The effect of UK fiscal policy on output: a historical narrative approach

The effect of UK fiscal policy on output: a historical narrative approach

Birkbeck, University of London

João Soares Coelho Pinto de Sousa

A thesis submitted for the degree of Doctor of Philosophy in Economics

June 2024

Declaration

Except where specific reference is made to the work of others, the contents of this thesis are my own and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other university.

Abstract

This thesis contributes to the debate on the effectiveness of fiscal policy in stimulating output, investigating how to measure that effectiveness, what it has been in the UK over the 140 years from 1879-80 to 2018-19, and under what conditions these effects might differ. This thesis makes use of original archival research to identify fiscal shocks from UK parliamentary documents, both in terms of discretionary spending and fiscal stance.

Chapter 2 uses data from the Jordà-Schularick-Taylor macrohistory database to illustrate the sensitivity of multiplier estimates to modelling decisions when using the Gordon-Krenn transformation. This transformation consists of dividing output and government spending by potential output, and this choice of potential output estimation method can lead to multiplier estimates as low as -0.04 and as a high as 0.70. This huge parameter uncertainty for what is a seemingly innocuous choice and which results in very similar potential output estimates. Instead, I propose returning to estimating an output elasticity with respect to government spending and then multiplying it by a conversion ratio such as the inverse of the share of government in total output. This is not only more transparent, but it also results in narrower variation than that induced by the different methods of estimating potential output.

Chapter 3 makes use of a novel dataset comprising of government spending shocks going back 140 years to estimate UK-specific multipliers. This dataset is compiled from archival research in the UK Parliament, consisting of changes between the estimate for government spending at the beginning of the financial year and the estimate in the subsequent budget of how much has actually been spent. This effectively presents a series of intra-year, discretionary spending shocks — excluding cyclical components, that is, social security and debt interest — which are unlikely to be anticipated, both given the UK's idiosyncratic budget process and statistical testing. The results point to a cumulative multiplier of 0.44 on impact and 0.47 in the long-run, as well some evidence of larger stimulative effects of civil spending relative to military spending at short horizons. This chapter's results also support theoretical and empirical findings of falls in household consumption in response to increases in government consumption, as well as higher multipliers in times of high slack — as measured unemployment considerably above the natural rate — but not for different regimes such as the debt-to-GDP ratio, openness to trade and exchange rate regimes.

Chapter 4 augments the historical analysis of chapter 3 by adding further archival research to include tax changes since 1879-80, and combines these two strands of government policy to create a combined series of changes to the discretionary fiscal stance over the course of 140 years. This allows for the estimation of a historical impact of a 1% of GDP increase in the fiscal balance on output of -0.24% in-year

and -0.38% by year 4. This chapter also provides evidence of a stronger effect of fiscal policy on output in times of high slack and of a stronger increase in household consumption as a response to a fiscal tightening in times of fiscal distress, as well as weak evidence to support asymmetric effects between expansionary and contractionary fiscal policy.

Acknowledgements

First and foremost, I would express my deepest gratitude to my supervisor, Pedro Gomes, for his unwavering support, mentorship and encouragement, and for providing me with much needed direction and challenge at the right points in this journey. Pedro's understanding of the pressures of my work alongside the part-time endeavour that was this PhD was critical in getting me over the line. I am forever grateful for all the discussions we have had and that I hope we will continue to have in years to come, and look forward to working together even after concluding this thesis.

I am also grateful to Ron Smith, my second supervisor, and Yunus Aksoy for helping me navigate the econometric and macroeconomic literature and providing advice that made this thesis much more focused and robust. I also extend my thanks to other academics and fellow students of the Economics Department at Birkbeck — I really valued our interactions, and this journey was all the richer for it. I feel that Birkbeck's flexible study ethos really supported me in being able to undertake this research in a way that made it possible to juggle work and study. I am grateful to the College for its existence and its tireless work to widen access and opportunity, and I will forever be a champion of its aims and successes in enriching the academic community and the UK.

I first started compiling historical data from Hansard and the Parliamentary Archives in February 2020, as part of my role then at the Office for Budget Responsibility, initially with the intention of putting the 2020 Spring Budget's fiscal package in historical context. Little did I know that this passion project would turn into a fully fledged PhD pursuit, and it would not have done so without the encouragement and support of many colleagues there. Special mentions must go to Dave Nolan, my manager then who saw my passion for the subject and supported my decision and application process; Stephen Farrington and Jim Ebdon, both of whom always showed interest in my progress and in providing me with the right amount of space to conduct my research alongside day-to-day responsibilities; and Charlie Bean, whose experience and feedback was invaluable in convincing me of the value of this work and of this pursuit. This research would also not have been possible without the financial support from the OBR from 2020 to 2023, for which I am very grateful. This thesis, however, is my own work and does not necessarily reflect the OBR's institutional views nor that of any of the members of its Budget Responsibility Committee.

I would also like to thank colleagues at my current employer, the Fraser of Allander Institute at the University of Strathclyde, and in particular Mairi Spowage, for allowing me the opportunity to continue to work on and finalise this thesis alongside my day job. I could not have been better supported through a big career and personal move to this wonderful city of Glasgow.

The work in this thesis has benefitted immensely from feedback from numerous colleagues, including but not limited to: Jagjit Chadha, Stephen Millard, Ben Caswell, Paula Bejarano Carbo and other colleagues at the National Institute of Economic and Social Research; participants at the 5th Workshop on Macroeconomic Research at the Cracow University of Economics; participants at the 50th Anniversary of the Birkbeck Economics Department's PhD Alumni Conference; members of the Bank of England/OBR/HM Treasury Fiscal and Macroeconomic Policy Interactions Working Group; participants at the 2022 Birkbeck School of Business, Economics and Informatics' PhD Student Conference; and colleagues at the University of Strathclyde's Department of Economics and Fraser of Allander Institute, with special thanks to Hannah Randolph for reviewing part of the thesis near its completion.

Thank you to my family, especially my parents and grandparents, who have always encouraged me all the way from Portugal and believed I could get to this place.

And the biggest thank you of all goes to my partner Anna and to Oscar, our beloved floof ball. Anna, I could not have got here without you backing me every step of the way. When in the summer of 2020 I was debating whether to even apply, and whether it would all be too much to juggle, I remember you saying "well, you'll enjoy it anyway, won't you?" And you were — as always — quite right, and much more aware than even I was of how much I would get out of all this work. Not only that, you always helped me believe that I could do it and that the hard work I put in throughout was worth it, even at the times when I was flagging. And Oscar, you were the purrfect ever-present company in those many hours of archival research, writing code and running regressions — thanks for being by my side, buddy!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the main defining characteristics of the pre-Great Financial Crisis consensus was the primacy of monetary policy as a macroeconomic stabilisation, embodied in influential papers such as Romer and Romer (1994) and Auerbach (2002), which concluded that there was a lack of evidence for the effects of fiscal policy on output. It was fully part of the so-called 'Jackson Hole consensus', epitomised by Feldstein's (2002) suggestion that it "confirms views that are now well-established and widely held in the profession." Even the sign of the effect of fiscal stimulus on output was contested at the time (Giavazzi et al., 2000).

At the same time, a different set of economists was starting a strand of literature that would become one of the standard ways of estimating the effects of fiscal policy on output. Ramey and Shapiro's (1998) work on pinpointing the timing of exogenous shocks used a narrative approach, which took information qualitative sources such as government documents or press articles to construct a timeline that helps establish causality (Romer and Romer, 2023). This approach is much more labour intensive than data-driven approaches, but has had success in identifying robust estimates of output effects of fiscal policy, and has now been used for monetary policy estimation as well.[1]

The narrative literature has consistently provided estimates of positive impacts on output from stimulative fiscal policies, be they through higher spending (Ramey and Shapiro, 1998; Ramey, 2011; Ramey and Zubairy, 2018) or through lower taxes (Romer and Romer, 2010). And there have also been some important methodological advances in the estimation of fiscal multipliers. Mountford and Uhlig's (2009) contribution to define the multiplier as the area under the curve of the cumulative impulse response function has been broadly accepted since then, and the adoption of the flexible local projections-instrumental variables (LP-IV) framework (Jordà, 2005; Stock and Watson, 2018) has also facilitated a renewed bout of literature on non-linear effects (Ramey and Zubairy, 2018) and multiplier decomposition (Cloyne et al., 2020).

But there are some areas where progress has been less unequivocal. One is in the calculation of the multiplier, and specifically how to convert econometric estimates into multiplier estimates. Earlier practice — for example, in Blanchard and Perotti (2002) — was to estimate an output elasticity and then multiply it by a

[1] It is a measure of how much the debate has changed that Romer and Romer, who in their 1994 paper concluded that monetary policy should have primacy over fiscal policy, titled their 2023 paper "Does Monetary Policy Matter?".

sample-based conversion ratio of the inverse of the government spending as a share of GDP (Y/G)[2]. Gordon and Krenn (2010) proposed a different way of calculating it, specifically tailored to the Great Depression in the US, where output took a long time to recover, and which consisted of dividing both output (on the left-hand side of the econometric equation) and government spending (on the right-hand side) by a measure of potential output. This meant both would be measured in the same units, and therefore would allow the direct retrieval of multiplier estimates from the regression outputs.

Ramey and Zubairy (2018) take this approach and apply it to a much longer sample, which spans from 1889 to 2015, arguing that using the Y/G conversion ratio biases multipliers upwards when using historical data due to changes in the average of the ratio over time. But using potential output as the denominator opens up the question of what is an appropriate way of estimating it, and whether different choices of method can influence the results. That is the subject of chapter 2, which takes a set of one- and two-sided methods of estimating potential output and shows that they generate multiplier estimates as low as -0.04 and as high as 0.70 for the same post-1946 period. This is huge parameter uncertainty, especially when it comes from such a seemingly obscure modelling decision — and much wider than the variation induced by the variation in Y/G. This leads me to recommend returning to using the conversion ratio as a more transparent alternative, and one which leaves less judgement in the hands of the econometrician — and it is this approach that I use in subsequent chapters.

The other area where there has been a distinct lack of progress has been in applying the narrative approach to non-US contexts. Narrative approaches are an important source of unanticipated shocks for use as instruments in estimating multipliers, but they are costly to assemble as the require a lot of archival research. This has meant their use has mostly been limited to US data. Cloyne (2013) used archival data to identify tax shocks in the UK, but there is no analogue on the spending side. This has contributed to a situation where official institutions such as the Office for Budget Responsibility have had to rely on US-based studies to make assumptions about the effects of government spending on output.

Chapter 3 is an attempt to fill that gap on the spending side by using historical data from the UK Parliamentary Archives to compile a series with 140 years' worth of intra-year policy changes, given that the UK's budget-setting process is ideally suited to provide this kind of data. For nearly 150 years, near or just after the beginning of each financial year, the Chancellor of the Exchequer has been required to present their estimate of how much the Exchequer spent in the previous financial year and how much they forecast spending to be in the coming year. In the UK's parliamentary system, the ability to pass money bills is a pre-condition of the government standing, which means that it is not negotiated in public, making the comparison between the beginning of the year's position and that at the end of the year an estimate of intra-year shock to spending policy.

This is a rich dataset, which contains large shocks from increased military expenditure, as might be expected (the Second Boer War, First and Second World Wars and the Iraq War), as well as unanticipated falls in military spending when those wars ended at an unanticipated time. But it also encompasses other important

[2]Note that $\partial \ln Y/\partial \ln G = \partial Y/\partial G \times G/Y$, and so $\partial Y/\partial G = \partial lnY/\partial \ln G \times Y/G$, which is the multiplier effect of government spending (G) on output (Y)

spending shocks, including the post-First World War adjustment, the cost overruns in the NHS immediately after its launch in 1948, the sterling crisis in 1967, the loss of control over inflation and public spending in the 1970s, the impact of the coal miners' strike on coal imports, and the introduction of the Conservative-Liberal Democrat coalition government's austerity programme midway through the 2010-11 financial year.

I use this dataset to estimate multipliers across the period from 1879-80 to 2018-19, finding cumulative multipliers around 0.4 to 0.5 over a 5-year horizon, with some evidence of larger effects of civil spending shocks on output than military spending ones at short horizons. I also use this dataset to test what the effect of government spending has been on other macroeconomic variables, finding that consumption generally falls in response to an increase in government expenditure, while employment rises. The effect on the policy rate is not significant, which together with a positive effect on inflation points to the Bank of England accommodating government spending increases in this period.

I also test whether there is evidence of differences in multipliers in different states of slack, and whether regime differences such as a high and low debt stock, a more or less open economy or different exchange rate regimes are associated with a different effectiveness of fiscal policy. I find some, though relatively weak, evidence that multipliers are higher in states of high slack, but no evidence of state-dependent effects otherwise.

And finally in chapter 4, I take a broader view of fiscal policy, considering the effect of shocks to the discretionary fiscal stance on output. This is motivated by an inconclusive literature on the effect of fiscal contractions on output, especially in the face of actual or potential distress in the public finances. This is a literature that had a resurgence around the time the UK embarked on a fiscal consolidation project in 2010, at the same time that the Euro Area debt crisis was unfolding.

Chapter 4 builds on chapter 3's data on government spending shocks and uses further Parliamentary Archives data to augment it with changes to tax policy, allowing me to create a series of discretionary fiscal stance shocks over the same period of 140 years — again exploiting the UK's long-established budget process, but this time to estimate the output cost of a 1% of GDP increase in the fiscal balance. The shocks capture a lot of the same events as in chapter 3, but also broader consolidations and loosenings such as the People's Budget in 1909; the aggressive fiscal tightening during the Great Depression; a cycle of loosening and subsequent tightening in the 1960s and 1970s; and the controversial 1981 and 2010 budgets, which provoked much discussion about whether a restrictive policy was desirable and what effect they ultimately had.

The results I obtain provide a relatively nuanced view of the effects of fiscal consolidations, indicating that output does fall in response to a fiscal tightening, but that consumption rises — although not by enough to offset the fall in other GDP components. This corroborates both the canonical Keynesian view of the effect of fiscal policy on output and some of the non-Keynesian mechanisms from the expansionary fiscal contraction literature regarding effects on the private sector. I also find stronger evidence that fiscal policy as a whole is more effective in states of high slack in terms of stimulating output than when considering spending only.

Using this dataset, I develop a set of measures of fiscal distress to capture episodes of fast debt accumulation, spikes in interest rates and rapid increases in the interest burden on national income, as well as a composite measure of all three. The results corroborate those of Giavazzi and Pagano (1990), showing that although there are no statistically significant differences in output effects, private-sector consumption expands more quickly in response to a budget consolidation in times of fiscal distress than in normal times. I also find some weak evidence that fiscal contractions might have a smaller effect on output than fiscal expansions.

Chapter 2

One step back, two steps forward: reassessing the appropriateness of conversion ratios in multiplier estimates

Abstract

I show that the choice of method of estimating potential output for use in the Gordon-Krenn transformation — which divides output (Y) and government spending (G) by potential output — can lead to wider variation in multiplier results than using the older conversion ratio method of Y/G. Using US data, I estimate that post-1946 in-year multipliers using different filtering methods vary between -0.04 and 0.70, whereas using the variation in Y/G between the 5th and 95th percentile gives a range between 0.32 and 0.53, which is a quarter as wide as the Gordon-Krenn interval. This narrower range is obtained using a method that is more transparent and leaves less discretion for an econometrician's seemingly small decision to influence results. I then extend this framework to the 18-country panel in the Jordà-Schularick-Taylor macrohistory dataset, finding that the wide dispersion of results using the Gordon-Krenn method is replicated, while using the conversion ratio method shows the expected statistically significantly higher multipliers pre-1946 than after that date.

2.1 Introduction and motivation

Government spending multipliers are an important concept for policymakers intending to use their fiscal stance for macroeconomic management, and their importance has only increased in the last fifteen years given monetary policy constraints and the needs to respond to shocks such as the Great Financial Crisis and the Covid-19 pandemic. The resurgence in interest in fiscal policy has led to an increase in work on multipliers, and to substantial development in the empirical strategies employed to estimate their size (Ramey, 2019). But how certain can we be of these estimates? I show that some of the developments in the last few years have actually widened the uncertainty about estimates.

As Ramey and Zubairy (2018) discuss, there are four main considerations when estimating multipliers. The first is how to identify exogenous shocks, as government spending and output are simultaneously determined. In the absence of an exogenous shock that could be used as an instrument, an ordinary least squares estimator would suffer from attenuation bias, that is, the size of the effect would be underestimated. To solve this issue, there are two main competing approaches. One, pioneered by Blanchard and Perotti (2002), is attempts to derive orthogonalised shocks from the data — for example, by applying restrictions on which variables can respond in-period. The other approach, pioneered by Ramey and Shapiro (1998), is also called the 'narrative' approach, focuses on identifying shocks from separate, uncorrelated sources. A classic example of the narrative approach is the outbreak of wars, which are generally thought to be as close to an exogenous event as one can find in macroeconomics. Ramey and Zubairy (2018) also explore an over-identified specification in which both these approaches are combined, which I use as the benchmark in my study.

The second consideration is which econometric approach to follow. Older approaches such as Blanchard and Perotti (2002) and Mountford and Uhlig (2009) use structural vector autoregressions (SVARs), while some more recent papers such as Ramey and Zubairy (2018) use local projections with instrumental variables (LP-IV) — based largely on Jordà (2005) and illustrated in the synthesis of Stock and Watson (2018). Estimation of impulse response functions using local projections has become more commonplace in recent years, as this specification allows for straightforward implementation of non-linearities, although structural VAR-type approaches remain popular (Afonso and Leal, 2019; Ferrara et al, 2021). Despite much debate regarding the relative merits of VARs and local projections, Plagborg-Møller and Wolf (2021) show that they estimate similar impulse responses to VARs, though with differing finite-sample properties. Given that, I choose to use local projections, a flexible framework which allows me to easily test for non-linear, state-contingent multiplier effects.

The third issue is how to define the multiplier, and probably the one where most consensus exists. Earlier influential analyses such as Blanchard and Perotti (2002) calculated the multiplier as the difference between the peak and trough of the impact of a fiscal policy change on output, which ignores the path of output. Instead, Mountford and Uhlig (2009) propose a definition of the multiplier as the area under the curve of the cumulative impulse response function, which has since been broadly accepted (Ramey, 2019) and which I use in the course of this thesis.

The final consideration is how to convert econometric outputs into a multiplier estimate, and it is the main focus of the chapter. Prior to Gordon and Krenn (2010), regression estimates were generally conducted using an isoelastic specification, that is, with both output and government spending in logs. In the isoelastic case, $\partial \ln Y/\partial \ln G = \partial Y/\partial G \times G/Y$, and so $\partial Y/\partial G = \partial lnY/\partial \ln G \times Y/G$, which is the multiplier effect of government spending (G) on output (Y) — and generally proxied by the sample average of Y/G. Gordon and Krenn (2010) propose an alternative specification in which they divide both government spending and output by potential output, which converts them into the same units and means one can obtain estimates directly from the regression results.

Gordon and Krenn devised this method for the Great Depression in the US, with the justification that potential output was below actual output during the

Great Depression for a prolonged period. Ramey and Zubairy (2018) then apply the same method to a much longer time series (1889 to 2015), arguing that the variation of the conversion ratio (Y/G) over time biases multipliers upwards when using historical data. The reason they cite for there being a bias is that it is possible to get different multipliers from the same elasticity due to differences in the average of Y/G, and making them larger than if they were estimated using the Gordon-Krenn method. While this might well be the case in their sample, it is unclear that this bias is positive at all times.

In this paper I illustrate how using different methods of estimating potential output for the Gordon-Krenn transformation can actually lead to a wider range of multiplier estimates than the variation inherent in using the Y/G conversion ratio. I first illustrate this by replicating Ramey and Zubairy's (2019) approach for the United States, using data from the Jordà-Schularick-Taylor (JST) macrohistory dataset (Jordà, Schularick, Taylor, 2017; Jordà et al., 2019), using an LP-IV specification and ten commonly used ways of estimating potential output. I show that although they all generate similar paths of potential output, in-year multiplier estimates for the post-World War II period are as low as -0.04 and as high as 0.70. This is huge uncertainty, and much wider than the variation generated by using the conversion ratio method of Y/G in that time period, as figure 2.1 shows.

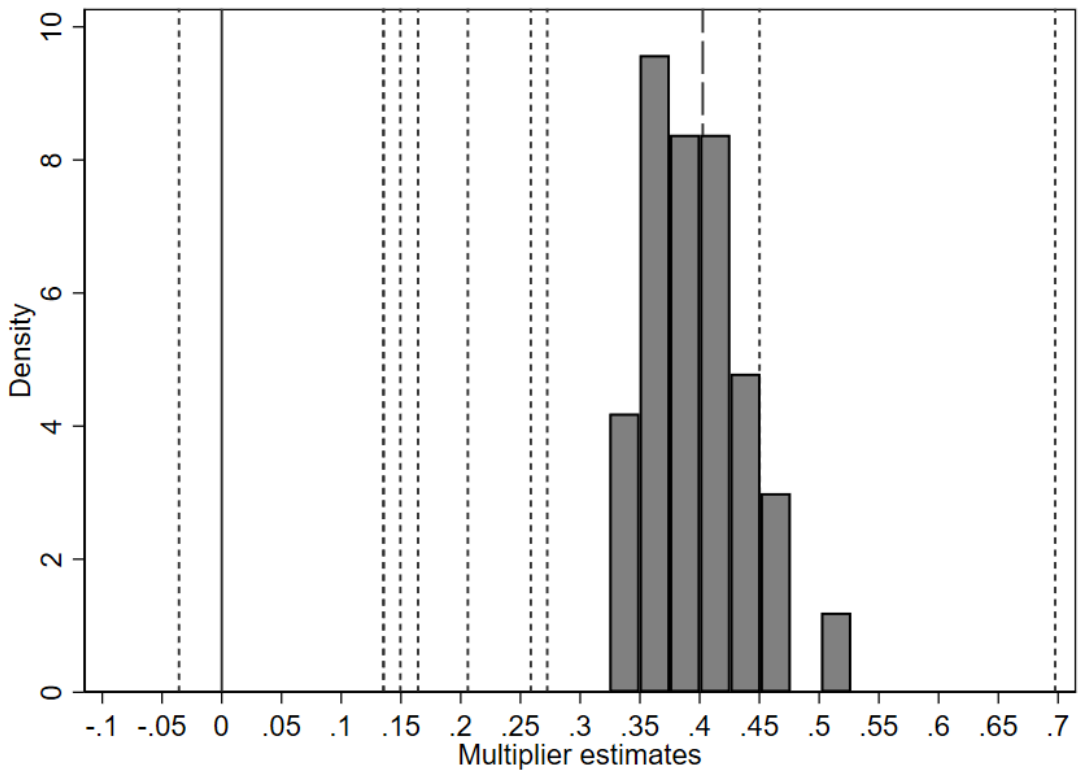


Figure 2.1: Histogram of point estimates of US in-year (horizon $h = 0$) multipliers using observed Y/G post-1946 between the 5th and 95th percentiles. The long-dashed vertical line uses the average Y/G post-1946 (5.3832). Short-dashed vertical lines are the point estimates of multipliers estimated using the Gordon-Krenn transformation with different methods of estimating potential output.

So instead of relying on such a small and untransparent modelling choice which can lead to very disparate results, I propose that we take a step back and reconsider the sample average of Y/G, especially in the context of the variability introduced by the Gordon-Krenn transformation. In section 2.2, I describe the methodology I employ in more detail and describe the dataset I use, and in section 2.3, I compare results obtained for the United States using the different methods for estimating potential output[1] to those obtained using the sample average of Y/G to make them more directly comparable with most studies. Figure 2.2 shows just how much Y/G has changed since 1871 as the size and scope of government have expanded. I therefore use the flexibility of the LP-IV framework to estimate separate multipliers pre- and post-World War II, treating time periods as states in the same vein as Ramey and Zubairy's (2018) use of a state-contingent formulation for monetary policy or slack states, and use Welch's (1947) t-test for populations with different variances to show that there is statistically significant evidence that US multipliers have changed over time.

I then apply the same methodology to the wider, 18-country panel, adding fixed effects to the LP-IV framework in section 2.4, showing that results are similar in flavour, although larger in magnitude than for the US. I then conclude in section 2.5, recommending the use a conversion ratio instead of the Gordon-Krenn transformation, as it produces a narrower range of estimates, with greater transparency and reduced degrees of freedom of the econometrician to influence results with their decisions.

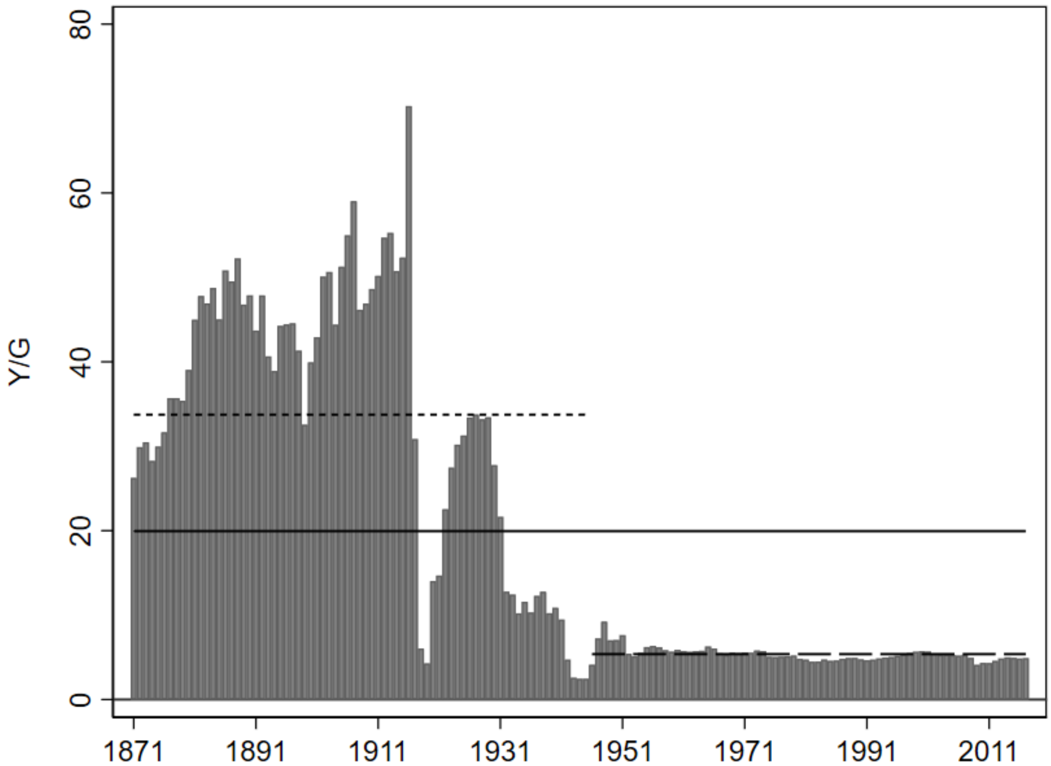


Figure 2.2: Average of Y/G for the US across the full sample (1871-2017). The solid line represents the average across the whole sample (19.9471). The short-dashed line represents the sample average before 1946 (33.7445). The long-dashed line represents the sample average from 1946 onwards (5.3832).

[1]Details of each of the methods are available in the appendix.

2.2 Methodology and data

2.2.1 Methodology for estimating a single multiplier across the sample

I will first focus on the US, and in section 2.4 I expand the specification to the 18-country panel from the JST dataset. I set up a model based on Ramey and Zubairy (2018). This means using the Mountford and Uhlig (2009) approach to calculating the multiplier, that is, calculating the area under the curve of the cumulative impulse response function. The estimation procedure itself employs the Jordà (2005) local projection method with instrumental variables (LP-IV), which means that, using the Gordon-Krenn transformation, the specification is as follows for each horizon h:

$$\sum_{j=0}^{h} \frac{y_{t+j}}{y^p_{t+j}} = \phi_h^{gk} L z_t + \beta_h^{gk} \sum_{j=0}^{h} \frac{g_{t+j}}{y^p_{t+j}} + \varepsilon^{gk}_{t+h} \tag{2.1}$$

where y_{t+j} is real output per capita at time $t+j$; y^p_{t+j} is a measure of potential output per capita at time $t+j$; g_{t+j} is real government consumption per capita at time $t+j$; Lz is a matrix of lagged controls, which in this case are the inflation rate and yield on short-term government bonds, as well as output and government spending per capita; and ε is an error term. If we define γ_h as the h-steps ahead multiplier, then for the Gordon-Krenn methodology, $\gamma_h^{gk} = \beta_h^{gk}$, that is, the multiplier estimate is obtained directly from the equation without requiring further transformation.

For the conversion ratio methodology, I use an isoelastic specification, that is, with both output per capita and government consumption per capita in logs — the dominant approach prior to Gordon and Krenn (2010) and Ramey and Zubairy (2018). The conversion ratio specification is as follows for each horizon h:

$$\sum_{j=0}^{h} \ln y_{t+j} = \phi_h^{cr} L z_t + \beta_h^{cr} \sum_{j=0}^{h} \ln g_{t+j} + \varepsilon^{cr}_{t+h} \tag{2.2}$$

where variables are as above, except they are in logs rather than transformed by potential output.

Unlike the Gordon-Krenn transformation, β_h^{cr} is an elasticity of output with respect to government spending and not a multiplier effect; to estimate γ_h^{cr}, we must use a conversion ratio, so that the estimate for the multiplier effect becomes:

$$\hat{\gamma}_h^{cr} = \hat{\beta}_h^{cr} \times \frac{\bar{y}}{\bar{g}} = \hat{\beta}_h^{cr} \times \frac{\sum_{t=0}^{T} y_t}{\sum_{t=0}^{T} g_t} \tag{2.3}$$

Finally, to estimate the β coefficients consistently, I follow Ramey and Zubairy's (2018) specification, combining both war dates and the Blanchard and Perotti shock as instruments for g_{t+j}, and estimate a regression equation at each horizon h using two-stage least squares. Ramey and Zubairy discuss the trade-offs involving in deciding for just one or both instruments; given that Blanchard-Perotti shocks are stronger at short horizons and war dates stronger at longer horizons, they opt for an overidentified specification, which I also use.

2.2.2 Methodology allowing for varying multipliers over time

Ramey and Zubairy's (2018) claim is that using the sample average of Y/G as a conversion ratio over a long time series biases multipliers upwards, citing the fact that it is possible to get different multipliers from the same elasticity just by virtue of a different Y/G. It is indisputable that given the formulation of equation (2.3), the multiplier estimate is a combination of the elasticity estimate and the conversion ratio — and that a higher Y/G (meaning lower G/Y, so a lower share of output that is government consumption) is associated with a higher multiplier, all else equal.

It is not particularly surprising to find an economic transaction for which its effect on output is lower when its base level is higher — in fact, the opposite would be more surprising. It is also consistent with evidence of crowding-out from government consumption, as found by Argimon et al. (1997).

But it does not immediately follow that estimates using a conversion ratio are biased upwards. It is plausible that multipliers calculated over a period which includes a period over which there were different levels of government's share of the economy reflect an average of its stimulative effect on output. So instead of assuming a constant elasticity or multiplier across time, I propose extending the methodology above to allow for variation in coefficients over time using a state-contingent framework. This is a straightforward extension, similar to Ramey and Zubairy's (2018) specification for differing economic conditions, only reinterpreting states as time periods.

In this framework, equation (2.1), using the Gordon-Krenn method, becomes equation (2.4) at each horizon h:

$$\begin{aligned}\sum_{j=0}^{h} \frac{y_{t+j}}{y^p_{t+j}} &= I_{t-1}\left[\phi^{gk}_{A,h} Lz_t + \beta^{gk}_{A,h} \sum_{j=0}^{h} \frac{g_{t+j}}{y^p_{t+j}}\right] \\ &+ (1 - I_{t-1})\left[\phi^{gk}_{B,h} Lz_t + \beta^{gk}_{B,h} \sum_{j=0}^{h} \frac{g_{t+j}}{y^p_{t+j}}\right] + \nu^{gk}_{t+h}\end{aligned} \tag{2.4}$$

where ν is an error term, and A (when $I = 1$) and B (when $I = 0$) are two different states, for which different coefficients can be estimated. Equation (2.4) is essentially a more general version of equation (2.1), which by assumption imposes that coefficients are equal across states A and B.

A similar transformation can be derived to transform equations (2.2) into equation (2.5) for the conversion ratio approach:

$$\begin{aligned}\sum_{j=0}^{h} \ln y_{t+j} &= I_{t-1}\left[\phi^{cr}_{A,h} Lz_t + \beta^{cr}_{A,h} \sum_{j=0}^{h} \ln g_{t+j}\right] \\ &+ (1 - I_{t-1})\left[\phi^{cr}_{B,h} Lz_t + \beta^{cr}_{B,h} \sum_{j=0}^{h} \ln g_{t+j}\right] + \nu^{cr}_{t+h}\end{aligned} \tag{2.5}$$

Elasticity estimates from equation (2.5) can then be converted into multiplier estimates using a similar method as in the single estimate case:

$$\hat{\gamma}^{cr}_{A,h} = \hat{\beta}^{cr}_{A,h} \times \frac{\bar{y}_A}{\bar{g}_A} = \hat{\beta}^{cr}_{A,h} \times \frac{\sum_{t=0}^{T} 1_{\{I_{t-1}=1\}} y_t}{\sum_{t=0}^{T} 1_{\{I_{t-1}=1\}} g_t} \tag{2.6}$$

$$\hat{\gamma}^{cr}_{B,h} = \hat{\beta}^{cr}_{B,h} \times \frac{\bar{y}_B}{\bar{g}_B} = \hat{\beta}^{cr}_{B,h} \times \frac{\sum_{t=0}^{T} 1_{\{I_{t-1}=0\}} y_t}{\sum_{t=0}^{T} 1_{\{I_{t-1}=0\}} g_t} \tag{2.7}$$

Simple eyeballing of figure 2.2 makes it quite clear how far apart the pre- and post-1946 averages of Y/G are. The aftermath of World War II (1946) is the most obvious place to start when it comes to dividing the full sample given the substantial change it heralded in terms of the size and scope of government intervention in the economy, as well as the increased degree of macroeconomic management in the post-war period. Running a Wald test for a structural break in Y/G allows me to reject the null hypothesis of no structural break ($p < 0.001$), though the structural break is so self-evident because of the war as to render the running of the test unnecessary.

Not only does the specification in equations (2.4) and (2.5) make it straightforward to estimate period-specific multipliers, it also makes it straightforward to test whether the difference between those estimates is statistically significant. This is similar to what Ramey and Zubairy (2018) do for state-contingent multipliers. But the large differences in Y/G between the two sub-samples (1871-1945 and 1946-2017) can plausibly lead to different variances. Because of that, in the appendix I use Welch's (1947) unequal variances t-test for the difference in means test, which allows for the testing of the null hypothesis that two populations with differences have the same mean.[2] Rejection of that null hypothesis at a standard level of statistical significance is then evidence that multipliers are different in each of the sub-samples.

2.2.3 Data

For both US-specific estimates and the wider 18-country panel, I use data from the Jordà-Schularick-Taylor macrohistory dataset (Jordà, Schularick, Taylor, 2017; Jordà et al., 2019). This is a rich dataset, containing (among other series) annual data on GDP, consumer prices, population, government spending and short-term interest rates for 18 advanced economies from 1870 to 2017 (or from independence where applicable).[3] I then augment this dataset by including war dates, which include the Ramey and Shapiro (1998) dates and dates before and after the window captured in that paper. The war dates and the sources I use are listed in table 2.1, with details in the appendix.

[2] Welch's unequal variances t-test is similar in flavour to Student's t-test, but calculations are slightly more complicated because of the absence of the simplifying assumption of equal variances. The test statistic for this specification is $t = \left(\hat{\gamma}^{cr}_{A,h} - \hat{\gamma}^{cr}_{B,h}\right) / \left(\sqrt{s_A^2 + s_B^2}\right) \sim t(\xi)$, where $\xi = \frac{\left(s_A^2/N_A + s_B^2/N_B\right)^2}{s_A^4/\left(N_A^2 \xi_A\right) + s_B^4/\left(N_B^2 \xi_B\right)}$, with $\xi_i = N_i - 1, i = A, B$ and where s is the sample standard error.

[3] These are Australia, Belgium, Canada, Germany, Denmark, Finland, France, Italy, Japan, the Netherlands, Norway, Portugal, the Republic of Ireland, Spain, Sweden, Switzerland, the United Kingdom and the United States. Estimation is conducted using data from 1871 onwards, as my preferred specification contains 1 lag — which is equivalent to the 4 quarter lags used by Ramey and Zubairy (2018).

War	Date	Source
Spanish-American war	1898	Library of Congress
Second Boer war	1899	The Gazette
World War I	1914	The Gazette
	1916	Library of Congress
Spanish civil war	1936	Library of Congress
Second Sino-Japanese war	1937	Crowley (1963)
World War II	1939	Imperial War Museum
	1940	Gordon and Krenn (2010)
Korean war	1950	Ramey and Shapiro (1998)
Vietnam war	1965	Ramey and Shapiro (1998)
Soviet-Afghan war	1980	Ramey and Shapiro (1998)
War on terror	2001	Ramey (2011a)

Table 2.1: War dates used for identification and sources (see appendix for more detail).

2.3 Results for the United States

2.3.1 Estimating a single multiplier across the sample

As discussed in the previous section, the Gordon-Krenn transformation involves dividing both output per capita and government spending per capita by potential output. The question then becomes how to estimate potential output, since it is unobservable. There is no standard agreement as to which method is best, but there are a number of common options in the literature that I test in this paper. I go through these methods below, with more detailed information in the appendix.

Methods for estimating potential output can broadly be divided into two types: one- and two-sided methods. Two-sided methods use data from both before and after the observation of GDP for which potential GDP is being estimated, whereas one-sided methods only use information up to the observation for which it is being estimated. Each type of method has its own merits and drawbacks. Two-sided methods generally have better fit and are relatively easy to use in standard econometric software. The commonly used Hodrick-Prescott (HP) filter (1981, 1997) is the most famous example, and its proliferation has led to claims that it has "withstood the test of time" and that it would "remain one of the standard methods for detrending" (Ravn and Uhlig, 2002). Others have disavowed its use for creating spurious cycles and for essentially assuming the answer (Hamilton, 2017). The latter is a general criticism of two-sided methods, which by using future GDP observation attenuate any calculation of multipliers toward zero, because they assume that changes in output would have happened anyway. One-sided methods' fit is generally not as good, but on the other hand they do not suffer from the attenuation issue.

Table 2.2 summarises the methods I have used in testing the impact of output gap estimation methods on results. All methods considered here extract trend output from the data in one way or another, and comprise some of the most popular methods of estimating potential output for good reason. They are flexible and importantly, they work well with long datasets, for which fine-tuning output gap models with external inputs is difficult. For example, contemporaneous business cycle indicators such as purchasing orders or confidence surveys, which can be used to create multivariate measures, are only available in more recent times. In fact, even employment and unemployment is unavailable as far back as the dataset goes,

One-sided methods	Two-sided methods
One-sided Hodrick-Prescott filter	(Two-sided) Hodrick-Prescott filter
Moving average filter	Butterworth filter
Kálmán filter	Baxter-King filter
Hamilton $t + h$ steps ahead method	Christiano-Fitzgerald filter
Polynominal trend	Exponential smoothing filter

Table 2.2: Methods used for estimating potential output.

and this lack of additional variables precludes broader approaches such as principal component analysis from being in scope of this paper.

Figure 2.3 shows the estimates of potential output for the United States using the different methods mentioned in table 2.2. It is immediately clear that despite the differences in methods, all come up with similar estimates, with the polynomial trend being the most different - but still not majorly different apart from the Great Depression. So our prior should be that multiplier estimated using any of these methods as the denominator in the Gordon-Krenn transformation should be similar.

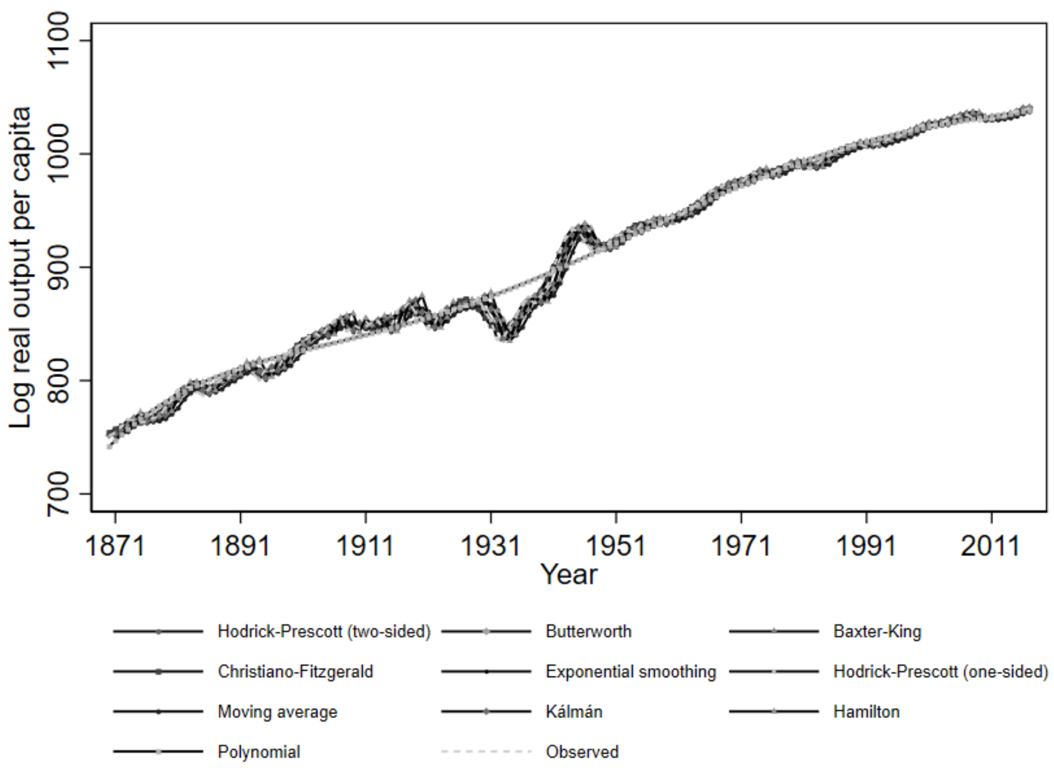


Figure 2.3: Estimated US potential log real output per capita using different methods.

That is not what I find. As figure 2.4 shows, there is a significant amount of variation between multiplier estimates depending on which estimate of potential output is used. The Hamilton method in-year estimate (0.76) is 1.5 to 4 times as large as that obtained by two-sided methods (all between 0.21 and 0.50); and between a quarter and three-fifths larger than the other one-sided methods (which are between 0.5 and 0.6), except for the one-sided HP filter, which is essentially zero (< 0.01). This is large parameter uncertainty, especially given how technical and inconsequential a point it may seem.

Table 2.3 summarises the results for the Gordon-Krenn methodology, obtained using equation (2.1) at each horizon h. It further highlights the fact that not only are the potential output estimates for all methods quite similar, but they also broadly yield statistically significant results for multiplier estimates. This makes it hard to distinguish between the different options for estimating potential output: they all look seemingly plausible, and produce statistically significant results, but those results differ significantly between one another, in some cases by large orders of magnitude. So can we find a better solution that has less parameter uncertainty?

As I discuss above, the wide range of multiplier estimates obtained using different yet equally defensible methods for estimating potential output is the main motivation for my reassessment of the merits of using the conversion ratio approach that was commonplace until 2010. I estimate equation (2.2) at each horizon and then convert them into multipliers using equation (2.3), with the results plotted below in figure 5 and detailed in table 2.4. Estimates are higher than using the Gordon-Krenn methodology, at 1.2 in-year, rising to 2.6 by year 6. While these are higher, they are still within the literature bounds (Ramey, 2019), and are all statistically significant. Their shape is also similar to that of Ramey and Zubairy (2018), implying persistent effects of government spending on output rather than tailing off over time.

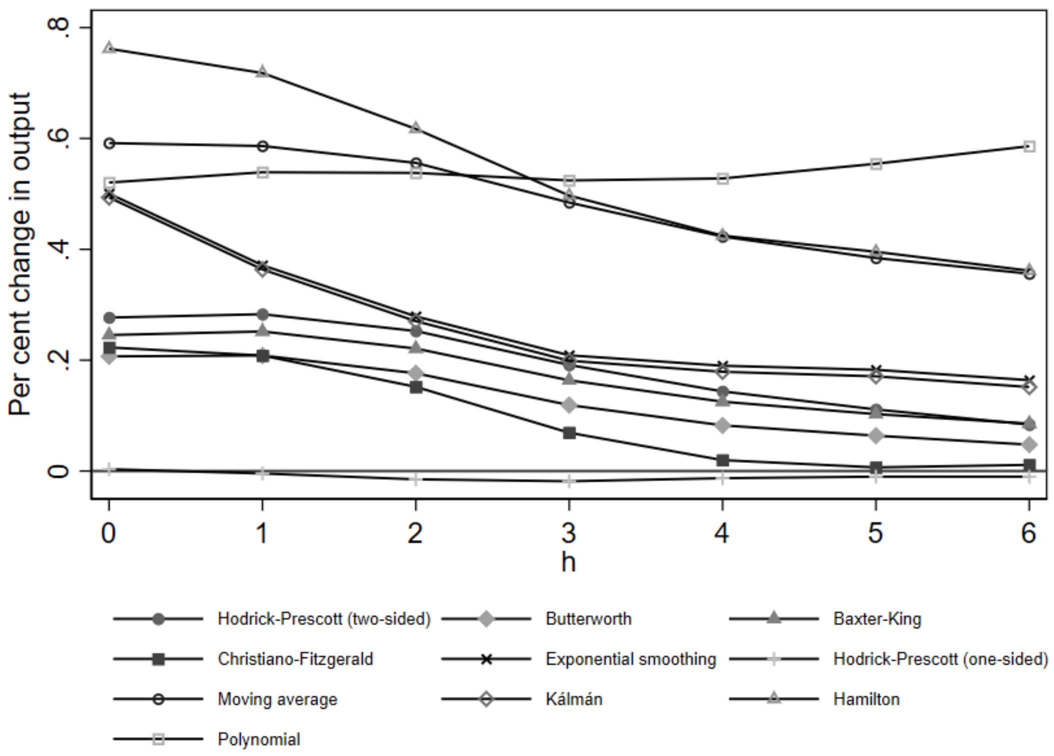


Figure 2.4: Cumulative impulse response functions (IRFs) of US output per capita in response to a 1% increase in government consumption per capita using the Gordon-Krenn transformation.

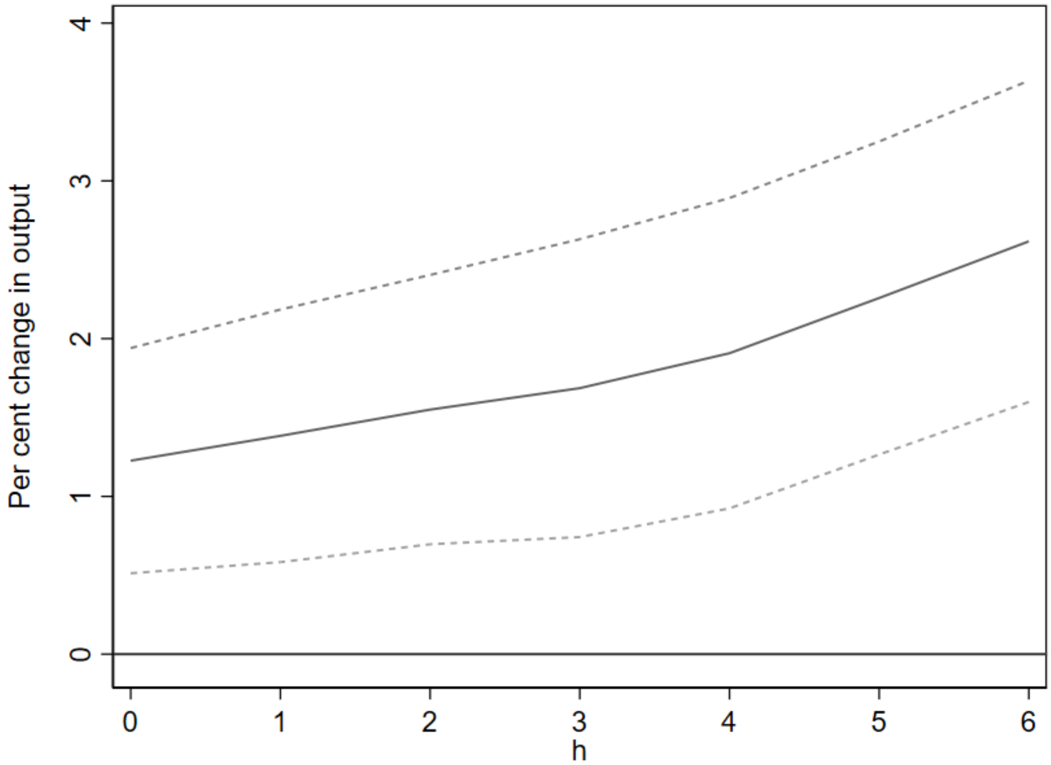


Figure 2.5: Cumulative impulse response functions (IRFs) of US output per capita in response to a 1% increase in government consumption per capita using the conversion ratio approach. Solid line represents point estimates. Dashed lines represent the 90% confidence interval for the estimates, calculated using HAC standard errors.

	Cumulative multiplier estimates						
Method	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Two-sided methods							
Hodrick-Prescott	0.2769***	0.2829***	0.2526***	0.1914***	0.1437***	0.1112***	0.0838***
	(0.0743)	(0.0435)	(0.0433)	(0.0370)	(0.0309)	(0.0223)	(0.0147)
Butterworth	0.2069***	0.2085***	0.1764***	0.1190***	0.0825***	0.0638***	0.0476***
	(0.0702)	(0.0366)	(0.0346)	(0.0253)	(0.0194)	(0.0127)	(0.0091)
Baxter-King	0.2454***	0.2518***	0.2210***	0.1637***	0.1251***	0.1027***	0.0856***
	(0.0708)	(0.0442)	(0.0456)	(0.0395)	(0.0341)	(0.0263)	(0.0199)
Christiano-Fitzgerald	0.2230***	0.2084***	0.1519***	0.0691**	0.0197	0.0066	0.0111
	(0.0733)	(0.0358)	(0.0297)	(0.0284)	(0.0182)	(0.0087)	(0.0127)
Exponential smoothing	0.5005***	0.3712***	0.2790***	0.2089***	0.1900***	0.1827***	0.1639**
	(0.1272)	(0.0812)	(0.0681)	(0.0780)	(0.0674)	(0.0645)	(0.0674)
One-sided methods							
Hodrick-Prescott	0.0034	-0.0043	-0.0149*	-0.0184*	-0.0129*	-0.0101*	-0.0101**
	(0.0189)	(0.0107)	(0.0090)	(0.0097)	(0.0075)	(0.0058)	(0.0047)
Moving average	0.5917***	0.5864***	0.5559***	0.4844***	0.4230***	0.3843***	0.3560***
	(0.1181)	(0.0922)	(0.0809)	(0.0895)	(0.0916)	(0.0934)	(0.0975)
Kálmán	0.4935***	0.3637***	0.2704***	0.1990**	0.1791***	0.1709***	0.1516**
	(0.1281)	(0.0813)	(0.0685)	(0.0782)	(0.0678)	(0.0656)	(0.0687)
Hamilton	0.7619***	0.7181***	0.6172***	0.4970***	0.4247***	0.3956***	0.3614***
	(0.1784)	(0.1255)	(0.1002)	(0.1033)	(0.1023)	(0.0925)	(0.0951)
Polynomial	0.5205***	0.5390***	0.5378***	0.5243***	0.5278***	0.5544***	0.5861***
	(0.0778)	(0.0654)	(0.0674)	(0.0816)	(0.0923)	(0.1031)	(0.1224)
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.3: Multiplier estimates for the United States using the Gordon-Krenn transformation with different methods for estimating potential output. Numbers in brackets are heteroscedasticity and autocorrelation consistent (HAC) standard errors.

Full sample (1871-2017)	In-year	Year 1	Year 2	Estimates Year 3	Year 4	Year 5	Year 6
Output elasticity	0.0615*** (0.0217)	0.0690*** (0.0243)	0.0769*** (0.0257)	0.0832*** (0.0283)	0.0937*** (0.0294)	0.1103*** (0.0295)	0.1271*** (0.0301)
Average Y/G	19.9471	20.0494	20.1535	20.2584	20.3644	20.4728	20.5847
Cumulative multiplier	1.2266*** (0.4338)	1.3840*** (0.4866)	1.5502*** (0.5188)	1.6861*** (0.5738)	1.9076*** (0.5979)	2.2576*** (0.6030)	2.6168*** (0.6189)
5th percentile Y/G conversion ratio	0.2667*** (0.0943)	0.2994*** (0.1053)	0.3336*** (0.1116)	0.3609*** (0.1228)	0.4062*** (0.1273)	0.4782*** (0.1277)	0.5513*** (0.1304)
95th percentile Y/G conversion ratio	3.1522*** (1.1149)	3.5387*** (1.2442)	3.9432*** (1.3196)	4.2666*** (1.4519)	4.8020*** (1.5050)	5.6528*** (1.5100)	6.5166*** (1.5413)
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.4: Multiplier estimates for the US using the conversion ratio approach using the full sample average of Y/G. Numbers in brackets are HAC standard errors.

2.3.2 Allowing for varying multipliers over time

I then use the state-contingent framework adapted from Ramey and Zubairy (2018), reinterpreting states as time periods, to estimate multipliers that vary across time. At each horizon h, this is obtained by estimating equation (2.4) if using the Gordon-Krenn method and equation (2.5) if using the conversion ratio method. The output elasticity with respect to government spending obtained from equation (2.5) is then converted into multiplier estimates using equations (2.6) and (2.7) for each state. As discussed in section 2.2, I use 1946 as the break point in the time series, reflecting the higher degree and scope of state intervention in the post-war period. The two periods I use are therefore 1871-1945 and 1946-2017.

Table 2.5 details the results from estimating different multipliers for the two periods. It is clear that point estimates for most methods results are higher before 1946 than after, which is exactly what we would expect, as discussed in section 2.2. It is consistent with diminishing marginal returns and crowding out of government spending, in a period where the scope and size of government is much larger than beforehand — average US central government spending is 18.9% of GDP post-1946, 3.3 times as large as the pre-1946 average of 5.7%.[4]

What also jumps out is that state-contingent estimation does not reduce the spread of estimates from Gordon-Krenn applications. Pre-1946 estimates of in-year multipliers range between 0.01 and 0.65, while post-1946 estimates range between -0.04 and 0.7. In fact, the range of estimates from different methods of estimating potential output is much wider than the variation in the conversion ratio Y/G. Figure 2.6 illustrates just that. It shows that converting the estimated post-1946 in-year output elasticity with respect to government spending into a multiplier with Y/G between the 5th and 95th percentiles generates estimates between 0.32 and 0.53 — an interval about a quarter as wide as that between the highest and lowest estimates using the Gordon-Krenn method.

What are we to make of this? It clearly illustrates that the results obtained using the Gordon-Krenn method — that is, transforming output and government

[4]See the appendix for testing of the differences across estimates for each method before and after the structural break.

Method	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
			Cumulative multiplier estimates				
			1871-1945				
Two-sided HP	0.2875***	0.2958***	0.2618***	0.2025***	0.1535***	0.1142***	0.0683***
	(0.0829)	(0.0615)	(0.0598)	(0.0484)	(0.0434)	(0.0335)	(0.0242)
Butterworth	0.2170***	0.2210***	0.1839***	0.1268***	0.0911***	0.0720***	0.0410**
	(0.0793)	(0.0543)	(0.0500)	(0.0344)	(0.0287)	(0.0207)	(0.0173)
Baxter-King	0.2525***	0.2632***	0.2304***	0.1748***	0.1327***	0.1058***	0.0714**
	(0.0821)	(0.0617)	(0.0610)	(0.0523)	(0.0492)	(0.0400)	(0.0320)
Christiano-Fitzgerald	0.2216***	0.2114***	0.1525***	0.0715**	0.0258	0.0176	0.0138
	(0.0849)	(0.0516)	(0.0415)	(0.0360)	(0.0260)	(0.0166)	(0.0252)
Exponential smoothing	0.4660***	0.1986**	0.1407	0.1924			
	(0.1631)	(0.0863)	(0.0973)	(0.1195)			
One-sided HP	0.0059	-0.0059	-0.0196	-0.0228*	-0.0149*	-0.0108	-0.0111
	(0.0263)	(0.0126)	(0.0119)	(0.0126)	(0.0091)	(0.0075)	(0.0071)
Moving average	0.5113***	0.5056***	0.4721***	0.4125***	0.3874***	0.4155***	0.4625**
	(0.1502)	(0.1300)	(0.1172)	(0.1290)	(0.1384)	(0.1549)	(0.1810)
Kálmán	0.4649***	0.3128***	0.1967**	0.1259	0.1385	0.1757*	0.1897
	(0.1638)	(0.1038)	(0.0867)	(0.1048)	(0.0976)	(0.1025)	(0.1198)
Hamilton	0.6505***	0.6000***	0.4840***	0.3679***	0.3266**	0.3673**	0.3861**
	(0.2028)	(0.1627)	(0.1290)	(0.1424)	(0.1581)	(0.1562)	(0.1789)
Polynomial	0.4658***	0.4679***	0.4335***	0.3888***	0.3634**	0.3828**	0.4278**
	(0.1072)	(0.0917)	(0.0983)	(0.1283)	(0.1479)	(0.1612)	(0.2011)
Conversion ratio	1.7239**	1.8704***	1.9768***	2.0790**	2.3789***	2.9051***	3.4330***
	(0.7907)	(0.7226)	(0.7396)	(0.8444)	(0.9148)	(0.9602)	(1.0312)
5th percentile Y/G	0.2185**	0.2371***	0.2505***	0.2635**	0.3015***	0.3682***	0.4351***
conversion ratio	(0.1002)	(0.0916)	(0.0937)	(0.1070)	(0.1159)	(0.1217)	(0.1307)
95th percentile Y/G	2.8086**	3.0472***	3.2206***	3.3871**	3.8757***	4.7329***	5.5929***
conversion ratio	(1.2882)	(1.1773)	(1.2050)	(1.3757)	(1.4904)	(1.5643)	(1.6799)
			1946-2017				
Two-sided HP	0.2060	0.2500***	0.2675***	0.2412**	0.1907**	0.1413*	0.1154**
	(0.1330)	(0.0886)	(0.0969)	(0.0943)	(0.0776)	(0.0753)	(0.0539)
Butterworth	0.1495	0.1770**	0.1764**	0.1473**	0.1026**	0.0621	0.0523*
	(0.1246)	(0.0723)	(0.0758)	(0.0673)	(0.0493)	(0.0489)	(0.0292)
Baxter-King	0.1356	0.1673**	0.1675**	0.1461*	0.1144	0.0807	0.1017*
	(0.1206)	(0.0796)	(0.0806)	(0.0792)	(0.0739)	(0.0788)	(0.0526)
Christiano-Fitzgerald	0.1643	0.1885*	0.1656	0.0812	-0.0102	-0.0562	-0.0260
	(0.1593)	(0.1060)	(0.1074)	(0.0708)	(0.0465)	(0.0365)	(0.0251)
Exponential smoothing	0.2724	0.2433*	0.1643	0.0823	-0.0179	-0.0871	-0.1085
	(0.2098)	(0.1288)	(0.1555)	(0.1177)	(0.0938)	(0.0917)	(0.0785)
One-sided HP	-0.0358	-0.0138	-0.0156	-0.0153*	-0.0179**	-0.0181*	-0.0104
	(0.0298)	(0.0138)	(0.0170)	(0.0085)	(0.0089)	(0.0105)	(0.0119)
Moving average	0.4501**	0.4750**	0.4490**	0.3200	0.1521	0.0069	-0.1082
	(0.2262)	(0.1979)	(0.2233)	(0.2096)	(0.1891)	(0.1881)	(0.1721)
Kálmán	0.2587	0.2268*	0.1443	0.0605	-0.0414	-0.1140	-0.1376*
	(0.2076)	(0.1281)	(0.1528)	(0.1165)	(0.0944)	(0.0914)	(0.0812)
Hamilton	0.6976**	0.6587**	0.6167**	0.4322*	0.2730	0.1549	0.1118
	(0.3399)	(0.2656)	(0.2494)	(0.2284)	(0.1778)	(0.1630)	(0.1456)
Polynomial	0.1352	0.1202	0.0514	-0.0572	-0.2020	-0.3967**	-0.5926***
	(0.1466)	(0.1412)	(0.1550)	(0.1631)	(0.1677)	(0.1765)	(0.1901)
Conversion ratio	0.4026*	0.6029***	0.8061***	0.9211***	0.9829***	1.0551***	1.0974**
	(0.2079)	(0.1972)	(0.2916)	(0.3346)	(0.3291)	(0.3798)	(0.4287)
5th percentile Y/G	0.3244*	0.4851***	0.6477***	0.7392***	0.7878***	0.8444***	0.8763**
conversion ratio	(0.1675)	(0.1586)	(0.2343)	(0.2685)	(0.2638)	(0.3040)	(0.3423)
95th percentile Y/G	0.5269*	0.7880***	1.0521***	1.2007***	1.2798***	1.3717***	1.4235**
conversion ratio	(0.2720)	(0.2577)	(0.3806)	(0.4362)	(0.4286)	(0.4938)	(0.5560)
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.5: Multiplier estimates for the United States for the 1871-1945 and 1946-2017 periods. Numbers in brackets are heteroscedasticity and autocorrelation consistent (HAC) standard errors.

spending by potential output — are very sensitive to choices about which method to use in estimating potential output. In fact, they are much more so than the variation induced by the variation in Y/G across time when using it as a conversion ratio, especially in the post-WW2 period.

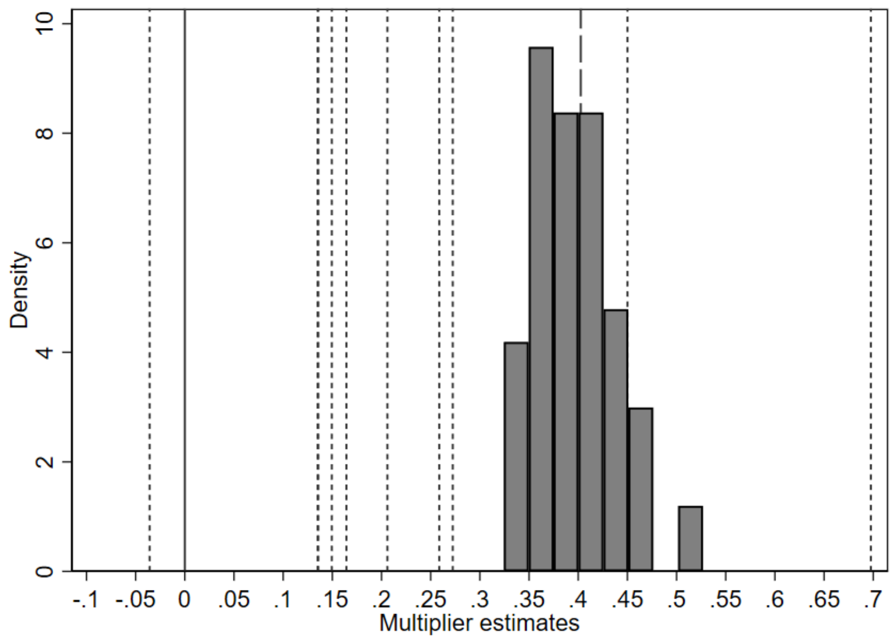


Figure 2.6: Histogram of point estimates of US in-year (horizon $h = 0$) multipliers using observed Y/G post-1946 between the 5th and 95th percentiles. The long-dashed line uses the average Y/G post-1946 (5.3832). Short-dashed vertical lines are the point estimates of multipliers estimated using the Gordon-Krenn transformation with different methods of estimating potential output.

Clearly when looking further back in time Y/G does have significant variation, which means that aggregating pre- and post-1946 observations will lead to higher multiplier estimates than just the post-1946 period, all else equal. But it is not surprising that is the case, nor does that mean that the answer obtained for the whole 1871-2017 period is wrong or unduly upwards biased. Instead, it reflects the very different macroeconomic conditions and policy paradigms in the two periods. In fact, it would be surprising if they were not different — an observation that is corroborated by my results in testing the difference in estimates between the two periods.

Finally, the conversion ratio method is also a more transparent way of calculating a multiplier. It is easy to report the output elasticity, Y/G and obtained multiplier, and it follows a well-established transformation using a log specification that is commonplace in both theoretical and applied macroeconomics. This means that the econometrician has fewer degrees of freedom with which to influence the results of their study — which can only be a good thing, especially given how hard it is to discern the difference in quality of fit between the different potential output estimation methods.

2.4 Extending the specification to a wider panel of countries

2.4.1 Estimating a single multiplier across time

The extension of equations (2.1)-(2.7) from a one-country setting to the 18-country panel from the JST dataset is trivial, and specifications are detailed in the appendix.

I choose to include country fixed effects, which is a standard approach for this kind of panel.

Table 2.6 summarises results for the 18-country panel, The panel results are similar to the US ones, with a broad dispersion of results for the different implementations of the Gordon-Krenn method. The in-year range (from -0.08 to 1.29) is actually wider in the panel setting than in the US-only case. The conversion ratio estimates for the US in the long run are considerably higher, which may reflect its greater freedom and use of accommodative monetary policy for large periods (Bianchi, 2012) — especially given the United States' centrality to the international monetary system for most of the sample.

	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Gordon-Krenn multiplier estimates							
Two-sided HP	0.2184**	0.1679***	0.1156***	0.0754**	0.0421	0.0124	0.0121
	(0.0919)	(0.0626)	(0.0413)	(0.0304)	(0.0259)	(0.0197)	(0.0149)
Butterworth	0.1682**	0.1072**	0.0620**	0.0370*	0.0198	0.0024	0.0086
	(0.0826)	(0.0495)	(0.0283)	(0.0201)	(0.0177)	(0.0133)	(0.0101)
Baxter-King	0.2058**	0.1469**	0.0900**	0.0565**	0.0314	0.0060	0.0060
	(0.0884)	(0.0574)	(0.0361)	(0.0271)	(0.0253)	(0.0207)	(0.0161)
Christiano-Fitzgerald	0.2173**	0.1519***	0.0955***	0.0560***	0.0303**	0.0130	0.0236**
	(0.0936)	(0.0582)	(0.0311)	(0.0197)	(0.0153)	(0.0119)	(0.0111)
Exponential smoothing	0.4692***	0.3355***	0.2219***	0.1393**	0.0989*	0.0767	0.0971**
	(0.1537)	(0.1120)	(0.0816)	(0.0651)	(0.0591)	(0.0494)	(0.0417)
One-sided HP	-0.0750**	-0.0320*	-0.0086	-0.0029	-0.0026	0.0014	0.0069
	(0.0326)	(0.0177)	(0.0122)	(0.0093)	(0.0068)	(0.0058)	(0.0053)
Moving average	0.6700***	0.6412***	0.5542***	0.4412***	0.3442**	0.2708**	0.2504**
	(0.2393)	(0.2022)	(0.1767)	(0.1531)	(0.1410)	(0.1296)	(0.1168)
Kálmán	0.4695***	0.3349***	0.2208***	0.1378**	0.0971	0.0746	0.0950**
	(0.1541)	(0.1121)	(0.0818)	(0.0654)	(0.0594)	(0.0497)	(0.0420)
Hamilton	1.2934***	0.9482***	0.7122***	0.5566***	0.4394***	0.3508***	0.3272***
	(0.2795)	(0.2168)	(0.1783)	(0.1505)	(0.1335)	(0.1204)	(0.1077)
Polynomial	0.4249***	0.4252***	0.3897***	0.3434**	0.2862*	0.2365	0.1989
	(0.1295)	(0.1338)	(0.1346)	(0.1383)	(0.1469)	(0.1538)	(0.1588)
Conversion ratio method estimates							
Output elasticity	0.1325***	0.1481***	0.1469***	0.1410***	0.1338***	0.1265**	0.1261**
	(0.0396)	(0.0460)	(0.0467)	(0.0482)	(0.0512)	(0.0535)	(0.0543)
Average Y/G	9.8162	9.8542	9.8934	9.9331	9.9741	10.0166	10.0602
Multiplier estimate	1.3006***	1.4592***	1.4529***	1.4404***	1.3345***	1.2667**	1.2690**
	(0.3891)	(0.4531)	(0.4623)	(0.4788)	(0.5106)	(0.5358)	(0.5464)
5th percentile Y/G	0.3410***	0.3811***	0.3780***	0.3638***	0.3459***	0.3285**	0.3288**
conversion ratio	(0.1020)	(0.1183)	(0.1203)	(0.1244)	(0.1323)	(0.1390)	(0.1416)
95th percentile Y/G	3.5371***	3.9724***	4.0200***	3.8708***	3.6746***	3.4872**	3.4925**
conversion ratio	(1.0583)	(1.2335)	(1.2792)	(1.3235)	(1.4059)	(1.4750)	(1.5038)
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.6: Multiplier estimates for the 18-country panel using the Gordon-Krenn transformation with different methods for estimating potential output and conversion ratio method. Numbers in brackets are heteroscedasticity and autocorrelation consistent (HAC) standard errors.

2.4.2 Allowing for varying multipliers over time

A further extension is to allow variation in multipliers over time in the panel setting, much in the same way as for the single country case presented in section 2.3 — details of the specifications are available in the appendix. Table 2.7 summarises the results from estimating different multipliers for 1871-1945 and 1946-2017 using both the Gordon-Krenn and conversion ratio methods.

Results for the 18-country panel are similar in flavour to those obtained just using US data. There continues to be broad dispersion of estimates for the different imple-

Method	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
			Cumulative multiplier estimates				
			1871-1945				
Two-sided HP	0.2674**	0.2145***	0.1574***	0.1129***	0.0722**	0.0289	0.0235
	(0.1084)	(0.0725)	(0.0470)	(0.0350)	(0.0296)	(0.0218)	(0.0168)
Butterworth	0.2088**	0.1416**	0.0894***	0.0609***	0.0386*	0.0106	0.0149
	(0.0969)	(0.0577)	(0.0324)	(0.0234)	(0.0206)	(0.0151)	(0.0116)
Baxter-King	0.2466***	0.1843***	0.1232***	0.0864***	0.0566**	0.0182	0.0132
	(0.1044)	(0.0668)	(0.0409)	(0.0308)	(0.0285)	(0.0228)	(0.0178)
Christiano-Fitzgerald	0.2274**	0.1510**	0.0889***	0.0511**	0.0289*	0.0094	0.0248**
	(0.1095)	(0.0665)	(0.0344)	(0.0220)	(0.0171)	(0.0132)	(0.0126)
Exponential smoothing	0.5329***	0.3890***	0.2506**	0.1474*	0.0863	0.0457	0.0714
	(0.1947)	(0.1414)	(0.1023)	(0.0812)	(0.0731)	(0.0593)	(0.0487)
One-sided HP	-0.0805**	-0.0365*	-0.0146	-0.0097	-0.0103	-0.0069	0.0018
	(0.0355)	(0.0186)	(0.0129)	(0.0099)	(0.0076)	(0.0067)	(0.0057)
Moving average	0.7296**	0.7113***	0.6318***	0.5073***	0.3872**	0.2829*	0.2463*
	(0.2972)	(0.2502)	(0.2211)	(0.1927)	(0.1774)	(0.1630)	(0.1464)
Kálmán	0.5339***	0.3893***	0.2506**	0.1470*	0.0854	0.0445	0.0701
	(0.1949)	(0.1414)	(0.1025)	(0.0814)	(0.0733)	(0.0594)	(0.0489)
Hamilton	1.2039***	0.9071***	0.6759***	0.5035***	0.3672**	0.2555	0.2285
	(0.3909)	(0.2838)	(0.2284)	(0.1932)	(0.1735)	(0.1588)	(0.1414)
Polynomial	0.3792**	0.3767**	0.3315**	0.2707*	0.1887	0.0931	0.0204
	(0.1542)	(0.1557)	(0.1566)	(0.1626)	(0.1752)	(0.1881)	(0.1984)
Conversion ratio	1.9526***	2.1372***	2.0446***	1.8726**	1.6490**	1.3918*	1.2776
	(0.6394)	(0.7352)	(0.7398)	(0.7568)	(0.8008)	(0.8383)	(0.8531)
5th percentile Y/G	0.4600***	0.5035***	0.4816***	0.4411**	0.3884**	0.3279*	0.3010
conversion ratio	(0.1506)	(0.1732)	(0.1743)	(0.1783)	(0.1886)	(0.1975)	(0.2010)
95th percentile Y/G	6.3578***	6.9591***	6.6574***	6.0976**	5.3693**	4.5319*	4.1600
conversion ratio	(2.0819)	(2.3941)	(2.4090)	(2.4642)	(2.6076)	(2.7298)	(2.7779)
			1946-2017				
Two-sided HP	0.0349	-0.0150	-0.0327	-0.0370	-0.0365	-0.0236	-0.0102
	(0.0560)	(0.0546)	(0.0454)	(0.0400)	(0.0306)	(0.0251)	(0.0204)
Butterworth	0.0298	-0.0196	-0.0331	-0.0319	-0.0272	-0.0133	-0.0011
	(0.0530)	(0.0482)	(0.0377)	(0.0316)	(0.0224)	(0.0183)	(0.0151)
Baxter-King	0.0479	-0.0111	-0.0435	-0.0445	-0.0449**	-0.0290	-0.0113
	(0.0484)	(0.0454)	(0.0368)	(0.0319)	(0.0227)	(0.0183)	(0.0156)
Christiano-Fitzgerald	0.2377***	0.2210***	0.1783***	0.1233***	0.0674***	0.0369***	0.0251***
	(0.0823)	(0.0783)	(0.0629)	(0.0452)	(0.0250)	(0.0134)	(0.0091)
Exponential smoothing	0.2500***	0.1651***	0.1411***	0.1222***	0.1198***	0.1243***	0.1166***
	(0.0831)	(0.0635)	(0.0504)	(0.0421)	(0.0418)	(0.0386)	(0.0355)
One-sided HP	-0.0417	-0.0050	0.0207	0.0244	0.0189	0.0184*	0.0167**
	(0.0281)	(0.0244)	(0.0198)	(0.0156)	(0.0124)	(0.0104)	(0.0080)
Moving average	0.5904***	0.5788***	0.5132***	0.4446***	0.3868***	0.3604***	0.3425***
	(0.1472)	(0.1464)	(0.1339)	(0.1180)	(0.1056)	(0.0987)	(0.0925)
Kálmán	0.2622***	0.1788***	0.1538***	0.1360***	0.1350***	0.1402***	0.1329***
	(0.0842)	(0.0651)	(0.0521)	(0.0440)	(0.0440)	(0.0410)	(0.0379)
Hamilton	1.6364***	1.2782***	1.0713***	0.9635**	0.8494**	0.7705**	0.7196**
	(0.3268)	(0.3673)	(0.3927)	(0.3909)	(0.3610)	(0.3390)	(0.3255)
Polynomial	0.5845***	0.5922***	0.5833***	0.5573***	0.5148***	0.5402***	0.5303***
	(0.0924)	(0.0985)	(0.1055)	(0.1093)	(0.1129)	(0.1196)	(0.1217)
Conversion ratio	0.4885***	0.5737***	0.6464***	0.6925***	0.7416***	0.8180***	0.8734***
	(0.1137)	(0.1269)	(0.1406)	(0.1509)	(0.1610)	(0.1748)	(0.1849)
5th percentile Y/G	0.2449***	0.2872***	0.3231***	0.3457***	0.3695***	0.4067***	0.4331***
conversion ratio	(0.0570)	(0.0635)	(0.0703)	(0.0753)	(0.0802)	(0.0869)	(0.0917)
95th percentile Y/G	0.9596***	1.1255***	1.2663***	1.3547***	1.4480***	1.5937***	1.6974***
conversion ratio	(0.2234)	(0.2489)	(0.2754)	(0.2951)	(0.3143)	(0.3405)	(0.3594)
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.7: Multiplier estimates for the 18-country panel for the 1871-1945 and 1946-2017 periods. Numbers in brackets are heteroscedasticity and autocorrelation consistent (HAC) standard errors.

mentations of the Gordon-Krenn method — in fact, even more so when comparing these results with those in table 2.5. Take the in-year multiplier in the post-1946 as an example. Gordon-Krenn implementations yield estimates as low as -0.04 (not statistically different from zero) and as high as 1.64 (statistically significant). Using the conversion ratio though, the dispersion is much narrower: the point estimate is 0.49 when using the sample average, and between 0.24 and 0.96 using the 5th and 95th percentiles. The range of estimates using the Y/G is less than half as wide as using different potential output estimation methods, which mirrors what I find for the US case. This highlights the robustness of the results to different geographies, while pointing to the relative sensitivity of the Gordon-Krenn methodology with respect to what seem to be minor modelling choices.

2.5 Conclusion

Government spending multipliers are crucial for understanding policy implications, but there are a number of considerations to take into account when estimating them to avoid common pitfalls: how to identify exogenous shocks, what econometric approach to follow, how to define the multiplier and how to convert econometric outputs into a multiplier estimate. In this chapter, I hold the first three constant and focus on the effect of different empirical decisions regarding the latter — specifically, how the Gordon-Krenn transformation of dividing output and government spending by potential output leads to a wide range of outcomes, wider still than the older approach of using Y/G as a conversion ratio, which Ramey and Zubairy (2018) criticise for biasing estimates upwards — a criticism that I do not find persuasive. It might result in higher estimates when taking into account data from further in the past, but that reflects a plausible assertion that there is a higher marginal output stimulation effect of government spending when the starting level of involvement of government in an economy is lower, which is what I find when estimating different multipliers before and after 1946.

To illustrate the impact of the choice of method to estimate potential output, I apply the methodology of Ramey and Zubairy (2018) first to US data from the Jordà-Schularick-Taylor (JST) macrohistory dataset (Jordà, Schularick, Taylor, 2017; Jordà et al., 2019). This means using local projections with instrumental variables to estimate the cumulative impulse response functions. I use the equivalent of Ramey and Zubairy's preferred specification of using two instruments: a Blanchard-Perotti-type shock and war dates. I test five two-sided and five one-sided methods for estimating potential output, and show that they all come up with reasonable and defensible estimates, with little discernible difference between them. But despite that similarity, I obtain very different results depending on which method I use. The in-year estimate using the Hamilton method (0.76) is 1.25 to 4 times as large as any of the other methods, apart from one which yields an essentially zero estimate. This is huge uncertainty for such a technical and seemingly inconsequential choice.

So instead I propose to estimate an output elasticity with respect to government spending and apply a conversion ratio of Y/G, as was common practice before Gordon and Krenn (2010). This produces a smaller dispersion of results, very much within the literature (Ramey, 2019), but with less discretion for an econometrician's small and untransparent decision to influence results. I then use Ramey and Zubairy's (2018) state-contingent specification to apply it to pre- and post-WW2 period. I show that converting the estimated post-1946 in-year elasticity with the Y/G conversion ratio between the 5th and 95th percentiles generates estimates between 0.32 and 0.53 — an interval about a quarter as wide as that between the highest and lowest estimates using the Gordon-Krenn method.

I then extend this framework to the wider, 18-country panel in the JST dataset, and find that results for this group of advanced economies are similar to those obtained just using US data. The broad dispersion of results using different methods when using the Gordon-Krenn method is actually wider, highlighting the sensitivity of results to these modelling choices. When using the conversion ratio method, however, I find that multipliers post-1946 between the 5th and 95th percentiles to be between 0.24 and 0.96 — less than half as wide a range as that obtained when using different Gordon-Krenn method implementations.

Appendices

2.A War dates

The list below details the sources for the war dates not obtained from Ramey and Shapiro (1998), Ramey (2011a) or Gordon and Krenn (2010) and referenced in table 2.1:

- 1898: sinking of USS Maine in February, which precipitates the Spanish-American war breaking out in April (Library of Congress, https://www.loc.gov/rr/hispanic/1898/intro.html, retrieved 8 August 2021);

- 1899: Boer ultimatum to the UK government, leading to the second Boer war breaking out on 11 October (The Gazette, https://www.thegazette.co.uk/all-notices/content/103822, retrieved 8 August 2021);

- 1914: Assassination of Archduke Franz Ferdinand in Sarajevo on 28 June, leading to successive declarations of war, resulting in the Allies and the Central Powers being at war by the end of August (The Gazette, https://www.thegazette.co.uk/all-notices/content/200, retrieved 8 August 2021);

- 1916: US President Woodrow Wilson progressively abandons effort to negotiate peace, and Germany attacks the US directly (Black Tom explosion on 30 July); there is expectation that the US will enter the war (Libary of Congress, https://blogs.loc.gov/maps/2017/04/wwi-era-terrorism/, retrieved 8 August 2021);

- 1936: A military uprising starts throughout Spain on 17 July, with a civil war breaking out (Library of Congress, https://www.loc.gov/item/today-in-history/july-17/, retrieved 8 August 2021);

- 1937: A minor battle between Chinese and Japanese troops near Beijing on 7 July devolves into a full-scale war by the end of the month (Crowley (1963)); and

- 1939: Germany invades Poland on 1 September; the UK and France declare war on 3 September (Imperial War Museum, https://www.iwm.org.uk/history/how-europe-went-to-war-in-1939, retrieved 8 August 2021).

2.B Methods of estimating potential output

This section describes in detail the methods used for estimating potential output, as well as the merits and drawbacks of each.

2.B.1 Two-sided methods

Two-sided methods, as their name suggests, use observations that lie both sides of each output observation y_t to split observed output into potential output and output gap. The most famous of these is the **Hodrick-Prescott, or HP, filter** (Hodrick and Prescott, 1981, 1997), which is a special case of the more general **Butterworth filter** (Goméz, 1999), which I also test. These belong to a general class of high-pass filters, which allow high-frequency above a certain threshold (λ) components to pass through it, while filtering out low-frequency below.

The HP filter was first popularised as a way of extracting the business cycle element of economic fluctuations and was the basis of the stylised business cycle facts upon which Kydland and Prescott (1982) and subsequent literature sought to build "real business cycle" (RBC) models. Both the HP and the Butterworth (in its Pollock (2000) form) filters solve:

$$\min_{\tau_{it}} \sum_{t=1}^{T} (y_{it} - \tau_{it})^2 + \lambda((\tau_{i,t+1} - \tau_{it}) - (\tau_{it} - \tau_{i,t-1}))^2 \tag{2.8}$$

where τ_{it} is a smoothed trend. The main difference is that for the Butterworth filter λ is a combination of two parameters:

$$\lambda = \left(\frac{1}{\tan \omega_d}\right)^{2n} \tag{2.9}$$

where ω_d, the cut-off point of the filter, and n, the order of the filter, whereas the HP filter uses λ on its own as a numerically-set parameter. Hodrick and Prescott proposed $\lambda = 1600$ for quarterly data, which has since been followed; Ravn and Uhlig (2002) derived $\lambda = 6.25$ for annual data on that basis, which I use. For the Butterworth filter, ω_d is derived from the NBER-defined length of a business cycle fluctuation, 6 to 32 quarters. From my testing, varying the order of the filter seems to have only a very small impact on estimates for multipliers.

Baxter and King (1995) and **Christiano and Fitzgerald** (2003) propose two similar alternatives to the HP filter. Both are approximations of the ideal band pass filter, meaning that they eliminate both slow-moving and very high-frequency components while retaining those in-between. The fluctuations within this spread then form the business cycle (Baxter and King, 1999). Both are weighted moving averages of observed output, solving for different optimal weights. Both minimise deviations from observed output and set the length of a business cycle to be between 6 and 32 quarters, but Baxter-King is a symmetric filter, with equal weights and equal lengths either side of contemporaneous output, whereas Christiano-Fitzgerald is a more general filter, allowing for both differening weights and asymmetric windows.

Finally, I also test an **exponential smoothing** filter. This was commonly used prior to the popularisation of the HP filter, and was used in famous economics papers such as Friedman (1957) on the permanent income hypothesis and Lucas (1980) on the quantity theory of money (Ladiray et al., 2003). It is a relatively simple method, which mathematically solves the following problem:

$$\min_{\{y^p_{it}\}^T_{t=1}} \sum_{t=1}^{T} \left[\left(y_{it} - y^p_{it}\right)^2 + \lambda \left(y^p_{it} - y^p_{i,t-1}\right)^2 \right] \tag{2.10}$$

Although it may not be immediately obvious from the optimisation problem, the first order condition depends on y^p_{t+1}, and so the exponential smoothing filter is effectively a two-sided procedure — and not that dissimilar from the HP filter, as shown by King and Rebelo (1993). King and Rebelo also show that the exponential smoothing filter is further away from the ideal band pass filter than the HP filter.

2.B.2 Merits and drawbacks of two-sided methods

Two-sided methods as a class are used very commonly, with the HP filter being the most popular of this class. Ravn and Uhlig (2002) posited that the HP filter had "withstood the test of time" and it was likely it would "remain one of the standard methods for detrending". But its proliferation is not without is not without criticism.

Two-sided methods include future values in the calculation of contemporaneous potential output estimates. This increases fit (as they minimise variation in potential output relative to observed output) but comes at the cost of in-period relevance for forecasting — future observations being by definition outside the information set at any point in time. The two-sided nature of the calculation also has an even more pernicious effect on long-term multipliers: as the path of GDP at time $t + 1$ affects potential GDP at time t, by construction the effect of an impulse at time t will be close to zero in the long-run. Indeed, that is what I find, with results statistically insignificant and point estimates around nil — but that is hardly supportive of a conclusion that the long-run multiplier is indeed close to zero, because we have implicitly assumed it to be so — and so it as an example of *petitio principii*, or assuming the conclusion. I would therefore posit that they are not well suited to estimating fiscal multipliers, especially in the long run.

Hamilton (2017) is a highly critical piece of the usage of the HP filter — so much so that it is titled "Why you should never use the Hodrick-Prescott filter". The paper deconstructs the example from which Hodrick and Prescott obtained the $\lambda = 1600$ rule of thumb, and shows that 1600 is very far from being an appropriate smoothing parameter that one would estimate from the data, with all of Hamilton's estimates coming in below 10 for quarterly data — and which would imply a much lower value for annual data than the commonly used 6.25. Using Ravn and Uhlig's (2002) methodology, the highest value of λ based on Hamilton's calculations would be 0.039. Figures 2.7 and 2.8 show estimated potential output and the output gap using the two values for λ, making it clear just how much flucutations are dampened by the lower value of the parameter[5].

Another major criticism from Hamilton (2017) is the fact that the HP filter can extract patterns even when none exist: filtering a random walk will yield a highly predictable series, which should not be the case if all it did was decompose the data generating process. A similar criticism regarding spurious cycles can be laid at the

[5] Assuming $\lambda^{quarterly} = 10$, we can use Ravn and Uhlig (2002) to derive $\lambda^{annual} = \frac{\lambda^{quarterly}}{4^4} = \frac{10}{256} = 0.039$.

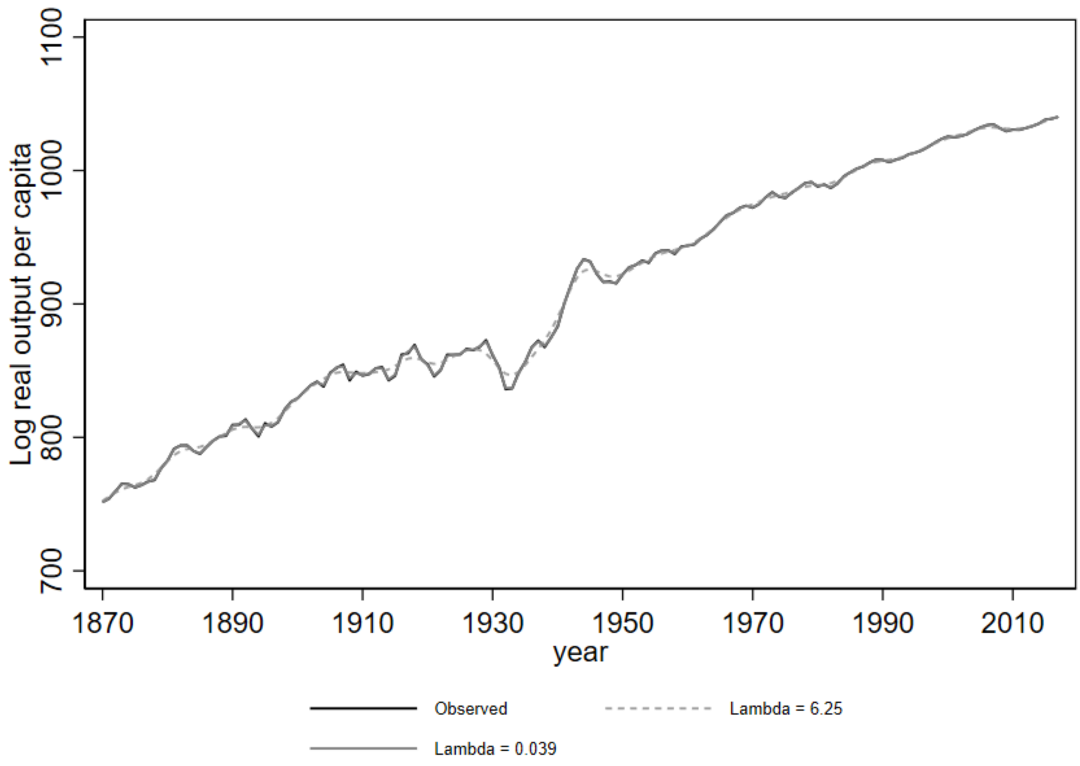


Figure 2.7: HP filter estimates of US potential log real output per capita using different values for λ.

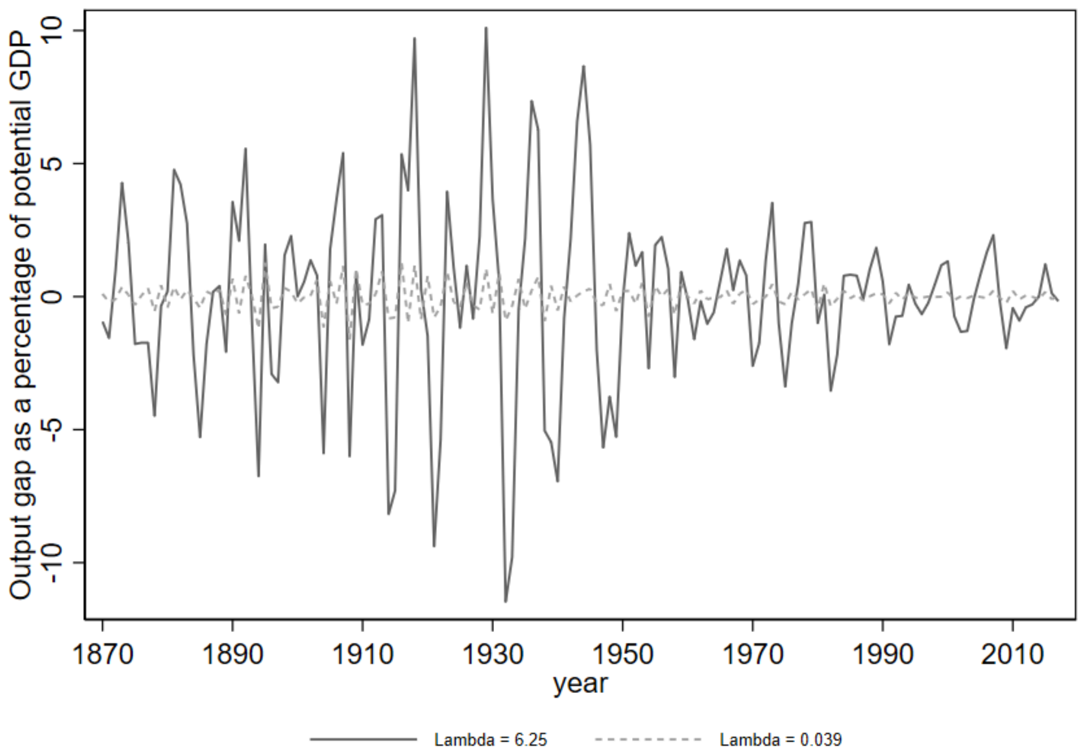


Figure 2.8: HP filter estimates of US and output gap using different values for λ.

feet of the Baxter-King and Christiano-Fitzgerald filters: Smith (2016) finds that this is the case too with digital band pass filters as a class, corroborating findings by Woitek (1998).

2.B.3 One-sided methods

Given the issues with two-sided methods, an intuitively appealing alternative would be to apply a **one-sided HP filter**. This would mean the filter would use only past and contemporaneous values, solving the following:

$$\min_{\tau_{it}} \sum_{t=1}^{T} (y_{i,t-1} - \tau_{i,t-1})^2 + \lambda((\tau_{it} - \tau_{i,t-1}) - (\tau_{i,t-1} - \tau_{i,t-2}))^2 \tag{2.11}$$

This restricts the information used in computing potential output to be contemporaneous and backward looking, which intuitively should be a solution to the problem and has been recommended for example as a way of detrending in the Basel III banking regulation framework.

A simpler way of estimating potential output is to use a **moving average (MA) filter**. MA filters are some of the simplest methods of smoothing out series, and are relatively flexible. In fact, the Baxter-King, Christiano-Fitzgerald and exponential smoothing filters all include moving averages, with the main difference being the weights attached to each observation. In this case, I test a one-sided, simple (unweighted) MA filter, as opposed to the two-sided filters mentioned before.[6]

I also test a **Kálmán filter**, initially developed for use in engineering in the late 1950s. It is a linear recursive estimator usually formulated in state-space form. The Kálmán filter has become relatively popular in macroeconomics as a way of extracting signals from noisy data, which includes output filtering to estimate potential output and hence the output gap. Its recursive form means that it uses Bayes' theorem to obtain a conditional probability density function for the unobservable potential output, which is done numerically in an iterative fashion. Mathematically, the Kálmán filter I estimate can be represented as below:

$$y^p_{it} = Dy_{it} + \nu_{it} \tag{2.12}$$

$$y_{it} = Ay_{i,t-1} + C\varepsilon_{it} \tag{2.13}$$

where $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$ and $\nu_{it} \sim \mathcal{N}(0, \Sigma_{i,\nu\nu})$, that is, both Gaussian white noise. The Kálmán filter is optimal if the observed variable (y_{it}) and the error terms are jointly normal; if not, it is the best in the class of linear filters (Parischa, 2006).

Hamilton (2017) proposes his own alternative to the HP filter which consists of running a local regression h steps ahead at each point t on the latest values of the variable itself, and using the residual as an estimate of output gap — which I will refer to henceforth as the **Hamilton method**. The intuition behind this is that the largest source of forecasting error in predicting most macroeconomic variables at a short horizon would be cyclical factors. Hamilton also shows that these simple forecasts can be estimated consistently for a wide range of non-stationary processes without knowing the true data generating process or having the correct forecast

[6]After testing, I chose an MA filter of order 3 for this comparison exercise.

model, which makes this potentially a very versatile tool. The Hamilton method was proposed for quarterly data, but it can easily be transformed into an annual setting, estimating the following equation:[7]

$$y_{i,t+h} = \beta_{i,0} + \sum_{q=1}^{p} \beta_{i,q} y_{i,t-q} + \nu_{i,t+h} \tag{2.14}$$

and then estimate potential output using the fitted values:

$$y_{it}^{p} = \hat{\beta}_{i,0} + \sum_{q=1}^{p} \hat{\beta}_{i,q} y_{i,t-q} \tag{2.15}$$

where $\hat{\beta}_{i,q}$ are the ordinary least squares estimates of the $\beta_{i,q}$ parameters in equation (10) for each country i. This is related to the Beveridge and Nelson (1981) decomposition of a non-stationary series into a random walk with drift and a serially correlated stationary process, as shown by Hodrick (2020).

And finally, I replicate Gordon and Krenn's (2010) approach of using a simple deterministic trend through observed output. Given different averages in trend growth over long cycles, a polynomial rather than a linear trend provides a much more useful fit, and it is indeed in line with what has been used in the literature (e.g. Ramey and Zubairy, 2018), so that output is assumed to follow the following process:

$$y_{it} = \alpha_{i,0} + \sum_{q=1}^{p} \alpha_{i,q} t^{q} + \mu_{it} \tag{2.16}$$

and then potential output is estimated using the fitted values:

$$y_{it}^{p} = \hat{\alpha}_{i,0} + \sum_{q=1}^{p} \hat{\alpha}_{i,q} t^{q} \tag{2.17}$$

where $\hat{\alpha}_{i,q}$ are the ordinary least squares estimates of the $\alpha_{i,q}$ parameters in equation (2.17) for each country i.

2.B.4 Merits and drawbacks of one-sided methods

Most of these methods — Hamilton, MA and deterministic polynomial trends — are much simpler than the filtering methods presented in the previous section. They also lead to larger deviations between observed and potential output. This should not be surprising, because they are less restrictive relative to observed output; all two-sided methods presented are constructed as minimising deviations from actual output as a way of calculating potential output. Such formulations imply that more of the shocks come from the supply side, and therefore leave less scope for demand management intervention on the demand side. In some sense, they provide a very RBC-consistent view of the world, and it is not surprising that the popularisation of the HP filter arose around the RBC modelling project. A more Keynesian view of business cycle fluctuations is more consistent with non-filtering methods, which

[7] I implement this using $p = 3$ and $h = 2$, the latter being the analogous to the 8 quarters used by Hamilton.

allow for large and sometimes relatively persistent output gap realisations, which can therefore allow for a more active demand management stance. The largest estimates for cyclical gaps comes from using polynomial trends, which is not surprising because I have estimated them with time-invariant coefficients. It is possible that testing of structural breaks might lead to more accurate estimates; however, with such a large dataset, this would be quite cumbersome. Filtering methods are much more tractable when it comes to obtaining estimates that take such variation into account.

Both the Kálmán filter and the determinstic polynomial trend method require a larger degree of judgement. In the case of the Kálmán filter, that is because it is a more general procedure than some of the other filters I have considered before. The filter's optimality (that is, minimum mean squared error) is conditional on the law of motion described, but that puts more weight on the law of motion, in which a degree of judgement must be exercised. In this case, I have only included observed log real output per capita as a determinant of potential output, as it is the most directly comparable law of motion to the other univariate methods I am using. But one might conduct an exercise of this form with other formulations of the law of motion — for example, including inflation as a determinant of potential output.

In the case of deterministic polynomial trends, judgement must be exercised in selecting the order of the polynomial for each country. One could conceivably choose the same order of polynomial for every country, but that would be too restrictive and would impose large estimated output gaps especially at times during which some countries in the sample had not started catch-up growth yet — the poor fit becomes obvious on eyeballing the data. Instead, I have applied different orders of polynomials to each country, depending on the fit of the data. Distinctions over the most appropriate order can at the margin be tricky, and using a statistical approach such as information criteria or likelihood ratios to determine which order to select might lead to some poor choices, such as the fact that some fitted values might mean that at the end of the sample potential output is falling for a number of years, especially for even-ordered polynomials. This is contrary to what one would expect, and I therefore take this into account when choosing the order p for each country.

Hodrick (2020) argues that in his simulations, the Hamilton method performs better than the HP filter for simple time series, but that the HP and Baxter-King filters are better suited for more complex data generating processes in which there are stylised facts for covariances between variables. While such filters may yield better fit in such cases, that would only be at the expense of a more opaque and less intuitive method, which relies on parameters that are difficult to interpret and on information unavailable at the time to identify potential output. And it is not completely clear that better fit is necessarily desirable, as discussed above regarding one's view of the determinants of the output gap and the actual relationships in the data between macroeconomic variables. When one uses Hamilton's empirical estimates to derive the HP filter's λ parameter, as figure 1 shows, estimated potential output fits incredibly well relative to observed output — but that means that business cycle fluctuations are severely dampened.

Regarding the one-sided HP filter, despite its intuitive appeal, Wolf et al. (2020) show that it does not in fact solve the problem: the lack of post-period data means it fails to remove low-frequency fluctuations as well as the two-sided version, and it also dampens fluctuations at all frequencies. This corroborates Hamilton's (2017) view that the HP filter's ability to seemingly predict future observations accurately comes only from its use of future values, and that when it cannot, the HP filter imposes its own dynamics on the time series it is attempting to decompose rather than matching the data generating process, and thus could not be observed in real time.

2.C First stage statistics

Figure 2.9 shows the Kleibergen-Paap Wald F-statistics at each horizon h for the Gordon-Kreen and conversion ratio estimates for the US. The objective is to gauge whether the instruments used (war dates and Blanchard-Perotti shocks) are strong enough for identification. Also plotted are the Montiel Olea and Pflueger (2013) critical values (23.1086, accounting for serial correlation) and the Staiger and Stock (1997) rule-of-thumb critical value (10). Similarly to Ramey and Zubairy (2018), these instruments easily pass these critical values at shorter horizons, and still over the Staiger and Stock values by year 6.

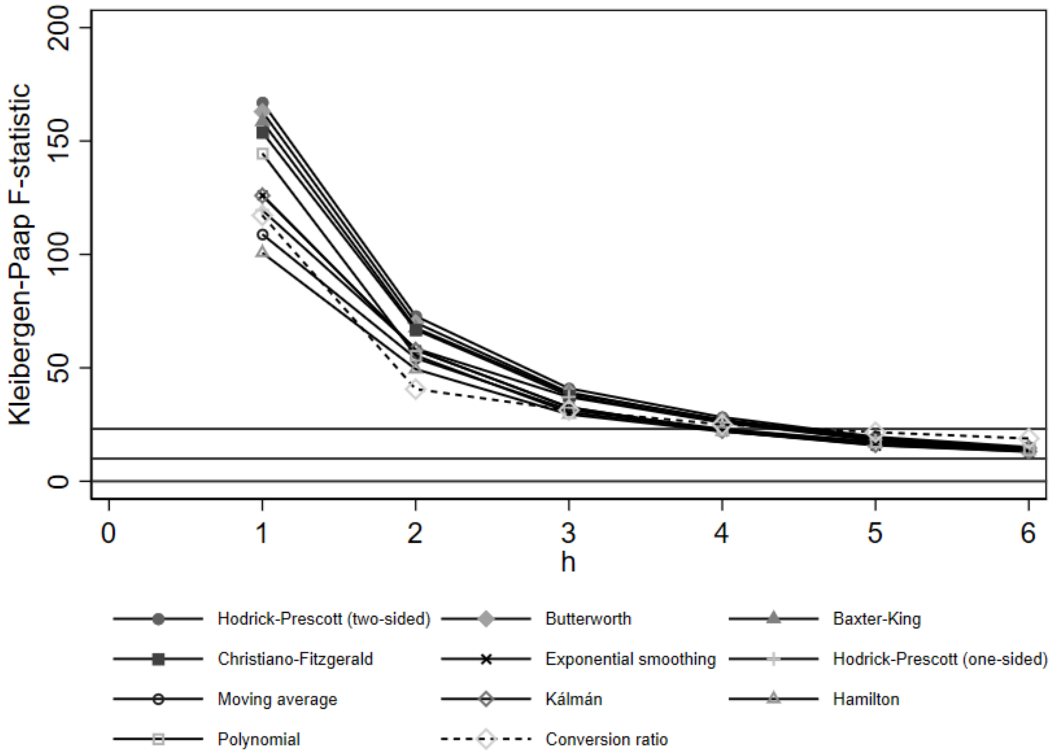


Figure 2.9: Kleibergen-Paap Wald F-statistics for the US-based estimates. Horizontal lines represent the Montiel Olea and Pflueger (23.1086) and Staiger and Stock (10) critical values.

Figure 2.10 replicates figure 2.9 for the 18-country panel, and shows the two instruments are even stronger in the panel setting. As with Ramey and Zubairy, I find that instrument strength does not help decide between models — they are all similar and generally pass the critical values.

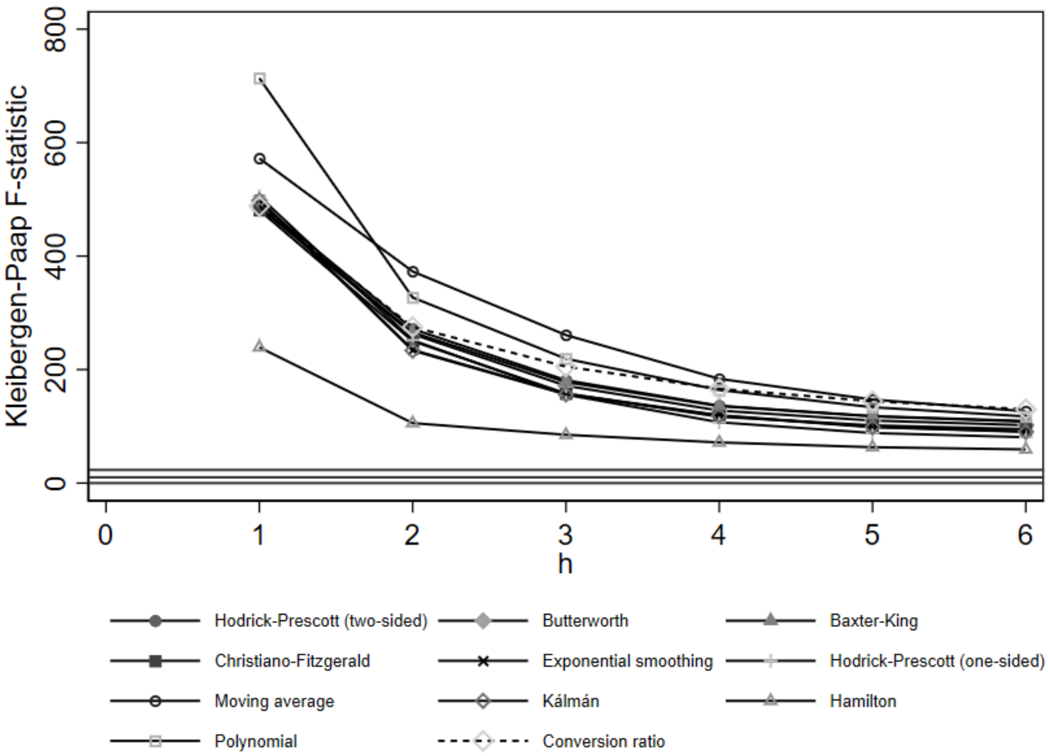


Figure 2.10: Kleibergen-Paap Wald F-statistics for the 18-country panel estimates. Horizontal lines represent the Montiel Olea and Pflueger (23.1086) and Staiger and Stock (10) critical values.

2.D Specification of the extension to a panel setting

The specification is for the 18-country panel is essentially identical to that which only includes only the US, with the addition of a country fixed effect term, f_i and with additional subscript i for each variable. So for the Gordon-Krenn method, this becomes:

$$\sum_{j=0}^{h} \frac{y_{i,t+j}}{y^p_{i,t+j}} = \phi^{gk}_h Lz_{i,t} + \beta^{gk}_h \sum_{j=0}^{h} \frac{g_{i,t+j}}{y^p_{i,t+j}} + f^{gk}_i + \mu^{gk}_{i,t+h} \tag{2.18}$$

$\gamma^{gk}_h = \beta^{gk}_h$, so no transformation is required to obtain the multiplier estimate in this case.

Similarly, for the conversion ratio method, the elasticity equation becomes:

$$\sum_{j=0}^{h} \ln y_{i,t+j} = \phi^{cr}_h Lz_{i,t} + \beta^{cr}_h \sum_{j=0}^{h} \ln g_{i,t+j} + f^{cr}_i + \mu^{cr}_{i,t+h} \tag{2.19}$$

These elasticity estimates can then be converted into multiplier estimates using the following:

$$\hat{\gamma}^{cr}_h = \hat{\beta}^{cr}_h \times \frac{\bar{y}}{\bar{g}} = \hat{\beta}^{cr}_h \times \frac{\sum_{i=0}^{N} \sum_{t=0}^{T} y_{i,t}}{\sum_{i=0}^{N} \sum_{t=0}^{T} g_{i,t}} \tag{2.20}$$

To allow for varying multiplier estimates over time, we can again simply add a country subscript and fixed effects to obtain the panel specification. So for the Gordon-Krenn estimate, that becomes:

$$\begin{aligned} \sum_{j=0}^{h} \frac{y_{i,t+j}}{y^p_{i,t+j}} &= I_{i,t-1} \left[\phi^{gk}_{A,h} Lz_{i,t} + \beta^{gk}_{A,h} \sum_{j=0}^{h} \frac{g_{i,t+j}}{y^p_{i,t+j}} + f^{gk}_{A,i} \right] \\ &+ (1 - I_{i,t-1}) \left[\phi^{gk}_{B,h} Lz_{i,t} + \beta^{gk}_{B,h} \sum_{j=0}^{h} \frac{g_{i,t+j}}{y^p_{i,t+j}} + f^{gk}_{B,i} \right] + \zeta^{gk}_{i,t+h} \end{aligned} \tag{2.21}$$

The following equation describes the equation for the conversion ratio method:

$$\begin{aligned} \sum_{j=0}^{h} \ln y_{i,t+j} &= I_{i,t-1} \left[\phi^{cr}_{A,h} Lz_{i,t} + \beta^{cr}_{A,h} \sum_{j=0}^{h} \ln g_{i,t+j} + f^{cr}_{A,i} \right] \\ &+ (1 - I_{i,t-1}) \left[\phi^{cr}_{B,h} Lz_{i,t} + \beta^{cr}_{B,h} \sum_{j=0}^{h} \ln g_{i,t+j} + f^{cr}_{B,i} \right] + \zeta^{cr}_{i,t+h} \end{aligned} \tag{2.22}$$

These estimates can then be converted into multiplier estimates as follows:

$$\hat{\gamma}^{cr}_{A,h} = \hat{\beta}^{cr}_{A,h} \times \frac{\bar{y}_A}{\bar{g}_A} = \hat{\beta}^{cr}_{A,h} \times \frac{\sum_{i=0}^{N} \sum_{t=0}^{T} 1_{\{I_{i,t-1}=1\}} y_{i,t}}{\sum_{i=0}^{N} \sum_{t=0}^{T} 1_{\{I_{i,t-1}=1\}} g_{i,t}} \tag{2.23}$$

$$\hat{\gamma}^{cr}_{B,h} = \hat{\beta}^{cr}_{B,h} \times \frac{\bar{y}_B}{\bar{g}_B} = \hat{\beta}^{cr}_{B,h} \times \frac{\sum_{i=0}^{N} \sum_{t=0}^{T} 1_{\{I_{i,t-1}=0\}} y_{i,t}}{\sum_{i=0}^{N} \sum_{t=0}^{T} 1_{\{I_{i,t-1}=0\}} g_{i,t}} \tag{2.24}$$

Figure 2.11 replicates figure 2.2 for the 18-country panel, showing a similar divergence between pre- and post-1946 averages for the conversion ratio Y/G, and shows that the differences across time periods are similar to those for the US only.

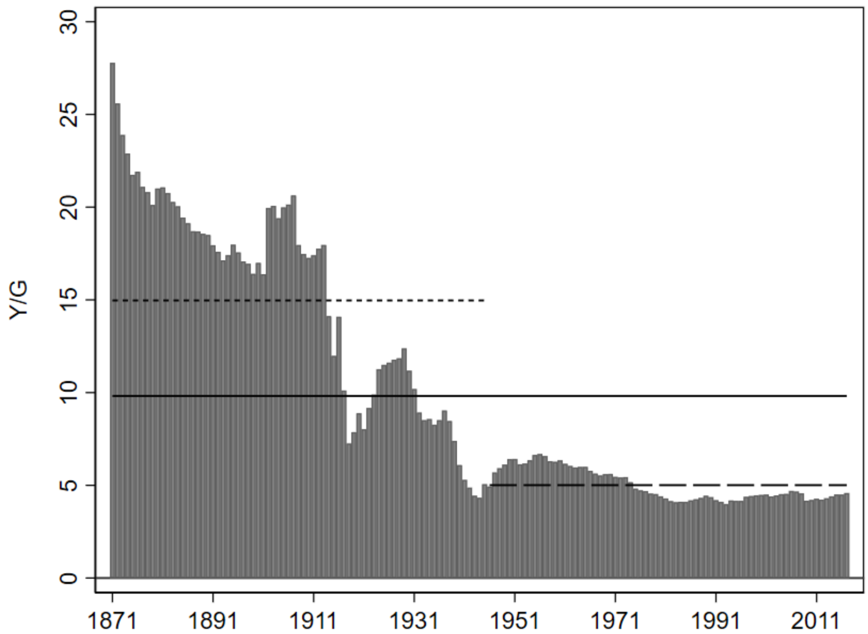


Figure 2.11: Annual average of Y/G across the full sample (1871-2017, 18 countries). Solid line represents the sample average across the whole sample (9.8162). Short-dashed line represents the sample average before 1946 (14.9701). Long-dashed line represents the sample average from 1946 onwards (5.0094).

2.E Additional analysis and results

Table 2.8 shows the results of the testing of statistically significant differences in multiplier estimates for the US in each state, using Welch's (1947) t-test of difference in means of populations with different variances. The results show that despite the clear pattern of positive differences (meaning larger pre-1946 point estimates than post-1946), a far smaller subset of those differences are statistically distinguishable from zero at conventional levels of significance. The exponential smoothing, moving average and Kálmán filters show higher pre-1946 multipliers in the long run, but no statistically significant differences in the short run. The polynomial method of estimating potential output is significantly different at all horizons, as is the conversion ratio method. The finding of no statistically significant difference between periods when using many of the potential output estimation methods suggests there may be attenuation bias with those methods — when using the conversion ratio method, I find multipliers to be larger pre-1946, which is what would be expected.

	Difference between 1871-1945 and 1946-2017 multipliers						
Method	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Two-sided HP	0.0815	0.0458	-0.0057	-0.0387	-0.0372	-0.0271	-0.0471
Butterworth	0.0675	0.0440	0.0076	-0.0204	-0.0115	0.0099	-0.0114
Baxter-King	0.1169	0.0958	0.0629	0.0287	0.0183	0.0251	-0.0303
Christiano-Fitzgerald	0.0573	0.0229	-0.0131	-0.0097	0.0361	0.0738**	0.0398
Exponential smoothing	0.3017*	0.1258	0.0330	0.0469	0.1510*	0.2344**	0.2184**
One-sided HP	0.0417	0.0079	-0.0041	-0.0075	0.0030	0.0073	-0.0007
Moving average	0.0612	0.0306	0.0232	0.0925	0.2353	0.4086**	0.5707**
Kálmán	0.2061	0.0860	0.0524	0.0653	0.1799*	0.2898**	0.3273**
Hamilton	-0.0471	-0.0587	-0.1327	-0.0643	0.0537	0.2123	0.2743
Polynomial	0.3306**	0.3477**	0.3820**	0.4460**	0.5653***	0.7795***	1.0204***
Conversion ratio	1.3213*	1.2675**	1.1707*	1.1579	1.3960*	1.8500**	2.3355**
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.8: Difference in the estimates of US multipliers using the 1871-1945 and 1946-2017 subsamples. p-values underlying the statistical significance calculations use Welch's t-test for the difference in means.

Table 2.9 shows the differences between pre- and post-WW2 multipliers for each method when using the 18-country panel. Quite a few of methods display statistically significant differences between pre- and post-1946 estimates, although not always in the same direction in both cases, which is surprising. Intuitively one would expect a much larger state to have a more limited marginal effect on output, or at least no smaller, and so the prior for differences in those tables would be for them to be positive. That is what I find consistently when using the conversion ratio, and those differences are statistically significant in both cases, especially in the short-run.

	Difference between 1871-1945 and 1946-2017 multipliers						
Method	In-year	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Two-sided HP	0.2324**	0.2294**	0.1901***	0.1500***	0.1087***	0.0525*	0.0037
Butterworth	0.1790*	0.1612**	0.1225***	0.0928***	0.0658**	0.0239	0.0161
Baxter-King	0.1987**	0.1953***	0.1667***	0.1309***	0.1015***	0.0473*	0.0245
Christiano-Fitzgerald	-0.0103	-0.0701	-0.0894	-0.0722*	-0.0385	-0.0274*	-0.0003
Exponential smoothing	0.2952*	0.1680	0.0723	0.0241	0.0189	0.0088	0.0462
One-sided HP	-0.0389	-0.0315	-0.0353*	-0.0341**	-0.0292**	-0.0253**	-0.0149*
Moving average	0.1391	0.1326	0.1186	0.0628	0.0004	-0.0775	-0.0963
Kálmán	0.2717	0.2105*	0.0968	0.0110	-0.0496	-0.0957*	-0.0628
Hamilton	-0.4326	-0.3712	-0.3954	-0.4600	-0.4822	-0.5149*	-0.4911*
Polynomial	-0.2053	-0.2053	-0.2518*	-0.2866*	-0.3261*	-0.4471**	-0.5099**
Conversion ratio	1.4641**	1.5636**	1.3982**	1.1801*	0.9074	0.5738	0.4042
Significance: *** - 1%; ** - 5%; * - 10%							

Table 2.9: Difference in the estimates of multipliers for the 18-country panel using the 1871-1945 and 1946-2017 subsamples. p-values underlying the statistical significance calculations use Welch's t-test for the difference in means.

Chapter 3

What can we learn about UK fiscal multipliers from 140 years of policy?

Abstract

I use a novel dataset compiled from archival research in the UK Parliament with 140 years (1879 to 2018) of in-year government spending shocks, which are unlikely to be anticipated due to the UK's idiosyncratic budget process — an assertion supported by statistical tests on the shocks. I find a multiplier of 0.44 on impact and 0.47 in the long-run, along with some evidence of larger stimulative effects from civil spending shocks at short horizons relative to military spending shocks. Effects on other macroeconomic variables support results from New Keynesian workhorse models, as well as negative consumption effects found in empirical studies using large military spending shocks. I also find evidence of larger multipliers in states of high slack as measured by unemployment considerably above the natural rate, but not for other measures of slack nor for broader measures of economic regimes such as levels of debt-to-GDP, openness to trade and exchange rate regimes.

3.1 Introduction

When deciding on the appropriate decisions for what fiscal policy to implement, it is imperative to consider what its effect will be on output. This effect is usually captured through a multiplier, but estimating that consistently is notoriously difficult. For example, fiscal aggregates such as public spending and tax revenues respond to economic conditions, making hard to disentangle the effects of policy decisions from automatic stabilisers. Policy decisions are also not independent of how the economy is performing — governments look at how they expect variables such as GDP growth, employment and inflation to evolve in order to set their fiscal policy, and so observed macroeconomic outcomes may belie different ones that might have occurred in the absence of policy interventions. And that is to say nothing of the fact that policy announcements might be anticipated. This kind of endogeneity permeates fiscal policy announcements.

As such, a lot of work has been conducted in attempting to find exogenous, unanticipated shocks that can isolate the effects of government spending on output. Narrative approaches are an important source of such shocks, but these are costly to assemble and have generally been limited to the United States. There is also limited literature for fiscal multipliers focused specifically in the United Kingdom, especially on the spending side. Cloyne (2013) uses a narrative approach to identify tax shocks from archival data, but no analogue exists on the spending side. Studies tend to use data-driven identification strategies instead of the narrative approach of the canonical multipliers literature such as Ramey and Shapiro (1998), Gordon and Krenn (2010), Romer and Romer (2010) and Ramey (2011b). Glocker et al. (2019) note that "[c]onstructing a direct measure of spending shocks for the UK from variations in defense spending would require a sizeable archive work and the use of historical data (most of the action in defense spending occurred before the 1970s)." This paper sets out to provide such a historical data series and to discern what can be learned from it when assessing the effect of government spending on output in the UK.

The reason why this matters is that most of the literature has focused heavily on the United States historically, but studies using US data do not necessarily translate into other countries, not least because of the size of the domestic market and the anchor role of the US dollar across the dollar allowing more monetary freedom. But despite that, institutions such as the Office for Budget Responsibility (OBR) — the UK's official fiscal forecaster — are required to make assumptions about the effects of government spending, and often have to rely on US-based estimates to do so.

Furthermore, the UK's budget-setting process provides an ideal source of data for this kind of econometric estimation. Near or just after the beginning of the each financial year — which in the United Kingdom runs from 1 April to 31 March — the Chancellor of the Exchequer is required (and has been for nearly 150 years) to present their estimate of how much the Exchequer actually spent in the previous and how much it forecasts spending to be in the coming year, which given the way the UK's parliamentary system works, presents a way of obtaining a best estimate of the discrepancy between forecast and actual discretionary spending[1] - essentially an intra-year unanticipated spending shock. This allows me to compile a dataset of shocks from 1877-78 to 2018-19, a total of 142 years[2]

I can then combine these shocks with broader UK annual macroeconomic data, which I obtain from the UK's Office for National Statistics for 1946 onwards and which I splice with pre-1946 consensus estimates from Thomas and Dimsdale (2017). I estimate output elasticities at each horizon from years 0 to 4 after the shock, which I convert to multipliers using Y/G as a conversion ratio. I find evidence of statistically significantly positive multipliers of 0.44 on impact and 0.47 in the long run. I also use the breakdown of the shocks into military and civil components, with some evidence of civil spending being associated with larger multipliers than military spending at short horizons. I test the effect of the shocks on other macroeconomic variables, which show results that are broadly consistent with New Keynesian model results, although not necessarily all the empirical literature. The fall in private consumption spending in particular mirrors other studies that include large military shocks as their identification strategy.

[1]As detailed in the data section, I remove debt interest and social security spending because of their inherent correlation with macroeconomic conditions as automatic stabilisers.

[2]Although I have 142 years in the dataset, I have to drop two observations from the regression equations, as I use inflation as one of the control variables (which means dropping 1877-78) and I use one lag of each macroeconomic variable (which means dropping 1878-79).

The local projections framework I use allows me to test whether there is evidence of state-dependent effects of fiscal policy on output across a number of measures of states. I only find strong statistical evidence of higher multipliers in states of unemployment considerably above the natural rate, but not for other measures of slack — nor do I find results to suggest differences in multiplier across regimes such as high and low debt stocks, a more or less open economy or different exchange rate regimes.

The rest of the chapter is structured as follows: section 3.2 reviews the literature on UK multipliers and where this paper fits in terms of gaps; section 3.3 summarises the data used and the shocks, testing them to show that there is no strong evidence of anticipation; section 3.4 details the methodology employed; section 3.5 presents the results for the full shock, the breakdowns of military and civil spending, the effect on other macroeconomic variables, and state-dependent estimates; and section 3.6 concludes.

3.2 Literature review

The literature on fiscal multipliers is heavily focused on the United States. Take Ramey (2011b) for example, which summarises significant contributions to the literature — the number of papers looking at the US far outweighs the few mentioned which look at other countries. Even in those cases, they tend to be done as panels rather than focusing on the specifics of a particular country. And that means they are more likely to be data-driven approaches, which are more practical for a panel (Ilzetzki et al., 2013). Constructing a series such as the ones used by Ramey (2009) for another country would require a lot of archival work, let alone for a number of countries.

As such, it is perhaps unsurprising that the literature on fiscal multipliers for the United Kingdom is generally quite sparse. On the tax side, Cloyne's (2013) seminal work employs a narrative approach by employing archival research to come up with discretionary tax changes on the basis of HM Treasury documents, and finding a multiplier of 0.6 on impact and 2.5 after three years.

But no analogue of Cloyne (2013) has been compiled on the spending side, for which even the limited number of studies conducted remains based on data-driven methods for identification rather than the narrative approach favoured by seminal multiplier papers such as Ramey and Shapiro (1998), Gordon and Krenn (2010), Romer and Romer (2010) and Ramey (2011b).

The UK does feature as part of some multi-country studies, with a lot of interest focused on such studies in the aftermath of the 2007-08 financial crisis and responses by government across the world in its aftermath. Barrell et al. (2012) finds a government consumption multiplier of 0.74, but it is a simulation study using the National Institute Global Econometric Model (NiGEM) and not an econometric one by itself. Cimadomo and Bénassy-Quéré (2012) find a spending multiplier for the UK of 0.28 on impact — in line with the 0.30 found by Perotti (2005) as well — though they find non-Keynesian effects in some parts of the time period they analyse (1971 to 2009). Baum et al. (2012), using data from 1970 to 2011, find even lower spending multipliers for the UK, at 0.2 after a year and 0.1 after two — considerably lower than for other countries in their sample.

Recent UK-specific studies of note include Rafiq (2014), who uses data a time-varying framework with data from 1959 to 2009, and finds a multiplier of 0.93 after six months — considerably higher than that found for the UK in empirical studies in which other countries were present. Rafiq's Bayesian time-varying VAR framework also points to the multiplier in the UK being above 1 since the 1990s, with no evidence of long-term effects on the level of GDP. This is in contrast to Shaheen and Turner's (2020) results, which show negative spending multipliers across their sample, to a low of -0.42 after three years — a result not in line with most of the recent literature. They do find positive multipliers for non-boom conditions, but even then they are low (0.25 at most).

Glocker et al. (2019) estimate the government spending multiplier for the UK to be 0.48 on average across states, although larger (1.21) when in recession and smaller (0.35) in non-recessionary periods. But they note that "[c]onstructing a direct measure of spending shocks for the UK from variations in defense spending would require a sizeable archive work and the use of historical data (most of the action in defense spending occurred before the 1970s)" and that it was beyond the scope of their paper. This paper sets out to provide such a historical data series and to determine what can be learned about historical data about the effects of government spending on UK output and other macroeconomic variables.

As for the effect of government spending on other macroeconomic variables, empirical results are mixed and not always in line with theoretical models. This is particularly well-documented in the case of private consumption, as discussed in detail by Ferrara et al. (2021). Theoretical models, including workhorse New Keynesian models, predict a fall in private real consumption as a result of a shock that increases government spending. This is corroborated by Ramey (2009, 2012), but contradicted by other studies, including Blanchard and Perotti (2002), Ravn et al. (2006), Ferrara et al. (2021) and — focusing on UK multiplier estimates — Rafiq (2014).

Effects on inflation and prices are also contradictory between the baseline New Keynesian model and empirical evidence, as detailed by Ferrara et al.'s (2021) and Jørgensen and Ravn's (2022) comprehensive literature reviews and Ramey (2019). New Keynesian models predict consistently an inflationary effect from an expansionary fiscal shock, including for increasing public spending, but the empirical evidence is mixed. Take the two recent studies mentioned above as an example: Jørgensen and Ravn (2022) find a negative effect of government spending on inflation across a number of specifications using a structural VAR (SVAR) with US data from 1951 to 2008; using a proxy SVAR for US data from 1964 to 2015, Ferrera et al. (2021) find instead a positive effect on inflation. These inconsistencies are reflected across a number of studies, as both the aforementioned papers detail. But there is also a large segment of the literature — including influential papers such as Mountford and Uhlig (2009) and Auerbach and Gorodnichenko (2012) — which find no significant lasting effect of fiscal policy on inflation.

The effect on interest rates found in the literature is also inconsistent with the standard New Keynesian framework, in which we would expect them to rise in response to an increase in government spending. Instead, empirical results generally find either no effect or a fall in interest rates (Murphy and Walsh, 2022).

The effect of fiscal policy on employment is the more consensual across the literature. Recent work on the topic including Monacelli et al. (2010), Nakamura and Steinsson (2014), Suárez Serrato and Wingender (2016) and Dupor and Guerrero (2017) all find positive employment effects from government spending, although of different magnitudes.

A growing literature has in recent years focused on state-dependent effects of fiscal policy. One of the main areas of interest is the very Keynesian idea of larger effectiveness of loose fiscal policy during recessions. Again, most of this analysis has been done on US data. Gordon and Krenn (2010) estimate that in a state of high slack — such as during the Great Depression but before the supply constraints of the Second World War effort came through — multipliers were relatively high, at around 1.8. Auerbach and Gorodnichenko (2012) find similarly that multipliers are considerably larger in recessions than in expansions. However, neither Ramey and Zubairy (2018) nor Ghassibe and Zanetti (2022) find strong support for this — with the latter showing large multipliers are associated with demand-driven recessions, but not supply-driven ones.

Ilzetzki et al. (2013) use a broader suite of measures to test state dependence. They find that higher openness to trade is associated with lower multipliers, which supports the idea that fiscal stimulus in open economies has a higher degree leakage; that fixed exchange rates are associated with higher multipliers, in line with predictions from the Mundell-Fleming model; and that countries with higher debt-to-GDP ratios tend to have smaller multipliers.

3.3 The data

3.3.1 Compiling data for the shocks

Fiscal multipliers are difficult to estimate consistently — randomised experiments are impossible and the numerous confounding factors going on at any point in the macroeconomy make it hard to isolate the effect of a government fiscal policy impulse. Long historical time series go some way towards combating this issue, as they provide us with a larger pool of shocks from which to draw inferences (Ramey and Zubairy, 2018), including large-scale wars: as Angus Deaton quipped in his response to Hall (1986), "nothing can be known *without* the wars." And while the use of long time series is not without its drawbacks — the size and scope of government has changed hugely since the 19th century, and consistent measurement becomes more of an issue the further back we go in time — I believe it to be a worthwhile endeavour.

To get around the difficulties in consistently estimating multipliers, one solution that has been employed frequently in the literature is to use a narrative approach to identify an exogenous shock, which is then used as an instrument for government spending in the reduced form equation. Ramey and Shapiro (1998) and Romer and Romer (2010) are two very famous uses of this approach, where they use dates for when it becomes anticipated that the US would enter into a war in the former and presidential and congressional records in the latter to identify the timing of shocks. Cloyne (2013) employs an approach such as this to UK data for tax changes.

My approach is similar to these in spirit. I have used records from the UK's Parliamentary Archives going back to 1877 to identify how much the government intended to spend in the forthcoming year and how much it estimates to have spent in the previous year. The difference between announced and actual spending is essentially an intra-year, unanticipated policy change in discretionary spending.

There is a lot in the last sentence of the previous paragraph, so it is worth outlining the rationale for arguing each of the points. First, I am only looking at discretionary spending, which means spending that the government has direct control over. That means that I have excluded two broad categories: debt interest, as it depends directly on the stock of debt outstanding and on market interest rates; and social security spending, as it depends directly on the state of the labour market. Secondly, because I use annual data, the shock occurs within the period, and so it is contemporaneous.

3.3.2 What do the shocks look like?

As both figure 3.1 and table 3.1 show, the shocks present a wide range of variation across time. This is not unexpected as they include the two World Wars — both their outbreak and their ending, neither of which would be predictable in terms of exact timing. The large positive shocks (i.e. spending above forecast) mean that the average shock is positive. The median, however, is pretty close to zero in all three cases (less than 0.1% of GDP).

	Full shock	Military only	Civil only
Average	1.1289	0.9872	0.1416
Standard error	5.2046	5.1684	0.7584
Minimum	-6.7093	-6.7849	-1.3303
Median	0.0520	0.0229	0.0142
Maximum	45.5077	45.5490	4.6169
N	140	140	140

Table 3.1: Summary statistics of the shocks in the dataset (1879-2018) as a share of GDP.

Shocks can broadly be divided into two categories. One is the most obvious, which is large-scale wars against foreign powers. This includes but is not limited to the First and Second World Wars, which were both long and extremely expensive and — crucially for the identification strategy — much more so than initially anticipated. Table 3.2 shows the most notable shocks related to wars with foreign powers. The Second Boer War (1899-1902) is the first large event in the series, and broke out in the October of 1899 — right in the middle of the financial year — just four months after the failure of the Bloemfontein Conference. This resulted in additional military spending[3], which continued during the unexpectedly protracted conflict (Miller, 2006). The lower than anticipated spending in 1902-03, on the other hand, reflects the end of the war on 31 May 1902, just two full months into the financial year.

The First World War contains by far the largest shocks in the series. The scale and expense of fighting the war was unprecedented, and manifested itself in very large increase of spending in-year. The immediate outbreak of war between Britain and Germany (and therefore the Central Powers) was midway through the financial

[3] Hansard record of House of Commons sittings of 18-20 October 1899

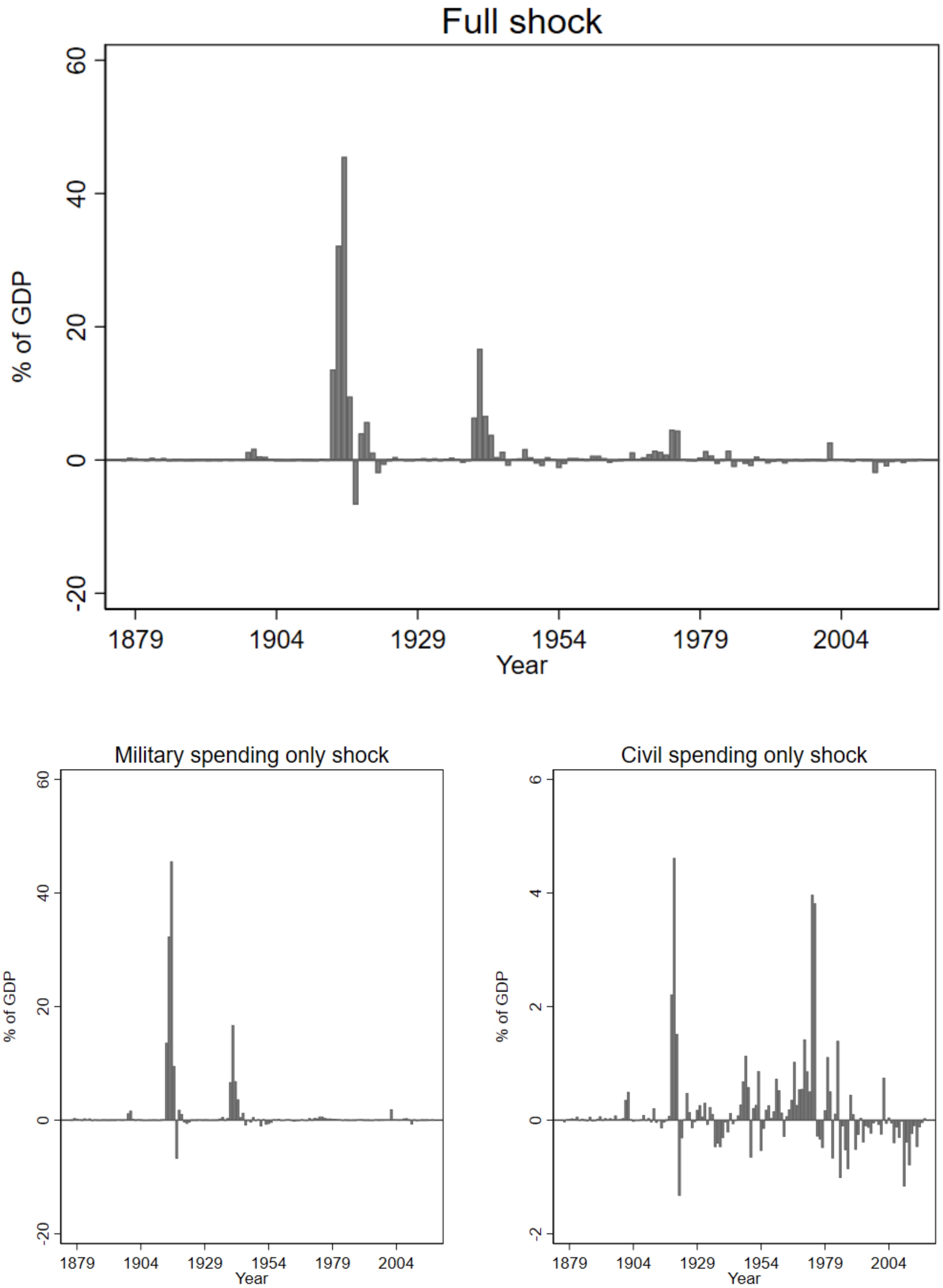


Figure 3.1: Discretionary spending shocks from 1879 to 2018 in the dataset.

year, on 4 August 1914, when Germany did not respond to the ultimatum of the Asquith Government over Belgian neutrality[4]. This was followed by an immediate vote of credit for additional military spending on 5 August, which would by no means be the last.

There is much contestation of how long governments expected the war in Europe

[4]Hansard record of the House of Commons sittings of 4-5 August 1914

to last (Halifax, 2010), but there is no doubt no one was prepared for the scale of the war that was to come. Spending over and above beginning of year estimates reached as high as 46% of GDP in 1916-17, with the financial burden of the war in Europe progressively shifting to the UK over successive war budgets (Allen, 1917). But the pattern of expenditure also reflects the largely unexpectedly quick collapse of the German Army from May to November 1918 (Deist, 1996), which led to an underspending of around 7% of GDP in the 1918-19 financial year.

The Second World War's pattern of in-year additions to spending bears some resemblance to the First World War's, but is much less pronounced. There was an initial increase in spending in the aftermath of the declaration of war, approved by the House of Commons on 1 September 1939[5]. But the largest expense over and above that in the budget was in 1940-41, which includes the German invasion of France in May, the Battle of Britain and almost the whole of the Blitz.

The financing of the Second World War was much more planned than the First — see for example Keynes (1940) — and that fed into the lower magnitudes of expenditure shocks, as well as the experience of a protracted large scale war feeding into plans into the mid-1940s. But as with any other war, the exact timing of its end was unpredictable in advance, causing the same pattern of underspending in the final year (1945-46).

After the end of the Second World War, military spending has become much less important as a source of large shocks in expenditure relative to forecasts. This is not particularly surprising, as Britain's role in world affairs has diminished since then, although it was involved in a number of military interventions such as Korean War in the 1950s and the proxy conflict in Afghanistan in the late 1970s and 1980s. The Iraq war is the sole exception to this lower level of importance, largely because of its timing: the United Nations inspections occurred in late 2002, with subsequent build-up of spending for the invasion launched on 20 March 2003, just 11 days before the end of the financial year. This was essentially a one-off occurrence — 2003-04 data shows pretty low errors in forecasts.

Event	Year	Full shock	*o/w military*	*o/w civil*
Second Boer War	1899-1900	1.20	*1.19*	*0.02*
	1900-01	1.69	*1.65*	*0.03*
	1901-02	0.58	*0.23*	*0.35*
	1902-03	0.48	*-0.16*	*0.50*
First World War	1914-15	13.61	*13.61*	*-0.01*
	1915-16	32.19	*32.33*	*-0.14*
	1916-17	45.51	*45.55*	*-0.04*
	1917-18	9.52	*9.52*	*0.00*
	1918-19	-6.71	*-6.78*	*0.08*
Second World War	1939-40	6.36	*6.68*	*-0.31*
	1940-41	16.70	*16.73*	*-0.04*
	1941-42	6.62	*6.84*	*-0.21*
	1942-43	3.78	*3.65*	*0.12*
	1943-44	0.44	*0.51*	*-0.07*
	1944-45	1.27	*1.28*	*-0.01*
	1945-46	-0.86	*-0.94*	*0.08*
Iraq War	2002-03	2.66	*1.92*	*0.75*
	2003-04	0.08	*0.14*	*-0.06*

Table 3.2: Notable shocks related to wars against foreign powers, expressed as percentage of GDP. Columns may not sum to totals due to rounding.

[5] Hansard record of the House of Commons sitting of 1 September 1939

Outside wars with foreign powers, there are also some large shocks — mostly on the civil side, but not exclusively — that are worth noting. The post-First World War period is the earliest of note, and it is a combination of issues. Some of it was higher-than-budgeted for pay settlements for both the military and civil servants, including war bonuses and the dealing with the aftermath of the influenza pandemic.[6] But it also included additional military spending abroad — including the occupation of Istanbul and of the League of Nations mandates — and in Ireland as part of the war of independence. The end of the Irish War of Independence part-way through 1921-22 also contributed to the fall in military spending in that year relative to the budget projections.

Post-Second World War, there were a number of large civil-led spending shocks. The first was immediately after the introduction of the National Health Service (NHS), whose costs immediately and severely overran in the first two years (Cutler, 2003), and which contributed significantly to overspending relative to budget in 1948-49 and 1949-50. 1967-68 was the next large increase in expenditure relative to plans during the year, as the government tried to inject demand into the economy during the summer to improve growth with the aim of avoiding devaluing sterling (Newton, 2010) — which it ended up having to do anyway on 18 November 1967.

This was followed by the most significant episode of loss of control over inflation and public spending in post-Second World War British history in the early to mid-1970s. The start of this episode predates the oil shock of 1973, though it no doubt was exacerbated by it. Inflation in Britain had been accelerating since 1968, and the government had resorted to an incomes policy to try to control it — essentially short-term limits on how much wages could rise by, imposed by the government and in the British case, agreed with trade unions and employer bodies. This of course is very different from today's consensus of the economics profession, which views inflation control as the job of monetary policy.

The incomes policy, in particular the 'stage three threshold payments' introduced in late 1973 that would become associated with the large loss of control of public spending in the face of high inflation. These threshold payments were to come into place if inflation rose above 7% — which proved to be the case, with average wages rising by nearly 15% between November 1973 and August 1974 (Ashenfelter and Layard, 1983). The Harold Wilson government then abolished the Pay Board[7], with a subsequent increase in wages of around 25% in the twelve months to August 1975. It was during this time that public spending rose well above the plans laid out at the beginning of the financial years, exceeding them by 4.6% and 4.4% of GDP in 1974-75 and 1975-76, respectively, in advance of the 1976 IMF crisis.

There are two other shocks of note in the series. One is related to the coal miners' strike in 1984, which meant the Government needed to import coal instead of procuring it domestically, adding to expenditure by around 1.4% of GDP[8]. And the second is the introduction of an austerity programme part-way through 2010-11, as part of the 22 June 2010 emergency budget, which severely restricted spending within year — by nearly 2% of GDP, across both military and civil spending.

[6] Revised Financial Statement (1919-20), Cmd. 377.

[7] Hansard record of the House of Commons sitting of 18 July 1974.

[8] Supply Estimates 1984-85, Supplementary Estimates (Classes I-XVIII), H.C. 7 (1984-85). Also see Hansard record of the 1985-86 budget statement, House of Commons sitting of 19 March 1985

Event	Year	Full shock	*o/w military*	*o/w civil*
Post-First World War adjustment: pay pressures, post-war occupations and war in Ireland	1919-20	4.02	*1.81*	*2.21*
	1920-21	5.70	*1.08*	*4.62*
	1921-22	1.13	*-0.39*	*1.52*
NHS costs overrun immediately after introduction	1948-49	1.65	*0.52*	*1.13*
	1949-50	0.43	*-0.16*	*0.58*
Currency crisis and devaluation	1967-68	1.15	*0.13*	*1.03*
Loss of control over inflation	1971-72	1.43	*0.01*	*1.42*
	1972-73	1.23	*0.37*	*0.86*
	1973-74	0.82	*0.31*	*0.50*
	1974-75	4.58	*0.61*	*3.97*
	1975-76	4.43	*0.62*	*3.82*
Coal miners' strike	1984-85	1.42	*0.03*	*1.40*
Austerity introduced mid-year	2010-11	-1.94	*-0.77*	*-1.17*

Table 3.3: Notable shocks not related to wars against foreign powers, expressed as percentage of GDP. Columns may not sum to totals due to rounding.

3.3.3 Are the shocks truly unanticipated?

This is the most important question, and in this case the UK's unique budget framework makes a more compelling case than other jurisdictions for that to be the case. The budget process is fairly well established in the UK. The Chancellor of the Exchequer must present a financial statement and budget report to the House of Commons each year in a session presided by the Chairman of Ways and Means[9]. In the Westminster system (often described as an elective dictatorship, or executive-dominant), there is no divided government — the government must have support in money bills (the budget) as a pre-condition for being in power or else it falls. The government also has the power to set the agenda of the Commons, effectively stopping any other source of money bills (Tsebelis, 2009). In the period in my sample, not a single budget failed to pass the House of Commons, and the only one (1909) which did fall in the Lords precipitated the 1911 Act which removed the Lords' veto over money bills — and was passed into law immediately after a general election. Party discipline, enforced by the whips' office, is also very strong and has been since the formation of modern political parties (1830-1860); and the combination of the House of Lords' convention not to oppose manifesto commitments and the Parliament Acts 1911 and 1949 means that the Government is effectively able to pass any legislation it brings forward.

The budget process itself also makes it less likely that the government will ignore information in its forecasting of spending for its own benefit. Because the passage of the government's budget is guaranteed by the government's very existence, there are no negotiations in public between parties. The Treasury's position as announced in the budget is taken as given, and then used for debt issuance and cash management.

[9]This is a holdover from before 1967, when the Committee of Ways and Means was abolished and full responsibility for all fiscal matters was formally transferred to the Chancellor of the Exchequer. Prior to that point, the Chancellor presented a financial statement with Government policy outlined, but formally any member of Parliament (MP) could put forward proposals for taxation and spending, although in practice the government's majority rendered this ineffective. The Chairman of Ways and Means is now the principal Deputy Speaker of the House of Commons. Since 1718, every Chancellor of the Exchequer has sat in the House of Commons rather than the Lords (with the exception of four brief interim periods), and the Commons' primacy over the Lords in money bills was formally put into law in the Parliament Act 1911.

Therefore, I would argue that any minuscule gains from gaming the forecasting system are effectively ignorable, especially in-year.

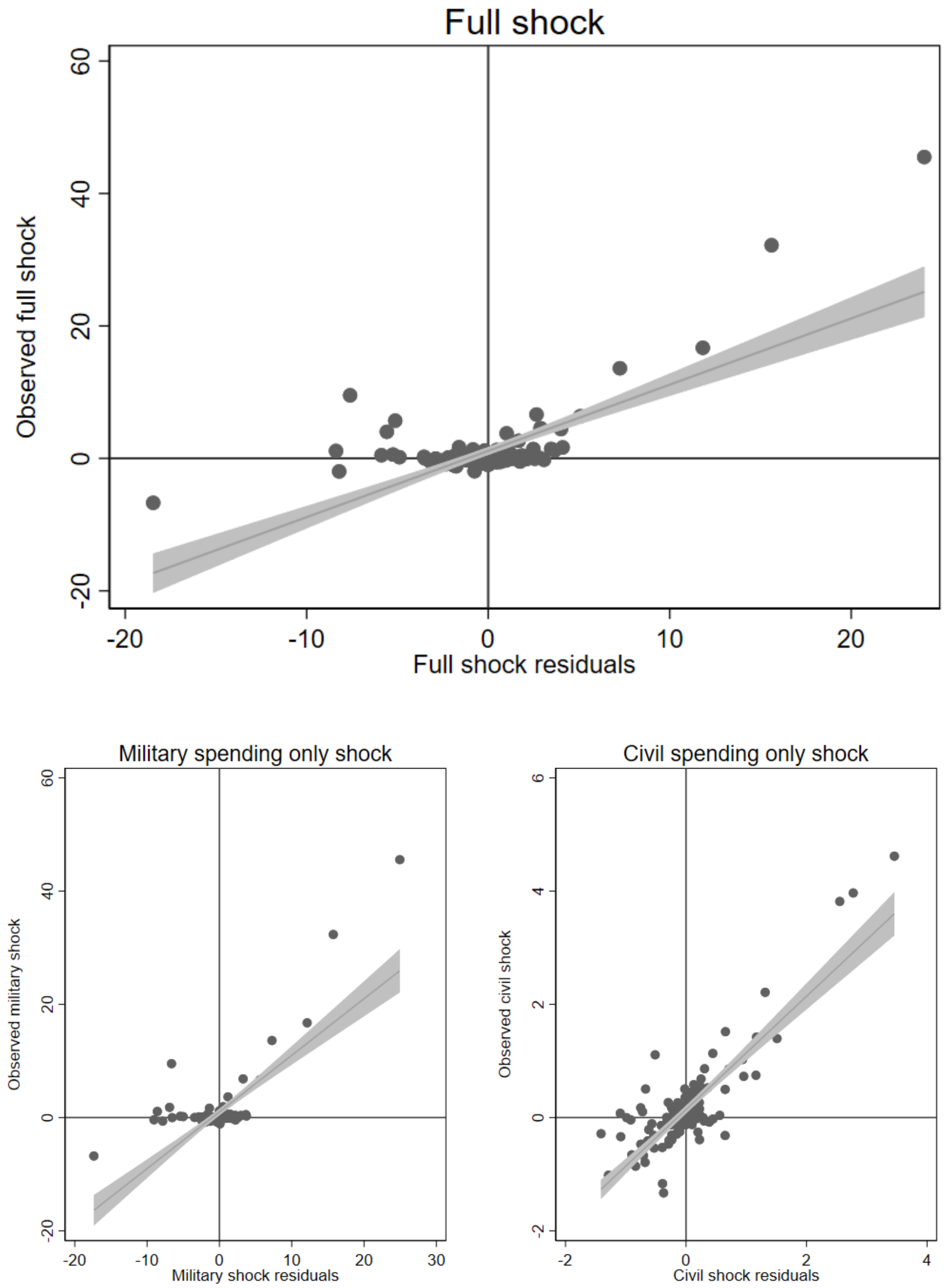


Figure 3.2: Mincer-Zarnowitz regressions of the civil and military spending shocks. Lines are the fitted Mincer-Zarnowitz regressions with 95% confidence intervals around them and represent the null hypothesis that the intercept is zero and the slope one — which I fail to reject.

Since at least 1877, the Chancellor has clearly laid out the latest estimate for expenditure for the year just gone, as well as their forecast for how much they expect to be spent in the year ahead[10]. It is these forecasts which I take from the financial statements and budget reports in the archives to compile my series of shocks, and which I am also able to break down into military and civil spending components.

Testing of the shocks using Mincer-Zarnowitz (1969) regressions supports the claim to them being exogenous. The idea behind using the Mincer and Zarnowitz method is that a shock should be unforecastable with all information available at the time at which they occur. In practice, this means using the same autoregressive structure — known at the time of the shock by definition — as in the main specification to try and forecast the shocks, and then estimating the residuals. If the shocks are unforecastable, we should not be able to reject the joint hypothesis that when regressing the shocks on their residuals, the intercept is 0 and the coefficient on the residuals is 1. Figure 3.2 shows the scatter plot of the Mincer-Zarnowitz regressions for the full shock and the military and civil spending breakdowns. In all three cases I fail to reject the null hypothesis, lending further weight to the argument for using them as an instrument in this specification.[11]

3.3.4 Other data

For macroeconomic variables, I combine two sources. The first one is the UK's Office for National Statistics (ONS), which compiles and publishes the UK's national accounts and has data on most series post-Second World War (1946 onwards). For pre-1946 data, I have spliced the ONS series with the series published by Thomas and Dimsdale (2017) as part of the Bank of England's *Millennium of macroeconomic data for the UK* project. This includes GDP, central government spending[12], household consumption, inflation as measured through the GDP deflator, the employment rate and the Bank of England's policy interest rate[13].

3.4 Methodology

3.4.1 Econometric specification

I use annual data from 140 financial years: 1879 to 2018[14] and apply a local projections with instrumental variables (LP-IV) approach, based initially on Jordà (2005)

[10]For almost all this time, the budget report was laid in the Commons in either March or April for the forthcoming financial year. More recently, this has been brought forward to the Autumn, but there is a Spring Statement in March, alongside which the Office for Budget Responsibility presents its latest forecasts, and that is what I use for the last few years of my dataset.

[11]See the appendix for detailed results of the Mincer-Zarnowitz regressions.

[12]This is the central government contribution to total managed expenditure (TME) for post-war observations and as consistent as possible with that definition before. TME is a public sector-wide metric, so it includes general government, public corporations and the Bank of England since nationalisation, but excludes public sector-owned commercial banks, and it is the most widely used metric of public sector expenditure used in the UK.

[13]Given the secular decline in real and nominal interest rates, I use a Kálmán filter to estimate the wedge between the fitted interest rate and the actual interest rate at each point in time.

[14]Although I have 142 years in the dataset, I drop two observations from the regression equations, as I use lagged inflation as one of the control variables (which means dropping 1877-78) and I use one lag of each macroeconomic variable (which means dropping 1878-79). UK financial years throughout the whole period start on 1 April, so that financial year 1879 runs from 1 April 1879 to 31 March 1880 and so on. The choice of financial years over calendar years is to take into account the Treasury budget cycle, and therefore ensure that shocks are being assinged to the

and following the synthesis of Stock and Watson (2018). I estimate the regression equations in logs at each horizon and which is a direct estimation[15] of the cumulative impulse response function (IRF), so that for each $h \epsilon \{0, 1, 2, 3, 4\}$ I estimate the following equation:[16]

$$\sum_{j=0}^{h} \ln y_{t+j} = \phi_h L z_{t-1} + \beta_h \sum_{j=0}^{h} \ln g_{t+j} + \varepsilon_{t+h} \tag{3.1}$$

where h is the horizon at which the cumulative IRF is estimated; Lz_{t-1} is a one lag operator of macroeconomic variables (output, government spending, policy interest rate, inflation, consumption, the employment rate and the debt-to-GDP ratio); g is government spending; and ε is an error term. I then convert the estimate of β_h, which is an output elasticity with respect to government spending, into a multiplier estimate using the sample average of Y/G:

$$\hat{\gamma}_h = \hat{\beta}_h \times \frac{\bar{y}}{\bar{g}} = \hat{\beta}_h \times \frac{\sum_{t=0}^{T} y_t}{\sum_{t=0}^{T} g_t} \tag{3.2}$$

For the remaining macroeconomic variables for which I estimate impulse responses (consumption, policy interest rate, inflation and employment rate) I estimate a similar equation at each horizon, but with the variable in question on the left-hand side.

Plagborg-Møller and Wolf (2021) show that the IRFs estimated using local projections are similar to those estimated using a VAR. However, the local projection framework makes it easier to estimate non-linear effects, which are of interest to me. To do so, I estimate the follow equation at each horizon h for each states A (when $I = 1$) and B (when $I = 0$):

$$\begin{aligned}\sum_{j=0}^{h} \ln y_{t+j} = I_{t-1} &\left[\phi_{A,h} L z_t + \beta_{A,h} \sum_{j=0}^{h} \ln g_{t+j}\right] \\ + (1 - I_{t-1}) &\left[\phi_{B,h} L z_t + \beta_{B,h} \sum_{j=0}^{h} \ln g_{t+j}\right] + \nu_{t+h}\end{aligned} \tag{3.3}$$

These output elasticity estimates can then be converted into multipliers in a similar way as in equation (3.2):

$$\hat{\gamma}_{A,h} = \hat{\beta}_{A,h} \times \frac{\bar{y}_A}{\bar{g}_A} = \hat{\beta}_{A,h} \times \frac{\sum_{t=0}^{T} 1_{\{I_{t-1}=1\}} y_t}{\sum_{t=0}^{T} 1_{\{I_{t-1}=1\}} g_t} \tag{3.4}$$

$$\hat{\gamma}_{B,h} = \hat{\beta}_{A,h} \times \frac{\sum_{t=0}^{T} 1_{\{I_{t-1}=0\}} y_t}{\sum_{t=0}^{T} 1_{\{I_{t-1}=0\}} g_t} \tag{3.5}$$

I then test for a number of different states, including high and low slack — measured by the unemployment rate, the wedge between the unemployment rate and an estimate of the natural rate of unemployment, and a measure of the output

correct financial year.

[15] See the appendix for discussion of the properties of this method of estimation.

[16] This horizon is roughly equivalent to the standard 20 quarters usually reported as "long run" in multiplier papers — for example, see Ilzetzki et al. (2013) and Ramey and Zubairy (2018).

gap — and regimes such as high and low debt-to-GDP ratios, high and low openness to trade, and fixed and flexible exchange rates.

I settle on a 1-lag structure for all the specifications. As I use annual data, this is similar to the 4-lag structure for quarterly data often used in these procedures — see Ilzetzki et al. (2013) and Ramey and Zubairy (2018), for example. This is also partly due to the number of observations I have available and the fact that I control for a number of variables. This is a criticism that can be levelled at this kind of analysis, of course. Going back this far in time — which I think can give us valuable insight — means using data from a time when quarterly national accounts were not available. It would be possible to interpolate data on a quarterly basis for before then — and some studies such as Ramey and Zubairy (2018) do so — but my judgement falls on the side of not introducing potentially further patterns into the data on the basis of a statistical procedure. And given that my identification strategy is well-aligned with annual data, I land on the side of using fewer observations and lags, but without interpolation.

3.4.2 Instrument strength and inference

Because of the endogeneity problem highlighted above, I use an instrumental variables approach and estimate equation (3.1) and (3.3) using two-stage least squares (2SLS). I instrument government spending using the shocks I compiled from the Parliamentary Archives, and estimate results using the full shock, with results for military and civil spending as robustness checks. I also compare the effects on other macroeconomic variables (consumption, policy interest rate, inflation and employment rate) with those in the theoretical and empirical literatures, and test whether this dataset supports recent empirical findings on state-dependence of multipliers in a number of contexts.

Inference using 2SLS estimates in an instrumental variables (IV) setting is asymptotically valid provided instruments are relevant and strong. Instrument strength is a concern in macroeconomics and in the narrative approach in particular, and many applications fail to clear the rule of thumb of a first-stage F-statistic above 10 suggested by Staiger and Stock (1997) (Ramey, 2016, 2019; Montiel Olea, Stock and Watson, 2021). Montiel Olea and Pflueger (2013) showed that a heteroscedasticity and autocorrelation robust (HAR) effective F-statistic has larger critical values — for the one instrument, just-identified case, the critical value for a worst case bias of 10% relative to the OLS estimate is 23.1085, well above the Staiger and Stock rule of thumb.

Recent work by Montiel Olea, Stock and Watson (2021) has provided a foundation for how to conduct inference in the case of weak instruments, building on work by Davidson and MacKinnon (2014) and on the recommendations of Lazarus et al. (2018). As my model is specified on a just-identified basis, I follow their approach of always using Anderson-Rubin (AR) confidence sets — that is, inverting the Anderson and Rubin (1949) test — regardless of the first-stage F-statistic rather than the 2SLS-based standard errors to compute confidence intervals. The non-parametric AR approach means that confidence sets will be identical to 2SLS ones with strong instruments, while computing still-valid sets in the presence of weak instruments.[17]

[17] I additionally opt for the quadratic spectral (QS) kernel to compute standard errors on which AR confidence sets are calculated, as Lazarus, Lewis and Stock (2021) show that the QS kernel

3.5 Results

3.5.1 Multiplier estimates

Figure 3.3 shows the multiplier estimates for the full shock — the cumulative IRF of output in response to a 1% of GDP increase in government spending. Table 3.4 details the results at each horizon up to 4 years after impact.

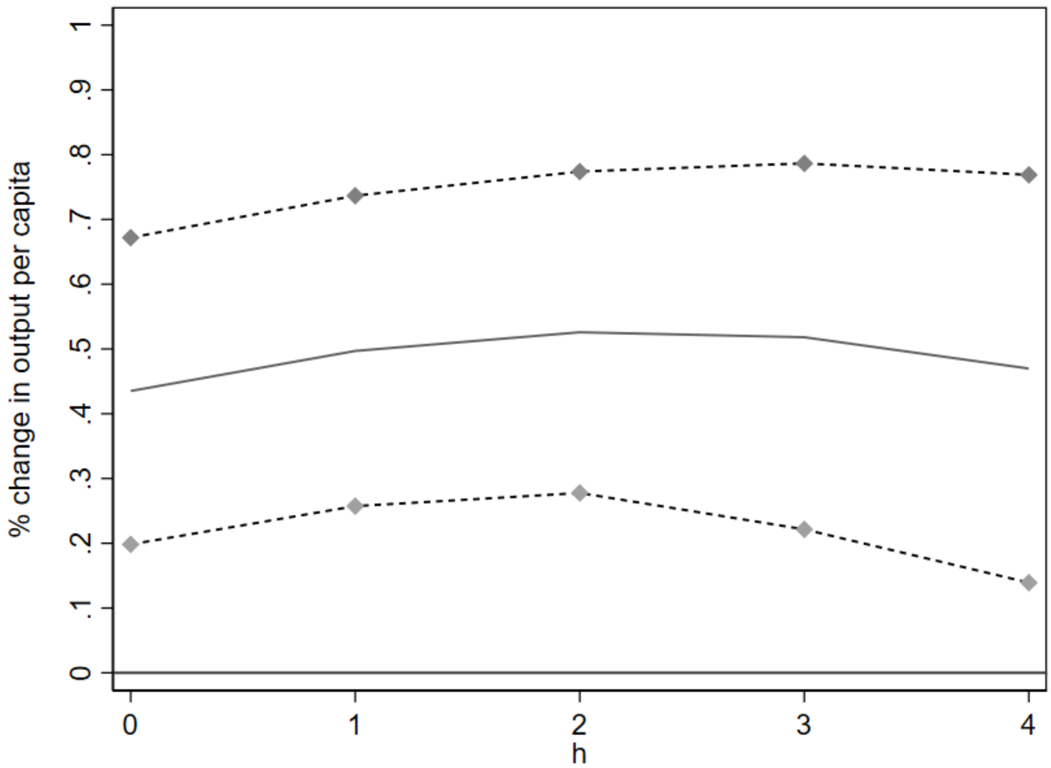


Figure 3.3: Cumulative IRFs of output per capita in response to a 1% of GDP increase in government spending per capita. Dashed lines represent the bounds of the 90% Anderson-Rubin confidence set, calculated using HAR standard errors using the QS kernel.

The multiplier estimate using the full shock is 0.44 on impact and then rises slightly, before dropping to a long-run value of 0.47. As an estimate, this is very much in line with results for the US (Ramey, 2019). Relative to UK-specific estimates, it is higher than most, though not all the literature, and more in line with Glocker et al. (2019). All estimates are statistically significant at the 1% level, save for year 4, which is significant at the 5% level.

Having the breakdown of the shock between military and civil spending allows the possibility of looking at whether each has a significantly different effect on output. As figure 3.4 shows, while the full shock and the military spending shock on its own are for the most part strong instruments (using the Kleibergen-Paap F-statistic as a measure), the civil spending shock by itself does not clear the critical values.

Nevertheless, using AR confidence sets allows me to say something about the relative size of the multipliers for the two shocks, even in the presence of a weak

achieves the size-power frontier. The decision to not use the Newey-West default kernel (Bartlett) nor its default bandwidth calculation is based on Müller (2014) and Lazarus et al. (2018), which document substantial evidence that doing so would lead to wrongly rejecting the null hypothesis too often, which in this case would mean finding statistically significant estimates too often.

	Impact (Year 0)	Year 1	Year 2	Year 3	Year 4
Output elasticity	0.0717*** (0.0247)	0.0816*** (0.0249)	0.0860*** (0.0257)	0.0844*** (0.0277)	0.0762** (0.0307)
Y/G	6.0667	6.0905	6.1148	6.1396	6.1653
Multiplier estimate	0.4352***	0.4970***	0.5258***	0.5158***	0.4698**
90% AR confidence set	[0.1984, 0.6719]	[0.2574, 0.7366]	[0.2777, 0.7739]	[0.2215, 0.7865]	[0.1392, 0.7689]
Kleibergen-Paap F-stat	47.8438	38.3113	30.0628	22.1509	16.3427
N	140	139	138	137	136
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.4: LP-IV estimates of the fiscal multiplier using the full shock. Values in brackets are heteroscedasticity and autocorrelation robust (HAR) standard errors estimated using the QS kernel. 90% confidence sets are calculated based on the inverted Anderson-Rubin test, in line with Davidson and MacKinnon (2014) and Montiel Olea, Stock and Watson (2021). Statistical significance is calculated based the whole Anderson-Rubin confidence sets being the same side of zero on the real line at the 90% (for the 10% significance level), 95% (for the 5% significance level) and 99% (for the 1% significance level) confidence levels.

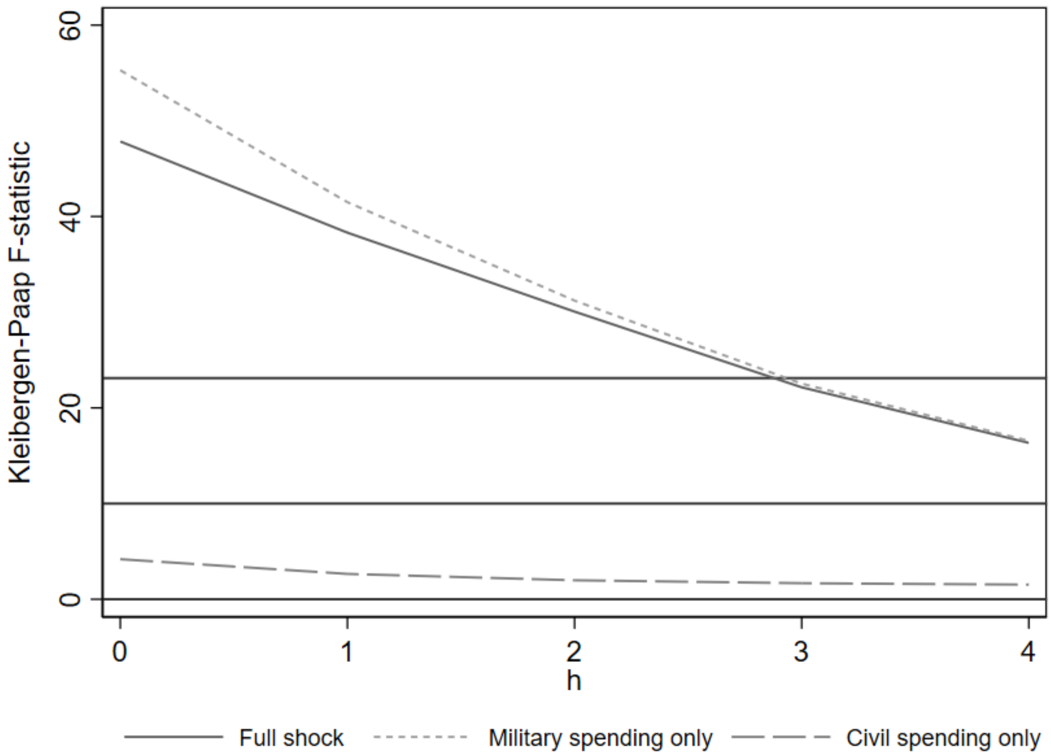


Figure 3.4: Kleibergen-Paap F-statistic at each horizon for each of the shocks.

instrument. Figure 3.5 plots the IRFs for both shocks. The military spending only shock is similar to the full shock, which is unsurprising given that military spending drives most of the large shocks (see figure 3.1). But the civil spending only shock — although underpowered and only significant at the 5% level on impact and at the 10% level after one year — has a much larger point estimate, and is statistically significantly higher than the estimate for the military spending only shock at those short horizons.

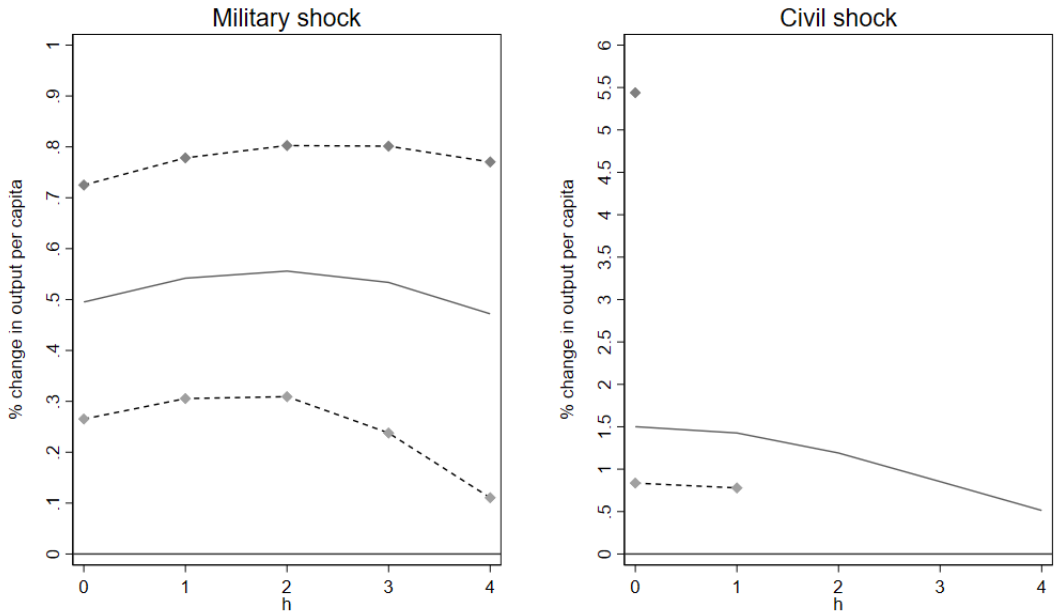


Figure 3.5: Cumulative IRFs of output per capita in response to a 1% of GDP increase in government spending per capita using the military spending only and civil spending only shocks. Dashed lines represent the bounds of the 90% Anderson-Rubin confidence set, calculated using HAR standard errors using the QS kernel.

3.5.2 Effect on other macroeconomic variables

The LP-IV framework allows me to estimate IRFs for other macroeconomic variables, much in the same way as one would in a VAR. Figure 3.6 shows the results for consumption, the policy interest rate, the employment rate and inflation of a 1% change in government spending.

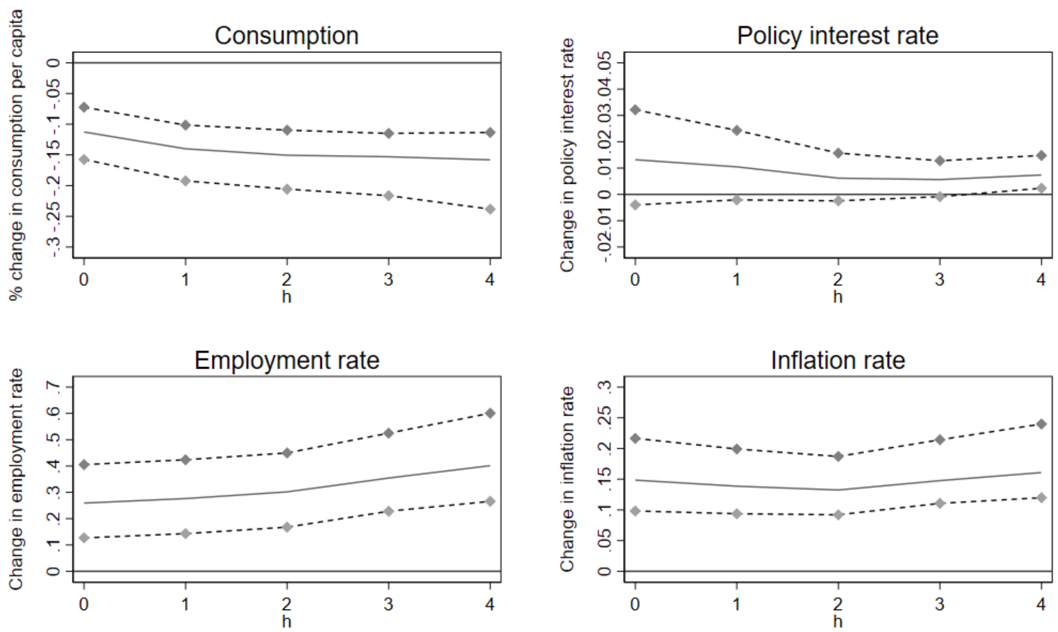


Figure 3.6: Cumulative IRFs of various macroeconomic variables in response to a 1% increase in government spending per capita using the full shock. Red dashed lines represent the bounds of the 90% Anderson-Rubin confidence set, calculated using HAR standard errors using the QS kernel.

Consumption falls in response to an increase in government spending, corroborating the insights from the New Keynesian models and the results from Ramey (2009, 2012) for the US. This result might be related to the fact that, much like Ramey's work, this paper's results are in large part driven by military spending shocks. But using the civil spending only shock also produces negative effects on consumption — significant at short horizons — which might indicate broader crowding-out effects of government spending.

Interest rates are a policy decision, and can respond differently to similar fiscal policy shocks, depending on the monetary policy setting. Accordingly, I find no evidence that monetary policy systematically accommodates or counteracts the fiscal shocks. Employment rate effects are strongly positive and statistically significant, in line with the literature, and are not only long-lasting but increasing over time. The estimates for the effect on inflation are positive and statistically significant — corroborating the findings of Ferrara et al. (2021), as well as the predictions of New Keynesian models.

3.5.3 State-dependent multipliers

I use the framework presented in equations (3.3)-(3.5) to estimate different fiscal multipliers across states of the economy, which can then be tested to ascertain whether they are statistically different from one another.

	Years after impact				
Multiplier	0	1	2	3	4
Actual employment rate					
High	0.45	0.67**	0.89***	1.05***	1.22***
Low	0.46***	0.54***	0.60***	0.60***	0.54***
Difference	-0.01	0.12	0.29	0.45	0.68
Unemployment rate relative to estimated NAIRU					
High	2.11***	3.22***	3.72***	3.54***	3.01***
Low	0.39***	0.51***	0.58***	0.62***	0.60***
Difference	1.72***	2.71***	3.14***	2.92**	2.41**
Output gap estimate					
Below -0.5%	1.34**	2.28**	2.57**	2.51**	2.09**
Above -0.5%	0.43***	0.53***	0.58***	0.59**	0.56**
Difference	0.90	1.75	1.99	1.92	1.53
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.5: Summary of multiplier estimates across states of slack. Significance calculations reflect the Anderson-Rubin confidence sets for single estimates and tests of restrictions on coefficients using HAR standard errors.

Table 3.5 summarises multiplier estimates across states of high and low slack using different measures. The output gap measure is based on a polynomial trend of potential output, for which I define high slack as an output gap of more than 0.5% below estimated potential GDP. The unemployment rate-based measure defines a state of high slack as one with an unemployment rate above the average of the whole period. And the NAIRU-based measure uses a Kálmán filter to estimate the non-accelerating inflation rate of unemployment, and defines high slack as being an unemployment rate more than 0.5 percentage points higher than the estimated NAIRU at any given point.

In all cases except the NAIRU-based measure, the difference between the multiplier estimates in states of high and low slack is not statistically significant at conventional levels. Results are robust to different choices of thresholds and to methods of estimating potential output.

If we define high slack as unemployment being more than 0.5 percentage points above the estimated NAIRU, the results suggest significantly larger multipliers at high slack (1.2 on impact and as high as 2.1 after two years) than in low slack states (0.2 on impact and no higher than 0.3 at any horizon). The significant difference between states is sensitive to thresholds, but is present above 0.4 percentage points.

These results are not particularly strong — they represent some evidence of higher multipliers in specific definitions of high slack states, but otherwise display no statistical evidence of state-dependent differences. In that sense, they are weaker than the results of Gordon and Krenn (2010) and Auerbach and Gorodnichenko (2012) — which find strong evidence large multipliers in times of recession — but stronger than those of Ramey and Zubairy (2018) and Ghassibe and Zanetti (2022), with the caveat that all those are US-focused.

	Years after impact				
Multiplier	0	1	2	3	4
			Debt-to-GDP ratio		
High	0.36***	0.39***	0.40***	0.36***	0.28***
Low	0.29	0.33	0.32	0.34	0.33
Difference	0.07	0.06	0.08	0.02	-0.05
			Openness to trade		
High	0.13	0.21	0.25	0.30	0.31
Low	0.46**	0.50***	0.50***	0.44**	0.32*
Difference	-0.33	-0.29	-0.26	-0.15	-0.01
			Exchange rate regime		
Flexible	0.28***	0.33***	0.36***	0.37***	0.35***
Fixed	-0.01	0.05	0.01	-0.04	-0.08
Difference	0.30	0.28	0.35	0.41	0.43
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.6: Summary of multiplier estimates across regimes. Significance calculations reflect the Anderson-Rubin confidence sets for single estimates and tests of restrictions on coefficients using HAR standard errors.

Table 3.6 looks at the suite of indicators of regimes for differences in multipliers used by Ilzetzki et al. (2013). Unlike that paper, I find no statistically significant difference between multipliers across each of the states. Unlike other states of interest (e.g. the zero-lower bound, which was only really hit in the UK after the 2007-08 crisis), the first-stage F-statistics do not indicate that the shocks are particularly underpowered in any of the states.

In all three cases (high/low debt, open/closed economy and flexible/fixed exchange rates), multiplier estimates for one of the main states are not statistically different from zero. That is the case for low debt states, higher openness states and fixed exchange rates. High debt (over 60% of GDP), lower openness (below the historic average of 36% of GDP) and flexible exchange rates, on the other hand, are all associated with positive estimates that are statistically different from zero. The test of difference in coefficients however does not provide enough evidence to reject the null hypothesis that the two parameters are the same — weaker results than those found by Ilzetzki et al., although that is a cross-country study which is likely to have more variation in regimes than one country across time.

3.6 Conclusion

Fiscal multipliers are notoriously tricky to estimate, and one of the main issues is how to identify exogenous shocks that allow the establishment of causality. Narrative approaches from seminal papers such as Ramey and Shapiro (1998), Gordon and Krenn (2010), Romer and Romer (2010) and Ramey (2011b) have increased our understanding of fiscal multipliers for US data, but data for other countries is costly to assemble and in many cases impractical. Cloyne (2013) used an archival source to conduct a narrative study focused on UK tax policy, but no analogue exists on the spending side. This paper seeks to provide that spending dataset and to ascertain what can be learned from a long time series approach to estimating spending multipliers for the UK.

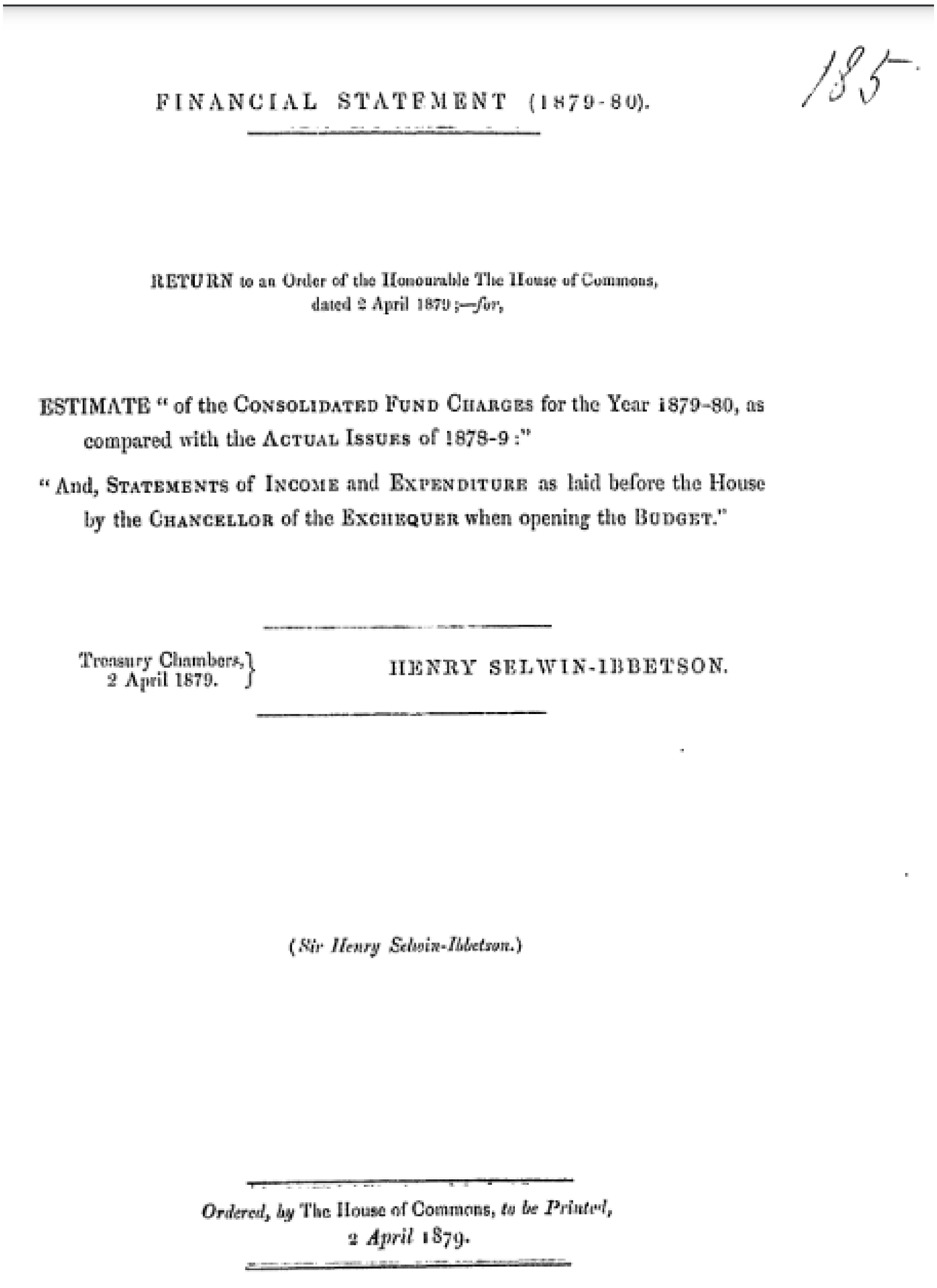
In this paper, I have used a novel dataset which I have compiled from archival research in the UK Parliament, and which allows me to create a 140-year series of in-year, unanticipated discretionary spending shocks, which exploit the UK's idiosyncratic budget process and therefore are unlikely to be anticipated. This is further supported by running Mincer-Zarnowitz regressions, which fail to reject the hypothesis that the shocks are unanticipated.

Using a local projections method, I find the multiplier to be statistically significantly greater than zero at all horizons, with an estimate of 0.44 on impact and 0.47 in the long run. This is not dissimilar to results for the US, although higher than the majority of the estimates for the UK in the literature. I also find some evidence of a larger multiplier associated with civil spending shocks than with military spending ones at short horizons. I also find that the effect of a positive spending shock on other macroeconomic variables broadly reflects results from New Keynesian models (lower private consumption, higher employment and higher inflation, although I find no significant effect on the policy interest rate) — which reflect some but not all the empirical literature. The fall in consumption in particular corroborates other studies which use large military spending shocks as their identification strategy — see Ramey (2009, 2011b).

In terms of state-dependent effects, I only find strong statistical evidence of higher multipliers in states with unemployment above the natural rate by large amounts — a finding that is robust for a number of threhsolds. I find no evidence of other state-dependent differences, be they other measures of slack or broader economic conditions (high/low debt, open/closed economy or flexible/fixed exchange rates). The results for states of slack are less categorical than those found by Gordon and Krenn (2010) and Auerbach and Gorodnichenko (2012), but somewhat stronger those of Ramey and Zubairy (2018) and Ghassibe and Zanetti (2022) — although all those studies are US-focused. The results for the broader suite of economic regimes are weaker than those found by Ilzetzki et al. (2013), although that study uses cross-country data, and therefore may include more variation than available in following one country across time.

Appendices

3.A Examples of documents used for archival research



185

FINANCIAL STATEMENT (1879-80).

RETURN to an Order of the Honourable The House of Commons, dated 2 April 1879;—*for*,

ESTIMATE "of the CONSOLIDATED FUND CHARGES for the Year 1879-80, as compared with the ACTUAL ISSUES of 1878-9:"

"And, STATEMENTS of INCOME and EXPENDITURE as laid before the House by the CHANCELLOR of the EXCHEQUER when opening the BUDGET."

Treasury Chambers, 2 April 1879. HENRY SELWIN-IBBETSON.

(*Sir Henry Selwin-Ibbetson.*)

Ordered, by The House of Commons, *to be Printed*, 2 *April* 1879.

Figure 3.7: Cover of the Financial Statement 1879-80, the first used for the estimation process. This is illustrative of all covers running until Budget 1997. Until then, the statement of income and expenditure for the Government was always signed and presented to the House of Commons by the Financial Secretary to the Treasury, rather than the Chancellor of the Exchequer, although of course all estimates were signed off by the Chancellor.

EXPENDITURE.

ESTIMATE for 1879–80, coupled with ACTUAL EXCHEQUER ISSUES in 1878–9.

	Estimate for 1879–80.	Actual Exchequer Issues in 1878–9.
	£.	£.
Permanent Charge of Debt	28,000,000	28,000,000
Interest on Local Loans	440,000	314,932
Interest on Supply, Exchequer Bonds and Loan to India	220,000	129,375
Charge of Suez Loan	200,000	199,876
Other Consolidated Fund Charges	1,760,000	1,624,422
	30,620,000	30,268,605
Army	15,645,700	17,653,473
Home Charges of Forces in India	1,100,000	1,080,000
Navy	10,586,894	11,962,316
Vote of Credit (South African War)	- -	1,500,000
Abyssinian Expedition	- -	17,865
Civil Services	15,084,851	14,974,760
Customs and Inland Revenue	2,865,383	2,791,169
Post Office	3;368,825	3,273,000
Telegraph Service	1,115,195	1,109,000
Packet Service	766,725	777,600
TOTAL £.	81,153,573	85,407,788

NET DECREASE of Estimated Expenditure in 1879–80 - - - £. 4,254,215.

Figure 3.8: Table from the Financial Statement 1879-80 detailing the forecast at the beginning of the year for the 1879-80 financial year, as well as actual spending (labelled 'actual Exchequer issues') for the previous financial year of 1878-79. This is the type of table that is present in Financial Statements all the way until the introduction of spending controls in 1988, under the original 'planning total' guise. These are the longest standing breakdowns, and therefore the ones I use for military and civil spending purposes. The debt charge and any subsequent forms of interest payments are not used in the calculations for the shocks, which instead focus only on discretionary spending.

Service.	Estimate, 1914–15 (November 1914).	Estimate, 1915–16.	Estimate for 1915–16, more (+) or less (—) than 1914–15.
I. Consolidated Fund Services.			
National Debt Services :	£	£	£
Inside the Fixed Debt Charge :—			
Interest and Management	17,580,000	24,500,000	+ 3,750,000
Repayment of Capital	3,170,000		
	20,750,000	24,500.000	+ 3,750,000
Outside the Fixed Debt charge :— Interest and Management	3,443,000	‡ 15,750,000	+12,307,000
	24.193,000	40,250,000	+16,057,000
Road Improvement Fund	1,545,000	1,431,000	— 114,000
Payments to Local Taxation Accounts, &c.	9,885,000	9,406,000	— 479,000
Other Consolidated Fund Services	1,706,000	1,697,000	— 9,000
Total Consolidated Fund Services	37,329,000	52,784,000	+15,455,000
II. Supply Services.			
Army (including Ordnance Factories)	28,885,000	§ *15,000*	*—28,870,000*
Navy	51,550,000	§ *17,000*	*—51,533,000*
Civil Services*	58,885,000	59,018,000	+ 133,000
Customs and Excise, and Inland Revenue Departments	4,741.000	4,788,000	+ 47,000
Post Office Services	26,227,000	26,836,000	+ 609,000
Total Supply Services	170,288,000	*90.674,000*	*—79,614,000*
Votes of Credit	325,000,000	*250,000,000*	*—75,000,000*
Total	532,617,000	393,458,000	—139,159,000

Figure 3.9: Table from the Financial Statement 1915-16, during a period which saw some of the largest in-year policy changes due to the First World War. There are two things of particular interest. One is the large amount of the votes of credit, which allowed the government to plan in additional spending as it came without needing to vote it through the estimates process and apportioning it to different funding streams. This was very common in UK history in times of war, and gave the government flexibility in financing those wars. The second interesting point is the forecast for a much reduced cost of the war in 1915-16 compared with 1914-15. In fact, 1915-16 would be the second largest in-year increase in spending relative to forecast, totalling over 30% of GDP, surpassed only by the 46% of GDP increase the following year.

	1915-16.			
	Total Expenditure provided for in the Budget as revised in September 1915.	Additional Expenditure for which Supplementary Estimates were presented.	Total Estimated Expenditure.	Amount issued to meet Total Expenditure
I. Consolidated Fund Services.	£	£	£	£
National Debt Services : Inside the Fixed Debt Charge - -	22,055,000	—	22,055,000	20,338,000
Outside the Fixed Debt Charge - -	45,030,000	—	45,030,000	39,911,000
	67,085,000	—	67,085,000	60,249,000
Road Improvement Fund - - - -	525,000	—	525,000	694,000
Payments to Local Taxation Accounts, &c.	9,600,000	—	9,600,000	9,757,000
Other Consolidated Fund Services - -	1,800,000	—	1,800,000	2,788,000
Total Consolidated Fund Services -	79,010,000	—	79,010,000	73,488,000
II. Supply Services.				
Army (including Ordnance Factories) -	16,000	—	16,000	15,000
Navy - - - - - - - -	17,000	—	17,000	7,000
Ministry of Munitions (including Ordnance Factories).	—	2,000	2,000	2,000
Civil Services - - - - - -	59,039,000	196,000	59,235,000	54,718,000
Customs and Excise and Inland Revenue -	4,788,000	—	4,788,000	4,603,000
Post Office Services - - - - -	26,836,000	—	26,836,000	26,673,000
Total Supply Services - - -	90,696,000	198,000	90,894,000	86,018,000
Votes of Credit - - - - - -	1,420,000,000	—	1,420,000,000	1,399,652,000
GRAND TOTAL - - -	1,589,706,000	198,000	1,589,904,000	1,559,158,000

Figure 3.10: Table from the Financial Statement 1916-17, showing the actual expenditure in 1915-16. The duration of the war and its toll on the public finances was such that a revised Budget had to be issued in September 1915. Votes of credit became almost the whole of the government's financing in terms of parliamentary procedure, and I have apportioned them fully to military expenditure as they were war related. When compared with the previous table, it becomes clear that government spending was over four times the allocated amounts at the beginning of the financial year — a scale of intra-year shock that no other war has come close to.

1948–49

TABLE I.—Comparison of Exchequer Issues with Estimated Expenditure

1947–48 Exchequer Issues	—	1948–49 Total Expenditure provided for in the Budget	1948–49 Supplementary and Excess Votes subsequently granted	1948–49 Total Estimated Expenditure	1948–49 Exchequer Issues
	Ordinary Expenditure				
£000	CONSOLIDATED FUND SERVICES	£000	£000	£000	£000
502,626	Interest and Management of National Debt	500,000	—	500,000	477,176*†
24,314	Payments to Northern Ireland Exchequer	26,000	—	26,000	32,294‡
7,291	Other Consolidated Fund Services	8,000	—	8,000	9,501
534,231	Total	534,000	—	534,000	518,971
	SUPPLY SERVICES				
383,600	Army Votes	305,000	55,000	360,000	346,700
194,300	Navy Votes	153,000	15,500	168,500	162,700
181,900	Air Votes	173,000	16,400	189,400	186,900
94,050	Ministry of Supply (Defence)	61,000	*2,000*	59,000	56,300
331	Ministry of Defence ...	632	—	632	550
854,181	Total Defence	692,632	84,900	777,532	753,150
1,768,729	Civil Votes (excluding Ministry of Supply (Defence))	1,708,613	224,309	1,932,922	1,844,815
29,963	Customs and Excise, Inland Revenue and balance of Post Office Votes ...	40,434	150	40,584	35,846
2,652,873	Total Supply Services ...	2,441,679	309,359	2,751,038	2,633,811
3,187,104	**Total Ordinary Expenditure**	**2,975,679**	**309,359**	**3,285,038**	**3,152,782**
22,374	**Sinking Funds ... ...**	—	—	—	**22,824†**
3,209,478		**2,975,679**	**309,359**	**3,285,038**	**3,175,606**
	Self-Balancing Expenditure				
143,300	Post Office Expenditure corresponding to Revenue ...	150,200	—	150,200	152,700
23,183	Excess Profits Tax, Post-war refunds (part deducted for tax)	16,000	—	16,000	8,751
166,483	**Total Self-Balancing Expenditure ... ...**	**166,200**	—	**166,200**	**161,451**

Figure 3.11: Table from the Financial Statement 1949-50 detailing comparisons between initial estimates and actual spending for 1948-49. That year saw a large shock in civil spending due in no small part to overspending on the newly created health service.

3.B List of variables used

Variable	ONS code 1946-2018	Thomas and Dimsdale (2017) reference (Sheet.Column) 1878-1945	Notes
Population	UKPOP	A18.C until 1920 A18.D subsequently	Spliced series. Follows political boundaries
Nominal GDP	YBHA	A9.P	Spliced series
Real GDP	ABMI	A8.U	Monetary series calculated from A8.U index for splicing
GDP deflator	n/a	n/a	Calculated from nominal and real GDP
Employment	MGRZ	A50.C	Spliced series. Simplified assumption of employment divided by population to obtain employment rate
Official Bank Rate	n/a	n/a	Obtained from Bank of England (official history since 1694)
Unemployment	MGSC	A50.G	Used to calculate unemployment rate
Consumption	ABPF + ABNU	A12.F	Households + non-profits serving households. Spliced series
National debt	HF6W	A29.AJ	Spliced series
Central government managed expenditure	ECOD	A27.C	Spliced series
Exports	IKBL	A35.G	Spliced series
Imports	IKBK	A35.M	Spliced series

Table 3.7: List of variables used in the estimation process.

3.C Mincer-Zarnowitz regression results

Table 3.8 shows the regression coefficients estimated using the Mincer-Zarnowitz regressions, as well as the testing of the joint hypothesis that the coefficient of the residual of the shock is 1 and the constant is 0. In none of the three cases is this rejected, as discussed in the chapter, which lends further confidence to the shocks being unanticipated. y is output per capita, g is government spending per capita, r is the policy interest rate (Bank Rate and its predecessors), π is the inflation rate as measured by the GDP deflator, e is the employment rate, c is consumption per capita and d is the debt-to-GDP ratio. The residuals used in the bottom panel are obtained by subtracting the fitted values of the top panel from the observed values of the shocks.

	Full shock	Military spending shock	Civil spending shock
$\ln y_{t-1}$	-0.6021***	-0.5988***	-0.0033
	(0.0835)	(0.0869)	(0.0129)
$\ln g_{t-1}$	0.1495***	0.1502***	-0.0007
	(0.0199)	(0.0208)	(0.0031)
r_{t-1}	0.8212**	0.6155	0.2057***
	(0.3981)	(0.3987)	(0.0635)
π_{t-1}	-0.1351	-0.2049	0.0699***
	(0.1263)	(0.1318)	(0.0195)
e_{t-1}	0.0592**	0.0633**	-0.0041
	(0.0258)	(0.0271)	(0.0043)
$\ln c_{t-1}$	0.3281***	0.3241***	0.0039
	(0.0653)	(0.0680)	(0.0101)
d_{t-1}	-0.0947***	-0.0966***	0.0019
	(0.0135)	(0.0141)	(0.0021)
Constant	134.8807***	133.2610***	1.6197
	(19.5060)	(20.3385)	(3.0930)
N	140	140	140
F-stat	10.34	9.05	5.73
Residual of shock	1.0000***	1.0000***	1.0000***
	(0.1017)	(0.1007)	(0.0642)
Constant	1.1289	0.9872	0.1416*
	(0.7433)	(0.7399)	(0.0787)
N	140	140	140
F-stat	95.33	97.30	239.05
Joint hypothesis test	2.31	1.78	3.23
p-value	0.3156	0.4105	0.1984
Significance: *** - 1%; ** - 5%; * - 10%			

Table 3.8: Mincer-Zarnowitz regression outputs and testing of joint hypothesis that the coefficient on the residual of the shock is 1 and the constant is 0. Numbers in brackets are HAR standard errors, calculated using the QS kernel.

3.D Regression results for the linear case

Table 3.9 shows the results of the cointegration tests between $\sum_h \ln g_{t+j}$ and the relevant left-hand side variables at different horizons. As the results show, there is strong evidence of a common stochastic trend between output and government spending, as well as between consumption and government spending, such that we can confidently reject the null hypothesis that a linear combination of them is I(1) using the first step in the Engle-Granger procedure.

	Horizon h				
Cointegration with $\sum_h \ln g_{t+j}$	0	1	2	3	4
	$\sum_h \ln y_{t+j}$				
Test statistic	-3.121	-3.603	-3.971	-4.152	-4.202
p-value	0.025	0.006	0.002	<0.001	<0.001
	$\sum_h \ln c_{t+j}$				
Test statistic	-3.523	-4.064	-4.461	-4.552	4.491
p-value	0.007	0.001	<0.001	<0.001	<0.001
	$\sum_h r_{t+j}$				
Test statistic	-11.077	-12.251	-11.815	-7.959	-6.046
p-value	<0.001	<0.001	<0.001	<0.001	<0.001
	$\sum_h \pi_{t+j}$				
Test statistic	-5.104	-5.652	-5.925	-5.559	-5.374
p-value	<0.001	<0.001	<0.001	<0.001	<0.001
	$\sum_h e_{t+j}$				
Test statistic	-3.708	-4.537	-4.858	-4.988	-4.956
p-value	0.004	<0.001	<0.001	<0.001	<0.001

Table 3.9: Results from the first step of the Engle-Grange procedure, the augmented Dickey-Fuller test on the residuals of the two variables. Null hypothesis is the presence of an I(1) process, and therefore rejection means evidence of a joint I(0) process — in other words, cointegration.

The specification in equation (3.1) essentially amounts to an instrumental variables-augmented implementation of an autoregressive distributive lag (ARDL) model, a procedure that is robust to cointegrated variables by including contemporaneous variables on the right-hand side (Pesaran and Shin, 1995). I choose this implementation rather than an error correction model (ECM) because my interest in less in the short-run dynamics of the variables — something an ECM is more suited to — and more in what the cumulative and long-run effect is on output and consumption. Implementing a deterministic trend, an error correction term or running a canonical ARDL model all yield very similar results to what is essentially an analogous implementation to Blanchard and Perotti's (2002) stochastic trend model, who also found little difference between implementing and not implementing the error correction term. Running the PSS bounds test (Pesaran, Shin and Smith, 2001) after an ARDL model strongly supports the use of contemporaneous and lagged levels in the regression.

Table 3.10 summarises the regression results for the linear case when using the full shock as instrument for $\sum_h \ln g_{t+j}$. These are the outputs of the estimation of equation (3.1) across each horizon h, with the coefficient on $\sum_h \ln g_{t+j}$ being the output elasticity with respect to government spending of interest. Tables 3.11 to 3.14 show the effects on consumption, policy interest rates, the employment rate and inflation, whereas tables 3.15 and 3.16 show the results for the military and civil spending shocks, respectively.

$\sum_h \ln y_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.0717***	0.0816***	0.0860***	0.0844***	0.0762**
	(0.0247)	(0.0249)	(0.0257)	(0.0277)	(0.0307)
$\ln y_{t-1}$	1.1084***	2.2648***	3.4289***	4.6398***	5.9678***
	(0.0461)	(0.1369)	(0.2668)	(0.4225)	(0.5910)
$\ln g_{t-1}$	-0.0816***	-0.1882***	-0.2976***	-0.4025***	-0.5034***
	(0.0283)	(0.0595)	(0.0948)	(0.1353)	(0.1792)
r_{t-1}	-0.3592*	-1.0817**	-1.6637*	-2.1073	-2.6349
	(0.2034)	(0.5389)	(0.9288)	(1.2948)	(1.6062)
π_{t-1}	-0.0533	-0.0850	-0.0951	-0.0910	-0.0418
	(0.0620)	(0.1695)	(0.3078)	(0.4629)	(0.6297)
e_{t-1}	-0.0443***	-0.1349***	-0.2628***	-0.4118***	-0.5706***
	(0.0134)	(0.0410)	(0.0820)	(0.1329)	(0.1899)
$\ln c_{t-1}$	-0.0794**	-0.1821*	-0.2759	-0.3851	-0.5467
	(0.0352)	(0.1053)	(0.2067)	(0.3308)	(0.4687)
d_{t-1}	0.0129*	0.0411*	0.0823**	0.1394**	0.2203**
	(0.0074)	(0.0214)	(0.0411)	(0.0646)	(0.0904)
Constant	-2.3889	-4.0403	-3.9737	-15.8448	-60.7634
	(10.2600)	(30.5216)	(60.1374)	(96.7004)	(138.9287)
N	140	139	138	137	136
Kleibergen-Paap F-stat	47.844	38.311	30.063	22.151	16.343
AIC	601.044	858.973	1009.820	1108.692	1178.854
SBIC	627.519	885.384	1036.166	1134.972	1205.068
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.10: Regression estimates for output using the full shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h \ln c_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	-0.1127***	-0.1401***	-0.1505***	-0.1530***	-0.1580***
	(0.0256)	(0.0273)	(0.0289)	(0.0305)	(0.0358)
$\ln y_{t-1}$	0.1837***	0.5705***	1.1380***	1.8626***	2.4797***
	(0.0455)	(0.1411)	(0.2720)	(0.4334)	(0.6045)
$\ln g_{t-1}$	0.0976***	0.2199***	0.3141***	0.3766***	0.4318**
	(0.0299)	(0.0663)	(0.1061)	(0.1446)	(0.1987)
r_{t-1}	-0.1392	-0.4059	-0.6970	-1.0653	-1.5230
	(0.1975)	(0.5190)	(0.8842)	(1.3487)	(1.5984)
π_{t-1}	-0.1594**	-0.3871**	-0.6076*	-0.8227	-1.0463
	(0.0647)	(0.1883)	(0.3541)	(0.5138)	(0.6297)
e_{t-1}	-0.0313**	-0.0858**	-0.1693**	-0.2917**	-0.4494**
	(0.0131)	(0.0411)	(0.0840)	(0.1423)	(0.1984)
$\ln c_{t-1}$	0.8483***	1.5481***	2.1303***	2.6010***	2.9608***
	(0.0347)	(0.1086)	(0.2124)	(0.3415)	(0.4841)
d_{t-1}	0.0068	0.0250	0.0611	0.1123*	0.1788*
	(0.0076)	(0.0227)	(0.0437)	(0.0678)	(0.0970)
Constant	-8.9235	-40.1515	-96.5602	-165.8469	-251.9462*
	(9.9859)	(30.3955)	(60.4976)	(102.3041)	(144.7959)
N	140	139	138	137	136
Kleibergen-Paap F-stat	47.780	38.260	30.180	24.686	16.453
AIC	609.023	880.428	1032.279	1130.411	1207.363
SBIC	635.498	906.838	1058.624	1156.691	1233.577
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.11: Regression estimates for consumption using the full shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h r_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.0132	0.0105	0.0062	0.0056	0.0074**
	(0.0109)	(0.0079)	(0.0055)	(0.0041)	(0.0036)
$\ln y_{t-1}$	0.0341***	0.0857***	0.1169***	0.1256***	0.1231**
	(0.0120)	(0.0275)	(0.0383)	(0.0467)	(0.0572)
$\ln g_{t-1}$	-0.0175	-0.0283*	-0.0256	-0.0280*	-0.0353*
	(0.0114)	(0.0163)	(0.0167)	(0.0168)	(0.0185)
r_{t-1}	0.2701**	-0.0768	-0.3763*	-0.5531***	-0.4595**
	(0.1176)	(0.2067)	(0.2159)	(0.1900)	(0.1833)
π_{t-1}	0.0031	-0.0080	-0.0116	0.0198	0.0244
	(0.0245)	(0.0492)	(0.0608)	(0.0656)	(0.0714)
e_{t-1}	-0.0049	-0.0146*	-0.0238*	-0.0303*	-0.0355*
	(0.0036)	(0.0088)	(0.0128)	(0.0162)	(0.0201)
$\ln c_{t-1}$	-0.0278***	-0.0767***	-0.1109***	-0.1245***	-0.1359***
	(0.0097)	(0.0220)	(0.0306)	(0.0374)	(0.0458)
d_{t-1}	0.0031	0.0056	0.0057	0.0060	0.0043
	(0.0021)	(0.0045)	(0.0062)	(0.0075)	(0.0091)
Constant	-1.8729	-0.0589	5.1406	10.4163	19.5396
	(3.1549)	(7.1885)	(10.1351)	(12.4143)	(15.0495)
N	140	139	138	137	136
Kleibergen-Paap F-stat	47.780	38.260	29.983	22.070	16.270
AIC	401.059	538.740	563.543	564.907	578.276
SBIC	427.534	565.150	589.889	591.187	604.490
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.12: Regression estimates for the policy interest rate using the full shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h e_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.2592***	0.2764***	0.3017***	0.3543***	0.4014***
	(0.0838)	(0.0844)	(0.0849)	(0.0893)	(0.0960)
$\ln y_{t-1}$	0.1355	0.2499	0.1117	-0.3012	-0.8926
	(0.1210)	(0.3453)	(0.6462)	(1.0341)	(1.4906)
$\ln g_{t-1}$	-02767***	-0.5517***	-0.8255***	-1.1694***	-1.5092***
	(0.0911)	(0.1819)	(0.2686)	(0.3637)	(0.4771)
r_{t-1}	-0.1259	-1.5701	-4.0579	-6.6950*	-9.331*
	(0.6868)	(1.7552)	(2.9036)	(3.9968)	(4.9784)
π_{t-1}	-0.0875	-0.3892	-0.7076	-1.1236	-1.7435
	(0.1989)	(0.5436)	(0.9621)	(1.4230)	(1.9188)
e_{t-1}	0.9143***	1.6968***	2.3599***	2.9075***	3.3901***
	(0.0380)	(0.1119)	(0.2141)	(0.3358)	(0.4944)
$\ln c_{t-1}$	-0.0810	-0.1696	-0.0671	0.1955	0.5123
	(0.0956)	(0.2748)	(0.5160)	(0.8257)	(1.1871)
d_{t-1}	0.0279	0.0613	0.0945	0.1210	0.1269
	(0.0205)	(0.0567)	(0.1045)	(0.1650)	(0.2373)
Constant	-4.1587	48.6952	173.512	392.6746	695.1702
	(29.4783)	(85.7425)	(165.3337)	(267.9368)	(393.7849)
N	140	139	138	137	136
Kleibergen-Paap F-stat	47.780	38.260	29.983	22.257	16.702
AIC	953.810	1182.876	1319.055	1409.737	1474.250
SBIC	962.285	1209.286	1345.400	1436.017	1500.464
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.13: Regression estimates for the employment rate using the full shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h \pi_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.1485***	0.1386***	0.1324***	0.1477***	0.1608***
	(0.0356)	(0.0317)	(0.0286)	(0.0297)	(0.0328)
$\ln y_{t-1}$	0.0720	0.1881	0.2956	0.3944	0.5104
	(0.0502)	(0.1405)	(0.2386)	(0.3742)	(0.5363)
$\ln g_{t-1}$	-0.1346***	-0.2150***	-0.2677***	-0.3739***	-0.4863***
	(0.0386)	(0.0705)	(0.0954)	(0.1307)	(0.1744)
r_{t-1}	0.4347	0.7558	0.6836	0.5646	0.4726
	(0.3095)	(0.7036)	(0.9696)	(1.2336)	(1.4814)
π_{t-1}	0.7156***	1.0108***	1.1238***	1.1720**	1.1514*
	(0.0832)	(0.2081)	(0.3347)	(0.4883)	(0.6638)
e_{t-1}	-0.0419***	-0.1144***	-0.1897**	-0.2679**	-0.3320*
	(0.0154)	(0.0440)	(0.0778)	(0.1232)	(0.1778)
$\ln c_{t-1}$	-0.0873**	-0.2766**	-0.4932***	-0.7385**	-1.0329**
	(0.0397)	(0.1103)	(0.1884)	(0.29734)	(0.4303)
d_{t-1}	0.0021	-0.0128	-0.0396	-0.0709	-0.1097
	(0.0085)	(0.0227)	(0.03824)	(0.0592)	(0.0848)
Constant	18.3787	75.1869**	151.7024***	246.6820***	352.0389***
	(12.2049)	(33.3398)	(58.0202)	(91.3697)	(132.7762)
N	140	139	138	137	136
Kleibergen-Paap F-stat	47.862	39.408	29.983	22.070	16.270
AIC	721.759	928.500	1026.705	1109.318	1178.127
SBIC	748.234	954.911	1053.050	1135.597	1204.341
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.14: Regression estimates for the inflation rate using the full shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h \ln y_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.0816***	0.0890***	0.0909***	0.0869***	0.0765**
	(0.0240)	(0.0246)	(0.0256)	(0.0276)	(0.0307)
$\ln y_{t-1}$	1.1120***	2.2695***	3.4312***	4.6391***	5.9672***
	(0.0454)	(0.1363)	(0.2664)	(0.4225)	(0.5916)
$\ln g_{t-1}$	-0.0923***	-0.2032***	-0.3112***	-0.4106***	-0.5045***
	(0.0275)	(0.0588)	(0.0945)	(0.1352)	(0.1792)
r_{t-1}	-0.3723*	-1.1085**	-1.6936*	-2.1279	-2.6384
	(0.2045)	(0.5426)	(0.9338)	(1.2997)	(1.6121)
π_{t-1}	-0.0473	-0.0759	-0.0877	-0.0879	-0.0416
	(0.0614)	(0.1691)	(0.3078)	(0.4629)	(0.6296)
e_{t-1}	-0.0455***	-0.1375***	-0.2658***	-0.4140***	-0.5710***
	(0.0133)	(0.0409)	(0.0820)	(0.1328)	(0.1896)
$\ln c_{t-1}$	-0.0813**	-0.1856*	-0.2759	-0.3876	-0.5471
	(0.0348)	(0.1048)	(0.2067)	(0.3303)	(0.4680)
d_{t-1}	0.0141*	0.0426*	0.0833**	0.1395**	0.2202**
	(0.0073)	(0.0214)	(0.0411)	(0.0646)	(0.0904)
Constant	-3.0243	-4.0614	-2.5397	-13.6872	-60.1836
	(10.1632)	(30.4694)	(60.1489)	(96.6211)	(138.5870)
N	140	139	138	137	136
Kleibergen-Paap F-stat	55.287	41.498	31.214	22.531	16.553
AIC	600.408	859.012	1010.012	1108.659	1178.794
SBIC	626.883	885.422	1036.357	1134.939	1205.008
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.15: Regression estimates for output using the military spending shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

	Horizon h				
$\sum_h \ln y_{t+j}$	0	1	2	3	4
$\sum_h \ln g_{t+j}$	0.2474*	0.2343*	0.1948	0.1390	0.0832
	(0.1015)	(0.1163)	(0.1116)	(0.0965)	(0.0891)
$\ln y_{t-1}$	1.1719***	2.3607***	3.4789***	4.6233***	5.9563***
	(0.0791)	(0.2151)	(0.3469)	(0.4530)	(0.6173)
$\ln g_{t-1}$	-0.2716**	-0.4996**	-0.5973*	-0.5765*	-0.5270
	(0.1106)	(0.2417)	(0.3192)	(0.3290)	(0.3347)
r_{t-1}	-0.5927*	-1.6359*	-2.3231*	-2.5512	-2.7044
	(0.3422)	(0.9135)	(1.3863)	(1.6480)	(1.9130)
π_{t-1}	0.0537	0.10354	0.0694	-0.02406	-0.0392
	(0.1148)	(0.3009)	(0.4591)	(0.5408)	(0.6300)
e_{t-1}	-0.0662***	-0.1880**	-0.3295***	-0.4581***	-0.5776***
	(0.0247)	(0.0736)	(0.1238)	(0.1621)	(0.2018)
$\ln c_{t-1}$	-0.1125*	-0.2553	-0.3589	-0.4387	-0.5547
	(0.0577)	(0.1660)	(0.2761)	(0.3560)	(0.4656)
d_{t-1}	0.0330**	0.0733*	0.1040*	0.1408**	0.2181**
	(0.0158)	(0.0398)	(0.0584)	(0.0703)	(0.0940)
Constant	-13.6892	-4.4765	27.6496	30.6183	-49.2032
	(17.2534)	(46.5171)	(84.1165)	(128.177)	(190.8629)
N	140	139	138	137	136
Kleibergen-Paap F-stat	4.186	2.649	1.984	1.682	1.531
AIC	720.713	975.062	1086.889	1132.167	1178.106
SBIC	747.188	1001.472	1113.234	1158.447	1204.320
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.16: Regression estimates for output using the civil spending shock at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

3.E Regression results with the full shock and different states of slack

Table 3.17 shows the regression results for high and low slack states as defined by whether the observed unemployment rate is more or less than 0.5 percentages above the estimated NAIRU (non-accelerating inflation rate of unemployment, also called the natural rate of unemployment) — the latter being estimated using a Kálmán filter. This is the only state-dependent estimate for which I estimate significant differences in the coefficients on the cumulative change in government spending. Details of the remaining regression outputs are available on request.

	Horizon h				
$\sum_h \ln y_{t+j}$	0	1	2	3	4
$I_{t-1}\sum_h \ln g_{t+j}$	0.3373***	0.5150***	0.5953***	0.5652***	0.4819***
	(0.0919)	(0.1383)	(0.1662)	(0.1643)	(0.1492)
$(1-I_{t-1})\sum_h \ln g_{t+j}$	0.0654**	0.0842***	0.0957***	0.1009***	0.0981***
	(0.0246)	(0.0266)	(0.0273)	(0.0283)	(0.0293)
I_{t-1}	68.5526**	347.5754***	753.4731***	1106.8980***	1352.4500***
	(33.3170)	(123.8984)	(250.5096)	(364.6650)	(459.7882)
$I_{t-1}\ln y_{t-1}$	0.95568***	1.3834***	1.4164	1.6836	2.4599
	(0.1261)	(0.4448)	(0.9140)	(1.3485)	(1.7132)
$(1-I_{t-1})\ln y_{t-1}$	1.0653***	2.1915***	3.3467***	4.5710***	5.8900***
	(0.0444)	(0.1293)	(0.2444)	(0.3787)	(0.5268)
$I_{t-1}\ln g_{t-1}$	-0.2661***	-0.6989***	-1.0889***	-1.2665***	-1.2426***
	(0.0672)	(0.1729)	(0.2847)	(0.3530)	(0.3768)
$(1-I_{t-1})\ln g_{t-1}$	-0.0732***	-0.1925***	-0.3283***	-0.4705***	-0.6089***
	(0.0298)	(0.0654)	(0.0990)	(0.1317)	(0.1638)
$I_{t-1}r_{t-1}$	-0.0360	-0.8635	-2.0477	-3.3558	-3.5415
	(0.4226)	(1.2044)	(2.1218)	(2.9343)	(3.3660)
$(1-I_{t-1})r_{t-1}$	-0.8293***	-2.2190***	-3.2086***	-3.7167**	-4.0676**
	(0.2270)	(0.6495)	(1.1478)	(1.5771)	(1.8715)
$I_{t-1}\pi_{t-1}$	-0.1805**	-0.3976*	-0.5769	-0.7368	-1.0098
	(0.0787)	(0.2104)	(0.3819)	(0.5457)	(0.6787)
$(1-I_{t-1})\pi_{t-1}$	0.0560	0.1565	0.2246	0.2316	0.2623
	(0.0623)	(0.1807)	(0.3520)	(0.5258)	(0.6744)
$I_{t-1}e_{t-1}$	0.0432	0.1770*	0.3202*	0.3854	0.3145
	(0.0280)	(0.0936)	(0.1849)	(0.2736)	(0.3384)
$(1-I_{t-1})e_{t-1}$	-0.0593***	-0.1725***	-0.3226***	-0.4887***	-0.6537***
	(0.0130)	(0.0402)	(0.0796)	(0.1276)	(0.1771)
$I_{t-1}\ln c_{t-1}$	-0.1151	-0.1108	0.0784	0.1961	0.1106
	(0.0730)	(0.2265)	(0.4561)	(0.6798)	(0.8825)
$(1-I_{t-1})\ln c_{t-1}$	-0.0355	-0.0951	-0.1602	-0.2650	-0.4132
	(0.0326)	(0.0973)	(0.1917)	(0.3050)	(0.4329)
$I_{t-1}d_{t-1}$	-0.0391**	-0.1766***	-0.3716***	-0.5250***	-0.6073**
	(0.0169)	(0.0632)	(0.1325)	(0.1988)	(0.2567)
$(1-I_{t-1})d_{t-1}$	0.0165**	0.0548***	0.1123***	0.1865***	0.2814***
	(0.0068)	(0.0195)	(0.0366)	(0.0566)	(0.0794)
Constant	3.7502	2.4678	-4.3250	-2.2284	-78.6029
	(9.6087)	(28.5691)	(55.9154)	(88.8015)	(124.8896)
Difference in coefficients	0.2719	0.4308	0.4995	0.4643	0.3838
Test statistic	8.67	9.74	9.09	8.07	6.70
p-value	0.003	0.002	0.003	0.005	0.010
N	140	139	138	137	136
Significance: *** - 1%; ** - 5%; * - 10%					

Table 3.17: Regression estimates for output using the full spending shock at each horizon h for different states of slack. I represents high slack (unemployment rate more than 0.5 percentage points higher than the estimated NAIRU) and $(1-I)$ represents low slack. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

3.F Shocks used in the first stage regression

Table 3.18 shows the raw values of the shocks used for the estimation process. These are in cash values (£ million) and in nominal terms, and so have been subsequently divided by GDP for normalisation.

Year	Full shock	Military	Civil	Year	Full shock	Military	Civil
1879-80	3.337	3.089	0.247	1949-50	52.395	-19.213	71.608
1880-81	1.103	0.727	0.376	1950-51	-68.696	17.564	-86.260
1881-82	0.046	-0.274	0.321	1951-52	-130.921	-161.762	30.841
1882-83	4.681	3.861	0.820	1952-53	68.612	26.520	42.092
1883-84	0.701	0.775	-0.074	1953-54	13.433	-132.262	145.695
1884-85	3.971	3.641	0.330	1954-55	-215.466	-118.644	-96.822
1885-86	-2.009	-1.999	-0.010	1955-56	-118.657	-89.300	-29.357
1886-87	0.200	0.459	-0.259	1956-57	64.614	26.400	38.214
1887-88	0.432	-0.378	0.811	1957-58	67.902	9.416	58.486
1888-89	-1.155	-0.855	-0.300	1958-59	57.398	49.200	8.198
1889-90	-0.082	0.182	-0.264	1959-60	11.933	-26.435	38.368
1890-91	0.525	0.170	0.355	1960-61	171.839	-21.830	193.669
1891-92	0.632	-0.351	0.983	1961-62	181.000	33.000	148.000
1892-93	-0.071	-0.027	-0.044	1962-63	84.000	45.000	39.000
1893-94	0.645	0.144	0.501	1963-64	-139.000	-46.000	-93.000
1894-95	0.200	0.004	0.196	1964-65	-69.000	-92.000	23.000
1895-96	2.001	1.499	0.502	1965-66	6.000	-64.000	70.000
1896-97	0.792	0.561	0.231	1966-67	85.000	-57.000	142.000
1897-98	1.576	0.201	1.375	1967-68	493.000	55.000	438.000
1898-99	1.281	1.069	0.212	1968-69	85.000	-39.000	124.000
1899-1900	22.725	22.388	0.337	1969-70	216.000	-62.000	278.000
1900-01	32.869	32.207	0.662	1970-71	517.000	200.000	317.000
1901-02	11.371	4.436	6.935	1971-72	929.000	8.000	921.000
1902-03	9.443	-0.310	9.753	1972-73	901.000	272.000	629.000
1903-04	3.477	3.196	0.281	1973-74	691.000	265.000	426.000
1904-05	-0.279	0.266	-0.545	1974-75	4,498.000	602.000	3,896.000
1905-06	-1.249	-1.052	-0.197	1975-76	5,343.000	744.000	4,599.000
1906-07	-2.698	-2.466	-0.232	1976-77	142.000	549.000	-407.000
1907-08	-0.886	-0.923	0.037	1977-78	-106.000	456.000	-562.000
1908-09	1.087	-0.750	1.837	1978-79	-415.000	533.000	-948.000
1909-10	0.374	0.465	-0.091	1979-80	942.000	545.000	397.000
1910-11	0.248	-0.529	0.777	1980-81	3,593.000	632.000	2,961.000
1911-12	-2.407	-1.576	-0.831	1981-82	2,000.000	500.000	1,500.000
1912-13	5.522	0.491	5.031	1982-83	-1,900.000	300.000	-2,200.000
1913-14	1.498	2.635	-1.137	1983-84	100.000	-300.000	400.000
1914-15	356.797	357.001	-0.204	1984-85	5,500.000	100.000	5,400.000
1915-16	1,001.701	1,006.186	-4.485	1985-86	-4,400.000	-100.000	-4,300.000
1916-17	1,672.150	1,673.665	-1.515	1986-87	-800.000	-300.000	-500.000
1917-18	427.725	427.800	-0.075	1987-88	-2,900.000	-200.000	-2,700.000
1918-19	-348.082	-352.000	3.918	1988-89	-5,100.000	-200.000	-4,900.000
1919-20	225.463	101.328	124.135	1989-90	3,300.000	500.000	2,800.000
1920-21	325.488	61.799	263.689	1990-91	1,600.000	900.000	700.000
1921-22	52.129	-18.025	70.154	1991-92	-3,600.000	100.000	-3,700.000
1922-23	-83.666	-27.079	-56.587	1992-93	-2,200.000	-300.000	-1,900.000
1923-24	-29.283	-16.211	-13.072	1993-94	200.000	-100.000	300.000
1924-25	-0.212	-0.426	0.214	1994-95	-4,200.000	-1,000.000	-3,200.000
1925-26	19.487	-1.136	20.623	1995-96	-1,400.000	-500.000	-900.000
1926-27	6.089	0.130	5.959	1996-97	-500.000	700.000	-1,200.000
1927-28	-3.847	2.325	-6.172	1997-98	-2,220.000	40.000	-2,260.000
1928-29	-2.581	-1.130	-1.451	1998-99	-220.000	310.000	-530.000
1929-30	8.496	0.390	8.106	1999-2000	300.000	500.000	-200.000
1930-31	11.941	0.435	11.506	2000-01	-100.000	800.000	-900.000
1931-32	0.184	-2.355	2.539	2001-02	-2,200.000	700.000	-2,900.000
1932-33	11.106	-1.334	12.440	2002-03	32,100.000	23,100.000	9,000.000
1933-34	-4.486	-1.074	-3.412	2003-04	1,000.000	1,800.000	-800.000
1934-35	10.236	0.159	10.077	2004-05	1,900.000	1,300.000	600.000
1935-36	17.610	12.699	4.911	2005-06	-300.000	500.000	-800.000
1936-37	4.633	27.821	-23.188	2006-07	-4,700.000	1,300.000	-6,000.000
1937-38	-22.333	-1.018	-21.315	2007-08	2,400.000	4,400.000	-2,000.000
1938-39	-6.583	18.941	-25.524	2008-09	100.000	5,000.000	-4,900.000
1939-40	383.013	401.963	-18.950	2009-10	100.000	500.000	-400.000
1940-41	1,217.323	1,220.001	-2.678	2010-11	-31,400.000	-12,500.000	-18,900.000
1941-42	566.681	585.000	-18.319	2011-12	-6,200.000	400.000	-6,600.000
1942-43	351.601	340.000	11.601	2012-13	-16,700.000	-3,000.000	-13,700.000
1943-44	43.409	50.000	-6.591	2013-14	-5,800.000	-1,500.000	-4,300.000
1944-45	124.219	125.000	-0.781	2014-15	-1,700.000	200.000	-1,900.000
1945-46	-82.229	-90.000	7.771	2015-16	-9,300.000	-200.000	-9,100.000
1946-47	13.555	-13.225	26.780	2016-17	-2,000.000	500.000	-2,500.000
1947-48	27.506	-45.150	72.656	2017-18	-1,000.000	0.000	-1,000.000
1948-49	192.132	60.518	131.614	2018-19	1,600.000	900.000	700.000

Table 3.18: Annual values of the in-year discretionary spending shocks — full, military only and civil only — used in the first stage of the regressions. Values are in £ million in current prices.

Chapter 4

The historical effect of UK fiscal policy on output: what can Parliamentary Archives data tell us?

Abstract

I use archival research data from the UK Parliament to construct a novel dataset with 140 years (1879-80 to 2018-19) of shocks to the discretionary fiscal stance to estimate what the effect of fiscal policy on output has been. The UK's fiscal policymaking process makes a compelling case for these being unanticipated shock within the year, and I use that to produce estimates using the shocks as instruments in a local projections framework. I find that a 1% of GDP increase in the fiscal balance is associated with a 0.24% fall in GDP in-year, which reaches 0.38% by year 4, despite private consumption increasing strongly in line with the expansionary fiscal contraction literature. I find evidence that fiscal policy is more powerful in times of high slack and that consumption increases more strongly with a fiscal tightening in times of fiscal distress, as well as weak evidence to support asymmetric effects between expansionary and contractionary fiscal policy.

4.1 Introduction

The canonical, Keynesian-inspired view of the main effect of fiscal policy on output is its effect on aggregate demand, as that determines output in the short run (Hemming et al., 2002). With the supply potential of the economy largely fixed in the short run, and prices rigid, a change in fiscal policy that reduces the fiscal balance — which I will call fiscal loosening, that is, a higher deficit (or "deficit spending", as it is commonly referred to) — will lead to higher output through higher disposable income, which translates into higher demand for goods and services in the private sector and therefore higher consumption. The opposite is true: a tightening of fiscal policy (that is, reducing the deficit) will reduce private demand for goods and services (including investment goods) through a reduction in disposable income, which will in turn reduce output. In one form or another, this has been the basis for macroeconomic thinking in most circles since the end of the Second World War — it was the basis of the Neo-Keynesian synthesis, and it is still embedded in today's

New Keynesian models which most official institutions around the world use.

But after the breakdown of the Neo-Keynesian synthesis in the 1970s (Mankiw, 1986), a competing view emerged. This was originally summarised by Fels and Frölich (1987), who described it as the "German view" — and has since become the basis for an area of work pioneered by Giavazzi and Pagano (1990) and Alesina and Perotti (1997). This "expansionary fiscal contraction" (EFC) hypothesis postulates that many fiscal contractions have led to increased growth in output and in particularly in some of their components, consumption and investment being the two main ones (Alesina et al., 2002). This appears to run counter to the results from most mainstream theoretical general equilibrium models, and much attention has been devoted to exploring the mechanism in operation for this to be the dominant effect. Giavazzi and Pagano themselves proposed a Ricardian equivalence argument in which reductions in expenditure increase the likelihood of future tax cuts, and therefore induce a wealth effect in agents — which in turn leads them to increase activity in the private sector that more than makes up for the gap in aggregate demand. Blanchard (1990) and Bertola and Drazen (1993) posit that a relatively small adjustment today to stabilise debt can be expansionary if it avoids the need for a large adjustment in the future. Additionally, Alesina et al. (1998) point to a situation in which a country whose rapidly increasing or very high level of debt can create a risk premium in the market for its sovereign bonds — and a credibility-enhancing fiscal consolidation can lead to a permanent reduction in real interest rates, which then increases consumption and investment.

Both views reflect real mechanisms operating in a change in the fiscal stance, and so which effect dominates and under what conditions becomes an empirical question. Since Giavazzi and Pagano's results were published in 1990, a growing literature has shown increases in private consumption in particular in response to fiscal tightenings, leaving authors like Barry and Devereux (1995, 2003) and Hogan (2004) to attempt to synthesise and reconcile those results with the prevailing macroeconomic paradigm. Further interest in the effect of fiscal consolidations developed in the 2010s, as Euro Area economies (Portugal, Spain, Italy, Ireland, Greece and Cyprus) suffered sovereign debt crises, while the UK also embarked in a large fiscal consolidation plan to "place [the UK's] fiscal credibility beyond doubt," in the words of then-Chancellor of the Exchequer George Osborne when presenting the June 2010 Emergency Budget to the House of Commons. Much has since been argued about the effects in both the Eurozone and the UK (De Grauwe and Ji, 2013; Riley and Chote, 2014; Wren-Lewis, 2015; Jordà and Taylor, 2016; Alesina et al., 2018, 2019, 2020; Afonso et al., 2022 being some of the most significant studies), without consensus on the effect that the programmes had on output growth.

My contribution to this renewed debate is to take a broader historical view and to use the richness of a dataset of fiscal policy changes going back to the 1870s to assess whether Keynesian or non-Keynesian effects have dominated empirically for the UK. This is related but slightly different to previous work on estimating government spending multipliers. Government spending multipliers are typically estimated independently of the revenue side of the ledger, whereas this framework considers the net effect of tax and spending policy. I also look at a broad range of metrics to estimate state-contingent effects of the fiscal stance on output. From the earliest point in the EFC literature (Giavazzi and Pagano, 1990), there has been an argument that these are more likely to prevail in periods of distress, and so I attempt to identify conditions in history that reflect such distress and estimate whether in fact those have been associated with an expansionary effect. I also test whether expansionary and contractionary policies have asymmetric effects, as found by Alesina et al. (2018).

The structure of the remainder of this chapter is as follows: section 4.2 reviews the literature on the impact of changes in fiscal stance on output and the mechanisms underlying them; section 4.3 presents the methodology and the data, setting out how the shocks are collected, the arguments for them not being anticipated, and the econometric specification — a modified version of the local projections with instrumental variables (LP-IV) (Jordà, 2005) that has become popular for estimating multipliers; section 4.4 covers the regression results for the linear case, discussing of effects on output, consumption and other macroeconomic variables; section 4.5 then estimates these effects in a state-contingent framework, looking at whether there is evidence of differences under high and low slack, fiscal distress and expansionary versus contractionary fiscal policy; and section 4.6 concludes.

4.2 Literature review

The Great Financial Crisis of 2007-08 prompted a reassessment of the role of fiscal policy in macroeconomic stabilisation in developed economies (Ramey, 2011a; Romer, 2011), particularly as policy interest rates were cut to near zero levels. Around that time, a number of European economies (Ireland, Greece, Portugal, Cyprus, Spain and Italy) also came under pressure to quickly reduce their budget deficits, while other such as the United Kingdom undertook restrictive fiscal policy to — in the words of then-Chancellor of the Exchequer George Osborne — "place [the UK's] fiscal credibility beyond doubt.[1] There was much debate as to the merits of these consolidations and how much effect they would have on output and other macroeconomic variables, a debate that has not yet reached consensus.

From a broader perspective, the canonical Keynesian view of the effect of fiscal policy on output in the short run is through its effect on aggregate demand (Hemming et al., 2002). Prices are rigid at short horizons and the supply potential is largely fixed. Therefore a looser fiscal policy on net — either through cutting taxes, increasing spending or a combination of the two — increases aggregate demand and therefore output; while the opposite is true for a tightening of fiscal policy (that is, reducing the deficit). There are several mechanism through which this increase in aggregate demand might propagate itself, with different levels of effectiveness: for example, a tax cut might be spent differently by different types of households; and direct purchases of goods and services by the government might have a more direct effect on output than transferring money to households, but it might also crowd out more private sector activity. Regardless of these nuances, a positive effect of fiscal loosening on output has been a pretty standard result in macroeconomic models for a long time: it was the basis of the post-Second World War Neo-Keynesian synthesis, and it is still a result in today's workhorse New Keynesian models.

The 1970s, however, provided a challenge to that synthesis (Mankiw, 1986), as both high inflation and high unemployment coexisted. A competing view emerged, which was summarised originally by Fels and Frölich (1987) as the “German view”.

[1] Hansard record of House of Commons sitting of 22 June 2010.

They argued that the large accumulation of debt in West Germany in the late 1970s and early 1980s — despite a healthy economy, and particularly the acceleration in the growth of the deficit — had led to a widespread view that the deficit must be cut in order to avoid the debt being monetised. But by the time the federal budget was passed by the Bundestag in late 1981, the economy was in recession, and so the West German government ended up launching into what appeared to be a heavily procyclical fiscal policy. Fels and Frölich argue that the economy instead recovered quickly, with rapid growth in private consumption and investment. They advance two possible explanations for this non-Keynesian effect. One is based on the German Council of Economic Experts (the Sachverständigenrat, or SVR), which posited the concept of "expectations-induced crowding out" — and that the fiscal tightening had been broadly supported by the public, and that a negative fiscal impulse might raise private demand if it is in line with private sector preferences. The second — which they put more stock on — is that the consolidation was done through lower expenditure growth rather than tax rises, which they posit allowed the private sector to expand in its place.

This was a controversial view at the time. Blanchard (1987), in his immediate response, dismissed the effects of this consolidation as deepening the recession in West Germany, and Miller (1987) also pointed to the SVR view being exactly the criticism levelled at Keynes by the UK Treasury in the 1930s, based on Middleton (1985). Nevertheless, Fels and Frölich's argument would come to be the seed of a renewed literature on negative fiscal multipliers and expansionary fiscal contractions, or EFCs, which may be present under certain conditions — and particularly when tightening policy.

Giavazzi and Pagano (1990) pick up from Fels and Frölich (1987) and Hellwig and Neumann (1987) — the latter reiterating the SVR view of the presence of negative fiscal multipliers in West Germany — by pointing to Denmark and the Republic of Ireland, which implemented the two most extreme fiscal consolidations in Europe in the 1980s. Giavazzi and Pagano focus on the effect of consolidations on private demand (consumption plus investment), showing it to be negatively correlated with changes in government spending. The authors posit this as evidence for the 'German view' (although it should be noted that private demand is only a subset of national income), and lay the success of these consolidations at the feet of positive effects on consumer confidence and interest rates.

Giavazzi and Pagano's proposed mechanism is a type of Ricardian equivalence argument, in which reductions in expenditure increase the likelihood of future tax cuts, and therefore induce a wealth effect in agents — which in turn leads them to increase activity in the private sector that more than makes up for the gap in aggregate demand. Blanchard's (1990) response to Giavazzi and Pagano (1990) and Bertola and Drazen (1993) propose a refinement of this mechanism, proposing theoretical models in which a relatively small adjustment today to stabilise debt can be expansionary if it avoids the need for a larger adjustment in the future.

Alesina and Perotti's (1997) contributions included the coining of the EFC terminology, as well as the strict differentiation between expenditure-based and tax-based fiscal consolidations. Alesina and Perotti's analysis covers theoretical and empirical arguments, and provides a more robust terminology and characterisation of the types of channels through which fiscal policy affects the broader macroeconomy. In addition to the wealth effect, they also argue that there can be a credibility-enhancing consolidations through the lowering of interest rates, particularly for high-debt countries, by crowding in private investment and consumption of durables.

Alesina and Perotti then use a panel of fiscal consolidation episodes to estimate the response of GDP growth, investment and consumption in before, during and after episodes of consolidation, which they classify into successful and unsuccessful. They find that successful ones in their metric — leading to significant falls in the debt-to-GDP ratio and in the cyclically adjusted primary deficit — are associated with stronger growth in all three of GDP, investment and consumption growth.

Perotti (1999) develops and test a model that formalises non-Keynesian effects of fiscal consolidations in good and bad times. The empirical results show in times of fiscal stress — either high debt or low expected tax revenues — a positive government spending shock is associated with a -0.51 change in private consumption. Alesina et al. (2002) focus on profits and business investment, and find similar non-Keynesian effects for those. Hogan (2004) argues that while these components might show an increase, only in extreme cases would they be large enough to make the output effect positive.

The Euro Area debt crisis that broke out in 2009 caused a revival of the EFC literature. Alesina et al. (2015, 2017) do simulations and empirical estimations of output, consumption and investment responses to episodes of consolidations from 1981 to 2014. Results are less supportive of the strongest version of the EFC hypothesis than in the 1990s studies, with negative or insignificant effects on GDP growth and consumption. Estimates of the investment response, however, are generally positive. Alesina et al. (2018, 2019) also develop and estimate a model based on composition differences, arguing that "expenditure-based consolidations" (spending cuts) are less costly than tax-based ones — and some estimates put the output cost at levels that are not statistically significant.

For their part, Afonso and Carvalho (2022) provide empirical evidence for Ricardian wealth effects in European consolidations, and Afonso et al. (2022) provide a model to step through the cases where adjustment might be justified — essentially focusing on large cyclically adjusted primary deficits leading to exploding debt. These non-Keynesian credibility effects echo the channels proposed by Gupta et al. (2018).

Given the significant austerity programme launched in the UK during 2010, there was also significant interest in what its effects were in terms of output costs. De Grauwe and Ji (2013), Riley and Chote (2014), Wren-Lewis (2015), Jordà and Taylor (2016) all identify a negative effect on output growth from the austerity programme introduced in the UK from 2010 onwards, although magnitudes differ. Riley and Chote view the Office for Budget Responsibility's (OBR) estimate of the effect of the consolidation on output growth from conventional multiplier estimates (c.1 percentage point a year in 2010-11 and 2011-12) as probably too high given the economic conditions then, whereas Wren-Lewis and Jordà and Taylor come up with much larger estimates and disagree with judgement about the lack of fiscal space.

4.3 Methodology and data

4.3.1 The approach

This paper's approach and contribution is to take a broad historical view of UK fiscal policy, and particularly to use the richness of a novel dataset going back to the late 1870s on unanticipated discretionary fiscal loosening and tightening. This is a period that encompasses different exchange rate regimes, differnt government aims in managing the macroeconomy, different conditions, large shocks such as large-scale wars and periods of consolidation such as after each of the World Wars and in the 2010s.

To do so, I employ a narrative approach based on archival research from the UK's Parliamentary Archives going back to 1877, similar in spirit to Ramey and Shapiro (1998), Romer and Romer (2010) and Cloyne (2013). This allows me to identify discretionary changes in the fiscal stance, measured through a combination changes to tax policy that take immediate effect and changes in spending relative to forecast at the beginning of the year.

On the expenditure side, this is the same dataset used in chapter 3, which also discusses in detail the arguments for these shocks not being anticipated in the context of the UK's unique budget framework. The dataset includes all forms of discretionary government spending — meaning I exclude the two most cyclical elements of expenditure, social security and debt interest — which I use to construct a measure of the intra-year surprise in spending. This is obtained by comparing the forecast at the beginning of the financial year (presented in the budget to the House of Commons) with the first estimate of how much was spent, which has throughout been published in the subsequent year's budget.[2] This creates a series of contemporaneous shocks, which are unlikely to be anticipated, and will reflect in large part changes in policy throughout the course of the financial year.

On the revenue side, I use a similar approach to construct the series by collecting in-year policy changes to tax as announced in the budget scorecards. Tax policy is inherently more difficult to analyse than discretionary spending, because the eventual tax take depends on the performance of the economy, and that is why I prefer to (i) focus on policy changes, which to some extent mitigate the effect of the economy on the tax base; and (ii) I use the *a priori* revenue forecast to be collected from those changes, which is the intended fiscal impulse by the government and also avoids the challenge of decomposing tax receipts into policy and underlying changes after the fact.[3]

Note that I only use announcements of tax policy changes that come in either with immediate effect or in the course of the year, and not announcements beyond that. The main reason is to do with needing a series of unanticipated shocks, which are so key to identification. By definition, tax policy changes announced years in advance will have been anticipated by the time they come into place, and so including them as fiscal impulses in those years is out of the question. Single-year costings are the norm for most of the period I use (multi-year policymaking has only become commonplace since the 1990s), and so I opt to use those in-year changes

[2] Or in the equivalent fiscal statement in the Spring.

[3] See for example the OBR's *Forecast Evaluation Report* from October 2023, page 31, which illustrates how policy changes can be dwarfed by changes in economic factors. It would be impractical to try and implement this for every policy change.

only. An improvement on this might be to use the data to construct a discounted series for more recent times, although that would require some judgement as to the appropriate discount rate.

4.3.2 The shocks to the discretionary fiscal stance

I combine revenue and expenditure shocks to produce a single series of discretionary spending shocks. This allows me to capture the overall change to the fiscal stance, and therefore create a measure of how tight or loose fiscal policy ran relative to announcements prior to the beginning of the financial year.

Figure 4.1 shows the discretionary fiscal stance shocks over time, as well as the breakdown between revenue and spending shocks. All of these are expressed as changes to the budget balance (defined as revenue minus expenditure), and therefore a positive number means a tighter policy (higher revenues or lower spending), while a negative number means a looser policy.

The patterns emerging from looking at these data, especially in earlier years, are very consistent with patterns relating to war efforts: large spending shocks during the war, which the government finances in part through higher taxation but also through debt build-up. But they also illustrate just how differently the UK government reacted to the Second World War relative to the First — especially on the tax side, as a result of Keynes' influence (Cooley and Ohanian, 1997). This meant much larger tax increases as part of the war financing, and contributed to much smaller shocks to the fiscal stance on net. The First World War, on the other hand, looks much more like an exception — an extremely expensive war with little in the way of attempting to finance it through high taxes, save for the 1916 Budget, which saw a large hike in income tax.

Event	Year	Total shock	o/w revenue	o/w spending
Second Boer War	1899-1900	-1.16	0.05	-1.20
	1900-01	-1.06	0.63	-1.69
	1901-02	-0.02	0.56	-0.58
	1902-03	-0.22	0.26	-0.48
	1903-04	-0.72	-0.54	-0.18
First World War	1914-15	-12.68	0.93	-13.61
	1915-16	-32.09	0.10	-32.19
	1916-17	-43.44	2.07	-45.51
	1917-18	-8.94	0.58	-9.52
	1918-19	8.02	1.31	6.71
Second World War	1939-40	-4.18	2.18	-6.36
	1940-41	-13.58	3.12	-16.70
	1941-42	-4.87	1.76	-6.62
	1942-43	-2.08	1.70	-3.78
	1943-44	0.61	1.05	-0.44
	1944-45	-1.23	0.04	-1.27
	1945-46	-2.11	-2.96	0.86
Iraq War	2002-03	-2.66	0.00	-2.66
	2003-04	-0.07	0.01	-0.08

Table 4.1: Notable shocks to the budget balance related to wars against foreign powers, expressed as percentage of GDP. Columns may not sum to totals due to rounding.

Since the Second World War, the UK's involvements in wars against foreign powers has not consumed anywhere near the same level of resources. The exception to this was the Iraq war in 2003, which due to timing led to an increase in expendi-

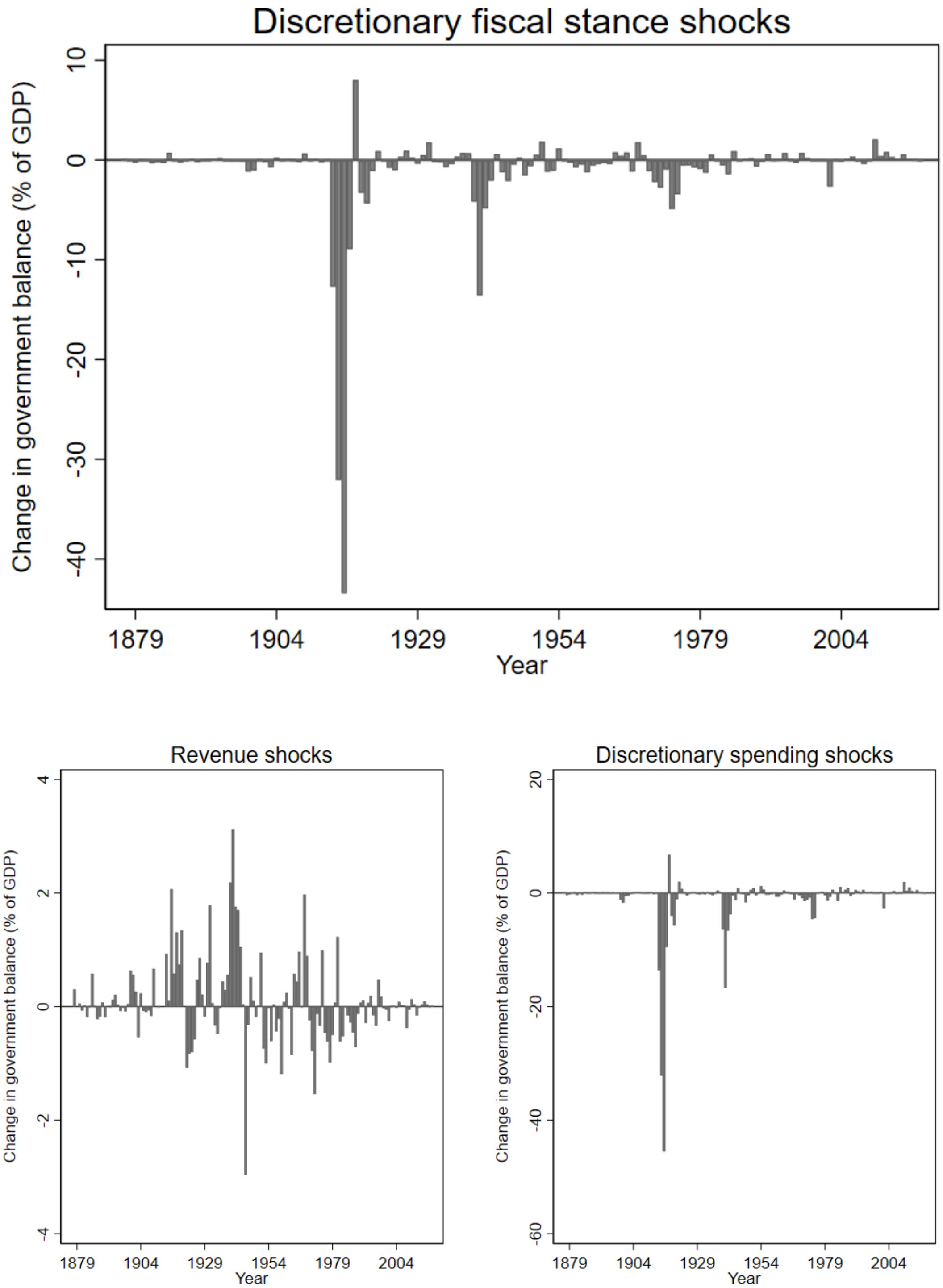


Figure 4.1: Discretionary fiscal stance shocks and breakdowns from 1879 to 2018 in the dataset.

ture relative to plans in 2002-03, and which the government chose to accommodate instead of increasing taxes.

But there have also been other significant, non-war related shocks. The People's budget of 1909 is the earliest one of these, in which a large increase in the taxation of higher incomes was announced — and which precipitated the fall of the government, as well as the formal establishment of the primacy of the House of Commons in financial matters through the Parliament Act 1911.

Event	Year	Total shock	o/w revenue	o/w spending
People's Budget	1909-10	0.65	0.67	-0.02
Post-First World War	1919-20	-3.28	0.74	-4.02
	1920-21	-4.36	1.34	-5.70
	1921-22	-1.13	0.00	-1.13
Great Depression budgets	1930-31	0.50	0.77	-0.27
	1931-32	1.78	1.79	-0.00
NHS cost overruns	1948-49	-1.55	0.10	-1.65
	1949-50	-0.61	-0.18	-0.43
1959 pre-election budget	1959-60	-1.24	-1.19	0.05
Maudling's "dash for growth"	1963-64	-0.41	-0.85	0.44
Wilson government's fiscal crisis	1966-67	0.75	0.97	-0.21
	1967-68	-1.16	0.00	-1.15
	1968-69	1.79	1.97	-0.18
	1969-70	0.47	0.89	-0.42
"Barber boom"	1972-73	-2.77	-1.54	-1.23
Loss of control of spending	1974-75	-4.92	-0.34	-4.58
	1975-76	-3.44	0.99	-4.43
Howe's controversial budget	1981-82	0.55	1.22	-0.67
Osborne's austerity programme	2010-11	2.07	0.13	1.94

Table 4.2: Notable shocks to the budget balance not related to wars against foreign powers, expressed as percentage of GDP. Columns may not sum to totals due to rounding.

The post-First World War period was also one of significant fiscal pressures, with spending overruns as high as 5.7% of GDP during 1920-21 as a result of domestic (higher pay awards) and foreign affairs (League of Nations mandates and war in Ireland). But there was some attempt to offset this through increases in taxes, with a broad increase in rates of excise duties and additional taxes on profits.

The 1930-31 and 1931-32 financial years are particularly notable for their timing. Britain was affected by the Great Depression after a decade of poor economic performance, in which governments continually pursued large surpluses to attempt to maintain fiscal sustainability (Crafts, 2018). As tax revenues fell considerably, then-Chancellor of the Exchequer Philip Snowden raised taxes to try and bring about a balanced budget, in line with orthodox thinking. Cloyne et al. (2023) point out that these are two very different years in tone — the 1930 budget is not one of emergency, but the two in 1931 were. The combined in-year increase in taxes in 1931-32 is astonishingly high — nearly 2% of GDP — especially given the dire economic situation of the time.

The next large, non-war related shock would be the overrun of costs in the first couple of years of the National Health Service, during which expenditure quickly surpassed what had been budgeted for — and which the government chose to broadly accommodate through a worsening fiscal balance. It would be another ten years until another large shock, this time a pre-election budget in 1959, when Chancellor Derick Heathcoat-Amory introduced substantial cuts to income tax, beer duty and the purchase tax, adding up to 1.2% of GDP.

The 1960s and 1970s represented periods of see-sawing of fiscal policy, with large tax cutting budgets — Reginald Maudling in 1963, Anthony Barber in 1972 — followed by Harold Wilson-led governments having to tighten fiscal policy considerably to deal with external pressures on the pound and on the balance of payments. Of particular note are the 1967 in-year stimulus attempt, which could not in the end avoid the devaluation of sterling in November (Newton, 2010); and the 1975-76 in-year tax raising announcements equivalent to 1% of GDP, alongside two years of spending exceeding plans by around 4.5% of GDP. The latter was largely due to increases in public sector wages, which immediately preceded the IMF crisis of 1976 (see chapter 3).

The last two shocks of special note are two highly tightening fiscal announcements, and which generated much controversy. Geoffrey Howe (1981) and George Osborne (2010) went against the established Keynesian principles of countercyclical policy, and instead used their budgets to announced large tax increases and large cuts to expenditure during the year, respectively. The 1981 budget was infamously decried by 364 economists in an open letter to The Times (Neild, 2014), and the 2010 "emergency" budget was also highly criticised (Barrell, 2014). 1981 and 2010 are also the largest examples of UK governments setting their fiscal policy stance on the basis of non-Keynesian effects dominating Keynesian effects — and therefore of particular interest to this paper.

4.3.3 To what extent are these shocks anticipated?

The timing of the shocks and the extent to which they are anticipated is something that all empirical work on fiscal policy must grapple with (Ramey, 2011a). Of course, it is impossible to be sure that they are truly unanticipated, but there are a number of arguments which increase confidence in this assertion.

For one, the UK's fiscal policy framework is more suited to this kind of analysis than most. The Westminster system relies on the government being able to pass money bills in the House of Commons as a pre-condition, or else it would fall. The combination of this with the convention — enshrined in law since 1911 — that the House of Commons has a primacy on money bills and the strong party discipline over the whole sample means that a government can effectively pass any major fiscal legislation it puts before the House.

The process of designing the budget is also conducted essentially behind closed doors, with the Treasury deciding in conjunction with the Prime Minister what fiscal policy should be, while spending departments depend on allocations from the Treasury to know what their budget will be. Of course, some politicians will have more clout than others, but the Chancellor of the Exchequer is not equal in status to other ministers, and is clearly in charge of fiscal decisions.

Chapter 3 highlighted a number of arguments for the in-year discretionary spending shocks being unanticipated, as they are the difference between the initial forecast for expenditure and how much was actually spent in the course of the financial year. Given that the government has the ability to pass the legislation behind these estimates and that it uses it for financial management, it seems reasonable to assume that such shocks would be less likely to be anticipated. Testing using Mincer-Zarnowitz (1969) regressions supports that assertion.

There is more of a case for there to be some anticipation of what the fiscal stance as a whole like given the current state of the economy and the public finances, even within the year. And that is indeed what I find. While testing separately tax and spending shocks means I fail to reject the null hypothesis that they are forecastable, when combining them into a single fiscal stance indicator, I reject the

Mincer-Zarnowitz null hypothesis at the 10% level (but not at the 5% level).[4] Using the residuals, however, allows me to decompose the errors into anticipated and unanticipated components. For the residuals, I fail to reject the Mincer-Zarnowitz null at conventional levels of significance. Running the regressions with either the raw or residual shock yields indistinguishable results in the second stage of the 2SLS estimation.[5]

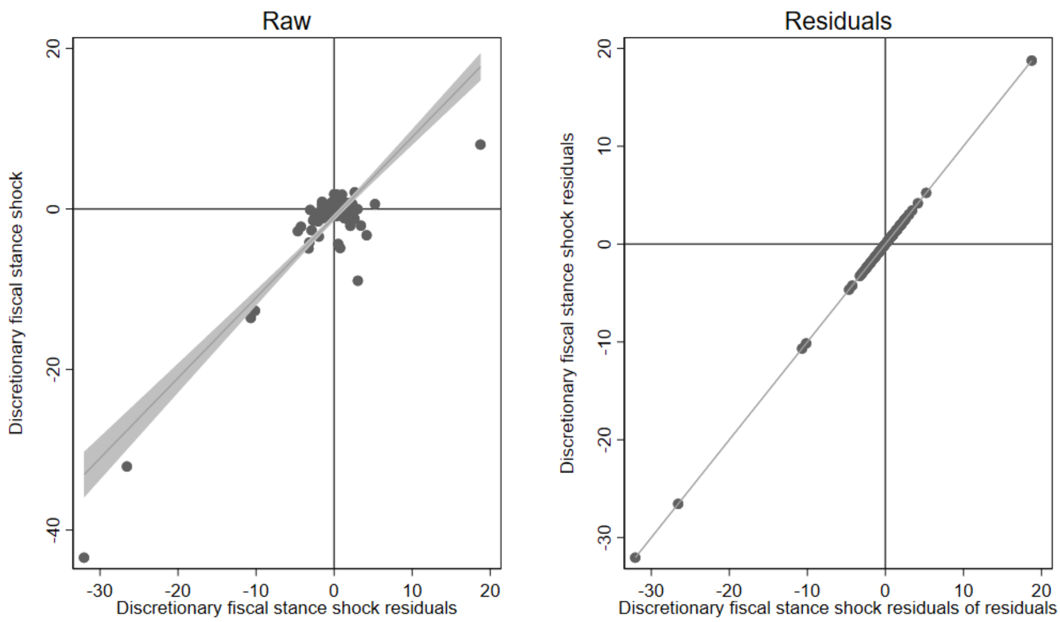


Figure 4.2: Mincer-Zarnowitz tests of the raw and residual shocks to the discretionary fiscal stance, with the regression line fitted to the shocks and 95% confidence intervals around them. I reject the null hypothesis on the left-hand side at the 10% significance level (p-value = 0.0776) but fail to reject it on the right-hand side (p-value = 0.1963).

4.3.4 Econometric specification

I estimate the cumulative impulse response functions (IRFs) directly using localm projections with instrumental variables (LP-IV), a method based on Jordà (2005) and Stock and Watson (2018), and similar to chapters 2 and 3, as well as Ramey and Zubairy (2018). Plagborg-Møller and Wolf (2021) show that they estimate equivalent IRFs to a vector autoregression in the linear case, while allowing for flexible modelling of non-linear effects.

In the linear case, I estimate the following equation at each horizon h:

[4] As discussed in chapter 3, using the Mincer-Zarnowitz regressions to test how forecastable shocks are means that I try to forecast them means using an autoregressive structure with the same variables as in the main regression. I then compute the fitted values and the residuals, and regress the shocks on their residuals with a constant. I then test the joint hypothesis that the constant is 0 and the coefficient on the residuals is 1. If I fail to reject that joint hypothesis, then — according to the Mincer-Zarnowitz approach — the shocks are unforecastable.

[5] This is not unexpected given that the reason for rejecting the Mincer-Zarnowitz null hypothesis at the 10% is the intercept being statistically different from zero, and so using the residuals essentially eliminates the average bias by constructing it to zero — but it does so without penalising the slope so much that it being equal to 1 no longer holds.

$$\sum_{j=0}^{h} \ln y_{t+j} = \phi_h L z_t + \beta_h \sum_{j=0}^{h} \frac{\tau_{t+j} - g_{t+j}}{y_{t+j}} + \varepsilon_{t+h} \tag{4.1}$$

where y is output, τ are tax revenues, g is government spending, Lz_t is a lagged set of control variables (real interest rates, inflation, employment rate and debt-to-GDP ratio); and ε is an error term. This allows the retrieval of β_h as an estimate of the semi-elasticity of output with respect to the government budget balance at each horizon h. The budget balance in each period t is defined as follows:

$$BAL_t = \frac{\tau_t - g_t}{y_t} \tag{4.2}$$

Because this is measured as a share of GDP, β_h provides the percentage effect on GDP of a 1% of GDP increase in the budget balance, meaning that both sides are measured in the same units and therefore no transformation is required to interpret the results in an intuitive manner. Essentially, it gives us the equivalent of a multiplier effect for a 1% of GDP tightening of fiscal policy.

But the budget balance is clearly simultaneously determined with GDP, and therefore a classic example of a situation in which the OLS estimator will produce biased and inconsistent estimates. To get around this issue, I estimate equation (4.1) using 2SLS, treating the budget balance as the endogenous variable and using the above described discretionary fiscal stance shock as the instrument.

To limit any issues with potential weak instrument bias — which can be an issue with annual macroeconomic data — I calculate Anderson-Rubin (1949) (AR) confidence sets for $\hat{\beta}_h$, following the approach of Montiel Olea, Stock and Watson (2021) which recommends always using weak instrument robust inference in the just-identified case.[6]

The setup for the two-state case is very similar to that of equation (4.1), with an added indicator variable for each state:

$$\begin{aligned} \sum_{j=0}^{h} \ln y_{t+j} = I_{t-1} \left[\phi_{A,h} L z_t + \beta_{A,h} \sum_{j=0}^{h} \frac{\tau_{t+j} - g_{t+j}}{y_{t+j}} \right] \\ + (1 - I_{t-1}) \left[\phi_{B,h} L z_t + \beta_{B,h} \sum_{j=0}^{h} \frac{\tau_{t+j} - g_{t+j}}{y_{t+j}} \right] + \nu_{t+h} \end{aligned} \tag{4.3}$$

where all variables in common with (4.1) are the same; A and B are two mutually exclusive states; I_{t-1} is an indicator variable which is 1 if state A is applicable and 0 otherwise; and ν is an error term. The ease of implementing these non-linear effect is one of the advantages of using an LP-IV framework, and it is something I use heavily to test under what conditions non-Keynesian effects might dominate.

I also use the equivalent of equations (4.1) and (4.3) with other macroeconomic variables, such as household consumption, policy interest rates, the employment rate and inflation. I simply replace y_t with the relevant variable to estimate the IRF for that variable.

[6] As Montiel Olea, Stock and Watson show, if instruments are strong, confidence sets using the AR method are identical to those obtained using 2SLS; and if they are weak, the non-parametric AR setup means that one does not assume the confidence interval is centred on the point estimate of the parameter.

4.3.5 Additional macroeconomic data

I combine two sources of macroeconomic data. One is the Office for National Statistics (ONS), the UK's official statistics body, which publishes national accounts on a regular basis and has estimates for most series from 1946 onwards. For data prior to 1946, I splice the ONS series with the Thomas and Dimsdale (2017) consensus estimates on a politically consistent basis.[7] This includes GDP, the fiscal balance[8], household consumption, inflation as measured through the GDP deflator, the employment rate and the Bank of England's policy interest rate[9].

4.4 Linear estimates

Figure 4.3 and table 4.3 summarise the results obtained for the linear case, in which I estimate a single parameter for the output cost per 1% of GDP tightening over the whole sample. The estimates indicate that over the whole period, Keynesian effects dominate, as a 1% of GDP increase in the budget balance is associated with a 0.24 decrease in GDP, with the effect rising to a 0.37 decrease in the long run.[10] The effect on GDP is statistically significant at the 1% level at all horizons.

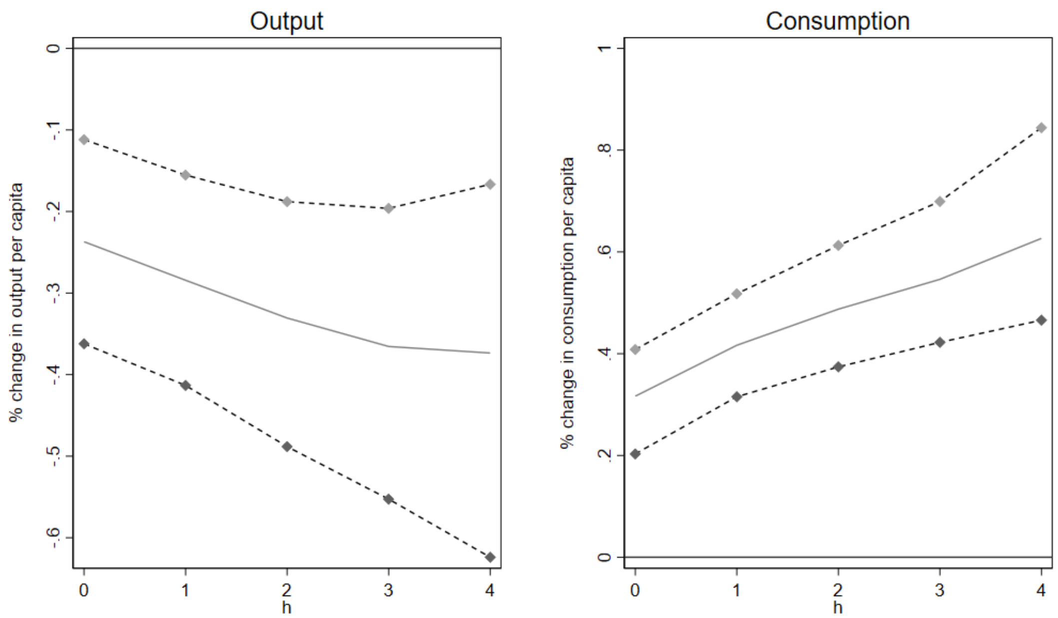


Figure 4.3: Cumulative IRFs of output and consumption to a 1% of GDP increase in the budget balance. Dashed lines represent 90% confidence sets calculated using the inverted Anderson-Rubin test.

[7]This accounts for Irish independence, and it takes into account the political boundaries over which the UK's fiscal policy is decided at each point in time.

[8]I use the reciprocal of central government net borrowing. This is the most consistent measure of the fiscal balance over time, and also the one most tightly controlled by the Treasury, as well as being by far the largest component of the public sector balance in the UK.

[9]Given the secular decline in real and nominal interest rates, I use a Kálmán filter to estimate the wedge between the fitted interest rate and the actual interest rate at each point in time.

[10]Four years after impact is roughly equivalent to the twenty quarters used in other studies for long-term effects (e.g. Ramey and Zubairy, 2018).

	Impact (Year 0)	Year 1	Year 2	Year 3	Year 4
Output	-0.2370***	-0.2843***	-0.3306***	-0.3655***	-0.3735***
	(0.0793)	(0.0817)	(0.0904)	(0.1073)	(0.1310)
90% AR confidence set	[-0.3622, -0.1119]	[-0.4132, -0.1553]	[-0.4882, -0.1879]	[-0.5527, -0.1962]	[-0.6239, -0.1667]
Consumption	0.3164***	0.4166***	0.4875***	0.5461***	0.6266***
	(0.0650)	(0.0641)	(0.0718)	(0.0877)	(0.1138)
90% AR confidence set	[0.2030, 0.4081]	[0.3153, 0.5178]	[0.3741, 0.6128]	[0.4223, 0.6991]	[0.4659, 0.8441]
Kleibergen-Paap F-stat	80.4592	70.7179	50.6575	31.8652	19.8391
N	140	139	138	137	136
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.3: LP-IV estimates of the effect of a 1% of GDP increase in the fiscal balance on output per capita and consumption per capita. Values in brackets are heteroscedasticity and autocorrelation robust (HAR) standard errors estimated using the QS kernel. 90% confidence sets are calculated based on the inverted Anderson-Rubin test, in line with Davidson and MacKinnon (2014) and Montiel Olea, Stock and Watson (2021). Statistical significance is calculated based the whole Anderson-Rubin confidence sets being the same side of zero on the real line at the 90% (for the 10% significance level), 95% (for the 5% significance level) and 99% (for the 1% significance level) confidence levels.

These results do bear out some important non-Keynesian effects, especially on consumption, in line with the findings of the EFC literature started with Giavazzi and Pagano (1990). Consumption per capita increases by 0.32% on impact after a 1% of GDP tightening, with the effect rising to 0.63% by year 4, with statistically significant effects throughout. It is however not large enough to turn overall effect of a fiscal tightening positive when measured in output — a point made by Hogan (2004).

The consumption effect is still relatively strong, as it is by far the largest component of GDP (around two-thirds across the whole time period). A 0.32% of GDP increase in consumption is equivalent to a 0.20% increase in GDP, implying that the combined effect of the remaining components is to drive down GDP by 0.44% for every 1% of GDP tightening. Assume the 1% of GDP tightening were done just through reducing government consumption, that effect is similar to the fiscal multiplier estimated in chapter 3 — a check that provides some comfort in terms of the consistency of the results.

As for effects of a fiscal tightening other macroeconomic variables (figure 4.4), the effects broadly follow the literature. The effect on policy interest rates is generally insignificant, with a small negative and statistically significant effect by year 4 — similar to what Murphy and Walsh (2022) found when surveying the literature. The effect on the employment rate is much larger, with a 1% of GDP tightening linked to a over a 1 percentage point fall in the employment rate in the long run, consistent with the findings of Monacelli et al. (2010) and Dupor and Guerrero (2017). And the results from these estimates bear out the findings of Ferrera et al. (2021) that a fiscal expansion is inflationary, and by analogy, a contraction is disinflationary.

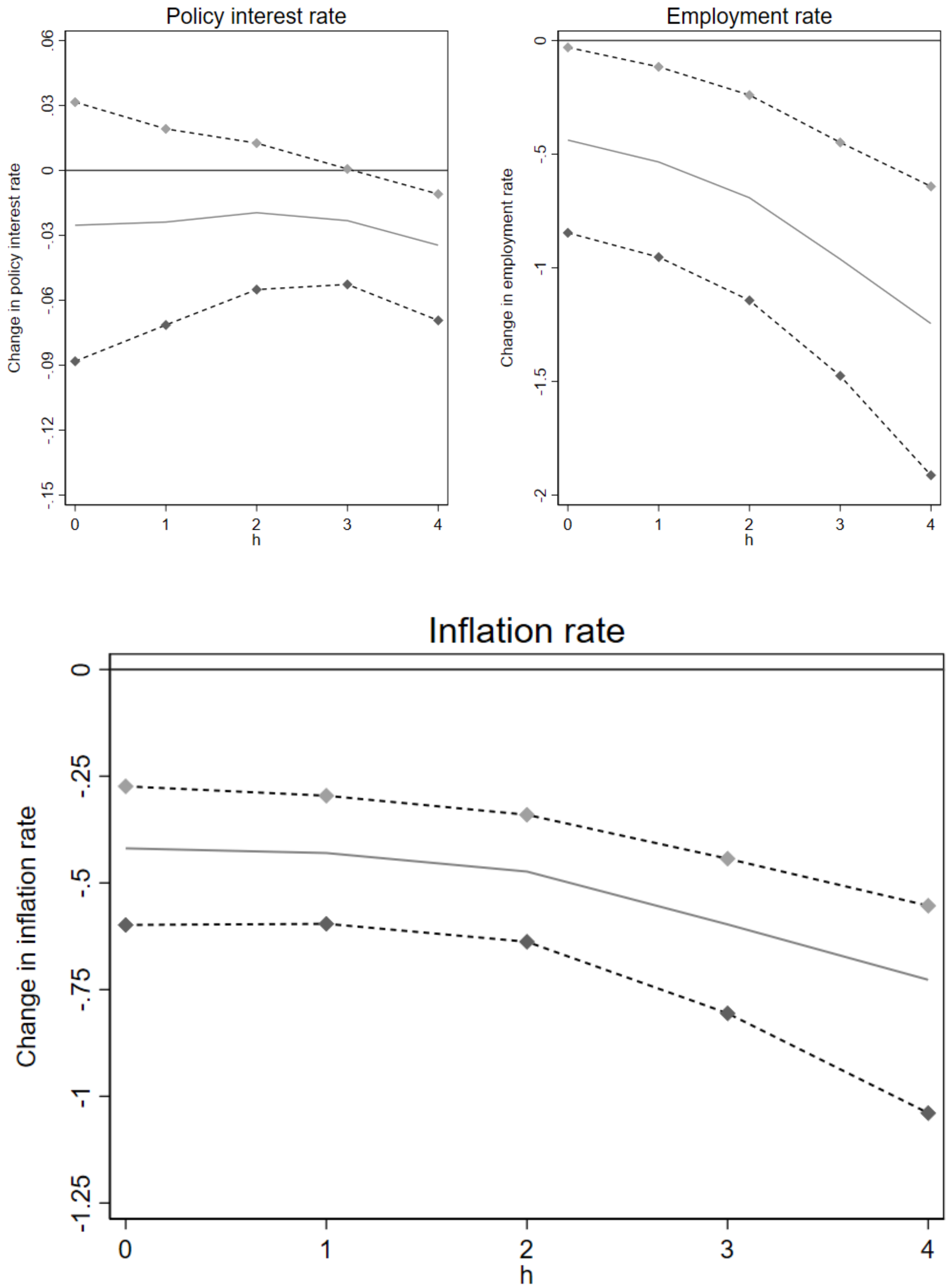


Figure 4.4: Cumulative IRFs of policy interest rates, the employment rate and inflation rate to a 1% of GDP increase in the budget balance. Dashed lines represent 90% confidence sets calculated using the inverted Anderson-Rubin test.

4.5 State-contingent estimates

4.5.1 Do results differ in conditions of high and low slack?

An interesting question is whether there are particular conditions under which the effect of a fiscal tightening might have different effects. A logical place to start would be to look at times of high and low slack, as there is plenty of literature suggesting that fiscal policy might be more effective when slack is high (Ramey and Zubairy, 2018).

Effect of 1% of GDP fiscal tightening		Years after impact 0	1	2	3	4
Unemployment rate						
Above average	Y	-0.7905***	-0.7258***	-0.8580***	-1.0010***	-1.1979***
	C	0.3369	0.4206*	0.4440*	0.4461	0.4352
	F-stat	76.7	97.5	57.8	36.0	24.6
Below average	Y	-0.2377***	-0.2962***	-0.3534***	-0.4043***	-0.4296***
	C	0.3150***	0.4168***	0.4891***	0.5435***	0.6188***
	F-stat	101.3	97.2	70.7	42.0	25.6
Difference	Y	-0.5528*	-0.4296*	-0.5047**	-0.5966**	-0.7682**
	C	0.0219	0.0038	-0.0451	-0.0974	-0.1835
Recession						
Recession	Y	-0.8642***	-1.1479***	-1.6722**	-1.8286*	-1.7813
	C	0.5379**	0.1858	-0.0383	0.0838	0.2878
	F-stat	24.7	15.1	8.2	6.0	6.3
Expansion	Y	-0.1743**	-0.2184***	-0.2789***	-0.3357***	-0.3759**
	C	0.2057***	0.3336***	0.4560***	0.5439***	0.6474***
	F-stat	111.8	64.2	35.3	18.6	9.6
Difference	Y	-0.6899**	-0.9295*	-1.3933	-1.4929	-1.4054
	C	0.3322	-0.1478	-0.4943	-0.4601	-0.3596
Significance: *** - 1%; ** - 5%; * - 10%						

Table 4.4: Percentage changes in output per capita (Y) and consumption per capita (C) in response to a 1% of GDP increase in the budget balance across different measures of slack. Significance calculations reflect the Anderson-Rubin confidence sets for single estimates and tests of restrictions on coefficients using HAR standard errors. F-statistics are calculated using the Kleibergen-Paap method.

Table 4.4 uses two measures of slack — the unemployment rate relative to the sample average and the economy being in recession — as states across which effects of a fiscal tightening are allowed to vary. All estimates of the output effect are negative, and all but one are statistically significant. There is some evidence of a higher output cost to tightening fiscal policy when slack is high — or put it another way, fiscal policy has a larger stimulative effect in times of high slack. The difference in the output effect is not statistically significant at longer horizons for the recession and expansion states, which is unsurprising given the low power of the shock in recession states in later years. This also likely drives the lack of statistical significance of the point estimate of the output effect in year 4 for a recession, despite it the large point estimate in absolute terms. On the other hand, there is no statistically significant difference in the effects of a fiscal tightening on consumption, suggesting that the mechanism driving the divergent output effects.

4.5.2 Are effects of fiscal tightenings or loosening different under conditions of fiscal distress?

But the essence of the "German view" of the SVR in the 1980s, of Fels and Frölich (1987) and Hellwig and Neumann (1987), and of the subsequent EFC literature is less focused on the effectiveness of fiscal policy in general situations and more in times of distress or constrained fiscal space. So a more relevant question to assess whether the UK experience reflects that literature's results might be whether fiscal policy effects differ in times of fiscal constraints.

To that effect, I construct four separate measures of fiscal distress[11]:

- A period of quick accumulation of debt, reflecting an unsustainable path of fiscal policy in the recent past. I define this as an accumulation of 10% of GDP in debt over five years;
- A relatively quick increase in the cost of government borrowing, defined by the short-term interest rate on gilts increasing by 0.75 percentage points over three years;
- A relatively rapid rise in the interest burden in servicing the debt, reflecting a combined effect of the debt stock and the interest rate effect. I define this as an increase in debt interest as a share of GDP of more than 0.5 percentage points over three years;
- And a composite measure of the three, defined as 1 if any of the three measures of distress is 1 in any year.

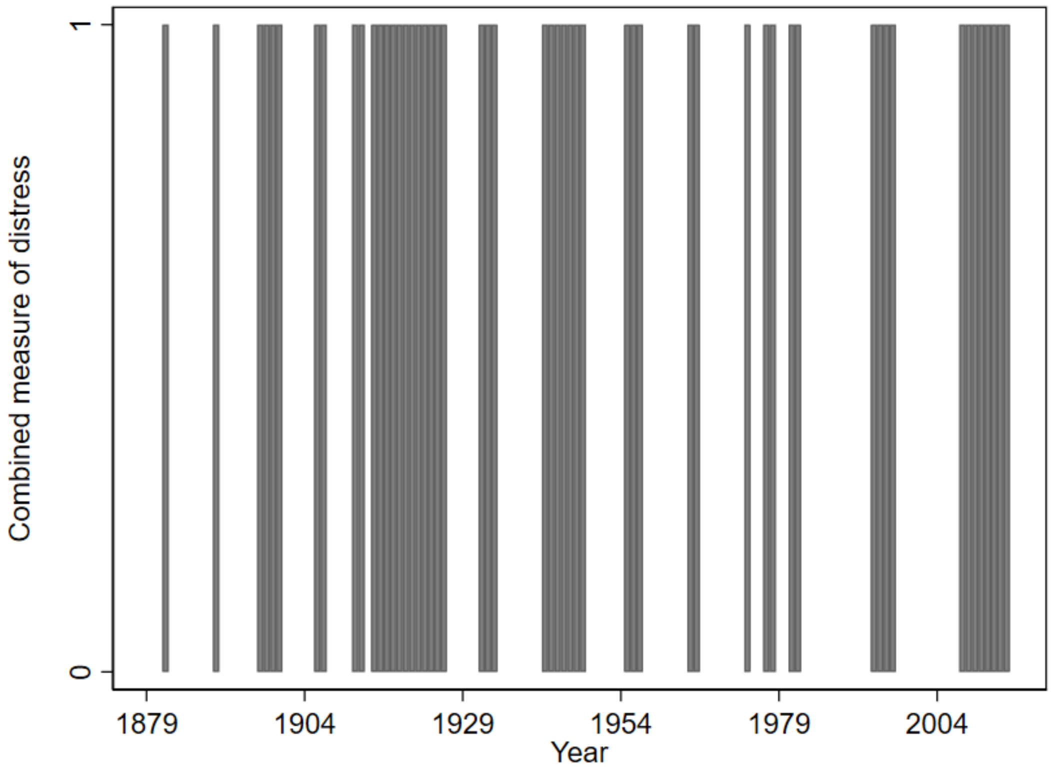


Figure 4.5: Combined measure of fiscal distress over the sample (1879-2018). 1 indicates a period of high distress under the measure, 0 indicates low distress.

Figure 4.5 shows the incidence of the combined measure of distress over the sample, with each of the individual measures' values available in the appendix. Each

[11] I have tested different thresholds — generally with similar results, but of course the stricter the cut-off, the less power the first stage will have, which is a normal trade-off in this situation. The measures presented here have good instrument strength, and my judgement is that they combine useful metrics with strength of the first stage in a meaningful way

individual measure captures slightly different times of distress: for example, there are five main periods of high debt accumulation (First and Second World Wars, the Great Depression, the 1990s and the post-Great Financial Crisis), but there are other periods in which UK fiscal policy was constrained — for example, due to spikes in interest rates (1966-68 and 1974-75) or due to the combination of accumulated debt and interest rates making it significantly more expensive to service the debt (1977-79 and 1981-83).

Effect of 1% of GDP fiscal tightening		Years after impact 0	1	2	3	4
		Debt accumulation				
High	Y	-0.1857	-0.2515	-0.3525**	-0.4227*	-0.4689
	C	0.4649**	0.6464***	0.6998***	0.7554***	0.9292***
	F-stat	27.1	38.3	49.5	46.0	34.7
Low	Y	-0.1852**	-0.2082***	-0.2260***	-0.2541**	-0.2545**
	C	0.1306*	0.2186***	0.3306***	0.4305***	0.5014***
	F-stat	316.9	170.4	87.2	47.2	25.9
Difference	Y	-0.0004	-0.0432	-0.1264	-0.1685	-0.2145
	C	0.3343**	0.4278***	0.3692***	0.3249*	0.4279
		Interest rate distress				
High	Y	-0.0233	-0.0959	-0.0942	-0.1039	-0.1031
	C	0.0915	0.0863	0.1763	0.2367	0.2687
	F-stat	1363.9	305.5	121.4	75.9	55.8
Low	Y	-0.2550***	-0.2993***	-0.3505***	-0.3923***	-0.4026***
	C	0.3298***	0.4524***	0.5258***	0.5868***	0.6818***
	F-stat	74.0	74.0	62.3	42.2	25.3
Difference	Y	0.2318	0.2034	0.2564	0.2884*	0.2995
	C	-0.2383	-0.3660***	-0.3495***	-0.3481***	-0.4131***
		Interest burden distress				
High	Y	-0.2337	-0.2876*	-0.3581**	-0.4043*	-0.4035
	C	0.4946***	0.6638***	0.7179***	0.7859***	0.9740***
	F-stat	44.2	77.6	133.1	143.6	128.5
Low	Y	-0.2342***	-0.2637***	-0.2856***	-0.3119***	-0.3186**
	C	0.1271*	0.2267***	0.3453***	0.4414***	0.5051***
	F-stat	298.0	166.3	89.8	51.5	28.8
Difference	Y	0.0005	-0.0239	-0.0725	-0.0924	-0.0849
	C	0.3674***	0.4371***	0.3726***	0.3445***	0.4689***
		Combined indicator of distress				
High	Y	-0.1437	-0.2443**	-0.3258***	-0.3697***	-0.3848**
	C	0.4822***	0.5498***	0.5858***	0.6220***	0.7257***
	F-stat	59.2	69.2	57.7	36.8	23.6
Low	Y	-0.2128***	-0.2276***	-0.2512**	-0.2925**	-0.2961*
	C	0.1351	0.2545***	0.3771***	0.4881***	0.5754***
	F-stat	331.0	231.4	142.3	72.6	33.0
Difference	Y	0.0691	-0.0167	-0.0747	-0.0772	-0.0887
	C	0.3471***	0.2953***	0.2086**	0.1339	0.1503
Significance: *** - 1%; ** - 5%; * - 10%						

Table 4.5: Percentage changes in output per capita (Y) and consumption per capita (C) in response to a 1% of GDP increase in the budget balance across different measures of fiscal distress. Significance calculations reflect the Anderson-Rubin confidence sets for single estimates and tests of restrictions on coefficients using HAR standard errors. F-statistics are calculated using the Kleibergen-Paap method.

The results in table 4.5 using different measures of fiscal distress generally point to no statistical difference between times of distress and normal times in terms of the effect on output. However, there are some interesting results that bear out some of the findings of the EFC literature in terms of consumption. For all cases apart from sharp increases in interest rates by themselves, the effect of a fiscal tightening is to increase private consumption by a statistically significantly larger amount in times of distress. This is borne out by the composite measure of fiscal distress as well.

This is strong evidence of not just an expansion in private demand in response to a fiscal tightening, but a stronger one in times of distress — very similar to the findings of Giavazzi and Pagano (1990). Delving into the components of GDP, I estimate that the fall in government spending itself as part of a fiscal tightening more than explains the fall in output, while the increase in imports also contributes to it. These are then offset by the other components of expenditure GDP: consumption, investment and exports, but not to enough of an extent that GDP changes turn positive. Figure 4.6 provides an approximation of the contributions of each component of GDP.

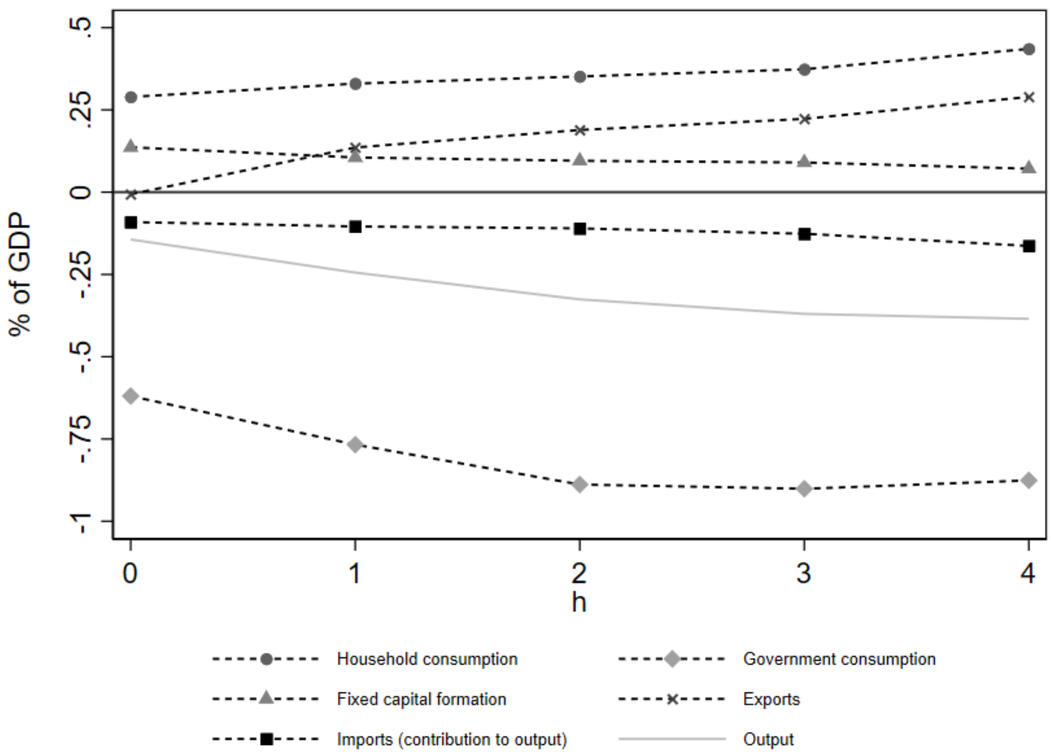


Figure 4.6: Approximation of output contributions by each of the components of expenditure GDP in response to a 1% of GDP tightening during times of distress under the combined indicator, calculated using estimated IRFs and sample average weights for each component. Note that this is only an approximation, but illustrates the relative magnitude of each component. Output IRF was estimated directly and therefore is not strictly a sum of its components.

4.5.3 Is there evidence of differences between tightenings and loosenings?

There is an open question about whether fiscal tightenings and loosenings are inherently different in their effects. Estimating their effects together increases the number of episodes over which we can estimate coefficients, but it also means we might miss that asymmetry altogether by lumping them into a combined fiscal indicator. This is what Alesina et al. (2018) refer to when they say that their analysis "focuses only on fiscal contractions: we have nothing to say about expansionary fiscal policies."

To do so, I estimate fiscal tightenings and loosening as different states, and then estimate the effects on output and consumption. Table 4.6 summarises these results.

Effect of 1% of GDP fiscal balance increase		Years after impact 0	1	2	3	4
Tightening	Y	0.6863	0.1953	0.1448	0.1840	0.2737
	C	0.2442	0.2128	0.0992	0.0157	0.0321
	F-stat	21.0	37.3	28.4	16.3	9.7
Loosening	Y	-0.2798***	-0.3035***	-0.3426***	-0.3739***	-0.3881***
	C	0.3034***	0.3824***	0.4666***	0.5378***	0.6177***
	F-stat	88.9	80.0	56.2	33.8	21.8
Difference	Y	0.9660*	0.4988	0.4874	0.5579	0.6618
	C	-0.0592	-0.1696	-0.3675	-0.5221	-0.5856
Significance: *** - 1%; ** - 5%; * - 10%						

Table 4.6: Percentage changes in output per capita (Y) and consumption per capita (C) in response to a 1% of GDP increase in the budget balance across fiscal tightenings and loosenings. Significance calculations reflect the Anderson-Rubin confidence sets for single estimates and tests of restrictions on coefficients using HAR standard errors. F-statistics are calculated using the Kleibergen-Paap method.

The results provide some, though weak, evidence that there might be a difference between tight and expansionary policies in terms of their effect on output. The effect of improving the fiscal balance on output is statistically significantly less severe in-year than that of loosening it at the 10% level, though not for the remainder of the horizon. But none of the coefficients on output or consumption are statistically significant at conventional levels for fiscal tightening, even if I was able to estimate significant effects with similar first-stage F-statistics — meaning that it is unlikely to be just due to the shock being less powered. So despite the test of restriction on coefficients generally not being significant, we can confidently say that loosenings have an expansionary effect on output, but not that tightenings have a restrictive — nor expansionary, for that matter — effect on output.

4.6 Conclusion

The canonical Keynesian view of the effect of a policy to increase the fiscal balance is that it would have a restrictive effect on output, and vice-versa. This is still present in today's mainstream New Keynesian models, but a literature developed in the 1980s and 1990s around the potential for fiscal consolidations to increase output — so-called "expansionary fiscal contractions", or EFCs. This was based on fiscal policies ran in a number of European countries in the 1980s, with a particular focus on West Germany, Denmark and Ireland, and focused on particular conditions it was claimed there was general agreement that policy was unsustainable and needed restricting.

With the Euro Area crisis and the enactment of a pre-emptive deficit reduction programme in the UK after the Great Financial Crisis of the late 2000s, this literature was given a new lease of life. Across both geographies, there was and still is significant discussion about the relative merits and costs of the austerity policies introduced, and there is no consensus regarding how big or small their effect on output was — or even the direction of that effect.

This chapter takes a broad view of history and attempts to quantify what the output cost of tightening and loosening fiscal policy has been over the course of the 140 years between 1879-80 and 2018-19. To do so, I use a novel dataset, which combines the in-year discretionary spending shocks from chapter 3 with new archival data from the UK Parliament cataloguing in-year tax policy changes over the same period, allowing me to construct a series of changes to the discretionary fiscal stance. The UK's fiscal policymaking process makes it more suited to these not being anticipated, as budget policy is conducted within the Treasury, with no divided government and therefore no need for negotiation out in the open about measures. I run the results with the raw shocks and the residuals of Mincer-Zarnowitz regressions, and obtain indistinguishable estimates. The shocks identify a number of very large intra-year changes to the fiscal stance, not only due to wars — the First and Second World Wars being the most significant — but also the People's Budget of 1909, the 1930 and 1931 tightening during the Great Depression, various injections of demand succeeded by restrictive policies in the 1950s, 1960s and 1970s, and two large contractionary budgets in 1981 and 2010.

I estimate that the average effect of a 1% of GDP increase in the fiscal balance is to reduce GDP by 0.24% in the year of impact, with the effect increasing to 0.38% by year 4. This is despite the non-canonically Keynesian effect of household consumption increasing — although not by enough to offset the fall in GDP from government consumption. Allowing coefficients to vary across states of slack supports the idea that fiscal policy is more powerful in times of high slack.

I use a set of measures of fiscal distress, including rapidly accumulating debt, interest rate spikes and rapid increases in the interest burden. My results do not provide evidence that output increases in response to a contractionary fiscal policy, but they do show similar patterns to Giavazzi and Pagano (1990) and other work in the EFC literature from that era, which showed higher increases in consumption as a response to fiscal tightening in times of distress. Finally, I find some, although relatively weak, evidence to support the Alesina et al. (2018) finding of asymmetric effects from tightenings and loosenings of fiscal policy. I find not statistically significant output costs of tightenings, but I do find an expansionary effect from fiscal loosenings.

Appendices

4.A Shocks used in the first stage regressions

Table 4.7 shows the raw values of the shocks used for the estimation process. These are in cash values (£ million) and in nominal terms, and so have been subsequently divided by GDP for normalisation.

Year	Total change	Tax	Spending	Year	Total change	Tax	Spending
1879-80	-3.337	0.000	-3.337	1949-50	-74.645	-22.250	-52.395
1880-81	-0.403	0.700	-1.103	1950-51	70.496	1.800	68.696
1881-82	-0.936	-0.890	-0.046	1951-52	269.321	138.400	130.921
1882-83	-4.434	0.247	-4.681	1952-53	-185.237	-116.625	-68.612
1883-84	-3.152	-2.451	-0.701	1953-54	-182.903	-169.470	-13.433
1884-85	-3.993	-0.022	-3.971	1954-55	211.341	-4.125	215.466
1885-86	9.509	7.500	2.009	1955-56	-0.193	-118.850	118.657
1886-87	-0.216	-0.016	-0.200	1956-57	-57.114	7.500	-64.614
1887-88	-3.452	-3.020	-0.432	1957-58	-165.552	-97.650	-67.902
1888-89	-1.305	-2.460	1.155	1958-59	-107.898	-50.500	-57.398
1889-90	1.182	1.100	0.082	1959-60	-306.933	-295.000	-11.933
1890-91	-3.321	-2.796	-0.525	1960-61	-150.059	21.780	-171.839
1891-92	-0.632	0.000	-0.632	1961-62	-113.000	68.000	-181.000
1892-93	0.047	-0.024	0.071	1962-63	-94.000	-10.000	-84.000
1893-94	1.105	1.750	-0.645	1963-64	-130.000	-269.000	139.000
1894-95	3.019	3.219	-0.200	1964-65	269.000	200.000	69.000
1895-96	-1.501	0.500	-2.001	1965-66	158.000	164.000	-6.000
1896-97	-2.067	-1.275	-0.792	1966-67	301.000	386.000	-85.000
1897-98	-1.892	-0.316	-1.576	1967-68	-494.000	-1.000	-493.000
1898-99	-2.786	-1.505	-1.281	1968-69	838.000	923.000	-85.000
1899-1900	-21.855	0.870	-22.725	1969-70	241.000	457.000	-216.000
1900-01	-20.552	12.317	-32.869	1970-71	-656.000	-139.000	-517.000
1901-02	-0.371	11.000	-11.371	1971-72	-1,437.000	-508.000	-929.000
1902-03	-4.293	5.150	-9.443	1972-73	-2,031.000	-1,130.000	-901.000
1903-04	-13.977	-10.500	-3.477	1973-74	-801.000	-110.000	-691.000
1904-05	4.829	4.550	0.279	1974-75	-4,831.000	-333.000	-4,498.000
1905-06	-0.301	-1.550	1.249	1975-76	-4,150.000	1,193.000	-5,343.000
1906-07	0.673	-2.025	2.698	1976-77	-796.000	-654.000	-142.000
1907-08	-0.514	-1.400	0.886	1977-78	-917.000	-1,023.000	106.000
1908-09	-4.507	-3.420	-1.087	1978-79	-1,501.000	-1,916.000	415.000
1909-10	13.826	14.200	-0.374	1979-80	-2,094.000	-1,152.000	-942.000
1910-11	-0.248	0.000	-0.248	1980-81	-3,404.000	189.000	-3,593.000
1911-12	2.312	-0.095	2.407	1981-82	1,642.000	3,642.000	-2,000.000
1912-13	-5.522	0.000	-5.522	1982-83	-117.000	-2,017.000	1,900.000
1913-14	-1.498	0.000	-1.498	1983-84	-1,970.000	-1,870.000	-100.000
1914-15	-332.497	24.300	-356.797	1984-85	-5,540.000	-40.000	-5,500.000
1915-16	-998.601	3.100	-1,001.701	1985-86	3,745.000	-655.000	4,400.000
1916-17	-1,596.200	75.950	-1,672.150	1986-87	-490.000	-1,290.000	800.000
1917-18	-401.625	26.100	-427.725	1987-88	575.000	-2,325.000	2,900.000
1918-19	415.882	67.800	348.082	1988-89	1,020.000	-4,080.000	5,100.000
1919-20	-184.013	41.450	-225.463	1989-90	-4,110.000	-810.000	-3,300.000
1920-21	-248.838	76.650	-325.488	1990-91	-1,135.000	465.000	-1,600.000
1921-22	-52.129	0.000	-52.129	1991-92	4,345.000	745.000	3,600.000
1922-23	37.816	-45.850	83.666	1992-93	85.000	-2,115.000	2,200.000
1923-24	-4.867	-34.150	29.283	1993-94	310.000	510.000	-200.000
1924-25	-33.838	-34.050	0.212	1994-95	5,725.000	1,525.000	4,200.000
1925-26	-44.427	-24.940	-19.487	1995-96	85.000	-1,315.000	1,400.000
1926-27	13.961	20.050	-6.089	1996-97	-2,620.000	-3,120.000	500.000
1927-28	41.827	37.980	3.847	1997-98	6,820.000	4,600.000	2,220.000
1928-29	12.016	9.435	2.581	1998-99	1,975.000	1,755.000	220.000
1929-30	-16.376	-7.880	-8.496	1999-2000	-605.000	-305.000	-300.000
1930-31	21.859	33.800	-11.941	2000-01	-500.000	-600.000	100.000
1931-32	73.816	74.000	-0.184	2001-02	-720.000	-2,920.000	2,200.000
1932-33	-8.581	2.525	-11.106	2002-03	-32,090.000	10.000	-32,100.000
1933-34	-9.467	-13.953	4.486	2003-04	-935.000	65.000	-1,000.000
1934-35	-31.206	-20.970	-10.236	2004-05	-2,165.000	-265.000	-1,900.000
1935-36	-18.720	-1.110	-17.610	2005-06	1,475.000	1,175.000	300.000
1936-37	17.142	21.775	-4.633	2006-07	5,060.000	360.000	4,700.000
1937-38	37.483	15.150	22.333	2007-08	-2,100.000	300.000	-2,400.000
1938-39	36.933	30.350	6.583	2008-09	-6,095.000	-5,995.000	-100.000
1939-40	-251.593	131.420	-383.013	2009-10	-1,015.000	-915.000	-100.000
1940-41	-990.160	227.163	-1,217.323	2010-11	33,515.000	2,115.000	31,400.000
1941-42	-416.321	150.360	-566.681	2011-12	6,825.000	625.000	6,200.000
1942-43	-193.601	158.000	-351.601	2012-13	13,990.000	-2,710.000	16,700.000
1943-44	59.091	102.500	-43.409	2013-14	5,525.000	-275.000	5,800.000
1944-45	-120.419	3.800	-124.219	2014-15	2,425.000	725.000	1,700.000
1945-46	-201.771	-284.000	82.229	2015-16	11,070.000	1,770.000	9,300.000
1946-47	-45.555	-32.000	-13.555	2016-17	2,655.000	655.000	2,000.000
1947-48	27.494	55.000	-27.506	2017-18	965.000	-35.000	1,000.000
1948-49	-180.832	11.300	-192.132	2018-19	-1,362.401	237.599	-1,600.000

Table 4.7: Annual values of the shocks to the discretionary fiscal stance used in the first stage of the regressions, along with the breakdown into tax and spending. Values are in £ million in current prices.

4.B List of variables used

Variable	ONS code 1946-2018	Thomas and Dimsdale (2017) reference (Sheet.Column) 1878-1945	Notes
Population	UKPOP	A18.C until 1920 A18.D subsequently	Spliced series. Follows political boundaries
Nominal GDP	YBHA	A9.P	Spliced series
Real GDP	ABMI	A8.U	Monetary series calculated from A8.U index for splicing
GDP deflator	n/a	n/a	Calculated from nominal and real GDP
Employment	MGRZ	A50.C	Spliced series. Simplified assumption of employment divided by population to obtain employment rate
Official Bank Rate	n/a	n/a	Obtained from Bank of England (official history since 1694)
Unemployment	MGSC	A50.G	Used to calculate unemployment rate
Consumption	ABPF + ABNU	A12.F	Households + non-profits serving households. Spliced series
National debt	HF6W	A29.AJ	Spliced series
Central government net lending (budget balance)	NMFJ	A27.AO	Spliced series
Exports	IKBL	A35.G	Spliced series
Imports	IKBK	A35.M	Spliced series
Gross fixed capital formation	NPQT	A12.K	Spliced series
Short-term interest rate	n/a	n/a	JST dataset spliced with Officer (2024)
Central government debt interest	NMFX	A27.H	Spliced series

Table 4.8: List of variables used in the estimation process.

4.C Regression results for the linear case

The specification in equation (4.1), as discussed in chapter 3, essentially amounts to an instrumental variables-augmented implementation of an autoregressive distributive lag (ARDL) model. This is a particularly useful setup in this chapter as tests indicate that output per capita is an I(1) series, whereas the fiscal balance is I(0). An ARDL-type model is particularly suited for situations where the two main variables have different orders of integration.

Table 4.9 summarises the regression results for the linear case when using the discretionary fiscal stance shock as instrument for $\sum_h BAL_{t+j}$. These are the outputs of the estimation of equation (4.1) across each horizon h, with the coefficient on $\sum_h BAL_{t+j}$ being the effect of a 1% of GDP increase in the fiscal balance on output — the main coefficient of interest. Tables 4.10 to 4.13 show the effects on consumption, policy interest rates, the employment rate and inflation.

			Horizon h		
$\sum_h \ln y_{t+j}$	0	1	2	3	4
$\sum_h BAL_{t+j}$	-0.2370***	-0.2843***	-0.3306***	-0.3655***	-0.3735***
	(0.0793)	(0.0817)	(0.0904)	(0.1073)	(0.1310)
$\ln y_{t-1}$	1.0820***	2.2037***	3.3576***	4.5652***	5.8360***
	(0.0266)	(0.0758)	(0.1506)	(0.2360)	(0.3432)
$\ln BAL_{t-1}$	0.2977***	0.7551***	1.3324***	1.9733***	2.6147***
	(0.0768)	(0.1485)	(0.2276)	(0.3108)	(0.4013)
r_{t-1}	-1.3459**	-1.0817**	-2.2956**	-3.2555***	-4.4546***
	(0.2050)	(0.5341)	(0.9144)	(1.2684)	(1.5509)
π_{t-1}	-0.0451	-0.0319	0.0368	0.1177	0.1883
	(0.0559)	(0.1507)	(0.2679)	(0.4084)	(0.5546)
e_{t-1}	-0.0290**	-0.0881**	-0.1777**	-0.2875**	-0.3996**
	(0.0126)	(0.0368)	(0.0820)	(0.1173)	(0.1693)
$\ln c_{t-1}$	-0.0755***	-0.1815**	-0.3081**	-0.4806	-0.7156**
	(0.0273)	(0.0777)	(0.1545)	(0.2422)	(0.3524)
d_{t-1}	0.0061*	0.0233**	0.0545***	0.0971***	0.1452***
	(0.0034)	(0.0101)	(0.0212)	(0.0328)	(0.0465)
Constant	4.8906	14.1050	25.4733	37.3151	47.8284
	(3.4900)	(10.5537)	(22.8565)	(35.7568)	(52.2310)
N	140	139	138	137	136
Kleibergen-Paap F-stat	80.459	70.718	50.658	31.865	19.839
AIC	597.187	850.181	996.724	1092.821	1157.927
SBIC	623.662	876.591	1023.069	1119.101	1184.141
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.9: Regression estimates for output at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

	Horizon h				
$\sum_h \ln c_{t+j}$	0	1	2	3	4
$\sum_h BAL_{t+j}$	0.3164***	0.4166***	0.4875***	0.5461***	0.6266***
	(0.0650)	(0.0817)	(0.0718)	(0.0877)	(0.1138)
$\ln y_{t-1}$	0.1390***	0.3944***	0.7240***	1.1182***	1.5458***
	(0.0236)	(0.0655)	(0.1296)	(0.2121)	(0.3203)
$\ln BAL_{t-1}$	-0.2347***	-0.5483***	-0.8260***	-1.0415***	-1.2466***
	(0.0634)	(0.1187)	(0.1853)	(0.2644)	(0.3612)
r_{t-1}	-0.2589	-0.7092*	-1.2120**	-1.8056*	-2.4864**
	(0.1637)	(0.4012)	(0.6799)	(0.9627)	(1.2337)
π_{t-1}	-0.1593***	-0.4251***	-0.7115***	-0.9807***	-1.2156**
	(0.0479)	(0.1272)	(0.2300)	(0.3547)	(0.5094)
e_{t-1}	-0.0268**	-0.0703**	-0.1262*	-0.2003*	-0.2838*
	(0.0113)	(0.0319)	(0.0652)	(0.1069)	(0.1567)
$\ln c_{t-1}$	0.8663***	1.6182***	2.3002***	2.9230***	3.5169***
	(0.0243)	(0.0674)	(0.1335)	(0.2182)	(0.3293)
d_{t-1}	0.0003	-0.0029	-0.0072	-0.0107	-0.0153
	(0.0031)	(0.0089)	(0.0183)	(0.2182)	(0.0438)
Constant	3.3197	9.4302	16.2181	25.3259	35.3133
	(3.2277)	(9.3428)	(19.7621)	(32.9781)	(49.1578)
N	140	139	138	137	136
Kleibergen-Paap F-stat	80.458	70.680	50.601	31.817	19.779
AIC	556.567	799.389	943.512	1044.681	1126.117
SBIC	583.041	825.799	969.858	1070.960	1152.331
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.10: Regression estimates for consumption at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

	Horizon h				
$\sum_h r_{t+j}$	0	1	2	3	4
$\sum_h BAL_{t+j}$	-0.0253	-0.0239	-0.0195	-0.0232	-0.0346**
	(0.0360)	(0.0273)	(0.0204)	(0.0169)	(0.0167)
$\ln y_{t-1}$	0.0202**	0.0615***	0.0958***	0.1145***	0.1433***
	(0.0089)	(0.0201)	(0.0286)	(0.0357)	(0.0455)
$\ln BAL_{t-1}$	0.0289	0.0548	0.0692	0.1022**	0.1479***
	(0.0347)	(0.0482)	(0.0483)	(0.0472)	(0.0514)
r_{t-1}	0.2907**	-0.0438	-0.3636	-0.5708***	-0.4988***
	(0.1190)	(0.2115)	(0.2249)	(0.2004)	(0.1910)
π_{t-1}	-0.0088	-0.0270	-0.0247	0.0162	0.0309
	(0.0214)	(0.0435)	(0.0549)	(0.0184)	(0.0680)
e_{t-1}	-0.0032	-0.0112	-0.0202	-0.0262	-0.0341
	(0.0041)	(0.0099)	(0.0146)	(0.0184)	(0.0231)
$\ln c_{t-1}$	-0.0217**	-0.0661***	-0.1035***	-0.1250***	-0.1560***
	(0.0089)	(0.0203)	(0.0290)	(0.0362)	(0.0462)
d_{t-1}	0.0003	0.0006	0.0012	0.0026	0.0051
	(0.0010)	(0.0025)	(0.0037)	(0.0048)	(0.0062)
Constant	1.8370*	6.3666**	11.4197***	15.5331***	19.3000***
	(1.0505)	(2.6555)	(5.4044)	(32.9781)	(7.1446)
N	140	139	138	137	136
Kleibergen-Paap F-stat	80.276	70.557	50.494	31.889	19.853
AIC	399.639	538.521	566.743	569.840	587.601
SBIC	426.114	564.932	593.089	596.120	613.815
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.11: Regression estimates for the policy interest rate at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h e_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h BAL_{t+j}$	-0.4381*	-0.5344**	-0.6917**	-0.9619***	-1.2456***
	(0.2584)	(0.2652)	(0.2861)	(0.3253)	(0.3824)
$\ln y_{t-1}$	0.0576	0.1698	0.2102	0.2875	0.5400
	(0.0929)	(0.2672)	(0.4974)	(0.7610)	(1.0716)
$\ln BAL_{t-1}$	0.2199	0.4323	0.8481	1.6270*	2.5531**
	(0.2509)	(0.4864)	(0.7203)	(0.9579)	(1.2025)
r_{t-1}	0.5308	0.0562	-1.4878	-3.1456	-4.9719
	(0.6228)	(1.5639)	(2.5562)	(3.4501)	(4.1305)
π_{t-1}	-0.2905	-0.8063	-1.2505	-1.7242	-2.4062
	(0.1925)	(0.5282)	(0.9218)	(1.2910)	(1.6239)
e_{t-1}	0.9014***	1.6592***	2.2868***	2.7634***	3.1269***
	(0.0445)	(0.1306)	(0.2477)	(0.3839)	(0.5328)
$\ln c_{t-1}$	-0.0273	-0.0712	-0.0004	0.0783	0.0037
	(0.0955)	(0.2748)	(0.5114)	(0.7801)	(1.0951)
d_{t-1}	0.0093	0.0388	0.1007	0.2065**	0.3416**
	(0.0010)	(0.1048)	(0.0670)	(0.1048)	(0.1471)
Constant	15.1620	57.7323	121.3121*	207.0446*	313.3744*
	(12.2872)	(36.9729)	(72.3683)	(116.0389)	(166.2218)
N	140	139	138	137	136
Kleibergen-Paap F-stat	80.276	70.557	50.494	31.889	19.853
AIC	925.475	1175.441	1311.388	1391.159	1441.708
SBIC	951.950	1201.852	1337.733	1417.438	1467.922
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.12: Regression estimates for the employment rate at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

$\sum_h \pi_{t+j}$	Horizon h 0	1	2	3	4
$\sum_h BAL_{t+j}$	-0.4189***	-0.4299***	-0.4733***	-0.5972***	-0.7269***
	(0.1029)	(0.0951)	(0.0943)	(0.1090)	(0.1392)
$\ln y_{t-1}$	0.1106***	0.3679***	0.6978***	1.1297***	1.7017***
	(0.0341)	(0.0935)	(0.1654)	(0.2645)	(0.4058)
$\ln BAL_{t-1}$	0.2889***	0.4756***	0.7259***	1.1998***	1.7372***
	(0.1009)	(0.1766)	(0.2418)	(0.3295)	(0.4503)
r_{t-1}	0.6652**	1.2250*	1.2522	1.2146	1.0876
	(0.2822)	(0.6297)	(0.9176)	(1.1891)	(1.4505)
π_{t-1}	0.6871***	1.0098***	1.2042***	1.3313***	1.3378**
	(0.0685)	(0.1747)	(0.2896)	(0.4343)	(0.6231)
e_{t-1}	-0.0519***	-0.1482***	-0.2532***	-0.3761***	-0.5026**
	(0.0161)	(0.0452)	(0.1698)	(0.1320)	(0.1987)
$\ln c_{t-1}$	-0.0992***	-0.3406***	-0.6586***	-1.0821***	-1.6561***
	(0.0348)	(0.0959)	(0.1698)	(0.2715)	(0.4164)
d_{t-1}	0.0076	0.0178	0.0297	0.0500	0.0722
	(0.0046)	(0.0131)	(0.0234)	(0.0373)	(0.0559)
Constant	8.1815*	26.2578*	48.4187*	75.8342*	107.0774*
	(4.7737)	(13.6992)	(25.1223)	(40.8736)	(62.6345)
N	140	139	138	137	136
Kleibergen-Paap F-stat	80.287	70.646	50.555	31.840	19.687
AIC	691.739	901.424	1008.253	1095.937	1174.679
SBIC	718.213	927.834	1034.598	1122.217	1200.893
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.13: Regression estimates for the inflation rate at each horizon h. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

4.D Results under the combined measure of distress

Table 4.14 shows the raw figures used to calculate the combined measure of distress. These are the 5-year cumulative change in the debt-to-GDP ratio, the 3-year cumulative change in the short-term interest rate and the 3-year cumulative change in the interest burden (measured as debt interest payments as a share of GDP).

Year	5-yr Δ debt	3-yr Δ int. rate	3-yr Δ int. burden	Combined measure	Year	5-yr Δ debt	3-yr Δ int. rate	3-yr Δ int. burden	Combined measure
1882-83	-4.56	1.37*	-0.12	1	1951-52	-85.18	-0.27	-0.51	0
1883-84	-6.49	0.43	-0.07	0	1952-53	-77.10	0.65	-0.21	0
1884-85	-6.07	-0.60	-0.09	0	1953-54	-60.91	0.63	-0.16	0
1885-86	-1.49	-1.06	-0.21	0	1954-55	-60.61	0.55	-0.22	0
1886-87	-1.29	-0.58	-0.13	0	1955-56	-59.35	1.15*	-0.27	1
1887-88	-3.94	0.13	-0.23	0	1956-57	-49.55	1.92*	-0.39	1
1888-89	-6.06	0.35	-0.19	0	1957-58	-43.00	1.60*	-0.41	1
1889-90	-9.78	0.74	-0.38	0	1958-59	-40.78	-0.40	-0.31	0
1890-91	-11.98	0.98*	-0.31	1	1959-60	-38.10	-1.47	-0.19	0
1891-92	-10.96	-0.14	-0.19	0	1960-61	-31.97	-0.07	0.03	0
1892-93	-6.17	-1.32	-0.06	0	1961-62	-26.16	0.25	-0.14	0
1893-94	-4.46	-1.53	-0.02	0	1962-63	-21.19	-0.26	-0.19	0
1894-95	-4.22	-1.34	-0.09	0	1963-64	-21.52	-1.58	-0.32	0
1895-96	-4.47	-0.57	-0.15	0	1964-65	-24.06	-0.60	-0.43	0
1896-97	-5.86	-0.26	-0.14	0	1965-66	-22.49	1.11*	-0.36	1
1897-98	-7.86	1.05*	-0.10	1	1966-67	-21.52	1.22*	-0.29	1
1898-99	-9.18	1.77*	-0.13	1	1967-68	-20.48	0.37	-0.05	0
1899-1900	-8.39	1.70*	-0.22	1	1968-69	-22.35	0.34	0.04	0
1900-01	-4.54	1.47*	-0.32	1	1969-70	-23.21	0.36	-0.16	0
1901-02	0.03	0.29	-0.17	0	1970-71	-25.75	-0.86	-0.41	0
1902-03	3.11	-0.25	0.12	0	1971-72	-24.73	-3.27	-0.44	0
1903-04	5.40	-0.31	0.25	0	1972-73	-29.77	-2.99	-0.22	0
1904-05	6.74	-0.47	0.13	0	1973-74	-25.10	0.28	-0.03	0
1905-06	2.57	-0.17	-0.12	0	1974-75	-11.37	1.15*	0.10	1
1906-07	-2.25	0.87*	-0.25	1	1975-76	-4.17	-2.08	0.18	0
1907-08	-5.11	1.24*	-0.27	1	1976-77	-3.71	-6.30	0.49	0
1908-09	-4.62	-0.63	-0.17	0	1977-78	-2.43	-9.36	0.52*	1
1909-10	-4.81	-1.62	-0.09	0	1978-79	-1.99	-6.25	0.50*	1
1910-11	-5.75	-0.86	-0.08	0	1979-80	-10.44	-3.10	0.35	0
1911-12	-5.78	0.73	-0.15	0	1980-81	-11.13	-1.10	0.46	0
1912-13	-5.93	1.23*	-0.15	1	1981-82	-9.47	-4.81	0.63*	1
1913-14	-7.73	0.84*	-0.12	1	1982-83	-7.34	-8.43	0.55*	1
1914-15	8.53	-0.10	0.21	0	1983-84	-5.37	-8.54	0.31	0
1915-16	37.28*	-0.03	1.28*	1	1984-85	-2.02	-5.56	0.24	0
1916-17	79.41*	0.52	2.64*	1	1985-86	-4.15	-2.86	0.14	0
1917-18	102.15*	0.44	3.21*	1	1986-87	-5.58	-2.41	0.03	0
1918-19	115.68*	-1.64	3.03*	1	1987-88	-7.76	-2.93	-0.31	0
1919-20	96.02*	-2.07	2.30*	1	1988-89	-13.19	-2.99	-0.55	0
1920-21	63.10*	0.31	1.60*	1	1989-90	-16.40	-0.91	-0.64	0
1921-22	56.27*	0.27	1.93*	1	1990-91	-16.19	-1.58	-0.78	0
1922-23	51.86*	-1.13	1.80*	1	1991-92	-13.26	-4.98	-0.82	0
1923-24	42.27*	-2.13	1.64*	1	1992-93	-5.48	-7.85	-0.65	0
1924-25	40.11*	-0.02	0.24	1	1993-94	5.02	-8.25	-0.26	0
1925-26	43.53*	1.70*	-0.33	1	1994-95	11.13*	-5.11	0.31	1
1926-27	12.85*	1.63*	-0.12	1	1995-96	14.90*	-2.90	0.51*	1
1927-28	-11.78	0.72	-0.23	0	1996-97	14.49*	-1.30	0.41	1
1928-29	-15.67	0.31	-0.24	0	1997-98	10.15*	-1.00	0.20	1
1929-30	-12.02	0.20	-0.64	0	1998-99	4.06	-1.30	-0.15	0
1930-31	-3.98	-1.32	-0.34	0	1999-2000	-1.79	-1.56	-0.53	0
1931-32	4.82	-1.14	0.00	0	2000-01	-8.08	-1.52	-0.73	0
1932-33	20.04*	-3.02	-0.15	1	2001-02	-9.05	-1.92	-0.86	0
1933-34	20.28*	-1.92	-1.31	1	2002-03	-6.78	-1.70	-0.69	0
1934-35	10.77*	-2.25	-2.15	1	2003-04	-4.21	-2.00	-0.52	0
1935-36	-2.03	-0.68	-2.05	0	2004-05	0.96	-1.10	-0.16	0
1936-37	-23.06	-0.09	-1.09	0	2005-06	6.26	-0.82	0.12	0
1937-38	-34.92	-0.21	-0.67	0	2006-07	7.20	-0.91	0.16	0
1938-39	-37.69	0.10	-0.53	0	2007-08	5.79	-1.39	0.24	0
1939-40	-28.28	0.42	-0.55	0	2008-09	18.04*	-2.74	0.12	1
1940-41	-12.52	-0.07	-0.84	0	2009-10	31.28*	-4.46	0.14	1
1941-42	4.80	-0.51	-0.78	0	2010-11	36.80*	-4.88	0.82*	1
1942-43	25.79*	-1.01	-0.21	1	2011-12	39.45*	-4.07	1.01*	1
1943-44	48.77*	-0.55	0.63*	1	2012-13	42.59*	-2.67	0.73*	1
1944-45	79.96*	-0.38	1.17*	1	2013-14	31.27*	-2.33	-0.15	1
1945-46	89.68*	-0.46	1.37*	1	2014-15	16.96*	-1.55	-0.50	1
1946-47	96.18*	-0.79	1.20*	1	2015-16	10.36*	-0.84	-0.51	1
1947-48	59.02*	-0.91	0.51*	1	2016-17	9.95	-0.62	-0.28	0
1948-49	16.02*	-0.88	-0.46	1	2017-18	5.18	-1.00	0.08	0
1949-50	-20.48	-0.57	-0.92	0	2018-19	1.10	-1.41	0.06	0
1950-51	-50.68	-0.53	-0.90	0					

* - threshold exceeded for individual measure of distress

Table 4.14: Indicators of distress and combined measure. The thresholds used are 10 per cent of GDP for the 5-year cumulative increase in the debt-to-GDP ratio; 0.75 percentage points for the 3-year cumulative increase in the short-term interest rate; and 0.5 per cent of GDP for the 3-year cumulative increase in the interest burden as a share of GDP. The combined measure of distress is 1 if any of the three thresholds are met, and 0 otherwise.

Tables 4.15 and 4.16 show the regression results for high and low distress states using the combined measure of distress. I find no statistically significant difference for the output effect of a 1% tightening or loosening of the fiscal balance, although I do find a stronger consumption response in times of high distress at short horizons.

	Horizon h				
$\sum_h \ln y_{t+j}$	0	1	2	3	4
$I_{t-1}\sum_h BAL_{t+j}$	-0.1437	-0.2443**	-0.3258***	-0.3697***	-0.384**
	(0.1085)	(0.1003)	(0.1085)	(0.1327)	(0.1647)
$(1-I_{t-1})\sum_h BAL_{t+j}$	-0.2128***	-0.2276***	-0.2512**	-0.2925**	-0.2961*
	(0.0808)	(0.0862)	(0.1004)	(0.1225)	(0.1553)
I_{t-1}	16.0450**	24.5622	17.4838	-6.6933	-36.5581
	(8.1416)	(22.2965)	(42.6360)	(66.0035)	(95.6150)
$I_{t-1}\ln y_{t-1}$	1.0689***	2.1429***	3.2416***	4.3942***	5.6070***
	(0.0372)	(0.1023)	(0.1917)	(0.3027)	(0.4266)
$(1-I_{t-1})\ln y_{t-1}$	1.0949***	2.2301***	3.3947***	4.6550***	5.9801***
	(0.0381)	(0.1101)	(0.2278)	(0.3843)	(0.5824)
$I_{t-1}BAL_{t-1}$	0.1819*	0.6043***	1.1531***	1.7318***	2.3261***
	(0.1053)	(0.1820)	(0.2601)	(0.3528)	(0.4455)
$(1-I_{t-1})BAL_{t-1}$	0.2653**	0.6686***	1.2568***	2.0693***	2.8372***
	(0.1060)	(0.2410)	(0.4246)	(0.6535)	(0.8902)
$I_{t-1}r_{t-1}$	0.0293	-0.1134	0.1645	0.7076	0.8049
	(0.3225)	(0.8401)	(1.4628)	(2.1165)	(2.7824)
$(1-I_{t-1})r_{t-1}$	-0.5800**	-1.8968***	-3.4305***	-4.9508***	-6.5002***
	(0.2311)	(0.5964)	(1.0182)	(1.4282)	(1.8433)
$I_{t-1}\pi_{t-1}$	-0.0445	0.0094	0.1324	0.2702	0.4058
	(0.0639)	(0.1698)	(0.3031)	(0.4526)	(0.6131)
$(1-I_{t-1})\pi_{t-1}$	-0.0418	0.0333	0.2101	0.2750	0.2331
	(0.0878)	(0.2432)	(0.4560)	(0.7049)	(1.0285)
$I_{t-1}e_{t-1}$	-0.0062***	-0.1697***	-0.2996***	-0.4313***	-0.5473***
	(0.0173)	(0.0478)	(0.0902)	(0.1461)	(0.1987)
$(1-I_{t-1})e_{t-1}$	0.0150	0.0311	0.0199	-0.0413	-0.1325
	(0.0181)	(0.0513)	(0.1008)	(0.1640)	(0.2328)
$I_{t-1}\ln c_{t-1}$	-0.0525	-0.1108	-0.1269	-0.2107	-0.3567
	(0.0379)	(0.1044)	(0.1958)	(0.3086)	(0.4369)
$(1-I_{t-1})\ln c_{t-1}$	-0.1019***	-0.2496**	-0.4220*	-0.6780*	-0.9869*
	(0.0380)	(0.1103)	(0.2301)	(0.3889)	(0.5853)
$I_{t-1}d_{t-1}$	0.0069	0.0353**	0.0877***	0.1499***	0.2072***
	(0.0059)	(0.0163)	(0.0311)	(0.0505)	(0.0686)
$(1-I_{t-1})d_{t-1}$	0.0016	0.0061	0.0172	0.0390	0.0785
	(0.0048)	(0.0134)	(0.0265)	(0.0434)	(0.0577)
Constant	-2.9840	-2.3337	5.8934	22.4491	39.5958
	(4.6139)	(13.5345)	(28.7140)	(47.3115)	(66.5191)
Difference in coefficients	0.0691	-0.0167	-0.0747	-0.0772	-0.0887
Test statistic	0.37	0.03	0.48	0.36	0.31
p-value	0.545	0.868	0.489	0.550	0.578
N	136	135	134	133	132
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.15: Regression estimates for output at each horizon h for the combined measure of distress. I represents times of high distress and $(1-I)$ represents low distress. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

			Horizon h		
$\sum_h \ln c_{t+j}$	0	1	2	3	4
$I_{t-1} \sum_h BAL_{t+j}$	0.4822***	0.5498***	0.5858***	0.6220***	0.7258***
	(0.0964)	(0.0906)	(0.0980)	(0.1195)	(0.1586)
$(1 - I_{t-1}) \sum_h BAL_{t+j}$	0.1351	0.2545***	0.3771***	0.4881***	0.5754***
	(0.0730)	(0.0789)	(0.0908)	(0.1104)	(0.1487)
I_{t-1}	7.3042	11.6608	13.2971	8.7438	1.1254
	(7.6822)	(21.5554)	(41.0916)	(64.9085)	(92.8709)
$I_{t-1} \ln y_{t-1}$	0.1699***	0.4549***	0.7923***	1.1874***	1.6486***
	(0.0344)	(0.0961)	(0.1770)	(0.2795)	(0.4209)
$(1 - I_{t-1}) \ln y_{t-1}$	0.1467***	0.3991***	0.6996***	1.0426***	1.4063**
	(0.0350)	(0.1033)	(0.2081)	(0.3484)	(0.5788)
$I_{t-1} BAL_{t-1}$	-0.3761***	-0.7282***	-0.9697***	-1.1088***	-1.2709***
	(0.0942)	(0.1673)	(0.2421)	(0.3268)	(0.4400)
$(1 - I_{t-1}) BAL_{t-1}$	-0.0117	-0.1499	-0.4619	-0.9433*	-1.4417*
	(0.0935)	(0.2147)	(0.3729)	(0.5696)	(0.8635)
$I_{t-1} r_{t-1}$	-0.0662	-0.2267	-0.2995	-0.3713	-0.6999
	(0.2939)	(0.7784)	(1.3554)	(1.9622)	(2.6258)
$(1 - I_{t-1}) r_{t-1}$	-0.2985	-1.0015*	-1.9124**	-2.7988**	-3.5343**
	(0.1993)	(0.5168)	(0.8816)	(1.2480)	(1.6483)
$I_{t-1} \pi_{t-1}$	-0.2237***	-0.6335***	-1.0806***	-1.4253***	-1.7484***
	(0.0605)	(0.1641)	(0.2926)	(0.4352)	(0.5978)
$(1 - I_{t-1}) \pi_{t-1}$	-0.1871**	-0.3857*	-0.4885	-0.5763	-0.6524
	(0.0794)	(0.2212)	(0.4095)	(0.6452)	(0.9654)
$I_{t-1} e_{t-1}$	-0.0282*	-0.0534	-0.0850	-0.1488	-0.2313
	(0.0160)	(0.0448)	(0.0830)	(0.1308)	(0.2017)
$(1 - I_{t-1}) e_{t-1}$	-0.0394**	-0.1134**	-0.1987**	-0.2844*	-0.3852*
	(0.0168)	(0.0484)	(0.0934)	(0.1501)	(0.2316)
$I_{t-1} \ln c_{t-1}$	0.8276***	1.5335***	2.1909***	2.8084***	3.3667***
	(0.0350)	(0.0977)	(0.1799)	(0.2853)	(0.4294)
$(1 - I_{t-1}) \ln c_{t-1}$	0.8652***	1.6322***	2.3545***	3.0308***	3.6888***
	(0.0349)	(0.1029)	(0.2087)	(0.3491)	(0.5797)
$I_{t-1} d_{t-1}$	-0.0047	-0.0191	-0.0331	-0.0390	-0.0540
	(0.0053)	(0.0147)	(0.0275)	(0.0347)	(0.0687)
$(1 - I_{t-1}) d_{t-1}$	-0.0026	-0.0065	-0.0090	-0.0105	-0.0098
	(0.0043)	(0.0123)	(0.0233)	(0.0369)	(0.0584)
Constant	2.8642	11.6217	22.1419	35.8509	55.1712
	(4.2117)	(12.5166)	(25.7988)	(41.9347)	(65.9234)
Difference in coefficients	0.3471	0.2953	0.2086**	0.1339	0.1503
Test statistic	11.36	10.43	4.95	1.48	1.06
p-value	<0.001	0.001	0.026	0.225	0.304
N	136	135	134	133	132
Significance: *** - 1%; ** - 5%; * - 10%					

Table 4.16: Regression estimates for consumption at each horizon h for the combined measure of distress. I represents times of high distress and $(1 - I)$ represents low distress. Numbers in brackets are HAR standard errors, calculated using the QS kernel. Significance calculated for the endogenous variable on the basis of the inverted Anderson-Rubin confidence test and using the t-test otherwise.

Chapter 5

Conclusion

This thesis presents some contributions to the empirical literature on the effects of fiscal policy, both on methodological issues and in applications to UK data. In this conclusion, I summarise the key contributions of my work, their implications for practice and the public policy debate — particularly for those in independent fiscal institutions — and some unanswered questions on which further research might be desirable.

5.1 Key contributions

The research contained in this thesis contains several new contributions.

Chapter 2 documents a detailed assessment of the difficulties in estimating fiscal multipliers when using Gordon and Krenn's (2010) proposed transformation of dividing both output and government spending by potential output. This is because largely indistinguishable potential output estimates can lead to widely varying multiplier estimates, from as low as -0.04 to as high as 0.70 — huge parameter uncertainty. This leads me to reassess the merits of applying the previously commonplace method of estimating an output elasticity with respect to government spending, which is then converted into a multiplier by using the sample average of Y/G — that is, the inverse of the share of government spending on output.

There are two main reasons why I recommend returning to the conversion ratio approach. The first is that I find the dispersion of potential different Y/G ratios is considerably narrower than that associated with different methods of estimating potential output. The second is related to the point about dispersion, as the conversion ratio is both more transparent and less easily influenced by choices by the econometrician, especially when they are relatively small and seemingly inconsequential.

Chapter 3 contributes to the results in the applied literature, especially for that of the UK. An important way in which it does so is to make use of a novel database consisting of 140 years of intra-year shocks to discretionary government in the UK. This is the result of my own archival research by comparing the estimate for government spending at the start of the financial year with the estimate in the subsequent budget, and which I have compiled going back to 1879-80.

Britain's idiosyncratic budget process, featuring virtually no public negotiation prior to its announcement to the House of Commons, as well as the executive dominance underpinning its government and parliamentary system — which makes being able to pass a budget a pre-condition of the government being in place — make the UK an ideal candidate for the use of this identification strategy, as the shocks are unlikely to be anticipated. This is an assertion that is strengthened by statistical testing, and this series of shocks provides the exogenous variable that I use to instrument government spending in estimating the fiscal multiplier.

This contribution is particularly important because so much of the fiscal multipliers literature is US-centric, especially when it comes to the narrative approach. Many seminal papers have used narrative series to identify fiscal shocks — Ramey and Shapiro (1998), Romer and Romer (2010), Ramey (2011a), Ramey and Zubairy (2018) to name but a few — but these are costly to assemble and therefore have generally been limited to the US. On the revenue side, Cloyne (2013) has estimated UK-specific tax multipliers using the narrative record of post-war tax changes, but no analogue exists on the spending side; the few recent UK-focused papers of Rafiq (2014), Glocker et al. (2019) and Shaheen and Turner (2020) are all data-driven rather than narrative. This lack of a UK-focused literature has meant that in many cases, official bodies such as the Office for Budget Responsibility (OBR) have to rely on US-based estimates to make forecasting assumptions regarding the effects of government spending.

The results in chapter 3 provide a number of new findings, including the shape of the cumulative UK multiplier, which I estimate to reach 0.44 in-year, before peaking at 0.53 two years ahead and falling to a long-run value of 0.47. This implies that the main effect of a government spending shock on output occurs on impact, which is an important finding and consideration for policymakers.

The findings on the effect of government spending shocks on other macroeconomic variables also contribute to the general literature and in particular add to the UK-specific literature. I estimate that an increase in government spending is associated with a fall in household consumption, a result that bears out both the insights of the workhorse New Keynesian model and empirical results from studies that also have large military spending shocks, such as Ramey (2009, 2012). I also find positive effects on employment and inflation — again, corroborating the New Keynesian model and parts of the empirical literature — but generally small and insignificant effects on the policy rate, implying accommodation by the monetary authority of government spending shocks through higher inflation in the time period.

The other way in which chapter 3 contributes to the literature is by examining whether different shocks, economic conditions and broader measures of regimes are associated with different magnitudes of multipliers. I find some evidence of multipliers from civil spending shocks being larger than those from military spending at short horizons — but the civil spending shocks do not have enough power to reliably predict long-run effects. I also find some evidence of higher multipliers in states of high slack, but no statistically significant differences in estimates for different regimes such as high and low debt, higher or lower openness to trade and flexible or fixed exchange rates.

Chapter 4 contributes to the literature by taking a wider view of the effect of government decisions on output, as I consider changes to the fiscal stance through the budget balance rather than simply increasing or decreasing government spending. This is a particular contribution to the evidence base regarding historical UK economic outcomes. I once again use archival research to compile a series of shocks to the discretionary fiscal stance — that is, changes to the budget balance excluding debt interest and social security spending. To do so, I combine the shocks to government spending used in chapter 3 with further data on tax shocks, these coming from changes to policy announced that come in immediately after a statement to the House of Commons or in the middle of the financial year.

Having compiled that series allows me to estimate the effect of a change in the fiscal balance on output, a formulation that is closely related to the literature on expansionary fiscal contractions and under what conditions fiscal tightening is less or more costly. The results in chapter 4 quantify the effect of a 1% of GDP increase in the fiscal balance as a reduction in output of around 0.23% on impact, increasing to 0.37% in the long run. On the other hand, my estimates of the effect of such a tightening on household consumption are that the latter increases by 0.32% in-year, rising to 0.63% in the long run — corroborating the insights from the EFC literature on private activity, while also showing that this increase in activity is not large enough to compensate for the fall in GDP from lower government spending.

There are three main results that add to the literature from this chapter. One is the finding that output responds more strongly to changes in the fiscal balance in times of high slack (both measured by employment and the output gap), validating the Keynesian view of fiscal policy being more effective when there is spare capacity. The second is the finding that private consumption increases more strongly in response to a fiscal tightening in times of fiscal distress than in normal times, corroborating Giavazzi and Pagano's (1990) results to a large extent. And finally, I find some — though relatively weak — evidence of asymmetric effects of fiscal stance changes on output, as fiscal tightenings might not have as large an impact on GDP as fiscal loosenings.

5.2 Implications for practice and the public policy debate

There are two main audiences for whom the results and insights of this thesis might have important implications.

The first is for practitioners in applied fiscal policy research, particularly those engage in estimating fiscal multipliers. As I mentioned in chapter 1, there have been very important and unambiguously positive methodological steps over the last couple of decades in the field, including now broadly accepted definitions of the multiplier and widespread use of local projections as a more flexible way of modelling non-linear effects.

However, my results in chapter 2 raise questions as to whether the widespread application of the Gordon and Krenn (2010) transformation of dividing both output and government by potential output, which has become commonplace since then, is an unequivocal advance. My results point to this seemingly small and innocuous modelling decision having potentially very large effects on multiplier estimates while using the same data and with different potential output estimation methods being very similar in fit. This could point to the value in revisiting the use of such specifications in the future, and I suggest that returning to a conversion ratio of Y/G might be both more transparent and leave fewer degrees of freedom for the econometrician to influence the results.

The second audience for which the results in this thesis have potentially important implications is the public policy and official institutions in the UK, including policymakers. The literature on UK multipliers is much more sparse than that for the US, with narrative approach studies being restricted to Cloyne's (2013) tax multipliers work. The results in chapter 3 provide a solid evidence base for the effect of government spending on output, being the expenditure analogue to the tax work and supplying UK-based estimates that are more directly relevant for use in a British context than having to repurpose US-based estimates — especially when the two economies are so different, and so is their relative reliance and ability to use fiscal policy. They also align with the annual, financial year basis on which forecasts are produced by the OBR, therefore providing a direct read-across for how they could be implemented in the OBR's multipliers framework.

Finally, the results from chapter 4 would appear to lend further weight to several 2010s studies (e.g. Wren-Lewis, 2016; Jordà and Taylor, 2016) that pointed towards the austerity programme introduced in 2010 having a negative effect on growth, particularly given the very high historical levels of unemployment the UK had at that time — which I find are associated with higher output costs. The 2% of GDP tightening announced for 2010-11, at a point when unemployment was over 7%, might have cost as much as 1.7% of GDP in that year and over 3% of cumulative GDP three years after — a huge cost, and an important finding that will be important to bear in mind should a similar crisis occur in future.

5.3 Areas for future research

There are a number of avenues opened up by the research presented in this thesis and which might be productively pursued.

The empirical results in chapter 2 point to a large sensitivity of government spending multiplier estimates from different choices of method of estimating potential output. The methodology applied is consistent across the different potential output methods and the conversion ratio approach, but it would be interesting to characterise the variation in estimates in an environment with a simulated data generation process, which might shed more light on the drivers of that variation.

In chapters 2 and 3, I have opted for a conversion ratio approach, which takes the previously commonplace approach of estimating an output elasticity with respect to government spending and multiplying it by a factor that allows the retrieval of the effect of a 1% of GDP increase in government spending on output. However, in chapter 4 I have used a slightly tweaked approach, as I have used the fiscal balance as a percentage of GDP as the endogenous variable and the log of GDP as the dependent variable. This means that the coefficient retrieved the from the regression outputs in chapter 4 is already essentially measured in the same unit as the dependent variable. This might be usefully applied to the methodologies in chapters 2 and 3, and it would be interesting to see how it compares with the approach I have taken.

My archival research approach has naturally been limited by the breakdowns that existed consistently throughout the whole span of time I have collected data over. This is a natural trade-off given my need to have a large enough sample size when using annual data — a decision that itself is intrinsically linked to my identification strategy, as discussed in chapter 3. The main drawback of this approach is the fact that the identification of investment as a separate category of government spending is a relatively recent addition to UK budgets in this sample.

Although it is possible to estimate how much total government spending consisted of investment, as Thomas and Dimsdale (2017) have, the main issue is in the identification of how much of the shocks was from investment and current spending. The compilation of the shocks relies on comparing publication from the time at which they took place. It might be possible, through further and much more detailed archival research of Main Estimates and Consolidated Reports, to estimate this, but it would be beyond the scope of the papers in this thesis to do so. Nonetheless, it would be very interesting — either through this or a different methodology altogether — to estimate how the effect of government investment and current spending on output might differ.

References

Afonso, António; Alves, José; Jalles, João Tovar. 2022. "To consolidate or not to consolidate? A multi-step analysis to assess needed fiscal sustainability." *International Economics*, volume 172, pages 106-123.

Afonso, António; Carvalho, Francisco Tiago. 2022. "Time-varying cyclicality of fiscal policy: The case of the Euro area." *North American Journal of Economics and Finance*, volume 62, 101778.

Afonso, António; Leal, Frederico Silva. 2019. "Fiscal multipliers in the Eurozone: an SVAR analysis." *Applied Economics*, volume 51, pages 5577-5593.

Alesina, Alberto; Ardagna, Silvia; Perotti, Roberto; Schiantarelli, Fabio. 2002. "Fiscal Policy, Profits, and Investment." *American Economic Review*, volume 92, number 3, pages 571-589.

Alesina, Alberto; Azzalini, Gualtiero; Favero, Carlo; Giavazzi, Francesco; Miano, Armando. 2018. "Is it the "How" or the "When" that Matters in Fiscal Adjustments?" *IMF Economic Review*, volume 66, pages 144-188.

Alesina, Alberto; Barbiero, Omar; Favero, Carlo; Giavazzi, Francesco; Paradisi, Matteo. 2017. "The Effects of Fiscal Consolidations: Theory and Evidence." *NBER Working Paper 23385.*

Alesina, Alberto; Favero, Carlo; Giavazzi, Francesco. 2015. "The output effect of fiscal consolidation plans." *Journal of International Economics*, volume 96, supplement 1, pages S19-S42.

Alesina, Alberto; Favero, Carlo; Giavazzi, Francesco. 2019. "Effects of Austerity: Expenditure- and Tax-based Approaches." *Journal of Economic Perspectives*, volume 33, number 2, pages 141-162.

Alesina, Alberto; Perotti, Roberto. 1997. "Fiscal Adjustments in OECD Countries: Composition and Macroeconomic Effects." *IMF Staff Papers*, volume 44, number 2, pages 210-248.

Allen, J.E. 1917. "War Budgets." *Fortnightly review, May 1865-June 1934*, volume 101, issue 605, pages 827-837.

Anderson, T.W.; Rubin, Herman. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics*, volume 20, number 1, pages 46-63.

Andrews, Donald W.K. 1993. "Tests for Parameter Instability and Structural Change With Unknown Change Point." *Econometrica*, volume 61, number 4, pages 821-856.

Argimon, Isabel; González-Paramo, José M.; Roldán, José M. 1997. "Evidence of public spending crowding-out from a panel of OECD countries." *Applied Economics*, volume 29, issue 8, pages 1001-1010.

Ashenfelter, Orley; Layard, Richard. 1983. "Incomes Policy and Wage Differentials." *Economica*, New Series, volume 50, number 198, pages 127-143.

Auerbach, Alan J. “Is There a Role for Discretionary Fiscal Policy?”, *NBER Working Paper 9306*.

Auerbach, Alan J.; Gorodnichenko, Yuriy. 2012. “Measuring the Output Responses to Fiscal Policy.” *American Economic Journal: Economic Policy*, volume 4, number 2, pages 1-27.

Barrell, Ray. 2014 “Macroeconomic policy and the 1981 Budget: changing the trend.” in *Expansionary Fiscal Contraction: The Thatcher Government’s 1981 Budget in Perspective*, eds. Needham, D. and A. Hotson, Cambridge University Press, pages 181-203.

Barrell, Ray; Holland, Dawn; Hurst, Ian. 2012. “Fiscal Consolidation: Part 2. Fiscal Multipliers and Fiscal Consolidations.” *OECD Economics Department Working Papers No. 933*.

Baum, Anja; Poplawski Ribeiro, Marcos; Weber, Anke. 2012. “Fiscal Multipliers and the State of the Economy.” *IMF Working Paper No. 2012/286*.

Baxter, Marianne; King, Robert G. 1995. “Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series.” *NBER Working Paper number 5022*.

Baxter, Marianne; King, Robert G. 1999. “Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series.” *The Review of Economic Statistics*, volume 81, issue 4, pages 575-593.

Bertola, Giuseppe; Drazen, Allan. 1993. “Trigger Points and Budget Cuts: Explaining the Effects of Fiscal Austerity.” *American Economic Review*, volume 83, number 1, pages 11-26.

Bianchi, Francesco. 2012. “Evolving Monetary/Fiscal Policy Mix in the United States.” *American Economic Review*, volume 102, number 3, pages 167-172.

Blanchard, Olivier; Perotti, Roberto. 2002. “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output.” *The Quarterly Journal of Economics*, volume 117, number 4, pages 1329-1368.

Bova, Elva; Kolerus, Christina; Tapsoba, Sampawende J.A.. 2015. “A fiscal job? An analysis of fiscal policy and the labor market.” *IZA Journal of Labor Policy*, volume 4, article 13.

Christiano, Lawrence J.; Fitzgerald, Terry J. 2003. “The Band Pass Filter.” *International Economic Review*, volume 44, issue 2, pages 435-465.

Cimadomo, Jacopo; Bénassy-Quéré, Agnès. 2012. “Changing patterns of fiscal policy multipliers in Germany, the UK and the US.” *Journal of Macroeconomics*, volume 34, issue 3, pages 845-873.

Cloyne, James. 2013. “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom.” *American Economic Review*, volume 103, number 4, pages 1507-1528.

Cloyne, James; Dimsdale, Nicholas; Postel-Vinay, Natacha. 2023. “Taxes and Growth: New Narrative Evidence from Interwar Britain.” *The Review of Economic Studies*, pages 1-33.

Cloyne, James; Jordà, Òscar; Taylor, Alan M.. 2020. “Decomposing the Fiscal Multiplier.” *NBER Working Paper 26939*.

Cooley, Thomas F.; Ohanian, Lee E.. 1997. “Postwar British Economic Growth and the Legacy of Keynes.” *Journal of Political Economy*, volume 105, number 3, pages 439-472.

Crafts, Nicholas. 2018. “Walking wounded: The British economy in the aftermath of World War I." in *The Economics of the Great War: A Centennial Perspective*, eds. Broadberry, S. and M. Harrison, CEPR Press, Paris and London.

Crowley, James B. 1963. "A Reconsideration of the Marco Polo Bridge Incident." *The Journal of Asian Studies*, volume 22, number 3, pages 277-291.

Cutler, Tony. 2003. "Dangerous Yardstick? Early Cost Estimates and the Politics of Financial Management in the First Decade of the National Health Service." *Medical History*, volume 47, issue 2, pages 217-238.

Davidson, Russell; MacKinnon, James G.. 2014. "Confidence sets based on inverting Anderson-Rubin tests." *The Econometrics Journal*, volume 17, issue 2, *Special issue: Advances in robust and flexible inference in econometrics, a special issue in honour of Joel L. Horowitz*, pages S39-S58.

De Grauwe, Paul; Ji, Yuemei. 2013. "From Panic-Driven Austerity to Symmetric Macroeconomic Policies in the Eurozone." *Journal of Common Market Studies*, volume 51, pages 31-41.

Deist, Wilhelm (translated by E.J. Feuchtwanger). 1996. "The Military Collapse of the German Empire: The Reality Behind the Stab-in-the-Back Myth." *War in History*, volume 3, issue 2, pages 186-207.

Dupor, Bill; Guerrero, Rodrigo. 2017. "Local and aggregate fiscal policy multipliers." *Journal of Monetary Economics*, volume 92, pages 16-30.

Feldstein, Martin. 2002. "The Role for Discretionary Fiscal Policy in a Low Interest Rate Environment." *NBER Working Paper 9203.*

Fels, Gerhard; Frölich, Hans-Peter; Blanchard, Oliver; Miller, Marcus. 1987. "Germany and the world economy: a round table Discussion." *Economic Policy*, volume 2, number 4, pages 177-205.

Ferrara, Laurent; Metelli, Luca; Natoli, Filippo; Siena, Daniele. 2021. "Questioning the puzzle: Fiscal policy, real exchange rate and inflation." *Journal of International Economics*, volume 133.

Friedman, Milton. 1957. "The Permanent Income Hypothesis." *A Theory of the Consumption Function*, Princeton University Press, ISBN 0-691-04182-2.

Ghassibe, Mishel; Zanetti, Francesco. 2022. "State dependence of fiscal multipliers: the source of fluctuations matters." *Journal of Monetary Economics*, volume 132, pages 1-23.

Giavazzi, Francesco; Pagano, Marco. 1990. "Can Severe Fiscal Contractions Be Expansionary? Tales of Tow Small European Countries." *NBER Macroeconomics Annual*, volume 5, pages 75-111.

Giavazzi, Francesco; Jappelli, Tullio; Pagano, Marco. 2000. "Searching for nonlinear effects of fiscal policy: Evidence from industrial and developing countries." *European Economic Review*, volume 44, issue 7, pages 1259-1289.

Glocker, Christian; Sestieri, Giulia; Towbin, Pascal. 2019. "Time-varying government spending multipliers in the UK." *Journal of Macroeconomics*, volume 60, pages 180-197.

Gómez, Víctor. 1999. "Three Equivalent Methods for Filtering Finite Nonstationary Time Series." *Journal of Business & Economic Statistics*, volume 17, number 1, pages 109-116.

Gordon, Robert J.; Krenn, Robert. 2010. "The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers." *NBER Working Paper 16380.*

Gupta, Sanjeev; Jalles, João Tovar; Mulas-Granados, Carlos; Schena, Michela. 2018. "Planned fiscal adjustments: Do governments fulfil their commitments?" *Eu-*

ropean Union Politics, volume 19, issue 3, pages 383-407.

Halifax, Stuart. 2010. "'Over by Christmas': British popular opinion and the short war in 1914." *First World War Studies*, volume 1, issue 2, pages 103-121.

Hall, Robert E.. 1986. "The Role of Consumption in Economic Fluctuations." *The American Business Cycle: Continuity and Change*, ed. Robert J. Gordon, pages 237-266.

Hamilton, James D. 2017. "Why You Should Never Use the Hodrick-Prescott Filter." *NBER Working Paper 23429.*

Hellwig, Martin; Neumann, Manfred J.T.; Vaubel, Roland; Miller, Marcus. 1987. "Economic Policy in Germany: Was There a Turnaround?" *Economic Policy*, volume 2, number 5, pages 103-145.

Hemming, Richard; Kell, Michael; Mahfouz, Selma. 2002. "The Effectiveness of Fiscal Policy in Stimulating Economic Activity — A Review of the Literature." *IMF Working Paper WP/02/208.*

Hodrick, Robert J. 2020. "An Exploration of Trend-Cycle Decomposition Methodologies in Simulated Data." *NBER Working Paper number 26750.*

Hodrick, Robert J.; Prescott, Edward C. 1981. "Post-War U.S. Business Cycles: An Empirical Investigation." *Northwestern University Center for Mathematical Studies in Economics and Management Science Working Paper number 451.*

Hodrick, Robert J.; Prescott, Edward C. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, volume 29, number 1, pages 1-16.

Ilzetzki, Ethan; Mendoza, Enrique G.; Végh, Carlos A. 2013. "How big (small?) are fiscal multipliers?" *Journal of Monetary Economics*, volume 60, issue 2, pages 239-254.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, volume 95, number 1, pages 161-182.

Jordà, Òscar; Knoll, Katharina; Kuvshinov, Dmitry; Schularick, Moritz; Taylor, Alan M.. 2019. "The Rate of Return on Everything, 1870–2015." *The Quarterly Journal of Economics*, volume 134, issue 3, pages 1225-1298.

Jordà, Òscar; Schularick, Moritz; Taylor, Alan M. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual 2016*, volume 31, eds. Martin Eichenbaum and Jonathan A. Parker. Chicago: University of Chicago Press.

Jørgensen, Peter L.; Ravn, Søren H. 2022. "The inflation response to government spending shocks: A fiscal price puzzle?" *European Economic Review*, volume 141, 103982.

Keynes, John Maynard. 1940. "How to Pay for the War." In: *Essays in Persuasion*, pages 367-439.

King, Robert G.; Rebelo, Sérgio. 1993. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Economic Dynamics and Control*, volume 17, issue 1-2, pages 207-231.

Kydland, Finn E.; Prescott, Edward C. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica*, volume 50, number 6, pages 1345-1370.

Ladiray, Dominique; Mazzi, Gian Luigi; Sartori, Fabio. 2003. "Statistical Methods for Potential Output Estimation and Cycle Extraction." *Eurostat Working Paper Series.*

Lazarus, Eben; Lewis, Daniel J.; Stock, James H.; Watson, Mark W.. 2018.

"HAR Inference: Recommendations for Practice." *Journal of Business and Economic Statistics*, volume 36, issue 4, pages 541-559.

Lazarus, Eben; Lewis, Daniel J.; Stock, James H.. 2021. "The Size-Power Tradeoff in HAR Inference." *Econometrica*, volume 89, issue 5, pages 2497-2516.

Lucas, Jr., Robert E. 1980. "Two Illustrations of the Quantity Theory of Money." *American Economic Review*, volume 70, issue 5, pages 1005-1014.

Mankiw, Gregory N.. 1986. "Issues in Keynesian macroeconomics: A review essay." *Journal of Monetary Economics*, volume 18, issue 2, pages 217-223.

Middleton, Roger. 1985. *Towards the Managed Economy: Keynes, the Treasury and the fiscal policy debate in the 1930s.*, Routledge, London.

Miller, Stephen M. 2006. "Slogging across the Veldt: British volunteers and the guerilla phase of the South African War, 1899-1902." *Journal of the Society for Army Historical Research*, volume 84, number 338, pages 158-174.

Mincer, Jacob A.; Zarnowitz, V. 1969. "The Evaluation of Economic Forecasts." *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, ed. Jacob A. Mincer, pages 3-46.

Monacelli, Tommaso; Perotti, Roberto; Trigari, Antonella. 2010. "Unemployment fiscal multipliers." *Journal of Monetary Economics*, volume 57, issue 5, pages 531-553.

Montiel Olea, José L.; Pflueger, Carolin. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics*, volume 31, issue 3, pages 358-369.

Montiel Olea, José L.; Stock, James H.; Watson, Mark W.. 2021. "Inference in Structural Vector Autoregressions identified with an external instrument." *Journal of Econometrics*, volume 225, issue 1, pages 74-87.

Mountford, Andrew; Uhligh, Harald. 2009. "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, volume 24, issue 6, pages 960-992.

Müller, Ulrich. 2014. "HAC Corrections for Strongly Autocorrelated Time Series." *Journal of Business and Economic Statistics*, volume 32, issue 3, pages 311-322.

Murphy, Daniel; Walsh, Kieran James. 2022. "Government spending and interest rates." *Journal of International Money and Finance*, volume 123.

Nakamura, Emi; Steinsson, Jón. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review*, volume 104, number 3, pages 753-792.

Neil, Robert. 2014. "The 1981 statement by 364 economists." in *Expansionary Fiscal Contraction: The Thatcher Government's 1981 Budget in Perspective*, eds. Needham, D. and A. Hotson, Cambridge University Press, pages 1-9.

Newton, Scott. 2010. "The Sterling Devaluation of 1967, the International Economy and Post-War Social Democracy." *The English Historical Review*, volume CXXV, issue 515, pages 912-945.

Officer, Lawrence H.. 2024. "What Was the Interest Rate Then?", MeasuringWorth, http://www.measuringworth.com/interestrates/ (Accessed 12 May 2024).

Parischa, Gurnain K. 2006. "Kalman Filter and its Economic Applications." *MPRA Paper number 22734.*

Perotti, Roberto. 1999. "Fiscal Policy in Good Times and Bad." *Quarterly Journal of Economics*, volume 114, number 4, pages 1399-1436.

Perotti, Roberto. 2005. "Estimating the effects of fiscal policy in OECD countries." *CEPR Discussion Paper No. 4842.*

Pesaran, M. Hashem; Shin, Yongcheol. 1995. *An autoregressive distributed lag approach to cointegration analysis*, volume 9514, Department of Applied Economics, University of Cambridge.

Pesaran, M. Hashem; Shin, Yongcheol; Smith, Richard J.. 2001. "Bounds Testing Approaches to the Analysis of Level Relationships." *Journal of Applied Econometrics*, volume 16, number 3, Special Issue in Memory of John Denis Sargan, 1924-1996: Studies in Empirical Macroeconometrics, pages 289-326.

Plagborg-Møller, Mikkel; Wolf, Christian K.. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica*, volume 89, issue 2, pages 955-980.

Pollock, D.S.G. 2000. "Trend estimation and de-trending via rational squarewave filters." *Journal of Econometrics*, volume 99, issue 2, pages 317-334.

Rafiq, Sohrab. 2014. "UK Fiscal Multipliers in the Postwar Era: Do State Dependent Shocks Matter?" *CESifo Economic Studies*, volume 60, issue 1, pages 213-245.

Ramey, Valerie A. 2009. "Identifying Government Spending Shocks: It's All in the Timing." *NBER Working Paper 15464.*

Ramey, Valerie A.. 2011a. "Identifying Government Spending Shocks: It's all in the Timing." *The Quarterly Journal of Economics*, volume 126, issue 1, pages 1-50.

Ramey, Valerie A.. 2011b. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature*, volume 49, number 3, pages 673-685.

Ramey, Valerie A. 2012. "Government spending and private activity." *NBER Working Paper 17787.*

Ramey, Valerie A.. 2016. "Macroeconomic Shocks and Their Propagation." *Handbook of Macroeconomics*, chapter 2, volume 2, pages 71-162.

Ramey, Valerie A.. 2019. "Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives*, volume 33, number 2, pages 89-114.

Ramey, Valerie A.; Shapiro, Matthew D.. 1998. "Costly capital reallocation and the effects of government spending." *Carnegie-Rochester Conference Series on Public Policy*, volume 48, pages 145-194.

Ramey, Valerie A.; Zubairy, Sarah. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy*, volume 126, number 2, pages 850-901.

Raven, Morten; Schmitt-Grohé, Stephanie; Uribe, Martín. 2006. "Deep Habits." *The Review of Economic Studies*, volume 73, number 1, pages 195-218.

Ravn, Morten O.; Uhlig, Harald. 2002. "On adjusting the Hodrick-Prescott filter for the frequency of observations." *The Review of Economics and Statistics*, volume 84, issue 2, pages 371-375.

Riley, Jon; Chote, Robert. 2014. "Crisis and consolidation in the public finances." *OBR Working Paper number 7.*

Romer, Christina D.; Romer, David H.. 1994. "What Ends Recessions?" *NBER Working Paper 4765.*

Romer, Christina D.; Romer, David H.. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review*, volume 100, number 3, pages 763-801.

Romer, Christina D.; Romer, David H. 2023. "Does Monetary Policy Matter? The Narrative Approach after 35 Years." *NBER Working Paper 31170*.

Romer, David H.. 2011. "What Have We Learned about Fiscal Policy from the Crisis?" *IMF Conference on Macro and Growth Policies in the Wake of the Crisis*.

Shaheen, Rozina; Turner, Paul. 2020. "Fiscal multipliers and the level of economic activity: a structural threshold VAR model for the UK." *Applied Economics*, volume 52, issue 17, pages 1857-1856.

Smith, Jacob. 2016. "Spurious Periodicity in Christiano-Fitzgerald Filtered Time Series." *SSRN Paper*.

Staiger, Douglas; Stock, James H.. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, volume 65, number 3, pages 557-586.

Stock, James H.; Watson, Mark W.. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *The Economic Journal*, volume 128, issue 610, pages 917-948.

Suárez Serrato, Juan Carlos; Wingender, Philippe. 2016. "Estimating Local Fiscal Multipliers." *NBER Working Paper No. 22425*.

Thomas, Ryland; Dimsdale, Nicholas. 2017. "A Millennium of UK data." Bank of England OBRA dataset.

Tsebelis, George. 2009. "Agenda Setting and Executive Dominance in Politics." *Parlamente, Agendasetzung und Vetospieler*, pages 13-24.

Welch, B.L. 1947. "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved." *Biometrika*, volume 34, issue 1-2, pages 28-35.

Woitek, Ulrich. 1998. "A Note on the Baxter-King Filter." *University of Glasgow School of Business (Economics) Working Paper number 9813*.

Wolf, Elias; Mokinski, Frieder; Schueler, Yves S. 2020. "On Adjusting the One-sided Hodrick-Prescott Filter." *Deutsche Bundesbank Discussion Paper number 11/2020*.

Wren-Lewis, Simon. 2015. "The Macroeconomic Record of the Coalition Government." *National Institute Economic Review*, volume 231, pages R5-R16.