Innovation Powered Narrative Inference*

Régis Barnichon[(a)] *and* *Geert Mesters*[(b)]

[(a)] Federal Reserve Bank of San Francisco and CEPR

[(b)] Universitat Pompeu Fabra, Barcelona School of Economics and CREI

July 9, 2024

Abstract

The narrative identification approach is often only able to identify a small number of exogenous events accurately, leading to consistent but high variance estimates of dynamic causal effects. In contrast, the time series model identification approach —e.g., recursive, long-run, max-share, etc— leads to low variance but possibly inconsistent estimates because of omitted variables. In this work, we introduce a new inference method —innovation powered narrative inference—, which combines the benefits of both identification approaches by using innovations from time series models to boost the power of narrative methods while preserving consistency. The method is illustrated using the narrative monetary series of Romer and Romer (1989, 2023), documenting considerable reductions in confidence intervals for the causal effects of monetary policy.

JEL classification: C14, C32, E32, E52, N10.

Keywords: Narrative methods, Time series, Prediction powered, Monetary policy.

*We thank seminar participants at BI Norwegian, Brown, Norges Bank, SETA2024 and Princeton. Mesters acknowledges financial support from the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/50110001103 and the 2021 ERC Starting Grant POLICYMETRICS. The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

1 Introduction

In recent years, the narrative identification approach has become a cornerstone of causal identification in macroeconomics (Romer and Romer, 2023).[1] In a nutshell, it consists in exploiting real time narrative records (e.g., newspapers, government reports, or policy meeting transcripts) in order to identify exogenous movements in a variable of interest —construct a *narrative* instrumental variable— that can be used to learn a causal effect of interest in an IV regression.[2] Intuitively, the idea is that the information contained in the narrative accounts allow to overcome reverse causality and omitted variable biases and thereby more firmly establish causal linkages.

The narrative approach has been a major step forward for identification in macro, but it is constrained by an important limitation: instances of clear and indisputable exogenous variations are *rare*, and the narrative instrument series contain many zeros. In the context of monetary policy for instance, Romer and Romer (2023) isolate only 10 exogenous monetary policy events in 70 years of monetary history, the rest are coded as zeros. More generally, the typical narrative instrument features about 80% of zeros (see Table 1). While this situation does not invalidate the approach — the instrument is still valid—, it can lead to an efficiency problem, i.e., large and uninformative confidence bands and/or estimates that are overly sensitive to specification choices. In other words, narrative instruments are often weak.

In this work, we propose a method that can solve, or at least considerably alleviate, the low power of the narrative identification approach, and we do so by leveraging another popular, though arguably less convincing, identification approach in macro: the time series identification approach pioneered by Sims (1980).

Our key innovation is to view identification in macro as a missing data problem — instances of clear and indisputable shocks are rare in the narrative records, and most shocks escape detection—. Instead of imputing zeros for the narrative instrument when the records are inconclusive, we treat inconclusive dates as *missing* entries and use time series identification —short-run, long-run, max share, etc... (e.g., Ramey, 2016)— to *predict* the missing entries —the missing shocks— with time series *innovations*.[3] By explicitly predicting the

[1] Popular narratively identified series include the oil shocks of Hamilton (1983, 1985), the monetary event series of Romer and Romer (1989), the government spending news shocks of Ramey (2011), the tax shocks of Romer and Romer (2010); Cloyne (2013), and the financial shocks of Reinhart and Rogoff (2009), Jalil (2015) and Romer and Romer (2017), among many others.

[2] Going beyond the estimation of dynamic causal effects, narrative instruments can also be used to tackle other important macro questions, such as the estimation of structural equations (Barnichon and Mesters, 2020; Lewis and Mertens, 2023) or counter-factual policy analyses (Barnichon and Mesters, 2023; McKay and Wolf, 2023).

[3] Crucially, we will *not assume* that the time series identification scheme is correct. We therefore refer to these time series predictions as *innovations* —not shocks—.

Table 1: Zero Events in Narrative Studies

Paper	# obs	# zeros	% zeros
Bi and Zubairy (2023)	590	464	79%
Carriere-Swallow, David and Leigh (2021)†	28	22	81%
Cloyne (2013)	248	94	38%
Cloyne, Dimsdale and Postel-Vinay (2023)	89	69	77%
Fieldhouse and Mertens (2023)	292	228	78%
Guajardo, Leigh and Pescatori (2014)†	32	22	68%
Gil et al. (2019)	120	87	72%
Hamilton (1985)	140	121	86%
Jalil (2015)	90	83	92%
Ramey and Shapiro (1998)	200	197	98%
Ramey (2011)	280	194	69%
Ramey and Zubairy (2018)	504	396	79%
Romer and Romer (2023)	852	842	99%
Romer and Romer (1989)	852	845	99%
Romer and Romer (2010)	252	207	82%
Romer and Romer (2017)†	91	81	88%
Rojas, Vegh and Vuletin (2022)†	106	86	81%
Average			80%

Notes: Number and percentage of zeros across different narrative series. Here † indicates that the reported values are per unit averages from the panel data setting considered. We omit narrative accounts reported in books, e.g. Reinhart and Rogoff (2009) and Alesina, Favero and Giavazzi (2019), as they include several datasets. Inspection reveals that the fractions of zeros in such texts are not different from the ones reported.

missing data instead of systematically imputing zeros, we can more efficiently handle the missing data and ultimately improve the efficiency of the narrative identification method.

The intuition underlying our approach is simple: by construction a time series identification exploits the entire time series variation —variation over the entire sample period— and thus has typically much higher power than a narrative method. At the same time however, a time series identification method may not be entirely convincing and can lead to a biased causal effect estimate. Fortunately, and this is key, the bias can be estimated from the narrative identification approach: on the dates where the narrative records speak clearly, we can estimate the bias of the time series identification approach and thus ultimately correct for it. In other words, by combining the narrative and the time series identification approach, it is possible to construct an estimator that is both consistent *and* efficient.

Specifically, we show that our "innovation-powered narrative estimator can be cast in a GMM framework where the moment conditions include both the conventional moments based on the narrative instrument as well as the (bias corrected) innovation powered moment conditions based on using the innovations as instruments. The additional moment conditions do not have a first order effect on the estimator but reduce the asymptotic variance of the baseline GMM estimator under mild assumptions and provide a way to merge the credible, but limitedly available, narrative evidence with evidence from conventional time series methods, where identification is often perceived as less credible, but the strategies can be adopted for all time periods.

The GMM framework is completed by a model for the selection probability for observing an informative narrative shock. This allows the arrival of narrative shocks to depend on the state of the economy. Further, despite adding the innovation-powered moment conditions the narrative instruments often remain weak instruments. Therefore our preferred inference approach follows the weak instrument literature and we develop an innovation powered version of the subset Anderson and Rubin (1949) statistic, see also Stock and Wright (2000) and Andrews and Guggenberger (2019).

We illustrate innovation powering by revisiting the seminal work of Romer and Romer (1989, 2023). We reevaluate the original narrative series and set dates where the narrative evidence is uninformative to missing, leaving only the identified dates of Romer and Romer (2023) and the periods where no policy meeting occurred as observed. We combine this series with an innovation series that is constructed by assuming a conventional timing assumption on the impact of monetary policy. Together this allows to implement our innovation powering method and infer the effect of monetary policy on inflation and output. We find large reductions in the length of the confidence intervals relative to using the current approach which includes the many zeros in the instrument. These differences are robust to numerous specification changes and various reasonable choices for the innovations.

Our paper relates to two strands of literature. First, an alternative approach for incorporating narrative evidence in time series models is presented in Ludvigson, Ma and Ng (2017) and Antolín-Daz and Rubio-Ramírez (2018) who suggest restricting the shocks in a structural VAR to conform with the narrative evidence. A prior robust Bayesian version of this idea, with frequentist guarantees, is developed in Giacomini, Kitagawa and Read (2023), see also Giacomini, Kitagawa and Read (2022) for more discussion. Relative to this approach innovation powered narrative inference addresses the following limitations (some were also highlighted in Plagborg-Møller (2022)): (i) we do not require specifying an invertible SVAR model nor a likelihood, (ii) we allow for narrative instruments that are contaminated with measurement error and (iii) we account for selection and relax the assumption that narrative events arrive at random.

Second, our approach is inspired by viewing identification in macro as a missing data problem — clearly interpretable shocks are often missing from the historical narrative accounts —. As such our methodology builds on the missing data literature (e.g. Little and Rubin, 2020), and most closely on the augmented inverse probability weighted (AIPW) estimators developed in Robins, Rotnitzky and Zhao (1994), see also Robins and Rotnitzky (1995), Rotnitzky, Robins and Scharfstein (1998), Bang and Robins (2005), Wooldridge (2007), Kang and Schafer (2007), Rotnitzky et al. (2012) and Jing Qin and Leung (2017) for numerous refinements of this approach. More generally, such estimators are widely used for estimating treatment effects in the causal inference literature. The main difference in our setting is that we work in an aggregate macro time series setting where instruments are often weak allowing us to combine the main idea from the AIPW estimator with time series prediction routines and weak identification robust methods (e.g. Andrews, Stock and Sun, 2019). Further, in our setting the instrumental variable is missing and not the explanatory or outcome variable, which relaxes the implications of the missing-at-random assumption, see also Chaudhuri and Guilkey (2016); Abrevaya and Donald (2017).

The remainder of this paper is organized as follows. The next section presents a simple illustrative example that highlights the main ideas. Section 3 presents the general dynamic framework and Section 4 discusses the empirical results. Section 5 concludes.

2 Motivating example

We informally introduce innovation powered narrative inference by means of a simple static example where the macro environment is given by the simultaneous equations model

$$\underbrace{\begin{pmatrix} p_t \\ y_t \end{pmatrix}}_{=\mathbf{w}_t} = \underbrace{\begin{pmatrix} 1 & \rho \\ \theta & 1 \end{pmatrix}}_{=\mathbf{A}} \underbrace{\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}}_{=\boldsymbol{\eta}_t}, \qquad \text{for } t \in \mathcal{N}\ , \tag{1}$$

with $\mathbf{w}_t = (y_t, p_t)'$ the observed macro variables, $\boldsymbol{\eta}_t = (\varepsilon_t, u_t)'$ the structural shocks and the non-singular matrix $\mathbf{A}$ includes the structural parameters θ and ρ. We emphasize that the existence of an invertible SVAR(0) model is only imposed for exposition purposes and in the main treatment no such underlying model is defined. The shocks are assumed to be serially and mutually independent with mean zero and diagonal variance matrix $\mathbf{D}$. Time periods are indexed by t and the available periods are collected in the set $\mathcal{N}$ with $n = |\mathcal{N}|$. We are interested in learning θ: the causal effect of ε_t on y_t, and note that without further assumptions θ is not point identified.

For expositional purposes, we can think of y_t as the output gap and p_t as the central bank policy rate and ε_t as a monetary shock.

The narrative identification approach

The narrative identification approach consists in using narrative accounts (official records, newspaper articles, speeches, etc..) in order to isolate dates (events) with uncontroversial movements in ε_t. Formally, we can view the narrative identification approach as a function $f(\cdot)$ that constructs a narrative instrument time series z_t defined as

$$z_t = f(\varepsilon_t) + \zeta_t \ , \qquad t \in \mathcal{N} \ , \tag{2}$$

where $f(\varepsilon_t)$ is a function of the shock of interest (ε_t), and ζ_t captures any additional measurement error, which is independent from the structural shocks.

In its current practice, the narrative approach $f(\cdot)$ constructs z_t as follows: (i) on dates when the narrative account is sufficiently informative, construct an estimate for ε_t,[4], (ii) on all other dates, i.e., when the narrative account is inconclusive, set z_t to zero. Together, these two steps ensure that z_t is a valid instrument for θ: $\mathbb{E}(z_t p_t) \neq 0$ and $\mathbb{E}(z_t(y_t - p_t\theta)) = 0$, and we can estimate θ from the regression

$$y_t = \theta p_t + e_t$$

using z_t as instrument variable for p_t.

The narrative approach has been a major step forward for identification in macro, but it is constrained by an important limitation: instances of clear and indisputable ε_t shocks are *rare*. This has a number of implications. First, the narrative records are inconclusive in most cases, and most values of z_t end up coded as zeroes. In the context of monetary policy for instance, Romer and Romer (2023) isolate 10 exogenous monetary policy events, such that 99 percent of the values of z_t are coded as zeroes. While this does not invalidate using z_t as an instrumental variable, this means that researchers exploit only a small share of the variance of p_t in order to infer θ. This can lead to an *efficiency* problem, i.e., large error bands.

Second, with few non-zero entries, a narrative instrument may suffer from finite sample confounding, i.e., an unlucky finite sample correlation between some correctly identified subset of shocks ε_t and other structural shocks. For instance, Hoover and Perez (1994) argue that the few uncontroversial monetary shocks of Romer and Romer (1989) coincide (by chance) with oil shocks, see also the discussion in Nakamura and Steinsson (2018).

Third, even with extensive documentation, researchers engaged in narrative identification must often use judgment calls, which benefit from the benefit of hindsight. This could lead

[4] Should p_t be some aggregate variable like taxes or government spending, one can construct an estimate for a shock to a sub-component of p_t, for instance individual income tax (Mertens and Ravn, 2013) or defense spending (Ramey and Zubairy, 2018).

to subconscious endogeneity biases, see for instance the critique of the Friedman-Schwartz dates in Romer and Romer (1989) and the discussion in Romer and Romer (2023).

Together, low power and endogeneity biases can compound each other: a researcher willing to only entertain uncontroversial dates without possible confounders or judgment calls can be left with very few exogenous variations, leading to even larger share of zeroes and more efficiency problems.

From zeros to missings

In this paper, we propose a method to improve the efficiency of the narrative approach. Our key starting idea is to view identification in macro as a missing data problem, which will ultimately allow us to improve inference by more efficiently handling the missing data —the missing shocks—, than imputing zeroes.

Formally, we introduce the notation

$$s_t = \begin{cases} 0 & \text{if} \quad t \in \mathcal{B} = \{t \in \mathcal{N} : z_t \text{ missing}\} \\ 1 & \text{if} \quad t \in \mathcal{G} \end{cases} , \tag{3}$$

where on the "bad" periods $\mathcal{B}$ the entry for z_t is treated as missing and on the "good" periods $\mathcal{G}$ the value for z_t is regarded as informative and credible. As such the indicator s_t records when the informative and reliable narrative events occur.[5]

Importantly, we will not assume that missing/non-missing entries arrive at random, and we explicitly allow the event of observing z_t to depend on the structural shocks. Formally, we let π_t denote the probability of narratively detecting a shock ε_t and assume that

$$\pi_t = P(s_t = 1|\boldsymbol{\eta}_t, \zeta_t) = P(s_t = 1|\boldsymbol{\eta}_t) \ . \tag{4}$$

This rules out that z_t is set to missing based on the measurement error ζ_t. At the same time it allows s_t to be a function of all structural shocks in the economy. For instance it is likely easier to identify exogenous movements in the central bank policy rate after times of persistently high inflation, such as the Volcker nomination by President Carter or the beginning of the post-COVID tightening cycle (Romer and Romer, 2023). For clarity of exposition, we treat the probability of detecting an ε_t as a given function of structural shock, and we postpone the estimation of a model for π_t to the general treatment.

With missing entries for z_t, we can still estimate θ consistently using the narrative in-

[5]We have that $\mathcal{G} \cup \mathcal{B} = \mathcal{N}$, $\mathcal{G} \cap \mathcal{B} = \emptyset$, $n_{\mathcal{G}} = |\mathcal{G}|$ and $n_{\mathcal{B}} = |\mathcal{B}|$. As Table 1 highlights, in practice we have that $n_{\mathcal{G}} \ll n_{\mathcal{B}}$.

strument z_t from the moment conditions moment condition becomes

$$\mathbb{E}(z_t(y_t - \theta p_t)) = \mathbb{E}(s_t z_t(y_t - \theta p_t)/\pi_t) = 0 \ , \quad \text{and} \quad \mathbb{E}(z_t p_t) \neq 0 \ . \tag{5}$$

where the first equality follows directly from the imposed assumption on the missing instrument mechanism (4). From this the simple IV estimate is computed as

$$\hat{\theta}_{\mathcal{G}} = \frac{\sum_{t\in\mathcal{N}} s_t z_t y_t/\pi_t}{\sum_{t\in\mathcal{N}} s_t z_t p_t/\pi_t} = \frac{\sum_{t\in\mathcal{G}} z_t y_t/\pi_t}{\sum_{t\in\mathcal{G}} z_t p_t/\pi_t} \ , \tag{6}$$

which is a conventional IV estimate, except perhaps for the weighting by the selection probabilities. When the narrative events arrive at random (i.e. $\pi_t = \pi$) the IV estimate (6) is standard. Including additional periods where $z_t = 0$ does not change the estimate.

However, and this is low-efficiency problem inherent to the narrative approach, the variance of $\hat{\theta}_{\mathcal{G}}$ will typically be large as only $n_{\mathcal{G}}$ observations are used to construct the estimate, i.e. the conventional variance of $\hat{\theta}_{\mathcal{G}}$ is of order $O(n_{\mathcal{G}}^{-1})$. The objective of innovation powered inference is to reduce the variance of the simple IV estimate without introducing any inconsistencies.

From missings to innovations

From the perspective of the data imputation literature, imputing zeros for all missing shocks is not the most efficient approach to address the missing data problem, see e.g., Little and Rubin (2020) for the limitations of mean imputation. Instead, we propose to use the time series identification approach to predict the missing shocks.

By filling the missing values with predictions about the likely values of ε_t instead of systematically imputing zeros, we can improve estimation efficiency for θ inference. Importantly, while the time series identification may not be correct and thus introduce a bias if used in isolation, we will see that it is possible to combine the narrative and time series identification methods to construct an estimator that is both consistent and more efficient than the traditional narrative estimator based on z_t alone.

To illustrate our approach, consider a simple prediction model for the structural shock ε_t given by

$$v_t = p_t - \beta y_t \ , \qquad \text{for } t \in \mathcal{N} \ , \tag{7}$$

where β is considered fixed for simplicity. We can think about v_t as a prediction for ε_t that is obtained by regressing p_t on the observable macro variables and taking the residual.

In our general treatment below such predictions are obtained from identified time series models and make use of popular identification schemes such as short-run, long-run or max share schemes as typically adopted in VAR or local projection models (e.g., Plagborg-Møller and Wolf, 2019). The example in equation (7) can be viewed as a special case of a short run identification. We refer to v_t as an *innovation* to make clear that we *do not* require the time series identification to be correct. Indeed v_t in (7) generally depends on u_t and not only on the shock of interest ε_t, making it an invalid instrument for identifying the causal effect of ε_t.

We define two IV estimates based on the innovation v_t:

$$\hat{\theta}^*_{\mathcal{N}} = \frac{\sum_{t\in\mathcal{N}} v_t y_t}{\sum_{t\in\mathcal{N}} v_t p_t} \qquad \text{and} \qquad \hat{\theta}^*_{\mathcal{G}} = \frac{\sum_{t\in\mathcal{G}} v_t y_t/\pi_t}{\sum_{t\in\mathcal{G}} v_t p_t/\pi_t}\,. \tag{8}$$

These two estimators have two features. First, the estimators $\hat{\theta}^*_{\mathcal{N}}$ and $\hat{\theta}^*_{\mathcal{G}}$ will in general not be consistent and lead to biased estimates of θ. However, and this is key, the two estimators will have the same bias. By exploiting this property, we can build a consistent estimator.

Second, the first estimator uses the entire sample period, so that its variance can be much smaller than the variance of the baseline narrative estimator $\hat{\theta}_{\mathcal{G}}$ in (6), typically an order of magnitude smaller for the typical narrative instruments containing about 80 percent missing entries.

Innovation-Powered IV

The key idea in this paper is to combine the best features of the estimators above; combining the consistency of the narrative instrument with the efficiency of the time series identification.

To see that, consider the baseline *Innovation Powered Instrumental Variable* (IPIV) estimator

$$\hat{\theta} = \hat{\theta}^*_{\mathcal{N}} + \hat{\theta}_{\mathcal{G}} - \hat{\theta}^*_{\mathcal{G}} \tag{9}$$

The innovation powered estimate starts from the inconsistent but precise estimate $\hat{\theta}^*_{\mathcal{N}}$ and applies the bias correction $\hat{\theta}_{\mathcal{G}} - \hat{\theta}^*_{\mathcal{G}}$ to ensure that $\hat{\theta}$ is consistent.

Further, while the variance of the IPIV estimator is dominated by the bias correction term $\hat{\theta}_{\mathcal{G}} - \hat{\theta}^*_{\mathcal{G}}$ this variance can be smaller when compared to the variance of the simple IV estimator $\hat{\theta}_{\mathcal{G}}$ whenever the innovations v_t are sufficiently accurate for the narrative series z_t. In such case the variance of the IPIV estimate will be smaller when compared to the variance of the simple IV estimate. To see this in a simple way, suppose for a moment that $v_t = z_t$ for $t \in \mathcal{G}$, e.g. the narrative identification coincides perfectly with say a recursive identification strategy, then the bias correction $\hat{\theta}_{\mathcal{G}} - \hat{\theta}^*_{\mathcal{G}}$ cancels and the variance of the IPIV estimator coincides with the low variance of the full sample estimator $\hat{\theta}_{\mathcal{N}}$.

The baseline IPIV estimate intuitively illustrates how we can boost the power of the narrative instrument approach using the time series identification method. In practice, we do not advocate using the estimator (9), as it is not the most efficient way to combine narrative and time series identification. To see this note that (9) is determined by three moment conditions: (i) $\mathbb{E}(s_t z_t(y_t - \theta p_t)/\pi_t) = 0$ yielding $\hat{\theta}_{\mathcal{G}}$, (ii) $\mathbb{E}(v_t(y_t - \theta^* p_t)) = 0$ yielding $\hat{\theta}^*_{\mathcal{N}}$ and (iii) $\mathbb{E}(s_t v_t(y_t - \theta^* p_t))/\pi_t = 0$ yielding $\hat{\theta}^*_{\mathcal{G}}$. Each of the corresponding sample moment conditions is solved separately and given equal weight in (9). We can improve efficiency by (a) solving the moment conditions jointly and (b) optimally weighting them. We note that conditions (ii) and (iii) must receive equal weight as otherwise the bias term does not cancel out. This leaves us to choose the weight on (ii)-(iii) vs (i).

Specifically, we may freely choose $\tilde{\gamma}$ in

$$\underbrace{\mathbb{E}((s_t z_t(y_t - \theta p_t)/\pi_t)}_{\text{baseline IV}} + \tilde{\gamma}\underbrace{\mathbb{E}((1 - s_t/\pi_t)v_t(y_t - \theta p_t))}_{\text{innovation powering}} = 0 \ , \tag{10}$$

which is a weighted average of the baseline moment conditions noting that the innovation powering term is the sum of (ii)-(iii). Our objective is to choose $\tilde{\gamma}$ such that the resulting moment estimate for θ has low variance. In line with the intuition presented in the previous section we propose to choose $\tilde{\gamma}$ such that v_t explains the most variation in z_t.

To see under which conditions this is an efficient choice consider the estimator $\hat{\theta}(\tilde{\gamma})$ that is defined by solving the sample moments for an arbitrary fixed $\tilde{\gamma}$:

$$\frac{1}{n}\sum_{t=1}^{n} s_t z_t(y_t - \hat{\theta}(\tilde{\gamma})p_t)/\pi_t + \tilde{\gamma}(1 - s_t/\pi_t)v_t(y_t - \hat{\theta}(\tilde{\gamma})p_t) = 0 \ .$$

The asymptotic variance of $\hat{\theta}(\tilde{\gamma})$ is of the form $\text{Avar}(\hat{\theta}(\tilde{\gamma})) = V(\tilde{\gamma})/G^2$, where G is the expected gradient that does not depend on $\tilde{\gamma}$ and $V(\tilde{\gamma})$ can be written as

$$\begin{aligned} V(\tilde{\gamma}) &= \text{Var}[z_t(y_t - \theta p_t) - (1 - s_t/\pi_t)(z_t - v_t\tilde{\gamma})(y_t - \theta p_t)] \\ &= \text{Var}[z_t(y_t - \theta p_t)] + \mathbb{E}[\mathbb{E}\{(z_t - v_t\tilde{\gamma})^2|\mathbf{w}_t\}(y_t - \theta p_t)^2(1 - \pi_t)/\pi_t] \ . \end{aligned}$$

To minimize $V(\tilde{\gamma})$ with respect to $\tilde{\gamma}$ we need to solve a problem of the form $\min_\mu \mathbb{E}\{(z_t - \mu)^2|\mathbf{w}_t\}$ for which the solution is $\hat{\mu} = \mathbb{E}(z_t|\mathbf{w}_t)$. In our work we propose to approximate this conditional expectation by γv_t, where γ is the linear projection coefficient of projecting z_t on v_t, i.e. $\gamma = \mathbb{E}(v_t^2)^{-1}\mathbb{E}(v_t z_t)$. It follows that choosing such γ in (10) minimizes the asymptotic variance of $\hat{\theta}(\tilde{\gamma})$ whenever $\mathbb{E}(z_t|\mathbf{w}_t) = \gamma v_t$.

Combining our preferred moment conditions take the form

$$\mathbb{E}g(\mathbf{d}_t, \psi) = 0 \ , \qquad g(\mathbf{d}_t, \psi) = \begin{bmatrix} s_t z_t(y_t - \theta p_t)/\pi_t + \gamma(1 - s_t/\pi_t)v_t(y_t - \theta p_t) \\ s_t v_t(z_t - v_t\gamma)/\pi_t \end{bmatrix} \ , \tag{11}$$

where $\mathbf{d}_t = (y_t, p_t, z_t, s_t, v_t)'$ collects all data and the second moment condition identifies γ as the desired projection estimate. Based on these moments we can compute innovation powered GMM estimates and conduct weak identification robust inference using conventional methods (e.g. Hall, 2005; Andrews, Stock and Sun, 2019). The advantage of such moment estimates over the baseline IPIV estimate is that the correlation between the narrative instruments and the innovations is maximized and hence the variance of the estimate for θ is minimized.

3 General framework

We discuss innovation powered inference for a general macro environment where we are interested in estimating a dynamic causal effect using a narrative series as instrument.

3.1 Model and assumptions

Let $\mathbf{w}_t = (y_t, p_t, w_{3t}, \ldots, w_{Kt})'$ denote a vector of macro variables that includes, among others, the outcome variable y_t and the explanatory variable of interest p_t. The object of interest is the dynamic causal effect $\theta_h \in \Theta \subset \mathbb{R}$ defined in the linear model

$$y_{t+h} = \theta_h p_t + \boldsymbol{\beta}_h' \mathbf{x}_t + u_{t+h} \ , \qquad \text{for } t \in \mathcal{N} \ , \tag{12}$$

where $\mathbf{x}_t$ denotes a vector of pre-determined control variables whose effect is captured by $\boldsymbol{\beta}_h$. Typically, $\mathbf{x}_t$ includes a constant and some lags of $\mathbf{w}_t$. Equation (12) can be derived from an underlying structural vector moving average model (e.g. Stock and Watson, 2018), but such additional structure is not necessary here. Similar as above we identify θ_h using the narrative instrument z_t which is treated as missing whenever it takes value zero or when there is a suspicion that z_t depends on u_{t+h}. The indicator for this event is given by

$$s_t = \mathbf{1}(t \in \mathcal{G}) = \mathbf{1}(z_t \text{ observed}) \ , \tag{13}$$

where $\mathcal{G} \subset \mathcal{N}$ are the good periods.

For exposition purposes we start by projecting out the control variables to obtain

$$y_{t+h}^{\perp} = \theta_h p_t^{\perp} + u_{t+h}^{\perp} \ , \qquad \text{for } t \in \mathcal{N} \ , \tag{14}$$

where $l_{t+j}^{\perp} = l_{t+j} - \text{Proj}(l_{t+j}|\mathbf{x}_t)$ for $l = y, p, u, z$ and $j = 0, 1, 2, \ldots$, with $\text{Proj}(a|b)$ denoting the linear projection of a on b. Throughout we treat the projected variables as observed, noting that no changes occur when replacing the population projection by its empirical counterpart. The narrative instrument $z_t^\perp$ is assumed to satisfy the following exogeneity assumption.

Assumption 1. $\mathbb{E}(z_t^\perp u_{t+h}^\perp) = 0$ *for all* t *with* $u_{t+h}^\perp$ *as defined in* (14).

This imposes that the narrative instrument is uncorrelated with the error term after projecting out the control variables (e.g. Stock and Watson, 2018).

Let $\mathbf{v}_t$ denote the vector of innovations used for innovation powering (a formal definition is given below). The assumption for the mechanism that generated the (un)informative narrative events is as follows.

Assumption 2. *Let* $\mathbf{d}_t^\pi = (y_{t+h}^\perp, p_t^\perp, \mathbf{v}_t')'$. *We have*

$$\pi(\mathbf{d}_t^\pi) \equiv P(s_t = 1|\mathbf{d}_t^\pi) = P(s_t = 1|\mathbf{d}_t^\pi, z_t) \ ,$$

and $\pi(\mathbf{d}_t^\pi) > 0$ *with probability 1. Further,*

$$\pi(\mathbf{d}_t^\pi) = \kappa(\mathbf{d}_t^\pi; \boldsymbol{\gamma}_\pi) \ , \tag{15}$$

where κ *is some known function that is differentiable with respect to* $\boldsymbol{\gamma}_\pi \in \Gamma_\pi \subset \mathbb{R}^{d_\gamma}$.

In contrast to the simple example we state the missing assumption directly in terms of the observable variables and not in terms of the underlying structural shocks. Nonetheless, the condition has the same implications: the probability of having a good narrative period is allowed to depend on the endogenous variables and only arrives at random conditional on these variables. Further, we assume a parametric model for the selection probabilities, where the most common modeling choice would be to take $\kappa(\mathbf{d}_t^\pi; \boldsymbol{\gamma}_\pi)$ as a logit function.[6]

Innovations from time series models

Next, we discuss the construction of the innovations $\mathbf{v}_t \in \mathbb{R}^{d_v}$. The requirements are that $\mathbf{v}_t$ depends only on observable variables and is available for all time periods. Further, ideally the innovations should be good predictors for $z_t^\perp$. While this is inherently an application specific task, there is a general route that follows naturally from the macro literature.

Specifically, since $z_t^\perp$ usually aims to capture a specific type of structural shock, say ε_t denoting e.g. a monetary, fiscal or oil shock, we can use virtually all structural time series models to obtain innovations. From fully fledged DSGE to structural VARs with short run restrictions any of these can provide useful proxies for the structural shock of interest.

[6]In principal, non-parametric methods can also be adopted for estimating $\pi(\mathbf{d}_t^\pi)$. However, for most narrative studies the sample size is prohibitively small to do so in a reliable manner.

We will not discuss all possible options but confine ourselves to a class of local projection and VAR models where the identifying restrictions can be implemented using control variables, see Plagborg-Møller and Wolf (2021) for examples. Specifically, consider the prototypical macro regression $\mathbf{w}_{t+h}^s = \boldsymbol{\delta}_h \varepsilon_t + \tilde{\boldsymbol{\beta}}_h^{s\prime} \mathbf{x}_t^s + \mathbf{u}_{t+h}^s$ where $\mathbf{w}_{t+h}^s$ and $\mathbf{x}_t^s$ are selected outcome and control variables. Usually, the literature provides guidance on which variables are affected by the shock at which horizons. Further, since shocks are not observed a wealth of time series identification strategies have been proposed to identify $\boldsymbol{\delta}_h$, several of these can be implemented by selecting $\mathbf{x}_t^s$ appropriately. We propose to exploit these strategies and reverse the regression to obtain good predictions for ε_t, i.e. we consider

$$\varepsilon_t = \boldsymbol{\gamma}_h^{s\prime} \mathbf{w}_{t+h}^s + \boldsymbol{\beta}_h^{s\prime} \mathbf{x}_t^s + u_{t+h}^s \ ,$$

where u_{t+h}^s is the error error term that is uncorrelated (by construction) with $(\mathbf{w}_{t+h}^s, \mathbf{x}_t^s)$. Note that this reverse model is a projection model that has no causal interpretation and the coefficients $\boldsymbol{\gamma}_h^s$ are the best linear projection coefficients. The innovations $\mathbf{v}_t$ are taken as the predictors from this model after projecting out the controls. By construction these should be good predictors for ε_t.

In practice, we would like to exploit predictors from different horizons, e.g. for monetary policy the contemporaneous interest rate may be a good predictor for a monetary policy shock, but inflation eight quarters ahead may also be a good predictor. To accommodate including both we define

$$\mathbf{v}_t = (v_{1t}, \dots, v_{d_v t})' \ , \qquad \text{with} \quad v_{it} \in \{\mathbf{w}_{t+h}^s - \text{Proj}(\mathbf{w}_{t+h}^s | \mathbf{x}_t^s), h = 0, 1, \dots, H\} \ . \tag{16}$$

We note that the theory places no restrictions on the specific innovations needed. In our empirical work below we always consider innovations from the class (16), but this is motivated by good practical performance not theoretical necessity.

The innovations model is formalized in the following (non-structural) assumption.

Assumption 3. $\mathbb{E}(\mathbf{v}_t(z_t^{\perp} - \boldsymbol{\gamma}_v' \mathbf{v}_t)) = \mathbf{0}$ *for all* t *with* $\mathbf{v}_t$ *as defined in* (16).

The assumption defines $\boldsymbol{\gamma}_v$ as the prediction coefficients that minimize the mean squared error for explaining $z_t^{\perp}$; our narrative proxy for ε_t. The innovation parameters will be estimated jointly with the other unknown parameters that we collect in $\boldsymbol{\psi} = (\theta_h, \boldsymbol{\gamma}')' \in \Psi = \Theta \times \Gamma$ with $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_\pi', \boldsymbol{\gamma}_v')' \in \Gamma = \Gamma_\pi \times \Gamma_v$.

3.2 Innovation powered inference

We discuss innovation powered inference for θ_h as defined in (14) subject to assumptions 1-3. The required moment conditions are collected in

$$\mathbb{E}g(\boldsymbol{\psi}; \mathbf{d}_t) = \mathbf{0} \ ,$$

with observations $\mathbf{d}_t = (\mathbf{d}_t^{\pi\prime}, s_t, z_t^{\perp})'$ and

$$g(\boldsymbol{\psi}; \mathbf{d}_t) = \begin{bmatrix} s_t z_t^{\perp}(y_{t+h}^{\perp} - \theta_h p_t^{\perp})/\kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi) + (1 - s_t/\kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi))\boldsymbol{\gamma}_v' \mathbf{v}_t (y_{t+h}^{\perp} - \theta_h p_t^{\perp}) \\ s_t \mathbf{v}_t (z_t^{\perp} - \boldsymbol{\gamma}_v' \mathbf{v}_t)/\kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi) \\ \kappa^{(1)}(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi)(s_t - \kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi))/(\kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi)(1 - \kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi))) \end{bmatrix} \in \mathbb{R}^{d_g} \ , \tag{17}$$

with $\kappa^{(1)}(\mathbf{d}_t^{\pi}; \gamma_\pi) = \partial\kappa(\mathbf{d}_t^{\pi}; \gamma_\pi)/\partial\gamma_\pi$. The moments can be intuitively separated into two parts. The first and second lines are exactly as in the simple example (see equation (11)), except that the innovations are now vector valued and the selection probability is parametrized, i.e. $\pi_t = \kappa(\mathbf{d}_t^{\pi}; \boldsymbol{\gamma}_\pi)$. The third line is new and can be recognized as the score function for the binary response model based on assumption 2. These moments allow to identify the parameters $\boldsymbol{\gamma}_\pi$ of the selection model.

To obtain parameter estimates and confidence sets we first define the sample average $\widehat{g}_n(\boldsymbol{\psi}) = \frac{1}{n}\sum_{t=1}^{n} g(\boldsymbol{\psi}; \mathbf{d}_t)$ and we let $\widehat{\Omega}_n(\boldsymbol{\psi})$ denote any consistent estimate for $\Omega_n(\boldsymbol{\psi}) = \mathrm{Var}(n^{-1/2}\sum_{t=1}^{n} g(\boldsymbol{\psi}; \mathbf{d}_t))$, which is usually obtained use some HAC type estimator (e.g. Newey and West, 1987; Andrews, 1991).

Innovation powered GMM estimation

Given the moment conditions we adopt conventional GMM methods for parameter estimation. Our preferred estimator, which aligns with our approach for constructing confidence bands, is the continuous updating GMM estimator of Hansen, Heaton and Yaron (1996) here defined as

$$(\hat{\theta}_{nh}^{\mathrm{IP}}, \hat{\boldsymbol{\gamma}}')' = \hat{\boldsymbol{\psi}} = \underset{\tilde{\boldsymbol{\psi}} \in \Psi}{\operatorname{argmin}}\, \widehat{g}_n(\tilde{\boldsymbol{\psi}})' \widehat{\Omega}_n^{-1}(\tilde{\boldsymbol{\psi}}) \widehat{g}_n(\tilde{\boldsymbol{\psi}}) \tag{18}$$

where $\hat{\theta}_{nh}^{\mathrm{IP}}$ is the innovation powered dynamic causal effect estimate. Under the assumption of strong instruments and some regularity conditions this estimate is consistent and asymptotically normal for θ_h. In practice, strong instrument assumptions are questionable for macro applications and our preferred inference approach will be based on weak instrument robust methods.

Innovation powered weak identification robust confidence bands

We discuss how to construct weak identification robust confidence bands for θ_h by inverting hypothesis tests for

$$H_0 : \theta_h = \bar{\theta}_h \qquad \text{against} \qquad H_1 : \theta_h \neq \bar{\theta}_h \ , \tag{19}$$

for some constant $\bar{\theta}_h$. Our preferred test statistic is an innovation powered version of the subvector Anderson and Rubin (1949) statistic which was developed for GMM by Stock and Wright (2000).

The IP-AR test statistic is given by

$$\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) = \min_{\tilde{\gamma}\in\Gamma} n\widehat{g}_n(\bar{\theta}_h, \tilde{\gamma})'\widehat{\Omega}_n^{-1}(\hat{\theta}_h, \tilde{\gamma})\widehat{g}_n(\hat{\theta}_h, \tilde{\gamma}) \ , \tag{20}$$

where θ_h is fixed at its null value and the minimum over the nuisance parameters $\boldsymbol{\gamma}$ is computed numerically. Under mild regularity conditions that we formalize in Assumption 4 in the appendix we have the following result.

Proposition 1. *Given assumptions 1-4 we have that under* H_0

$$\lim_{n\to\infty} P_\psi(\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) > c_{\chi^2(1),\alpha}) = \alpha$$

where $c_{\chi^2(1),\alpha}$ *denotes the 1-*α *quantile of the chi-squared distribution with 1 degree of freedom.*

The proof is collected in the appendix. The proposition shows that the null rejection probability is equal to the nominal level of the test α when we reject the null using critical values from the chi-squared one distribution. This result can be used to define the following confidence set for θ_h.

$$\mathrm{CS}_{nh}^{\mathrm{IP}} = \{\theta_h \in \Theta : \mathrm{AR}_n^{\mathrm{IP}}(\bar{\psi}_n) \leq c_{\chi^2(1),\alpha}\} \ . \tag{21}$$

In practice, we compute $\mathrm{CS}_{nh}^{\mathrm{IP}}$ by searching over the parameter space Θ and collecting all values for which the IP-AR test statistic takes values below $c_{\chi^2(1)}$.

3.3 Simulation study

We summarize the results from a simulation study that we conducted to evaluate the gains in efficiency from innovation powering.

Simulation design. We simulate data from a conventional vector autoregressive model for monetary policy resembling either a quarterly or monthly design. For the quarterly design we fit the model with 4 lags to GDP growth g_t, PCE inflation π_t and the short term interest rate i_t over the 1959Q1-2007Q2 period using US data. For the monthly model we consider the same variables but now replace GDP growth by the unemployment rate and use 12 lags. For both models the structural shocks are denoted by $\boldsymbol{\varepsilon}_t = (\varepsilon_t^g, \varepsilon_t^\pi, \varepsilon_t^{mp})'$, where ε_t^{mp} is the monetary policy shock. This shock is identified by imposing a lower triangular assumption on the impact matrix of the SVAR model. Importantly, when implementing the innovation powering methods we will not make use of this short run restriction.

Given the fitted models we simulate time series of length $n = 200$ (quarterly model) or $n = 400$ (monthly model), and we are interested in estimating θ_h in model (12) where $y_{t+h} = g_{t+h}$, $p_t = i_t$ and $\mathbf{x}_t$ includes lags of g_t, π_t, i_t where the number of lags is chosen using the Akaike information criteria. The narrative series is observed with probability π_t which is modeled by using a logit specification and the value is equal to the monetary policy shock, possibly contaminated by measurement error.

$$z_t = \begin{cases} \varepsilon_t^{mp} + \zeta_t & w.p. \quad \pi_t \\ \text{missing} & w.p. \quad 1 - \pi_t \end{cases} ,$$

where we consider two different choices for π_t:

$$\pi_t^1 = 0.2 \;, \qquad \text{and} \qquad \pi_t^2 = \frac{\exp(\gamma_{1\pi} + \gamma_{2\pi}|\varepsilon_t^{mp}|)}{1 + \exp(\gamma_{1\pi} + \gamma_{2\pi}|\varepsilon_t^{mp}|)} \;,$$

where the second choice implies that the narrative is more likely to be observable when the monetary policy shock is large.

Importantly the selection model that we implement is always based on fitting a logit model using observables $(1, y_{t+h}^\perp, p_t^\perp, \mathbf{v}_t')'$. As such the selection model is either not parsimonious (for π_t^1) or only an approximation of the true selection model (for π_t^2). The parameters for generating π_t^2 are chosen to ensure that on approximately 80% of the time periods z_t is missing: $\boldsymbol{\gamma}_\pi = (-2, 0.5)'$. The measurement error ζ_t is iid normal with variance calibrated such that the effective F-statistic of the first stage is equal to 10.

To implement the innovation powering method we need to choose which innovations $\mathbf{v}_t$ are used. While ad hoc choices are in principle fine choosing the innovations in a smart way can reduce the mse of the IP estimates. To illustrate we consider the following two choices: (i) $\mathbf{v}_t = i_t^\perp$ and (ii) $\mathbf{v}_t = g_t^\perp$. Intuitively, (i) which uses interest rate innovations is more promising as the monetary policy shock is typically more pronounced in these series. In contrast, (ii) uses GDP innovations and variance decompositions show that monetary policy explains little variation in output. Note that neither choice makes use of the lower triangular ordering in the SVAR used for simulating the data.

We compare the mean absolute error (MAE) of $\hat{\theta}_h^{\mathrm{IP}}$ to the conventional narrative IV estimator that uses z_t with the missing values replaced by zeros as instrument.[7] We note

[7] Due to the relatively weak instruments in the simulation design some outliers in the parameter estimates

that (i) z_t with imputed zeros remains a valid instrument as it is only a function of ε_t^{mp} and independent measurement error, and (ii) even though z_t includes many zeros, after projecting out $\mathbf{x}_t$ these zeros are replaced by linear combinations of the lagged variables.

Further, we compare the innovation powered confidence sets $\mathrm{CS}_{nh}^{\mathrm{IP}}$ to the conventional weak IV robust confidence set that do not use innovation powering. Specifically, since we have a single instrument for a single endogenous variable we use the conventional Anderson and Rubin (1949) statistic which is efficient in this setting. For each model specification we generate $M = 5000$ data sets. We report the mean squared error estimate, the empirical size of the test under H_0 and the average confidence set length.

Results. The results are shown in Table 2: the top panel shows the results for the quarterly model whereas the bottom panel shows the monthly model results.

For both specifications the mean absolute error estimates indicate that the innovation powered GMM estimates that use the residuals of the contemporaneous interest rate as innovations $\mathbf{v}_t = i_t^{\perp}$ are nearly always more accurate when compared to the conventional IV estimates. Moreover, the gains in accuracy can be very large, often reducing the errors two to three times.

In contrast when using $\mathbf{v}_t = g_t^{\perp}$ (or $\mathbf{v}_t = u_t^{\perp}$ in the monthly model) we find that innovation powering generally leads to a small increase in the MAE relative to the conventional IV estimate. This finding corresponds to our intuition as $g_t^{\perp}$ does not depend on the monetary policy shock and effectively only introduces noise in the estimation routine. The take away point is that a smart choice for the innovations is important for the success of innovation powering.

When comparing the mean absolute errors across the different selection models we find little differences. The mild mis-specification for π_t^2 does not seem to materially affect the results.

The empirical rejection frequencies of the different Anderson-Rubin tests are always close to the nominal level of the test. This holds regardless of which innovations are chosen and reassures us that with on average 80% of instruments missing the IP-AR test still has a reliable rejection probability.

The key advantage of innovation powering is revealed when looking at the $\mathrm{CS}_{nh}^{\mathrm{IP}}$ confidence bands. When using $\mathbf{v}_t = i_t^{\perp}$ as innovations the confidence bands are typically 2-3 times smaller when compared to the conventional AR-based confidence bands that include zeros instead of missing values for the instruments. This finding holds regardless of the specification considered.

occur (notably in the conventional IV estimator). This makes the comparison based on mean squared error somewhat distorted.

4 Empirical study: US monetary policy

In this section we revisit the seminal work of Romer and Romer (1989, 2023) which studies the effects of monetary policy. This is an inherently difficult task as omitted variable bias is a central concern: both monetary policy actions and macro outcome variables, e.g. inflation and unemployment, are likely to be influenced by other variables, e.g. expectations, fiscal policy, financial stress and so on.

To identify the effect of monetary policy actions Romer and Romer (1989) conducted a narrative study that identified periods where the Federal Reserve changed the interest rate for reasons unrelated to current or prospective real economic activity. These periods are labeled as policy "shocks" as they are not driven by output or other factors affecting output. The study was updated in Romer and Romer (2023) resulting in a binary series z_t that contains ten monetary events, i.e. ten ones, that are associated with periods in which exogenous changes in monetary policy occurred. This series is subsequently used in a autoregressive distributed lag model Romer and Romer (1989), or local projection Romer and Romer (2023), to capture the effects of monetary policy on output and prices.

In our work we will follow recommendations in Stock and Watson (2018) and use z_t as an instrumental variable to identify exogenous variation in the interest rate i_t. This approach avoids possible bias from measurement error and quantifies the effects of monetary policy in terms of policy relevant interest rate changes. As the interest rate we use the Fed Funds rate which implies that our monthly data sample starts in 1954M7 and is taken until 2016M12. This has the consequence that we do not include the October 1947 monetary shock, but this omission does not lead to much changes.[8]

The model that we consider, i.e. equation (12), is completed by selecting either inflation or unemployment as the dependent variable y_{t+h}, setting $p_t = i_t$ and defining the controls $\mathbf{x}_t$ as a constant and twelve lags of p_t and y_t.

To implement the innovation powered narrative method we treat the uninformative periods of the Romer and Romer (2023) narrative series as missing. Specifically, our baseline specification uses the original series, but sets all zero months to missing except for the months where there was no FOMC meeting. For these months it seems reasonable to assume that there was no monetary policy news and hence $z_t = 0$ is an informative value. This gives us $n_{\mathcal{G}} = 156$ good periods and $n_{\mathcal{B}} = 593$, which implies that 79% of values are treated as missing. Robustness to alternative treatments of the narrative information are presented below.

Our baseline specification uses the contemporaneous interest rate as the innovation $\mathbf{v}_t$ after projecting out $\mathbf{x}_t$. This simple approach is intuitive as it amounts to using the residual

[8] One could argue that dropping the October 1947 shock may be preferable as this shock closely coincides with the December 1947 oil shock of Hamilton (1985), see Hoover and Perez (1994).

from a simple backward looking Taylor rule as the predictor for the monetary policy shocks. Note that this implies that $\mathbf{v}_t = p_t^{\perp}$ and the projected interest rate is used twice: once as the endogenous variable in the IV model and once as the innovation, which implies that on the missing periods we are simply using least squares and we correct for the endogeneity bias using the innovation powered inference method.

The baseline model is completed by the selection model for which we use the same specification as in the simulation study: a logit model for s_t with explanatory variables $(1, y_{t+h}^{\perp}, \mathbf{v}_t)'$. The robustness of the choices for the innovations and selection model are investigated below.

Main results

Figure 1 shows the results for the baseline specification; the left panels shows the results for inflation as measured by normalized year-on-year CPI differences and the right panels shows the results for unemployment. The top row show the point estimates $\hat{\theta}_{hn}^{\mathrm{IP}}$ from equation (18) together with the 67% and 95% confidence sets $\mathrm{CS}_{nh}^{\mathrm{IP}}$ computed as in (21). The middle row compares the $\mathrm{CS}_{nh}^{\mathrm{IP}}$ confidence sets to the conventional Anderson and Rubin (1949) confidence sets that use the original narrative series as an instrument. The bottom row, plots the differences in the length of the confidence sets.

Qualitatively our findings are similar as in Romer and Romer (2023): the inflation response shows a small price puzzle but becomes negative after 24 months whereas the unemployment response is positive and peaks after approximately 30 months.

Crucially however, when we compare the innovation powered confidence sets to conventional weak instrument robust confidence sets we find that innovation powering reduces the length width of the confidence set by a large amount. For inflation conventional confidence range between -20 and +20, whereas the innovation powered sets are between -10 and 10. Similar differences are found for the unemployment response. We conclude that without innovation powering little can be learned from the Romer and Romer (2023) series when taking weak identification serious, yet innovation powering largely restores the identification power of the narrative series.

Alternative usage of the narrative instrument

We consider various alternative specifications for the narrative instrument. First, as a variation on baseline approach we consider replacing the occurrences of $z_t = 1$ by residuals from a standard Taylor rule regression that includes inflation, unemployment and the lags of these variables. In this set-up innovation powering allows the researcher to use a standard short run restriction, but only believe its validity on the narrative dates.

The second row of Figure 2 shows the innovation powered dynamic causal effects for this specification (the first row repeats the baseline specification). We find minor differences in terms of the shape of the responses: the unemployment response peaks slightly earlier and the inflation response becomes negative earlier and for a longer period. The most notable change is found in the confidence bands; when using the more informative residuals as instruments the confidence bands shrink considerably.

Second, Hoover and Perez (1994) show that some of the Romer and Romer (1989) dates coincide with the oil shocks identified by Hamilton (1985). To investigate the consequences of such co-occurrence we consider an additional specification where the coinciding October 1979 date is also set to missing. Recall that we already removed the October 1947 dates which is also close to the oil shock of Hamilton (1985).

The third row of Figure 2 shows that removing this date increases the uncertainty in the response substantially. For inflation the 95% confidence bands never exclude zero whereas the 67% bands dip below zero only for a short period. The unemployment response remains significantly positive similar as we found in the baseline specification.

Finally, we consider using only the nine dates identified by Romer and Romer (2023) as observable (omitting October 1947) and set all other dates to missing including the instances where there was no meeting. To ensure that there still is variation in z_t we replace the nine ones by the aforementioned Taylor rule residuals.

Figure 2-row four shows that the treating the remaining zeros as missing does not materially change the results. The confidence bands only increase very slightly when compared to the second row. This is understandable as the zeros provide little information anyway and treating them as missing has little influence. Nonetheless, with only nine observations the central limit theorem that is used to obtain the asymptotic approximation in Proposition 1 becomes questionable and we discuss an modification of the innovation powering method for fixed number of narrative events below.

Alternative selection and innovation models

Innovation powering requires the researcher to specify a model for the selection probabilities and decide which innovations to use. In our baseline approach we used a logit model that is linear in $\mathbf{d}_t^\pi = (1, y_{t+h}^\perp, \mathbf{v}_t)'$. Possibly such specification is too restrictive and we consider a quadratic specification to investigate. On the other hand, the existing literature assumes that narrative events arrive at random which would suggest that a constant would suffice for the selection model. We also explore this option. Further, we used $\mathbf{v}_t = p_t^\perp$ as the innovation whereas here we consider adding future values of inflation and unemployment as additional innovations, see equation (16). Specifically, we add the innovation $y_{t+30}^\perp$ to see if we can improve inference.

The results are shown in Figure 3 where the top row again shows the baseline dynamic causal effects. With quadratic selection probabilities the confidence bands become slightly wider, notably for inflation, but the overall shape remains the same. For a constant selection model the confidence bands shrink slightly, but no major differences occur. The final row shows the dynamic causal effect estimates for $\mathbf{v}_t = (p_t^{\perp}, y_{t+30}^{\perp})'$. We find that the confidence bands reduce quite a bit for inflation but little differences are found for unemployment.

5 Conclusion

We introduced a new inference method —innovation-powered narrative inference—, which combines the benefits of two well known identification approaches by using innovations from time series models to boost the power of narrative methods while preserving consistency. We introduced the method for the purpose of conducting inference on a dynamic causal effect in a linear model specification.

Given that narrative shocks can be used to identify structural equation like the Phillips curve or the Euler equation (Barnichon and Mesters, 2020; Lewis and Mertens, 2023), it is natural to adopt the innovation powering approach for inference in these problems as well. Further, while linearity assumptions are common when estimating dynamic causal effects in macro linearity is not a requirement for innovation powering and the method can be easily adopted when considering state dependent or other nonlinear moment equations.

We have illustrated innovation powering using the narrative monetary series of Romer and Romer (1989, 2023). Clearly the method can be used to improve inference for other narrative studies as well. Moreover, using narrative series as the baseline is not necessary; any instrument that is informative on a subset of good periods can be innovation powered.

References

Abrevaya, Jason, and Stephen G. Donald. 2017. "A GMM Approach for Dealing with Missing Data on Regressors." *The Review of Economics and Statistics*, 99(4): 657–662.

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. *Austerity: When It Works and When It Doesn't.* Princeton University Press.

Anderson, Theodore W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Ann. Math. Statist.*, 20(1): 46–63.

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica*, 59(3): 817–858.

Andrews, Donald W. K., and Patrik Guggenberger. 2019. "Identification- and singularity-robust inference for moment condition models." *Quantitative Economics*, 10(4): 1703–1746.

Andrews, Isaiah, James Stock, and Liyang Sun. 2019. "Weak Instruments in IV Regression: Theory and Practice." *Annual Review of Economics*, 11: 727–753.

Antolín-Daz, Juan, and Juan F. Rubio-Ramírez. 2018. "Narrative Sign Restrictions for SVARs." *American Economic Review*, 108(10): 2802–29.

Bang, Heejung, and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics*, 61(4): 962–973.

Barnichon, Regis, and Geert Mesters. 2020. "Identifying Modern Macro Equations using Old Shocks." *Quarterly Journal of Economics*, 135: 22552298.

Barnichon, Regis, and Geert Mesters. 2023. "A Sufficient Statistics Approach for Macroeconomic Policy." *American Economic Review*, 113(11): 2809–45.

Bi, Huixin, and Sarah Zubairy. 2023. "Public Pension Reforms and Retirement Decisions: Narrative Evidence and Aggregate Implications." *American Economic Journal: Economic Policy*, 15(4): 142–82.

Carriere-Swallow, Yan, Antonio C. David, and Daniel Leigh. 2021. "Macroeconomic Effects of Fiscal Consolidation in Emerging Economies: New Narrative Evidence from Latin America and the Caribbean." *Journal of Money, Credit and Banking*, 53(6): 1313–1335.

Chaudhuri, Saraswata, and David K. Guilkey. 2016. "GMM with Multiple Missing Variables." *Journal of Applied Econometrics*, 31(4): 678–706.

Cloyne, James. 2013. "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom." *American Economic Review*, 103(4): 1507–28.

Cloyne, James, Nicholas Dimsdale, and Natacha Postel-Vinay. 2023. "Taxes and Growth: New Narrative Evidence from Interwar Britain." *The Review of Economic Studies*, rdad081.

Fieldhouse, Andrew J., and Karel Mertens. 2023. "The Returns to Government R&D: Evidence from U.S. Appropriations Shocks." working paper.

Giacomini, Raffaella, Toru Kitagawa, and Matthew Read. 2022. "Narrative Restrictions and Proxies." *Journal of Business & Economic Statistics*, 40(4): 1415–1425.

Giacomini, Raffaella, Toru Kitagawa, and Matthew Read. 2023. "Identification and Inference Under Narrative Restrictions." working paper.

Gil, Paula, Francisco Martí, Richard Morris, Javier J. Pérez, and Roberto Ramos. 2019. "The output effects of tax changes: narrative evidence from Spain." *SERIEs*, 10(1): 1869–4195.

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary austerity? International evidence." *Journal of the European Economic Association*, 12(4): 949–968.

Hall, Alastair R. 2005. *Generalized Method of Moments.* Oxford University Press.

Hamilton, James D. 1983. "Oil and the Macroeconomy since World War II." *Journal of Political Economy*, 91(2): 228–248.

Hamilton, James D. 1985. "Historical Causes of Postwar Oil Shocks and Recessions." *The Energy Journal*, 6(1): 97–116.

Hansen, Lars Peter, John Heaton, and Amir Yaron. 1996. "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business & Economic Statistics*, 14(3): 262–280.

Hoover, Kevin D, and Stephen J Perez. 1994. "Post hoc ergo propter once more an evaluation of does monetary policy matter?in the spirit of James Tobin." *Journal of monetary Economics*, 34(1): 47–74.

Jalil, Andrew J. 2015. "A New History of Banking Panics in the United States, 1825-1929: Construction and Implications." *American Economic Journal: Macroeconomics*, 7(3): 295–330.

Jing Qin, Biao Zhang, and Denis H.Y. Leung. 2017. "Efficient Augmented Inverse Probability Weighted Estimation in Missing Data Problems." *Journal of Business & Economic Statistics*, 35(1): 86–97.

Kang, Joseph D. Y., and Joseph L. Schafer. 2007. "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data." *Statistical Science*, 22(4): 523 – 539.

Lewis, Daniel J., and Karel Mertens. 2023. "Dynamic Identification Using System Projections and Instrumental Variables." working paper.

Little, Roderick J. A., and Donald B. Rubin. 2020. *Statistical Analysis with Missing Data.* John Wiley & Sons, Inc.

Ludvigson, Sydney C, Sai Ma, and Serena Ng. 2017. “Shock Restricted Structural Vector-Autoregressions.” National Bureau of Economic Research Working Paper 23225.

McKay, Alisdair, and Christian Wolf. 2023. “What Can Time-Series Regressions Tell Us About Policy Counterfactuals?” *Econometrica.* forthcoming.

Mertens, Karel, and Morten O. Ravn. 2013. “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States.” *American Economic Review*, 103(4): 1212–1247.

Nakamura, Emi, and Jón Steinsson. 2018. “Identification in Macroeconomics.” *Journal of Economic Perspectives*, 32(3): 59–86.

Newey, Whitney K., and Kenneth D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55(3): 703–708.

Plagborg-Møller, Mikkel. 2022. “Discussion of Narrative Restrictions and Proxies by Raffaella Giacomini, Toru Kitagawa, and Matthew Read.” *Journal of Business & Economic Statistics*, 40(4): 1434–1437.

Plagborg-Møller, Mikkel, and Christian K Wolf. 2019. “Local projections and VARs estimate the same impulse responses.” working paper.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. “Local Projections and VARs Estimate the Same Impulse Responses.” *Econometrica*, 89(2): 955–980.

Ramey, Valerie. 2016. “Macroeconomic Shocks and Their Propagation.” In *Handbook of Macroeconomics.* , ed. J. B. Taylor and H. Uhlig. Amsterdam, North Holland:Elsevier.

Ramey, Valerie A. 2011. “Identifying Government Spending Shocks: It’s all in the Timing.” *The Quarterly Journal of Economics*, 126(1): 1–50.

Ramey, Valerie A., and Matthew D. Shapiro. 1998. “Costly capital reallocation and the effects of government spending.” *Carnegie-Rochester Conference Series on Public Policy*, 48: 145–194.

Ramey, Valerie A., and Sarah Zubairy. 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” *Journal of Political Economy*, 126.

Reinhart, Carmen M, and Kenneth S Rogoff. 2009. “This time is different.” In *This Time Is Different.* princeton university press.

Robins, James M., and Andrea Rotnitzky. 1995. “Semiparametric Efficiency in Multivariate Regression Models with Missing Data.” *Journal of the American Statistical Association*, 90(429): 122–129.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. “Estimation of Regression Coefficients When Some Regressors Are Not Always Observed.” *Journal of the American Statistical Association*, 89(427): 846–866.

Rojas, Diego, Carlos Vegh, and Guillermo Vuletin. 2022. “The macroeconomic effects of macroprudential policy: Evidence from a narrative approach.” *Journal of International Economics*, 139: 103644.

Romer, Christina D, and David H Romer. 1989. “Does monetary policy matter? A new test in the spirit of Friedman and Schwartz.” *NBER macroeconomics annual*, 4: 121–170.

Romer, Christina D., and David H. Romer. 2010. “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks.” *American Economic Review*, 100(3): 763–801.

Romer, Christina D., and David H. Romer. 2017. “New Evidence on the Aftermath of Financial Crises in Advanced Countries.” *American Economic Review*, 107(10): 3072–3118.

Romer, Christina D., and David H. Romer. 2023. “Presidential Address: Does Monetary Policy Matter? The Narrative Approach after 35 Years.” *American Economic Review*, 113(6): 1395–1423.

Rotnitzky, Andrea, James M. Robins, and Daniel O. Scharfstein. 1998. “Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse.” *Journal of the American Statistical Association*, 93(444): 1321–1339.

Rotnitzky, Andrea, Quanhong Lei, Mariela Sued, and James M. Robins. 2012. “Improved double-robust estimation in missing data and causal inference models.” *Biometrika*, 99(2): 439–456.

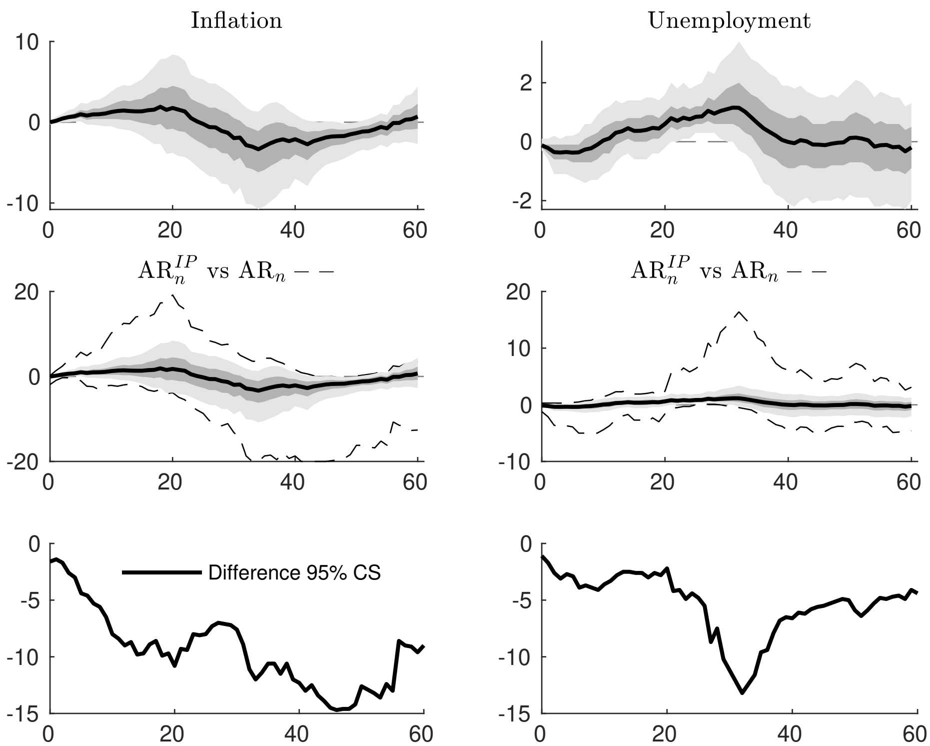
Sims, Christopher A. 1980. “Macroeconomics and reality.” *Econometrica*, 1–48.

Stock, James H., and Jonathan H. Wright. 2000. “GMM with Weak Identification.” *Econometrica*, 68(5): 1055–1096.

Stock, James H., and Mark W. Watson. 2018. “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” *The Economic Journal*, 128(610): 917–948.

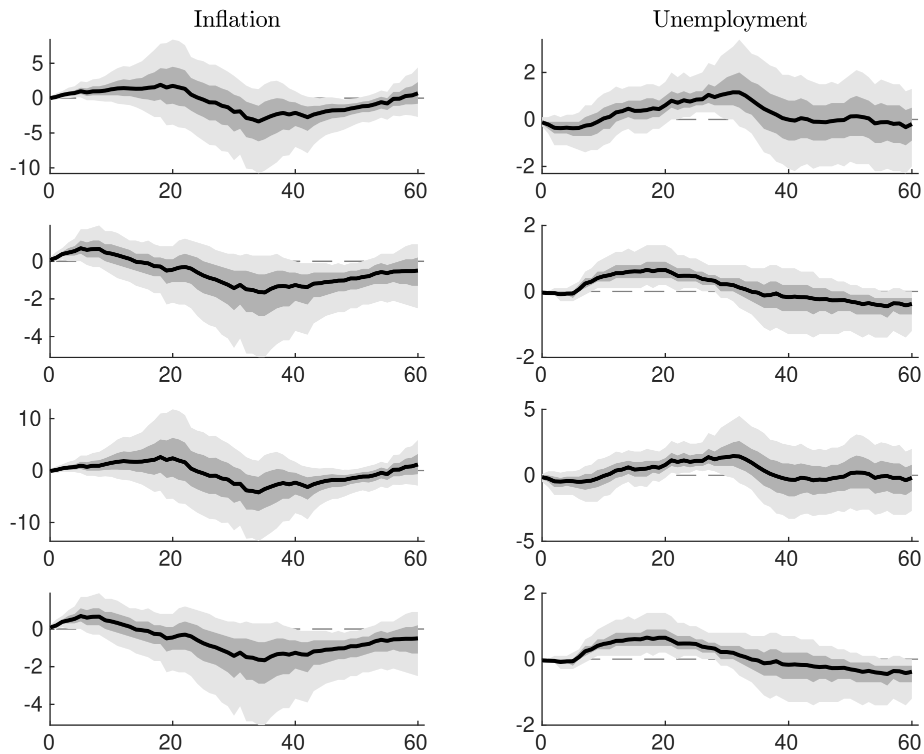
Wooldridge, Jeffrey M. 2007. “Inverse probability weighted estimation for general missing data problems.” *Journal of Econometrics*, 141(2): 1281–1301.

Figure 1: US Monetary Policy: Romer and Romer (1989, 2023)



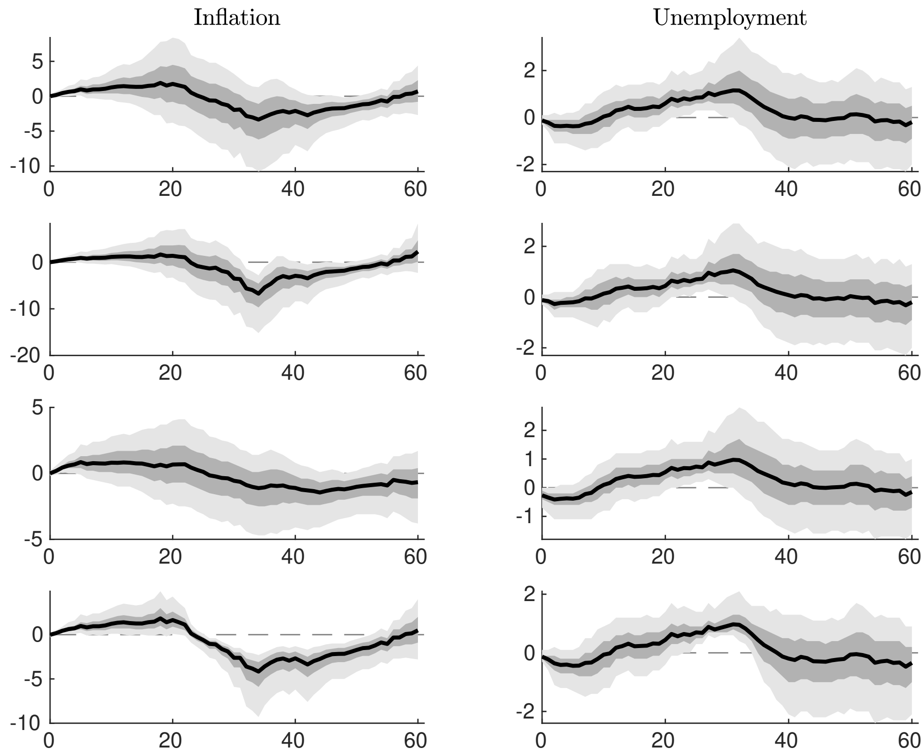
Notes: The top row shows the $\mathrm{CS}_{nh}^{\mathrm{IP}}$ 67% and 95% confidence bands. The middle row compares the $\mathrm{CS}_{nh}^{\mathrm{IP}}$ 95% bands as based on the $\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h)$ statistic, to the conventional Anderson-Rubin bands that use the narrative instrument with zeros. The bottom row report the difference in the length of the 95% bands.

Figure 2: Alternative Narrative Treatments



Notes: The figures show the $\mathrm{CS}_{nh}^{\mathrm{IP}}$ 67% and 95% confidence bands for various specifications of the narrative instrument. The top row shows the baseline from Figure 1. The second row replaces $z_t = 1$ by the residuals from a Taylor rule. The third row drops the October 1947 monetary policy shock. The final row sets all zero to missing except for the nine Romer and Romer dates.

Figure 3: Alternative Selection and Innovation models



Notes: The figures show the $\mathrm{CS}^{\mathrm{IP}}_{nh}$ 67% and 95% confidence bands for various specifications of the selection and innovations model. The top row shows the baseline from Figure 1. The second row considers a quadratic logit model. The third row considers the constant logit model. The fourth row considers $\mathbf{v}_t = (1, p_t^{\perp}, y_{t+30}^{\perp})$.

Table 2: Innovation Powering – Simulation Evidence

	Quarterly Monetary VAR – $n = 200$, $p = 4$							
	π_t^1				π_t^2			
	$h = 0$	$h = 5$	$h = 10$	$h = 15$	$h = 0$	$h = 5$	$h = 10$	$h = 15$
MAE								
$\hat{\theta}_{hn}^{\mathrm{IV}}$	0.224	0.592	0.774	0.799	0.186	0.480	0.656	0.730
$\hat{\theta}_{hn}^{\mathrm{IP}} - (i)$	0.075	0.183	0.254	0.274	0.079	0.202	0.248	0.297
$\hat{\theta}_{hn}^{\mathrm{IP}} - (ii)$	0.241	0.626	0.861	0.876	0.222	0.539	0.738	0.841
ERP								
$\mathrm{AR}_n(\bar{\theta}_h)$	0.041	0.043	0.042	0.045	0.040	0.039	0.041	0.044
$\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) - (i)$	0.067	0.066	0.068	0.072	0.076	0.063	0.065	0.064
$\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) - (ii)$	0.045	0.051	0.052	0.058	0.041	0.039	0.039	0.041
wCS								
CS_{nh}	3.242	4.513	4.814	4.856	3.004	4.358	4.706	4.698
$\mathrm{CS}_{nh}^{\mathrm{IP}} - (i)$	1.379	2.268	2.626	2.733	1.387	2.296	2.664	2.754
$\mathrm{CS}_{nh}^{\mathrm{IP}} - (ii)$	2.560	3.761	4.090	4.167	2.296	3.520	3.890	3.886

	Monthly Monetary VAR – $n = 400$, $p = 12$							
	π_t^1				π_t^2			
	$h = 0$	$h = 20$	$h = 40$	$h = 60$	$h = 0$	$h = 20$	$h = 40$	$h = 60$
MAE								
$\hat{\theta}_{hn}^{\mathrm{IV}}$	0.039	1.333	1.668	1.584	0.033	1.103	1.400	1.341
$\hat{\theta}_{hn}^{\mathrm{IP}} - (i)$	0.041	0.560	0.701	0.665	0.044	0.591	0.685	0.637
$\hat{\theta}_{hn}^{\mathrm{IP}} - (ii)$	0.039	1.329	1.758	1.691	0.034	1.169	1.506	1.377
ERP								
$\mathrm{AR}_n(\bar{\theta}_h)$	0.041	0.048	0.049	0.042	0.042	0.047	0.052	0.046
$\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) - (i)$	0.049	0.054	0.042	0.066	0.047	0.054	0.045	0.062
$\mathrm{AR}_n^{\mathrm{IP}}(\bar{\theta}_h) - (ii)$	0.055	0.058	0.067	0.061	0.043	0.061	0.066	0.065
wCS								
CS_{nh}	0.338	2.296	2.603	2.597	0.301	2.122	2.374	2.386
$\mathrm{CS}_{nh}^{\mathrm{IP}} - (i)$	0.231	1.251	1.416	1.421	0.239	1.255	1.395	1.403
$\mathrm{CS}_{nh}^{\mathrm{IP}} - (ii)$	0.257	1.977	2.267	2.257	0.227	1.811	2.045	2.049

Notes: The table reports the mean absolute error (MAE), Empirical Rejection Probability (ERP) and the width of the confidence interval (wCS) for the standard IV estimator with AR confidence set and the innovation powered versions given in equations (18) and (21).