

When to Raise Taxes: State Dependence in the UK Tax Multiplier

Alessandro Conway

Working Paper AC-WP-2026-06

Version 1.0 (September 2026)

Not peer reviewed. Comments welcome.

Abstract

Using UK narrative tax measures from 1918 to 2018 and quarterly household consumption from 1955, I estimate the response of consumption to tax changes that were not anticipated, and split it by whether unemployment was above or below its own recent average when the change took effect. Against a tax rise worth one per cent of annual GDP, consumption falls by 7.3 per cent at its trough when there is slack and by 3.2 per cent when there is not, so the same rise costs roughly twice as much in a slack economy. The difference is estimated directly, with its own standard error, and is significant at eight of seventeen horizons. Consumption is also the better outcome to study: it responds more than output, on three times as many horizons, and unlike the output response it survives the removal of the single Budget on which the published UK multiplier depends. The shock series reproduces Cloyne (2013) exactly before it is extended, which settles the dating conventions the estimate relies on. The aggregate record is too thin to separate one tax from another or to detect whether pre-announcement weakens the effect, and whose spending falls is a question for household data, which the paper identifies.

Keywords: fiscal multipliers, narrative tax shocks, state dependence, local projections, household consumption

JEL classification: E21, E32, E62, H20, H30

Suggested citation: Conway, Alessandro (2026). “When to Raise Taxes: State Dependence in the UK Tax Multiplier.” Working Paper AC-WP-2026-06, Version 1.0.

© Alessandro Conway 2026. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

When to Raise Taxes: State Dependence in the UK Tax Multiplier

Alessandro Conway*

First draft September 2026

Abstract

Using UK narrative tax measures from 1918 to 2018 and quarterly household consumption from 1955, I estimate the response of consumption to tax changes that were not anticipated, and split it by whether unemployment was above or below its own recent average when the change took effect. Against a tax rise worth one per cent of annual GDP, consumption falls by 7.3 per cent at its trough when there is slack and by 3.2 per cent when there is not, so the same rise costs roughly twice as much in a slack economy. The difference is estimated directly, with its own standard error, and is significant at eight of seventeen horizons. Consumption is also the better outcome to study: it responds more than output, on three times as many horizons, and unlike the output response it survives the removal of the single Budget on which the published UK multiplier depends. The shock series reproduces Cloyne (2013) exactly before it is extended, which settles the dating conventions the estimate relies on. The aggregate record is too thin to separate one tax from another or to detect whether pre-announcement weakens the effect, and whose spending falls is a question for household data, which the paper identifies.

Keywords: fiscal multipliers, narrative tax shocks, state dependence, local projections, household consumption.

JEL classification: E21, E32, E62, H20, H30.

1 Introduction

A Chancellor who has decided to raise taxes still has to decide when. On that second decision the empirical literature offers surprisingly little, because most of it reports a single number: the average response of output to a tax change, estimated over whatever sample is available. That number is a useful guide to timing only if the response is roughly the same in good years and bad, and this paper tests the assumption on the UK record. It does not hold, since the same tax rise is followed by roughly twice the fall in household spending when the economy has slack.

*Working paper. I thank James Cloyne for making the underlying narrative dataset public, and Cloyne, Dimsdale, Postel-Vinay and Hürtgen for depositing the interwar measures and the macroeconomic panel used here, without which this paper could not exist. A companion paper, *The anticipation content of fiscal policy*, builds the measure-level record this paper estimates on.

The exercise is possible because of a long narrative record. Romer and Romer (2010) identify legislated US tax changes from the contemporaneous documents and code the motivation behind each, so that changes made for reasons unrelated to the state of the economy can be separated from those made in response to it. Cloyne (2013) builds the corresponding UK account for 1945 to 2009, and it has since been extended back into the interwar period by Cloyne et al. (2023) and forward to 2020 by Hürtgen et al. (2024), with the combined series used at century length by Cloyne et al. (2025). A companion paper (Conway, 2021) extends the measure-level record from Budget 2004 to Autumn 2018. It shows that the interval between announcing a UK tax change and its taking effect has roughly quadrupled since 1945, and that the share of the tax impulse arriving with more than four months of notice has risen from 0.18 to 0.70.

That finding limits what can be estimated here. A measure announced two years before it takes effect is no surprise on the day it does, and treating it as one would estimate something other than a causal response. The projections below therefore use only measures whose announcement and implementation nearly coincide. They are a minority of the modern record, which costs a good deal of precision, but without the restriction the estimate would have no causal reading.

Whether the multiplier depends on the state of the economy has been asked before. Auerbach and Gorodnichenko (2012) find larger output responses to fiscal policy in US recessions, and Ramey and Zubairy (2018) dispute the finding for government spending, using a longer sample and a different treatment of the news. Neither paper settles the question for tax changes or for the United Kingdom. This paper gives a UK answer, on household consumption rather than output, from a narrative series that reproduces the published one exactly before extending it.

Against a tax rise worth one per cent of annual GDP, household consumption falls by 7.3 per cent at its trough when unemployment is above its own recent average, and by 3.2 per cent when it is not. Per pound raised, the same rise costs roughly £4.50 of household spending in a slack economy and roughly £2 in a tight one. The gap between the two is estimated directly rather than inferred from two separate regressions, and it is significant at eight of seventeen horizons.

The aggregate record also has limits, which Section 5 sets out. Most instruments amount to one or two episodes, so the record cannot separate one tax from another. Whether pre-announcement weakens the response cannot be settled either, since the test lacks the power to detect an effect even as large as the one it is compared against. The question of who bears the cost is harder still. Splitting each measure by the incidence of its instrument produces two series that move together almost perfectly, so the answer needs households observed directly, and the section names the survey that can provide them.

2 Data

2.1 The narrative record

Three public sources, all carrying announcement dates, implementation dates, a revenue costing, and the same motive classification, are chained into one measure-level record.

The identifying assumption is the exogeneity restriction of the original coders, so a measure

Table 1: The measure-level record

Period	Source	Datable measures	Of which exogenous
1918–1938	Cloyne et al. (2023), Zenodo, CC-BY	204	89
1945–2009	Cloyne (2013)	1,221	1,221
2004–2018	Conway (2021)	173	173

Exogenous counts are of measures usable in estimation. No source covers 1939 to 1944, and 1921 contains no datable measure, so a century-long specification has gaps that the estimation has to allow for. The interwar block has no tax-type field and so cannot enter any instrument-level specification.

is used only where the coder judged its motivation unrelated to the current state of the economy. The companion paper does without the restriction, since it studies how long the policy process takes, but a causal estimate cannot.

One design choice departs from the companion paper, which restricted attention to household-relevant measures using Cloyne’s **Group** field, a field his own documentation disclaims. The interwar block has no such field, so the restriction cannot be carried back, and dropping it also matches the construction of the published exogenous series. I built the household-restricted variant alongside, and the choice does not change any result below.

2.2 Replication

Small differences in how a narrative series is dated and aggregated matter more than they look, so before extending anything I made the construction reproduce Cloyne (2013) exactly.

His aggregation code, published with his dataset, pushes an implementation date forward by 45 days before taking its calendar quarter, replaces an implementation date that precedes its own announcement with the announcement date, treats a measure as a surprise when it takes effect within 90 days, and divides by annualised nominal GDP. Rebuilding his 1955 to 2009 surprise series from his raw file under those rules returns it at a correlation of 1.0000, with a maximum difference of zero across all 220 quarters.

The same exercise shows how much each convention matters. Table 2 changes one convention at a time from his baseline and reports the correlation of the result with his published series.

Table 2: Sensitivity of the series to its dating conventions

Construction	Correlation with the published series
His conventions exactly	1.0000
A 120-day surprise threshold instead of 90	0.9236
A half-month quarter rule instead of his 45-day shift	0.9316
Retroactive dates kept rather than replaced	0.9648
A different nominal GDP vintage	0.9990
All four differences together	0.8643

The GDP vintage makes almost no difference and the threshold makes the most. The companion paper uses a 120-day rule, for reasons specific to the British fiscal calendar; this paper keeps the 90-day convention so that its estimates are comparable with the published multiplier literature.

2.3 Outcomes

The macroeconomic panel is the one Cloyne et al. (2025) estimate on, taken from their replication deposit: quarterly real and nominal GDP, the deflator, unemployment, Bank Rate, the exchange rate, prices, revenue, expenditure, and the deficit, from 1918 to 2020, sourced from Thomas and Dimsdale (2017). Using their outcomes rather than assembling my own keeps every comparison with the published estimates like for like, and their narrative government spending shocks over the same span enter as a control.

Household consumption is not in that deposit, so I take it from Thomas and Dimsdale (2017) directly; since that is the source underlying the panel, consumption and output come from a single vintage. Their real GDP and the panel’s agree at a correlation of 1.0000 on levels and 0.9999 on growth. The expenditure components begin in 1955Q1 and version 3.1 ends in 2016Q4, so the consumption estimates use 248 quarters, against 1945 to 2018 for output.

2.4 Identifying variation

Table 3 shows how scarce the identifying variation is in the modern period.

Table 3: Standard deviation of the quarterly shock, per cent of annual GDP

Era	Quarters	Unanticipated	Non-zero quarters	News
1920–38	76	0.227	14	0.012
1945–79	140	0.259	41	0.317
1980–99	80	0.158	40	0.125
2000–18	76	0.046	35	0.142

The number of quarters carrying a surprise stays roughly constant while the amplitude of the series falls by four-fifths, so surprises continued to occur but became small. This is the companion paper’s central finding seen from the side of estimation, and it is the reason no result below is reported for the modern period alone.

3 Consumption and output

For horizon h , the projection is

$$100 \times (\log y_{t+h} - \log y_{t-1}) = \alpha_h + \beta_h s_t + \Gamma_h(L) + \varepsilon_{t+h},$$

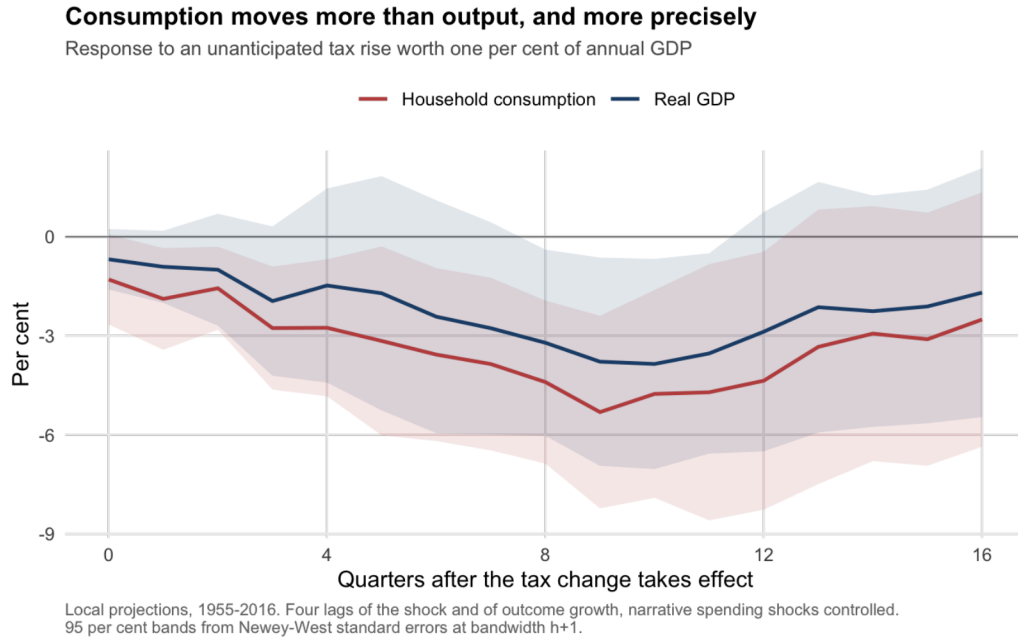
following Jordà (2005), where s_t is the unanticipated tax shock as a per cent of annual GDP, and the controls are four lags of the shock, four lags of outcome growth, and the contemporaneous narrative spending shock. Standard errors are Newey-West at bandwidth $h + 1$. A positive shock is a tax rise, so a contractionary response is $\beta_h < 0$, and β_h is the multiplier in the sense of Romer and Romer (2010).

Output and consumption respond quite differently (Table 4). Output falls by 3.86 per cent at its trough, significant at four of seventeen horizons, and as the next paragraph shows the estimate depends largely on one Budget. Consumption falls further, by 5.31 per cent, and the fall is significant at twelve horizons.

Table 4: Response to an unanticipated tax rise worth one per cent of annual GDP

Outcome	Trough	Quarter	95 per cent interval	t	Significant horizons
Real GDP	-3.86	10	$[-7.04, -0.67]$	-2.37	4 of 17
Household consumption	-5.31	9	$[-8.23, -2.40]$	-3.57	12 of 17

Consumption is 61 per cent of GDP and falls by more than output in absolute terms, so households account for more than their share of the fall. Figure 1 shows both paths.

**Figure 1:** Responses of household consumption and real GDP to an unanticipated tax rise, with 95 per cent bands.

The stronger reason to prefer consumption is that it does not depend on a single Budget, while the published UK tax multiplier largely does. In Cloyne's own 1955 to 2009 surprise series the June 1979 Budget, which raised VAT to 15 per cent, is the largest single observation, at +1.596 per cent of GDP. Removing that quarter takes his aggregate output response from $t = -2.45$, with four significant horizons, to $t = -1.46$, with none, and the same happens with the series built here and with the 2025 revision. The consumption response survives the same test: without that quarter the trough is -3.78, with $t = -2.27$, so the sign, most of the magnitude, and the significance all remain.

The same Budget is also coded very differently across vintages, at +1.596 per cent of GDP in the 2013 coding and +0.377 in the 2025 revision. The largest postwar observation in this literature therefore shrank by a factor of four between two releases of the same authors' data. To the extent that an estimate depends on it, the estimate measures a coding decision, and I would give more weight to results that survive its removal.

4 State dependence

The state variable is an indicator for unemployment above its own seven-year backward moving average. Because it looks only backwards, it uses nothing a policymaker would not have known at the time, and it needs no estimate of the output gap and no filter. On this definition 149 of the 248 usable quarters are in slack, and Table 5 reports the response in each state.

Table 5: Consumption response by the state of the economy

State	Trough	Quarter	95 per cent interval	t	Significant horizons
Economy has slack	-7.31	11	$[-9.55, -5.06]$	-6.38	14 of 17
Economy does not	-3.16	8	$[-6.53, +0.20]$	-1.84	4 of 17

The narrative spending control is identically zero within these subsamples and is dropped from them. Removing the largest single quarter in each state leaves the slack trough at -7.37 with $t = -4.32$.

Two estimates that differ from each other do not by themselves show that the difference is significant, so Table 6 estimates it directly, by interacting the shock with the state indicator, which gives the gap its own standard error.

Table 6: The difference between the two states, estimated directly

Quarters after the change	Slack minus tight	Standard error	t
4	-0.41	1.78	-0.23
8	-3.50	1.95	-1.80
9	-5.44	1.69	-3.22
11	-8.25	2.14	-3.86
12	-8.16	2.30	-3.54
16	-6.47	2.46	-2.63

The difference is significant at eight of seventeen horizons. It opens from about the fifth quarter and is widest at the eleventh, where the slack response also reaches its trough.

Figure 2 shows the two paths. They are indistinguishable for the first four quarters and then separate. The response in a slack economy keeps falling and stays down, while the response in a tight one reaches a shallow trough at two years and returns to zero.

Taking consumption at 61 per cent of GDP, a tax rise of one per cent of GDP is followed at its worst by a fall in household spending of about 4.5 per cent of GDP when there is slack, against about 2 per cent when there is not. In both states the trough comes two to three years after the change, so the result concerns the medium run.

4.1 Notice and the state of the economy

The multiplier does appear to fall as pre-announcement becomes the norm, from a trough of -4.52 over 1955 to 1985 to -2.97 over 1986 to 2016. The later interval spans zero and the two overlap heavily, however, so the most this supports is a tentative direction.

There is a second possible link with the companion paper, though it matters less than it might at first appear. If tax changes are now scheduled years ahead, a Chancellor may not know the state of the economy in which a measure will take effect, and combined with the result

The same tax rise costs twice as much in a slack economy

Consumption response to an unanticipated tax rise worth one per cent of annual GDP

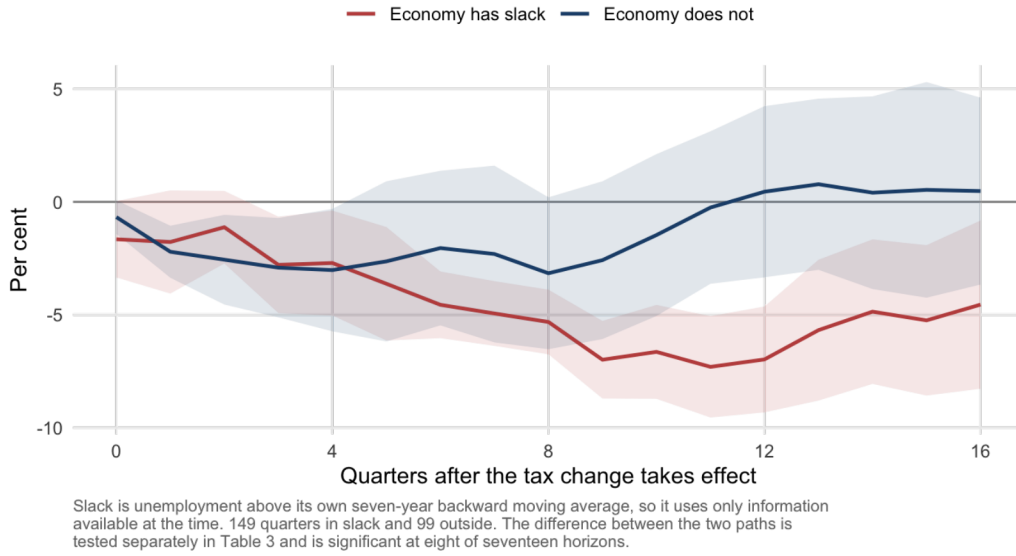


Figure 2: Consumption response by state, with 95 per cent bands.

above that would be a first-order cost of pre-announcement. The correlation between slack at announcement and slack at implementation does fall with notice, from 0.995 for measures taking effect within 90 days to 0.806 for those taking more than a year, and the typical state moves by 0.74 points compared with 0.03. The loss of foresight is real but modest, since unemployment is persistent enough that even two years ahead a Chancellor has a fair idea of the state.

5 Limits of the aggregate record

Several natural extensions of this paper lie beyond what the aggregate record can support, and for each of them the figure that settles the matter is given below.

Individual taxes. On measure counts, instrument-level multipliers look estimable, with 449 income tax measures, 350 excise measures, and 175 corporate ones, but the variation behind them is far thinner. Eighty-four per cent of the absolute variation in the VAT surprise series sits in three quarters, the largest of them June 1979, and the corresponding figures are 64 per cent for capital gains and 47 per cent for corporation tax. Each resulting estimate is in effect a single episode extrapolated to a shock worth one per cent of GDP, which for capital gains means an extrapolation of fifty standard deviations, and the estimates duly come out at implausible magnitudes. A tax-by-tax decomposition is therefore beyond what the UK narrative record can support.

Pre-announcement. The obvious question to carry over from the companion paper is whether pre-announced tax changes move consumption less, because households have already adjusted by the time they take effect. The announcement-dated series shows no detectable response at any horizon, but this null is uninformative. The minimum detectable effect of the design,

meaning the smallest true response it would find four times in five, is 10.5 per cent, against a surprise response of 5.3. The test therefore cannot distinguish the absence of an announcement effect from one exactly as large as the surprise.

Incidence. If a pound taken from a household with no savings cuts spending by more than a pound taken from one with savings, the aggregate response above conceals a distribution. Two tax packages of the same size can then cost different amounts depending on whom they fall on. That is the mechanism Cloyne et al. (2025) and the heterogeneous-agent literature would predict. The ONS Effects of Taxes and Benefits tables report, for every year from 1977, what each income decile pays in income tax, National Insurance, council tax, VAT, and each duty. On that cash incidence 1,036 measures carrying 83 per cent of gross value can be placed. Corporation tax, capital gains, inheritance, and North Sea measures cannot, because the tables never allocate them to households.

The incidence gradient itself is clear and in the expected order. The bottom five deciles bear 0.36 of duties, 0.31 of VAT, 0.19 of stamp duty, 0.16 of National Insurance, and 0.15 of income tax, so the share of duties borne by the lower half of the distribution is more than twice the share of income tax (Figure 3). Splitting each measure into the part borne by the bottom five deciles and the part borne by the top five, and projecting consumption on each, gives troughs of -14.3 and -5.1 , with the right sign and the expected ordering. They are not an estimate of heterogeneity, however: the two series correlate at 0.93 because they are slices of the same measures, the interval for the bottom-borne part runs from -31.3 to $+2.7$, and standardised to a one standard deviation shock the two responses are 0.60 and 0.65 per cent of consumption. Since nothing in the aggregate data distinguishes one slice from the other, the identification has to come from households observed directly.

The Living Costs and Food Survey and its predecessor, the Family Expenditure Survey, record household expenditure from 1961 and allow the consumption response to be estimated by position in the income or liquidity distribution. The specification is written and the aggregate it decomposes is estimated above, so the household estimation is the next step of this work.

6 Conclusion

On UK data from 1955 to 2016, the same unanticipated tax rise is followed by roughly twice the fall in household consumption when unemployment is above its recent average as when it is below. The difference is large enough, and estimated precisely enough, to bear on when a government chooses to consolidate, so the timing of a tax rise appears to be an economic question as well as a political one.

The result should be read with its limits in mind. It concerns the medium run, since the gap opens after a year and is widest at eleven quarters, and it says nothing about the quarter of impact. It is also a result about the average tax rise, because the record cannot separate one instrument from another and cannot yet say whose spending falls; the household surveys named in Section 5 are the natural next step.

The work also turned up something I did not set out to find. The published UK tax multiplier depends on a single Budget, in June 1979, whose coded value shrank by a factor

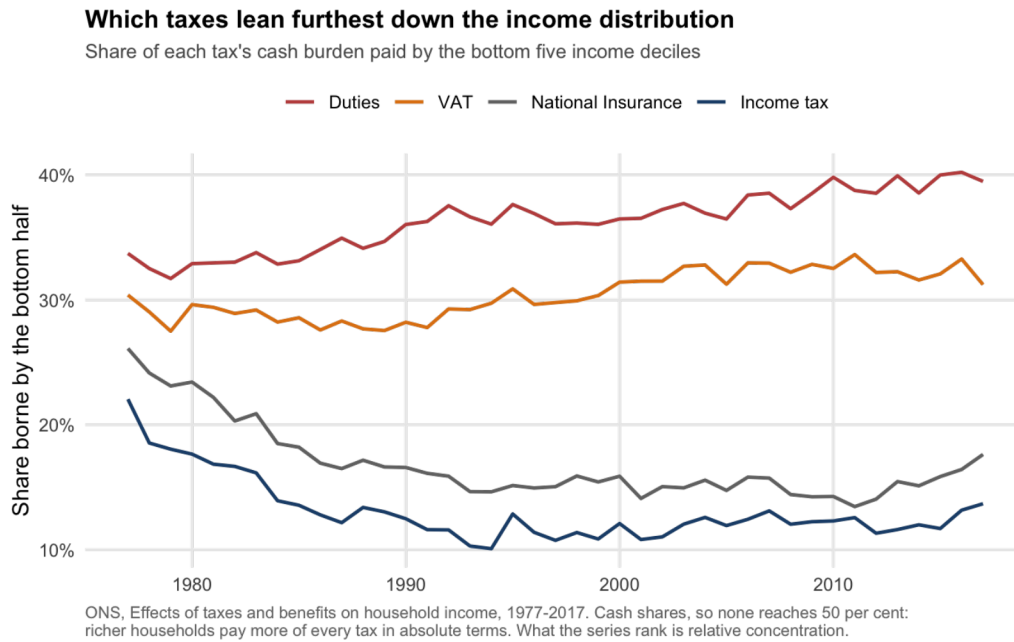


Figure 3: Share of each tax's cash burden borne by the bottom five income deciles.

of four between two vintages of the same dataset. The output response disappears when that Budget is removed, while the consumption response holds. Results on this record are therefore worth checking against its removal before they are read as statements about the British economy.

References

- Auerbach, A. J. and Gorodnichenko, Y. (2012), 'Measuring the output responses to fiscal policy', *American Economic Journal: Economic Policy* **4**(2), 1–27.
- Cloyne, J. (2013), 'Discretionary tax changes and the macroeconomy: new narrative evidence from the United Kingdom', *American Economic Review* **103**(4), 1507–1528.
- Cloyne, J., Dimsdale, N. and Postel-Vinay, N. (2023), 'Taxes and growth: new narrative evidence from interwar Britain', *Review of Economic Studies*.
- Cloyne, J., Hürtgen, P. and Dimsdale, N. (2025), 'Are tax cuts contractionary at the zero lower bound? Evidence from a century of data', *Journal of Political Economy* **133**(2), 568–603.
- Conway, A. (2021), 'The anticipation content of fiscal policy: UK tax measures, 1945–2019'. Working Paper AC-WP-2021-01, Version 2.0.
- Hürtgen, P., Cloyne, J., Dimsdale, N. and Postel-Vinay, N. (2024), Tax changes in the United Kingdom 2009–2020: a new narrative account and dataset, Technical report, SSRN Working Paper 4909051.
- Jordà, O. (2005), 'Estimation and inference of impulse responses by local projections', *American Economic Review* **95**(1), 161–182.

- Ramey, V. A. and Zubairy, S. (2018), ‘Government spending multipliers in good times and in bad: evidence from US historical data’, *Journal of Political Economy* **126**(2), 850–901.
- 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.
- Thomas, R. and Dimsdale, N. (2017), A millennium of macroeconomic data for the UK, Technical report, Bank of England OBRA dataset, version 3.1.