Differences from Differencing: Should Local Projections with Observed Shocks be Estimated in Levels or Differences?

Jeremy Piger[*] Thomas Stockwell[†]

This Draft: June 1, 2023

Abstract: We show there are substantial small sample gains from estimating local projections (LPs) in a cumulated differences specification vs. a specification in levels when the impulse response of interest is to an externally identified ("observed") shock. The cumulated differences specification reduces estimation bias and improves confidence interval coverage over LPs specified in levels, with the improvement increasing for more persistent processes, at longer horizons, and for smaller sample sizes. For processes with persistence ranging from zero to moderately high, LPs estimated in cumulated differences display consistently negligible bias, while LPs estimated in levels display increasing bias as persistence rises. For near unit root and unit root processes, LPs estimated in cumulated differences are biased, but the improvement in bias over the levels specification is largest in these cases. We demonstrate these results using simulation evidence as well as analytic results for the example case of an autoregression. Overall, the cumulated differences specification appears to be an effective approach to reduce bias and improve the accuracy of confidence intervals in LPs estimated with observed shocks, regardless of the true integration properties of the data.

Keywords: impulse response function, vector autoregression, monetary policy

JEL Classifications: C22, E32, E52

[*]Department of Economics, University of Oregon, (jpiger@uoregon.edu)

[†]Department of Economics, Sykes College of Business, University of Tampa, (tstockwell@ut.edu)

1 Introduction

Following Jordá (2005), local projections (LPs) have become a popular approach to estimate impulse response functions. In the empirical macroeconomics literature specifically, LPs are now widely viewed as a viable alternative to the usual impulse response functions estimated via vector autoregressive (VAR) models. LPs offer some well publicized potential advantages over VAR models. First, they are simple to estimate and draw inference on, since LPs can be implemented via univariate linear regressions. Second, since LPs place less structure on the assumed data generating process, they are in principle more robust to misspecification than VAR models. Third, LPs can more easily accommodate state-dependent and non-linear specifications, making them especially popular in these applications.[1] As local projections have increased in popularity, there has been a growing theoretical literature studying the asymptotic properties of LPs and their relation to VAR models.[2]

It is common to find differences in the literature in the way the response variable is specified in LP regressions, with some studies specifying the LP regression in (log) levels, and others using a cumulated differences specification. In the literature constructing impulse response functions from VARs, the levels specification has increasingly been considered the safer route to estimate impulse response functions when the true integration properties of the data is unknown (Ramey (2016)). The argument typically proceeds as follows: Estimation in differences can provide a reduction in bias and improved efficiency if the system contains unit roots. However, if the process is instead stationary, differencing will introduce non-invertibilities and may hide long-run relationships that create issues for recovering structural shocks of interest. At the same time, estimation in levels retains long-run relationships and does not introduce non-invertible disturbances, while techniques have been developed for near-unit root or unit-root processes to provide appropriate inference (Gospodinov et al.

[1] See, e.g., Ramey and Zubairy (2018), Auerbach and Gorodnichenko (2013), and Tenreyro and Thwaites (2016).

[2] Examples include Olea and Plagbørg-Moller (2021), Plagbørg-Moller and Wolf (2021), Gonçalves et al. (2023) and Xu (2023)

(2013)). Further, in the LP literature, recent results from Olea and Plagbørg-Moller (2021) demonstrate that standard estimators applied to lag-augmented LPs, such as OLS with HAC standard errors, have asymptotic normal distributions that are invariant to the underlying persistence properties of the data, including the unit root case. This result seems to provide a justification for use of the levels specification in LPs.

At the same time, there is a growing literature that shows standard OLS estimates of impulse response functions via LPs are biased and produce incorrect confidence intervals in finite samples, particularly in the relatively small sample sizes used in the empirical macroeconomics literature. Using simulations, Kilian and Kim (2011) find asymptotic confidence intervals from LPs are less accurate than bias-adjusted VAR bootstrap confidence intervals. Herbst and Johannsen (2022) document that LPs are in practice often used with very small samples in the time dimension, and that point estimates of impulse response functions from LPs are severely biased on these sample sizes. This is especially true when the process under consideration is persistent, which is the case with most macroeconomic series of interest. Building on these results, a small number of papers have presented attempts to reduce finite-sample bias and improve the accuracy of confidence intervals in LP regressions. Herbst and Johannsen (2022) use an approximate bias function to characterize and partially account for the bias in the LP regression. Olea and Plagbørg-Moller (2021) consider lag-augmented LPs, which use lags of the regressors as controls. Using simulations, they find that bootstrapped lag-augmented LPs generate improved confidence interval accuracy in finite samples.

These simulation studies finding finite sample bias in LP regressions have focused on LPs specified in levels, and have not considered the performance of LPs specified in cumulated differences. It is unclear whether the lessons from the VAR literature regarding the relative merits of estimating in levels vs. differences apply to the LP setting, especially in the common case where local projections are estimated with an externally identified shock of interest (Stock and Watson (2018)). Also, while the results of Olea and Plagbørg-Moller (2021) provide a compelling asymptotic justification for the levels specification, the demonstrated poor performance of the levels LP specification in empirically relevant sample sizes leaves open the possibility that differenced specifications may provide improvements. To our knowledge, the finite sample performance of LPs estimated in levels vs. differences has not been studied previously.

We fill this gap by conducting a simulation study to evaluate the finite sample performance of LPs specified in levels vs. differences. Consistent with Herbst and Johannsen (2022), we focus on the empirically relevant case where we have an externally identified, observed, shock available for which we wish to estimate the impulse response function via LPs. We begin with the example of an AR(1) with i.i.d. disturbances, and demonstrate analytically that differencing should substantially reduce a particular source of small sample bias that exists in levels LPs when the true data generating process is stationary, but persistent. Then, using a wide variety of data generating processes for empirically relevant sample sizes, we show using simulations that the difference specification can substantially reduce bias and improve confidence interval accuracy over LP regressions specified in levels for persistent processes, regardless of whether the true process contains a unit root. Further, even for data that is less persistent, the differences specification does not demonstrate any apparent disadvantages over the levels regression. Overall, the differences specification appears to be an effective approach to reduce bias and improve the accuracy of confidence intervals in LP estimation of impulse response functions where there is an observed shock, regardless of the true integration properties of the data. As noted above, this stands in contrast to the conclusions of an existing literature using structural VARs with internally identified shocks, such as Gospodinov et al. (2013).

As an application, we consider the question of the effects of U.S. monetary policy shocks occurring during recessions. There is a significant literature that investigates whether U.S. monetary policy shocks have different effects on output and inflation when occurring during recessions vs. expansions, with a number of recent papers using LPs to tackle this question.[3]

[3]See, for example, Tenreyro and Thwaites (2016).

A criticism of this literature is the small sample size available to estimate the effects of policy shocks occurring during recessions, since recessions account for only approximately 10% of the post-World War II sample. Such a small sample size is a case where we would expect to see differences between levels and differences specifications of LPs. We do find significant differences in this setting, with the output effects of monetary policy shocks occurring during recessions having larger and more persistent effects when estimated in differences than when estimated in levels.

The rest of this paper proceeds as follows: Section 2 reviews the local projection approach to estimate impulse response functions with externally identified, observed, shocks and discusses standard inference techniques used in the literature. Section 3 uses the stylized example of an AR(1) data generating process to demonstrate the intuition for some of the improvements in estimation bias that come from the differences specification. We then move to our simulation study in Section 4 that considers estimation bias and confidence interval accuracy for a variety of data generating processes and practical estimation considerations. Section 5 shows the results of the application to estimation of the state-dependent effects of monetary policy shocks. Section 6 concludes.

2 Local Projections

Suppose one has an observed shock of interest, labeled ε_t, and a response variable of interest, labeled y_t. We wish to measure the impulse response at horizon h, up to some maximum horizon H:

$$\beta_h = \frac{\partial y_{t+h}}{\partial \varepsilon_t}$$

A local projection to estimate β_h is simply a direct multi-step ahead prediction:

$$y_{t+h} = \beta^h \varepsilon_t + \rho_1^h y_{t-1} + \rho_2^h y_{t-2} + \cdots + \rho_p^h y_{t-p} + \left(\gamma^h\right)' X_t + v_{t+h} \tag{1}$$

In most applications of local projections, lagged values of the response variable appear as controls, and we have explicitly allowed for p lags of the response variable in equation (1).[4] Additional controls can appear in the vector X_t, and usually include deterministic terms, such as a constant or deterministic time trends. In some applications, lags of variables other than the response variable are also included. Since the left hand side is specified in the levels of the response variable, we refer to equation (1) as the "levels" specification.[5]

We can alternatively estimate β^h using a cumulated differences specification. To begin, consider a local projection where the response variable is the first difference of y_{t+h}:

$$\Delta y_{t+h} = \widetilde{\beta}^h \varepsilon_t + \widetilde{\rho}_1^h \Delta y_{t-1} + \widetilde{\rho}_2^h \Delta y_{t-2} + \cdots + \widetilde{\rho}_p^h \Delta y_{t-p} + (\widetilde{\gamma}^h)' \widetilde{X}_t + \widetilde{v}_{t+h} \tag{2}$$

where $\widetilde{\beta}^h$ is the impulse response of the first difference of y_{t+h} to the shock ε_t. We can then recover β^h as:

$$\beta^h = \sum_{i=0}^{h} \widetilde{\beta}^i \tag{3}$$

One could estimate β^h by first estimating equation (2) and then forming the h-period sum in equation (3). However, as pointed out by Stock and Watson (2018), we can instead first sum equation (2), providing the following equation to estimate β^h directly:

$$y_{t+h} - y_{t-1} = \beta^h \varepsilon_t + \theta_1^h \Delta y_{t-1} + \theta_2^h \Delta y_{t-2} + \cdots + \theta_p^h \Delta y_{t-p} + (\alpha^h)' X_t^D + u_{t+h} \tag{4}$$

We refer to equation (4) as the "differences" specification, though the left hand side of this equation is in terms of the h-period difference of y_{t+h}, rather than the first difference.

While the impulse responses at alternative horizons could be estimated by treating the H equations as a seemingly unrelated regression that is estimated jointly, it is common in

[4] Olea and Plagbørg-Moller (2021) refer to LPs with lagged response variables as "lag-augmented" LPs. All of the LPs we consider in this paper will be lag-augmented in the sense of Olea and Plagbørg-Moller (2021).

[5] As discussed in Stock and Watson (2018), in most applications ε_t is likely better considered as an instrument for the true shock of interest rather than the shock itself. However, to stay consistent with a signifiant existing literature, here we follow the common specification of including ε_t in the local projection as the observed shock.

the applied LP literature to estimate via equation by equation OLS. Also, as discussed in Jordá (2005), the disturbance terms in equations (1) and (2) are serially correlated and follow a moving average (MA) process. Because of this, much of the literature makes use of robust standard errors to compute confidence intervals on the impulse response β^h, with the Newey-West methodology being a popular choice. The disturbance term in equation (4) is further complicated by the summation of errors from equation (2). In the remainder of this paper we will evaluate the performance of equation by equation OLS estimation of the LP in both the levels and differences specification, as well as the performance of the the Newey-West methodology for computing standard errors.

3 An Illustrative Example Based on an AR(1) Model

In this section we consider a specific data generating process (DGP), a stationary autoregressive model of order 1 (AR(1)) with i.i.d. disturbances. This DGP will allow us to illustrate analytical results that will aid our intuition regarding the relative effectiveness of estimating LPs with observed shocks via the levels vs. differences specification. Specifically, assume the true DGP is:

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t,$$

where ε_t is independent and identically distributed with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$. We begin by assuming that $|\phi| < 1$, thereby focusing attention on the stationary case where the benefits of differencing are *a priori* dubious. Later in this section we will consider the unit root case where $\phi = 1$.

Suppose that ε_t is the observed shock of interest in the LP.[6] The correctly specified levels specification for the LP is then:

$$y_{t+h} = c_h^L + \beta^h \varepsilon_t + \rho_1^h y_{t-1} + v_{t+h} \tag{5}$$

[6] Of course, if ε_t is observed, the parameter ϕ can be trivially solved from the AR(1) equation and there is no need for estimation. In practice however, the econometrician does not know the true DGP.

where $\beta^h = \phi^h$, $\rho_1^h = \phi^{h+1}$ and $v_{t+h} = \varepsilon_{t+h} + \phi\varepsilon_{t+h-1} + \phi^2\varepsilon_{t+h-2} + \cdots + \phi^{h-2}\varepsilon_{t+2} + \phi^{h-1}\varepsilon_{t+1}$. At first look, the levels LP appears well specified since the regressor ε_t is independent of each of the values of $\varepsilon_{t+j}, j > 0$ that sum to form the regression disturbance v_{t+h}. However, despite this independence, the OLS estimate of β^h from 5 will be biased because of an expected non-zero sample correlation between ε_t and v_{t+h} in finite samples. This non-zero expected sample correlation arises because of the interaction of the sample means of ε_t and v_{t+h} in the sample covariance formula. Specifically, the Appendix shows that the expected sample covariance, $S_{\varepsilon_t, v_{t+h}}$ is:

$$\begin{aligned} E\left(S_{\varepsilon_t, v_{t+h}}\right) &= -\sum_{i=0}^{h-1} \phi^i E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h-i}\right) \\ &= -\frac{\sigma^2}{(T-h)^2}\left[\sum_{i=0}^{h-1} \phi^i (T - 2h + i)\right] \end{aligned} \tag{6}$$

Equation 6 provides several elements of intuition regarding the expected bias in the OLS estimate of β^h. First, the size of the expected covariance between ε_t and v_{t+h} depends on the value of $\beta^i = \phi^i$ for $i = 0, \ldots h-1$. In other words, the expected covariance depends on the value of the true IRF at all horizons up to horizon $h-1$.[7] The more persistent the IRF, the larger will be these terms in absolute value, which increases the covariance in absolute value. Second, the expected covariance will grow in absolute value with the horizon h. Third, the sample size influences the size of the expected covariance. As T grows, the denominator grows with respect to the numerator and shrinks the size of the covariance.

We now consider how this source of potential bias may be mitigated by estimating the differences LP. The correctly specified differences LP for the AR(1) DGP is:

$$y_{t+h} - y_{t-1} = c_h^D + \beta^h \varepsilon_t + \theta_1^h \Delta y_{t-1} + \cdots + \theta_h^h \Delta y_{t-h} + u_{t+h},$$

[7] This result is consistent with Herbst and Johannsen (2022), who use a higher order expansion to characterize the small-sample bias in the OLS estimator of the levels specification of the LP. They show that the bias in the LP estimator at horizon h is a function of the true (population) impulse responses at other horizons.

where again, $\beta^h = \phi^h$ and:

$$u_{t+h} = (\varepsilon_{t+h} - \varepsilon_{t-1}) + \phi(\varepsilon_{t+h-1} - \varepsilon_{t-2}) + \phi^2(\varepsilon_{t+h-2} - \varepsilon_{t-3})$$
$$+ \cdots + \phi^{h-1}(\varepsilon_{t+1} - \varepsilon_{t-h}) - \phi^h \varepsilon_{t-h-1}$$

As shown in the Appendix, the expected sample covariance between ε_t and u_{t+h} is:

$$\begin{aligned} E(S_{\varepsilon_t, u_{t+h}}) &= \phi^h E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t-h-1}\right) - \sum_{i=0}^{h-1} \phi^i E\left(\bar{\varepsilon}_t(\bar{\varepsilon}_{t+h-i} - \bar{\varepsilon}_{t-1-i})\right) \\ &= \frac{\sigma^2}{(T-2h-1)^2}\left[\phi^h[T-3h-2] - \sum_{i=0}^{h-1} \phi^i\left[1-h+2i\right]\right] \end{aligned} \tag{7}$$

The expected sample covariance from the differences specification in 7 will in general be much smaller than that from the levels specification in 6. In other words, the observed shock, ε_t will display less expected correlation with the regression disturbance in the differences LP than the levels LP. Figure 1 displays the expected sample covariance from 6 and 7 for the case where $T = 100$, $\sigma^2 = 1$, and for three values of persistence, $\phi = \{0.7, 0.9, 0.95\}$. The figure shows that the expected sample covariance between ε_t and the levels LP regression disturbance is increasing in absolute value in both horizon and persistence, whereas this is not the case for the differences LP regression. Also, the expected sample covariance is larger in absolute value for the levels regression for all horizons beyond $h = 1$. Figure 2 shows the same set of experiments for the case where $T = 200$. Here the expected sample covariance term is lower for the levels LP regression than when $T = 100$, but displays the same pattern and remains significantly larger than that for the differences specification.

The source of the reduction in the expected sample covariance term can be seen through comparison of equations 6 and 7. In equation 6, each of the expectations $E(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h-i})$, $i = 0, 1, \ldots, h-1$, creates $[T-h+i]$ non-zero terms due to overlap between the samples used to calculate $\bar{\varepsilon}_t$ and $\bar{\varepsilon}_{t+h-i}$. By contrast, in 7, each of the expectations $E\left(\bar{\varepsilon}_t(\bar{\varepsilon}_{t+h-i} - \bar{\varepsilon}_{t-1-i})\right)$, $i = 0, 1, \ldots, h-1$, creates only $(1-h+2i) << (T-h+i)$ non-zero terms, with this reduction due to cancelation of terms caused by the differencing in the expectation. In the end, regardless of the value of h, equation 7 includes only a single expectation that does not include such a difference, that being $\phi^h E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t-h-1}\right)$. By contrast, equation 6 has h such terms. As such, the reduction in the expected sample covariance will be larger for larger h. Also, since these terms in equation 6 are scaled by ϕ^i, $i = 0, 1, \ldots, h-1$, the reduction in the expected sample covariance seen in equation 7 will be larger for higher values of ϕ.

The discussion above has focused on the case of the stationary AR model. In the unit root case, we would not expect to see a mitigation in bias of the type discussed above from use of the differences specification. To see this, note that in the case where $\phi = 1$, the correct levels specification is:

$$y_{t+h} = c_h^L + \beta^h \varepsilon_t + \rho_1^h y_{t-1} + v_{t+h}$$

where $\beta^h = 1$, $\rho_1^h = 1$ and $v_{t+h} = \varepsilon_{t+h} + \varepsilon_{t+h-1} + \varepsilon_{t+h-2} + \cdots + \varepsilon_{t+2} + \varepsilon_{t+1}$. When $\phi = 1$ the correct differences specification is:

$$y_{t+h} - y_{t-1} = c_h^D + \beta^h \varepsilon_t + u_{t+h}$$

where $\beta^h = 1$, $c_h^D = c_h^L$, and $u_{t+h} = v_{t+h} = \varepsilon_{t+h} + \varepsilon_{t+h-1} + \varepsilon_{t+h-2} + \cdots + \varepsilon_{t+2} + \varepsilon_{t+1}$. Thus, in the unit root case, the regression disturbance is the same for the levels vs. differences specification and thus there is no difference in the finite sample expected correlation between ε_t and the regression disturbance term from using one specification vs. the other. With that being said, we would still expect better finite sample performance from the differences specification in this case, since it correctly imposes the restriction $\rho_1^h = 1$. As we will see in the simulations below, the gains from imposing this restriction are very large in practice.

The results in this section suggest that when the true data generating process is a stationary AR(1) with i.i.d. disturbances, and the AR(1) disturbance is the observed shock of interest, the differences LP should generate a reduction in bias for the impulse response estimate vs. the levels estimate. This reduction in bias comes through a mitigation of the correlation between the observed shock of interest and the LP regression disturbance in the differences vs. the levels specification. In the unit root case, this mitigation disappears. However, we would still expect improved performance from the differences specification in this case, as it correctly enforces restrictions imposed by the integration properties of the DGP. With this illustrative example as motivation, in the next section we conduct simulation experiments to investigate the relative performance of the differences LP vs. the levels LP across a range of data generating processes and persistence levels.

4 Simulation Evidence

In this section we turn to results of a simulation study using a variety of different data generating processes (DGP) to evaluate the performance of the levels (equation (1)) and differences (equation (4)) LP specifications. For each of the DGPs considered, we assume that the true DGP is not known. However, we assume that the shock of interest, labeled ε_t in all cases, is externally identified and available. We will consider both univariate and multivariate DGPs.

We set the control variables in equations (1) and (4) as follows: We include p_L lags of the level of y_t in the levels specification and p_D lags of the first difference of y_t in the differences specification. For the levels specification, X_t includes a constant and linear time trend for univariate data generating processes, and additionally contains p_L lags of the level of additional endogenous variables beyond y_t for multivariate data generating processes.[8] For the differences specification, X_t^D contains a constant for univariate DGPs, and additionally contains p_D lags of the first difference of additional endogenous variables beyond y_t for multivariate DGPs. When estimating each version of the LP models on the simulated data, we conduct data-based lag selection to select p_L and p_D via a test-down procedure. Specifically, a p_{max} is selected and then a test statistic is formed for the coefficient on the lagged variable

[8]Our qualitative results are robust to the exclusion of the deterministic trend in the levels specification in cases where the true DGP does not include a deterministic trend.

corresponding to p_{max} using Newey-West standard errors. If this test statistic is greater than two, then the number of lags is set to p_{max}. Otherwise, p_{max} is lowered by one and the process is repeated. We set the initial value of p_{max} to equal 8.

For each DGP the results are based on 1000 simulations. We consider two sample sizes, $T = 100$ and $T = 200$, corresponding to 25 and 50 years of quarterly data, which are typical sample sizes in studies of U.S. macroeconomic data.[9] We assess the accuracy of both the OLS point estimates and Newey-West coverage intervals for impulse responses at horizons up to and including a maximum horizon of $H = 20$. In constructing the Newey-West standard errors the maximum autocorrelation lag is set to $H + 1$ following Jordá (2005).

4.1 Autoregressive Model

The first DGP considered is a simple Gaussian autoregressive model of order 1 (AR(1)):

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim \text{i.i.d. } N(0, 1).$$

We explore three different calibrations for this model, which differ in their level of persistence. The first specification features a process that is persistent, but clearly stationary in that unit root tests will have very high power to detect the null of stationarity ($\phi = 0.70$), the second is a very persistent, though still stationary process ($\phi = 0.95$), while the third is a unit root process ($\phi = 1$). In all cases, we set the intercept $\alpha = 0$.[10]

Figure 3 shows the results of level and differences specification LPs applied to estimate the impulse response for data generated from the AR(1) model where the sample size is $T = 100$. For each value of the autoregressive parameter considered, the figure contains

[9] Herbst and Johannsen (2022) survey a large number of recent empirical papers utilizing LPs and find that the median value of T across these studies is 95.

[10] In this DGP, the shock of interest for which we are estimating impulse responses, ε_t, is the only source of stochastic variation. We have also considered cases where an additional source of stochastic variation is added to the process as in Herbst and Johannsen (2022). Results for this case are similar to those reported here.

two sets of results. The left panel shows the true impulse response function (solid line) and the average estimated impulse response function over the simulations by both the differences (dash-circle line) and levels (dashed line) specification. The right panel shows the proportion of simulations where the true value of β^h is contained inside of a 90% confidence interval constructed via the differences specification (dash-circle line) and levels specification (dashed line).

The figures provide a striking conclusion - for all three persistence levels for the AR(1) model, the differences specification has less bias and more accurate coverage intervals than the levels specification. As the persistence of the system increases, the better the performance of the differences specification becomes relative to the levels specification. Also, the relative improvement from the differences specification increases as the horizon of the impulse response function increases. It is worth emphasizing that the improvement in the differences specification is still visible even with a process that is clearly stationary.

With this general conclusion in place, we turn to the results in more detail. For the two stationary cases, the differences specification is approximately unbiased, with coverage intervals that are somewhat undersized (between 80% and 90% for all horizons). The levels specification performs reasonably well in the $\phi = 0.7$ case, though it still displays noticeable downward bias and less accurate coverage intervals than the differences specification. When $\phi = 0.95$, the performance of the levels specification deteriorates significantly, with estimates displaying very high levels of bias and coverage intervals that are far below their stated levels. These inaccuracies become larger as the horizon of the impulse response increases. Finally, in the unit root case, there is some bias introduced in the differences specification, and coverage intervals fall to between 60% and 80%. However, the differences specification vastly outperforms the levels specification in this case. Indeed, the levels specification in the unit root case has abysmal performance, with estimated impulse responses at the longest horizons that are less than half of their true value and with 90% Newey-West coverage intervals that are around 20%.

It is useful to discuss the relative performance of the levels vs. differences specification for the AR(1) model in terms of the analytic results from Section 3. From those results, in the stationary case, the levels specification should generate biased estimates of β^h in finite samples due to an expected small sample correlation between ε_t and the levels regression disturbance. We would also expect this source of bias to be substantially mitigated in the differences specification. Both of these results are born out in the results presented in Figure 3. Section 3 also demonstrated that this source of bias mitigation from the differences specification should be eliminated in the unit root case, which is consistent with the biased IRF estimates we see for the differences specification in Figure 3 when $\phi = 1$. However, in this case, the relative performance of the differences specification vs. the levels specification in mitigating overall bias is strongest. This is likely due to the correct restriction imposed by the differences specification in the unit root case, which eliminates the typical finite sample biases associated with dynamic regressions estimated in levels when there are unit roots. Overall, these results suggest that the differences specification has advantages over the levels specification regardless of the integration properties of the data.

Figure 4 shows the results when the sample size is increased to $T = 200$. These results are qualitatively similar to the $T = 100$ case. As would be expected, the performance of the levels specification improves, in terms of both bias and coverage. However, the differences specification maintains a distinct performance advantage in all cases considered. Also, it is worth noting that for the stationary cases, the differences specification now appears to have very close to correct confidence interval coverage, whereas it was somewhat undersized in the $T = 100$ case.

4.2 Alternative Univariate Models

4.2.1 ARMA(1,1) Model

In this section, we explore the performance of the levels and differences specifications for univariate DGPs beyond the AR(1) case. We begin with an ARMA(1,1) model:

$$y_t = \alpha + \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim \text{i.i.d. } N(0,1)$$

We again explore three different calibrations for this model, a clearly stationary process ($\phi = 0.70$), a very persistent, but still stationary, process ($\phi = 0.95$), and a unit root process ($\phi = 1.0$). In all calibrations, $\alpha = 0$ and $\theta = -0.5$.

Figure 5 shows the results of the simulations for the ARMA(1,1) model when $T = 100$, which are very similar to those for the AR(1) model. In particular, the levels specification displays significant estimation bias for the impulse response functions and very inaccurate coverage intervals, with the performance of the levels specification deteriorating as both the persistence of the process and the horizon of the impulse response increases. The differences specification performs much better than the levels specification at every level of persistence considered. In absolute terms, the differences specification is approximately unbiased and has somewhat undersized coverage intervals at all horizons for the stationary calibrations. In the unit root case, the differences specification again displays some bias and coverage intervals that fall further below their nominal level, but still displays large improvements over the levels specification in this case.

Figure 6 shows the results when the sample size is increased to $T = 200$. Again, these results are qualitatively similar to the $T = 100$ case, with the performance of the levels specification improving, but the differences specification maintaining a clear performance advantage. Also, it is again the case that for the stationary cases, the differences specification has very close to correct confidence interval coverage in the $T = 200$ case.

4.2.2 Trend Stationary Unobserved-Components Model

Next we consider a Trend-Stationary Unobserved-Components (UC) model:

$$y_t = T_t + C_t$$

$$T_t = \mu + T_{t-1}$$

$$C_t = \phi_1 C_{t-1} + \phi_2 C_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim \text{i.i.d. } N\left(0, \sigma^2\right)$$

We calibrate the model based on maximum likelihood estimation of this trend-stationary UC model on log quarterly U.S. GDP, measured from 1969:Q1 to 2007:Q4. This estimation produced the following calibration:

$$\mu = 0.77; \phi_1 = 1.22; \phi_2 = -0.3; \sigma = 0.76$$

Figure 7 contains the results of the simulations based on the trend stationary UC model, and shows again that the differences specification outperforms the levels specification in all aspects considered. Once again, the impulse response estimates from the differences specification exhibits very little bias over the entire horizon while the levels specification has a large downward bias. Also, the differences specification produces confidence intervals with true coverage much closer to the nominal coverage. It is again notable that the differences specification provides such large improvements despite the fact that the underlying process is stationary. Figure 8 shows the results when $T = 200$, which are again qualitatively similar to the $T = 100$ case.

4.2.3 Stochastic Trend Unobserved-Components Model

The final univariate model considered is the Stochastic Trend UC Model detailed below:

$$y_t = T_t + C_t$$

$$T_t = \mu + T_{t-1} + v_t$$

$$C_t = \phi_1 C_{t-1} + \phi_2 C_{t-2} + \varepsilon_t$$

$$v_t \sim WN\left(0, \gamma^2\right)$$

$$\varepsilon_t \sim N\left(0, \sigma^2\right)$$

We again calibrate the model based on maximum likelihood estimation of this stochastic trend unobserved components model on log quarterly U.S. GDP, measured from 1969:Q1 to 2007:Q4. This estimation produced the following calibration:

$$\mu = 0.77; \phi_1 = 1.55; \phi_2 = -0.6; \gamma = 0.5783; \sigma = 0.443$$

Figure 9 contains the results of the simulations based on the stochastic trend UC model, which again show the differences specification outperforming the levels specification. The estimates of the IRF from the differences specification display very little bias at any horizon, while the levels specification produces estimates that display significant bias over most of the horizon. The differences specification again produces a confidence intervals with noticeably better coverage properties than the levels specification. Figure 10 shows the results when $T = 200$, which are again qualitatively similar to the $T = 100$ case.

4.3 VAR Models

In this section, we consider a DGP matching the bivariate VAR model considered in Kilian and Kim (2011):

$$Y_t = (x_t, y_t)'$$

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + W_t$$

$$W_t \sim N(0, \Sigma)$$

where:

$$\Phi_0 = \begin{bmatrix} \phi_1^0 \\ \phi_2^0 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} \phi_{11}^1 & 0 \\ \phi_{12}^1 & \phi_{22}^1 \end{bmatrix}$$

and:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$

The structural shocks are known and equal to:

$$U_t = Q * W_t$$

where Q is the inverse of the Cholesky factorization of Σ. Define the components of U_t as $U_t = (\varepsilon_t, u_t)'$. Our interest is then on the response of the second variable to the first structural shock:

$$\beta^h = \frac{\partial y_{t+h}}{\partial \varepsilon_t}$$

For all calibrations, we set: $\phi_1^0 = 0; \phi_2^0 = 0; \phi_{12}^1 = 0.5; \phi_{22}^1 = 0.5; \sigma_1^2 = 1, \sigma_{12} = 0.3, \sigma_2^2 = 1$.

We consider three different values for ϕ_{11}^1:

$$\phi_{11}^1 = 0.50$$
$$\phi_{11}^1 = 0.95$$
$$\phi_{11}^1 = 0.99$$

Figure 11 show the results of the simulations for the VAR model when $T = 100$. The results for the VAR DGP are very similar to the univariate models that we have seen thus far. The levels specification has a small downward bias at the lowest calibration of ϕ_{11}^1, with the bias increasing as ϕ_{11}^1 increases and as the horizon increases. The differences specification has much less bias than the levels impulse response function for all three levels of persistence. The differences specification produces confidence intervals that are consistently undersized for all values of ϕ_{11}^1. However, the coverage of these intervals is much closer to the stated 90% size than those produced from the levels specification, which are extremely undersized. Figure 12 shows the results when $T = 200$, which are again qualitatively similar to those when $T = 100$. As was the case in earlier simulations, when $T = 200$ the differences specification produces confidence intervals with close to correct size.

5 Measuring the Effects of Monetary Policy Shocks Occurring During Recessions

The previous results have demonstrated that for LPs with an observed shock of interest, the differences specification of the LP has less bias and better confidence interval coverage than the levels specification. Among other things, this improvement is increasing for smaller sample sizes. In this section, we provide an example of the estimation differences one can obtain in empirical practice from the levels vs. differences specification.

We focus on a long-standing question of empirical interest, the state dependent effects of U.S. monetary policy shocks. There is substantial interest in whether the effects of monetary policy are symmetric across multiple dimensions. A significant portion of this literature has focused on business cycle asymmetry, namely differences in the effects of U.S. monetary policy when the economy is in an expansion vs. a recession. The ability of LPs to easily incorporate asymmetries have made LPs a popular empirical approach in the recent literature on this topic. However, given that the number of U.S. recessions over the sample typically studied is relatively low, one might expect substantial bias from the levels specification when estimating the effects of monetary policy shocks occurring during recessions, with a commensurate large expected improvement from using the differences specification.[11]

We begin with a levels specification of the LP, which closely follows Tenreyro and Thwaites (2016):

$$\begin{aligned} y_{t+h} &= F_t(\beta_r^h \varepsilon_t + \rho_{1,r}^h y_{t-1} + \rho_{2,r}^h y_{t-2} + \cdots + \rho_{p,r}^h y_{t-p} + (\gamma_r^h)' X_t) \\ &+ (1 - F_t)(\beta_e^h \varepsilon_t + \rho_{1,e}^h y_{t-1} + \rho_{2,e}^h y_{t-2} + \cdots + \rho_{p,e}^h y_{t-p} + (\gamma_e^h)' X_t) + v_{t+h} \end{aligned} \tag{8}$$

where y_{t+h} is output measured in log levels at time horizon h, F_t is an indicator variable indicating if the US economy is in a recession or an expansion, ε_t is the externally identified monetary policy shock, and X_t is a control vector. The coefficients of interest are β_r^h, indicating the response of output at horizon h to monetary policy shocks occurring during recessions, and β_e^h, indicating this response at horizon h for monetary policy shocks occurring during expansions.

We can alternatively estimate these state dependent effects using the differences specification:

$$\begin{aligned} y_{t+h} - y_{t-1} &= F_t(\beta_r^h \varepsilon_t + \theta_{1,r}^h \Delta y_{t-1} + \theta_{2,r}^h \Delta y_{t-2} + \cdots + \theta_{p,r}^h \Delta y_{t-p} + (\alpha_r^h)' X_t^D) \\ &+ (1 - F_t)(\beta_e^h \varepsilon_t + \theta_{1,e}^h \Delta y_{t-1} + \theta_{2,e}^h \Delta y_{t-2} + \cdots + \theta_{p,e}^h \Delta y_{t-p} + (\alpha_e^h)' X_t^D) + u_{t+h}. \end{aligned} \tag{9}$$

Following Tenreyro and Thwaites (2016), the control vector X_t will contain a constant, a linear time trend, and one lag of the Federal Funds Rate. We define the control vector X_t^D to contain a constant and one lag of the Federal Funds Rate. We also follow Tenreyro and Thwaites (2016) in their use of non-linear Romer and Romer (2004) shocks, which allows

[11] Recent work by Gonçalves et al. (2023) has called into question the suitability of local projections for estimating state-dependent macroeconomic effects of policy shocks when the state in question evolves endogenously to macroeconomic shocks. This is certainly applicable for the literature investigating the output effects of monetary policy shocks during recessions. Our focus here is on the results given by the state-dependent LP estimated in levels relative to the LP estimated in differences, and we do not have reason to expect that the Gonçalves et al. (2023) critique would apply more to one specification vs. the other. However, the results of Gonçalves et al. (2023) should be kept in mind when interpreting the absolute IRF estimates for either specification.

for the reaction function of the Federal Reserve to be non-linear over the business cycle. We use monthly U.S. industrial production as our measure of output. The monthly National Bureau of Economic Research (NBER) recession indicator will be used to define F_t in both the levels and differences specification. We compute impulse response functions for five years of monthly responses ($h = 60$). The monthly sample period runs from 1969:03-2008:12. The sample period ends prior to the onset of the Great Recession, since the interest rate was near the zero lower bound for most of the duration and aftermath of the recession. Over this sample period, only 77 months correspond to an NBER recession.

Figure 13 contains the impulse response of industrial production to a one standard-deviation positive Romer and Romer (2004) monetary policy shock occurring during recessions. The dashed line shows the estimated IRF from the levels specification, while the dash-circle line shows the estimated IRF from the differences specification. There are significant differences in the impulse response estimates produced by the levels vs. differences specifications. For approximately the first 15 months of the horizon, there is little difference between the levels and differences specifications. After that point, the difference between the two specifications increases dramatically, and the estimated response of output in the differences specification is much larger in absolute value than the estimated response of output in the levels specification. Overall, the estimates from the differences specification suggests that the effects of monetary policy shocks occurring during recessions are larger and more persistent than is suggested by the estimates from the levels specification. This result is consistent with the bias toward zero observed in the simulation results for the levels specification of the LP.

6 Conclusion

The local projection methodology has become a popular alternative to VAR models for the calculation of impulse response functions. However, there is growing evidence that standard approaches to estimate local projections have significant bias in the sample sizes typically utilized for estimation of these models. There are also discrepancies in the literature with whether local projections are estimated in the log levels of response variables vs. cumulated differences, with a common assumption being that models estimated in levels are more reliable.

In this paper, we have conducted a simulation experiment to compare the performance of local projections estimated in levels vs. cumulated differences on a variety of different data generating processes including ARMA models, unobserved components models, and VAR models. We focused on the case where the econometrician has an externally identified, observed, shock of interest for which they wish to compute impulse response functions.[12] The simulations show the estimates from the levels specification are severely biased and have confidence intervals that are significantly undersized, with these deficiencies growing larger as both the persistence of the process and the horizon of the impulse response increases. In contrast, the cumulated differences specification produces striking improvements over the levels specification in both the amount of bias and confidence interval coverage. In absolute terms, for most data generating processes and impulse response function horizons considered, the differences specification produces close to unbiased estimates and confidence intervals with close to correct coverage. Importantly, the differences specification provides improved inference even in cases where the underlying DGP is well inside the stationary region. In other words, these results suggest that the preference for the differences specification does not hinge on the data containing a unit root.

[12] An interesting avenue for future research is to extend the results presented here to the case where an externally identified, observed shock is used as an instrument for an unobserved shock of interest as in Stock and Watson (2018).

References

Auerbach, A. and Y. Gorodnichenko (2013). Fiscal Multipliers in Recession and Expansion. In A. Alesina and F. Giavazzi (Eds.), *Fiscal Policy after the Financial Crisis*, Chicago. University of Chicago Press.

Gonçalves, S., A. M. Herrera, L. Kilian, and E. Pesavento (2023). State-Dependent, Local Projections. *Working Paper*.

Gospodinov, N., A. M. Herrera, and E. Pesavento (2013). Unit Roots, Cointegration, and Pretesting in VAR Models. *VAR Models in Macroeconomics - New Developments and Applications: Essays in Honor of Christopher A. Sims*, 81–115.

Herbst, E. P. and B. K. Johannsen (2022). Bias in Local Projections. *Working Paper*.

Jordá, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review 95*(1), 161–182.

Kilian, L. and Y. J. Kim (2011). How Reliable are Local Projection Estimators of Impulse Responses? *The Review of Economics and Statistics 93*(4), 1460–1466.

Olea, J. L. M. and M. Plagbørg-Moller (2021). Local Projection Inference is Simpler and More Robust Than You Think. *Econometrica*.

Plagbørg-Moller, M. and C. K. Wolf (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica 89*(2), 955–980.

Ramey, V. and S. Zubairy (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. *Journal of Political Economy April*, 850–901.

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.

Stock, J. H. and M. W. Watson (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal 128*(610), 917–948.

Tenreyro, S. and G. Thwaites (2016). Pushing on a String: U. S. Monetary Policy Is Less Powerful in Recessions. *American Economic Journal: Macroeconomics 8*(4), 43–74.

Xu, K.-L. (2023). Local Projection Based Inference under General Conditions. *Working Paper, Center for Applied Economics and Policy Research*.

Appendix

In this appendix we provide additional detail behind the derivation of equations 6 and 7. Again, we consider an AR(1) data generating process:

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t$$

where ε_t is independent and identically distributed with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$. The correctly specified levels LP for the AR(1) case is:

$$y_{t+h} = c_h^L + \beta_h \varepsilon_t + \phi^{h+1} y_{t-1} + v_{t+h}$$

where $\beta_h = \phi^h$ and $v_{t+h} = \varepsilon_{t+h} + \phi\varepsilon_{t+h-1} + \phi^2\varepsilon_{t+h-2} + \cdots + \phi^{h-2}\varepsilon_{t+2} + \phi^{h-1}\varepsilon_{t+1}$.

Assume we have T total realizations of the random variable ε. We use these observations to create two series of observations, $\varepsilon_t, t = 1, 2, \ldots, T-h$ and $v_{t+h}, t = 1, 2, \ldots, T-h$. We are interested in the expectation of the sample covariance:

$$E(S_{\varepsilon_t, v_{t+h}}) = E\left(\frac{1}{T-h}\sum_{t=1}^{T-h}(\varepsilon_t - \bar{\varepsilon}_t)(v_{t+h} - \bar{v}_{t+h})\right)$$

where:

$$\bar{\varepsilon}_t = \frac{1}{T-h}\sum_{t=1}^{T-h}\varepsilon_t$$

$$\bar{v}_{t+h} = \frac{1}{T-h}\sum_{t=1}^{T-h}v_{t+h}$$

Expanding and noting that $E(\varepsilon_t v_{t+h}) = 0$ for $h \neq 0$ we have:

$$E(S_{\varepsilon_t, v_{t+h}}) = -E(\bar{\varepsilon}_t \bar{v}_{t+h})$$

Now, from the definition of v_{t+h}:

$$-E\left(\bar{\varepsilon}_t \bar{v}_{t+h}\right) = -E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h}\right) - \phi E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h-1}\right) - \phi^2 E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h-2}\right) - \cdots -$$
$$\phi^{h-2} E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+2}\right) - \phi^{h-1} E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+1}\right)$$

It is straightforward to show that:

$$E\left(\bar{\varepsilon}_t \bar{\varepsilon}_{t+h-i}\right) = \frac{\sigma^2}{(T-h)^2}\left[T - 2h + i\right]$$

Combining gives us equation 6:

$$-E\left(\bar{\varepsilon}_t \bar{v}_{t+h}\right) = -\frac{\sigma^2}{(T-h)^2} \sum_{i=0}^{h-1} \phi^i \left[T - 2h + i\right]$$

Turning to the cumulated differences LP, the correctly specified LP for the AR(1) DGP is:

$$y_{t+h} - y_{t-1} = c_h^D + \phi^h \varepsilon_t + \theta_1^h(\Delta y_{t-1}) + \cdots + \theta_h^h(\Delta y_{t-h}) + u_{t+h}$$

where:

$$u_{t+h} = (\varepsilon_{t+h} - \varepsilon_{t-1}) + \phi(\varepsilon_{t+h-1} - \varepsilon_{t-2}) + \phi^2(\varepsilon_{t+h-2} - \varepsilon_{t-3}) + \cdots + \phi^{h-1}(\varepsilon_{t+1} - \varepsilon_{t-h}) - \phi^h \varepsilon_{t-h-1}$$

As before, assume we have T total realizations of ε. We use these observations to create two series of observations, $\varepsilon_t, t = h+2, h+3, \ldots, T-h$ and $u_{t+h}, t = h+2, h+3, \ldots, T-h$. Each of these series contains $T-2h-1$ observations. We are again interested in the expectation of the sample covariance:

$$E(S_{\varepsilon_t, u_{t+h}}) = E\left(\frac{1}{T-2h-1} \sum_{t=h+2}^{T-h} (\varepsilon_t - \bar{\varepsilon}_t)(u_{t+h} - \bar{u}_{t+h})\right)$$

where:

$$\bar{\varepsilon}_t = \frac{1}{T-2h-1}\sum_{t=h+2}^{T-h}\varepsilon_t$$
$$\bar{u}_{t+h} = \frac{1}{T-2h-1}\sum_{t=h+2}^{T-h}u_{t+h}$$

Using similar calculations as for the levels case we have:

$$E(S_{\varepsilon_t,u_{t+h}}) = -E\left(\bar{\varepsilon}_t\bar{u}_{t+h}\right)$$

Now, from the definition of u_{t+h}:

$$\begin{gathered}
-E\left(\bar{\varepsilon}_t\bar{u}_{t+h}\right) = -E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+h}-\bar{\varepsilon}_{t-1}\right)\right) - \phi E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+h-1}-\bar{\varepsilon}_{t-2}\right)\right) - \phi^2 E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+h-2}-\bar{\varepsilon}_{t-3}\right)\right) \\
-\cdots-\phi^{h-2}E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+2}-\bar{\varepsilon}_{t-h+1}\right)\right) - \phi^{h-1}E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+1}-\bar{\varepsilon}_{t-h}\right)\right) \\
+\phi^h E\left(\bar{\varepsilon}_t\bar{\varepsilon}_{t-h-1}\right)
\end{gathered}$$

Also, it is straightforward to show that:

$$E\left(\bar{\varepsilon}_t\left(\bar{\varepsilon}_{t+h-i}-\bar{\varepsilon}_{t-1-i}\right)\right) = \frac{\sigma^2}{(T-2h-1)^2}\left[1-h+2i\right]$$

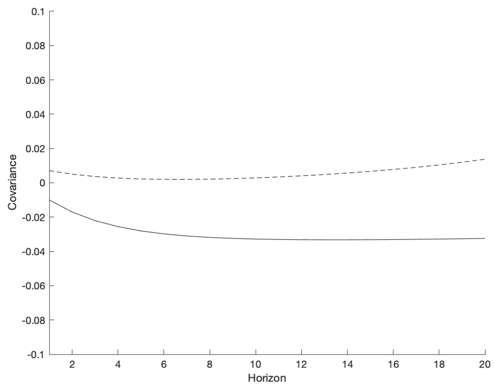
and:

$$E\left(\bar{\varepsilon}_t\bar{\varepsilon}_{t-h-1}\right) = \frac{\sigma^2}{(T-2h-1)^2}\left[T-3h-2\right]$$

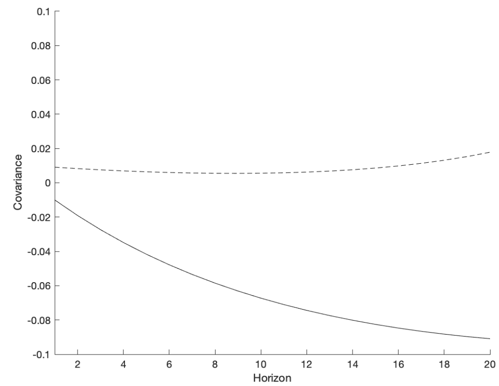
Substituting and rearranging we have equation 7:

$$-E\left(\bar{\varepsilon}_t\bar{u}_{t+h}\right) = \frac{\sigma^2}{(T-2h-1)^2}\left[\phi^h[T-3h-2] - \sum_{i=0}^{h-1}\phi^i\left[1-h+2i\right]\right]$$

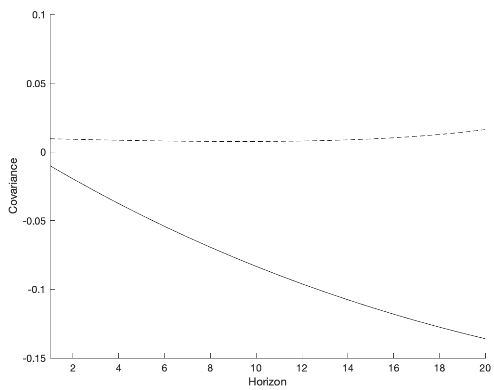
Figure 1
Expected Sample Covariance between Observed Shock and LP Regression Disturbance
$T = 100$



(a) $\phi = 0.7$



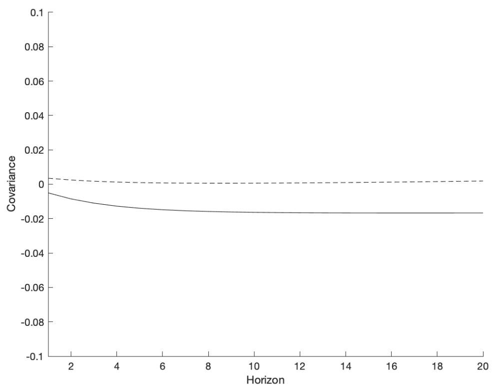
(b) $\phi = 0.9$



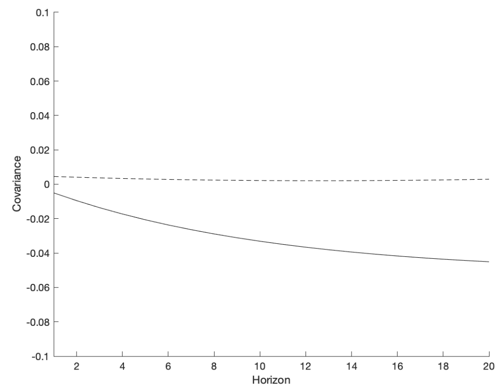
(c) $\phi = 0.95$

Notes: This figure displays the expected sample covariance from equations 6 and 7 when $T = 100$, $\sigma^2 = 1$, and $\phi = \{0.7, 0.9, 0.95\}$. In each sub-figure, the black solid line is the expected sample covariance from the levels specification of the LP, while the black dashed line is the expected sample covariance from the cumulated differences specification of the LP.

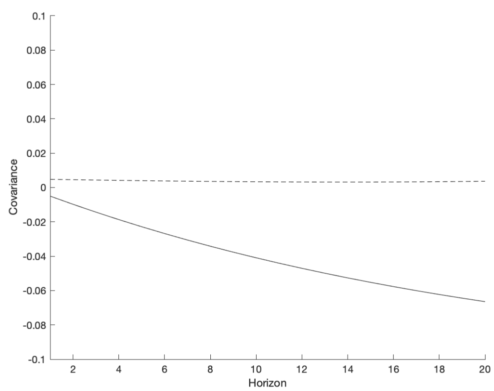
Figure 2
Expected Sample Covariance between Observed Shock and LP Regression Disturbance
$T = 200$



(a) $\phi = 0.7$



(b) $\phi = 0.9$

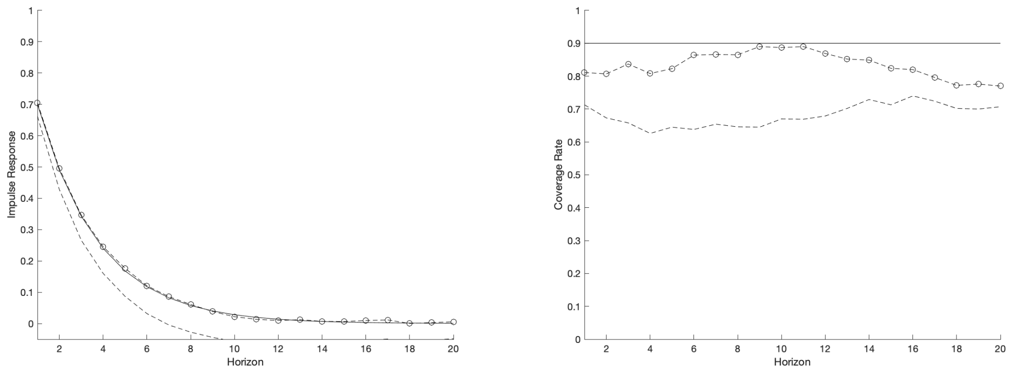


(c) $\phi = 0.95$

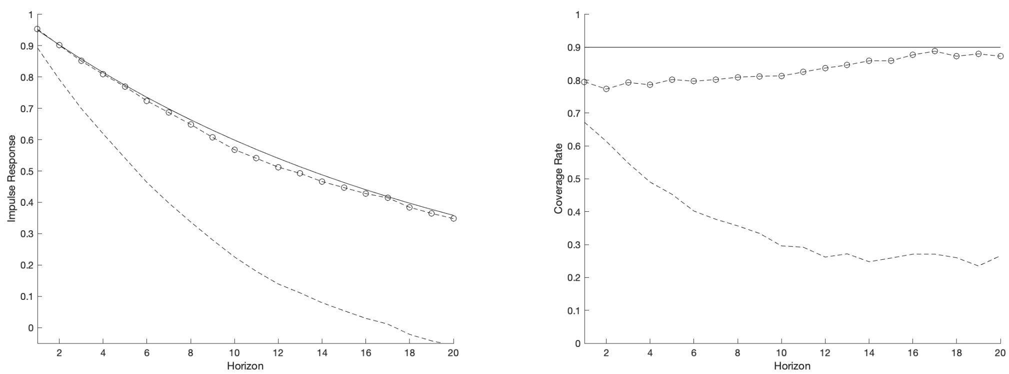
Notes: This figure displays the expected sample covariance from equations 6 and 7 when $T = 200$, $\sigma^2 = 1$, and $\phi = \{0.7, 0.9, 0.95\}$. In each sub-figure, the black solid line is the expected sample covariance from the levels specification of the LP, while the black dashed line is the expected sample covariance from the cumulated differences specification of the LP.

Figure 3
AR Model ($T = 100$)

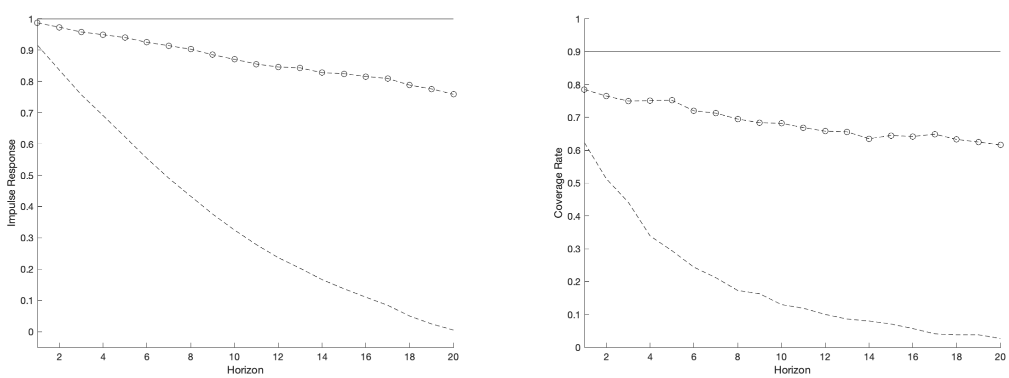
$\phi = 0.7$



$\phi = 0.95$



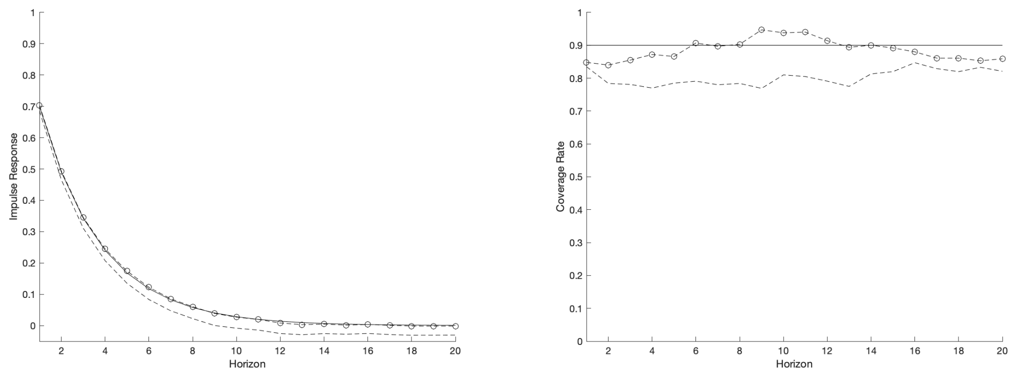
$\phi = 1.0$



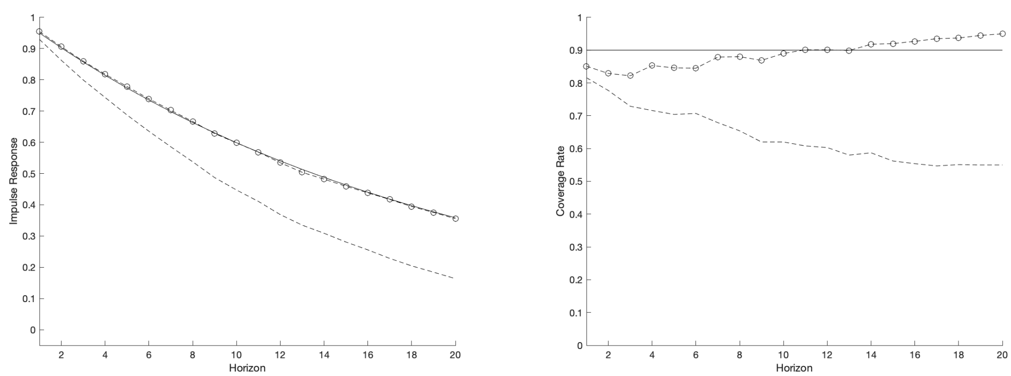
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is an AR(1) model and $T = 100$. Results for three alternative values of the autoregressive parameter ($\phi = \{0.7, 0.95, 1.0\}$) are displayed. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 4
AR Model ($T = 200$)

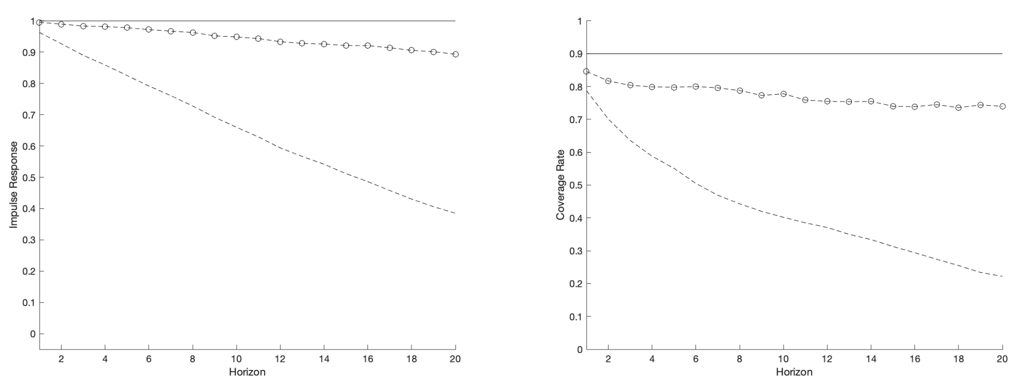
$\phi = 0.7$



$\phi = 0.95$



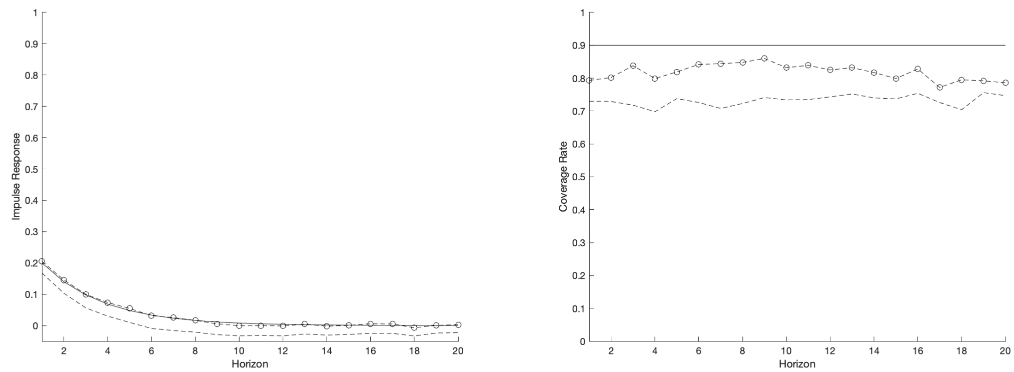
$\phi = 1.0$



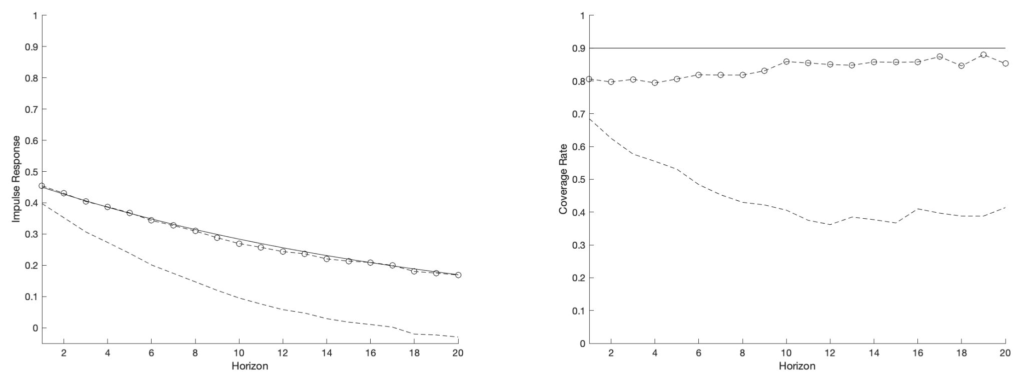
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is an AR(1) model and $T = 200$. Results for three alternative values of the autoregressive parameter ($\phi = \{0.7, 0.95, 1.0\}$) are displayed. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 5
ARMA(1,1) Model ($T = 100$)

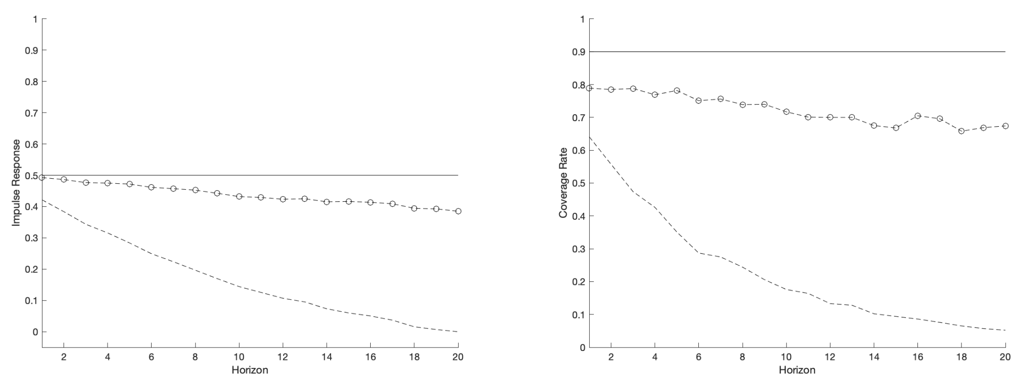
$\phi = 0.7$



$\phi = 0.95$



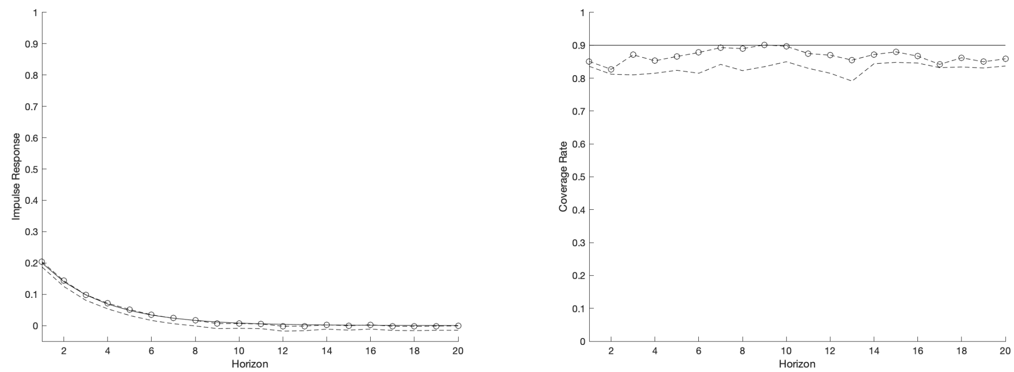
$\phi = 1.0$



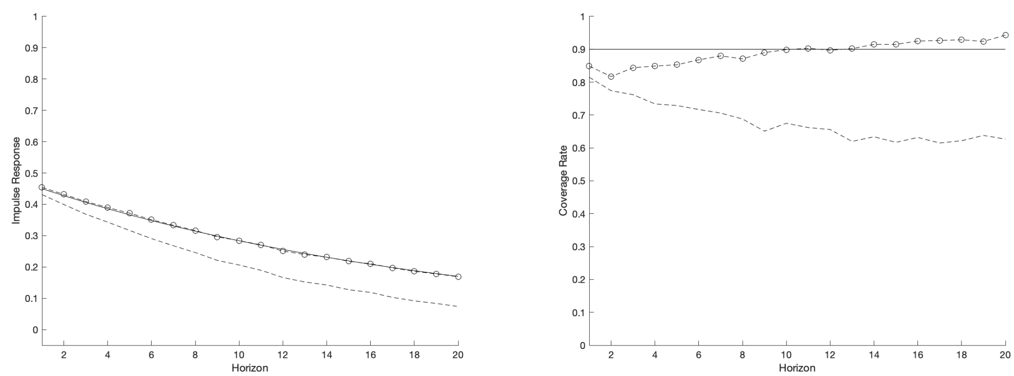
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is an ARMA(1,1) model and $T = 100$. Results for three alternative values of the autoregressive parameter ($\phi = \{0.7, 0.95, 1.0\}$) are displayed. The moving average parameter is set to -0.5. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 6
ARMA(1,1) Model ($T = 200$)

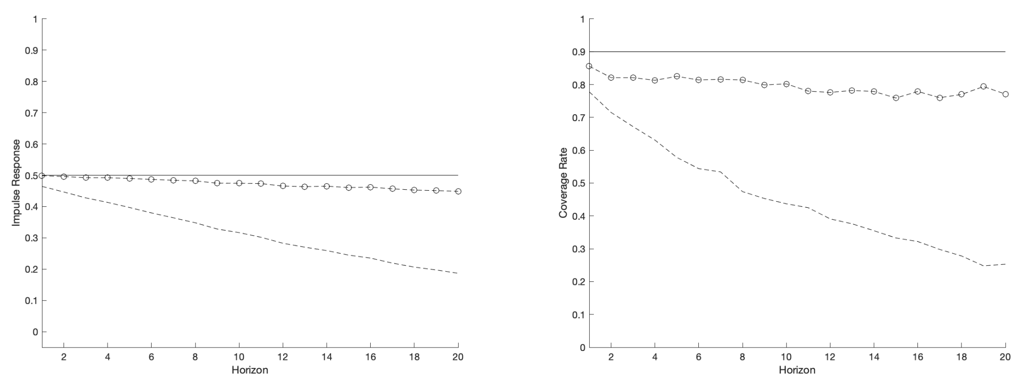
$\phi = 0.7$



$\phi = 0.95$

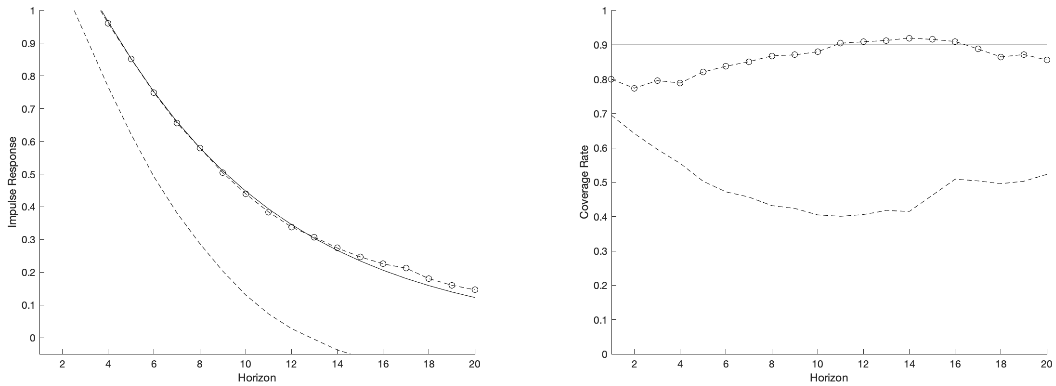


$\phi = 1.0$



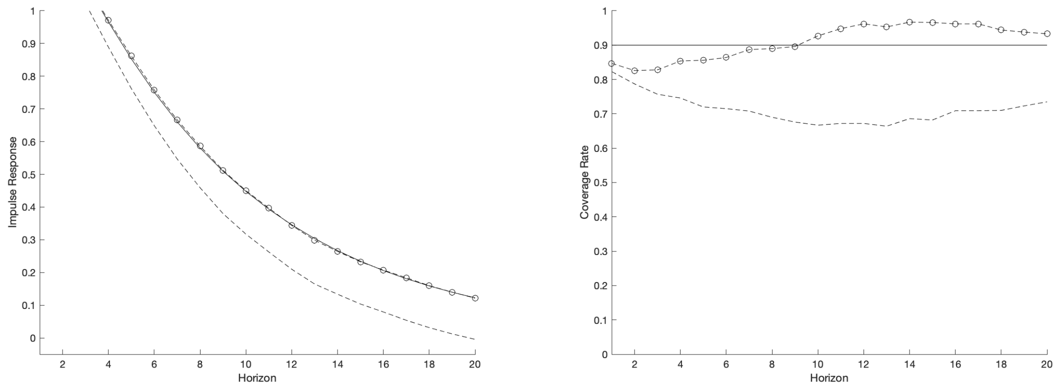
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is an ARMA(1,1) model and $T = 100$. Results for three alternative values of the autoregressive parameter ($\phi = \{0.7, 0.95, 1.0\}$) are displayed. The moving average parameter is set to -0.5. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 7
Trend Stationary UC Model ($T = 100$)



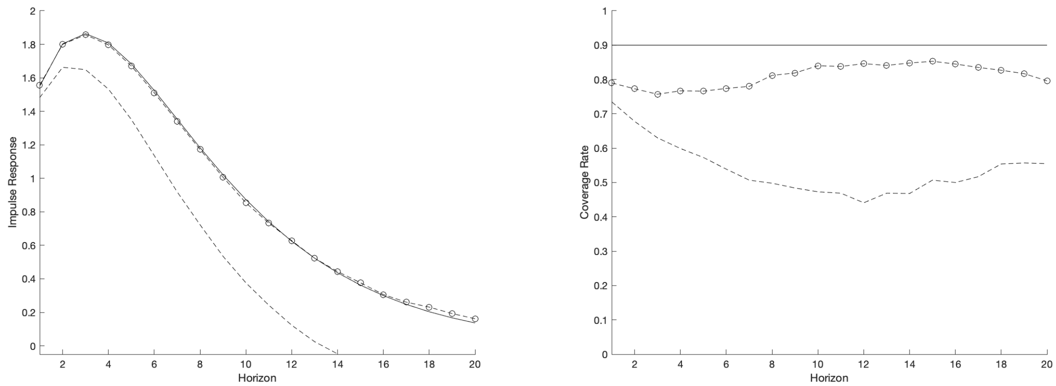
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a Trend Stationary Unobserved Components model where $\mu = 0.77, \phi_1 = 1.22, \phi_2 = -0.3$, and $\sigma = 0.76$ and $T = 100$. These parameter values were obtained by calibrating the model based on estimations of quarterly real GDP from 1969:Q1 to 2007:Q4. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 8
Trend Stationary UC Model ($T = 200$)



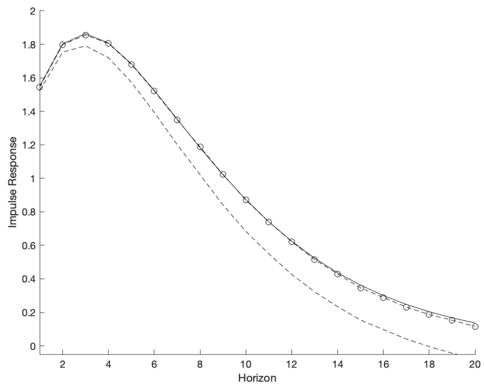
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a Trend Stationary Unobserved Components model where $\mu = 0.77, \phi_1 = 1.22, \phi_2 = -0.3$, and $\sigma = 0.76$ and $T = 200$. These parameter values were obtained by calibrating the model based on estimations of quarterly real GDP from 1969:Q1 to 2007:Q4. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

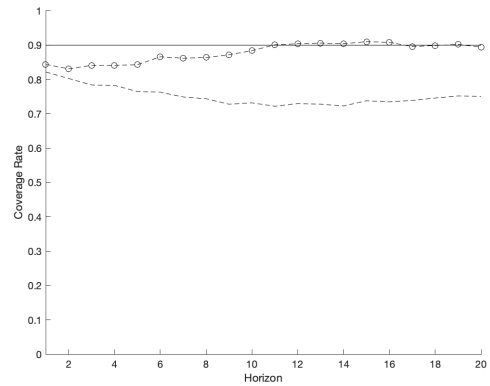
Figure 9
Stochastic Trend UC Model ($T = 100$)



Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a Stochastic Trend Unobserved Components model where $T = 100$ and the parameters are set to $\mu = 0.77, \phi_1 = 1.55, \phi_2 = -0.6, \gamma = 0.5783$, and $\sigma = 0.443$. These parameter values were obtained by calibrating the model based on estimations of quarterly real GDP from 1969:Q1 to 2007:Q4. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 10
Stochastic Trend UC Model ($T = 200$)

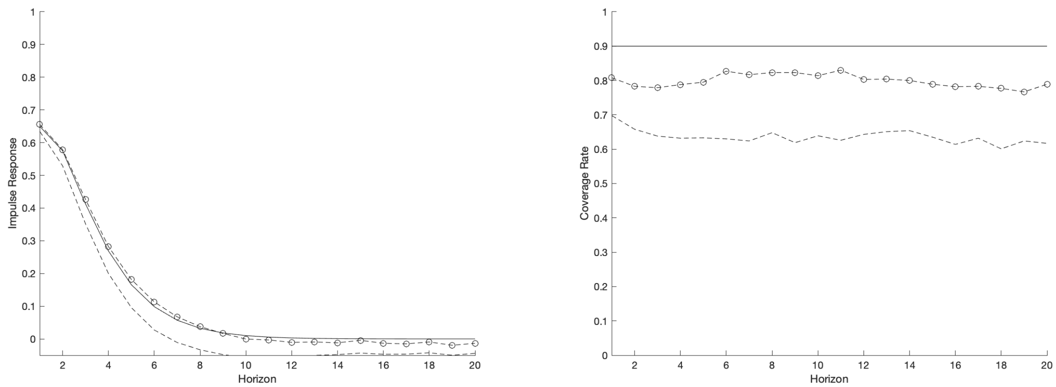




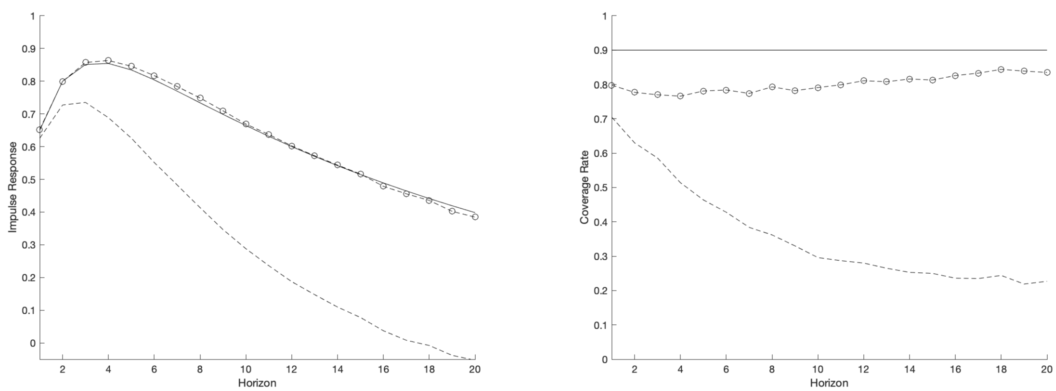
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a Stochastic Trend Unobserved Components model where $T = 200$ and the parameters are set to $\mu = 0.77, \phi_1 = 1.55, \phi_2 = -0.6, \gamma = 0.5783$, and $\sigma = 0.443$. These parameter values were obtained by calibrating the model based on estimations of quarterly real GDP from 1969:Q1 to 2007:Q4. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 11
VAR Model ($T = 100$)

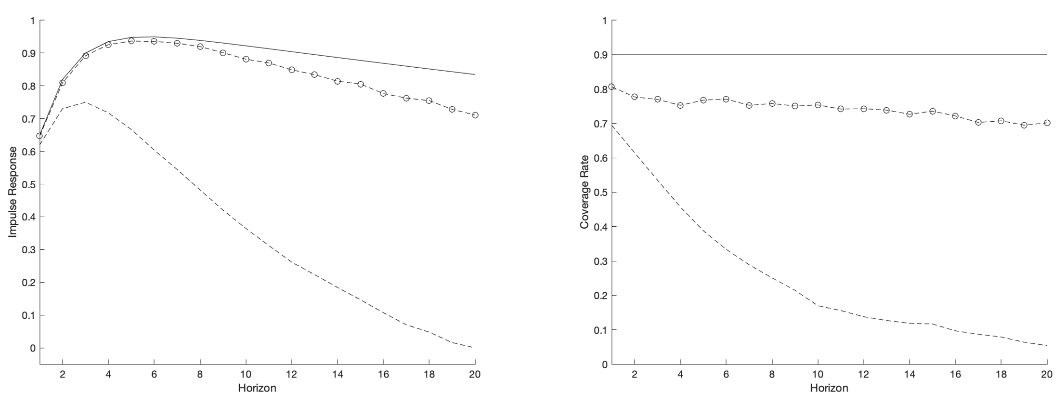
$$\phi_{11}^{1} = 0.50$$



$$\phi_{11}^{1} = 0.95$$



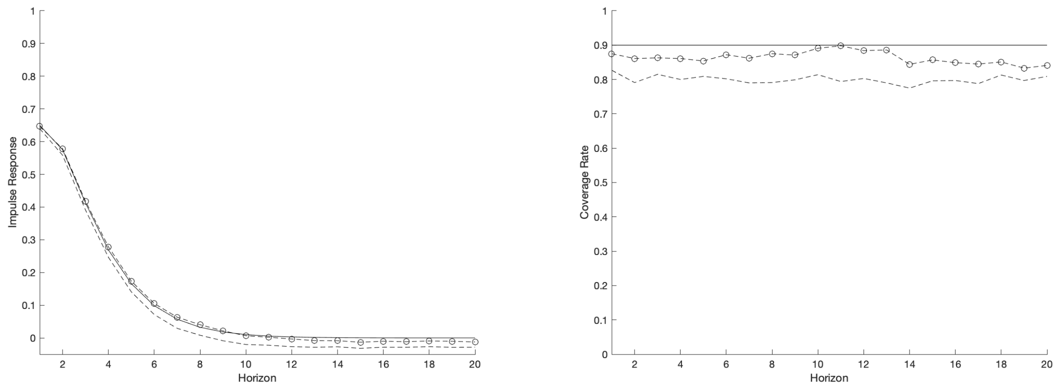
$$\phi_{11}^{1} = 0.99$$



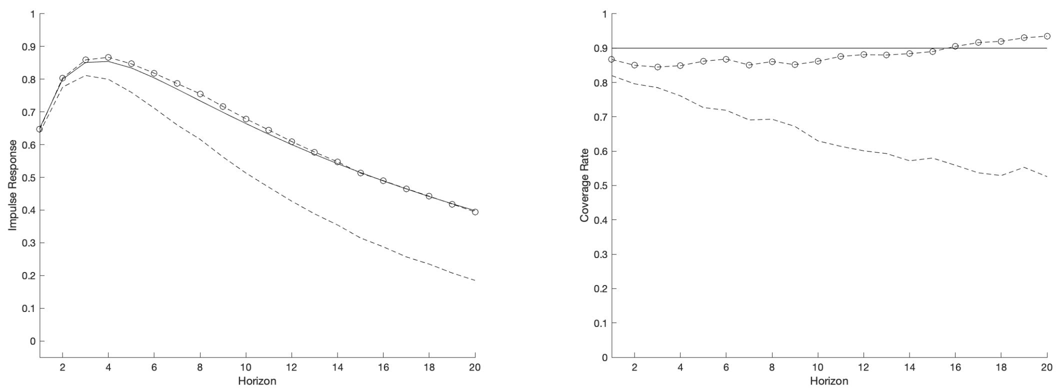
Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a VAR model and $T = 100$. Results are shown for three alternative values of $\phi_{11}^{1} = \{0.50, 0.95, 0.99\}$. Other parameters are set as described in Section 4.3. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 12
VAR Model ($T = 200$)

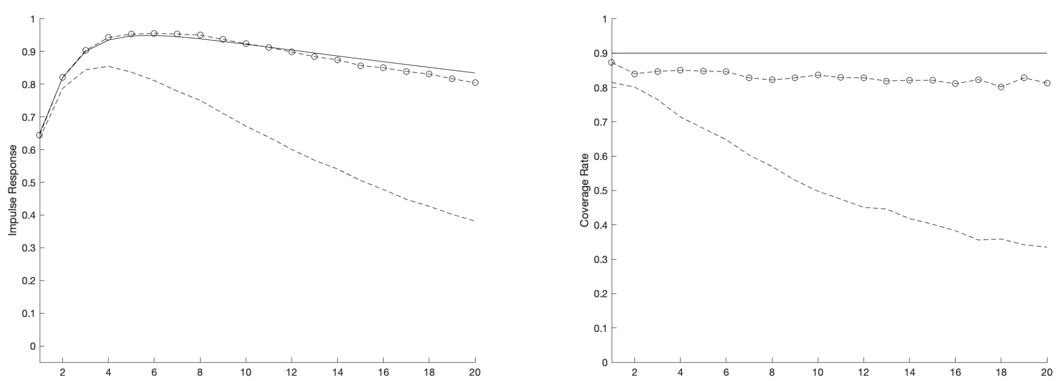
$\phi_{11}^{1} = 0.50$



$\phi_{11}^{1} = 0.95$

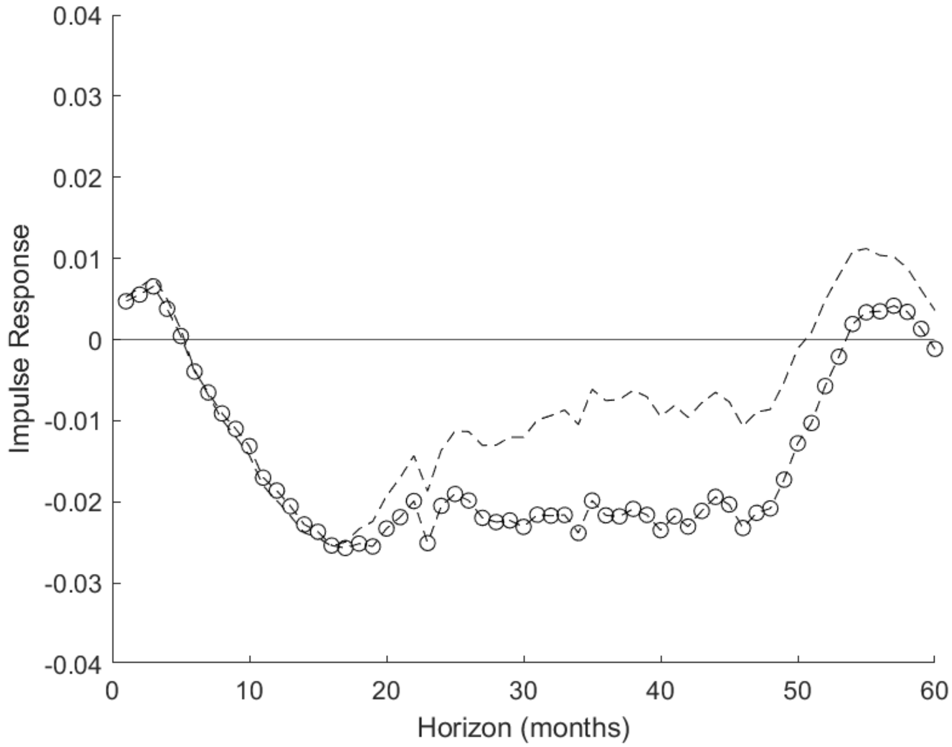


$\phi_{11}^{1} = 0.99$



Notes: This figure displays simulation results from estimation of the levels and differences specification of the LP when the true DGP is a VAR model and $T = 200$. Results are shown for three alternative values of $\phi_{11}^{1} = \{0.50, 0.95, 0.99\}$. Other parameters are set as described in Section 4.3. The left column shows the average impulse response function estimate for the levels specification (dashed line) and differences specification (dash-circle line), as well as the true impulse response function (solid line). The right column shows the 90% confidence interval coverage of the true impulse response function for the levels specification (solid line) and differences specification (dash-circle line).

Figure 13
Impulse Response Function of Industrial Production to Monetary Policy Shock Occurring During Recession



Notes: This figure shows the impulse response function of U.S. Industrial Production to a one standard-deviation positive Romer and Romer (2004) monetary policy shock occurring during an NBER recession month. The estimated response using the levels specification is the dashed line, while the estimated response using the differences specification is the dash-circle line. The monthly sample period used for estimation extends from 1969:03-2008:12.