Are Tax Cuts Still Expansionary at the Zero Lower Bound? Corrected Local Projections and Bootstrap Inference for UK Narrative Tax Reforms

Master Final Work

Applied Econometrics and Forecasting

Justus J.M. KLEINSMAN

(L62921)

30 June 2026

Lisbon School of Economics & Management
Universidade de Lisboa

Thesis Committee:

Prof. Adriana Cornea-Madeira (Supervisor)

Abstract

This paper studies whether tax reforms intended to reduce tax liabilities remain expansionary when monetary policy is constrained by the zero lower bound. New Keynesian theory predicts that some distortionary tax cuts can become contractionary at the ZLB because lower expected inflation raises the real interest rate when the nominal interest rate cannot fall further. The analysis starts from the UK narrative tax reform series from Cloyne et al. (2025) and compares regular level local projections, cumulative change local projections, two corrected local-projection specifications, and fixed regressor residual moving block bootstrap inference. The main cumulative change results do not support the contractionary ZLB prediction.

The estimated ZLB response is positive at the selected horizons and remains positive after the corrected LPs are applied. The level LP gives a less stable response, and the bootstrap intervals make the longer horizon evidence more cautious. The results therefore suggest that the sign of the ZLB response is robust: tax reforms do not appear to become contractionary when monetary policy is constrained. However, the statistical strength of this result is less robust, since it depends on the specification of the local projection and on whether inference is pointwise or simultaneous. The ZLB response is also not shown to be statistically larger than the non-ZLB response.

Keywords: narrative tax reforms, zero lower bound, local projections, corrected local projections, bootstrap inference, fiscal policy

Contents

1 Introduction

Tax cuts are usually expected to stimulate output by increasing disposable income, private demand, or incentives to work and invest (Romer and Romer, 2010; Mertens and Ravn, 2013). At the zero lower bound (ZLB), this usual intuition can change. The New Keynesian ZLB literature shows that fiscal policy can work differently when nominal interest rates cannot adjust downward. Eggertsson (2011) shows that in this setting some distortionary tax cuts can become contractionary.

The main empirical starting point for this paper is Cloyne et al. (2025). They study this prediction using a long run UK dataset of narrative-identified tax reforms and find that tax cuts remain expansionary at the ZLB. This paper uses their finding as the starting point and asks whether the positive ZLB response remains supported when the empirical specification and inference method are changed.

The main research question is therefore whether, in the United Kingdom setting, tax reforms intended to reduce tax liabilities continue to increase output when monetary policy is constrained.

The United Kingdom is a well-suited setting because the sample contains several episodes in which interest rates were close to the lower bound, such as the interwar period and the period after the 2008 financial crisis, while also containing many normal monetary-policy periods (Cloyne et al., 2025).

This paper makes both a methodological and an empirical contribution. The methodological contribution is that corrected-LP ideas are adapted to the narrative tax reform setting. The first correction is a Lusompa-style corrected-left-hand-side specification related to Lusompa (2023), while the second is a future-residual control specification related to Breitung and Brüggemann (2023). Both versions use scalar horizon zero GDP residuals rather than the full system of innovations used in the original estimators. They are therefore adaptations of these corrected-LP ideas, rather than exact replications of the original procedures. The bootstrap inference is also adapted to this setting by using a fixed regressor residual moving block bootstrap with calendar aware blocking, so that residual blocks and HAC covariance pairs do not cross the excluded World War II period.

The empirical contribution is to show where the expansionary ZLB result of Cloyne et al. (2025) is robust and where it becomes more fragile. The cumulative change ZLB response remains positive across the main specifications and robustness checks. However, the strength of the evidence depends on two choices that the single specification in Cloyne et al. (2025) could not separate. The first is the dependent variable transformation, since the cumulative change LP gives a more stable positive response than the level LP in the ZLB regime. The second is the inference method, since the pointwise intervals support positive responses at selected horizons while the simultaneous bands make the evidence more cautious. The contribution is therefore not only to check whether the original result can be reproduced, but to show the conditions under which it remains supported.

The main results suggest that the sign of the ZLB response is robust, while its statistical strength is less robust. The cumulative change ZLB response remains positive, even after applying the corrected LPs. The results therefore do not support the New Keynesian prediction that tax reforms become contractionary at the ZLB. At the same time, the evidence is more cautious in the level LP and under bootstrap inference, and the ZLB response is not statistically larger than the non-ZLB response.

The remainder is structured as follows. Section 2 discusses the relevant literature on tax policy at the ZLB, fiscal reform identification, local projections, and corrected LP inference. Section 3 sets out the regular LP, the cumulative change LP, the corrected LPs, and the bootstrap logic. Section 4 describes the data and the way in which the ZLB regimes are constructed. Section 5 presents the empirical results. Section 6 relates the robustness checks in this paper to those in Cloyne et al. (2025). Finally, Section 7 concludes.

2 Literature Review

The literature is discussed in four connected parts. The first part discusses why tax policy may work differently when interest rates are close to the lower bound. The second part discusses narrative fiscal reforms and why the motivation behind a tax change matters for identification. The third part discusses local projections and their relation to VARs. The fourth part discusses corrected local projections and inference.

2.1 Tax Policy at the Zero Lower Bound

The ZLB changes the usual interaction between fiscal and monetary policy. In normal times monetary policy can respond to changes in demand by adjusting nominal interest rates. When rates are already close to the effective lower bound, this response is limited because nominal rates cannot be lowered much further. Fiscal policy may therefore work differently from normal times (Eggertsson and Woodford, 2003).

Tax cuts are usually expected to increase output because they can raise disposable income, reduce tax distortions, or improve incentives to work and invest. Empirical work using narrative tax changes generally supports this expansionary view. Romer and Romer (2010) find large positive output effects of exogenous tax cuts, while Mertens and Ravn (2013) also find expansionary effects from personal and corporate income tax cuts.

Eggertsson (2011) provides a New Keynesian theoretical framework in which distortionary tax cuts can become contractionary when the nominal interest rate is constrained by the ZLB. This goes against the intuition that lower taxes should stimulate the economy, but the reasoning becomes clear once the role of expected inflation is considered.

If a tax cut lowers marginal costs, firms can produce at lower cost. This can put downward pressure on prices, meaning that current and expected inflation may fall. In normal times the central bank could respond by lowering the nominal interest rate. At the ZLB this response is limited however, so the lower expected inflation raises the real interest rate through the Fisher relation in equation (1). This higher real interest rate can reduce demand and output, since saving becomes relatively more attractive than spending and borrowing becomes more costly in real terms. Hence a policy that would normally be expected to stimulate output can become contractionary when monetary policy is constrained (Eggertsson, 2011).

The mechanism can be summarised using the standard approximate Fisher relation for real interest rates:

$$r_t \approx i_t - E_t \pi_{t+1}, \tag{1}$$

where r_t is the real interest rate, i_t is the nominal interest rate, and $E_t \pi_{t+1}$ is expected inflation. Away from the ZLB, the nominal interest rate can fall when expected inflation falls. At the ZLB, this adjustment is limited. Therefore a fall in expected inflation can increase the real interest rate, which can reduce demand and output (Eggertsson, 2011).

This mechanism can be summarised schematically as follows. This is not a separate model, but a visual summary of the mechanism described in Eggertsson (2011).

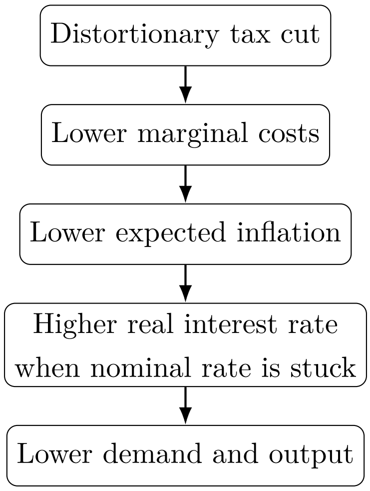

Figure 1: Schematic summary through which a tax cut can become contractionary at the ZLB, based on Eggertsson (2011).

Related theoretical work also shows that tax policy can have a special role at the zero bound. Correia et al. (2013) study how tax instruments can change incentives and relative prices when conventional monetary policy is constrained. This is not the same question as whether tax cuts are empirically contractionary, but it shows why tax policy at the ZLB is theoretically important.

The main empirical reference is Cloyne et al. (2025). They test the ZLB tax-cut prediction using a long run UK dataset of narrative-identified tax changes from 1918 to 2020. They find that tax cuts are expansionary both at and away from the ZLB, challenging the New Keynesian prediction that distortionary tax cuts should become contractionary when monetary policy is constrained.

This also relates to the broader question of whether fiscal policy has the same effect in every state of the economy. In simple terms, a fiscal multiplier describes how much output changes after a fiscal policy change. The relevant question is therefore not only whether fiscal policy affects output, but whether this effect differs between normal times, recessions, periods with unused economic capacity, or periods where monetary policy is constrained. Ramey and Zubairy (2018) study this question for government spending and ask whether the output effect is larger in weak economic states or at the ZLB.

Jordà and Taylor (2016) study fiscal consolidations, meaning periods where governments try to reduce deficits through spending cuts or tax increases. They find that these consolidations can reduce output, especially when economic conditions are already weak.

These papers do not study exactly the same policy as Cloyne et al. (2025), since their focus is not specifically on tax cuts at the ZLB. However, they are still relevant because they show that fiscal policy effects should not automatically be treated as constant across all periods. The effect may depend on the economic regime in which the policy is implemented. For the current analysis, the relevant regime split is ZLB versus non-ZLB periods.

2.2 Narrative Fiscal Reforms

The problem with tax policy changes is that they are often endogenous, meaning that they are partly determined by the state of the economy itself. Governments may change taxes because the economy is already weakening, because inflation is high, because debt is rising, or because they want to stimulate demand. This means that observed tax changes can partly be a response to the state of the economy, rather than only causing changes in the economy.

Using tax changes without separating their motivation can therefore be misleading. For example, if the government introduces a tax cut because the economy is already weakening, the movement in GDP that follows may partly reflect the weakening economy that caused the tax cut, rather than the effect of the tax cut itself. In that case, the tax change is partly a reflection of the economic state in which it was made. The estimated relationship between tax changes and GDP may then reflect reverse causality, rather than the causal effect of fiscal policy (Romer and Romer, 2010; Cloyne, 2013).

To deal with this possible endogeneity problem, a narrative approach is used. This means that historical policy records are used to classify tax changes by their motivation. The goal is to separate tax changes made in response to current or expected economic conditions from tax changes motivated by longer run objectives. The identification strategy used by Cloyne et al. (2025) follows the narrative tax shock literature outlined by Romer and Romer (2010).

This does not mean that the reform series is perfect. Narrative classification requires judgement and historical policy documents can be interpreted in more than one way. However, the approach directly addresses the main endogeneity problem by focusing on the stated motivation behind the reform. This is why the narrative classification is central for interpreting the estimated tax reform responses as causal responses rather than simple correlations.

2.3 Local Projections (LPs)

The next step after identifying the tax reforms is to estimate the dynamic response of GDP. In other words, the question shifts to "how GDP changes in the quarters after a tax reform occurs". These responses are estimated using local projections (LPs), first introduced by Jordà (2005). Local projections estimate impulse responses directly by running a separate regression for each response horizon. The GDP response is thus built up horizon by horizon rather than estimated as one single coefficient.

This is useful because the effect of a tax reform does not have to be constant over time. A tax reform may have a different effect in the quarter in which it is introduced than several quarters later. The horizon specific tax reform coefficients can thus be placed after each other to form the impulse response and show how the estimated GDP response develops over time.

The main alternative would be to estimate a full dynamic system, such as a vector autoregression (VAR) model, and then derive the impulse responses from that system. This requires specifying how the variables move together over time. LPs avoid part of this problem by estimating the response for each horizon separately. For example, if the true dynamics differ across regimes, a single system that does not account for this may impose too much structure on the response. This misspecification can then introduce bias into the estimated impulse response. This makes LPs easier to adapt when the response is allowed to differ between states, such as between ZLB and non-ZLB periods. This state dependent LP structure is common in applied fiscal work, for example in Ramey and Zubairy (2018), who allow fiscal responses to differ between economic states.

The flexibility of LPs can reduce bias when the full dynamic system would otherwise be misspecified. The downside is that this flexibility can make the estimates noisier, especially at longer horizons and in the smaller samples, such as that of the ZLB. Recent work therefore describes the choice between local projections and VARs mainly as a finite sample bias-variance trade-off rather than as a difference in the population object being estimated (Plagborg-Møller and Wolf, 2021; Li et al., 2024).

Table 1: Local Projections Compared with VARs

	Local projections	VAR
IRF construction	Separate regression for each horizon	Derived from full dynamic system
Flexibility	Easy to allow state dependence	More structure imposed by system
Main advantage	Less need to specify full system	More efficient if system is correct
Main downside	Can be noisy at longer horizons	Can be biased if system is misspecified

For this paper, the LP versus VAR comparison helps explain why LPs are a useful starting point. They make it easier to estimate responses separately by horizon and by regime. At the same time, recent methodological work stresses that LPs still require choices about the lag length, dependent variable, regime definition, and standard errors (Jordà and Taylor, 2024; Montiel Olea et al., 2025). These choices can affect the estimated response, especially in finite samples. This is why the empirical analysis later compares several LP specifications and inference methods, rather than relying on one single response path.

2.4 Corrected Local Projections and Inference

The main reason for using local projections is that they estimate the response horizon by horizon. However, it is this horizon-by-horizon structure that also creates a practical problem. Since a separate regression is estimated for every horizon, the estimates can become noisy at longer horizons, especially in small samples, such as the ZLB sample.

A related issue is that longer horizon LP residuals can be serially correlated. An unexplained movement shortly after the reform can still be part of GDP several quarters later. This means that the horizon-h regression error may contain unexplained GDP movements from the quarters between the reform at time t and the GDP outcome at time $t+h$. The LP estimate is not necessarily incorrect because of this, but it can become less precise and the confidence intervals can become less stable. Lusompa (2023) shows this more formally by deriving a moving average structure for LP residuals. This motivates using intermediate residuals when estimating longer horizon responses.

In the methodology, this idea is implemented in two ways. The first version is closest to the correction logic in Lusompa (2023), since it subtracts a constructed residual-propagation term from the left hand side. The second version is related to the residual-augmentation logic of Breitung and Brüggemann (2023), since the intermediate residuals are included directly as additional regressors. Both are corrected-LP specifications, but neither is presented as an exact replication of the full estimator in the original papers. This is because the implementation used here adapts the correction logic to the tax reform LP setting. It uses scalar GDP residuals from the horizon zero projection rather than the full system of innovations used in the original estimators.

Inference is a separate issue because the point estimate itself is not enough. The LP may estimate a positive GDP response, but this estimate is based on a finite sample and therefore comes with uncertainty. The confidence interval shows the range of responses that are still plausible given this uncertainty. If this interval includes zero, the estimated positive response is statistically speaking less convincing. This is relevant here since the ZLB sample is relatively small and the macroeconomic variables are highly persistent, meaning that the estimated responses may be imprecise.

HAC/Newey-West standard errors are used because the regression errors may be heteroskedastic and serially correlated (Newey and West, 1987). Heteroskedasticity means that the size of the errors can change over time. Serial correlation means that an error in one quarter may be related to errors in nearby quarters, something which is common in time-series data. The standard errors therefore need to account for this dependence rather than treating all observations as fully independent.

Residual based block bootstrap inference is also considered as an additional way to construct confidence intervals. The idea of the bootstrap is to create many artificial samples that resemble the original sample. The LP is then re-estimated on each artificial sample, resulting in a distribution of estimated impulse responses which show how much the estimated response can vary across samples.

Because the data are time-series data, the artificial samples should not be created by resampling individual residuals one by one. Residuals from nearby quarters may be related, so individual resampling would break part of this time-series dependence. Instead, residuals are resampled in blocks of consecutive quarters. This keeps part of the dependence between nearby residuals in the artificial bootstrap samples (Gonçalves and Kilian, 2004; Brüggemann et al., 2016).

The literature therefore points to three concerns that are important for the empirical analysis. First, the tax reforms need to be plausibly external to short run economic conditions. Otherwise the estimated response can partly reflect reverse causality rather than the effect of the reform itself. Second, the response may differ between ZLB and non-ZLB periods, meaning that an average response can hide important differences between regimes. Third, LP estimates and confidence intervals may be sensitive to persistence, finite sample behaviour, and the way inference is constructed. These concerns motivate the comparison between level LPs, cumulative change LPs, corrected LPs, and bootstrap confidence intervals.

3 Methodology

3.1 Methodology Overview

The methodology starts from the narrative tax reform series constructed by Cloyne et al. (2025) and focuses on how the GDP response is estimated from that series.

The baseline control structure also follows Cloyne et al. (2025). The extensions in this paper therefore do not change the core tax reform series or the main control set. Instead, they change the LP transformation, the corrected-LP construction, and the inference method.

After identifying the narrative tax reforms, the next step is to estimate how GDP responds to these reforms over time. This is needed because the response does not have to be the same in every quarter after the reform. One coefficient is therefore not enough to describe the full dynamic response of GDP to a tax reform.

Local projections (LPs), first introduced by Jordà (2005), are used to estimate these responses. Local projections estimate the response directly by running a separate regression for every horizon, meaning that only the GDP observation used as the dependent variable moves further into the future, while the tax reform remains dated at time t. The regression at $h = 0$ therefore relates GDP during the reform quarter to the reform at time t, while the regression at $h = 4$ relates GDP one year later to that same reform.

In its simplest form, the LP can be written as:

$$y_{t+h} = \alpha_h + \beta_h x_t + \text{controls}_t + u_{t+h}. \tag{2}$$

This equation is only meant to show the basic timing. The tax reform remains dated at time t, while the GDP outcome moves forward with the horizon h. The coefficient β_h is estimated separately for each horizon, and the sequence of these coefficients forms the impulse response. The exact dependent variable and controls differ across the level, cumulative change, and corrected specifications described below.

The response is first estimated using a regular level LP, where the left hand side is future real GDP. Then a cumulative change LP is estimated, where the left hand side measures the change in GDP between the quarter before the reform and the chosen response horizon. This cumulative change specification is also used by Cloyne et al. (2025). Recent econometric work has also shown that level and cumulative change LPs can behave differently when the sample is limited and the variables are highly persistent. Therefore both versions are compared.

The two regular LPs are then followed by two corrected versions, which are estimated for both the level and cumulative change outcomes. The first estimates a correction term and subtracts this term from the left hand side before estimating the tax reform response. The second includes the same intermediate residual information directly as additional regressors. These corrections are included because an unexplained economic movement shortly after the reform can continue to form part of the regression error at later horizons. The corrected LPs do not change the underlying LP setup. They keep the same reform timing and control structure, but add a correction for intermediate horizon zero GDP residuals that may otherwise remain in later-horizon regression errors. The corrected-LP construction and its relation to Lusompa (2023) and Breitung and Brüggemann (2023) are described in Section 3.7.

Lastly, the response is allowed to differ between ZLB and non-ZLB periods. The two responses are estimated inside one joint regression, which makes it possible to calculate their covariance and to test whether the two responses are statistically different. The point estimates are obtained using Ordinary Least Squares (OLS). HAC/Newey-West standard errors and a fixed regressor residual moving block bootstrap are then used to describe the uncertainty around these estimates.

3.2 Notation and Variable Interpretation

3.2.1 GDP Variable

Real GDP is first transformed by taking its logarithm and is then multiplied by 100. The response variable is therefore defined as:

$$y_t = 100 \ln(\text{RealGDP}_t). \tag{3}$$

Using the logarithm makes changes in y_t interpretable as approximate percentage changes in real GDP. Multiplying the logarithm by 100 then places these changes directly into percentage units, meaning that an increase of one unit in y_t corresponds approximately to a one percent increase in real GDP.

3.2.2 Tax Reform Variable

The main explanatory variable is denoted by x_t and represents the narrative tax reform variable used in the empirical analysis. Positive values of x_t represent tax cuts, while negative values represent tax increases. This sign is the reverse of the original tax-change series, where a tax increase is positive. The sign is reversed so that the estimated response can be interpreted as the response to a tax cut.

Although this series is commonly referred to as a narrative tax shock in the literature, the underlying historical policy actions are tax reforms. The value of x_t does not measure an unexpected movement in GDP, tax revenue, or another macroeconomic variable. Instead, it measures the intended revenue size of the identified tax reform. In this setting, the word shock refers to the identification assumption that the reform is treated as plausibly unrelated to current or expected short run economic conditions.

To keep this distinction clear, x_t is referred to as the narrative tax reform variable throughout the empirical analysis. The term "narrative tax shock" is only used when referring to the terminology and identification approach used in the existing literature.

The variable x_t is dated at the quarter to which the identified tax change is assigned in the narrative dataset. Individual tax measures are assigned to the quarter in which they are implemented, while measures implemented during the second half of a quarter are assigned to the following quarter. One tax reform can contain several individual tax measures. When multiple measures are assigned to the same quarter, their intended revenue effects are combined into one net quarterly tax reform value.

Let $\mathcal{M}_t$ denote the set of narrative tax measures assigned to quarter t. For each measure m, let $\Delta T_{m,t}^{int}$ denote the intended revenue effect of that measure. This value is positive when the measure is intended to increase tax liabilities and negative when it is intended to reduce tax liabilities. The net intended tax change in quarter t, denoted by ΔT_t^{int}, is then:

$$\Delta T_t^{int} = \sum_{m \in \mathcal{M}_t} \Delta T_{m,t}^{int}. \tag{4}$$

This net intended tax change is divided by nominal GDP in the previous quarter, $NGDP_{t-1}$, and multiplied by 100 (Cloyne et al., 2025). This gives the intended tax change as a percentage of GDP:

$$\tau_t = 100 \frac{\Delta T_t^{int}}{NGDP_{t-1}}. \tag{5}$$

Here τ_t is positive for intended tax increases and negative for intended tax cuts. The sign is then reversed because the analysis is written in terms of tax cuts:

$$x_t = -\tau_t. \tag{6}$$

A value of $x_t = 1$ means that the tax measures assigned to quarter t were together intended to reduce total tax liabilities by an amount equal to one percent of GDP at that time. This does not mean that one particular tax rate, such as the income-tax rate, fell by one percentage point, nor does it mean that GDP was intended to increase by one percent or that GDP actually increased by one percent. The value of x_t only describes the intended revenue size of the reform relative to the economy at the time of implementation. The eventual response of GDP is not contained in x_t and is the outcome which the LP is intended to estimate.

3.2.3 Timing and Coefficient Interpretation

The quarter to which the tax reform is assigned is denoted by t, with h describing how many quarters after the reform the GDP response is measured. The reform remains dated at time t in every horizon specific LP, while the horizon h only changes the quarter in which GDP is measured. Since the data are quarterly, increasing h by one moves the GDP outcome forward by one quarter. Thus $h = 0$ measures the response during the same quarter as the reform, $h = 1$ measures the response one quarter later, and this continues until $h = 12$, where the response is measured three years after the reform.

The coefficient estimated for the tax reform variable at horizon h is denoted by β_h. The subscript h is used because a separate tax reform coefficient is estimated for every horizon. So β_0 measures the estimated GDP response during the reform quarter, β_1 measures the response one quarter later, continuing until β_{12} measures the response three years after the reform. Combining these thirteen coefficients from $h = 0$ to $h = 12$ gives the estimated impulse response.

The estimated coefficient is denoted by $\hat{\beta}_h$. Since the dependent variable is 100 times log real GDP, and x_t is measured as a percentage of GDP, β_h measures the approximate percentage response of real GDP at horizon h to a reform intended to reduce tax liabilities by one percent of GDP. The fitted contribution of a reform with observed size x_t can be written as:

$$\widehat{\text{Response}}_{t,h} = \hat{\beta}_h x_t. \tag{7}$$

For a reform with $x_t = 1$, the estimated response is equal to $\hat{\beta}_h$. This is also the coefficient plotted in the impulse response graphs, meaning that no additional rescaling of the coefficient is performed. For example, if $\hat{\beta}_h = 2$, a tax reform intended to reduce tax liabilities by one percent of GDP is associated with an approximate two percent increase in real GDP at horizon h. The LP imposes a linear structure, so the fitted response changes in proportion to the size of the reform. Therefore a reform of half the size would also scale the fitted response down by half. So a reform with $x_t = 0.5$ would imply an approximate one percent increase in real GDP.

The coefficient β_h is described as a semi-elasticity of GDP with respect to the tax reform. A conventional elasticity measures the percentage change in one variable resulting from a one percent change in another variable. This is not the interpretation here because x_t is not the logarithm or percentage change of tax liabilities. It instead measures the intended change in tax liabilities as a percentage of GDP. The semi-elasticity can therefore be written as:

$$\beta_h = \frac{\text{approximate percentage response of real GDP at horizon } h}{\text{intended tax reduction measured as a percentage of GDP}}. \tag{8}$$

This interpretation depends on the identifying assumption that the narrative tax reform variable is unrelated to current or expected short run economic conditions after accounting for the included controls. If reforms are still introduced in response to economic conditions which are not captured by these controls, $\hat{\beta}_h$ may partly reflect these conditions rather than only the GDP response caused by the reform.

3.2.4 Control Variables

The narrative government spending reform is denoted by s_t and is also measured relative to GDP. The coefficient on this variable is not interpreted as part of the main result, but the variable is included as a control because government spending reforms may also affect GDP. Without this control, part of the GDP movement related to spending policy could be incorrectly attributed to the tax reform.

The four macroeconomic variables used as controls are grouped into the vector Z_t:

$$Z_t = \begin{pmatrix} y_t \\ q_t \\ g_t \\ i_t \end{pmatrix}, \tag{9}$$

where y_t is $100 \ln(\text{RealGDP}_t)$, q_t is $100 \ln(\text{RealTaxRevenue}_t)$, g_t is $100 \ln(\text{RealGovernmentExpenditure}_t)$, and i_t is the Bank Rate. The vector Z_t is an abbreviated way of displaying these variables in formulas; each variable still enters the estimated regression separately.

Lagged values of these controls are included to account for the economic conditions which are present before the tax reform happens. Previous GDP, tax revenue, government expenditure, and the Bank Rate may all be related to the later development of GDP. Including them therefore helps separate the estimated tax reform response from the predictable movement of GDP based on existing economic conditions. If these previous conditions are related to both the tax reform and future GDP, leaving this information out could cause part of the effect of existing conditions to be incorrectly attributed to the tax reform.

Since the macroeconomic controls enter with a lag, they are measured before the quarter to which the tax reform is assigned. They therefore do not normally remove part of the subsequent GDP response in the way that a control measured after the reform could. One caveat is that some reforms are announced before they are implemented. In such cases, lagged macroeconomic variables may already contain responses to the expected reform, even though they are measured before its implementation date (Cloyne et al., 2025).

Adding more lags can only weakly improve the in-sample fit of the regression, since OLS can assign a coefficient of zero to a lag which adds no explanatory information. But this does not necessarily make the estimated tax reform response more reliable. Additional lags use up degrees of freedom and can increase the variance and standard errors of the estimated coefficients, especially when they add little information or are highly correlated with the other regressors. The chosen lag order therefore needs to include enough previous economic information without making the regression unnecessarily large.

The same four macroeconomic controls are used in both level and cumulative change LP specifications, but the form in which they enter differs. The level LP includes the previous levels of these variables, denoted by Z_{t-j}, where j indicates how many quarters the control is lagged. For example, Z_{t-1} contains real GDP, real tax revenue, real government expenditure, and the Bank Rate one quarter before the reform.

The cumulative change LP instead includes the lagged first differences of these variables, denoted by ΔZ_{t-j}. A first difference measures how a variable changes between two consecutive quarters and is defined as:

$$\Delta Z_{t-j} = Z_{t-j} - Z_{t-j-1}. \tag{10}$$

For example, ΔZ_{t-1} measures the change between quarters $t-2$ and $t-1$, meaning that this change is fully observed before the reform at time t. For real GDP, real tax revenue, and real government expenditure, which are measured as 100 times their natural logarithm, the first differences can be interpreted as approximate quarterly percentage changes. For the Bank Rate, the first difference measures its change in percentage points. The level LP therefore controls for the previous values of the macroeconomic variables, while the cumulative change LP controls for their previous quarterly movements. The controls themselves remain the same and between the two specifications it is only the transformation that changes.

3.3 Identification and OLS

Estimating the effect of a tax reform becomes difficult when the decision to change taxes is itself influenced by the economy. For example, a government may decide to lower taxes because output is already weakening. The movement in GDP observed after this reform may then partly reflect the economic weakness which caused the government to intervene, rather than only the effect of the tax reform itself. This creates reverse causality, since the economy affects the tax reform while the tax reform is simultaneously intended to affect the economy.

The narrative identification procedure is used to reduce this problem. Historical policy documents are examined to determine why each tax change was introduced. Tax reforms which were directly motivated by current or expected short run economic conditions are classified as endogenous and are excluded from the identified reform series. The remaining reforms are motivated by objectives which are less directly connected to short run economic developments, making them more plausibly external to the macroeconomic conditions surrounding their implementation (Cloyne et al., 2025).

Table 2: Narrative Classification of Tax Changes

Type of tax change	Motivation	Use in identification
Endogenous tax change	Response to current or expected conditions	Less suitable for causal interpretation
Narrative tax reform	Longer run policy motivation	Closer to external policy reform

Table 2 summarises the distinction used by the narrative identification strategy. The LP regressions use the narrative tax reforms, since these are the tax changes treated as more plausibly external to current and expected short run economic conditions. This avoids treating tax changes caused by the economy in the same way as reforms that are less directly tied to those conditions.

This classification leads to the main identifying assumption used in the LP regressions. After accounting for the included controls, the narrative tax reform variable should not be systematically related to the other unobserved disturbances affecting future GDP. This assumption can be written as:

$$\operatorname{Cov}\left(x_t, u_{t+h} \mid \mathcal{C}_t\right) = 0, \tag{11}$$

where x_t is the narrative tax reform variable, u_{t+h} contains the remaining unexplained factors affecting GDP at horizon h, and $\mathcal{C}_t$ represents the information included as controls. These controls contain the spending reform, previous tax reforms, previous macroeconomic

conditions, and the ZLB regime interactions used in the state dependent regressions.
The condition in Equation (11) means that, after accounting for this information, larger or smaller narrative tax reforms should not systematically occur together with positive or negative values of the remaining GDP disturbance. If this condition holds, the tax reform coefficient can be estimated using OLS and interpreted as the response associated with variation in the identified reform series.
No separate instrumental-variable step is used because x_t enters the LP directly as the observed policy-reform variable which is treated as exogenous. This is different from an LP-IV specification, where the narrative series would only be used as an instrument for another tax-policy variable which is itself treated as endogenous. Such an approach would be required if the aim were to isolate exogenous variation in another observed measure of tax policy rather than estimating the response to x_t directly.
The narrative classification provides the economic argument behind the identifying assumption, but it cannot prove that the selected reforms are completely exogenous. A Granger-style predictability test is therefore included as a diagnostic, following Cloyne et al. (2025). This test examines whether lagged macroeconomic changes and lagged spending reforms jointly help to predict the current narrative tax reform after controlling for earlier tax reforms.
Failing to reject the null hypothesis means that the test does not provide evidence that the included economic history predicts the narrative tax reform variable. This result is consistent with the identifying assumption, but it does not prove that the assumption is correct. The test only considers the variables and lag structure which are included and cannot rule out every unobserved economic condition, anticipated policy response, or judgement made during the narrative classification.

3.4 Baseline Local Projections

The analysis starts with a regular level LP with the future level of real GDP placed on the left hand side, providing the simplest baseline estimate of how GDP at horizon h is related to the tax reform at time t. This baseline version is useful since it shows the estimated response before any cumulative change transformation or correction is applied.
The regular level LP is estimated separately for every horizon:

$$y_{t+h} = \alpha_h + \beta_h x_t + \delta_h s_t + \sum_{j=1}^{p} \theta_{h,j} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j} Z_{t-j} + u_{t+h}. \tag{12}$$

The intercept α_h is allowed to differ across horizons. The coefficient β_h is the tax reform response of interest. The coefficient δ_h captures the relation between the current spending reform and future GDP. The lagged tax coefficients $\theta_{h,j}$ control for tax reforms that occurred j quarters earlier. The row vector $\Gamma_{h,j}$ contains the coefficients on the jth lag of GDP, tax revenue, government expenditure, and the Bank Rate. Lastly, u_{t+h} contains the part of future GDP which is not explained by the tax reform or by the included controls.
The level LP is then compared to a cumulative change, or long-difference, LP. In this

version the left hand side is the change in GDP between the quarter before the reform and horizon h, rather than only the future level of GDP. This is the specification used by Cloyne et al. (2025) and is therefore used for reproducing their original estimates.
The cumulative change outcome is:

$$\Delta_h y_{t+h} = y_{t+h} - y_{t-1}. \tag{13}$$

The value y_{t-1} is observed before the tax reform and is therefore predetermined relative to x_t. At $h = 0$, the dependent variable is $y_t - y_{t-1}$, so the regression captures GDP growth during the reform quarter. At $h = 4$, the left hand side is $y_{t+4} - y_{t-1}$, so the regression captures the cumulative change between the pre-reform quarter and one year after the reform.
The cumulative change LP is defined as:

$$y_{t+h} - y_{t-1} = \alpha_h^{\Delta} + \beta_h^{\Delta} x_t + \delta_h^{\Delta} s_t + \sum_{j=1}^{p} \theta_{h,j}^{\Delta} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j}^{\Delta} \Delta Z_{t-j} + u_{t+h}^{\Delta}. \tag{14}$$

The narrative tax and spending reforms are not first-differenced because they are already constructed as reform variables assigned to a particular quarter. The macroeconomic controls enter as lagged quarterly changes. This follows the main cumulative change LP construction used by Cloyne et al. (2025).
Including both the level and cumulative change LP is not only useful for comparing the estimates with Cloyne et al. (2025); there is also an econometric reason for comparing them. Real GDP is highly persistent, meaning that its current level is strongly related to its previous levels and that the effect of an economic movement can remain visible for many quarters. This persistence becomes especially relevant when estimating longer horizon responses using a relatively small number of observations, such as in the ZLB state.
In population, the level and cumulative change LP are intended to estimate the same impulse response when the narrative tax reform is treated as exogenous. Population in this context refers to the response which would be estimated if an unlimited amount of data generated by the economic process was available. Since the cumulative change LP subtracts GDP from the quarter before the reform, which is predetermined relative to the reform at time t, subtracting previous-quarter GDP should not change the population response under the identifying assumption.
In practice however, there are only limited historical samples available and the estimates from the level and cumulative change LP can therefore differ due to their finite sample properties. So the estimators can behave differently when they are applied to a limited number of observations rather than to the theoretical population.
Piger and Stockwell (2025) show that this distinction is important when the response is identified using an observed external reform and that the cumulative change specification can reduce finite sample bias and improve confidence interval coverage relative to the level LP. These improvements become especially relevant when the dependent variable is

persistent, the response horizon is long, or the available sample is small. This makes the cumulative change LP worth exploring beyond replicating the specification used by Cloyne et al. (2025).

3.5 State Dependent ZLB Local Projections

The response to a tax reform is allowed to differ between ZLB and non-ZLB periods, following the ZLB classification used by Cloyne et al. (2025). A base lower bound indicator, denoted by L_t, is first constructed from the Bank Rate. Bank Rate equal to 2 is treated as the lower bound value for the earlier historical lower bound episode, while Bank Rate at or below 0.5 captures the modern lower bound period. The indicator L_t equals one in these lower bound observations and zero otherwise.

The observation 1939Q3 is also classified as lower bound because Bank Rate briefly rises in that quarter, while the surrounding quarters remain at 2. Although 1939Q3 is part of the excluded WWII period and is not used as a regression observation, it is still used when constructing the lower bound indicator on the original calendar. The sustained-ZLB rule uses both the previous quarter and the quarter three periods ahead, making the classification depend on the lower bound status of nearby quarters on the original calendar. Treating 1939Q3 as lower bound prevents this one-quarter Bank Rate movement from breaking the classification of the surrounding lower bound episode. After the ZLB indicator is constructed, the WWII observations are set to missing and do not enter the LP regressions.

The sustained lower bound indicator is:

$$D_t = L_{t-1} L_{t+3}. \tag{15}$$

Therefore, $D_t = 1$ when the economy is in the lower bound regime during the quarter before the tax reform and remains at the lower bound through $t + 3$. There are four quarters between the two observations used in this classification. This follows the attempt by Cloyne et al. (2025) to avoid classifying reforms near the end of a lower bound episode as ZLB reforms when monetary policy becomes unconstrained shortly afterwards.

The replication data also contain supplied variables `Regime0`, `Regime1`, ..., `Regime12`. The preferred specifications use the manually reconstructed sustained-ZLB indicator. This is done because it applies the four-quarter lower bound rule directly to the Bank Rate data. The supplied `Regime0` classification behaves like a previous-quarter lower bound indicator rather than the sustained four-quarter indicator used in the preferred specification. This difference is documented in Appendix A.3.2. It is therefore not used in the main specifications, but is included in the ZLB-definition robustness check for the regular level and cumulative change LPs.

The ZLB and non-ZLB responses are estimated inside one joint regression rather than through two unrelated regressions. This makes it possible to estimate the covariance between the ZLB and non-ZLB coefficients and to test their difference directly.

Let $Y_{t,h}$ denote either the level outcome y_{t+h} or the cumulative change outcome $y_{t+h} - y_{t-1}$. The joint state dependent specification is:

$$\begin{aligned}
Y_{t,h} = {} & \alpha_h^N(1-D_t) + \alpha_h^Z D_t \\
& + \beta_h^N x_t(1-D_t) + \beta_h^Z x_t D_t + \delta_h s_t \\
& + \sum_{j=1}^{p} \left[\theta_{h,j}^N x_{t-j}(1-D_t) + \theta_{h,j}^Z x_{t-j} D_t\right] \\
& + \sum_{j=1}^{p} \left[\Gamma_{h,j}^N W_{t-j}(1-D_t) + \Gamma_{h,j}^Z W_{t-j} D_t\right] + u_{t+h}.
\end{aligned} \tag{16}$$

The superscript N denotes the non-ZLB regime and Z denotes the ZLB regime. For the level LP, $W_{t-j} = Z_{t-j}$. For the cumulative change LP, $W_{t-j} = \Delta Z_{t-j}$. The coefficients β_h^N and β_h^Z are the non-ZLB and ZLB tax reform responses. The regime intercepts, current tax reform coefficient, lagged tax reform coefficients, and lagged macroeconomic control coefficients are allowed to differ between regimes. This follows the state dependent LP structure used by Cloyne et al. (2025).

The current spending reform enters with one common coefficient δ_h. Interacting the spending reform by regime produces columns with little or no identifying variation at several longer horizons, resulting in highly influential observations and unstable covariance estimates. The common spending coefficient is therefore used in the main specifications, while the fully interacted version is retained as a robustness check.

The difference between the ZLB and non-ZLB responses is:

$$d_h = \beta_h^Z - \beta_h^N. \tag{17}$$

Its estimated variance is:

$$\widehat{Var}(d_h) = \widehat{Var}(\hat{\beta}_h^Z) + \widehat{Var}(\hat{\beta}_h^N) - 2\widehat{Cov}(\hat{\beta}_h^Z, \hat{\beta}_h^N). \tag{18}$$

This gives a direct statistical test of whether the two responses are different. This is needed because visually comparing two separate confidence intervals does not perform the same test as estimating the difference directly.

3.6 Persistence, Lag Length, and Lag-Augmented Inference

The persistence of GDP does not only matter for the estimated response itself, but also for the conclusions which can be drawn from this estimate. For example, finding a positive coefficient does not immediately mean that there is strong evidence of a positive response since the coefficient may still be estimated with a large amount of uncertainty. It is therefore also necessary to consider the standard error, confidence interval, and statistical significance surrounding the estimate. Using these measures to draw conclusions about the underlying GDP response is referred to as inference.

When GDP is highly persistent, the usual statistical approximations may not work as expected. The coefficient can still be estimated, while the standard error and confidence interval surrounding it may not correctly describe its sampling uncertainty. For example, a 90 percent confidence interval procedure should produce intervals which contain the true response in approximately 90 percent of repeated samples. If persistence is not properly accounted for, the actual coverage can differ from 90 percent, meaning that the reported statistical evidence can appear either stronger or weaker than it actually is.

Montiel Olea and Plagborg-Møller (2021) show that this problem can be addressed using lag augmentation. Lag augmentation means including one additional lag beyond the lag order used to describe the underlying dynamics. This extra lag protects the distribution of the test statistic against the persistence of the variables, allowing normal critical values to be used when constructing confidence intervals and significance tests. This helps the usual statistical tests remain reliable even when GDP is highly persistent.

The purpose of lag augmentation is therefore to make the standard errors, confidence intervals, and significance tests more reliable. It does not necessarily make the estimated response itself less biased. This is different from the argument made by Piger and Stockwell (2025), where the cumulative change specification is shown to reduce finite sample bias and improve confidence interval coverage. The two arguments should therefore be kept separate: lag augmentation is mainly used to obtain valid inference under persistence, while the cumulative change specification is used to improve the finite sample behaviour of the estimated response.

The main specifications use four lags, following Cloyne et al. (2025). Their reason for this choice is to control for one full year of quarterly economic history. Using four lags is therefore first of all a replication choice, allowing the resulting impulse responses to be compared with their estimates. It should not be interpreted as proving that four is the uniquely correct lag order.

The choice of four lags and the idea of lag augmentation are related, but they are not the same decision. The number of lags determines how much of the previous economic history is included in the regression. Lag augmentation instead concerns how the chosen lag structure can help to obtain valid statistical inference when the variables are highly persistent.

To investigate whether the results depend on the four-lag replication choice, the lag order is also examined using AIC and BIC (Akaike, 1974; Schwarz, 1978). In the current sample, AIC selects three lags for the level system and two lags for the first-difference system, while BIC selects one lag for both. The LPs are re-estimated using the AIC-selected orders as a robustness check, allowing the four-lag results to be compared with a more data dependent choice.

3.7 Corrected Local Projections

As discussed in Section 2.4, the motivation for the corrected LPs is that the error term at a longer horizon can contain intermediate economic movements that occurred between the reform date and the GDP outcome. This section focuses on how this correction is implemented.

The only future-dated objects used in the correction are the estimated residuals between the reform at time t and the outcome at time $t+h$. These residuals are not ordinary future macroeconomic controls, such as future GDP, future tax revenue, future government expenditure, or the future Bank Rate. Instead, they are the parts of GDP which were not explained by the horizon zero regression. Their role is to account for error propagation, not to control away the economic response to the tax reform itself (Lusompa, 2023).

Both corrected LPs begin by estimating the horizon zero projection. For the level specification this is:

$$y_t = a_0 + b_0 x_t + d_0 s_t + \sum_{j=1}^{p} c_{0,j} x_{t-j} + \sum_{j=1}^{p} G_{0,j} Z_{t-j} + \varepsilon_t. \tag{19}$$

The error ε_t is the part of current GDP which is not explained by the current tax reform, current spending reform, lagged tax reforms, and lagged macroeconomic variables. This equation at horizon zero is estimated using all complete observations rather than being split by regime.

The cumulative change LP at horizon zero is similarly written as:

$$y_t - y_{t-1} = a_0^{\Delta} + b_0^{\Delta} x_t + d_0^{\Delta} s_t + \sum_{j=1}^{p} c_{0,j}^{\Delta} x_{t-j} + \sum_{j=1}^{p} G_{0,j}^{\Delta} \Delta Z_{t-j} + \varepsilon_t^{\Delta}. \tag{20}$$

The error ε_t^{Δ} is the unexplained part of current GDP growth rather than the unexplained part of the GDP level. The estimated residuals are:

$$\hat{\varepsilon}_t = y_t - \hat{a}_0 - \hat{b}_0 x_t - \hat{d}_0 s_t - \sum_{j=1}^{p} \hat{c}_{0,j} x_{t-j} - \sum_{j=1}^{p} \hat{G}_{0,j} Z_{t-j}, \tag{21}$$

$$\hat{\varepsilon}_t^{\Delta} = y_t - y_{t-1} - \hat{a}_0^{\Delta} - \hat{b}_0^{\Delta} x_t - \hat{d}_0^{\Delta} s_t - \sum_{j=1}^{p} \hat{c}_{0,j}^{\Delta} x_{t-j} - \sum_{j=1}^{p} \hat{G}_{0,j}^{\Delta} \Delta Z_{t-j}. \tag{22}$$

At horizon h, the correction only uses residuals occurring after the reform at time t and before the outcome at time $t+h$:

$$\hat{\varepsilon}_{t+1}, \hat{\varepsilon}_{t+2}, \ldots, \hat{\varepsilon}_{t+h-1}. \tag{23}$$

These residuals are future residuals relative to the reform date, but they are still intermediate residuals because they occur before the final outcome y_{t+h}. The residual $\hat{\varepsilon}_t$ is not used, since it comes from the same quarter as the tax reform. The endpoint residual $\hat{\varepsilon}_{t+h}$ is also not used since it occurs in the same quarter as the dependent variable. The correction therefore only uses residuals from between the reform at time t and the outcome at time $t+h$. This timing also means that no correction is possible at $h=0$ and $h=1$, since there are no residuals between the reform and the outcome at these horizons. The regular and corrected estimators are therefore identical at $h=0$ and $h=1$. The first correction thus occurs at $h=2$, where only $\hat{\varepsilon}_{t+1}$ is available. This timing follows the intermediate residual structure derived by Lusompa (2023).

Suppose GDP at time $t+h$ is described by a dynamic model with lagged variables, such as variables dated $t+h-1$. If this relationship is substituted backwards, the outcome at $t+h$ can be written in terms of variables dated closer to the reform date t, together with the shocks that occur between t and $t+h$. These terms are not new tax reform effects. Rather, they come from the underlying dynamic relationships in the model. This is why a longer horizon LP error can contain movements that happened after the reform, but before the final GDP outcome. The Lusompa-style corrected-left-hand-side LP uses this idea by estimating how these intermediate residual movements propagate into later outcomes and then subtracting the predicted propagation from the left hand side.

3.7.1 Corrected-Left-Hand-Side LP

The first corrected version combines the intermediate residuals into one correction term. This correction term is subtracted from the left hand side before the tax reform coefficient is estimated. The purpose is to remove the part of the later GDP outcome which can already be predicted from earlier LP residuals.

The corrected-left-hand-side estimator requires coefficients which measure how the short run residual propagates into later GDP outcomes. These coefficients are denoted by B_j. For the level specification, B_j is estimated using:

$$y_{t+j} = a_j^B + B_j \hat{\varepsilon}_t + b_j^B x_t + d_j^B s_t + \sum_{k=1}^{p} c_{j,k}^B x_{t-k} + \sum_{k=1}^{p} G_{j,k}^B Z_{t-k} + \eta_{t+j}^B, \qquad j = 1, \dots, 11. \quad (24)$$

The coefficient B_j measures how strongly the unexplained GDP movement at time t is related to GDP j quarters later after controlling for the same tax reforms and macroeconomic history. It is not the tax reform response.

The cumulative change propagation coefficients are estimated using:

$$y_{t+j} - y_{t-1} = a_j^{B,\Delta} + B_j^{\Delta} \hat{\varepsilon}_t^{\Delta} + b_j^{B,\Delta} x_t + d_j^{B,\Delta} s_t + \sum_{k=1}^{p} c_{j,k}^{B,\Delta} x_{t-k} + \sum_{k=1}^{p} G_{j,k}^{B,\Delta} \Delta Z_{t-k} + \eta_{t+j}^{B,\Delta}. \quad (25)$$

The propagation coefficients are estimated horizon by horizon. Meaning that B_1 measures one-quarter propagation, B_2 measures two-quarter propagation, and so on.

The corrected level outcome is:

$$\widetilde{y}_{t+h} = y_{t+h} - \sum_{j=1}^{h-1} \widehat{B}_j \hat{\varepsilon}_{t+h-j}. \quad (26)$$

At $h = 2$, the correction is:

$$\widetilde{y}_{t+2} = y_{t+2} - \widehat{B}_1 \hat{\varepsilon}_{t+1}. \quad (27)$$

At $h = 3$, the correction becomes:

$$\widetilde{y}_{t+3} = y_{t+3} - \widehat{B}_1 \hat{\varepsilon}_{t+2} - \widehat{B}_2 \hat{\varepsilon}_{t+1}. \quad (28)$$

Meaning that the residual closest to the outcome receives the one-quarter propagation coefficient, while an earlier residual receives the coefficient corresponding to the greater distance over which it must propagate.

The corrected outcome is then used in the final LP:

$$\widetilde{y}_{t+h} = \alpha_h^{CLP} + \beta_h^{CLP} x_t + \delta_h^{CLP} s_t + \sum_{j=1}^{p} \theta_{h,j}^{CLP} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j}^{CLP} Z_{t-j} + v_{t+h}. \tag{29}$$

The coefficient β_h^{CLP} remains the tax reform response. The correction does not change what the coefficient is meant to measure, but only changes the dependent variable used in the final LP.

The same correction can be used for the cumulative change LP. The original specification in Cloyne et al. (2025) uses the uncorrected cumulative change outcome $y_{t+h} - y_{t-1}$. To keep the notation readable, the corrected cumulative change outcome is denoted by $\tilde{Y}_{t,h}^{\Delta}$. The tilde is introduced here only to indicate that the outcome has been adjusted by the corrected-LP term.

$$\tilde{Y}_{t,h}^{\Delta} = y_{t+h} - y_{t-1} - \sum_{j=1}^{h-1} \widehat{B}_j^{\Delta} \hat{\varepsilon}_{t+h-j}^{\Delta}. \tag{30}$$

Here $\hat{\varepsilon}_{t+h-j}^{\Delta}$ are the cumulative change horizon zero residuals and $\widehat{B}_j^{\Delta}$ measures how these residuals propagate into the later cumulative change outcome.

The corrected cumulative change outcome is then used in the cumulative change LP:

$$\tilde{Y}_{t,h}^{\Delta} = \alpha_h^{CLP,\Delta} + \beta_h^{CLP,\Delta} x_t + \delta_h^{CLP,\Delta} s_t + \sum_{j=1}^{p} \theta_{h,j}^{CLP,\Delta} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j}^{CLP,\Delta} \Delta Z_{t-j} + v_{t+h}^{CLP,\Delta}. \tag{31}$$

Compared to the corrected level LP, this version starts from the cumulative change outcome, $y_{t+h} - y_{t-1}$, rather than the level outcome, y_{t+h}. The correction term is then subtracted from this cumulative change outcome, giving the corrected outcome $\tilde{Y}_{t,h}^{\Delta}$. The macroeconomic controls also enter as lagged first differences, ΔZ_{t-j}, instead of lagged levels, Z_{t-j}.

For the state dependent results, the corrected level outcome or corrected cumulative change outcome is inserted into the same joint ZLB/non-ZLB regression as before. The residual series and propagation coefficients are estimated using the full sample, while the tax reform coefficient is allowed to differ between ZLB and non-ZLB periods.

The corrected-left-hand-side version used here is a plug-in implementation. This means that the residuals and propagation coefficients are estimated first, and then inserted into the correction term. The final LP is then estimated using this corrected outcome. The correction is therefore applied separately for each horizon.

This follows the error-propagation idea of Lusompa (2023), but it is not the exact sequential FGLS estimator. The difference being that the correction weights are not updated recursively from previously corrected horizons, but rather that the residuals and propagation coefficients are estimated in earlier steps and then used to construct the corrected left hand side.

3.7.2 Future-Residual Control LP

The second corrected version uses the same intermediate residual information, but in a different way. Instead of first combining the residuals into one correction term and subtracting this from the left hand side, the residuals are added directly to the right hand side of the horizon specific LP. Their coefficients are then estimated together with the tax reform coefficient.

For the level LP, this gives:

$$y_{t+h} = \alpha_h^{FR} + \beta_h^{FR} x_t + \delta_h^{FR} s_t + \sum_{j=1}^{p} \theta_{h,j}^{FR} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j}^{FR} Z_{t-j} + \sum_{m=1}^{h-1} \rho_{h,m} \hat{\varepsilon}_{t+m} + e_{t+h}^{FR}. \quad (32)$$

The residual $\hat{\varepsilon}_{t+m}$ is the horizon zero residual from an intermediate quarter after the reform, but before the final outcome at $t+h$. The coefficient $\rho_{h,m}$ is the weight placed on this residual in the horizon-h regression. These weights are estimated inside the final regression itself. Meaning that, unlike the corrected-left-hand-side version, no pre-estimated weighted correction term is forced into the dependent variable.

The coefficient β_h^{FR} remains the tax reform response. The future residuals are only included to account for the part of the later-horizon regression error that can be related to unexplained movements in the intermediate quarters. They should therefore not be interpreted as ordinary future macroeconomic controls.

The cumulative change version follows the same logic:

$$\begin{aligned} y_{t+h} - y_{t-1} = {} & \alpha_h^{FR,\Delta} + \beta_h^{FR,\Delta} x_t + \delta_h^{FR,\Delta} s_t \\ & + \sum_{j=1}^{p} \theta_{h,j}^{FR,\Delta} x_{t-j} + \sum_{j=1}^{p} \Gamma_{h,j}^{FR,\Delta} \Delta Z_{t-j} \\ & + \sum_{m=1}^{h-1} \rho_{h,m}^{\Delta} \hat{\varepsilon}_{t+m}^{\Delta} + e_{t+h}^{FR,\Delta}. \end{aligned} \quad (33)$$

Here $\hat{\varepsilon}_{t+m}^{\Delta}$ is the cumulative change horizon zero residual from the intermediate quarter $t+m$, and $\rho_{h,m}^{\Delta}$ is the coefficient placed on that residual. At $h=0$ and $h=1$, there are no residuals between the reform date and the outcome date. The future-residual control LP is therefore identical to the regular LP at these horizons.

For the state dependent results, the residual leads receive common coefficients across ZLB and non-ZLB periods. Interacting every residual lead by regime would add many generated regressors to the smaller ZLB sample and would create a separate state dependent correction. This could make the estimates much noisier. The tax reform coefficient and the lagged macroeconomic coefficients remain regime specific, following the joint ZLB/non-ZLB specification used earlier.

This version is related to the residual-augmentation logic of Breitung and Brüggemann

(2023), but it is not presented as the exact Breitung-Brueggemann GLS estimator. The exact estimator is derived from a full system of innovations, while the implementation used here relies on scalar GDP residuals from the horizon zero projection. The implementation here also only includes the intermediate residuals between $t+1$ and $t+h-1$, and does not subtract the endpoint residual from the dependent variable. The idea is therefore similar, since residual information from the periods between the reform and the final outcome is added to the regression, but the exact estimator is not replicated. For that reason, the specification is described as a future-residual control LP rather than as the exact Breitung-Brueggemann GLS estimator.

The lagged macroeconomic controls remain the same in the regular and corrected LPs. The corrected-left-hand-side version changes the dependent variable, while the future residual control version adds the intermediate residuals to the already existing right hand side. The corrected LPs therefore do not replace the regular LP, but add a correction to the same underlying specification. Meaning that differences between the regular and corrected LPs should come from the correction itself, rather than from changing the control variables or basic LP setup.

3.8 Sample Timing and Calendar Aware Estimation

The estimation keeps the original quarterly calendar when leads, lags, HAC standard errors, and bootstrap residual blocks are constructed. This matters because World War II is excluded from the estimation sample, creating a gap between 1938Q4 and 1946Q1. These two quarters should not be treated as if they directly follow each other.

The WWII observations therefore remain in the dataset as missing calendar rows, but they are not used in the regressions. This keeps the quarterly time structure intact. If these rows were deleted completely, R would place the last usable pre-war observation directly before the first usable post-war observation. This would not change the OLS coefficient itself, but it could affect the standard error by treating residuals separated by several years as if they were observed in consecutive quarters.

A calendar aware HAC calculation is therefore used. The original dataset row number is retained for every observation entering the regression. At HAC lag l, two residual contributions are only paired when their original row numbers are exactly l quarters apart. The usual Bartlett weights and finite sample adjustment remain unchanged, but no covariance pair is allowed to cross the WWII gap.

The same timing logic is used for the ZLB observations. The ZLB quarters come from separate historical episodes and are not rearranged into one artificial continuous ZLB sample. The joint state dependent regression keeps the original chronological order and uses the ZLB indicator to switch the regime specific regressors on and off. The non-ZLB quarters therefore still remain in the time series, even though they do not contribute to the ZLB-specific coefficient.

Keeping the complete chronological ordering is also important for the HAC standard errors, since serial correlation should be calculated using the actual sequence of quarters. Estimating a separate ZLB regression after removing all non-ZLB observations would incorrectly treat observations from different ZLB episodes as consecutive unless the time gaps were handled explicitly.

3.9 Statistical Inference

Estimating the tax reform response only gives the point estimate $\hat{\beta}_h$. This point estimate shows the estimated size of the response, but not how precise this estimate is. A standard error is therefore needed, since the standard error determines the width of the confidence interval around the estimate, where a larger standard error implies a less precise estimate. Cloyne et al. (2025) use standard heteroskedasticity-robust standard errors for their main results. This is connected to their lag-augmented LP specification, since Montiel Olea and Plagborg-Møller (2021) show that lag augmentation can make standard robust inference reliable even when the data are persistent and the response horizon is relatively long. Cloyne et al. (2025) also report that their conclusions about statistical significance are similar when using Newey-West standard errors. The analysis here uses HAC/Newey-West standard errors as the main analytical inference choice, while the fixed regressor block bootstrap is used as an additional inference check.

Conventional OLS standard errors assume that the residual variance is constant and that residuals from different observations are unrelated. These assumptions are questionable for quarterly macroeconomic data. The size of unexpected economic movements can change over time, which is referred to as heteroskedasticity. Residuals from nearby quarters can also be related, which is referred to as serial correlation.

Serial correlation is especially relevant for LPs because the future outcomes used by neighboring observations overlap. This overlap generally increases with the horizon. Lusompa (2023) shows more formally that the horizon-h LP residual follows a moving average structure involving the intermediate innovations.

HAC/Newey-West standard errors are therefore used to account for both heteroskedasticity and serial correlation in the standard error calculation. Newey-West does not change $\hat{\beta}_h$, but rather changes the estimated standard error, confidence interval, and statistical significance attached to this response (Newey and West, 1987).

3.9.1 Newey-West Bandwidth

The Newey-West calculation requires a bandwidth. This bandwidth determines how many quarters around each observation are allowed to contribute to the serial correlation correction. This means that the covariance calculation does not only use the squared residuals from the same quarter, but also covariance terms between residuals from nearby quarters. A larger bandwidth allows residual relationships over more quarters to enter the standard error calculation, while a smaller bandwidth only uses more local dependence. A very short bandwidth may therefore ignore relevant serial correlation, while a very long bandwidth may add many imprecisely estimated autocovariances and make the standard error noisy.

The exact bandwidth rule used is not a separate theoretical result, but rather an implementation choice. The automatic bandwidth from `bwNeweyWest` is used as the starting point, while $h-1$ is imposed as a lower bound because the LP residuals at horizon h can contain dependence from the intermediate quarters. This combines a data dependent bandwidth choice with a horizon based minimum bandwidth. The Newey-West correction itself follows Newey and West (1987). The data dependent bandwidth choice is related to the general bandwidth-selection approach in Andrews (1991), while the horizon based lower bound is motivated by the moving average structure of LP errors discussed by Lusompa (2023).

The term bw_h^{NW} denotes the automatic bandwidth selected by the `bwNeweyWest` function from the R package `sandwich`. This function uses the residuals from the estimated regression to choose a data dependent Newey-West bandwidth. The term T_h denotes the number of complete observations available at horizon h. The final bandwidth is:

$$\ell_h = \min\left\{T_h - 1,\ \max\left(\left\lfloor bw_h^{NW}\right\rfloor, h-1\right)\right\}. \tag{34}$$

The value $h-1$ prevents a long-horizon regression from using a bandwidth which ignores most of the dependence created by the horizon itself. The upper bound $T_h - 1$ prevents the bandwidth from becoming larger than the available sample.

A Bartlett kernel is used inside the Newey-West calculation. This means that residual covariances receive smaller weights when the observations are further apart. For example, the covariance between residuals one quarter apart receives more weight than the covariance between residuals four quarters apart. Residuals further apart than the chosen bandwidth receive no weight. The weight at lag l is:

$$w_l = 1 - \frac{l}{\ell_h + 1}, \tag{35}$$

where w_l is the weight given to the residual covariance at lag l and ℓ_h is the chosen bandwidth. This weighting is part of the HAC covariance calculation and is not a separate regression step.

Prewhitening is switched off, meaning that no additional autoregressive model is fitted before calculating the HAC covariance matrix. The standard errors are therefore calculated directly from the LP residuals. A finite sample adjustment is applied because the analysis deals with a finite sample and not the full population.

The bandwidth rule is also compared with two simpler alternatives. The first uses only the horizon based bandwidth $h-1$. The second uses the cube-root sample size rule, where the bandwidth is based on $\lfloor T_h^{1/3} \rfloor$. These alternatives are reported as robustness checks to see whether the confidence intervals depend strongly on the bandwidth rule.

3.9.2 Pointwise Confidence Intervals

The HAC/Newey-West standard error is used to construct the pointwise 90 percent analytical confidence interval:

$$\left[\hat{\beta}_h - z_{0.95}\widehat{SE}(\hat{\beta}_h), \quad \hat{\beta}_h + z_{0.95}\widehat{SE}(\hat{\beta}_h)\right], \tag{36}$$

Here $\widehat{SE}(\hat{\beta}_h)$ is the HAC/Newey-West standard error of the estimated tax reform coefficient at horizon h. The value $z_{0.95} \approx 1.645$ is used because the interval is two-sided and covers the middle 90 percent of the normal distribution. A larger standard error gives a wider interval, meaning that the estimated response is less precise.

The interval is pointwise, meaning that it is constructed for each horizon separately. For example, the interval at $h = 4$ describes the uncertainty around the estimated response one year after the reform. It should therefore not be interpreted as a confidence band for the complete impulse response path from $h = 0$ to $h = 12$.

3.9.3 Residual Moving Block Bootstrap

The corrected LPs contain residuals and propagation coefficients which are estimated before the final horizon specific regression. A normal HAC/Newey-West confidence interval is calculated from the final regression only. Meaning that it treats the correction terms as if they were already given, while in reality they are also estimated from the data. This is one reason why a bootstrap is used as an additional inference method.

The basic idea of the bootstrap is to create many artificial samples which have the same length as the original sample and still keep part of the time-series structure of the data. The LP procedure is then repeated in every artificial sample. This gives a distribution of possible tax reform coefficients, which can be used to construct bootstrap confidence intervals.

The bootstrap used here is a fixed regressor residual moving block bootstrap. This means that the tax reform variable, the spending reform variable, the lagged controls, the dates, and the ZLB classification are kept fixed. The bootstrap therefore does not create new tax reforms, does not move the reforms to different dates, and does not generate new control variables. This follows the fixed-design bootstrap logic, where the regressors are treated as given and the artificial variation is generated through the residuals (Gonçalves and Kilian, 2004). In this application, the bootstrap creates alternative dependent variable paths by resampling the horizon zero residuals.

The residuals are sampled in moving blocks rather than one observation at a time. This is done because macroeconomic residuals can be related over time. If individual residuals were sampled separately, the time ordering of the residuals would be lost. Drawing blocks keeps consecutive residuals together and therefore preserves part of the time-series dependence in the bootstrap samples (Gonçalves and Kilian, 2004; Brüggemann et al., 2016).

The residuals are centred within each drawn block. This means that the average residual inside the drawn block is subtracted from the residuals in that block. The purpose is to keep the bootstrap residuals centred around zero, so that the bootstrap does not add an artificial average movement to the generated dependent variable.

The default block length is:

$$L = \left\lfloor T^{1/3} \right\rfloor, \tag{37}$$

where T is the estimation-sample size and L is the number of consecutive quarters in one block. In the baseline this gives a block length of seven quarters. Block lengths of 4, 7, and 12 are compared as robustness checks, since the results should not depend too strongly on one specific block length choice.

Only genuinely consecutive quarters can form a block. This means that a block cannot start before World War II and continue after it, since the WWII observations are excluded from the estimation sample. The same logic is used throughout the analysis: observations are kept in their original calendar order, but excluded periods are not treated as if they were adjacent quarters.

No single bootstrap paper directly covers the exact procedure used here. The bootstrap is therefore described as an application-specific implementation. It combines residual block resampling with fixed regressors, the narrative tax reform LP structure, and the calendar restrictions created by the excluded WWII period. The residual-resampling idea is related to Gonçalves and Kilian (2004), the moving block structure is related to Brüggemann et al. (2016), and the studentised bootstrap-t interval follows the general bootstrap-t logic discussed by Hall (1992).

For every bootstrap repetition $b = 1, \ldots, R$, the following steps are used:

1. At horizon $h = 0$, obtain the level and cumulative change residuals $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_t^{\Delta}$ from Equations (21) and (22).

2. Choose a block length L and draw moving blocks from these residual series. The baseline uses $L = \lfloor T^{1/3} \rfloor$, while block lengths of four and twelve quarters are used as robustness checks.

3. Within each drawn block, subtract the mean of the residuals in that block. This gives the centred bootstrap residuals ε_t^* and $\varepsilon_t^{\Delta *}$, $t = 1, \ldots, T$. The block structure keeps part of the time-series dependence in the residuals, while the centering keeps the bootstrap residuals centred around zero.

4. Using the estimated horizon zero coefficients, the fixed regressors, and the centred bootstrap residuals, construct the bootstrap dependent variables:

$$y_t^* = \hat{a}_0 + \hat{b}_0 x_t + \hat{d}_0 s_t + \sum_{j=1}^{p} \hat{c}_{0,j} x_{t-j} + \sum_{j=1}^{p} \hat{G}_{0,j} Z_{t-j} + \varepsilon_t^*, \tag{38}$$

$$\Delta y_t^* = \hat{a}_0^{\Delta} + \hat{b}_0^{\Delta} x_t + \hat{d}_0^{\Delta} s_t + \sum_{j=1}^{p} \hat{c}_{0,j}^{\Delta} x_{t-j} + \sum_{j=1}^{p} \hat{G}_{0,j}^{\Delta} \Delta Z_{t-j} + \varepsilon_t^{\Delta *}. \tag{39}$$

5. Using y_t^* and Δy_t^*, construct the bootstrap outcomes needed for each horizon and re-estimate the same LP specifications as in the original sample. This gives bootstrap versions of the regular level LP, the Lusompa-style corrected-left-hand-side LP, the future-residual control LP, and the corresponding cumulative change versions. The correction residuals, propagation coefficients, corrected outcomes, final LP estimates, and HAC/Newey-West standard errors are recomputed inside each bootstrap draw.

6. Store the resulting bootstrap tax reform coefficient $\hat{\beta}_{h,b}^*$ for every horizon h and every specification. The baseline uses $R = 2000$ bootstrap replications.

7. For each bootstrap repetition, calculate the studentised bootstrap statistic:

$$t_{h,b}^* = \frac{\hat{\beta}_{h,b}^* - \hat{\beta}_h}{\widehat{SE}_{h,b}^*}. \tag{40}$$

8. For the symmetric pointwise bootstrap-t interval, use the absolute value of the bootstrap-t statistics. From the bootstrap statistics $t_{h,1}^*, \ldots, t_{h,R}^*$, take the 0.90 quantile of their absolute values. This critical value is denoted by c_h^*:

$$c_h^* = Q_{0.90}\left(\left|t_{h,1}^*\right|, \ldots, \left|t_{h,R}^*\right|\right). \tag{41}$$

   The pointwise 90 percent symmetric bootstrap-t interval is then:

$$\left[\hat{\beta}_h - c_h^* \widehat{SE}_h, \quad \hat{\beta}_h + c_h^* \widehat{SE}_h\right]. \tag{42}$$

Here $\hat{\beta}_h$ is the tax reform coefficient estimated in the original sample, $\hat{\beta}_{h,b}^*$ is the corresponding coefficient from bootstrap draw b, $\widehat{SE}_{h,b}^*$ is the HAC/Newey-West standard error from the bootstrap draw, and $\widehat{SE}_h$ is the HAC/Newey-West standard error from the original sample. The critical value c_h^* is based on the absolute bootstrap-t statistics, so the interval is centred around the original estimate $\hat{\beta}_h$. The bootstrap therefore changes the critical value used to scale the original standard error, but it does not shift the interval away from the original impulse response estimate.

3.10 Robustness Checks

Several robustness checks are used to see whether the main results depend strongly on one specific modelling choice:

1. The four-lag replication baseline is compared with the AIC-selected lag orders.

2. The manually reconstructed sustained-ZLB indicator is compared with a previous-quarter-only definition, a contemporaneous definition, and the supplied `Regime0` variable.

3. The current spending reform is interacted by regime as a diagnostic, since the main

specification uses one common spending reform coefficient.

4. Six large narrative tax reforms identified in the replication code are removed one at a time to test whether the estimated response is driven by individual historical reforms.

5. Alternative HAC bandwidth rules are compared.

6. Moving block lengths of 4, 7, and 12 quarters are compared.

The spending-interaction check is only used as a diagnostic. When the spending reform is interacted by regime, some horizons contain very little useful variation for estimating the spending coefficient. This makes the regression unstable at those horizons. The preferred specification therefore uses one common spending reform coefficient.

When a large reform is removed, it is set to zero at its original date. This also removes the reform when that date enters as a lagged tax reform in later observations.

The influence exercise is applied directly to the regular LP. Repeating it for the corrected LPs would require recomputing the horizon zero residuals and every propagation coefficient after each reform is removed. Therefore, changing only the final corrected regression would not be a complete corrected-LP robustness check.

3.11 Summary of Estimated Objects

The regular level LP estimates the response of future GDP, y_{t+h}. The cumulative change LP estimates the change in GDP between the pre-reform quarter and horizon h, $y_{t+h}-y_{t-1}$. Under the identifying assumption, both specifications are meant to estimate the same population response. However, because the sample is finite and GDP is highly persistent, they can behave differently in practice.

The corrected-left-hand-side LP subtracts the estimated propagation of intermediate GDP residuals from the outcome before the final tax reform coefficient is estimated. The future-residual control LP uses the same intermediate residual information, but instead estimates the weights on these residuals directly inside the horizon specific regression.

In every specification, the coefficient of interest remains the coefficient on the narrative tax reform at date t. A coefficient of two means that a reform intended to lower taxes by an amount equal to one percent of GDP is associated with an approximate two percent increase in real GDP at that horizon, conditional on the identifying assumption and the included controls. It should not be described as the effect of an unspecified "one percent tax shock", since the construction and units of the narrative tax reform variable determine its interpretation.

The ZLB and non-ZLB responses are estimated jointly in the preferred specifications. This makes it possible to estimate the two regime specific responses in one regression and to test their difference directly.

Analytical HAC intervals and pointwise bootstrap-t intervals are kept separate. The HAC intervals use the analytical standard errors from the estimated regression, while the bootstrap intervals use the distribution of studentised bootstrap statistics. They therefore provide two different ways of describing the uncertainty around the same estimated response.

Lastly, the two corrected estimators are described using names which match their actual construction. The corrected-left-hand-side version is a Lusompa-style horizon-by-horizon plug-in estimator rather than the exact sequential FGLS estimator. The direct version is described as a future-residual control LP rather than the exact Breitung-Brueggemann GLS estimator.

4 Data

4.1 Sample and Variables

The analysis uses quarterly macroeconomic data from the United Kingdom, obtained from the replication package of Cloyne et al. (2025). The raw dataset contains 411 quarterly observations from 1918Q1 till 2020Q3. The main estimation sample is restricted to the period from 1920Q1 till 2019Q4.

World War II is excluded from the analysis, since the war economy is unlikely to describe the normal relationship between fiscal reforms and output. This excluded period runs from 1939Q1 till 1945Q4. After restricting the sample from 1920Q1 till 2019Q4 and excluding World War II, there are 372 quarterly observations before leads and lags are constructed.

The main response variable is real GDP, denoted by `lnRealGDP`. The dataset reports the logged real variables multiplied by 100, meaning that the estimated responses can be interpreted approximately as percentage responses. The main tax reform variable is `X_TaxcutToGDP`. This variable records narrative tax reforms as a share of GDP, with positive values corresponding to tax cuts. The spending reform variable, `X_SpendToGDP`, is also included in the dataset and is used as a control in the LP regressions, since spending policy movements should not be attributed to the tax reform as discussed in more detail under "Control Variables" in Section 3.2.

The construction and interpretation of these variables are discussed in Section 3.2, while their role in the LP specification is shown in Section 3.4. The macroeconomic variables entering the control vector Z_t are real GDP, real tax revenue, real government expenditure, and the Bank Rate. These are denoted by `lnRealGDP`, `lnRealTaxRev`, `lnRealExp`, and `BankRate`. In the LP regressions, these macroeconomic controls enter with four lags, as described in Section 3.4.

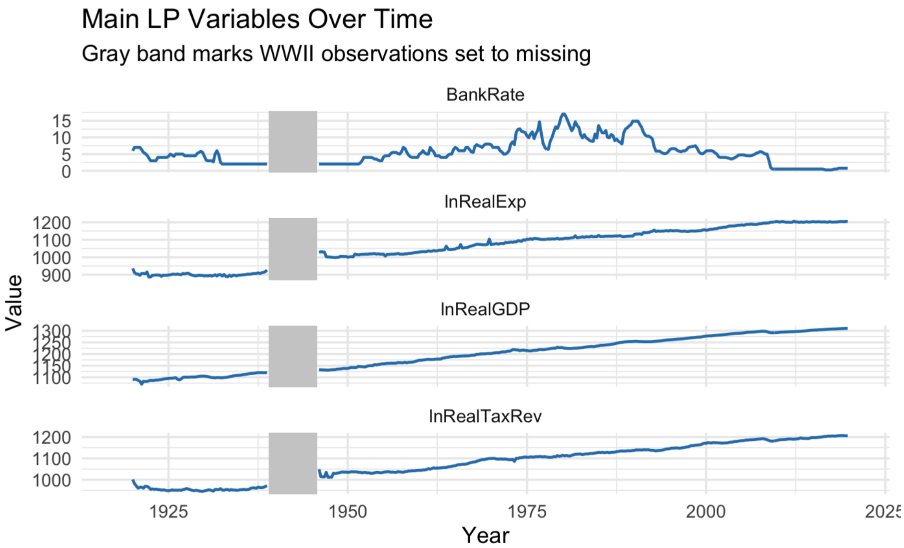

Figure 2: Main variables used in the LP analysis. The grey band marks World War II, which is excluded from the estimation sample.

Table 3: Descriptive Statistics of LP Variables

Variable	Non-missing	Mean	SD	Min	P_{25}	Median	P_{75}	Max
$\ln RealGDP$	372	1203.45	71.16	1070.48	1139.86	1214.22	1265.01	1310.14
$\ln RealTaxRev$	372	1084.30	82.96	945.62	1033.01	1099.83	1149.16	1206.85
$\ln RealExp$	372	1067.53	103.11	886.75	1012.22	1089.79	1150.77	1206.04
$BankRate$	372	5.55	3.77	0.25	2.25	5.00	7.00	17.00
$TaxReform$	372	0.05	0.22	-1.01	0.00	0.00	0.01	1.74
$SpendingReform$	372	-0.00	0.04	-0.45	0.00	0.00	0.00	0.30

The descriptive statistics suggest that both the tax and spending reforms are sparse. The median tax reform and spending reform are both zero, and the interquartile range is very small. This means that the typical observation is zero or close to zero. The identifying variation therefore does not come from continuous policy movements every quarter, but from a smaller number of narrative fiscal reforms. For the LPs, the implication is that a relatively small number of large reform episodes can have a visible influence on the estimated response.

4.2 Narrative Fiscal Reforms

The main reform variable is the tax reform scaled by GDP, where positive values represent tax cuts. The spending reform is not the main object of interest, but is included as a control. Most reform observations are equal to zero, while a small number of large narrative reforms provide much of the identifying variation. The largest tax cut reform occurs in 1953Q2, with a value of 1.738. Other large tax reform observations occur in 1988Q2, 1923Q2, 1930Q2, and 1947Q2. The largest spending reform observations are concentrated in the post-2010 period, especially around 2013 till 2018, which is part of the post-financial-crisis and quantitative-easing environment discussed in the robustness checks of Cloyne et al. (2025).

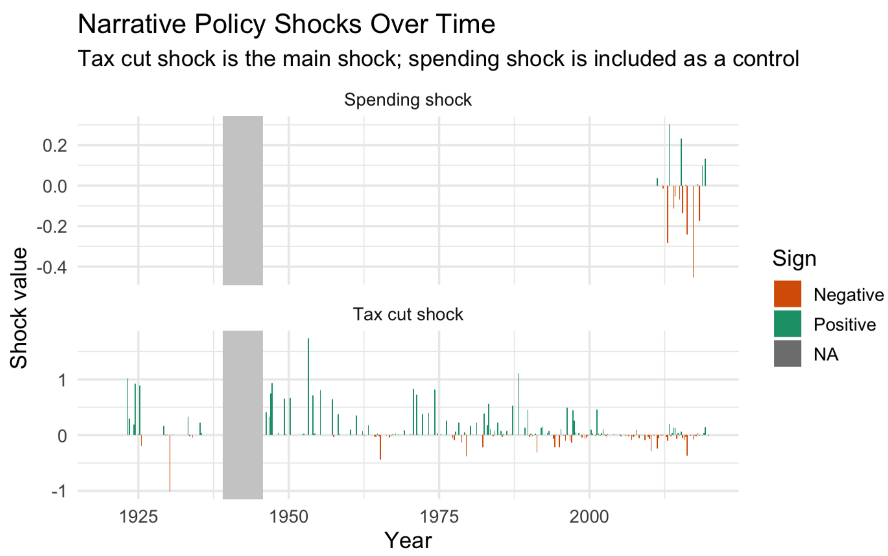

Figure 3: Narrative tax and spending reforms over time. The tax reform is the main reform variable, while the spending reform is included as a control.

4.3 ZLB Regimes

The analysis distinguishes between ZLB and non-ZLB periods using the manually reconstructed sustained-ZLB indicator defined in Section 3.5. The indicator is based on the Bank Rate. It treats the historical Bank Rate period at 2 percent and the modern period at or below 0.5 percent as lower bound episodes. The observation 1939Q3 is kept inside the historical lower bound episode, since it is a one-quarter interruption between surrounding quarters where Bank Rate is equal to 2. A quarter is then classified as ZLB only if the lower bound condition is satisfied according to the sustained four-quarter rule described in the methodology.

Using the manually reconstructed sustained-ZLB rule gives 78 ZLB observations and 291 non-ZLB observations before the LP lead and lag structure is imposed. The ZLB observations are spread across separate historical lower bound episodes rather than forming one continuous period. Once the LP regressions are estimated, the usable ZLB sample becomes smaller because the regressions require lagged controls and future GDP observations. This means that the ZLB response is estimated from fewer and more fragmented observations than the non-ZLB response, so the ZLB impulse responses should be expected to be less precise.

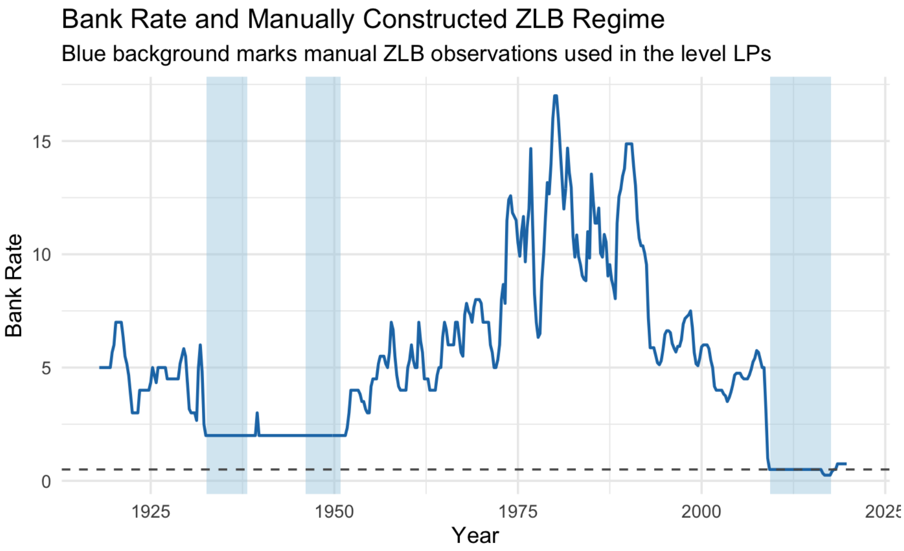

Figure 4: Bank Rate and manually constructed ZLB regime. The shaded areas indicate observations classified as ZLB in the level LP analysis.

Table 4: Usable LP Observations by Horizon

Horizon	Level total	Level non-ZLB	Level ZLB	Cum.-change total	Cum.-change non-ZLB	Cum.-change ZLB
0	361	287	74	359	286	73
4	356	286	70	354	285	69
8	348	282	66	346	281	65
12	340	281	59	338	280	58

Note: Cum.-change refers to the cumulative change LP. The total columns give the number of usable observations in the joint state dependent regression. The non-ZLB and ZLB columns split these observations by regime.

The ZLB count is smaller in the LP regressions than in the raw regime indicator because the regressions require lagged controls and future GDP observations. The first post-war ZLB quarters are lost because their lagged controls would cross the WWII gap. The cumulative change LP loses one additional post-war observation because it uses lagged first differences of the macroeconomic controls.

Table 5: Non-Zero Narrative Tax Reforms by Regime

Regime	Observations	Non-zero reform quarters	Tax cut quarters	Tax increase quarters	Largest tax cut	Largest tax increase
non-ZLB	287	165	110	55	1.738	-1.011
ZLB	74	42	19	23	0.938	-0.370

Note: The table uses the level-LP usable sample at horizon $h = 0$, after applying the lag structure.

Table 5 shows that the ZLB part of the sample contains fewer observations and fewer non-zero tax reform quarters than the non-ZLB part of the sample. This means that the ZLB impulse responses are estimated from a relatively small number of reform observations, making them more sensitive to individual reform quarters.

4.4 Descriptive Features Relevant for the LPs

The main macro variables are highly persistent. The first-order autocorrelation is approximately 0.9997 for real GDP, 0.9979 for real tax revenue, and 0.9975 for real expenditure.

The Bank Rate is also highly persistent, with a first-order autocorrelation of 0.9787. In contrast, the tax reform and spending reform variables are much less persistent. The corresponding autocorrelation figure is reported in Appendix 11.

The persistence of the macroeconomic variables is relevant for the LP analysis because movements in one quarter may also carry over into later quarters. This makes it more likely that regression errors are related across nearby quarters. This is one reason why lagged controls are included in the LP regressions and why HAC/Newey-West standard errors are used. The methodological implications of persistence are discussed in Section 3.4 and Section 3.9.

5 Results

The results are used to answer whether a tax reform intended to reduce tax liabilities by one percent of GDP is followed by an increase in real GDP when monetary policy is constrained by the ZLB. All reported responses can therefore be read as approximate percentage responses of real GDP to a tax reform of this size. The main specification is the cumulative change LP, since this is closest to the specification used by Cloyne et al. (2025). It is also useful here because real GDP is highly persistent and the ZLB sample is relatively small, meaning that the level and cumulative change specifications can behave differently in practice.

The results are discussed in five steps. First, the identification and correction diagnostics are briefly considered. Second, the main cumulative change LP results are presented. Third, the level LP is used as a comparison. Fourth, the ZLB and non-ZLB responses are compared directly. Lastly, the fixed regressor residual moving block bootstrap and the robustness checks are used to see how strong the results remain once alternative inference and specification choices are considered.

As discussed in Section 3.9, two types of confidence intervals are used in the results. The first are analytical asymptotic intervals based on HAC/Newey-West standard errors, referred to as HAC intervals in this section. The second are symmetric pointwise bootstrap-t intervals from the fixed regressor residual moving block bootstrap, referred to as bootstrap intervals in this section. These bootstrap intervals use absolute bootstrap-t critical values, meaning that the interval is centred around the original impulse response estimate. The caption of each figure or table states which type of interval is shown.

5.1 Identification and Correction Diagnostics

Before interpreting the impulse responses, two diagnostic points are useful. The first concerns the narrative tax reform variable. If this variable is to be treated as plausibly external to short run macroeconomic conditions, it should not be strongly predictable from the included macroeconomic history. The Granger-style predictability test is therefore used as a diagnostic check. It does not prove exogeneity, since no predictability test can rule out every possible omitted policy motivation, but it can show whether the included lagged variables strongly predict the current reform.

Table 6: Predictability Test for the Narrative tax reform Variable

Wald statistic	Degrees of freedom	p-value	Observations
23.88	20	0.25	362

This Wald test is close in spirit to the Granger-style predictability check in Cloyne et al. (2025), but it is not an exact replication of their VAR-based test. The test does not find strong evidence that the included lagged variables jointly predict the current tax reform, with a p-value of 0.25. This means that, using the included lagged macroeconomic variables and spending reforms, there is no strong evidence that the tax reform series is predictable from earlier economic conditions. This is consistent with treating the narrative tax reform variable as plausibly external to short run macroeconomic conditions, but it should still be read as a diagnostic rather than as proof that the reform variable is fully exogenous.

The second diagnostic point concerns the corrected LPs. The corrected LPs are useful only if intermediate LP residuals contain information about later outcomes. The estimated propagation coefficients are positive and sizeable, meaning that unexplained movements in GDP at short horizons are related to later GDP outcomes. This gives empirical motivation for comparing the regular LP with the corrected LPs.

5.2 Cumulative Change Responses

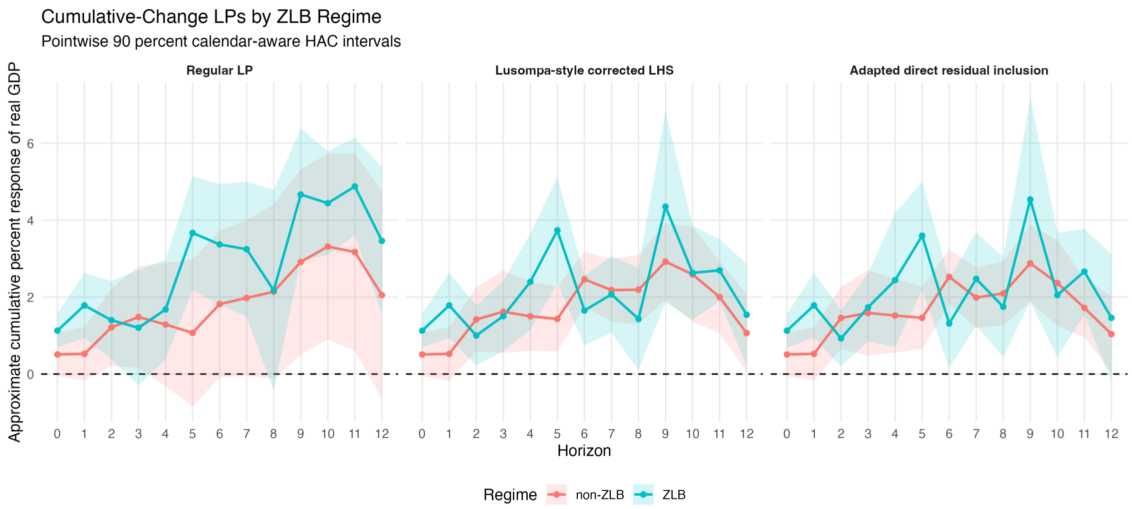

Figure 5: Cumulative change GDP responses following a narrative tax reform. The shaded areas show pointwise 90 percent calendar aware HAC confidence intervals.

Figure 5 shows the main cumulative change results. The dependent variable is $y_{t+h} - y_{t-1}$, meaning that the coefficient measures the change in real GDP between the quarter before the reform and horizon h. A positive coefficient therefore means that real GDP is estimated to be higher after a tax cut reform, relative to the pre-reform quarter.

Table 7: Selected Cumulative Change LP Results

Estimator	Regime	$h = 0$	$h = 4$	$h = 8$	$h = 12$
Regular LP	non-ZLB	0.51 [-0.06, 1.07]	1.29 [-0.32, 2.89]	2.14 [-0.13, 4.41]	2.06 [-0.65, 4.76]
Regular LP	ZLB	1.13 [0.69, 1.56]	1.68 [0.39, 2.97]	2.18 [-0.43, 4.79]	3.46 [1.55, 5.38]
Lusompa-style corrected LHS	non-ZLB	0.51 [-0.06, 1.07]	1.50 [0.61, 2.39]	2.19 [1.30, 3.09]	1.07 [0.06, 2.07]
Lusompa-style corrected LHS	ZLB	1.13 [0.69, 1.56]	2.40 [1.15, 3.65]	1.43 [0.11, 2.75]	1.54 [0.22, 2.86]
Future-residual control LP	non-ZLB	0.51 [-0.06, 1.07]	1.52 [0.56, 2.48]	2.10 [1.26, 2.93]	1.04 [0.05, 2.03]
Future-residual control LP	ZLB	1.13 [0.69, 1.56]	2.44 [0.70, 4.18]	1.75 [0.45, 3.05]	1.46 [-0.19, 3.11]

Note: Values in brackets are pointwise 90 percent HAC intervals.

At $h = 0$, the regular and corrected estimators coincide because there are no intermediate residuals between the reform date and the outcome date. The point estimate and confidence interval are therefore identical across the three estimators at this horizon. The regular cumulative change LP produces positive point estimates in both regimes. In the non-ZLB regime, the estimated response equals approximately 0.51 percent during the reform quarter, 1.29 percent after four quarters, 2.14 percent after eight quarters, and 2.06 percent after twelve quarters. The corresponding ZLB estimates are 1.13, 1.68, 2.18, and 3.46 percent. Thus, the main point estimate pattern does not suggest that tax reforms become contractionary at the ZLB.

The corrected LPs do not overturn the positive cumulative change response. Instead, they mainly change the shape and size of the response. At $h = 4$, the ZLB response increases from 1.68 in the regular LP to 2.40 under the Lusompa-style corrected-left-hand-side version and 2.44 under the future-residual control version. At $h = 8$ and $h = 12$, the corrected ZLB responses are smaller than the regular LP response, but remain positive.

The corrected LPs do not simply smooth the response path or lower the response at every horizon. Instead, they change the timing and size of the estimated response. In the ZLB regime, the corrected responses are larger than the regular LP at some intermediate horizons, but smaller at the later horizons from $h = 10$ to $h = 12$. The corrected LPs therefore reduce the large late horizon response in the regular cumulative change LP. This is consistent with the correction logic described in Section 3.7, since part of the regular late horizon response may reflect intermediate residual movements rather than only the tax reform coefficient itself.

Under HAC inference, the corrected cumulative change estimates are also more often statistically different from zero than the regular LP. For the ZLB response, the Lusompa-style corrected interval is above zero at every horizon, while the future-residual control interval is above zero through $h = 11$. The full horizon-by-horizon intervals are reported in Appendix A.6.

5.3 Comparison With the Level LP

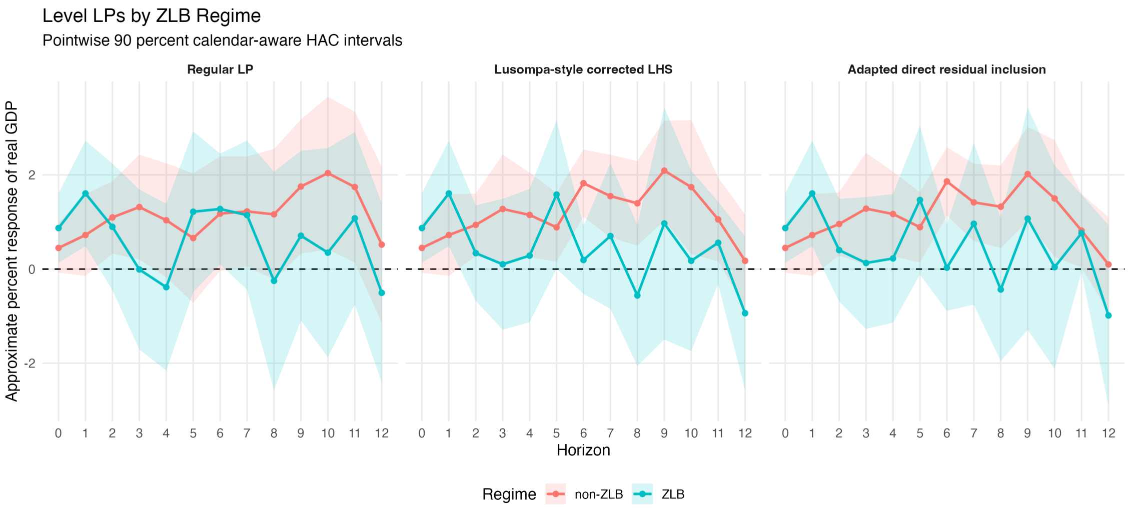

Figure 6: Level GDP responses following a narrative tax reform. The shaded areas show pointwise 90 percent calendar aware HAC confidence intervals.

The level LP gives a less stable picture than the cumulative change LP. Under the regular level LP, the ZLB response equals 0.87 percent during the reform quarter, but becomes -0.38 percent after four quarters, -0.25 percent after eight quarters, and -0.50 percent after twelve quarters. Every selected ZLB interval apart from the contemporaneous interval includes zero.

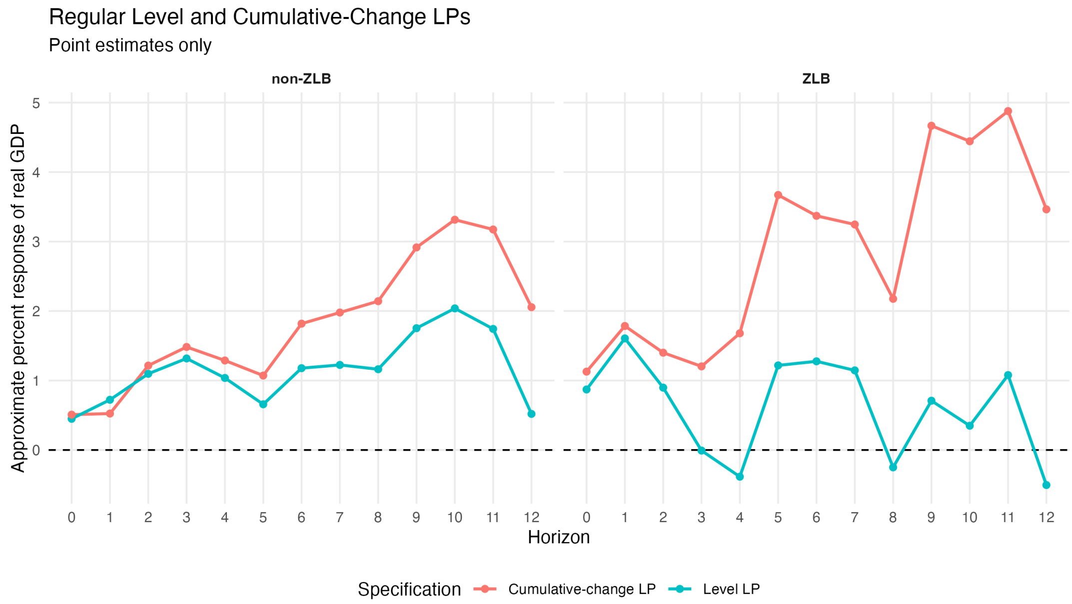

Figure 7: Comparison of the regular level and cumulative change LP point estimates. Both specifications use the same narrative tax reform series and sustained-ZLB classification.

Figure 7 shows that the difference between the level and cumulative change specification is especially visible in the ZLB regime. As a descriptive comparison, the correlation between the level and cumulative change point estimate paths is 0.84 in the non-ZLB regime, but only 0.17 in the ZLB regime. In the non-ZLB regime, both specifications have the same sign at all thirteen horizons, while in the ZLB regime the two specifications have the same sign at only nine horizons.

This difference affects the interpretation. It does not necessarily mean that one specification is simply right and the other is wrong. Rather, it shows that the dependent variable choice matters in this historical sample. Since real GDP is highly persistent and the ZLB sample is relatively small, the cumulative change LP is used as the main specification, while the level LP is used as a comparison.

5.4 Regime Differences

The positive ZLB estimates do not automatically mean that the ZLB response is statistically larger than the non-ZLB response. These are two separate questions. The ZLB response itself shows whether tax reforms are estimated to be expansionary or contractionary when monetary policy is constrained. The ZLB minus non-ZLB difference instead shows whether the estimated response is different from the response in normal periods.

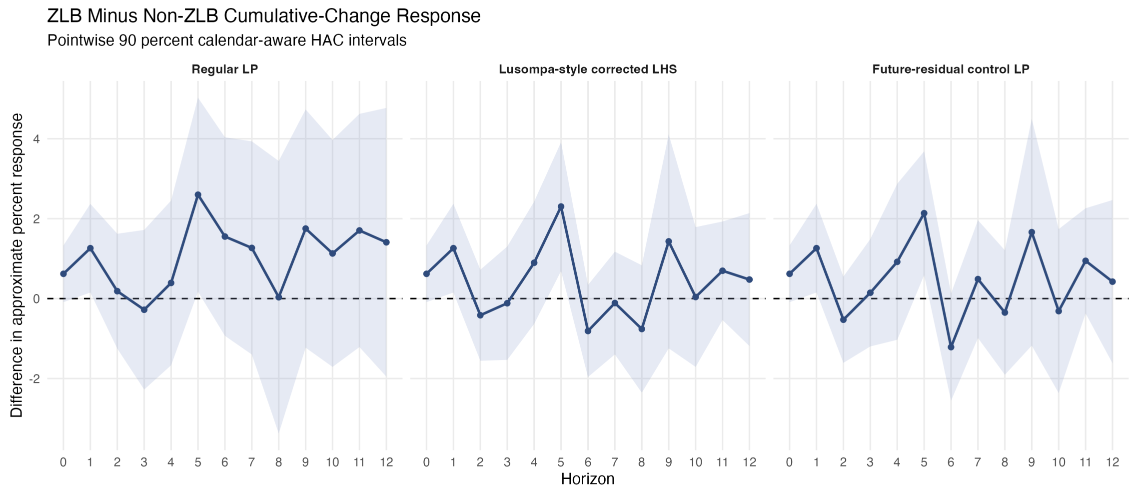

Figure 8: ZLB minus non-ZLB cumulative change response. Positive values mean that the estimated response is larger at the ZLB. The shaded areas show pointwise 90 percent calendar aware HAC confidence intervals.

Figure 8 shows the ZLB minus non-ZLB difference for the cumulative change specifications. In the regular cumulative change LP, the difference is positive at most horizons, meaning that the ZLB point estimate is usually larger than the non-ZLB point estimate. However, the HAC confidence intervals are wide and include zero at most horizons. The exceptions are around $h = 1$ and $h = 5$, where the pointwise HAC intervals exclude zero.

These pointwise differences are not confirmed by the bootstrap intervals, which include zero at all horizons. The simultaneous bands also include zero throughout, as reported in Appendix A.7. The full regime difference intervals are reported in Appendix A.6. The results therefore do not provide robust evidence that the ZLB response is statistically larger than the non-ZLB response. The main result is instead that the cumulative change estimates do not support the prediction that tax reforms become contractionary at the ZLB.

Appendix A.7 reports the corresponding bootstrap intervals for the ZLB minus non-ZLB difference. The point estimates are positive at some horizons, but the symmetric pointwise bootstrap intervals include zero at all horizons. The simultaneous bands also include zero throughout. The regime difference evidence is therefore not strong enough to conclude that the ZLB response is clearly different from the non-ZLB response over the full path.

5.5 Fixed Regressor Residual Moving Block Bootstrap

The bootstrap is used as an additional inference check. The preferred bootstrap uses 2,000 draws and a block length of seven quarters. The tax reforms, spending reforms, lagged controls, dates, and ZLB classifications are kept fixed. Blocks of horizon zero residuals are resampled and centred within each drawn block. The bootstrap dependent variables are then rebuilt and the LPs are re-estimated for every horizon.

The bootstrap therefore changes the dependent variable path, while keeping the regressors and the historical timing of the reforms fixed. The symmetric bootstrap-t intervals are centred around the original IRF estimate by construction, as discussed in Section 3.9.3. Table 8 reports the symmetric bootstrap-t intervals for the ZLB cumulative change response at the selected horizons. These are the same selected horizons used in the HAC table, so the bootstrap and analytical intervals can be compared directly.

Table 8: Symmetric Pointwise 90 Percent Bootstrap-t Intervals for the ZLB Cumulative Change Response

Estimator	$h = 0$	$h = 4$	$h = 8$	$h = 12$
Regular LP	[0.32, 1.94]	[-0.11, 3.47]	[-2.03, 6.38]	[-0.86, 7.78]
Lusompa-style corrected LHS	[0.32, 1.94]	[-0.02, 4.82]	[-0.36, 3.23]	[-0.35, 3.43]
Future-residual control LP	[0.32, 1.94]	[-0.64, 5.52]	[-0.07, 3.57]	[-0.79, 3.71]

At $h = 0$, the three estimators coincide because there are no intermediate residuals between the reform date and the outcome date. The bootstrap interval is therefore identical across the three rows and remains above zero. At the selected later horizons shown in the table ($h = 4$, $h = 8$, and $h = 12$), the point estimates remain positive, but all bootstrap intervals include zero. Figure 9 shows the full horizon path. Across the full path, some pointwise bootstrap intervals still exclude zero, so the bootstrap does not remove all evidence of a positive response. It does however make the evidence less uniform across horizons than the analytical HAC intervals.

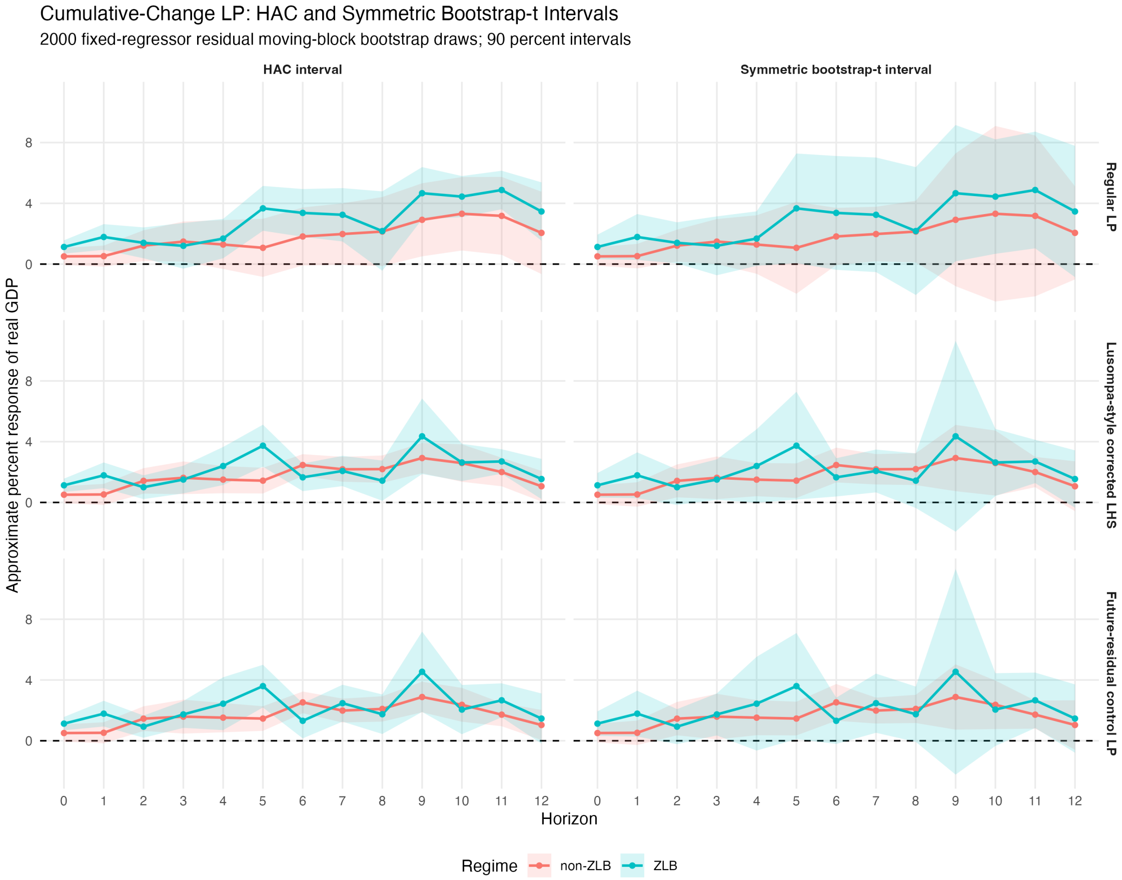

Figure 9: Symmetric pointwise bootstrap-t intervals for the cumulative change LP responses. The intervals are based on the fixed regressor residual moving block bootstrap with block length seven.

The symmetric bootstrap intervals are centred around the original IRF estimate by construction. They should therefore be read as uncertainty bands around the estimated response, rather than as a separate bootstrap response path. Compared with the HAC intervals, the bootstrap bands are generally wider, especially for the ZLB response at later horizons.

This makes the bootstrap evidence more cautious. The cumulative change ZLB point estimates remain positive at the selected horizons, but the symmetric bootstrap intervals include zero at several intermediate and later horizons. Some horizons still exclude zero under pointwise bootstrap inference, so the bootstrap does not remove all evidence of a positive response. It mainly shows that the evidence is less uniform across horizons and less precise than the analytical HAC intervals suggest.

The corrected LPs should also not be read as simply smoothing the response path. They do not lower the response at every horizon either. Instead, they change the timing and size of the estimated response. For the ZLB response, the corrected LPs are larger than the regular LP at some intermediate horizons, but smaller at the later horizons from $h = 10$ to $h = 12$. This suggests that the correction matters most in the part of the response

where intermediate residual propagation becomes more important.
Overall, the bootstrap still points to a positive ZLB response at selected horizons, but it also shows that the response is not estimated very precisely over the full three-year horizon. The positive result is therefore easier to defend horizon by horizon than for the full response path at once. Appendix A.7 reports the simultaneous sup-t bands, which are wider because they ask whether the whole response path is covered jointly rather than horizon by horizon.

5.6 Robustness Checks

The robustness checks show that the positive cumulative change pattern is not driven by one single modelling choice, although the size of the ZLB response changes across specifications. The detailed robustness figures are reported in Appendix A.3.
Appendix A.3.1 compares the four-lag baseline with the AIC-selected lag orders. The cumulative change ZLB response remains positive under the AIC-selected lag length, although the magnitude changes at longer horizons. This means that the sign of the main cumulative change result is not only created by the four-lag replication choice.
Appendix A.3.2 compares the sustained-ZLB baseline with alternative ZLB definitions. The supplied `Regime0` variable gives the same response as the previous-quarter-only definition, while the sustained four-quarter baseline gives different magnitudes. The positive cumulative change pattern remains visible, but the size of the ZLB response depends on how the ZLB regime is defined. This is consistent with the small number of ZLB observations discussed in Section 4.
Appendix A.3.3 reports the leave-one-reform-out exercise. After each selected large tax reform is set to zero separately, the cumulative change ZLB response remains positive at the selected horizons. The magnitude still moves, meaning that individual historical reforms influence the estimated response, but the positive ZLB response is not created by one single selected reform.
Appendix A.5 compares bootstrap block lengths of four, seven, and twelve quarters. The width of the bootstrap intervals changes with the block length, but the main interpretation remains the same. The bootstrap gives more cautious evidence than the analytical HAC intervals, especially for the later ZLB horizons.
Appendix A.4 reports the spending reform interaction diagnostic. The preferred specification uses one common spending reform coefficient. When the spending control is interacted by regime, the design becomes unstable at several horizons. This supports keeping the common spending coefficient in the main specification.

5.7 Main Result

The central result is that the positive ZLB response is robust in sign, but not in statistical strength. Across the regular cumulative change LP, both corrected cumulative change LPs, alternative lag lengths, alternative ZLB definitions, and the leave-one-reform-out exercise, the estimated ZLB response to a tax reform intended to reduce tax liabilities by one percent of GDP remains positive at the selected horizons. In this sense, the results do not support the prediction that tax reforms become contractionary at the ZLB.

The strength of the evidence, however, depends systematically on two choices that the single specification in Cloyne et al. (2025) could not separate. The first is the dependent variable transformation. The cumulative change LP gives a stable positive response, while the level LP is less stable and becomes negative at some ZLB horizons. Its estimated response path is also only weakly correlated with the cumulative change response path in the ZLB regime. The second is how uncertainty is measured. The response is statistically different from zero at selected horizons under pointwise HAC and bootstrap-t intervals, but the simultaneous sup-t bands are much wider and do not support a uniformly significant positive response over the full path. The positive ZLB response is therefore easier to defend horizon by horizon than as a jointly significant response over the full three-year horizon.

The ZLB response is also not shown to be statistically larger than the non-ZLB response. The main conclusion is therefore not that tax reforms work more strongly at the ZLB, but that the estimates do not support the prediction that tax reforms become contractionary when monetary policy is constrained.

6 Relation to Cloyne et al.'s Robustness Checks

The robustness checks in this paper focus on the parts of the analysis which are changed relative to Cloyne et al. (2025). These are the LP specification, the corrected LP implementation, the ZLB classification, the HAC bandwidth rule, the bootstrap block length, and the influence of selected large reforms. The results of these checks are summarised in Section 5.6, while the corresponding figures are reported in Appendix A.3. The full HAC and bootstrap interval tables are reported in Appendix A.6.

Some robustness checks in Cloyne et al. (2025) are discussed rather than reproduced because they concern the construction and timing of the narrative tax reform series itself. Since the present paper uses their constructed narrative series, these checks are not the main object of the analysis. For example, Cloyne et al. (2025) show that their results are robust when the sample is restricted to tax changes implemented within 90 days of announcement. This addresses the concern that anticipated reforms may affect the economy before the implementation date. They also show that their results are robust when tax shocks close to quantitative easing episodes are excluded from the post-2010 period, where unconventional monetary policy could otherwise make the interpretation less clean.

Cloyne et al. (2025) also report additional robustness checks which are not repeated one by one here. These include alternative ZLB-window definitions, alternative historical ZLB episode definitions, different lag and control sets, different spending control choices, alternative tax shock series, a skipped sample comparison, and comparisons between smoothed and raw responses. Some of these are covered in a related way in this paper, for example through the ZLB-definition check, the lag choice check, the spending control diagnostic, and the smoothed response appendix. However, the exact robustness grid from Cloyne et al. (2025) is not repeated here. This is because the contribution of this paper is different. The aim is not to rebuild every robustness exercise around the original narrative series, but to test whether the positive ZLB response remains visible when the LP construction, corrected-LP specification, and inference method are changed.

7 Conclusion

This paper set out to ask whether tax reforms intended to reduce tax liabilities still increase output when monetary policy is constrained by the ZLB. The theoretical motivation comes from the New Keynesian prediction that some distortionary tax cuts can become contractionary at the ZLB (Eggertsson, 2011). The empirical starting point is Cloyne et al. (2025), who find that tax cuts remain expansionary both at and away from the ZLB. The contribution here is to test whether this conclusion remains visible when the LP specification, corrected-LP construction, and inference method are changed.

The main cumulative change results do not support the contractionary ZLB prediction. A tax reform intended to reduce tax liabilities by one percent of GDP is followed by a positive estimated GDP response in the ZLB regime. This remains true for the regular cumulative change LP and for both corrected cumulative change LPs at the selected horizons. The corrected LPs change the timing and size of the response, especially at later horizons, but they do not turn the main ZLB cumulative change response negative. A separate result is that the dependent variable transformation changes the interpretation. Consistent with the discussion in Section 3.4, the cumulative change LP gives the clearest positive ZLB response, while the level LP is much less stable and becomes negative at some later horizons.

The results should still be interpreted with caution. The bootstrap intervals show that the longer horizon cumulative change estimates are not always estimated very precisely. The positive ZLB response is therefore easier to defend horizon by horizon than as a jointly significant full response path over all horizons.

The results also do not show clear evidence that tax reforms work more strongly in the ZLB state than in non-ZLB periods. The ZLB minus non-ZLB difference is often positive in the cumulative change specification, but the confidence intervals for this difference generally include zero. The main conclusion is therefore that the estimates do not suggest that tax reforms become contractionary when monetary policy is constrained. They also do not provide strong evidence that the ZLB response is statistically larger than the non-ZLB response.

The results are close to Cloyne et al. (2025), but with a more cautious interpretation. Tax reforms appear able to support output at the ZLB in the main cumulative change specification, but the strength of the evidence depends on the dependent variable choice and on the inference method. The cumulative change specification gives the clearest positive result, while the level LP and bootstrap inference make the final interpretation more cautious.

There are also several limitations. The ZLB sample is small and comes from separate historical episodes, making the estimates sensitive to individual reforms and to the exact ZLB definition. The narrative tax reform variable is sparse since most quarters contain no tax reform. This is expected for a narrative reform series, but it means that the tax reform coefficient is identified from a smaller number of non-zero reform observations.

As a result, the estimate can be less precise and more sensitive to individual reform quarters. The narrative tax reform series is taken from Cloyne et al. (2025). Their identification argument is that reforms classified as exogenous are not motivated by current or expected short run macroeconomic conditions, and they support this with Granger-style predictability tests. This paper therefore inherits their identifying assumption. The assumption remains imperfect however, since some reforms may still have been chosen using information about future economic conditions which is not captured by the included controls. If that information also affects future GDP, the reform variable could still be correlated with the remaining GDP disturbance. The narrative reform variable may also contain measurement or classification error, since both the intended revenue size and the motivation of historical reforms are constructed from policy records. The identifying assumption does not require the measure to be perfect, but it does require that this error is not systematically related to the remaining GDP disturbances after the controls are included (Stock and Watson, 2018). The corrected LPs are also implemented as Lusompa-style and future-residual-control adaptations, rather than as exact replications of the original theoretical estimators.

Several extensions follow from these limitations. The remaining endogeneity concern could be studied more directly, for example with an LP-IV setup. A further specification check would be to include deterministic trends or other low frequency controls. The main specifications follow Cloyne et al. (2025), but the sample is long and real GDP trends strongly upward over this period, so it would be useful to check whether the results are sensitive to adding a trend. Bootstrap procedures developed more directly for fixed regressor LP applications would also be useful, since the bootstrap intervals are especially relevant for the late horizon interpretation. An iterated or double bootstrap could be used to check whether the studentised bootstrap distribution itself behaves well in this finite sample (Hall, 1992). Another extension would be to allow for structural breaks or time varying parameters. The sample covers 1920Q1 to 2019Q4 and the ZLB observations come from historically different episodes, so this would test whether the estimated ZLB response is really comparable between the interwar lower bound period and the post-2008 lower bound period. Bias corrected LP estimators could also be considered, since persistence and finite sample behaviour remain important concerns in this paper (Herbst and Johannsen, 2021). A related extension would be to apply the corrected LPs to the leave-one-reform-out exercise, since the current influence check is applied to the regular LP, as discussed in Section 5.6.

A Additional Tables and Figures

A.1 Smoothed Response for Comparison with the Original Paper

The main results use the unsmoothed horizon specific LP estimates. A smoothed version of the regular cumulative change response is only included for comparison with the graphical presentation in Cloyne et al. (2025). It is therefore used as a visual comparison and not as the basis for the main inference.

The smoothing is applied after the horizon specific LP coefficients have been estimated. The dependent variable is not smoothed before estimation. For the interior horizons, the smoothed response is:

$$\bar{\beta}_h = \frac{\hat{\beta}_{h-1} + \hat{\beta}_h + \hat{\beta}_{h+1}}{3}, \qquad h = 1, \ldots, 11. \tag{43}$$

The endpoints $h = 0$ and $h = 12$ are left unchanged. The smoothed line should therefore be read only as a presentation device. It is not a separately estimated LP response. The unsmoothed confidence intervals are also not simply averaged, since confidence intervals for a smoothed response would require the covariance between coefficients estimated at different horizons.

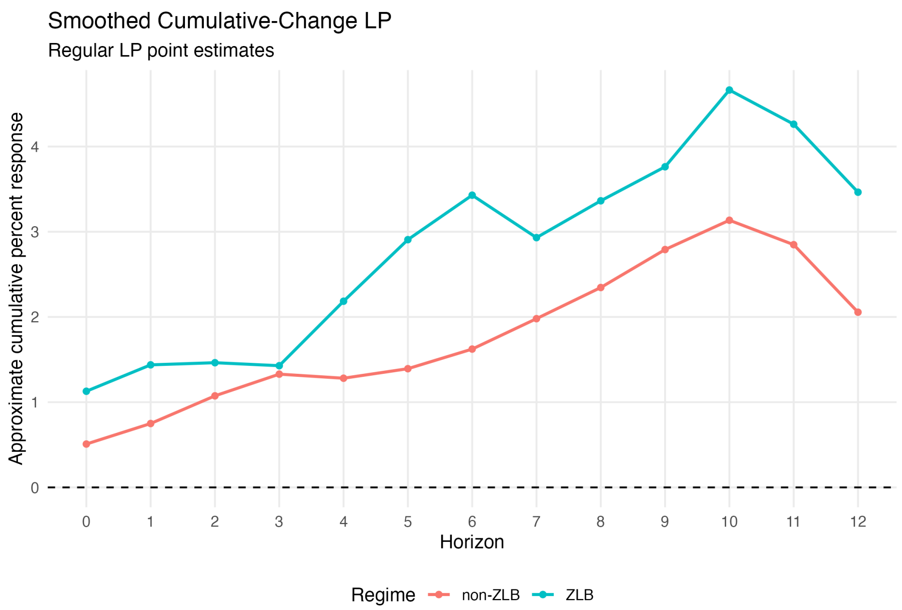

Figure 10: Smoothed regular cumulative change LP response, included for visual comparison with the graphical presentation in Cloyne et al. (2025).

Figure 10 shows the same general positive cumulative change pattern as the unsmoothed estimates. The empirical conclusions in the main text remain based on the unsmoothed responses.

A.2 Autocorrelation of LP Variables

Figure 11 shows the autocorrelation of the main variables used in the LP analysis. This figure is placed in the appendix because the main data section only needs the persistence point to motivate lagged controls and HAC/Newey-West inference.

Figure 11: Autocorrelation of the main variables used in the LP analysis. The macroeconomic variables and Bank Rate are highly persistent, while the narrative reform variables are much less persistent.

A.3 Robustness Figures

This appendix reports robustness figures which support the results section. They are kept in the appendix because they do not change the central result, but they show how sensitive the cumulative change response is to implementation choices.

A.3.1 Lag Choice

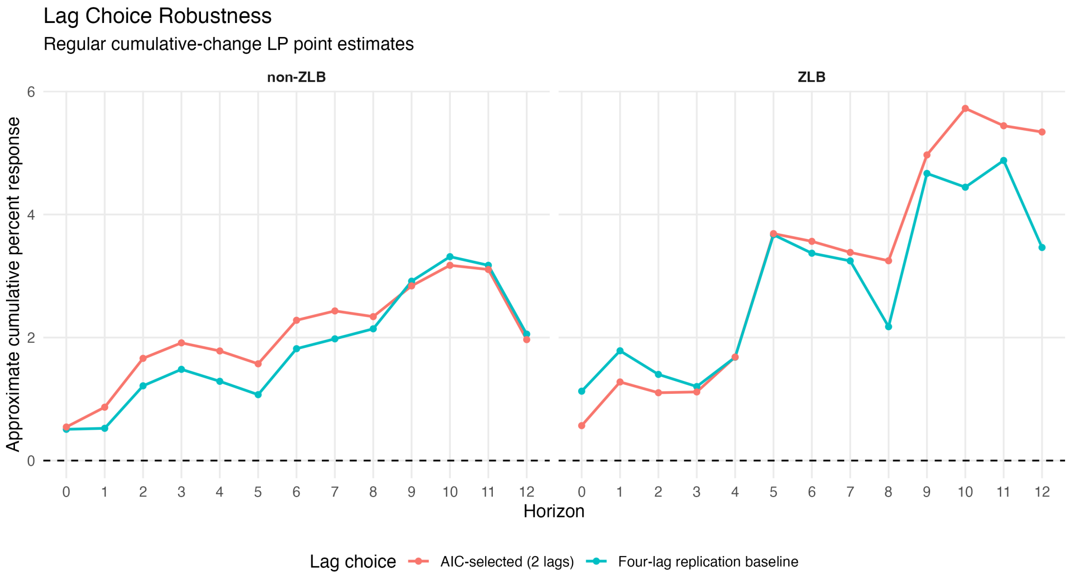

Figure 12: Lag choice robustness for the regular cumulative change LP. The four-lag baseline is compared with the AIC-selected lag specification.

Figure 12 shows that changing the lag length affects the size of the response, especially in the ZLB sample. However, it does not remove the positive cumulative change pattern. This supports using the four-lag specification as the replication baseline, while still checking a data dependent alternative.

A.3.2 ZLB Definition

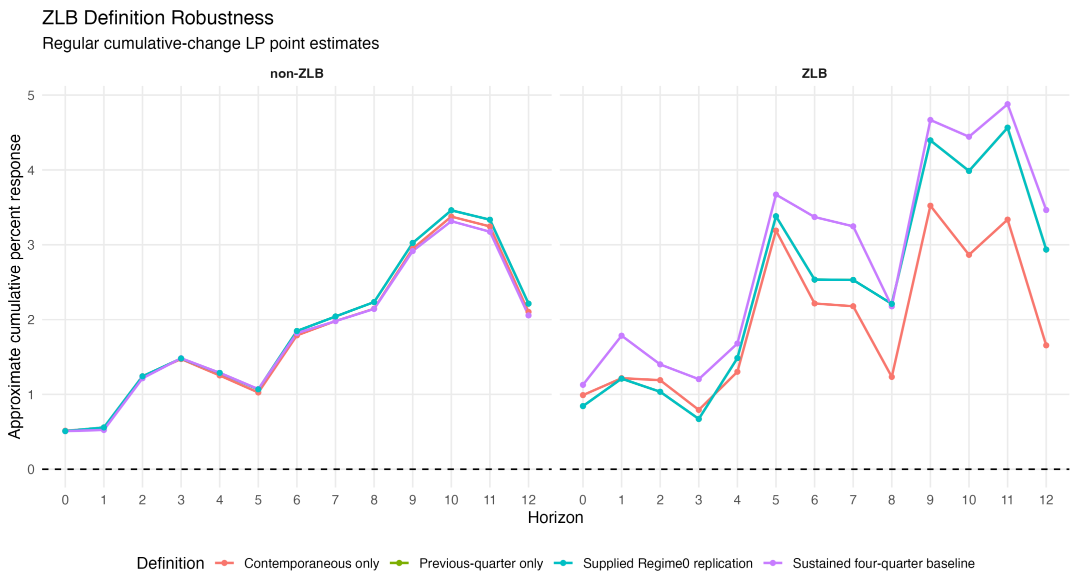

Figure 13: ZLB-definition robustness for the regular cumulative change LP. The sustained four-quarter baseline is compared with alternative ZLB classifications.

Figure 13 shows that the ZLB estimates are sensitive to the exact regime definition, which is expected given the small number of ZLB observations. The supplied `Regime0` classification gives the same response as the previous-quarter-only definition and is therefore hidden behind that line. This supports using the manually reconstructed sustained-ZLB indicator in the main analysis, while keeping `Regime0` as a robustness check.

A.3.3 Leave-One-Reform-Out

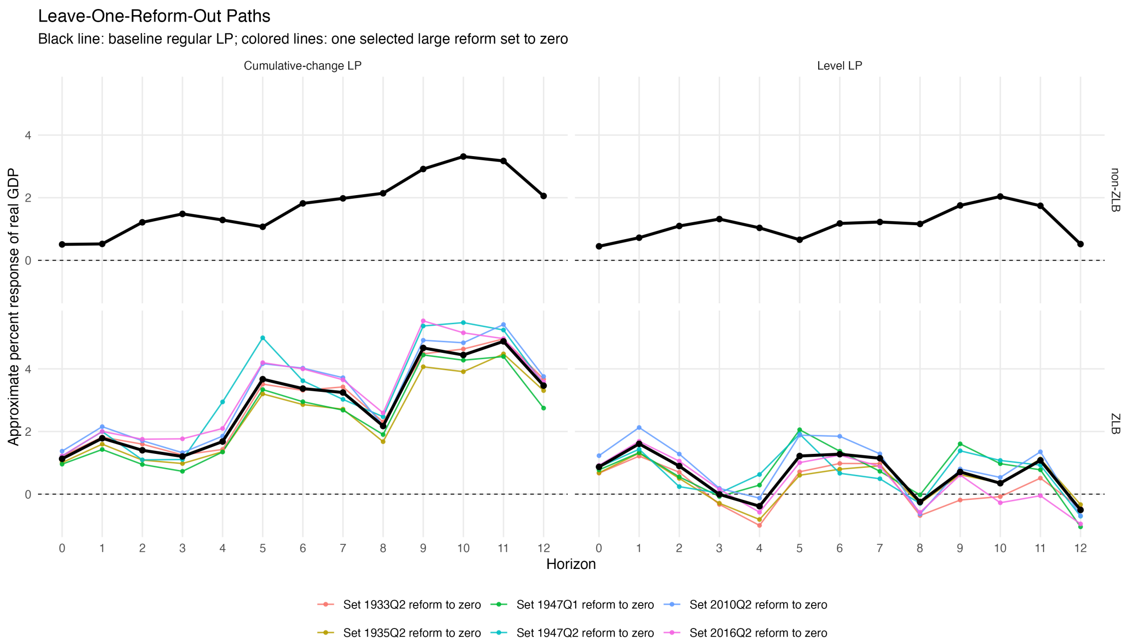

Figure 14: Leave-one-reform-out robustness check. The black line shows the baseline regular LP estimate. Each coloured line sets one selected large narrative tax reform to zero while keeping the rest of the sample unchanged.

Figure 14 shows the leave-one-reform-out exercise. The black line gives the baseline regular LP estimate, while each coloured line shows the result after setting one selected large narrative tax reform to zero. The observation itself is not deleted, meaning that GDP, the macroeconomic controls, the spending reform, and the ZLB classification remain unchanged. Only the tax reform value at that date is removed from the tax reform series.
For the cumulative change LP, the ZLB response remains positive across all selected horizons after each reform is removed. This suggests that the positive cumulative change ZLB pattern is not driven by one single large reform. The level LP remains less stable, which is consistent with the main results.

A.4 Spending-Control Interaction Diagnostic

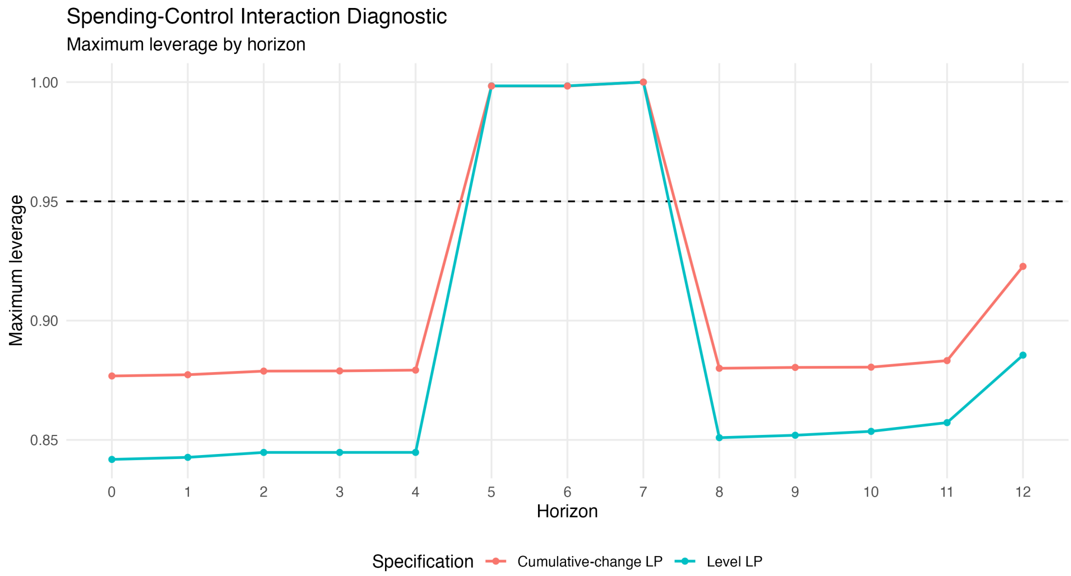

Figure 15: Maximum leverage when the spending reform control is interacted with the ZLB regime.

Figure 15 shows why the preferred specification uses one common spending reform coefficient. When the spending reform is interacted with the ZLB regime, maximum leverage becomes very high around horizons 5 to 7 and reaches one at horizon 7. From horizon 8 onward, the non-ZLB spending interaction has no variation and must be removed. This makes the fully interacted spending-control version too fragile to use as the main specification.

A.5 Bootstrap Block Length

The baseline bootstrap uses a block length of seven quarters. Figure 16 compares this baseline with block lengths of four and twelve quarters. The purpose is to check whether the bootstrap intervals depend strongly on one specific block length choice.

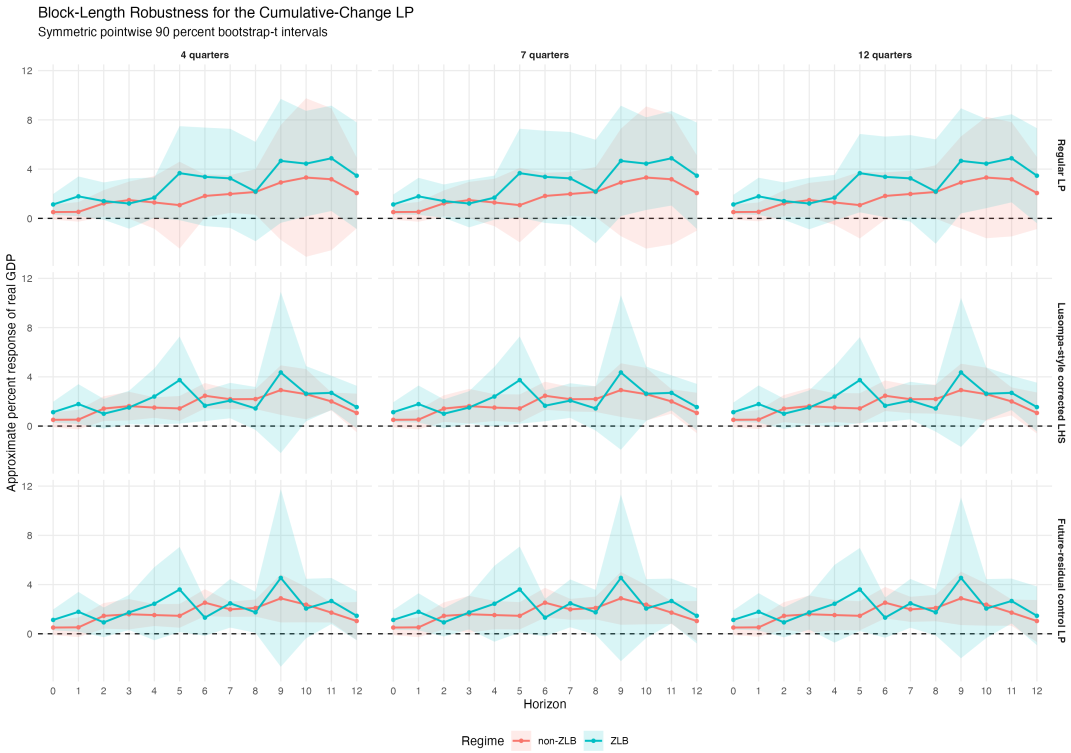

Figure 16: Symmetric pointwise bootstrap-t intervals for the cumulative change LP under alternative block lengths.

The intervals become wider or narrower at some horizons when the block length changes. However, the overall pattern is similar across the three choices. The cumulative change ZLB estimates remain positive at the selected horizons, while the intervals become wide enough at later horizons that the response is not estimated very precisely.

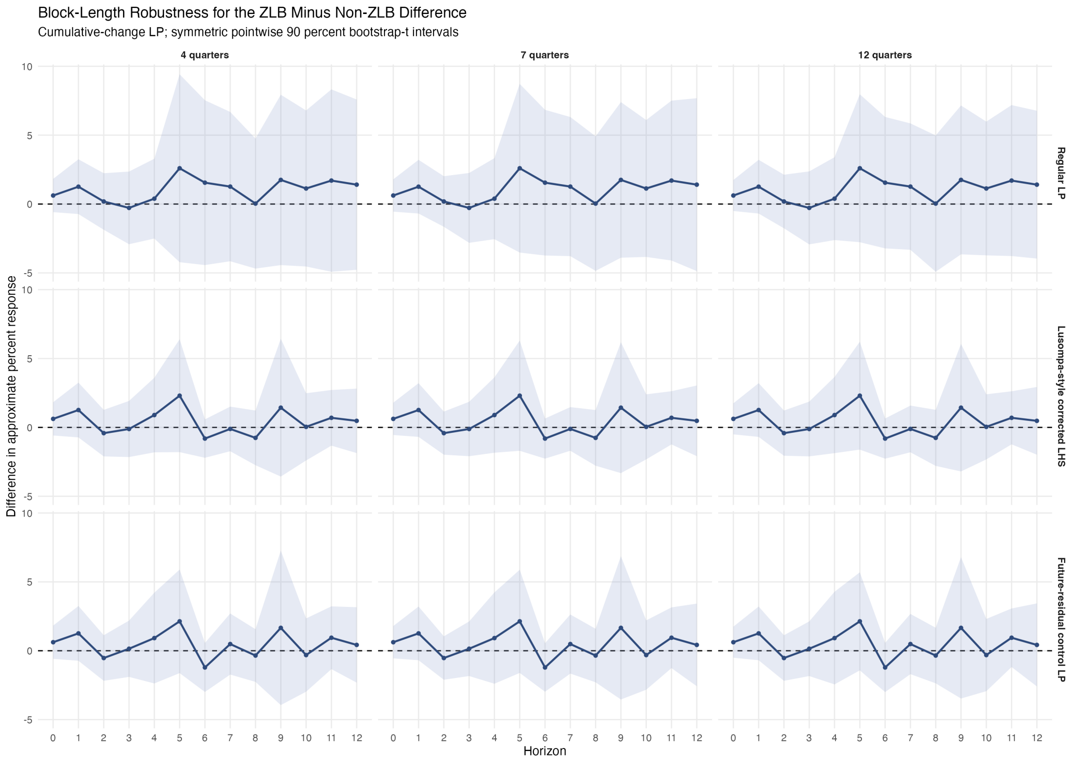

Figure 17: Symmetric pointwise bootstrap-t intervals for the ZLB minus non-ZLB response difference under alternative block lengths.

Figure 17 shows the same check for the ZLB minus non-ZLB difference. The regime difference remains sensitive to the block length and the intervals include zero at the selected horizons. This supports the main interpretation that the ZLB response is not shown to be statistically larger than the non-ZLB response.

A.6 Complete HAC and Bootstrap Interval Tables

The main text reports selected horizons to keep the discussion readable. Tables 9 to 14 report the full pointwise HAC and symmetric bootstrap-t intervals for every horizon, specification, estimator, and regime. These tables are included for completeness and reproducibility.

Table 9: Full Interval Table: Level LP, Regular LP

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.45	[-0.08, 0.98]	[-0.12, 1.02]	0.87	[0.13, 1.61]	[-0.13, 1.88]
1	0.72	[-0.14, 1.59]	[0.02, 1.42]	1.61	[0.49, 2.72]	[0.07, 3.14]
2	1.10	[0.33, 1.87]	[0.68, 1.51]	0.90	[-0.44, 2.24]	[-0.69, 2.48]
3	1.32	[0.21, 2.43]	[0.59, 2.04]	-0.01	[-1.71, 1.69]	[-1.76, 1.74]
4	1.04	[-0.17, 2.25]	[0.46, 1.61]	-0.38	[-2.16, 1.39]	[-1.95, 1.18]
5	0.66	[-0.71, 2.03]	[-0.22, 1.53]	1.22	[-0.49, 2.92]	[-0.88, 3.31]
6	1.18	[-0.04, 2.39]	[0.54, 1.82]	1.28	[0.10, 2.45]	[0.02, 2.54]
7	1.22	[0.06, 2.39]	[0.71, 1.74]	1.15	[-0.44, 2.73]	[-0.68, 2.97]
8	1.16	[-0.22, 2.55]	[0.59, 1.74]	-0.25	[-2.57, 2.07]	[-3.32, 2.82]
9	1.75	[0.33, 3.18]	[0.68, 2.83]	0.71	[-1.09, 2.51]	[-0.97, 2.39]
10	2.04	[0.42, 3.66]	[1.04, 3.03]	0.35	[-1.87, 2.57]	[-1.53, 2.23]
11	1.74	[0.14, 3.34]	[1.01, 2.47]	1.08	[-0.75, 2.91]	[-0.55, 2.71]
12	0.52	[-1.13, 2.17]	[-0.77, 1.81]	-0.50	[-2.41, 1.40]	[-3.40, 2.39]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

Table 10: Full Interval Table: Level LP, Lusompa-style corrected LHS

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.45	[-0.08, 0.98]	[-0.12, 1.02]	0.87	[0.13, 1.61]	[-0.13, 1.88]
1	0.72	[-0.14, 1.59]	[0.02, 1.42]	1.61	[0.49, 2.72]	[0.07, 3.14]
2	0.94	[0.28, 1.60]	[0.49, 1.39]	0.34	[-0.67, 1.35]	[-1.25, 1.92]
3	1.28	[0.12, 2.43]	[0.47, 2.08]	0.10	[-1.29, 1.49]	[-1.30, 1.50]
4	1.15	[0.25, 2.05]	[0.76, 1.54]	0.29	[-1.13, 1.70]	[-1.42, 1.99]
5	0.89	[0.16, 1.61]	[0.55, 1.23]	1.58	[0.01, 3.16]	[-0.71, 3.87]
6	1.82	[1.11, 2.54]	[1.14, 2.51]	0.19	[-0.53, 0.91]	[-0.88, 1.27]
7	1.55	[0.67, 2.43]	[0.99, 2.11]	0.70	[-0.84, 2.25]	[-1.07, 2.48]
8	1.40	[0.50, 2.30]	[1.04, 1.76]	-0.56	[-2.06, 0.93]	[-2.78, 1.66]
9	2.09	[1.03, 3.15]	[1.09, 3.09]	0.97	[-1.49, 3.43]	[-1.64, 3.57]
10	1.74	[0.32, 3.16]	[1.18, 2.30]	0.18	[-1.74, 2.09]	[-1.50, 1.85]
11	1.06	[0.16, 1.95]	[0.31, 1.80]	0.56	[-0.31, 1.43]	[-0.20, 1.32]
12	0.17	[-0.81, 1.16]	[-0.78, 1.13]	-0.94	[-2.57, 0.69]	[-3.85, 1.97]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

Table 11: Full Interval Table: Level LP, Future-residual control LP

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.45	[-0.08, 0.98]	[-0.12, 1.02]	0.87	[0.13, 1.61]	[-0.13, 1.88]
1	0.72	[-0.14, 1.59]	[0.02, 1.42]	1.61	[0.49, 2.72]	[0.07, 3.14]
2	0.96	[0.28, 1.63]	[0.51, 1.40]	0.40	[-0.69, 1.49]	[-1.22, 2.03]
3	1.28	[0.10, 2.47]	[0.47, 2.10]	0.13	[-1.27, 1.53]	[-1.25, 1.51]
4	1.17	[0.27, 2.07]	[0.78, 1.56]	0.23	[-1.14, 1.59]	[-1.36, 1.81]
5	0.89	[0.15, 1.63]	[0.54, 1.23]	1.47	[-0.11, 3.05]	[-0.72, 3.65]
6	1.86	[1.13, 2.59]	[1.13, 2.59]	0.03	[-0.89, 0.94]	[-1.41, 1.46]
7	1.42	[0.60, 2.24]	[0.96, 1.87]	0.96	[-0.76, 2.68]	[-0.84, 2.76]
8	1.32	[0.45, 2.20]	[0.98, 1.67]	-0.43	[-1.95, 1.09]	[-2.49, 1.62]
9	2.02	[1.02, 3.01]	[1.10, 2.93]	1.07	[-1.29, 3.43]	[-1.37, 3.51]
10	1.50	[0.26, 2.74]	[1.01, 1.98]	0.04	[-2.12, 2.20]	[-1.90, 1.98]
11	0.82	[0.04, 1.60]	[0.03, 1.61]	0.76	[-0.05, 1.58]	[0.03, 1.50]
12	0.10	[-0.91, 1.10]	[-0.94, 1.13]	-0.99	[-2.90, 0.93]	[-3.90, 1.93]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

Table 12: Full Interval Table: Cumulative change LP, Regular LP

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.51	[-0.06, 1.07]	[-0.11, 1.13]	1.13	[0.69, 1.56]	[0.32, 1.94]
1	0.52	[-0.17, 1.22]	[-0.28, 1.33]	1.79	[0.94, 2.63]	[0.28, 3.29]
2	1.22	[0.21, 2.22]	[0.16, 2.27]	1.40	[0.39, 2.42]	[0.05, 2.75]
3	1.48	[0.16, 2.80]	[0.00, 2.96]	1.20	[-0.30, 2.70]	[-0.73, 3.14]
4	1.29	[-0.32, 2.89]	[-0.64, 3.21]	1.68	[0.39, 2.97]	[-0.11, 3.47]
5	1.07	[-0.84, 2.98]	[-1.95, 4.09]	3.67	[2.20, 5.15]	[0.06, 7.28]
6	1.82	[-0.09, 3.73]	[-0.06, 3.70]	3.37	[1.80, 4.94]	[-0.37, 7.12]
7	1.98	[-0.04, 4.00]	[0.19, 3.77]	3.25	[1.49, 5.00]	[-0.52, 7.02]
8	2.14	[-0.13, 4.41]	[0.11, 4.17]	2.18	[-0.43, 4.79]	[-2.03, 6.38]
9	2.92	[0.51, 5.32]	[-1.45, 7.28]	4.67	[2.95, 6.38]	[0.18, 9.15]
10	3.31	[0.90, 5.73]	[-2.46, 9.09]	4.44	[3.09, 5.79]	[0.68, 8.21]
11	3.17	[0.61, 5.74]	[-2.12, 8.47]	4.88	[3.61, 6.14]	[1.04, 8.72]
12	2.06	[-0.65, 4.76]	[-1.00, 5.12]	3.46	[1.55, 5.38]	[-0.86, 7.78]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

Table 13: Full Interval Table: Cumulative change LP, Lusompa-style corrected LHS

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.51	[-0.06, 1.07]	[-0.11, 1.13]	1.13	[0.69, 1.56]	[0.32, 1.94]
1	0.52	[-0.17, 1.22]	[-0.28, 1.33]	1.79	[0.94, 2.63]	[0.28, 3.29]
2	1.42	[0.58, 2.25]	[0.34, 2.50]	1.00	[0.22, 1.78]	[-0.16, 2.17]
3	1.62	[0.55, 2.70]	[0.22, 3.02]	1.51	[0.60, 2.42]	[0.18, 2.83]
4	1.50	[0.61, 2.39]	[0.41, 2.59]	2.40	[1.15, 3.65]	[-0.02, 4.82]
5	1.43	[0.59, 2.28]	[0.30, 2.57]	3.74	[2.36, 5.11]	[0.19, 7.28]
6	2.46	[1.75, 3.17]	[1.32, 3.60]	1.65	[0.75, 2.56]	[0.39, 2.91]
7	2.18	[1.37, 3.00]	[1.19, 3.18]	2.07	[1.08, 3.07]	[0.66, 3.48]
8	2.19	[1.30, 3.09]	[1.15, 3.24]	1.43	[0.11, 2.75]	[-0.36, 3.23]
9	2.92	[1.92, 3.92]	[0.74, 5.10]	4.35	[1.87, 6.84]	[-1.92, 10.63]
10	2.59	[1.36, 3.82]	[0.45, 4.73]	2.63	[1.41, 3.85]	[0.42, 4.84]
11	2.00	[1.07, 2.94]	[1.01, 3.00]	2.70	[1.89, 3.50]	[1.27, 4.12]
12	1.07	[0.06, 2.07]	[-0.58, 2.71]	1.54	[0.22, 2.86]	[-0.35, 3.43]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

Table 14: Full Interval Table: Cumulative change LP, Future-residual control LP

Horizon	non-ZLB IRF	non-ZLB HAC CI	non-ZLB bootstrap CI	ZLB IRF	ZLB HAC CI	ZLB bootstrap CI
0	0.51	[-0.06, 1.07]	[-0.11, 1.13]	1.13	[0.69, 1.56]	[0.32, 1.94]
1	0.52	[-0.17, 1.22]	[-0.28, 1.33]	1.79	[0.94, 2.63]	[0.28, 3.29]
2	1.46	[0.65, 2.27]	[0.36, 2.55]	0.93	[0.20, 1.66]	[-0.22, 2.08]
3	1.59	[0.48, 2.70]	[0.09, 3.09]	1.73	[0.85, 2.62]	[0.35, 3.12]
4	1.52	[0.56, 2.48]	[0.37, 2.67]	2.44	[0.70, 4.18]	[-0.64, 5.52]
5	1.46	[0.65, 2.27]	[0.35, 2.57]	3.60	[2.20, 5.00]	[0.11, 7.09]
6	2.53	[1.82, 3.24]	[1.32, 3.73]	1.31	[0.16, 2.47]	[-0.21, 2.83]
7	1.99	[1.20, 2.77]	[1.14, 2.84]	2.47	[1.26, 3.69]	[0.53, 4.42]
8	2.10	[1.26, 2.93]	[1.17, 3.02]	1.75	[0.45, 3.05]	[-0.07, 3.57]
9	2.88	[1.88, 3.88]	[0.74, 5.03]	4.54	[1.90, 7.18]	[-2.23, 11.31]
10	2.37	[1.26, 3.48]	[0.77, 3.96]	2.05	[0.43, 3.67]	[-0.34, 4.45]
11	1.72	[0.96, 2.48]	[0.80, 2.64]	2.66	[1.55, 3.78]	[0.84, 4.49]
12	1.04	[0.05, 2.03]	[-0.59, 2.66]	1.46	[-0.19, 3.11]	[-0.79, 3.71]

Note: HAC CI refers to the pointwise 90 percent HAC/Newey-West interval. Bootstrap CI refers to the pointwise 90 percent symmetric bootstrap-t interval.

A.7 Simultaneous Bootstrap Bands

The main text reports symmetric pointwise bootstrap intervals, meaning that uncertainty is assessed horizon by horizon. A stronger check is to construct simultaneous sup-t bands, which ask whether the complete response path from $h = 0$ to $h = 12$ is covered jointly.

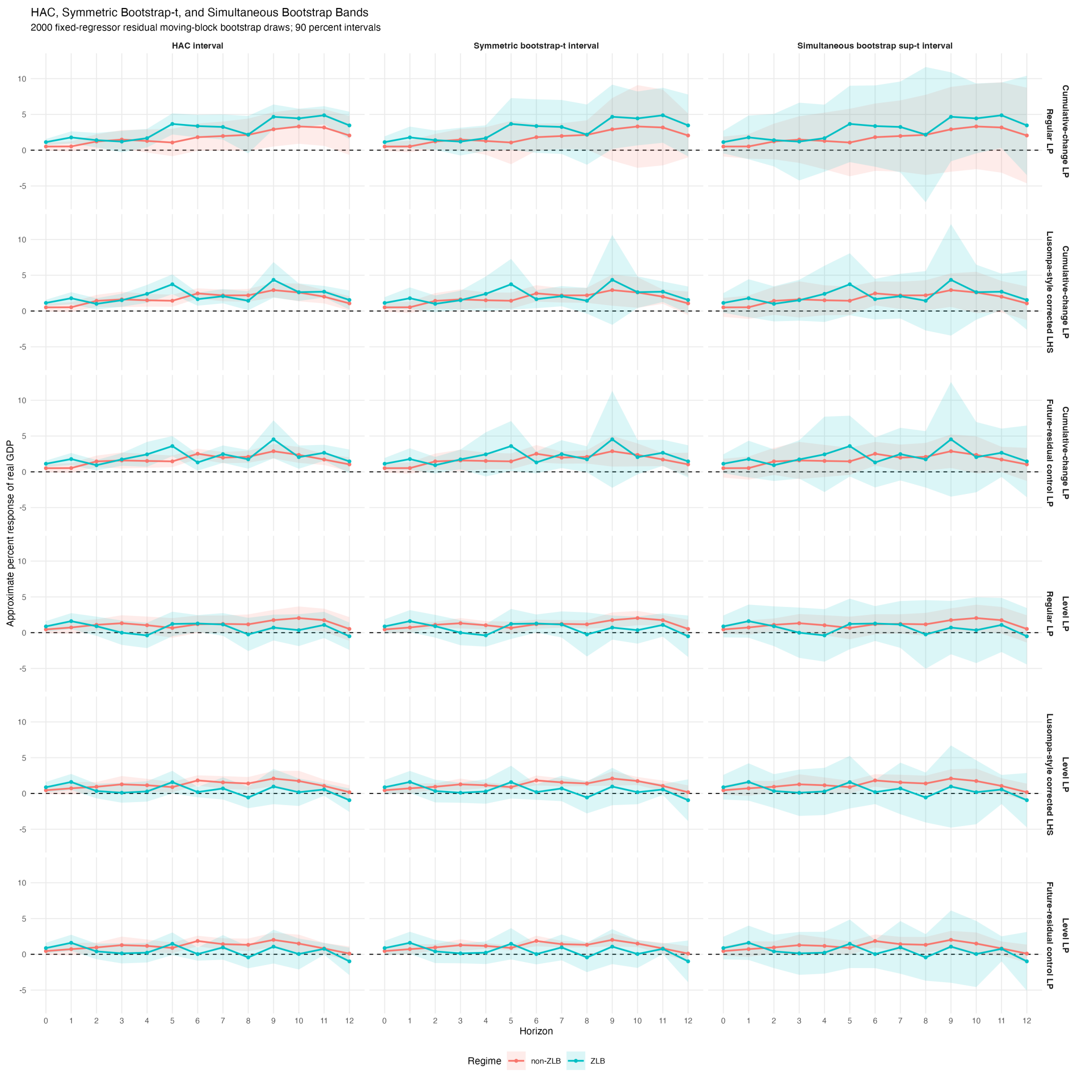

Figure 18: Comparison between symmetric pointwise bootstrap intervals and simultaneous sup-t bootstrap bands.

Figure 18 compares the symmetric pointwise bootstrap intervals with simultaneous sup-t bands. The simultaneous sup-t bands are much wider than the pointwise bootstrap intervals. In most specifications the simultaneous bands include zero over the full response horizon, even when some pointwise intervals exclude zero. This means that the response can be statistically different from zero at selected horizons, but the complete impulse response path is not shown to be jointly different from zero at the 10 percent level.

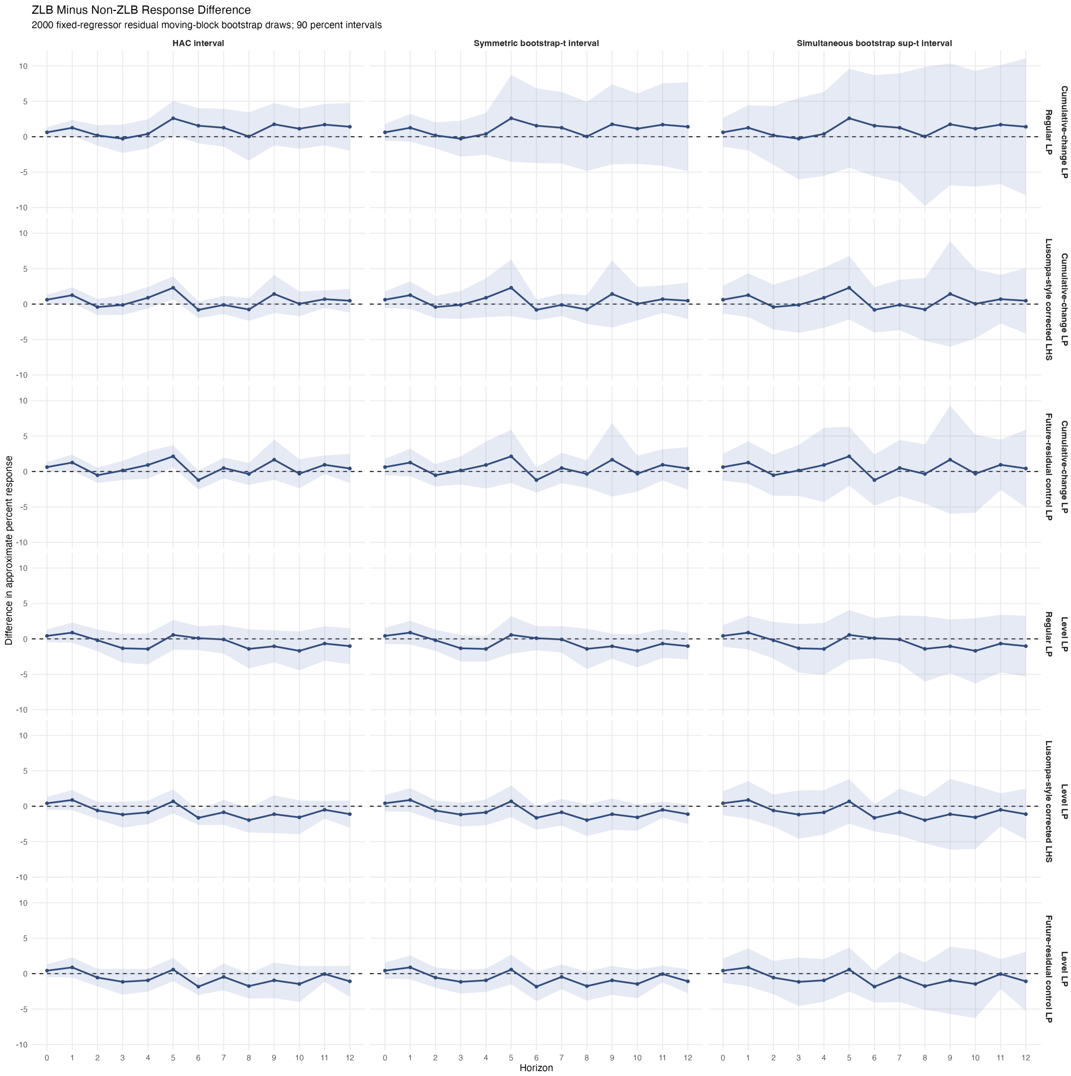

Figure 19: Bootstrap inference for the ZLB minus non-ZLB response difference. The figure compares HAC intervals, symmetric pointwise bootstrap intervals, and simultaneous sup-t bands.

The ZLB minus non-ZLB difference at horizon h is defined as:

$$\hat{d}_h = \hat{\beta}_h^Z - \hat{\beta}_h^N. \tag{44}$$

Since both coefficients are estimated inside the same joint regression, the standard error of the difference uses the covariance between the two coefficients:

$$\widehat{SE}(\hat{d}_h) = \sqrt{\widehat{Var}(\hat{\beta}_h^Z) + \widehat{Var}(\hat{\beta}_h^N) - 2\widehat{Cov}(\hat{\beta}_h^Z, \hat{\beta}_h^N)}. \tag{45}$$

This means that the confidence interval for the difference is not obtained by simply subtracting the separate ZLB and non-ZLB confidence intervals.

The same conclusion applies to the ZLB minus non-ZLB difference, namely that the simultaneous bootstrap bands for the regime difference include zero across the response path. This supports the interpretation that the ZLB response should not be treated as statistically larger than the non-ZLB response.

References

Akaike, Hirotugu (1974). “A New Look at the Statistical Model Identification”. In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

Andrews, Donald W. K. (1991). “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation”. In: *Econometrica* 59.3, pp. 817–858. DOI: 10.2307/2938229.

Breitung, Jörg and Ralf Brüggemann (2023). “Projection Estimators for Structural Impulse Responses”. In: *Oxford Bulletin of Economics and Statistics* 85.6, pp. 1320–1340. DOI: 10.1111/obes.12562.

Brüggemann, Ralf, Carsten Jentsch, and Carsten Trenkler (2016). “Inference in VARs with Conditional Heteroskedasticity of Unknown Form”. In: *Journal of Econometrics* 191.1, pp. 69–85. DOI: 10.1016/j.jeconom.2015.10.004.

Cloyne, James (2013). “Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom”. In: *American Economic Review* 103.4, pp. 1507–1528. DOI: 10.1257/aer.103.4.1507.

Cloyne, James, Nicholas Dimsdale, and Patrick Hürtgen (2025). “Are Tax Cuts Contractionary at the Zero Lower Bound? Evidence from a Century of Data”. In: *Journal of Political Economy* 133.2, pp. 568–603. DOI: 10.1086/732890.

Correia, Isabel, Emmanuel Farhi, Juan Pablo Nicolini, and Pedro Teles (2013). “Unconventional Fiscal Policy at the Zero Bound”. In: *American Economic Review* 103.4, pp. 1172–1211. DOI: 10.1257/aer.103.4.1172.

Eggertsson, Gauti B. (2011). “What Fiscal Policy Is Effective at Zero Interest Rates?” In: *NBER Macroeconomics Annual* 25.1, pp. 59–112. DOI: 10.1086/657529.

Eggertsson, Gauti B. and Michael Woodford (2003). “Zero Bound on Interest Rates and Optimal Monetary Policy”. In: *Brookings Papers on Economic Activity* 2003.1, pp. 139–233. DOI: 10.1353/eca.2003.0010.

Gonçalves, Sílvia and Lutz Kilian (2004). “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form”. In: *Journal of Econometrics* 123.1, pp. 89–120. DOI: 10.1016/j.jeconom.2003.10.030.

Hall, Peter (1992). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. New York: Springer. DOI: 10.1007/978-1-4612-4384-7.

Herbst, Edward P. and Benjamin K. Johannsen (2021). *Bias in Local Projections*. Finance and Economics Discussion Series 2020-010r1. Board of Governors of the Federal Reserve System. DOI: 10.17016/FEDS.2020.010r1.

Jordà, Òscar (2005). “Estimation and Inference of Impulse Responses by Local Projections”. In: *American Economic Review* 95.1, pp. 161–182. DOI: 10.1257/0002828053828518.

Jordà, Òscar and Alan M. Taylor (2016). “The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy”. In: *The Economic Journal* 126.590, pp. 219–255. DOI: 10.1111/ecoj.12332.

Jordà, Òscar and Alan M. Taylor (2024). *Local Projections*. Working Paper 2024-24. Federal Reserve Bank of San Francisco. DOI: 10.24148/wp2024-24.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf (2024). "Local Projections vs. VARs: Lessons from Thousands of DGPs". In: *Journal of Econometrics* 244, p. 105722. DOI: 10.1016/j.jeconom.2024.105722.

Lusompa, Amaze (2023). "Local Projections, Autocorrelation, and Efficiency". In: *Quantitative Economics* 14.4, pp. 1199–1220. DOI: 10.3982/QE1988.

Mertens, Karel and Morten O. Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States". In: *American Economic Review* 103.4, pp. 1212–1247. DOI: 10.1257/aer.103.4.1212.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021). "Local Projection Inference is Simpler and More Robust Than You Think". In: *Econometrica* 89.4, pp. 1789–1823. DOI: 10.3982/ECTA18756.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf (2025). "Local Projections or VARs? A Primer for Macroeconomists". Prepared for the NBER Macroeconomics Annual 2025.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. DOI: 10.2307/1913610.

Piger, Jeremy and Thomas Stockwell (2025). "Differences From Differencing: Should Local Projections With Observed Shocks Be Estimated in Levels or Differences?" In: *Journal of Applied Econometrics* 40.7, pp. 759–787. DOI: 10.1002/jae.70003.

Plagborg-Møller, Mikkel and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impulse Responses". In: *Econometrica* 89.2, pp. 955–980. DOI: 10.3982/ECTA17813.

Ramey, Valerie A. and Sarah Zubairy (2018). "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data". In: *Journal of Political Economy* 126.2, pp. 850–901. DOI: 10.1086/696277.

Romer, Christina D. and David H. Romer (2010). "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks". In: *American Economic Review* 100.3, pp. 763–801. DOI: 10.1257/aer.100.3.763.

Schwarz, Gideon (1978). "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2, pp. 461–464. DOI: 10.1214/aos/1176344136.

Stock, James H. and Mark W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments". In: *The Economic Journal* 128.610, pp. 917–948. DOI: 10.1111/ecoj.12593.