

Specification Choice in Local Projections: Evidence from Monetary Policy Shocks

Eric Fortier*

April 24, 2026

Abstract

This study examines how specification choices in local projections affect the estimation of impulse responses to monetary policy shocks. Using monthly U.S. data from 1983 to 2007 and the shock series of [Aruoba and Drechsel \(2024\)](#), I systematically compare levels and long-differences specifications across 12 control sets and multiple lag lengths. I evaluate the results both qualitatively—by benchmarking impulse responses against theory, standard beliefs, and prior evidence—and quantitatively, using information criteria (AIC, BIC, and CV). The findings show that the long-differences specification produces distorted long-run dynamics in its point estimates. By contrast, the levels specification, when paired with a robust control set, generates impulse responses that are consistent with expectations. I show that the distortion in the long-differences specification is driven by omitted variable bias induced by transforming the control variables into first-differences.

Keywords: Impulse response function, Monetary policy shocks, Local projections

JEL Codes: E32, E43, E52, E58

*Simon Fraser University PhD candidate (email: eric_fortier@sfu.ca). I thank Professors Bertille Antoine, Chenyu Hou, Luba Petersen, and Eric Swanson for their helpful feedback and suggestions on earlier drafts of this paper. Any remaining errors are my own.

1 Introduction

Jordà's (2005) local projection has become a standard tool for estimating impulse responses to monetary policy shocks. The approach involves estimating a sequence of regressions of an outcome variable at horizons h after time t on a shock and a set of controls. These controls—typically lags of macroeconomic variables—are meant to help the exogeneity of the measured shock and to improve precision by reducing the variance of the estimator (see Stock and Watson 2018).

In practice, applied researchers differ in both the choice of controls and the number of lags. For example, Ramey (2016) includes two months of lags for six key macroeconomic variables, whereas Romer and Romer (2023) include twelve months of lags of the outcome variable and the shock. Moreover, local projections can be estimated either in levels or in long-differences, and the choice between them can produce significantly different results (Piger and Stockwell, 2025; Jordà and Taylor, 2025).

This paper asks: *What is the most appropriate specification for local projections in the context of monetary policy shocks?* The goal is to identify choices over controls, lag length, and functional form (levels versus long-differences) that produce impulse responses that are most credible given theory, standard beliefs, and prior empirical evidence.

Reviews such as Ramey (2016), Jordà and Taylor (2025), and Inoue et al. (2026) typically recommend including lags of the outcome variable, lags of the shock, and lags of other key macroeconomic variables—sometimes along with a linear time trend if the data has trending behaviour. Additionally, Olea et al. (2025) and Olea et al. (2026) emphasize the importance of including lagged controls that strongly predict the outcome variable and/or the shock.

Most studies estimate local projections in levels, regressing the outcome variable at horizon h on the shock and a set of controls:

$$y_{t+h} = \alpha_h + \beta_h s_t + \gamma_h' X_t + u_{t+h}, \quad (1)$$

where y_{t+h} denotes the outcome variable, s_t the shock, and X_t a vector of controls. An alternative is to estimate local projections in long-differences, which instead regress the cumulative change in the outcome variable at horizon h on the shock and a set of controls:

$$y_{t+h} - y_{t-1} = \delta_h + \beta_h s_t + \theta'_h X_t^D + v_{t+h}, \quad (2)$$

where X_t^D is a vector of controls in first-differences. [Piger and Stockwell \(2025\)](#) show that estimating local projections in long-differences can reduce bias and deliver more accurate confidence intervals in small samples. For this reason, [Jordà and Taylor \(2025\)](#) prefer this specification. Not all researchers agree, however. [Olea et al. \(2025\)](#) argue that first-differencing the data in the long-differences specification can weaken the predictive power of lagged controls, potentially causing substantial efficiency losses, and note that further research on this issue would be welcome. This paper contributes to that effort.

To this end, I systematically compare a wide range of control sets and lag lengths across both levels and long-differences specifications. Using monthly U.S. data from 1983 to 2007, I employ the monetary policy shock identified by [Aruoba and Drechsel \(2024\)](#) and evaluate the response of key macroeconomic variables: real GDP, the PCE price index, and the unemployment rate. I compare 12 different control sets for each key macroeconomic variable, each estimated with 3, 6, 9, and 12 lags. The control sets are designed to reflect the most commonly used specifications in the literature.

To interpret and discriminate among empirical results, I explicitly lay out the expected shape and timing of each outcome variable based on theory, standard beliefs, and prior empirical evidence. I also supplement this evaluation with a quantitative approach using three information criteria: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the time-series Cross-Validation Criterion (CV). For each criterion, I compute the average score across the entire response.

The results show that the long-differences specification distorts long-run dynamics in

its point estimates. Output and unemployment display large and persistent effects, while the price level responds with a longer-than-expected lag. Under the levels specification, by contrast, output and unemployment exhibit hump-shaped responses, and all variables display lags consistent with expectations.¹ I show that this divergence is driven by omitted variable bias induced by transforming the control variables into first-differences. When the long-differences specification is estimated with controls in levels, the resulting impulse responses are virtually identical to those from the levels specification. Further, I show that first-differencing the controls weakens their predictive power for the outcome at longer horizons, while keeping the controls in levels does not.

For the levels specifications, a robust control set includes at least six lags of the federal funds rate, GDP, the unemployment rate, and the PCE price index, along with a linear time trend. With this control set, output and unemployment display classic hump-shaped responses and are both statistically significant at the 10 percent level. The response of the PCE price index shows no price puzzle and features a long-term and statistically significant decline. These well-behaved responses, consistent with theory and standard beliefs, were difficult to obtain in samples covering the Great Moderation (see [Barakchian and Crowe 2013](#) and [Ramey 2016](#)), but here they emerge clearly once the appropriate control set is used.

The key control variable in the levels specification is the federal funds rate. Failing to include it and relying only on lags of the shock can produce responses that deviate from expectations. I show that once the lags of the federal funds rate are included, there is no need to control for the lags of the shock; responses that include both sets of lags are virtually identical to those that include only the federal funds rate. This is useful for applied macroeconomists seeking to preserve additional degrees of freedom.

The remainder of this paper is organized as follows. Section 2 presents the criteria

¹The results use the monetary policy shocks from [Aruoba and Drechsel \(2024\)](#). When using [Romer and Romer \(2004\)](#) shocks, long-differences produces even more distorted responses relative to levels; see Appendix 7.3.

to evaluate and interpret the empirical results. In section 3, the estimation strategy is presented. Section 4 presents the data. In section 5, the results are presented and discussed. Section 6 examines the role of first-differencing the controls. Section 7 concludes with methodological recommendations.

2 Methodology

I follow the research strategy of *weak apriorism* (Hoover, 1994). This approach uses theory (or prevailing beliefs) to interpret and discriminate among empirical results. It is *weak* because an empirical result that does not conform to theory does not necessarily invalidate it. One can still make observations that require interpretation and may contribute to the development of new theories. To supplement this approach, I also employ information criteria.

2.1 Expected Shape

Theory: From a theoretical standpoint, the New-Keynesian model predicts that a contractionary monetary policy shock will cause an immediate (though temporary) decrease in both output and inflation. The mechanism operates through sticky prices: because prices cannot adjust immediately, an unexpected increase in the nominal interest rate increases the real interest rate, which decreases output (and by implication, employment). The decrease in output also means inflation falls. Lastly, because the price level cumulates inflation, a temporary disinflation pushes the price level onto a lower path relative to its pre-shock trend. Classic New-Keynesian analyses (e.g., Woodford 2003; Galí 2015) all produce this same qualitative result.

Standard Beliefs: Many economists and policymakers argue that output and inflation do not respond to a monetary policy shock immediately, as in the basic New-Keynesian

model, but instead adjust gradually (this has been recently highlighted by [Cochrane \(2024\)](#), [Cochrane \(2025\)](#)). This view is rooted in Milton Friedman’s dictum that monetary policy “has long and variable lags,” meaning that monetary policy takes time to affect the economy, and that this timing can vary. [Dupor \(2023\)](#) identifies two broad categories of frictions that can generate such lagged effects. First, contract lengths: many prices and quantities are set in advance (e.g., long-term contracts for wages or sales), so a change in the interest rate cannot immediately alter those preset arrangements. Second, inattention: firms and households may adjust prices or consumption infrequently.²

Empirical Evidence: A large body of empirical evidence supports the existence of long and variable lags in monetary policy. Comprehensive reviews by [Christiano et al. \(1999\)](#) and [Ramey \(2016\)](#), and more recent work by [Romer and Romer \(2023\)](#) document that the impulse responses of key macroeconomic variables often exhibit a hump-like shape: output declines modestly at first, falls more deeply after about a year, and then gradually returns to baseline. The price level tends to respond more sluggishly—in some studies it barely moves or even rises slightly in the short run (the “price puzzle”), before eventually falling with a lag.

Given the theoretical predictions of the New-Keynesian model, the standard beliefs among practitioners, and the broad empirical evidence, the core macroeconomic variables are expected to evolve as follows in response to a contractionary monetary policy shock:

1. **Output:** declines in a hump-shaped manner with a short lag, then gradually returns to its baseline level.
2. **Price Level:** decreases sluggishly and eventually stabilizes at a lower level relative to the pre-shock trend.

²To match these gradual responses formally, medium-scale New-Keynesian models incorporate additional sources of persistence such as price indexation and habit formation (e.g., [Christiano et al. 2005](#); [Smets and Wouters 2007](#)).

3. **Unemployment:** rises in a hump-shaped manner with a short lag, then gradually returns to its baseline rate.

2.2 Expected Lag

Milton Friedman found that the lag between monetary policy and its economic effect ranged between 29 months and four years. This is a broad range of time. In order to be more specific, I document the timing of the first significant effect for each core macroeconomic variable from a large number of empirical papers (based on the significance level chosen in each paper) (see Tables 1, 2, and 3). I use the review of the empirical results by [Ramey \(2016\)](#) and the main results from more recent papers ([Jarociński and Karadi, 2020](#); [Miranda-Agrippino and Ricco, 2021](#); [Bu et al., 2021](#); [Aruoba and Drechsel, 2024](#); [Bauer and Swanson, 2023](#); [Romer and Romer, 2023](#)). The lag varies across estimator, identification type, and sample period. Thus, I calculate the first and third quartile of this variable to provide an acceptable range of time. The expected lag for output, the price level, and unemployment are summarized as follows.

1. **Output:** 1 to 10 months after the shock
2. **Price Level:** 0 to 27 months after the shock
3. **Unemployment:** 5 to 12 months after the shock

Table 1: *Summary of output responses to monetary policy shocks*

Paper	Estimator	Identification Type	Sample	Lag (month)
Ramey (2016) replications:				
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1965m1-1995m6	19
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1983m1-2007m12	-
<i>Christiano et al. (1999)</i>	SVAR, omits money and reserves	Recursive VAR	1983m1-2007m12	19
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-1996m12	5
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1983m3-2007m12	-
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-2007m12	5
<i>Romer and Romer (2004)</i>	LP with recursiveness assumption	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	LP	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	LP factor augmented	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-1996m12	14
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-2007m12	14
<i>Gertler and Karadi (2015)</i>	Proxy SVAR	HF surprise	1990m1-2012m6	8
<i>Gertler and Karadi (2015)</i>	LP with recursiveness assumption	HF surprise	1990m1-2012m6	-
Jarocinski and Karadi (2020)	BVAR	Sign-restricted HF surprise	1984m2-2016m12	-
Miranda-Agrippino and Ricco (2021)	LP	Informationally robust	1979m1-2014m12	0
Miranda-Agrippino and Ricco (2021)	BVAR	Informationally robust	1979m1-2014m12	0
Miranda-Agrippino and Ricco (2021)	BLP	Informationally robust	1979m1-2014m12	0
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	1994m1-2017m12	0
Bu et al (2021)	LP	Fama and MacBeth (1973)	1994m1-2017m12	4
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	2008m1-2017m12	12
Bu et al (2021)	LP	Fama and MacBeth (1973)	2008m1-2017m12	7
Aruoba and Drechsel (2024)	BVAR	Greenbook	1984m2-2016m12	9
Aruoba and Drechsel (2024)	BVAR	Sign-restricted Greenbook	1984m2-2016m12	4
Bauer and Swanson (2023)	LP	HF surprise	1988m1-2020m2	3
Bauer and Swanson (2023)	Proxy SVAR	HF surprise	1973m1-2020m2	0
Romer and Romer (2023)	LP	Narrative	1947m12-2016m12	2

Notes: This table summarizes output responses across key studies estimating the effects of monetary policy shocks. “Lag (month)” indicates how many months before the response shows a significant negative effect.

Table 2: Summary of price level responses to monetary policy shocks

Paper	Estimator	Identification Type	Sample	Lag (month)
Ramey (2016) replications:				
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1965m1-1995m6	19
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1983m1-2007m12	-
<i>Christiano et al. (1999)</i>	SVAR, omits money and reserves	Recursive VAR	1983m1-2007m12	19
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-1996m12	27
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1983m3-2007m12	-
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-2007m12	27
<i>Romer and Romer (2004)</i>	LP with recursiveness assumption	Greenbook	1969m3-1996m12	38
<i>Romer and Romer (2004)</i>	LP	Greenbook	1969m3-1996m12	38
<i>Romer and Romer (2004)</i>	LP factor augmented	Greenbook	1969m3-1996m12	38
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-1996m12	0
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-2007m12	0
<i>Gertler and Karadi (2015)</i>	Proxy SVAR	HF surprise	1990m1-2012m6	26
<i>Gertler and Karadi (2015)</i>	LP with recursiveness assumption	HF surprise	1990m1-2012m6	42
∞				
Jarocinski and Karadi (2020)	BVAR	Sign-restricted HF surprise	1984m2-2016m12	0
Miranda-Agrippino and Ricco (2021)	LP	Informationally robust	1979m1-2014m12	0
Miranda-Agrippino and Ricco (2021)	BVAR	Informationally robust	1979m1-2014m12	0
Miranda-Agrippino and Ricco (2021)	BLP	Informationally robust	1979m1-2014m12	0
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	1994m1-2017m12	8
Bu et al (2021)	LP	Fama and MacBeth (1973)	1994m1-2017m12	7
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	2008m1-2017m12	6
Bu et al (2021)	LP	Fama and MacBeth (1973)	2008m1-2017m12	4
Aruoba and Drechsel (2024)	BVAR	Greenbook	1984m2-2016m12	-
Aruoba and Drechsel (2024)	BVAR	Sign-restricted Greenbook	1984m2-2016m12	-
Bauer and Swanson (2023)	LP	HF surprise	1988m1-2020m2	1
Bauer and Swanson (2023)	Proxy SVAR	HF surprise	1973m1-2020m2	0
Romer and Romer (2023)	LP	Narrative	1947m12-2016m12	10

Notes: This table summarizes the price level responses across key studies estimating the effects of monetary policy shocks. “Lag (month)” indicates how many months before the response shows a significant negative effect.

Table 3: *Summary of unemployment responses to monetary policy shocks*

Paper	Estimator	Identification Type	Sample	Lag (month)
Ramey (2016) replications::				
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1965m1-1995m6	11
<i>Christiano et al. (1999)</i>	SVAR	Recursive VAR	1983m1-2007m12	-
<i>Christiano et al. (1999)</i>	SVAR, omits money and reserves	Recursive VAR	1983m1-2007m12	11
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-1996m12	5
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1983m3-2007m12	5
<i>Romer and Romer (2004)</i>	Coibion VAR	Greenbook	1969m3-2007m12	-
<i>Romer and Romer (2004)</i>	LP with recursiveness assumption	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	LP	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	LP factor augmented	Greenbook	1969m3-1996m12	10
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-1996m12	14
<i>Romer and Romer (2004)</i>	Proxy SVAR	Greenbook	1969m3-2007m12	14
<i>Gertler and Karadi (2015)</i>	Proxy SVAR	HF surprise	1990m1-2012m6	-
<i>Gertler and Karadi (2015)</i>	LP with recursiveness assumption	HF surprise	1990m1-2012m6	-
Jarocinski and Karadi (2020)	BVAR	Sign-restricted HF surprise	1984m2-2016m12	-
Miranda-Agrippino and Ricco (2021)	LP	Informationally robust	1979m1-2014m12	2
Miranda-Agrippino and Ricco (2021)	BVAR	Informationally robust	1979m1-2014m12	2
Miranda-Agrippino and Ricco (2021)	BLP	Informationally robust	1979m1-2014m12	2
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	1994m1-2017m12	-
Bu et al (2021)	LP	Fama and MacBeth (1973)	1994m1-2017m12	-
Bu et al (2021)	SVAR	Fama and MacBeth (1973)	2008m1-2017m12	-
Bu et al (2021)	LP	Fama and MacBeth (1973)	2008m1-2017m12	-
Aruoba and Drechsel (2024)	BVAR	Greenbook	1984m2-2016m12	13
Aruoba and Drechsel (2024)	BVAR	Sign-restricted Greenbook	1984m2-2016m12	7
Bauer and Swanson (2023)	LP	HF surprise	1988m1-2020m2	-
Bauer and Swanson (2023)	Proxy SVAR	HF surprise	1973m1-2020m2	5
Romer and Romer (2023)	LP	Narrative	1947m10-2016m12	17

Notes: This table summarizes unemployment responses across key studies estimating the effects of monetary policy shocks. “Lag (month)” indicates how many months before the response shows a significant positive effect.

2.3 Information Criteria

Information criteria can help identify the preferred control set and lag length in an econometric model. I use three: AIC, BIC, and CV. For each criterion, I report its average score across the entire impulse response.

The AIC is given by

$$-2\ln(\hat{L}) + 2k, \quad (3)$$

where $\hat{L}$ is the maximized value of the likelihood function and k is the number of estimated parameters. The better the model explains the data, the smaller the first term becomes, reflecting a higher maximized likelihood. The penalty for an additional parameter is constant and equal to 2.

The BIC is given by

$$-2\ln(\hat{L}) + k\ln(n). \quad (4)$$

The BIC penalty term increases with larger sample sizes n . Since $\ln(n) > 2$ when $n \geq 8$, BIC penalizes additional parameters more heavily than AIC.

I use a time-series CV approach. The model is re-estimated using an expanding window. At each step t , the model is estimated on observations $\{1, \dots, t\}$ and predicts the outcome variable at time $t + 1$. A prediction error $\tilde{u}_{t+1}$ is recorded as the difference between the predicted and actual values of the outcome variable. The time-series CV score is the average of the squared prediction errors:

$$\frac{1}{T - t_0} \sum_{t=t_0}^{T-1} \tilde{u}_{t+1}^2, \quad (5)$$

where $t_0 = \lfloor T/2 \rfloor$ is the initial training window. A lower time-series CV score indicates better out-of-sample forecast accuracy.

3 Estimation Strategy

The impulse responses are estimated using Jordà's local projection (Jordà, 2005). In the empirical literature, there are two common forms of local projections: levels and long-differences. The levels specification is estimated as follows:

$$y_{t+h} = \tau_h t + \alpha_h + \beta_h s_t + \gamma_h' X_t + u_{t+h}. \quad (6)$$

The estimated coefficient β_h is the impulse response of y_{t+h} to a monetary policy shock s_t . X_t is a vector of controls, which are lagged values of macroeconomic variables, and t is a linear time trend. The levels specification is the most commonly used specification in the literature.

To derive the long-differences specification, consider first a local projection where the outcome variable is in first-differences:

$$y_{t+h} - y_{t+h-1} = \tilde{\alpha}_h + \tilde{\beta}_h s_t + \tilde{\gamma}_h' \tilde{X}_t + \tilde{u}_{t+h}. \quad (7)$$

The estimated coefficient $\tilde{\beta}_h$ is the impulse response of $y_{t+h} - y_{t+h-1}$ to a monetary policy shock s_t . Following Piger and Stockwell (2025), the control variables are also first-differenced, denoted as $\tilde{X}_t$.³ Further, since the outcome variable and controls are stationary, then there is no longer a need to control for a linear time trend. You can think of (6) as estimating the impulse response of the log price level, while (7) estimates the impulse response of the inflation rate.

By summing $\tilde{\beta}_i$ to a particular horizon h , its cumulative effect can be recovered:

$$\sum_{i=0}^h \tilde{\beta}_i = \beta_h. \quad (8)$$

³The exception is lags of the shock. When included as controls, they enter in levels rather than first-differences. This keeps their treatment consistent with the contemporaneous shock on the right-hand side of (7), which also enters in levels rather than first-differences.

Thus, by taking the sum of equation (7), β_h can be estimated alternatively by:

$$y_{t+h} - y_{t-1} = \delta_h + \beta_h s_t + \theta'_h X_t^D + v_{t+h}, \quad (9)$$

which is referred to as the long-differenced specification. The coefficient of interest β_h is the same as (6), but now the outcome variable is in cumulative differences, and the lagged controls are first-differenced. The long-differences specification has been shown to reduce bias and produce more accurate confidence intervals in small samples (Piger and Stockwell, 2025) and is the preferred specification by Jordà and Taylor (2025).

For both levels and long-differences specifications, I account for serial correlation in the error term, and compute Newey-West standard errors (Newey and West, 1987) setting the lag equal to h for each horizon.

4 Data

I use the monetary policy shock series from Aruoba and Drechsel (2024), which builds on the classic Romer and Romer (2004) study. Romer and Romer (2004) regress the change in the target federal funds rate on the Fed’s forecasts of inflation, output, and unemployment. These forecasts are taken from the Fed’s Greenbook and are meant to approximate the full information set available to the Federal Reserve at the time of its policy decision. The residual from this regression is interpreted as a monetary policy shock. Aruoba and Drechsel (2024) argue that the Greenbook also contains a lot of information that is captured by text. Their idea is to place the entire Greenbook on the right-hand side of a Romer-Romer-style regression. To do this, they used natural language processing and machine learning techniques. Natural language processing translates text into sentiment scores, and machine learning allows for a large number of independent variables in the regression.

The data are at a monthly frequency from October 1983 to December 2007. This sample period avoids pollution from the Volcker disinflation period and the Great Recession. The

core macroeconomic variables of interest are log real GDP (interpolated by [Barrett \(2025\)](#)), log PCE price index, and the unemployment rate.

5 Results

In this section, I present the impulse responses of each core macroeconomic variable to a monetary policy shock while varying the control set, lag length, and functional form.

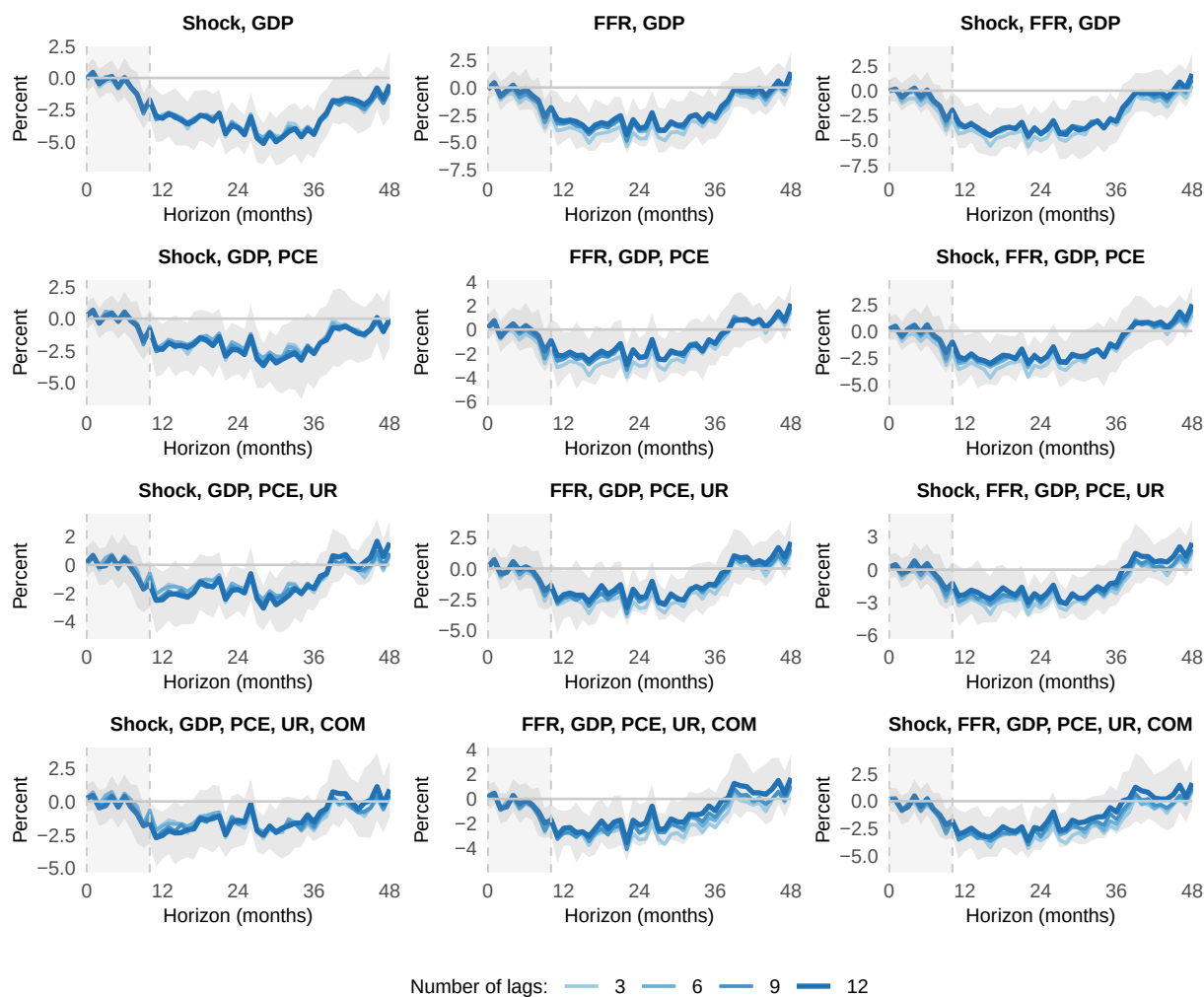
5.1 Output Responses

Figure 1 shows the impulse response of GDP to a monetary policy shock in levels. I compare a range of control sets, including several commonly used in the literature. The first control set includes only lags of the monetary policy shock and the outcome variable, reflecting the approach used by [Romer and Romer \(2023\)](#). The second control set includes lags of the outcome variable and the federal funds rate. The third control set includes lags of the outcome variable, the shock, and the federal funds rate. This pattern is repeated across the remaining rows and enables a direct comparison of the importance of controlling for lags of the monetary policy shock versus lags of the federal funds rate. The fourth, fifth, and sixth control sets add lags of log GDP and of log PCE price index. This reflects the control sets used in [Jordà and Taylor \(2025\)](#) and [Inoue et al. \(2026\)](#). The seventh, eighth, and ninth control sets add lags of the unemployment rate. The 10th, 11th, and 12th control sets add lags of log commodity price index. These last broader control sets are intended to replicate the controls used in [Ramey \(2016\)](#).

The overall pattern of the GDP response is broadly consistent with expectations. Following a lag of roughly eight months, GDP gradually declines, reaching its peak about 20 months after the shock. The peak decline is around -3% and is statistically significant at the 10 percent level. The response then gradually returns toward baseline, displaying a hump-like shape. The estimated responses are very similar across control sets and lag

lengths. Information-criterion results (Table 4) generally favour broader control sets that exclude the shock: AIC and CV prefer six or nine lags, while BIC prefers three lags.

Figure 1: *Response of GDP from an Aruoba-Drechsel monetary policy shock in levels*

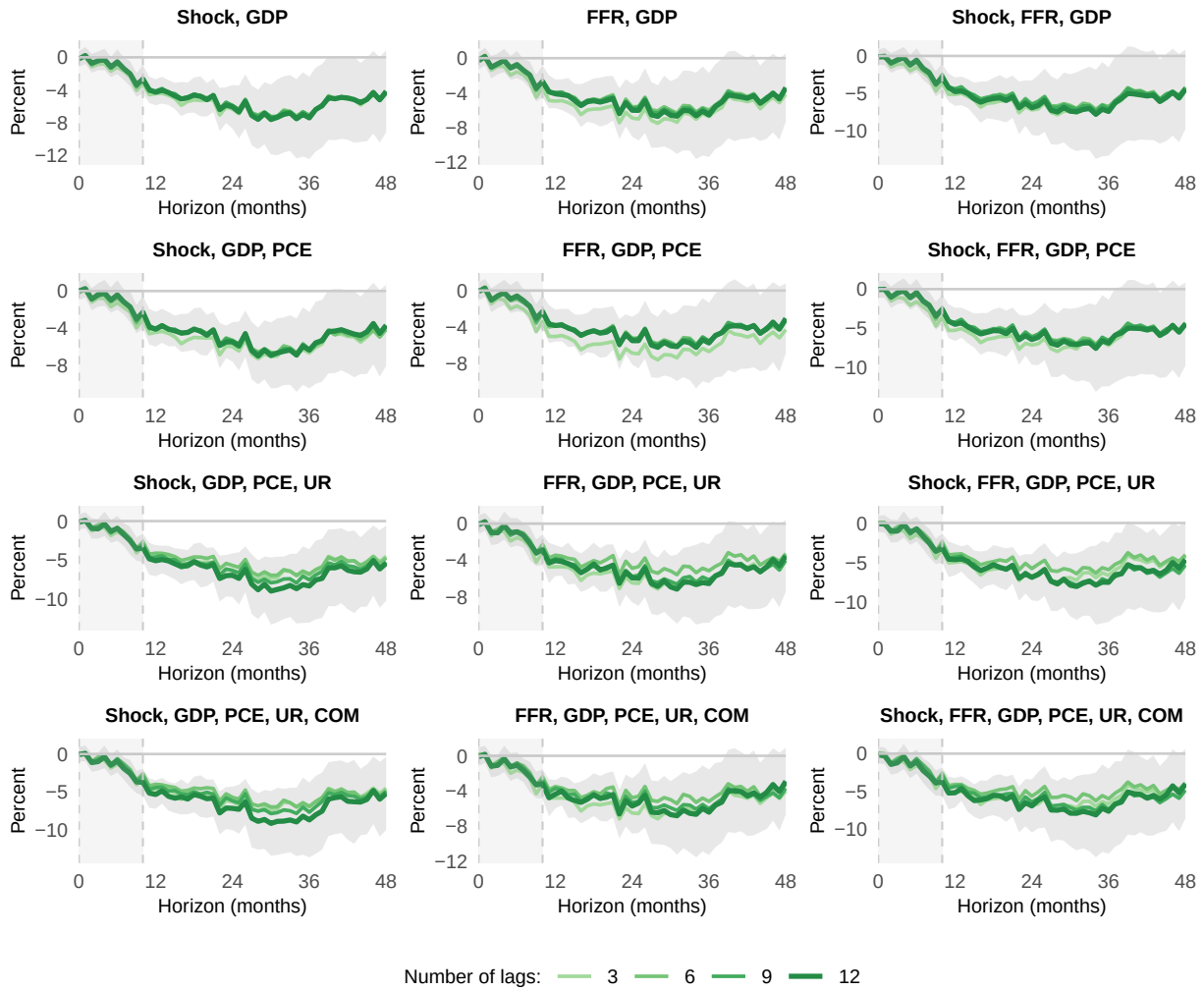


Notes: The response of GDP to a 1 percentage point Aruoba-Drechsel monetary policy shock (Aruoba and Drechsel, 2024). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 10$ mark the expected lag for the output response, based on prior empirical evidence.

In contrast, the GDP responses from the long-differences specification (Figure 2) imply excessive persistence. This specification produces a larger and more persistent decline, with the peak response of around -7% occurring roughly three years after the shock,

followed by a slower return to baseline. For the information criteria, AIC favours broader control sets with longer lags, whereas BIC and CV generally favour shorter lag lengths.

Figure 2: *Response of GDP from an Aruoba-Drechsel monetary policy shock in long-differences*



Notes: The response of GDP to a 1 percentage point Aruoba-Drechsel monetary policy shock ([Aruoba and Drechsel, 2024](#)). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 10$ mark the expected lag for the output response, based on prior empirical evidence.

Overall, the main source of variation in the results is not the choice of control set or lag length, but the choice between the levels and long-differences specifications. The long-differences specification produces a substantially larger and more persistent negative

Table 4: *Information criteria for an Aruoba-Drechsel shock on GDP*

Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, GDP	3	865	901	2.4	1020	1052	3.52
	6	850	907	2.18	1011	1065	3.76
	9	840	919	2.09	1011	1086	4.05
	12	842	942	2.11	1020	1117	4.33
FFR, GDP	3	857	893	2.57	1030	1062	3.68
	6	848	906	2.41	1028	1081	4.03
	9	842	921	2.31	1028	1103	4.31
	12	845	946	2.31	1036	1132	4.62
Shock, FFR, GDP	3	849	895	2.46	1022	1065	3.53
	6	837	916	2.29	1020	1095	3.91
	9	833	944	2.25	1025	1132	4.35
	12	840	983	2.32	1035	1175	4.76
Shock, GDP, PCE	3	816	862	2.02	1009	1052	3.37
	6	797	876	1.83	1000	1075	3.58
	9	788	899	1.81	1000	1108	4.02
	12	787	930	1.82	1012	1152	4.36
FFR, GDP, PCE	3	797	844	1.93	1019	1062	3.49
	6	777	856	1.7	1017	1092	3.8
	9	768	879	1.6	1019	1126	4.16
	12	767	910	1.58	1027	1167	4.53
Shock, FFR, GDP, PCE	3	792	849	1.87	1007	1061	3.31
	6	761	861	1.57	1004	1101	3.63
	9	750	893	1.5	1012	1151	4.18
	12	748	934	1.47	1022	1204	4.61
Shock, GDP, PCE, UR	3	764	821	1.52	996	1050	3.33
	6	749	849	1.4	988	1085	3.33
	9	743	886	1.38	993	1133	3.69
	12	742	928	1.4	1006	1189	4.09
FFR, GDP, PCE, UR	3	721	778	1.33	1004	1057	3.49
	6	712	813	1.25	994	1091	3.57
	9	710	854	1.25	991	1131	3.77
	12	714	900	1.28	996	1179	4.1
Shock, FFR, GDP, PCE, UR	3	723	791	1.35	1001	1065	3.42
	6	714	836	1.28	996	1114	3.61
	9	715	891	1.32	997	1169	3.93
	12	720	950	1.39	1002	1228	4.3
Shock, GDP, PCE, UR, COM	3	761	830	1.64	999	1064	3.56
	6	752	874	1.55	996	1114	3.73
	9	750	926	1.57	1005	1177	4.15
	12	748	978	1.6	1021	1246	4.66
FFR, GDP, PCE, UR, COM	3	717	786	1.43	1005	1070	3.74
	6	706	828	1.38	1001	1120	3.95
	9	704	880	1.41	1000	1172	4.14
	12	706	935	1.47	1003	1229	4.51
Shock, FFR, GDP, PCE, UR, COM	3	720	799	1.45	1003	1078	3.68
	6	710	853	1.41	1003	1143	4.05
	9	710	918	1.48	1007	1211	4.42
	12	713	985	1.59	1013	1282	4.93

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

effect in its point estimates, whereas the levels specification more closely matches the expected hump-shaped output response to a monetary policy shock.

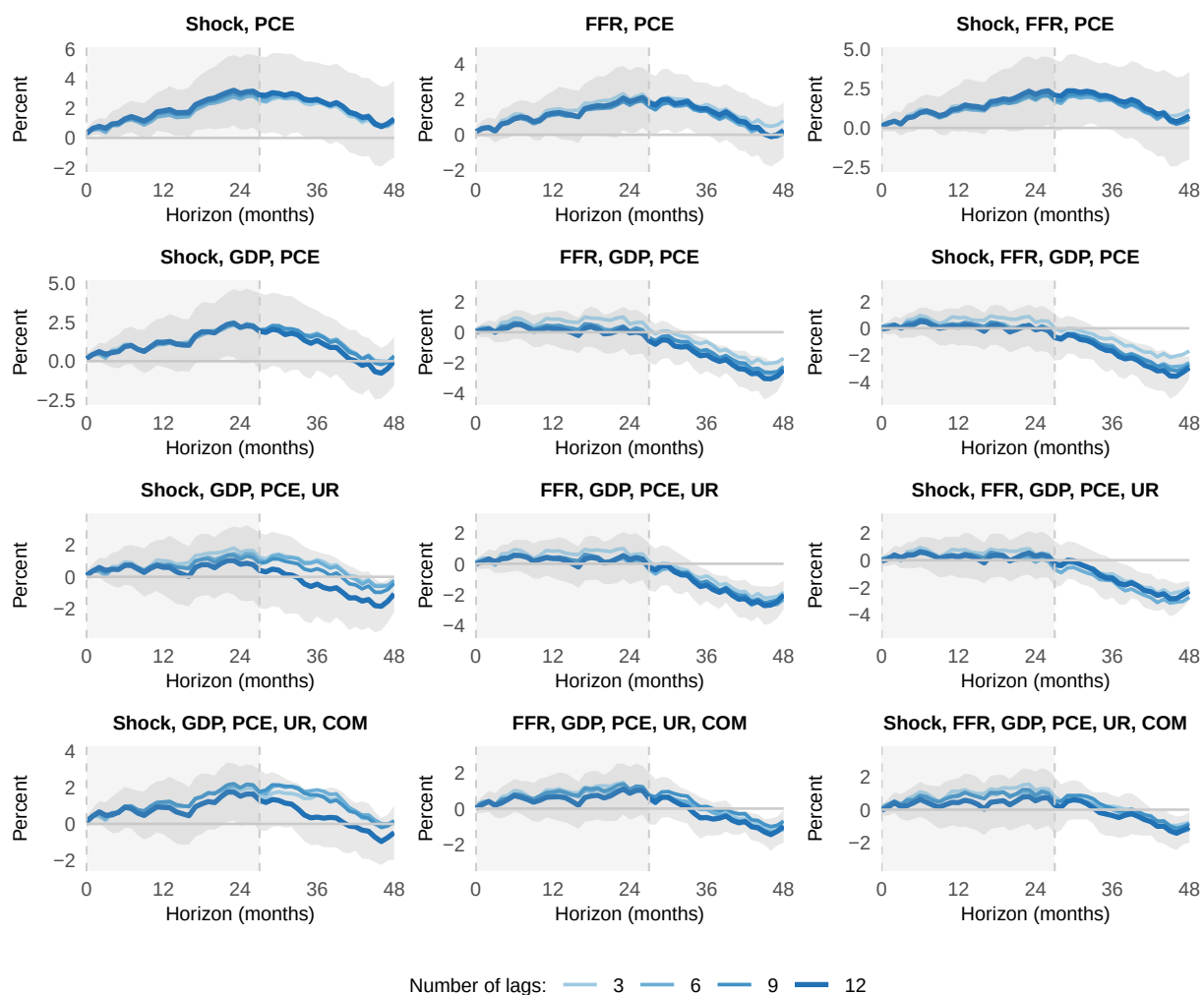
5.2 Price Level Responses

Figure 3 shows the impulse response of the PCE price index to a monetary policy shock in levels. The responses vary substantially. In particular, failing to control for the federal funds rate or GDP leads to a pronounced price puzzle and little to no negative effect. This is inconsistent with the expected response. Indeed, specifications that do not control for the federal funds rate or GDP are also associated with worse information criteria scores (see Table 5). By contrast, control sets that include at least six lags of the federal funds rate, the PCE price index, and GDP produce responses that behave as expected. These responses display a two-year lag, followed by a gradual and statistically significant decline at the 10 percent level, then an upward movement by the fourth year.

Figure 4, in contrast, shows the impulse response of the PCE price index in long-differences. Here, the shape is broadly similar across control sets and lag lengths, and price puzzles are largely absent. In general, the responses show a lag of three years, followed by a gradual but statistically insignificant decline, and then an upward movement by the fourth year. The information criteria reported in Table 5 also tend to favour narrower control sets.

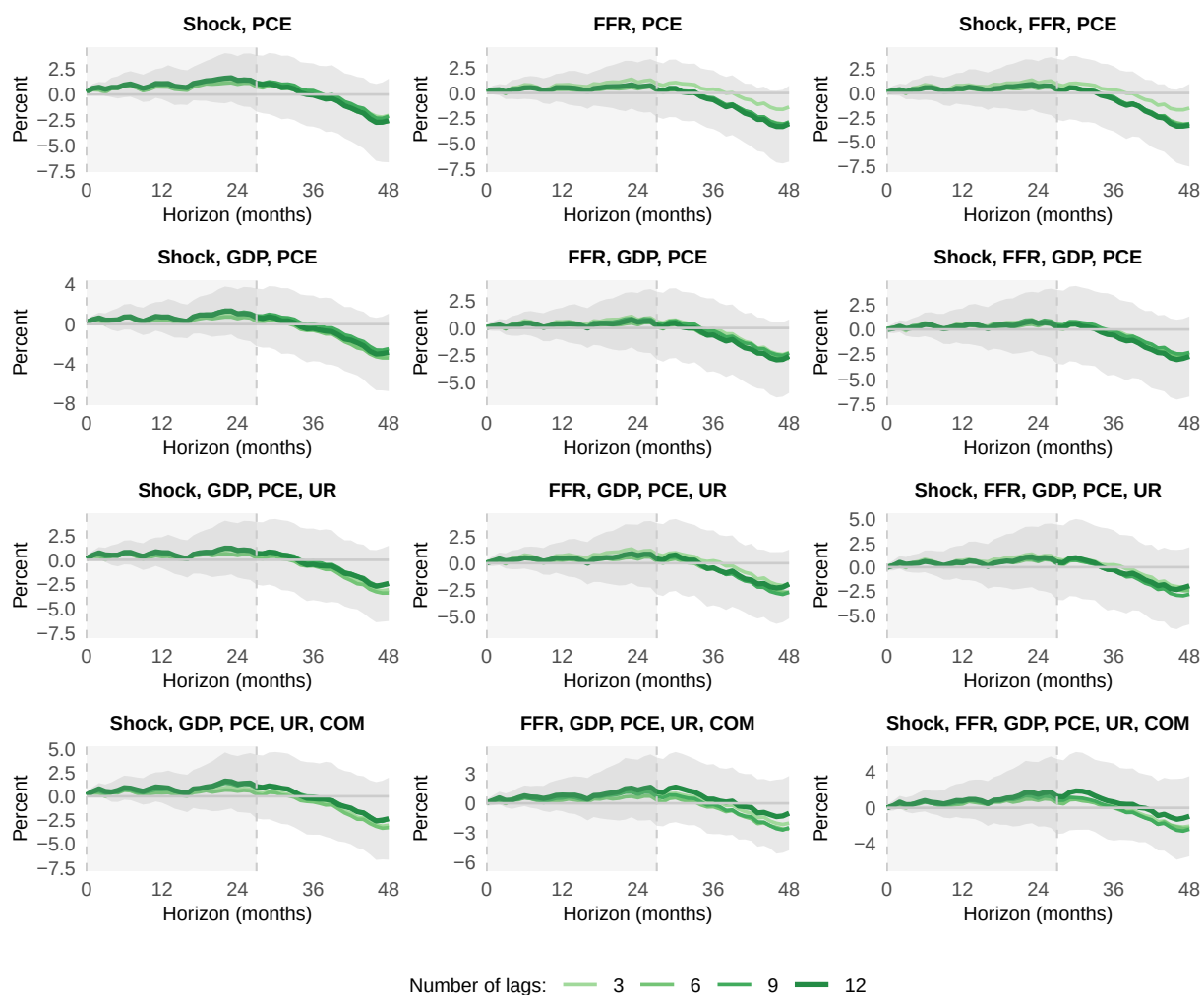
Overall, in the levels specification, different control sets produce different impulse responses. The response behaves as expected once six lags of the federal funds rate, GDP, and the PCE price index are included. As with GDP, there are also considerable differences between the levels and long-differences specifications. The long-differences specification produces a longer lag and is not statistically significant at the 10 percent level, whereas the levels specification more closely matches the expected lag of two years and is statistically significant at the 10 percent level.

Figure 3: *Response of the PCE price index from an Aruoba-Drechsel monetary policy shock in levels*



Notes: The response of the PCE price index to a 1 percentage point Aruoba-Drechsel monetary policy shock (Aruoba and Drechsel, 2024). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 27$ mark the expected lag for the price level response, based on prior empirical evidence.

Figure 4: *Response of the PCE price index from an Aruoba-Drechsel monetary policy shock in long-differences*



Notes: The response of the PCE price index to a 1 percentage point Aruoba-Drechsel monetary policy shock (Aruoba and Drechsel, 2024). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 27$ mark the expected lag for the price level response, based on prior empirical evidence.

Table 5: Information criteria for an Aruoba-Drechsel shock on the PCE price index

Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, PCE	3	752	788	1.71	893	926	3.35
	6	745	802	1.77	874	928	2.9
	9	748	827	1.91	872	947	2.77
	12	754	854	2.03	879	976	2.81
FFR, PCE	3	732	768	1.64	892	924	3.33
	6	717	775	1.53	872	926	2.92
	9	713	792	1.54	870	945	2.88
	12	714	814	1.6	877	974	3.01
Shock, FFR, PCE	3	734	781	1.7	898	941	3.44
	6	719	798	1.75	883	958	3.12
	9	720	831	1.92	885	993	3.13
	12	726	870	2.05	896	1036	3.32
Shock, GDP, PCE	3	729	776	2.06	895	938	3.39
	6	724	803	2	879	954	3.03
	9	721	832	1.98	881	989	2.96
	12	713	856	1.94	892	1032	3.07
FFR, GDP, PCE	3	645	691	1.16	893	936	3.33
	6	617	696	0.96	875	950	2.88
	9	597	708	0.84	876	984	2.81
	12	575	719	0.75	888	1028	3.01
Shock, FFR, GDP, PCE	3	650	707	1.18	899	953	3.43
	6	627	727	1.03	884	981	3.03
	9	609	753	0.97	892	1032	3.08
	12	588	774	0.93	908	1091	3.46
Shock, GDP, PCE, UR	3	687	745	1.59	899	952	3.52
	6	685	785	1.55	890	987	3.19
	9	682	825	1.55	896	1035	3.17
	12	671	857	1.51	911	1094	3.53
FFR, GDP, PCE, UR	3	641	699	1.18	895	949	3.36
	6	618	718	0.97	884	981	2.97
	9	597	741	0.85	890	1030	2.94
	12	578	764	0.77	905	1088	3.21
Shock, FFR, GDP, PCE, UR	3	646	714	1.2	900	964	3.43
	6	627	749	1.03	893	1011	3.12
	9	607	783	0.94	906	1078	3.25
	12	589	819	0.92	925	1151	3.68
Shock, GDP, PCE, UR, COM	3	642	710	1.36	904	969	3.66
	6	625	747	1.22	901	1019	3.46
	9	605	781	1.13	910	1083	3.58
	12	572	802	0.98	930	1156	4
FFR, GDP, PCE, UR, COM	3	567	635	0.81	901	965	3.48
	6	537	658	0.69	895	1013	3.19
	9	519	694	0.68	905	1077	3.25
	12	503	732	0.66	922	1148	3.6
Shock, FFR, GDP, PCE, UR, COM	3	568	647	0.81	905	981	3.57
	6	537	681	0.69	903	1043	3.4
	9	525	733	0.7	920	1125	3.67
	12	502	775	0.7	943	1212	4.2

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

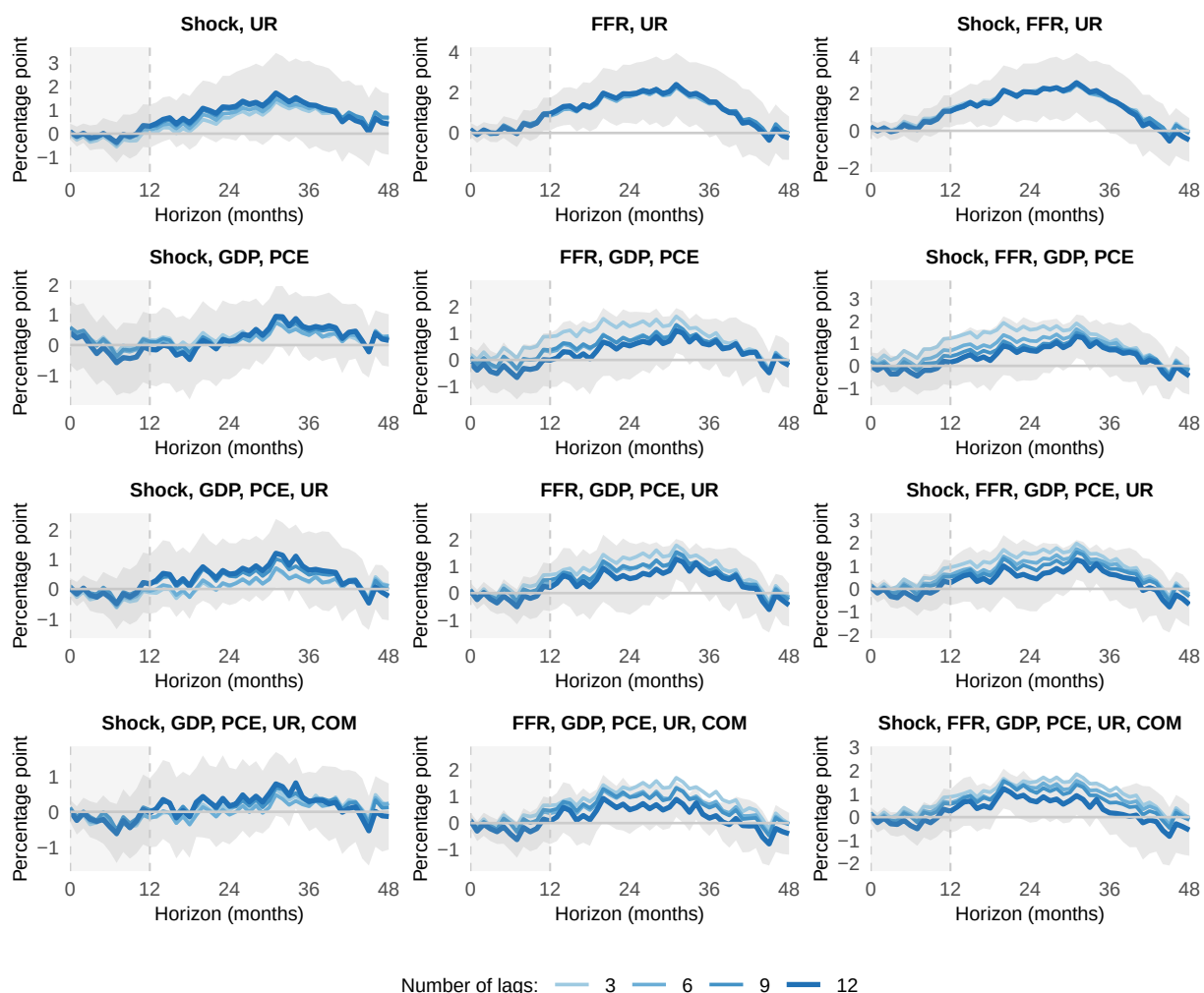
5.3 Unemployment Responses

Figure 5 shows the response of the unemployment rate to a monetary policy shock in levels. The information criterion results in Table 6 favour broader control sets that include the unemployment rate and exclude the shock: AIC and CV prefer six or nine lags, while BIC prefers three lags. These broader control sets also produce responses that align with expectations. These responses display a lag of about 12 months, after which the unemployment rate rises and peaks roughly 30 months after the shock. The peak is about 1 percentage point, after which the responses gradually return toward baseline. By contrast, control sets that exclude the federal funds rate tend to produce flatter and more distorted impulse responses.

Figure 6 shows the impulse response of the unemployment rate in long-differences. Here, the responses are very similar across control sets. The responses show a larger and more persistent increase in the unemployment rate, with point estimates peaking at around 2.5 percentage points roughly three years after the shock and remaining elevated thereafter. The information criterion results in Table 6 are less clear in this case. Although AIC, BIC, and CV generally favour narrower control sets, AIC also assigns relatively strong scores to some broader control sets that exclude the shock.

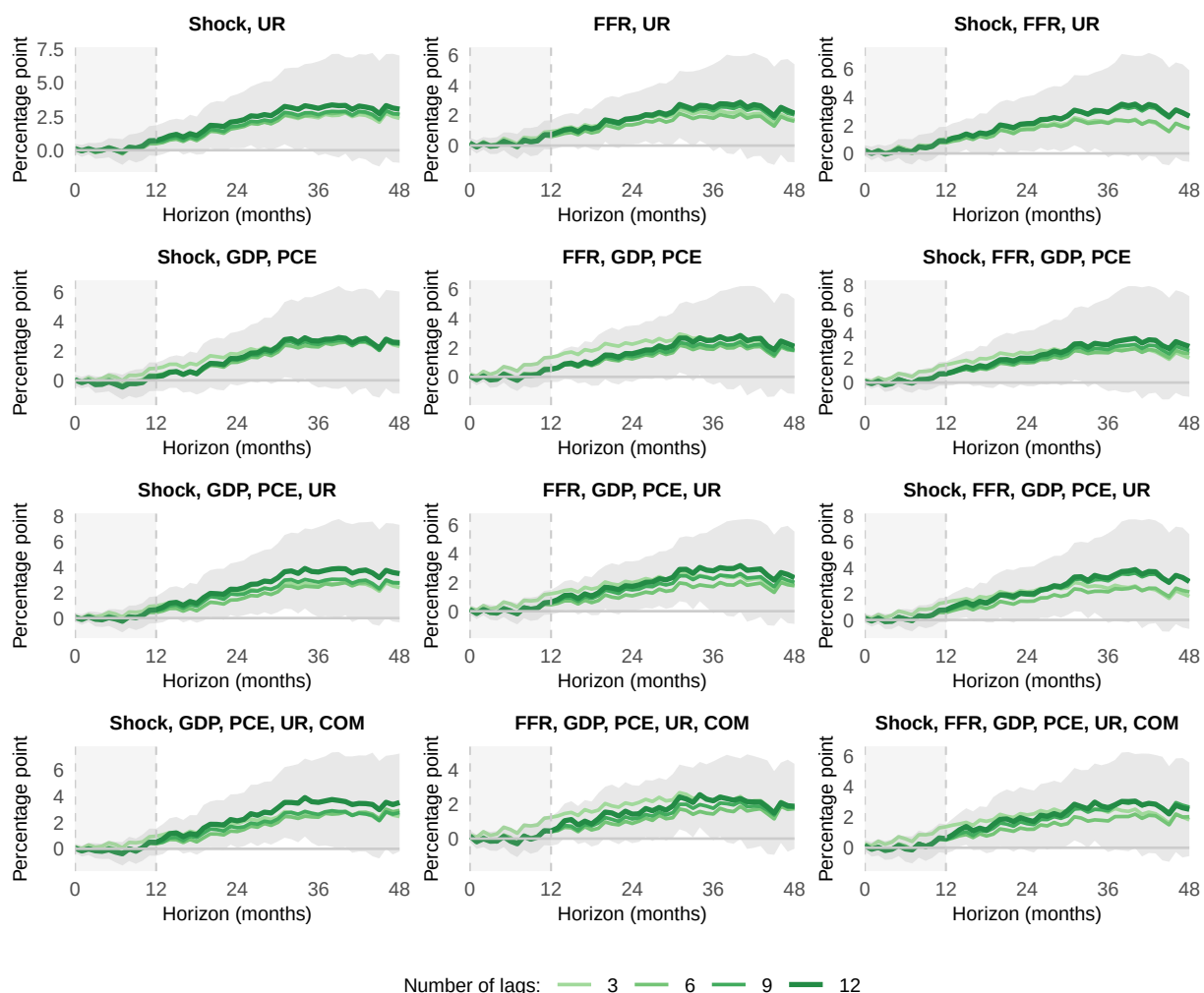
Overall, the levels specification produces unemployment rate responses that are sensitive to the choice of controls. In particular, failing to control for the federal funds rate produces a response that is less hump-shaped. More broadly, there are clear differences between the levels and long-differences specifications. In terms of point estimates, the long-differences specification produces a substantially larger and more persistent increase in the unemployment rate, with the response remaining elevated above baseline four years after the shock. The levels specification, by contrast, produces a more clearly hump-shaped response, with unemployment gradually returning toward baseline. This pattern is more consistent with the expected unemployment response to a contractionary monetary policy shock.

Figure 5: *Response of the unemployment rate from an Aruoba-Drechsel monetary policy shock in levels*



Notes: The response of the unemployment rate to a 1 percentage point Aruoba-Drechsel monetary policy shock (Aruoba and Drechsel, 2024). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 12$ mark the expected lag for the unemployment response, based on prior empirical evidence.

Figure 6: *Response of the unemployment rate from an Aruoba-Drechsel monetary policy shock in long-differences*



Notes: The response of unemployment rate to a 1 percentage point Aruoba-Drechsel monetary policy shock (Aruoba and Drechsel, 2024). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 12$ mark the expected lag for the unemployment response, based on prior empirical evidence.

Table 6: Information criteria for an Aruoba-Drechsel shock on unemployment rate

Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, UR	3	471	507	0.5	654	686	0.95
	6	457	515	0.49	654	707	0.99
	9	459	538	0.5	658	734	1.06
	12	463	563	0.52	664	761	1.17
FFR, UR	3	444	480	0.61	651	684	0.99
	6	427	484	0.6	648	702	1.03
	9	428	506	0.61	644	719	1.08
	12	429	530	0.61	641	737	1.16
Shock, FFR, UR	3	439	486	0.6	654	697	1
	6	424	503	0.61	656	731	1.06
	9	430	542	0.63	655	763	1.13
	12	439	582	0.65	653	793	1.19
Shock, GDP, PCE	3	537	583	0.69	665	708	1.2
	6	516	595	0.62	657	732	1.37
	9	506	617	0.6	664	772	1.56
	12	501	644	0.6	677	817	1.68
FFR, GDP, PCE	3	496	542	0.67	660	703	1.19
	6	458	537	0.56	654	729	1.37
	9	439	550	0.51	658	765	1.51
	12	428	572	0.49	662	802	1.61
Shock, FFR, GDP, PCE	3	483	540	0.63	656	710	1.19
	6	424	525	0.48	655	752	1.4
	9	399	542	0.44	664	804	1.54
	12	387	574	0.42	665	848	1.6
Shock, GDP, PCE, UR	3	418	475	0.52	648	702	1.18
	6	387	488	0.46	649	746	1.35
	9	384	527	0.44	664	803	1.53
	12	387	574	0.46	676	858	1.66
FFR, GDP, PCE, UR	3	350	408	0.39	646	700	1.17
	6	316	417	0.34	643	740	1.31
	9	315	459	0.34	648	787	1.39
	12	325	511	0.36	648	831	1.42
Shock, FFR, GDP, PCE, UR	3	351	419	0.39	648	713	1.18
	6	316	438	0.35	651	769	1.36
	9	319	495	0.35	658	830	1.42
	12	328	558	0.37	658	884	1.45
Shock, GDP, PCE, UR, COM	3	422	490	0.58	653	717	1.23
	6	391	513	0.5	656	775	1.44
	9	382	558	0.48	673	845	1.67
	12	375	604	0.47	689	915	1.86
FFR, GDP, PCE, UR, COM	3	338	407	0.4	651	715	1.22
	6	296	417	0.35	649	768	1.36
	9	289	464	0.34	652	824	1.45
	12	291	520	0.35	653	879	1.5
Shock, FFR, GDP, PCE, UR, COM	3	341	420	0.41	653	728	1.23
	6	301	445	0.36	658	797	1.43
	9	300	508	0.37	664	868	1.53
	12	302	575	0.4	668	937	1.62

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

6 The Role of Control Transformation in Local Projections

In this section, I explain why the levels and long-differences local projections produce different impulse responses.

6.1 The Difference Comes from First-Differencing the Controls

In the levels specification, the lagged controls enter in levels, whereas in the long-differences specification they are transformed into first-differences. To examine whether this transformation drives the difference in the results, I plot the impulse responses under both specifications while varying the transformation applied to the control variables.

Figure 7 shows the results. Notably, once the long-differences specification is estimated with controls in levels, the resulting impulse responses are virtually identical to those from the levels specification, which also uses controls in levels. This suggests that the difference in impulse responses between the levels and long-differences specifications is driven entirely by the transformation of the control variables, rather than by the transformation of the outcome variable.

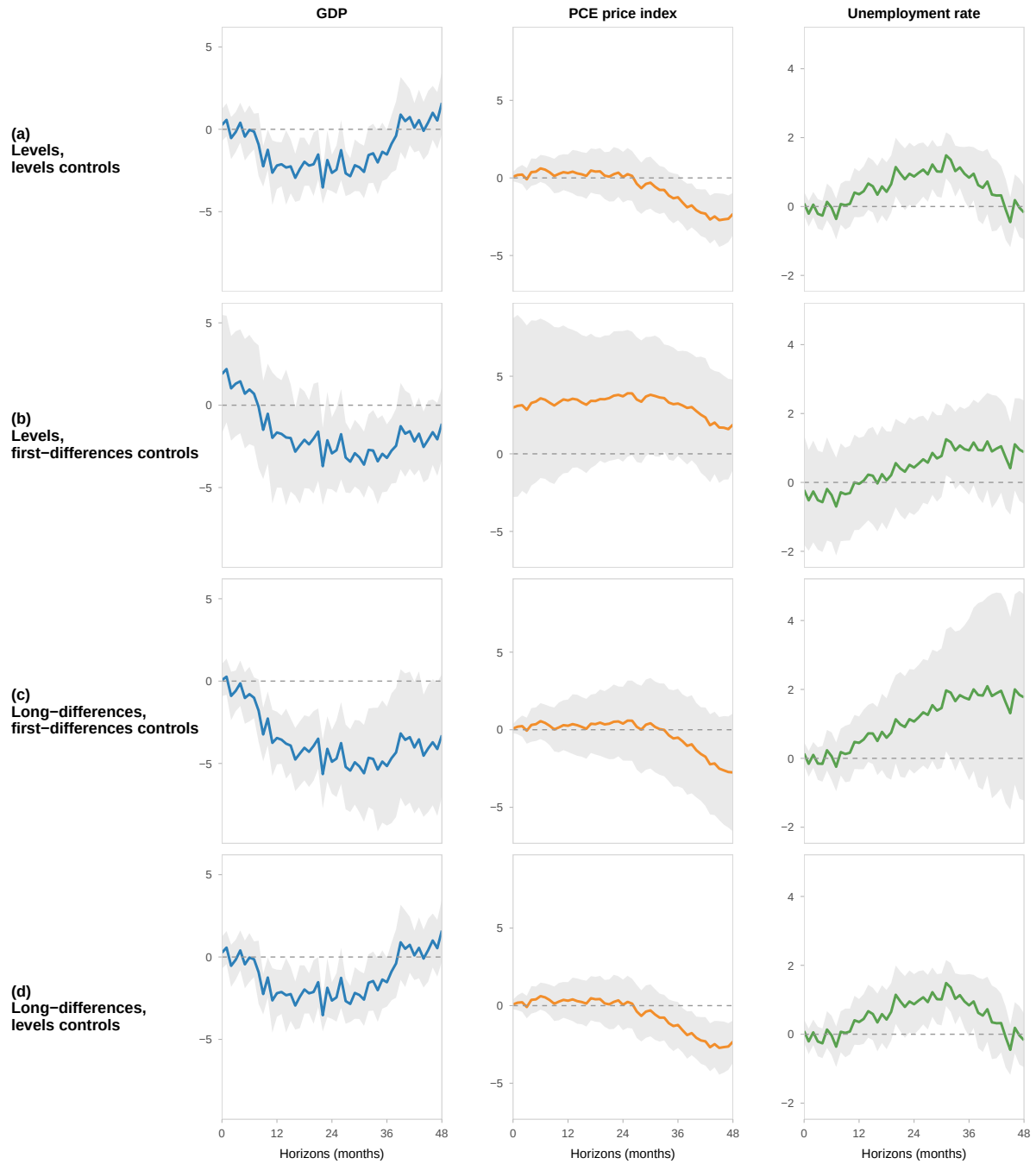
6.2 Omitted Variable Bias from First-Differenced Controls

A natural explanation for why the point estimates differ when level controls are used rather than first-differenced controls is that first-differencing the controls introduces omitted variable bias. Lags of macroeconomic variables in levels may be correlated with the shock and may also independently affect the outcome in important ways that the first-differences of those variables fail to capture.

More formally, suppose that the true model for the outcome is

$$y_{t+h} - y_{t-1} = \delta_h + \beta_h s_t + \theta'_h W_t + v_{t+h}, \quad (10)$$

Figure 7: *Levels vs. Long-Differences Under Alternative Control Transformations*



Notes: Impulse responses of log GDP, log PCE price index, and the unemployment rate to a 1 percentage point Aruoba–Drechsel monetary policy shock (Aruoba and Drechsel, 2024). Row (a) estimates local projections in levels with controls in levels. Row (b) estimates local projections in levels with controls in first-differences. Row (c) estimates local projections in long-differences with controls in first-differences. Row (d) estimates local projections in long-differences with controls in levels. All specifications use the same sample (October 1983 to December 2007) and the same control set: six lags of the federal funds rate, log GDP, log PCE, and the unemployment rate. All level control sets include a linear time trend. Shaded bands denote 90% confidence intervals.

where W_t denotes the controls needed for exogeneity, so that $E[\nu_{t+h} \mid s_t, W_t] = 0$.

The econometrician does not observe W_t and instead estimates

$$y_{t+h} - y_{t-1} = a_h + b_h s_t + c_h' X_t^D + \varepsilon_{t+h}, \quad (11)$$

where X_t^D is a vector of some lagged macroeconomic variables in first-differences. Let $\hat{\beta}_h$ denote the OLS coefficient on s_t in (11). The goal is to show that the associated OLS estimator is biased whenever X_t^D fails to fully capture the information in W_t .

By the Frisch–Waugh–Lovell theorem, the OLS estimator $\hat{\beta}_h$ from (11) can be written as

$$\hat{\beta}_h = \frac{\sum_t s_t^* y_{t+h}^*}{\sum_t (s_t^*)^2}, \quad (12)$$

where s_t^* and y_{t+h}^* denote the residuals from projecting s_t and $y_{t+h} - y_{t-1}$ onto a constant and X_t^D , respectively.

Substituting (10) into (12) yields

$$\hat{\beta}_h = \frac{\sum_t s_t^* (\beta_h s_t^* + \theta_h' W_t^* + \nu_{t+h}^*)}{\sum_t (s_t^*)^2},$$

where W_t^* and ν_{t+h}^* denote the residuals from projecting W_t and ν_{t+h} onto a constant and X_t^D , respectively.

Expanding this expression gives

$$\hat{\beta}_h = \beta_h + \theta_h' \frac{\sum_t s_t^* W_t^*}{\sum_t (s_t^*)^2} + o_p(1). \quad (13)$$

The first term recovers the true impulse response β_h ; the second is the omitted variable bias. A full derivation is provided in Appendix 7.5.4. The long-differences estimator is therefore biased whenever $\theta_h \neq 0$ and $\frac{\sum_t s_t^* W_t^*}{\sum_t (s_t^*)^2} \neq 0$, that is, whenever W_t affects the outcome and X_t^D fails to absorb every component of W_t that is correlated with s_t . Moreover, the bias

can remain large even when the shock is nearly exogenous. A small residual correlation between s_t and W_t , multiplied by a large θ_h , can still produce substantial bias.

6.3 Sources of Omitted Variable Bias

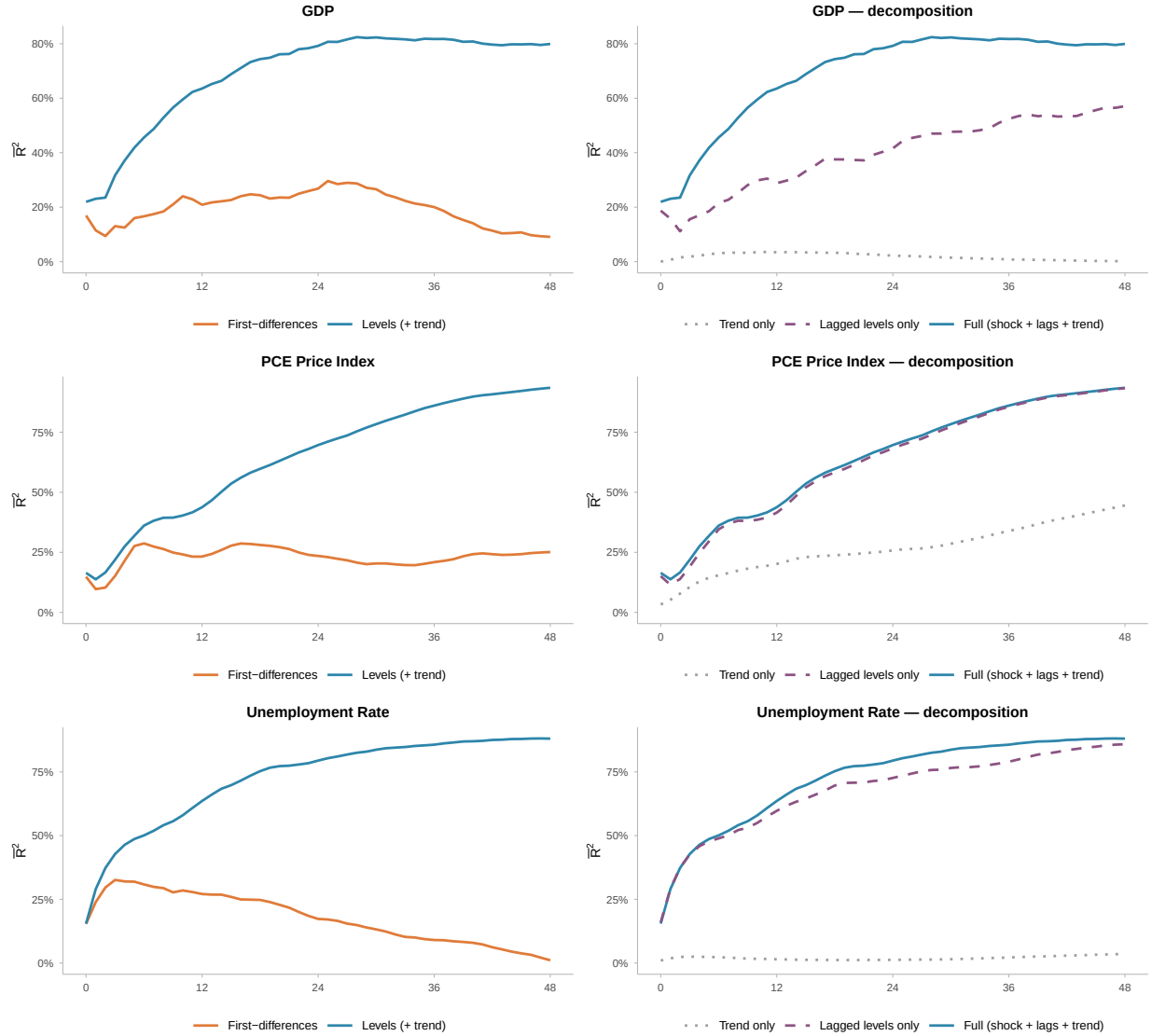
To assess whether first-differenced controls generate more omitted variable bias than level controls, I compare their predictive power for the shock and for the outcome variable.

I begin by examining their predictive power for the shock s_t . Specifically, I regress the shock on six lags of each control set and evaluate the results using the joint significance of the controls and the adjusted R^2 . The null hypothesis in the F -test is that all lagged controls are jointly zero. Table 7 reports the results. The F -statistics, p -values, and adjusted R^2 are nearly identical across the two columns for every control set. Thus, any difference in omitted-variable bias between levels and first-differenced controls does not appear to arise from differences in how much variation in the shock the controls absorb. In terms of the bias formula in (13), this implies that $\sum_t (s_t^*)^2$ is similar across the two specifications, so any difference in bias must operate through $\sum_t s_t^* W_t^*$.

I next examine the relative predictive power of the controls for the outcome variable, $y_{t+h} - y_{t-1}$. The left panels of Figure 8 plot the adjusted R^2 for each specification at horizons $h = 0, \dots, 48$. At short horizons, the two control sets provide a similar fit. As the horizon increases, however, controls in levels explain a substantially larger share of the variation in the outcome variable than first-differenced controls. The right panels of Figure 8 further show that this stronger fit is driven by the lagged macroeconomic variables rather than by the linear trend. In terms of the bias formula in (13), a lower adjusted R^2 suggests that first-differenced controls leave more of W_t unabsorbed, which implies a larger W_t^* . First-differenced controls are therefore more likely to generate omitted-variable bias.

Finally, to visualize the level information that first-differenced controls fail to absorb, I plot the level controls, X_t , against X_t^* , the residual from projecting X_t onto a constant and the first-differenced controls, X_t^D . Figure 9 shows that, for all four variables, X_t^* closely

Figure 8: *Adjusted R^2 across horizons: levels vs. first-differenced controls*



Notes: Left panels show the adjusted R^2 of the levels (blue) and first-differences (orange) local projections at each horizon $h = 0, \dots, 48$. The levels specification includes six lags of the federal funds rate, log GDP, log PCE price index, and the unemployment rate in levels, along with a linear time trend. The first-differences specification uses the same variables transformed into first-differences, with no trend. Right panels decompose the fit of the levels specification into three components, each estimated on the same sample: the adjusted R^2 from regressing $y_{t+h} - y_{t-1}$ on the linear trend alone (dotted grey), on the six lags of the macro controls alone (dashed purple), and on the full model including the shock, lags, and trend (solid blue).

Table 7: Joint F -test of lagged controls on the monetary policy shock

Control set	Levels			First-differences		
	F	p -value	$\bar{R}^2$	F	p -value	$\bar{R}^2$
Shock, GDP	1.83	0.043**	0.029	1.79	0.050**	0.030
FFR, GDP	3.19	<0.001***	0.075	3.12	<0.001***	0.075
Shock, FFR, GDP	3.04	<0.001***	0.105	2.96	<0.001***	0.103
Shock, GDP, PCE	1.68	0.041**	0.036	1.55	0.073*	0.031
FFR, GDP, PCE	2.67	<0.001***	0.086	2.50	<0.001***	0.079
Shock, FFR, GDP, PCE	2.61	<0.001***	0.110	2.46	<0.001***	0.102
Shock, GDP, PCE, UR	1.47	0.077*	0.033	1.32	0.151	0.024
FFR, GDP, PCE, UR	2.20	0.001***	0.082	2.08	0.003***	0.076
Shock, FFR, GDP, PCE, UR	2.21	<0.001***	0.104	2.10	0.001***	0.098
Shock, GDP, PCE, UR, COM	1.29	0.152	0.025	1.15	0.279	0.014
FFR, GDP, PCE, UR, COM	1.86	0.005***	0.074	1.75	0.011**	0.067
Shock, FFR, GDP, PCE, UR, COM	1.94	0.002***	0.098	1.83	0.004***	0.089

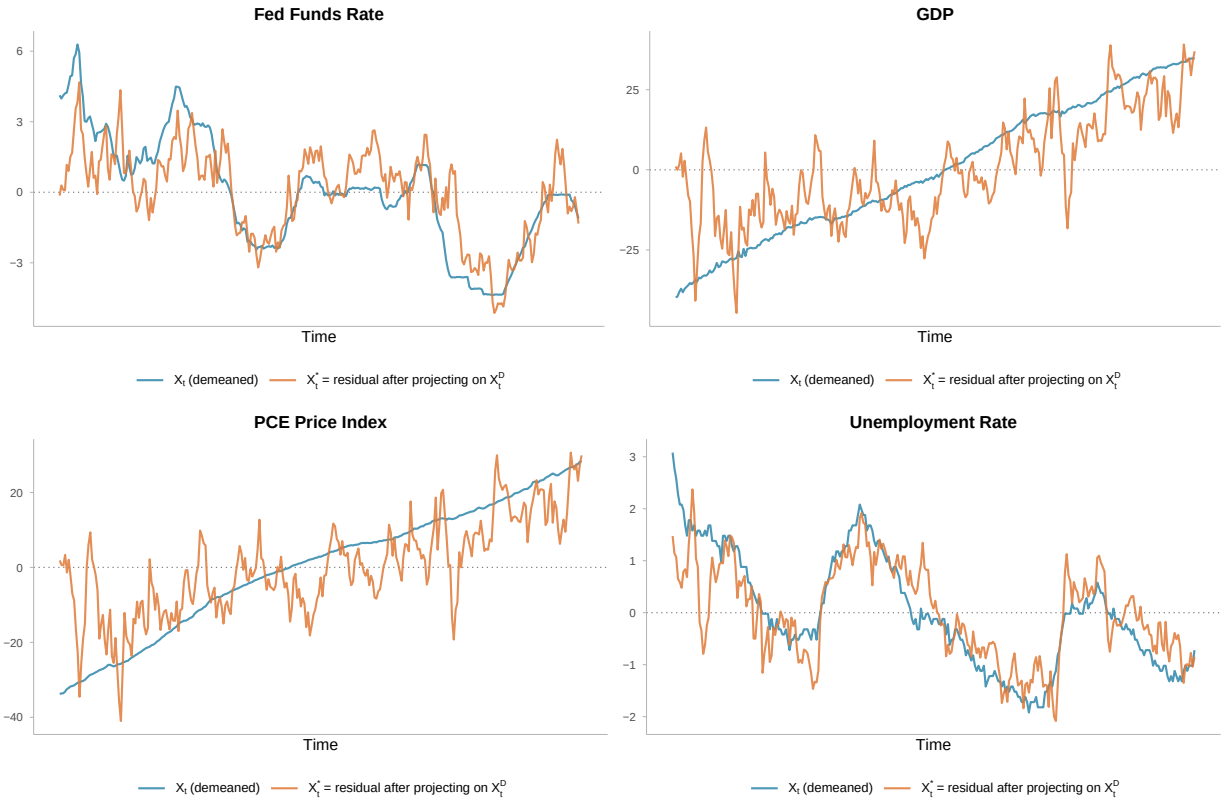
Notes: Each row reports the F -statistic, p -value, and $\bar{R}^2$ from regressing the [Aruoba and Drechsel \(2024\)](#) monetary policy shock on six lags of the indicated control set. The levels specification includes a linear time trend. Lags of the shock are always included in levels. The null hypothesis is that all lagged controls are jointly zero. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

tracks X_t , indicating that first-differenced controls fail to absorb important information contained in X_t . This likely reflects the level information that remains in W_t^* and gives rise to the omitted-variable bias in (13).

7 Summary and Concluding Remarks

The results show that the long-differences specification distorts long-run dynamics in its point estimates. Output and unemployment display large and persistent effects, while the price level responds with a longer-than-expected lag. I show that this divergence is driven by transforming the control variables into first-differences, which induces omitted variable bias. When the long-differences specification is estimated with controls in levels, the resulting impulse responses are virtually identical to those from the levels specification. Further, I show that first-differencing the controls weakens their predictive power for the outcome variable at longer horizons, while keeping the controls in levels does not.

Figure 9: *Level information that first-differenced controls fail to capture*



Notes: Each panel plots the demeaned level X_t (blue) against X_t^* (orange) for the federal funds rate, log GDP, log PCE price index, and the unemployment rate. X_t^* is the residual from projecting X_t onto a constant and six lags of the first-differenced controls, X_t^D .

For the levels specification, including lags of the federal funds rate, GDP, the PCE price index, and the unemployment rate, along with a linear time trend produces well behaved responses. Notable features of this control set—and others similar to it—are that output and unemployment responses display a lag of approximately eight to ten months and exhibit a hump-like shape. The price level response displays a lag of approximately two years, no price puzzle, and a decline that is statistically significant at the 10 percent level. These well-behaved dynamics were previously difficult to obtain in samples covering the Great Moderation (see [Barakchian and Crowe 2013](#) and [Ramey 2016](#)), but here they emerge clearly once the appropriate control set is used. In most cases, six months of lags is enough, as the estimated response tends to stabilize—meaning that increasing the lag to nine or twelve produces little to no change. Importantly, all three information criteria converge in favouring broader control sets for which lags of the federal funds rate are preferred over lags of the shock.

As local projections become more widely used by central banks and policymakers, selecting the right specification, along with appropriate controls and lag lengths, is essential for accurately estimating the timing and shape of responses in key macroeconomic variables. Inadequate controls or poorly chosen specifications can lead to misleading conclusions.

Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

The dataset is assembled from the FRED-MD database ([McCracken and Ng, 2016](#)), the monetary policy shock series of [Aruoba and Drechsel \(2024\)](#) and [Wieland and Yang \(2020\)](#), and the monthly real GDP series of [Barrett \(2025\)](#), each of which is publicly available at the

sources documented in Appendix 7.3. One series (the Bloomberg CRB All Commodities index) is proprietary. A complete replication package containing all data, code, and intermediate output is available from the author on request.

References

- Aruoba, S. B. and T. Drechsel (2024). Identifying monetary policy shocks: A natural language approach. *National Bureau of Economic Research*.
- Barakchian, S. M. and C. Crowe (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics* 60(8), 950–966.
- Barrett, P. (2025). Estimated monthly national accounts for the united states. *IMF Working Papers*.
- Bauer, M. D. and E. T. Swanson (2023). A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual* 37(1), 87–155.
- Bu, C., J. Rogers, and W. Wu (2021). A unified measure of fed monetary policy shocks. *Journal of Monetary Economics* 118, 331–349.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics* 1, 65–148.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1), 1–45.
- Cochrane, J. H. (2024). Expectations and the neutrality of interest rates. *Review of Economic Dynamics* 53, 194–223.
- Cochrane, J. H. (2025). Sticky prices as adjustment to equilibrium.
- Dupor, B. (2023). Examining long and variable lags in monetary policy. *The Regional Economist*.
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.
- Hoover, K. D. (1994). Econometrics as observation: the lucas critique and the nature of econometric inference. *Journal of Economic Methodology* 1(1), 65–80.
- Inoue, A., Ò. Jordà, and G. M. Kuersteiner (2026). Inference for local projections. *The Econometrics Journal*, utaf004.
- Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics* 12(2), 1–43.
- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review* 95(1), 161–182.
- Jordà, Ò. and A. M. Taylor (2025). Local projections. *Journal of Economic Literature* 63(1), 59–110.
- McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34(4), 574–589.
- Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics* 13(3), 74–107.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Olea, J. L. M., M. Plagborg-Møller, E. Qian, and C. K. Wolf (2025). Local projections or vars? a primer for macroeconomists.
- Olea, J. L. M., M. Plagborg-Møller, E. Qian, and C. K. Wolf (2026). Double robustness of local projections and some unpleasant varithmetic.
- Piger, J. and T. Stockwell (2025). Differences from differencing: Should local projections with observed shocks be estimated in levels or differences? *Journal of Applied Economet-*

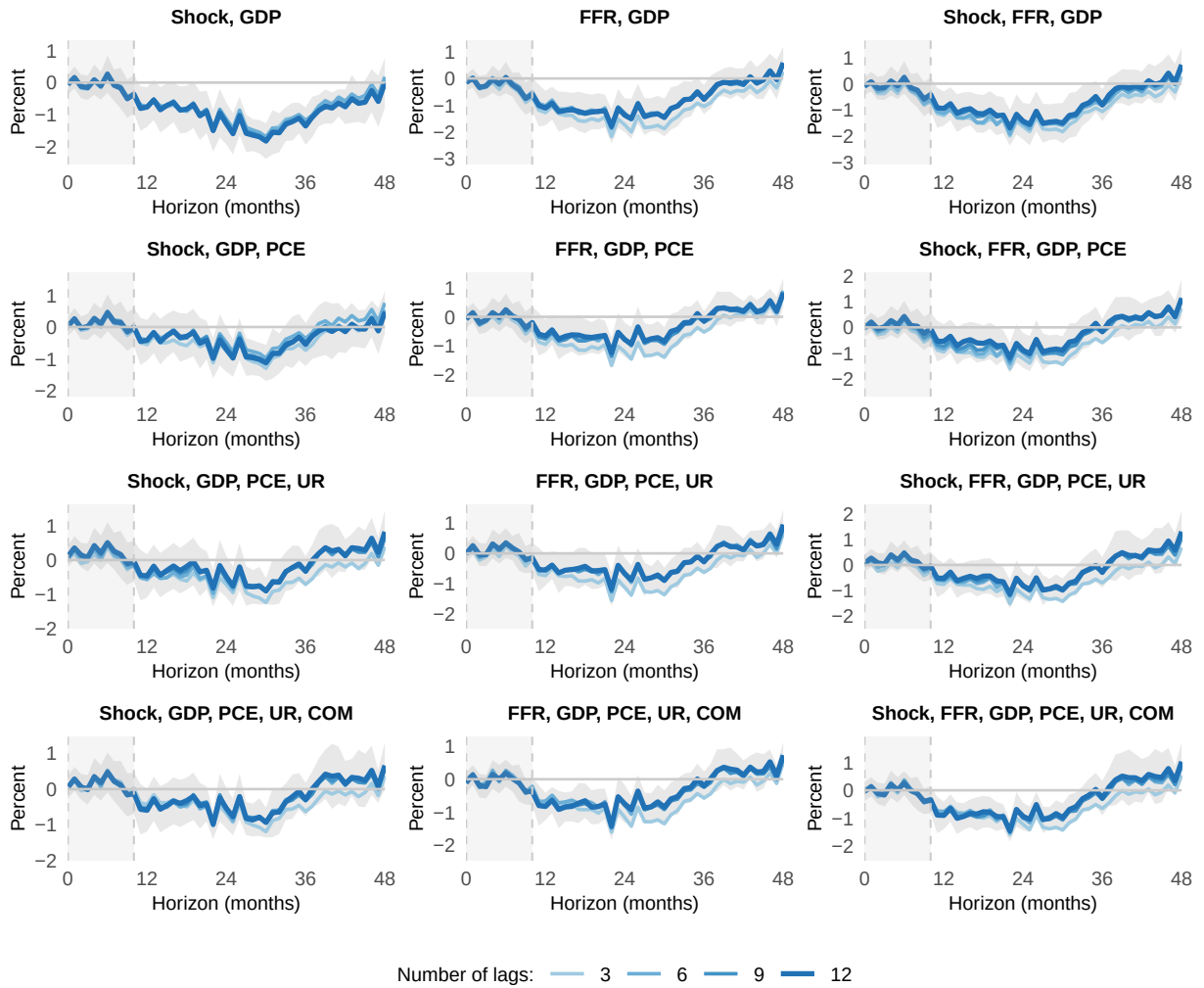
- rics 40(7), 759–787.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics* 2, 71–162.
- Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review* 94(4), 1055–1084.
- Romer, C. D. and D. H. Romer (2023). Presidential address: Does monetary policy matter? the narrative approach after 35 years. *American Economic Review* 113(6), 1395–1423.
- Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review* 97(3), 586–606.
- Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128(610), 917–948.
- Wieland, J. F. and M.-J. Yang (2020). Financial dampening. *Journal of Money, Credit and Banking* 52(1), 79–113.
- Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.

Appendix A. Romer-Romer Shock Analysis

7.1 Output Responses

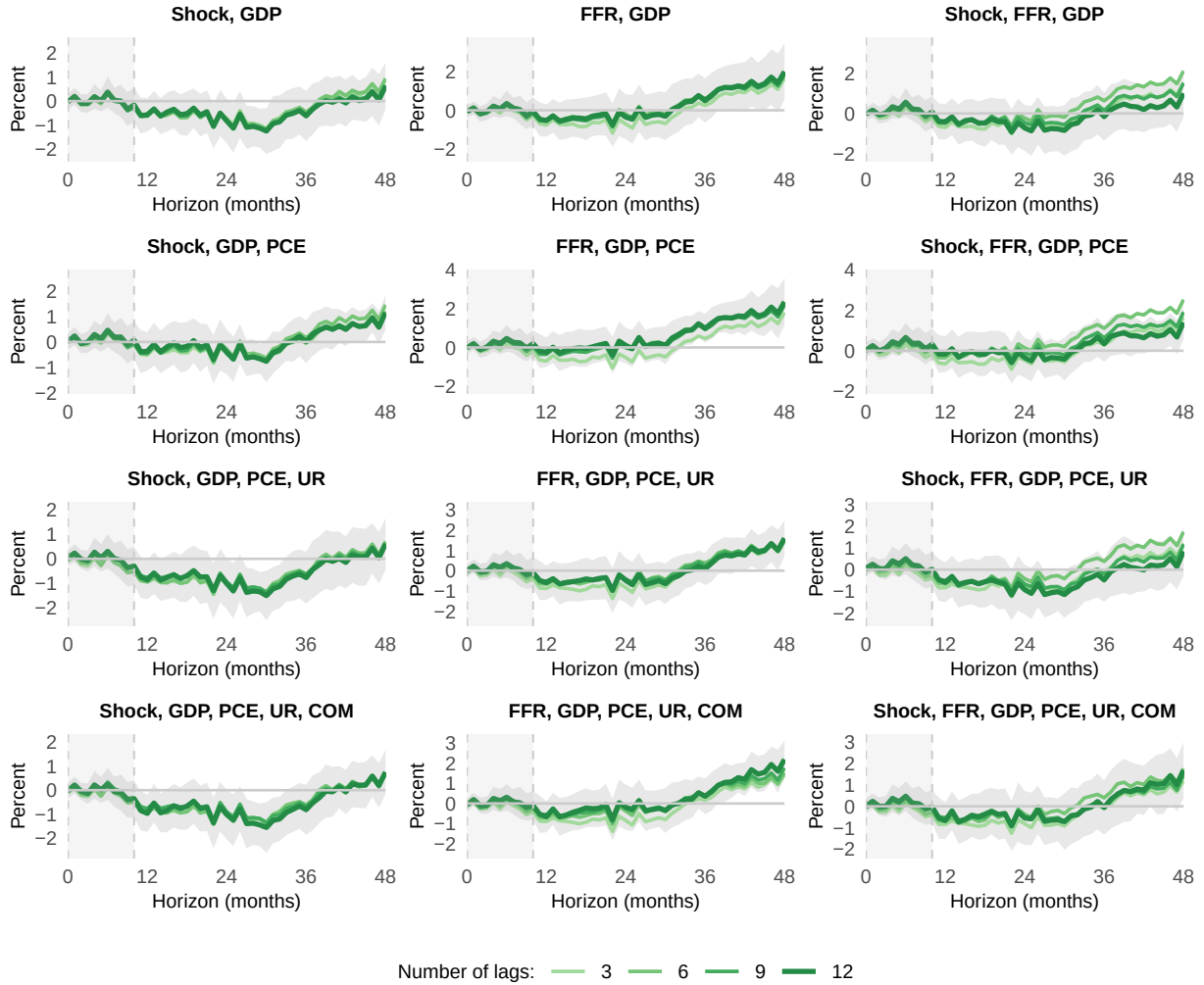
Figure 10 shows the impulse response of GDP to the Romer-Romer monetary policy shock in the levels specification, while Figure 11 shows the responses for the long-differences specification. Table 8 presents the information criteria for the Romer-Romer shock on GDP.

Figure 10: *Response of GDP from a Romer-Romer monetary policy shock in levels*



Notes: The response of GDP to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 10$ mark the expected lag for the output response, based on prior empirical evidence.

Figure 11: *Response of GDP from a Romer-Romer monetary policy shock in long-differences*



Notes: The response of GDP to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 10$ mark the expected lag for the output response, based on prior empirical evidence.

Table 8: *Information criteria for a Romer–Romer shock on GDP*

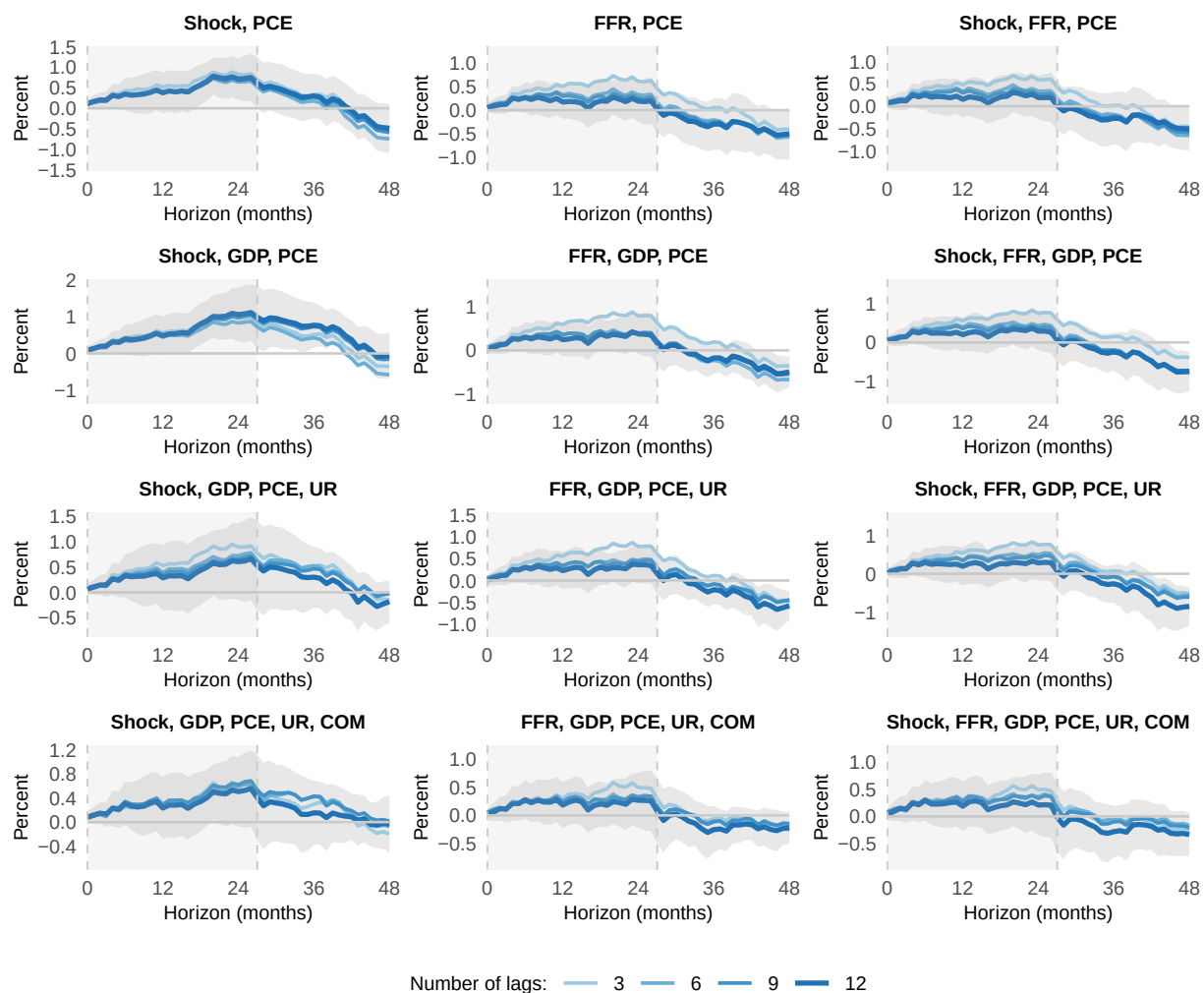
Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, GDP	3	865	901	2.4	1033	1066	3.8
	6	854	911	2.21	1034	1087	4.16
	9	845	924	2.12	1036	1111	4.55
	12	847	948	2.16	1043	1140	4.72
FFR, GDP	3	856	892	2.55	1033	1066	3.72
	6	848	906	2.41	1030	1084	4.05
	9	842	921	2.3	1031	1106	4.34
	12	844	945	2.3	1038	1135	4.66
Shock, FFR, GDP	3	849	896	2.45	1036	1079	3.76
	6	840	919	2.31	1036	1111	4.08
	9	835	947	2.29	1037	1144	4.32
	12	841	985	2.38	1045	1185	4.53
Shock, GDP, PCE	3	816	863	2.04	1022	1065	3.58
	6	800	879	1.88	1020	1095	3.89
	9	792	904	1.84	1023	1131	4.35
	12	793	936	1.88	1032	1172	4.6
FFR, GDP, PCE	3	796	843	1.93	1023	1066	3.53
	6	777	855	1.7	1019	1094	3.8
	9	767	879	1.6	1021	1129	4.18
	12	767	910	1.58	1029	1169	4.56
Shock, FFR, GDP, PCE	3	796	853	1.91	1026	1080	3.59
	6	775	876	1.67	1023	1120	3.78
	9	766	909	1.59	1022	1162	3.99
	12	766	952	1.58	1027	1210	4.2
Shock, GDP, PCE, UR	3	761	819	1.51	1010	1063	3.58
	6	748	848	1.39	1009	1106	3.79
	9	743	886	1.37	1019	1158	4.25
	12	743	929	1.4	1032	1215	4.6
FFR, GDP, PCE, UR	3	721	778	1.33	1007	1061	3.53
	6	713	813	1.26	997	1093	3.59
	9	711	854	1.25	995	1135	3.81
	12	714	900	1.28	1000	1182	4.16
Shock, FFR, GDP, PCE, UR	3	723	791	1.34	1009	1074	3.5
	6	716	838	1.28	1000	1118	3.5
	9	717	892	1.29	995	1168	3.64
	12	721	951	1.36	999	1225	3.95
Shock, GDP, PCE, UR, COM	3	760	828	1.64	1011	1076	3.8
	6	751	873	1.55	1015	1133	4.23
	9	749	924	1.55	1029	1202	4.82
	12	745	974	1.58	1045	1271	5.35
FFR, GDP, PCE, UR, COM	3	717	786	1.43	1009	1074	3.75
	6	707	829	1.39	1004	1122	3.95
	9	704	880	1.42	1003	1175	4.16
	12	705	935	1.47	1007	1233	4.55
Shock, FFR, GDP, PCE, UR, COM	3	719	798	1.44	1011	1087	3.72
	6	710	853	1.4	1006	1146	3.86
	9	709	917	1.47	1001	1206	3.94
	12	709	982	1.56	1001	1270	4.24

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

7.2 Price Level Responses

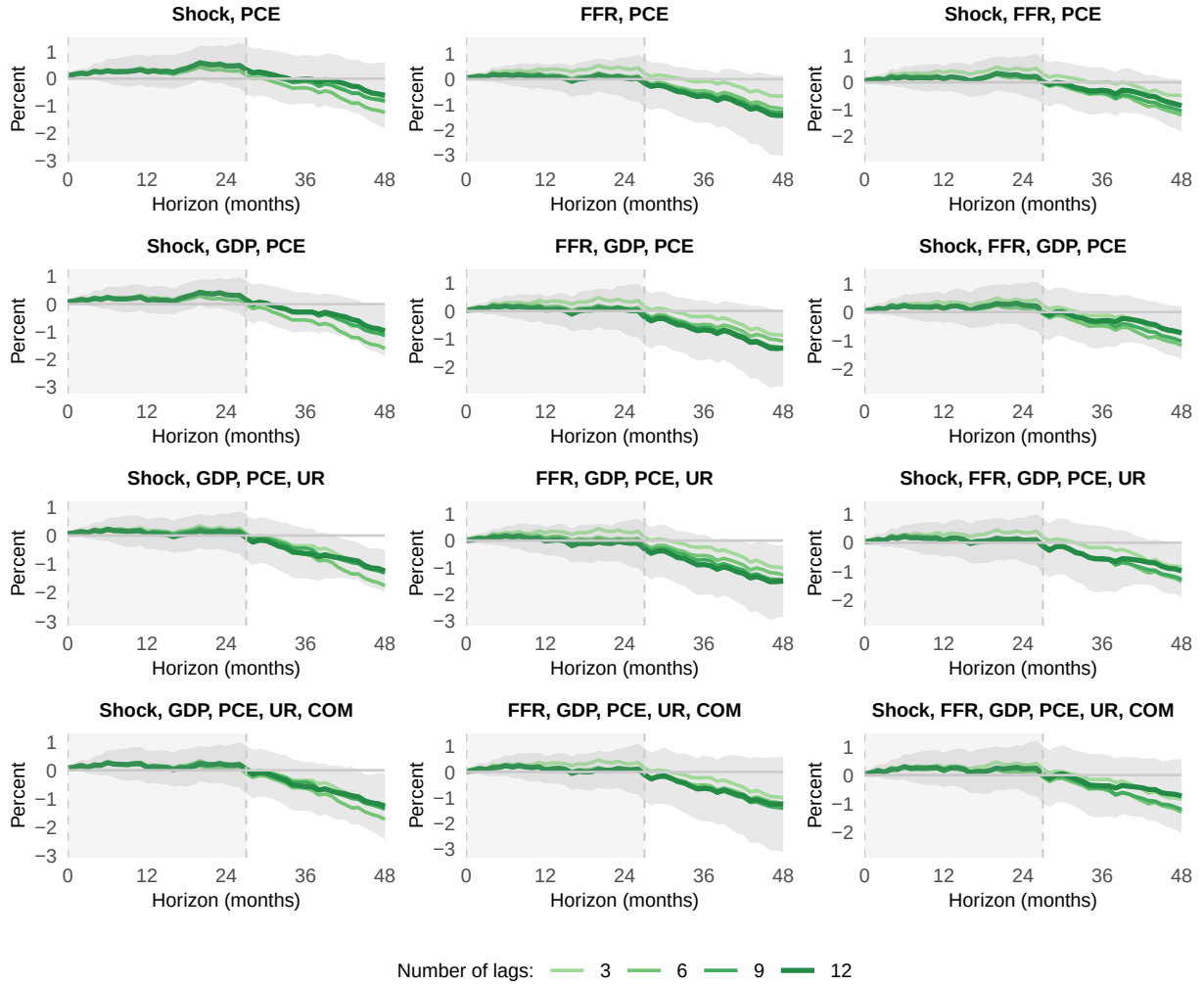
Figure 12 shows the impulse response of the PCE price index from a Romer-Romer monetary policy shock in the levels specification, while Figure 13 shows the responses for the long-differences specification. Table 9 presents the information criteria for the Romer-Romer shock on the PCE price index.

Figure 12: *Response of the PCE price index from a Romer-Romer monetary policy shock in levels*



Notes: The response of the PCE price index to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 27$ mark the expected lag for the price level response, based on prior empirical evidence.

Figure 13: *Response of the PCE price index from a Romer-Romer monetary policy shock in long-differences*



Notes: The response of the PCE price index to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 27$ mark the expected lag for the price level response, based on prior empirical evidence.

Table 9: *Information criteria for a Romer–Romer shock on the PCE price index*

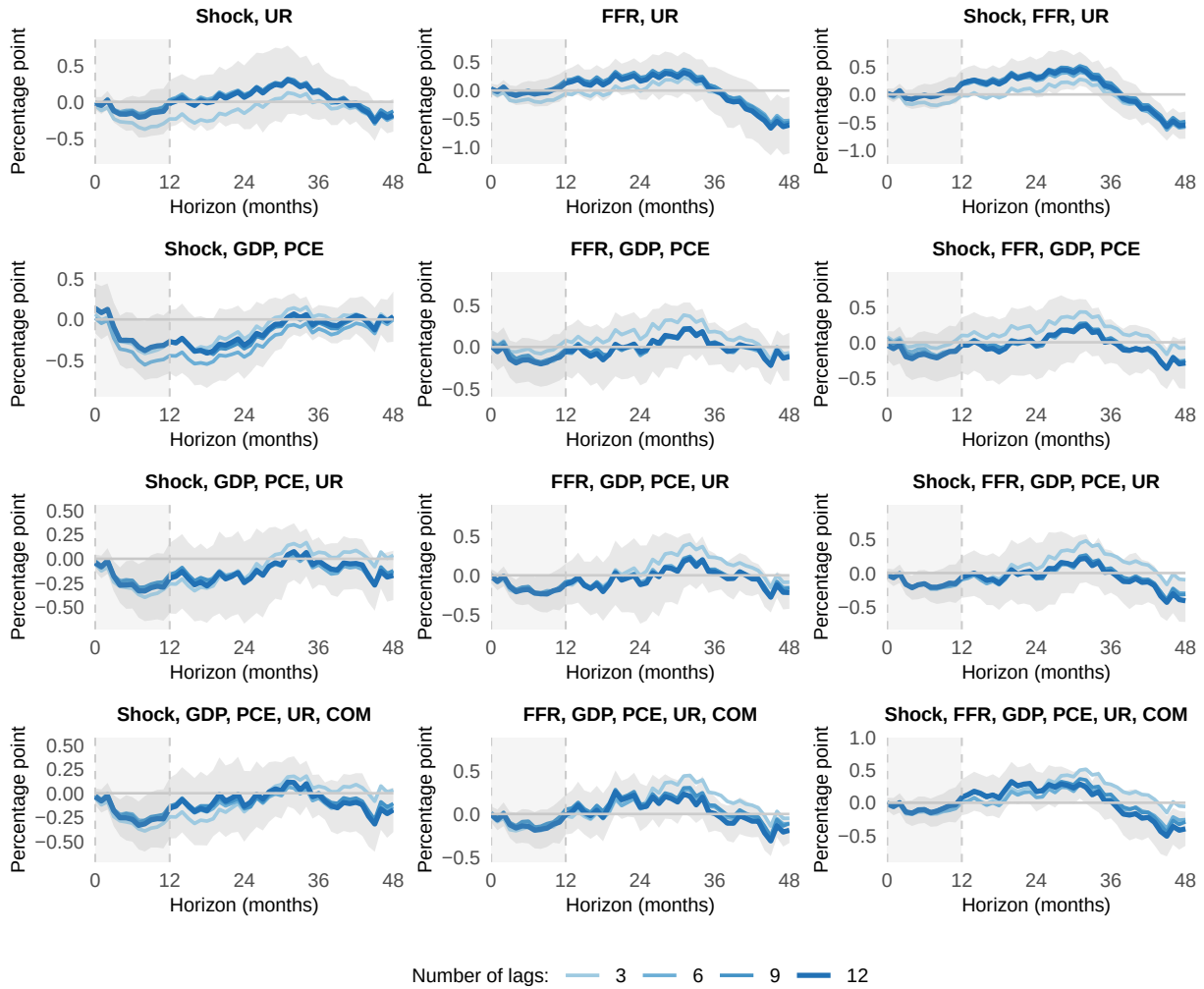
Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, PCE	3	752	788	1.68	892	924	3.28
	6	746	804	1.73	870	924	2.7
	9	748	827	1.81	864	939	2.42
	12	751	852	1.85	869	966	2.38
FFR, PCE	3	732	768	1.62	892	924	3.32
	6	717	775	1.53	872	926	2.91
	9	713	792	1.54	870	945	2.87
	12	715	815	1.61	877	974	3
Shock, FFR, PCE	3	737	783	1.68	897	940	3.41
	6	727	806	1.63	880	955	3
	9	729	840	1.72	878	986	2.87
	12	735	878	1.82	888	1028	2.95
Shock, GDP, PCE	3	725	772	2.03	893	936	3.33
	6	722	801	1.97	874	949	2.79
	9	718	829	1.9	871	979	2.51
	12	708	852	1.8	880	1020	2.47
FFR, GDP, PCE	3	643	690	1.16	893	936	3.33
	6	617	696	0.96	874	950	2.87
	9	597	708	0.84	876	984	2.8
	12	576	720	0.75	887	1027	2.99
Shock, FFR, GDP, PCE	3	645	702	1.15	898	952	3.43
	6	624	724	0.98	883	980	2.97
	9	610	753	0.88	886	1026	2.89
	12	588	775	0.82	900	1083	3.06
Shock, GDP, PCE, UR	3	682	740	1.56	897	951	3.43
	6	681	782	1.52	883	980	2.91
	9	680	824	1.5	885	1024	2.69
	12	670	856	1.44	895	1078	2.75
FFR, GDP, PCE, UR	3	640	697	1.17	895	949	3.36
	6	618	718	0.97	884	981	2.96
	9	597	741	0.85	890	1030	2.93
	12	578	764	0.77	904	1087	3.17
Shock, FFR, GDP, PCE, UR	3	642	711	1.17	900	965	3.44
	6	626	747	1	892	1011	3.04
	9	610	786	0.89	900	1072	3.06
	12	591	820	0.84	914	1140	3.27
Shock, GDP, PCE, UR, COM	3	642	710	1.36	903	967	3.59
	6	626	748	1.24	894	1013	3.19
	9	607	783	1.14	900	1072	3.1
	12	575	805	0.99	914	1140	3.24
FFR, GDP, PCE, UR, COM	3	567	635	0.81	900	965	3.48
	6	537	659	0.69	894	1013	3.18
	9	519	694	0.68	905	1077	3.24
	12	503	732	0.66	922	1148	3.56
Shock, FFR, GDP, PCE, UR, COM	3	570	649	0.82	906	981	3.59
	6	546	690	0.73	903	1043	3.32
	9	534	742	0.76	915	1119	3.49
	12	517	789	0.76	933	1202	3.89

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

7.3 Unemployment Responses

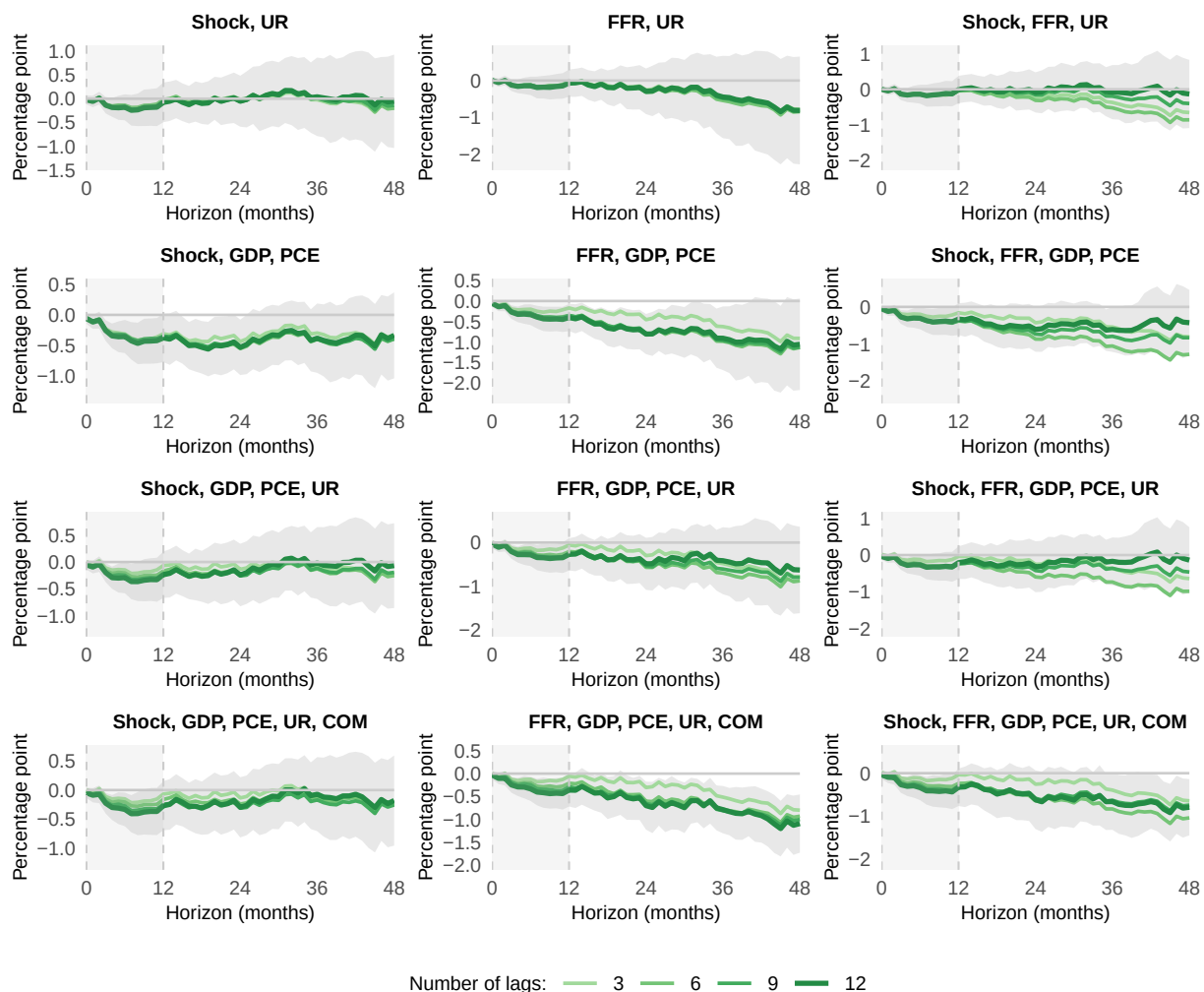
Figure 14 shows the impulse responses of the unemployment rate to the Romer-Romer monetary policy shock, while Figure 15 shows the response in long-differences. Table 10 presents the information criteria for the Romer-Romer shock on unemployment rate.

Figure 14: *Response of unemployment rate from a Romer-Romer monetary policy shock in levels*



Notes: The response of unemployment rate to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 12$ mark the expected lag for the unemployment response, based on prior empirical evidence.

Figure 15: *Response of unemployment rate from a Romer-Romer monetary policy shock in long-differences*



Notes: The response of unemployment rate to a 1 percentage point Romer-Romer monetary policy shock (Romer and Romer, 2004) updated by Wieland and Yang (2020). All specifications use the same sample from October 1983 to December 2007. The shaded bands around the impulse response indicate the 90 percent confidence intervals for the 12-lag specification. The dashed lines from $h = 0$ to $h = 12$ mark the expected lag for the unemployment response, based on prior empirical evidence.

Table 10: *Information criteria for a Romer–Romer shock on unemployment rate*

Controls	Lag	Levels			Long-differences		
		AIC	BIC	CV	AIC	BIC	CV
Shock, UR	3	470	505	0.51	658	690	1
	6	459	516	0.5	662	716	1.1
	9	464	543	0.51	671	747	1.19
	12	471	572	0.56	681	777	1.3
FFR, UR	3	446	482	0.62	652	684	0.99
	6	428	485	0.61	648	702	1.02
	9	429	508	0.61	645	720	1.07
	12	431	531	0.61	642	738	1.15
Shock, FFR, UR	3	449	495	0.63	655	698	0.97
	6	430	509	0.61	653	729	0.99
	9	427	538	0.6	648	756	0.98
	12	424	568	0.6	643	783	1.01
Shock, GDP, PCE	3	531	578	0.68	665	708	1.22
	6	503	582	0.59	657	732	1.43
	9	486	597	0.55	666	773	1.64
	12	478	621	0.55	679	818	1.77
FFR, GDP, PCE	3	496	543	0.67	661	704	1.19
	6	459	538	0.56	652	727	1.34
	9	439	550	0.51	656	763	1.5
	12	429	572	0.49	660	800	1.6
Shock, FFR, GDP, PCE	3	501	558	0.68	664	718	1.21
	6	465	565	0.57	653	750	1.31
	9	445	588	0.52	653	792	1.37
	12	434	621	0.5	652	835	1.4
Shock, GDP, PCE, UR	3	413	471	0.52	651	704	1.22
	6	384	484	0.45	654	751	1.45
	9	381	524	0.44	673	812	1.66
	12	387	573	0.46	687	870	1.79
FFR, GDP, PCE, UR	3	351	409	0.4	647	701	1.17
	6	317	417	0.35	643	739	1.3
	9	316	460	0.35	647	787	1.39
	12	325	511	0.36	649	831	1.43
Shock, FFR, GDP, PCE, UR	3	353	421	0.4	649	714	1.15
	6	323	445	0.35	643	762	1.25
	9	320	496	0.34	645	817	1.28
	12	325	554	0.35	644	870	1.3
Shock, GDP, PCE, UR, COM	3	418	486	0.58	656	720	1.27
	6	389	510	0.5	663	782	1.56
	9	380	555	0.47	684	856	1.84
	12	372	602	0.46	702	928	2.04
FFR, GDP, PCE, UR, COM	3	339	407	0.4	652	716	1.21
	6	296	418	0.35	648	767	1.35
	9	289	465	0.35	651	823	1.43
	12	291	520	0.35	652	877	1.49
Shock, FFR, GDP, PCE, UR, COM	3	341	420	0.41	653	728	1.19
	6	303	447	0.36	647	787	1.28
	9	292	500	0.35	639	843	1.26
	12	285	557	0.35	631	900	1.24

Notes: Each column is colored independently. Within each column, entries are ranked from best to worst fit: green shading indicates the top third of specifications (lowest criterion values), yellow the middle third, and red the bottom third. Lower values indicate better fit.

Appendix B. Data

7.4 Public Access Data

7.4.1 FRED-MD Dataset

The following data series were retrieved from the FRED-MD Database (McCracken and Ng, 2016), 2024/12 vintage (data file: FRED-MD 2024m12.csv). The last data point in the vintage is December 2024. Series ID and description as in the 2024/12 vintage.

FRED-MD 2024/12 Vintage ID	Description	Source
UNRATE	Civilian Unemployment Rate	(McCracken and Ng, 2016)
FEDFUNDS	Effective Federal Funds Rate	(McCracken and Ng, 2016)
PCEPI	Personal Cons. Expend.: Chain Index	(McCracken and Ng, 2016)

Link to dataset: <https://research.stlouisfed.org/econ/mccracken/fred-databases/>.

7.4.2 NBER US Recession Dates

The NBER US Recession Dates were retrieved from St. Louis Fed FRED Data Service (FRED ID: USREC). The series can be downloaded from <https://fred.stlouisfed.org/series/USREC>.

7.5 Data from Other Publications

7.5.1 Romer-Romer MP Shocks

The data for the Romer-Romer MP shocks updated to December 2007 are from Wieland and Yang (2020). The series is available to download at <https://www.openicpsr.org/openicpsr/project/135741/version/V1/view>.

7.5.2 Aruoba-Drechsel MP Shocks

The data for the Aruoba-Drechsel MP shocks are from Aruoba and Drechsel (2024). The series is available to download at <https://econweb.umd.edu/~drechsel/research.html>.

7.5.3 Commodity Price Series

The data for the commodity price index used in the paper were retrieved from Bloomberg (ticker: CRB CMDT Index). This source requires a subscription to the service.

7.5.4 Monthly GDP Series

The data for real GDP interpolated to monthly frequency are from Barrett (2025). The series is available to download at <https://sites.google.com/view/philip-barrett/research>.

Appendix C. Derivation of the Bias in the Long-Differences Specification

This appendix section provides a full derivation of the omitted variable bias in the long-differences local projection specification.

Suppose that the true model for the outcome is

$$y_{t+h} - y_{t-1} = \delta_h + \beta_h s_t + \theta'_h W_t + v_{t+h}, \quad (14)$$

where W_t denotes the controls needed for exogeneity, so that $E[v_{t+h} \mid s_t, W_t] = 0$.

The econometrician does not observe W_t and instead estimates

$$y_{t+h} - y_{t-1} = a_h + b_h s_t + c'_h X_t^D + \varepsilon_{t+h}, \quad (15)$$

where X_t^D is a vector of some lagged macroeconomic variables in first-differences. Let $\hat{\beta}_h$ denote the OLS coefficient on s_t in (15). The goal is to show that the associated OLS estimator is biased whenever X_t^D fails to fully capture the information in W_t .

Stack the sample observations into $\mathbf{s}$, $\mathbf{W}$, $\mathbf{X}^D$, and $\mathbf{v}_h$, and define $Z^D = [\mathbf{1}_T, \mathbf{X}^D]$, which combines the intercept and the first-differenced controls. The residual maker matrix associated with Z^D is

$$M_D = I_T - Z^D (Z^{D'} Z^D)^{-1} Z^{D'},$$

and the residualized shock is $s_t^* = [M_D \mathbf{s}]_t$. By construction, s_t^* is orthogonal to every column of Z^D , so in particular

$$\sum_t s_t^* = 0, \quad \sum_t s_t^* X_t^D = 0.$$

By the Frisch–Waugh–Lovell theorem, the OLS estimator of β_h in (15) can be written as

$$\hat{\beta}_h = \frac{\sum_t s_t^* y_{t+h}^*}{\sum_t (s_t^*)^2},$$

where $y_{t+h}^* = [M_D \mathbf{y}_h]_t$ is the projection residual of the long-difference outcome $y_{t+h} - y_{t-1}$ against Z^D . Applying M_D to both sides of (14) annihilates the intercept and gives:

$$y_{t+h}^* = \beta_h s_t^* + \theta_h' W_t^* + v_{t+h}^*,$$

where $W_t^* = [M_D \mathbf{W}]_t$ and $v_{t+h}^* = [M_D \mathbf{v}_h]_t$. Substituting,

$$\begin{aligned} \hat{\beta}_h &= \frac{\sum_t s_t^* (\beta_h s_t^* + \theta_h' W_t^* + v_{t+h}^*)}{\sum_t (s_t^*)^2} \\ &= \beta_h + \theta_h' \frac{\sum_t s_t^* W_t^*}{\sum_t (s_t^*)^2} + \frac{\sum_t s_t^* v_{t+h}^*}{\sum_t (s_t^*)^2}, \end{aligned}$$

which gives

$$\hat{\beta}_h = \beta_h + \underbrace{\theta_h' \frac{\sum_t s_t^* W_t^*}{\sum_t (s_t^*)^2}}_{\text{omitted-variable bias}} + \underbrace{\frac{\sum_t s_t^* v_{t+h}^*}{\sum_t (s_t^*)^2}}_{o_p(1)}. \quad (16)$$

It remains to show that the third term in (16) is $o_p(1)$.

$$\begin{aligned} \frac{\sum_t s_t^* v_{t+h}^*}{\sum_t (s_t^*)^2} &= \frac{\sum_t s_t^* (v_{t+h} - \psi_0 - \psi_1' X_t^D)}{\sum_t (s_t^*)^2} && (v_{t+h}^* = v_{t+h} - \psi_0 - \psi_1' X_t^D) \\ &= \frac{\sum_t s_t^* v_{t+h} - \sum_t s_t^* (\psi_0 + \psi_1' X_t^D)}{\sum_t (s_t^*)^2} \\ &= \frac{\sum_t s_t^* v_{t+h}}{\sum_t (s_t^*)^2} && (\sum_t s_t^* = 0, \sum_t s_t^* X_t^D = 0) \\ &= \frac{(1/T) \sum_t s_t^* v_{t+h}}{(1/T) \sum_t (s_t^*)^2} \\ &\xrightarrow{p} \frac{E[s_t^* v_{t+h}]}{E[(s_t^*)^2]} && (\text{LLN and CMT}) \\ &= \frac{E[E[s_t^* v_{t+h} \mid s_t, W_t, X_t^D]]}{E[(s_t^*)^2]} && (\text{Law of Iterated Expectations}) \\ &= \frac{E[s_t^* E[v_{t+h} \mid s_t, W_t, X_t^D]]}{E[(s_t^*)^2]} && (s_t^* \text{ is } (s_t, X_t^D)\text{-measurable}) \\ &= 0. && (E[v_{t+h} \mid s_t, W_t, X_t^D] = 0) \end{aligned}$$