimation, estimating the impulse response using local projections, and an influential analysis to remove influential observations.

Results appear fairly robust with most substantial changes in the significance of impact and peak responses coming from specifications with a small number of lags relative to the baseline used in Cloyne (2013)<sup>2</sup> We also find estimates some-

---

<sup>1</sup>Ultimately we matched the authors' by simply calculating the share of bootstrap simulations below zero.

<sup>2</sup>It would be more troublesome if adding additional lags substantially altered results, as we discuss below.

what sensitive to influential analysis, though the shape of impulse responses stays roughly stable. In general, the results are mainly robust to our tests with initial and peak responses remaining significant at the 95% level in 60% and 70% of cases respectively. The share of robustness checks that remain significant at the 90% level are 77% and 84% when using a more lenient 90% confidence interval. The average  $z$  statistic in our robustness tests is 2.10 for the impact response and 2.19 for peak response. Relative to the  $z$  statistic from the author's baseline these are 104% and 71% respectively. After adjusting the bootstrap method of the author to match those used in our analysis these relative  $z$ -scores are 103% and 97% respectively.

## 2 Reproduction

We choose Figure 3 of [Cloyne \(2013\)](#) as the main benchmark result we will test. This is the response of output (log GDP per capita<sup>3</sup>) to a decrease in taxes. In particular, we replicate estimates of the impact (contemporaneous) and peak response of the narrative fiscal policy shocks. The tax variable is scaled so that it is the percent of GDP so in all results that follow coefficients are the percent response of output to a 1 percent of GDP decrease in taxes. The estimating equation used in this baseline estimation is the vector autoregression (VAR) given by:

$$X_t = A_0 + A_1 t + B(L)X_{t-1} + C(L)d_t + e_t \quad (1)$$

Where  $X_t$  is a set of endogenous regressors,  $t$  is a time trend, and  $d_t$  is the exogenous policy shock.  $B(L)$  and  $C(L)$  are lag polynomials with  $p$  and  $q + 1$  lags, allowing that the lag structure differs between the endogenous regressors and the exogenous shock. In the baseline specification of [Cloyne \(2013\)](#) log per-capita consumption and investment are included as endogenous regressors along with log GDP per capita. The number of lags used for these endogenous regressors (i.e. the lag length of  $B(L)$ ) is 4 quarters, while the lag length for the exogenous shock series

---

<sup>3</sup>Though not reported here, using instead log GDP does not seem to alter results.

is 12 quarters.

## 2.1 Computational Reproduction

We first complete computational reproduction of all the main results of the paper, and using author-provided data and code, were able to reproduce the coefficients exactly. One note is that when replicating the p-values reported in the text for the main result we get slightly different estimates (these are not generated in the author-provided code). We believe that these have been calculated as the share of bootstrap replications below zero (which is conventional), and we get very similar results to the author, but which provides something like a one-sided test for significance, rather than a two-sided p-value we initially calculated. Column one of [Table 1](#) gives the estimates obtained using the author-provided code in Matlab. We conducted our robustness analysis in Stata and provide a reproduction of this result for the author’s baseline model in [Equation 1](#) using the same VAR methodology in Column 3. While these estimates are quite close there are some small differences, which we found puzzling.

While the point estimates are quite close (and we did not figure out a cause for the discrepancy) the difference in standard errors was more troubling, particularly for the peak response. After comparing the author’s bootstrap code to that described in the stata manual we concluded that the main difference here was the failure of the original code to sample bootstrap residuals jointly for all “K” variables. This is potentially important as it preserves contemporaneous correlation that may exist among the residuals of each endogenous regressor. Column 2 adjusts the author’s bootstrap routine to instead draw these jointly. This brings our estimates of standard error quite close together, with differences in significance now largely driven by the small changes in point estimates. We will use these bootstrap methods in all of our analyses that follows and believe the approach is quite strongly justified in this context. Given this change, we deem that our Stata results are quite similar for the impact response, and although the very narrow acceptance of

the 95% confidence interval for the impact now is rejected (extremely narrowly) we take this as replicating nearly identically.

Table 1: Reproduction

	Author code	Author code - Adj. SEs	Reproduction: Stata
Impact Response	0.599	0.599	0.567
S.E.	0.295	0.293	0.294
z	2.033	2.044	1.925
95% CIs	[0.033, 1.189]	[0.023,1.180]	[-0.010,1.144]
Peak Response	2.458	2.458	2.226
S.E.	0.797	1.012	0.981
z	3.083	2.267	2.269
95% CIs	[0.937, 4.009]	[0.385, 4.390]	[0.303,4.150]

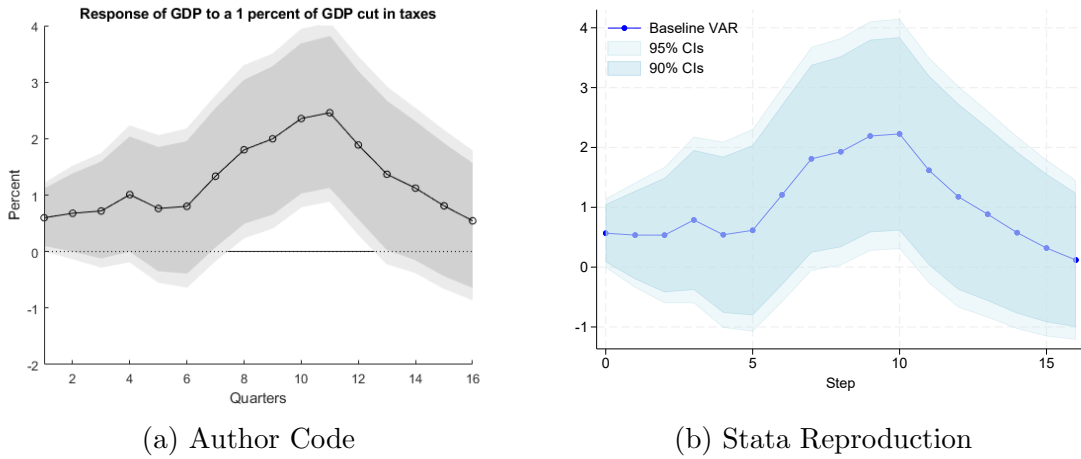
*Author code runs replication files exactly as provided. Because these yielded non-trivial differences in standard errors for peak response relative to our Stata reproduction, we adjusted the bootstrap procedure to better match that used in the Stata var irf command. These are reported in Author code- Adj.SEs.*

While we test both the impact and peak response, we note that the impact response, which just fails the 95% significance threshold here in our Stata code, is not consistently significant at this level in the author's robustness, so this result is secondary to the peak response, which is much stronger. As such we will also report 90% confidence intervals to see when, as in this case, results fail only marginally.

To further emphasize the similarity of these two estimations we graph both the main, author code, and reproduction of Figure 3 from [Cloyne \(2013\)](#) next to our own using the Stata *var* and *irf* commands. While [Cloyne \(2013\)](#) presents 95% and 68% confidence intervals (as was fairly conventional in this literature) we opt to show 90% and 95%. The shapes of these impulse response functions are nearly identical and while there are slightly larger errors (at longer horizons) and small differences in the point estimates we take this as a close replication. In what follows when we refer to the *baseline* model it will be the results of this replication in Stata.



Figure 1: Baseline IRF Reproduction



Reproduction of main results. Confidence intervals are 90% and 95%.

### 3 Robustness Results

#### 3.1 Lag Lengths

We begin by testing the robustness of the baseline model used in the paper to an equivalent specification with varied lag structure. The lags chosen for the exogenous policy shock are set at  $q = 12$ , in keeping with the lags used in [Romer and Romer \(2010\)](#), but sets  $p = 4$  saying this “is common, although I experiment with different  $p$  and the results are robust”. Because these experiments are not reported we will do some here in addition to checking robustness to lag structure in  $q$ , for the exogenous shock. In [Table 2](#) and [Table 3](#) we report the estimated impact and peak responses. Standard errors are calculated using bootstrap with 10000 replications and we report the z-statistic associated with this bootstrapped standard error and the corresponding 90 and 95% confidence intervals.

We find that while the response is sensitive to specifying a small number of lags,  $q$ , of the exogenous fiscal shock, there is very little sensitivity to changes in the lag structure of the endogenous regressors,  $p$ . In particular when only including the past four quarters of the exogenous shock half of the effects on impact and all of the peak responses become insignificant at the 10% level. However, conditional on using

Table 2: Robustness to Varying Lag Lengths

	(1)	(2)	(3)	(4)
q (Lag Exogenous Shock) = 4				
Lags, Endog (p=)	4	8	12	16
Impact Response	0.5650	0.4883	0.6686	0.6754
S.E.	(0.2961)	(0.2974)	(0.2958)	(0.3000)
Z	1.9080	1.6419	2.2602	2.2516
95% CIs	[-0.0154, 1.1454]	[-0.0946, 1.0713]	[0.0888, 1.2483]	[0.0875, 1.2633]
90% CIs	[0.0779, 1.0521]	[-0.0009, 0.9775]	[0.1820, 1.1551]	[0.1820, 1.1688]
Peak Response	0.9572	0.9968	1.0578	1.2278
S.E.	(0.7287)	(0.8308)	(0.9015)	(0.9525)
Z	1.3135	1.1997	1.1734	1.2891
95% CIs	[-0.4711, 2.3855]	[-0.6317, 2.6252]	[-0.7092, 2.8248]	[-0.6390, 3.0947]
90% CIs	[-0.2415, 2.1558]	[-0.3698, 2.3634]	[-0.4251, 2.5407]	[-0.3389, 2.7945]
q (Lag Exogenous Shock) = 8				
Lags, Endog (p=)	4	8	12	16
Impact Response	0.5580	0.4716	0.6466	0.6755
S.E.	(0.2930)	(0.2937)	(0.2963)	(0.2963)
Z	1.9042	1.6055	2.1825	2.2944
95% CIs	[-0.0163, 1.1323]	[-0.1041, 1.0473]	[0.0659, 1.2272]	[0.0985, 1.2525]
90% CIs	[0.0760, 1.0399]	[-0.0116, 0.9547]	[0.1593, 1.1339]	[0.1912, 1.1598]
Peak Response	1.7154	1.8444	1.8699	2.4137
S.E.	(0.9883)	(0.9995)	(1.0660)	(1.1323)
Z	1.7357	1.8453	1.7540	2.1318
95% CIs	[-0.2217, 3.6524]	[-0.1146, 3.8035]	[-0.2195, 3.9593]	[0.1946, 4.6329]
90% CIs	[0.0898, 3.3410]	[0.2003, 3.4885]	[0.1164, 3.6234]	[0.5514, 4.2761]

*Estimates use baseline specification as in our Stata reproduction in column three of Table 1 while varying the lag length of both the exogenous shock and endogenous regressors. Columns represent increasing quarterly lags of endogenous regressions, while upper and lower panels use various quarterly lags of the exogenous shock process from 4 to 8 respectively.*

$q \geq 8$  we find at least 90% significance in all but one estimate of the impact response and all estimates of the peak response, while clearing the 95% significance threshold in 2/3 of the impact specifications and 3/4 of the peak responses. Failing to fully control for the lag structure likely biases estimates; therefore the weak results found when  $q = 4$  are less concerning than would be the case if results that increased the lag length had a similar impact on robustness.

Plotting the full set of impulse responses in Figure 2 for these estimates suggests that there is some convergence toward a profile similar to the baseline case for all

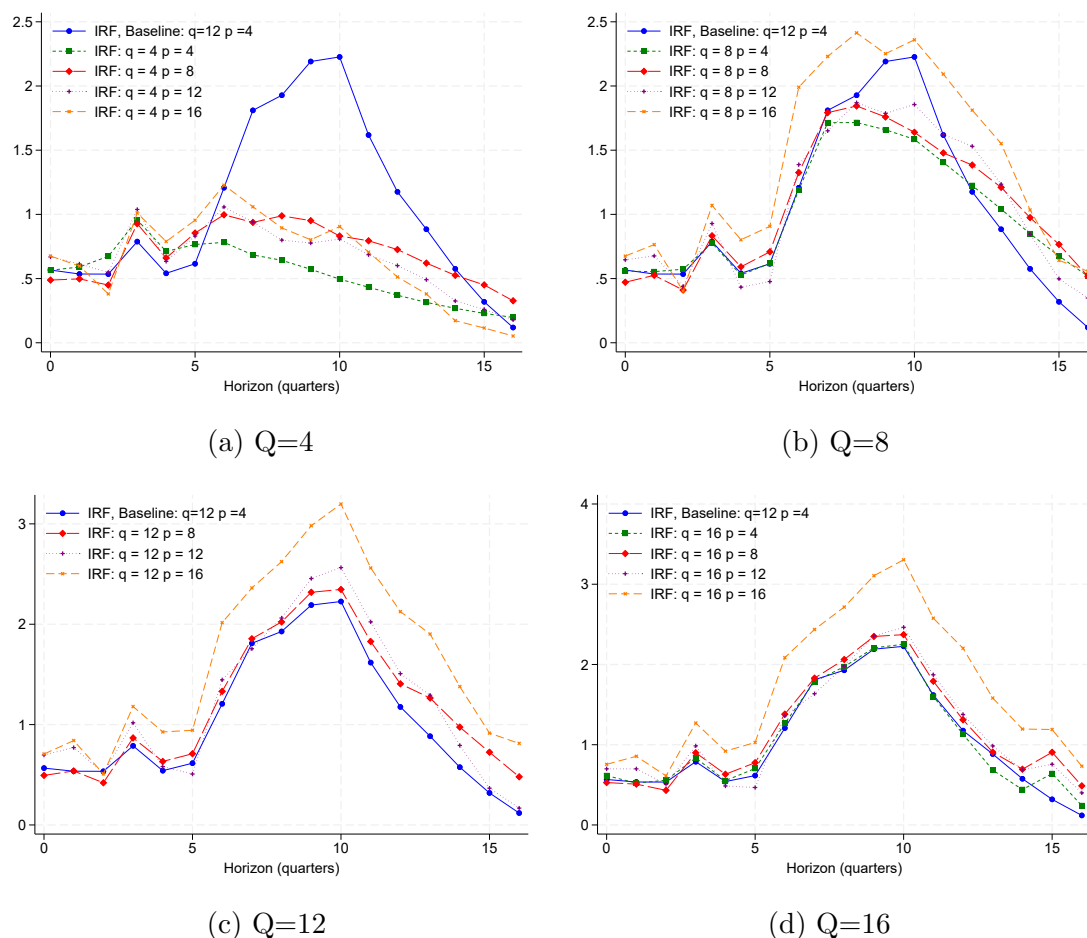
Table 3: Robustness to Varying Lag Lengths

	(1)	(2)	(3)	(4)
q (Lag Exogenous Shock) = 12				
Lags, Endog (p=)	4	8	12	16
Impact Response	0.5666	0.4938	0.6971	0.7066
S.E.	(0.2944)	(0.2950)	(0.2916)	(0.2883)
Z	1.9246	1.6741	2.3908	2.4507
95% CIs	[-0.0104, 1.1437]	[-0.0843, 1.0720]	[0.1256, 1.2685]	[0.1415, 1.2717]
90% CIs	[0.0824, 1.0509]	[0.0086, 0.9791]	[0.2175, 1.1767]	[0.2323, 1.1808]
Peak Response	2.2264	2.3467	2.5643	3.1980
S.E.	(0.9813)	(0.9863)	(0.9863)	(0.9863)
Z	2.2689	2.3792	2.4846	2.8639
95% CIs	[0.3032, 4.1497]	[0.4135, 4.2799]	[0.5415, 4.5872]	[1.0094, 5.3865]
90% CIs	[0.6124, 3.8405]	[0.7243, 3.9691]	[0.8667, 4.2620]	[1.3613, 5.0347]
q (Lag Exogenous Shock) = 16				
Lags, Endog (p=)	4	8	12	16
Impact Response	0.6153	0.5285	0.6996	0.7555
S.E.	(0.2905)	(0.2938)	(0.2892)	(0.2821)
Z	2.1180	1.7987	2.4194	2.6781
95% CIs	[0.0459, 1.1846]	[-0.0474, 1.1043]	[0.1328, 1.2664]	[0.2026, 1.3084]
90% CIs	[0.1374, 1.0931]	[0.0452, 1.0118]	[0.2240, 1.1753]	[0.2915, 1.2195]
Peak Response	2.2523	2.3720	2.4633	3.3046
S.E.	(0.9900)	(0.9898)	(1.0473)	(1.1006)
Z	2.2750	2.3964	2.3520	3.0026
95% CIs	[0.3119, 4.1928]	[0.4320, 4.3120]	[0.4106, 4.5160]	[1.1475, 5.4616]
90% CIs	[0.6238, 3.8808]	[0.7439, 4.0001]	[0.7406, 4.1860]	[1.4943, 5.1148]

*Estimates use baseline specification as in our Stata reproduction in column three of Table 1 while varying the lag length of both the exogenous shock and endogenous regressors. Columns represent increasing quarterly lags of endogenous regressions, while upper and lower panels use various quarterly lags of the exogenous shock process from 12 to 16 respectively.*

lag structures with  $q \geq 8$ . It thus seems reasonable to exclude specifications of lower lag lengths for the exogenous policy shock as lacking enough structure to capture the dynamics of the response of output to the policy shock, and the result appears to be quite strongly robust to changes in lag structure for either endogenous or exogenous controls once above this threshold.

Figure 2: Impulse Response Functions



Comparison of impulse response functions to baseline specification, varying lag lengths.  $Q$  refers to lags on the exogenous fiscal shock, while  $p$  refers to the lag length of endogenous variables in the VAR.

### 3.1.1 Local Projections

We next use the local projections method of estimating impulse response functions in a reduced form approach suggested in [Jordà \(2005\)](#). We report two separate specifications of standard errors, first using standard heteroskedastic robust standard errors, and then using newey-west standard errors to adjust for autocorrelation. Both of these are estimated using standard OLS methods and the same control set and lag structure as the baseline model in Figure 3 of the paper. We also here estimate the impulse response using a quantile regression. The coefficients, standard errors, and confidence intervals are reported in [Table 4](#). Using standard OLS to estimate the local projection we find impulse and peak responses quite close to,

though slightly smaller, than the baseline VAR, but with slightly wider confidence intervals that fail 90% significance on impact. The peak response remains strongly significant. The quantile regression has a much weaker impact response, but a similar peak response that also remains significant.

Table 4: Robustness Local Projections Methods

	(1)	(2)
	LP: Robust	LP: Newey West
Impact Response	0.5666	0.5666
S.E.	0.4049	0.4049
t	1.3994	1.3994
95% CIs	[-0.2270, 1.3603]	[-0.2270, 1.3603]
90% CIs	[-0.0994, 1.2327]	[-0.0994, 1.2327]
Peak Response	2.3969	2.3969
S.E.	0.9237	0.8651
t	2.5950	2.7706
95% CIs	[0.5865, 4.2073]	[0.7013, 4.0926]
90% CIs	[0.8775, 3.9164]	[0.9738, 3.8201]

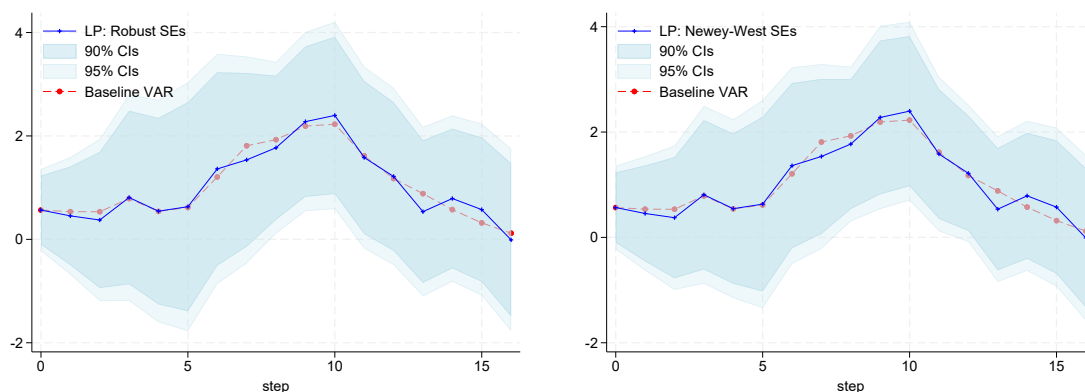
*Estimates replicate the baseline impulse response functions using local projections Jordà (2005). We report impact and peak response with standard errors estimated using both robust and Newey-West standard errors.*

Plotting these against the baseline model in [Figure 3](#) shows that the standard local projection follows the baseline VAR quite closely. This provides a useful robustness check as in principle these methods should estimate the same object, as shown in [Plagborg-Møller and Wolf \(2021\)](#). In small samples, of course, they can look quite different and so their nearly identical shape provides some support for the estimates reported in the paper. Our quantile regression finds less impact effect, but reaches a similar peak, and while there is more noise in the series the two curves follow a similar response.

### 3.1.2 Testing Control Set

[Cloyne \(2013\)](#) uses various sets of controls in robustness tests in the paper, particularly the inclusion of monetary policy rates and government spending, which have

Figure 3: Impulse Response Functions: Local Projections



(a) Local Proj: Robust SEs

(b) Local Proj: NW SEs

Comparison of impulse response functions to baseline specification, estimation by local projections.

no impact on their main result. The author also shows that the result is robust to splitting the sample based on monetary policy regimes or exchange regimes (the impact effect is weak during the fixed exchange sample). In this section we merge two new variables to the dataset: UK total stock market share, and UK ten-year interest rates. Our goal is to see if these forward-looking financial variables have any impact relative to the various macroeconomic controls already used. In addition, we run tests using some of the variables not reported for robustness in [Cloyne \(2013\)](#), but included in his data. In particular: nominal wages and employment<sup>4</sup>, a production index, and specifications that exclude consumption and investment. We also report the estimate using no endogenous controls at all. In keeping with the results found in the paper the result seems relatively insensitive to the specification of the endogenous controls. This is what should be the case if the narrative shock is properly specified.

[Table 5](#) includes combinations of the two additional endogenous controls from the baseline specification. First testing the specification with no additional controls<sup>5</sup> along with using only investment and consumption, rather than both as in

<sup>4</sup>The response of these variables are of note and studied in the paper, but the response of output is not discussed when they are included

<sup>5</sup>Which should be the “correct” specification if the narrative instrument is truly exogenous.

the baseline. There are quite small differences across these specifications, and if anything estimates become more strongly significant.

Table 5: Robustness: Exclusion of Baseline Controls

	(1)	(2)	(3)
	No Endog Controls	Only Investment	Only Consumption
Impact Response	0.7676	0.7796	0.8113
S.E.	0.2879	0.2982	0.2911
z	2.6658	2.6144	2.7867
95% CIs	[0.2032, 1.3319]	[0.1952, 1.3641]	[0.2407, 1.3819]
90% CIs	[0.2940, 1.2412]	[0.2891, 1.2701]	[0.3324, 1.2902]
Peak Response	2.2691	2.2565	2.5129
S.E.	1.0141	1.0280	1.0600
z	2.2376	2.1951	2.3706
95% CIs	[0.2816, 4.2566]	[0.2417, 4.2714]	[0.4352, 4.5906]
90% CIs	[0.6011, 3.9371]	[0.5656, 3.9475]	[0.7693, 4.2565]

*No endogenous controls estimates the response of output, with baseline lag structure (4 quarters) to the exogenous policy shock with no other controls. Only Investment includes only investment in addition as an endogenous control and only consumption includes log consumption. Including both brings the specification back to the baseline case. Standard errors estimated by bootstrap.*

We next check robustness to including other variables that are present in the replication data from [Cloyne \(2013\)](#). Some of these, such as employment and wage, are studied in [Cloyne \(2013\)](#), but the paper does not include the response of output when they are studied. We choose employment, employment and wage, and an index of production as these might contain indicators relevant to fiscal policymakers with respect to the business cycle, and therefore might pose a threat to the narrative identification if indeed they provide predictable information on the narrative shock. This does not appear to be the case in [Table 6](#), where results are slightly stronger than the baseline and remain significant at the 95% level.

Finally in [Table 7](#), we introduce two new variables to the data. These are chosen as indicators that are inherently forward-looking, and therefore might contain information that could pose a threat to identification if they contain relevant correlation that predicts the fiscal shock. We choose the 10-year interest rate, and the total value of UK stocks. These are included along with baseline controls individually

Table 6: Robustness: Other Controls in [Cloyne \(2013\)](#) Data

	(1)	(2)	(3)
Baseline +	Employment	Employment + Wage	Production Index
Impact Response	0.6231	0.6200	0.8320
S.E.	0.2851	0.3074	0.2860
z	2.1854	2.0173	2.9088
95% CIs	[0.0643, 1.1820]	[0.0176, 1.2224]	[0.2714, 1.3926]
90% CIs	[0.1541, 1.0921]	[0.1145, 1.1256]	[0.3615, 1.3024]
Peak Response	3.0243	3.3599	2.4970
S.E.	1.3457	1.4127	1.0902
z	2.2474	2.3783	2.2904
95% CIs	[0.3868, 5.6618]	[0.5910, 6.1287]	[0.3603, 4.6338]
90% CIs	[0.8108, 5.2377]	[1.0361, 5.6836]	[0.7038, 4.2902]

*We test the robustness of results to other variables that are provided in the replication data. We use baseline model including consumption and investment with 4 quarter lags on endogenous controls and 12 quarter lags on the fiscal shock while including: employment, employment and wages, and an index of industrial production.*

and together, in all cases both impact and peak response are larger and more significant. We note that both time series are taken from FRED [citation needed], and begin in 1960 so omit the first five years of the baseline sample.

We find no evidence that the choice of control set has a meaningful impact on the primary results in [Cloyne \(2013\)](#). We note that the author carries out many robustness checks of this (and other) types in the appendix of the paper, in particular controlling for macroeconomic policy (fiscal and monetary). These also appear to have little effect.

### 3.1.3 Influential Analysis

To conduct influential analysis we first estimate a quantile version of our local projections approach to the baseline model specification. This entails simply estimating each response horizon using quantile regression rather than standard OLS. This approach substantially alters the impact response of the policy, but peak responses are quite similar to those reported in the paper, and although somewhat noisy, the impulse response looks similar to the [Romer and Romer \(2010\)](#) response with which



Table 7: Robustness: Other Controls Forward Looking

	(1)	(2)	(3)
Baseline +	10yr Interest	Stock Market Valuation	10yr Interest & Stocks
Impact Response	0.8169	0.8072	1.4188
S.E.	0.2891	0.2889	0.2716
z	2.8260	2.7937	5.2246
95% CIs	[0.2503, 1.3835]	[0.2409, 1.3735]	[0.8866, 1.9511]
90% CIs	[0.3414, 1.2924]	[0.3319, 1.2824]	[0.9721, 1.8655]
Peak Response	2.8929	2.5477	3.3391
S.E.	0.9355	1.0412	1.0321
z	3.0924	2.4469	3.2352
95% CIs	[1.0594, 4.7264]	[0.5070, 4.5884]	[1.3162, 5.3620]
90% CIs	[1.3541, 4.4316]	[0.8351, 4.2603]	[1.6415, 5.0368]

*We consider two additional controls in the baseline estimation that are not included in the original replication data: 10 year interest rates on government bonds, and current stock market valuation. Column three includes both additional endogenous controls.*

Cloyne (2013) draws primary comparisons.<sup>6</sup> These impact and peak responses are reported in Column 1 of Table 8, with the full impulse response function reported in Figure 4a.

Consistent with our other robustness analyses we use the dfbeta statistic to purge outliers, using a standard cutoff of  $2/N^{-.5}$ . For local projections, we estimate this cutoff at each horizon as the observations vary slightly as the horizon increases. Since this statistic does not have a comparable VAR alternative we use the cutoff from the local projection at the peak horizon ( $h = 11$ ) for the baseline case, though cutoffs are similar at all horizons, this produces a cutoff for our policy shock of  $2/N^{0.5} = 0.1397$ . This results in a sample of  $N = 193$ , dropping 23 observations from the baseline estimation. This is a more extreme version of the outlier analysis done in Appendix E of Cloyne (2013), where the author accounts for two large changes in the exogenous tax series in 1979Q3-Q4 by adjusting the series to smooth out the large shocks. We report local projections using the horizon-specific cutoff in Column 2 of Table 8 and the VAR using the exclusion restriction from the peak  $h = 11$  horizon in Column 3 of Table 8. The full impulse response functions are

<sup>6</sup>Similarly that response had a zero impact effect, though here error bars are much larger

plotted against the baseline VAR in [Figure 4b](#) and [Figure 4c](#). While both retain somewhat similar profiles to the baseline case, the local projection now has wide confidence intervals and is not significant. The VAR specification remains significant at the 90% level at peak response (and narrowly misses the 0.95% threshold).

This cutoff substantially reduces the shock, bringing down the mean of the absolute value of the narrative shock from 0.09829 to 0.07487.<sup>7</sup> We are conscious that in the case of narrative fiscal shocks excluding roughly 10% of the estimation sample with a bias toward large values of the shock series is perhaps less reasonable than in other contexts as these narrative shocks are relatively rare (the series is zero in half of the sample), and large fiscal surprises perhaps contain the information most relevant to causal identification.

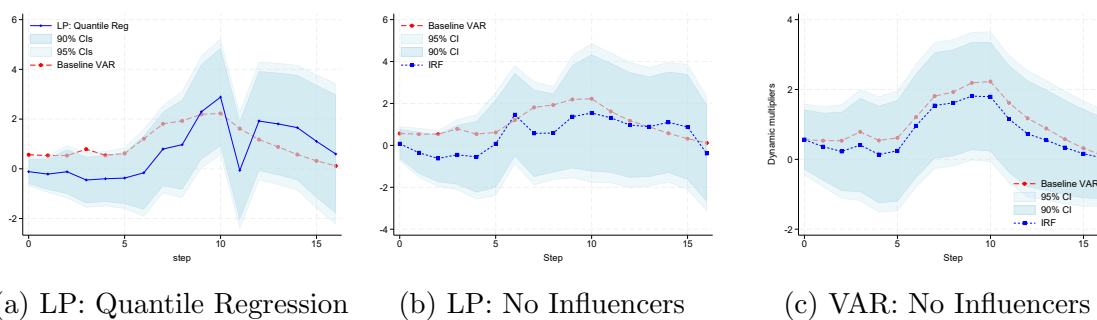
Table 8: Influential Analysis

	(1)	(2)	(3)
	LP: Quantile	Local Projection (baseline)	VAR p=4,q=12
Impact Response	-0.1128	0.0806	0.5668
S.E.	0.2970	0.4193	0.5219
t/z	-0.3797	0.1922	1.0861
95% CIs	[-0.6949, 0.4693]	[-0.7413, 0.9025]	[-0.4560, 1.5897]
90% CIs	[-0.6013, 0.3758]	[-0.6092, 0.7704]	[-0.2917, 1.4253]
Peak Response	2.2953	1.5446	1.8104
S.E.	1.1582	1.6936	0.9365
z	1.9817	0.9120	1.9331
95% CIs	[0.0252, 4.5655]	[-1.7748, 4.8641]	[-0.0252, 3.6460]
90% CIs	[0.3900, 4.2006]	[-1.2414, 4.3306]	[0.2698, 3.3510]

*Influential analysis is conducted by estimating three specifications of the model. First we consider quantile regression in the context of local projections. In columns 2 report estimates for impact and peak allowing for removal of influential observations. For VAR estimates we eliminate both impact and peak responses reflect estimates that exclude the influential observations as identified for the peak effect in the local projections estimation (column 2).*

<sup>7</sup>The mean absolute value of the exogenous series in the exclude data is 0.191, while the mean absolute for those that remain below is 0.07487

Figure 4: Baseline IRF Reproduction



(a) LP: Quantile Regression

(b) LP: No Influencers

(c) VAR: No Influencers

## 4 Conclusion

Overall, we find that the positive impact of a percentage reduction in tax rates is robust at the 95% confidence level in 70% of the 30 robustness checks we ran for peak effect, and 60% for impact effect. These rise to 84% and 77% respectively when using 90% confidence intervals, and in these cases, most of the remaining rejections come from specifications with limited lag structures that likely fail to account for the appropriate lag structure in the exogenous shock series. Our estimates for peak impact are notably less significant than those reported in the paper, with an average z-score 71% of that reported in the paper, but this appears to be entirely an artifact of choices on residual sampling in the bootstrap procedure of the author, with his baseline estimates falling nearly exactly at the average of those we find (and remaining statistically significant) when adjusting his replication file to match ours. The shape of the output response to these narrative tax shocks appears fairly robust to our tests. In general we find these results to be fairly robust.

## References

- Campbell, D., Brodeur, A., Dreber, A., Johannesson, M., Kopecky, J., Lusher, L. and Tsoy, N.: 2024, The robustness reproducibility of the american economic review, *Technical report*.
- Cloyne, J.: 2013, Discretionary tax changes and the macroeconomy: new narrative evidence from the united kingdom, *American Economic Review* **103**(4), 1507–1528.
- Jordà, Ò.: 2005, Estimation and inference of impulse responses by local projections, *American economic review* **95**(1), 161–182.
- Plagborg-Møller, M. and Wolf, C. K.: 2021, Local projections and vars estimate the same impulse responses, *Econometrica* **89**(2), 955–980.
- Romer, C. D. and 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.