Towards Robust Local Projections

Italo Morais Santos *†

Latest Revision: August 2024

Abstract

Local projections (LPs) with external instruments have become an increasingly prominent method to identify structural impulse responses in empirical macroeconomics. When instruments are noisy measures of structural shocks, most estimates are subject to attenuation bias. In this paper, I propose a Bayesian two-stage local project that is robust to noisy instruments. Additionally, the proposed method can sample the posterior distribution of the bias term which can be used to assess the exogeneity of the instruments. I apply this method to estimate the impulse responses to U.S. marginal income tax shocks using a medium-sized, yearly-frequency local projection. The findings indicate that while marginal tax shocks are contractionary, their effects on real activity and consumption dissipate within two years, which I attribute to capital-labor displacement effects. Additionally, I identify the effects of monetary policy shocks using high-frequency instruments and show that the instruments adopted by the literature are noisy and the estimated effects of monetary policy are larger once we account for it.

*Visiting Assistant Professor, Boise State University, Department of Economics. Email: italomoraissanto@boisestate.edu

†I'm grateful to Pooyan Amir-Ahmadi, Marcelo Madeiros, Dan Bernhardt, Greg Howard, Shihan Xie and Ji Hyung Lee at the University of Illinois at Urbana-Champaign who provided me with guidance and support. I'm also thankful for the comments and discussions many participants provided in the numerous seminars. In particular, I want to thank Collin Philipps, Sebastian Laumer, Vinicios Sant'Anna, Andrea Carolina Atencio De Leon, Cihang Wang, Humberto Martinez-Garcia. Finally, I also want to thank Min Chul Shin for the many important comments.

1 Introduction

Since the seminal work of Sims (1980), structural vector autoregressions (SVARs) have been the cornerstone for empirical analysis of macroeconomic fluctuations. Estimating the dynamic responses of outcomes such as output levels, inflation, and unemployment to structural shocks is the primary method macroeconomists use for causal inference. Jordà (2005) proposed Local Projections (LP) as an alternative approach to estimate impulse responses (IRs) by directly projecting future values of outcomes onto time series of structural shocks. This approach has been praised for its robustness to model misspecification compared to SVARs, though it comes with reduced estimation efficiency (Stock and Watson (2018), Plagborg-Møller and Wolf (2021) and Li et al. (2022)). Notably, LPs remain valid even under non-invertibility, that is, when there is no possible VAR representation such as in many dynamic stochastic general equilibrium models Sims (2002) and Sims (2012).

In many applications, direct observations of structural shocks are unavailable, requiring the use of proxies to identify structural impulse responses (Jordà et al. (2015)). For identification to be successful, two key assumptions must be satisfied: (1) the instrument must be orthogonal to all other shocks, both contemporaneously and at all leads and lags, and (2) the instrument must be strongly correlated with the structural shock. The validity of these assumptions depends on both the choice of instrument and the model specification.

One often overlooked aspect of identification is the possibility instruments are contaminated by noise or measurement error. In this case, the instrument fails to meet the exclusion restrictions in one stage local projections [1], even if the instrument is uncorrelated with other structural shocks. Specifically, when proxies are noisy measures of structural shocks, the estimated impulse responses are subject to an attenuation bias proportional to the variance of the noise component. In the limit case, it will fail to satisfy the relevance condition, leading to weak instrument problems. Some previous works have addressed this issue by proposing two-stage estimators that are robust to measurement error, such as Stock and Watson (2012), Jordà et al. (2015) and Stock and Watson (2018).

In this paper, I propose a Bayesian Local Projection (LP) regression in two stages, enabling the same robust identification of structural impulse responses (IRs) in the Bayesian case. This approach is grounded in the assumption that the data-generating process is covariance stationary and that fundamental shocks can be retrieved from a non-invertible form of a Gaussian vector moving average (VMA). These assumptions encompass a wide range of models, including linear univariate and multivariate time series, linear Gaussian state-space models, and linear approximations of dynamic stochastic general equilibrium (DSGE) models.

[1] I mean regressions of the outcome of interest in the instrument, known as ”reduced form” regressions in the applied microeconomics jargon.

The parametric structure of LP allows for well-defined Bayesian inference on structural impulse responses. Within this framework, I introduce a Gibbs sampler algorithm for drawing samples from the posterior distribution of structural IRs, the first stage parameter, and the remaining nuisance parameters.

In addition to correcting for bias, my proposed sampler can generate draws from the posterior distribution of the first-stage regression parameter. While classical two-stage least squares can correct for this bias without directly estimating the first-stage parameter, the posterior distribution provides valuable information about the quality of the instrumental variables being used. I demonstrate in this paper that, when the exclusion restriction holds and the instrument is in the same units as the dependent variable in the first stage, the first-stage parameter reflects the proportion of variation in the instrument driven exclusively by the structural shock. This allows researchers to quantify the bias introduced by single-equation regressions. Furthermore, since the bias term is constrained to the unit interval, the posterior distribution offers a way to evaluate a necessary (though not sufficient) condition for the exclusion restriction.

Additionally, my Bayesian Local Projections (LPs) offer several notable advantages. First, it addresses serial correlation in the residuals, which, when accounted for directly in the model, enhances efficiency (See Lusompa (2021)). This improvement helps mitigate one of the primary weaknesses of LPs compared to alternative methods. Second, Bayesian LPs benefit from shrinkage, allowing for a more richly parameterized model compared to traditional LPs. This capability enables Bayesian LPs to incorporate a broader set of control variables, even with small sample sizes. Since instruments are often correlated with other shocks, control variables are crucial for accurate identification. Third, concerns about weak instruments are managed within the Bayesian framework, unlike classic LPs, which require alternative inferential methods. In Bayesian inference of instrumental variable regressions, a weak instrument does not invalidate the posterior sampler; instead, it reduces the rate at which data updates the posterior distribution. Using dispersed priors over impulse responses helps ensure that credible sets accurately reflect the information deficit in weak instrument scenarios. When instruments do not provide sufficient information for meaningful results, Bayesian LPs can incorporate prior information from economic theory. Such prior restrictions refine inference in the presence of weak instruments (Hirano and Porter (2015), Andrews and Armstrong (2017)).

To illustrate the method, I conduct two empirical exercises. First, I apply the two-stages BLP to identify the impulse responses of marginal income tax rates in the U.S. economy from 1948 to 2012. The instrument is constructed from variations in average marginal tax rates driven by tax reforms and revenue acts, which are influenced by lags in aggregate income. In this case, the exogeneity of the instrument hinges on the model's forecasting accuracy. To address this, I propose a medium-sized yearly-frequency Bayesian Local Projection model with 10 variables and 4 lags. By applying shrinkage to the nuisance parameters, this richly parameterized model is able to identify impulse responses despite the limited yearly sample size.

Since this proposed instrument is a noisy measure of true structural tax shocks, the attenuation bias should lies in the unit interval. Using this necessary condition, I assess the quality of instruments commonly used in the literature. By itself, the statutory variation induced by reforms is an endogenous instrument. When interacted with a narrative-driven selection of reforms (Romer and Romer (2009), Montiel Olea and Plagborg-Møller (2021)) the instrument meets the necessary condition.

The results show marginal income tax shocks are contractionary, with small effects on the impact period that builds up during the next two or three years. The effect on aggregate income is long-lasting but fades by the third year in GDP and consumption due to capital-labor displacement effects I'm able to identify.

In the second exercise, I identify monetary policy shocks using high-frequency monetary policy surprises as instruments, following the approaches of Gertler and Karadi (2015), Bauer and Swanson (2023), and others. These instruments are derived from event studies conducted within narrow windows around FOMC announcements. However, since fluctuations unrelated to the announcements are also captured in these event studies, the instruments are contaminated by noise. Using my first-stage regression, I evaluate the degree of contamination across different monetary policy instruments. The share of structural shock variance explained by these high-frequency instruments ranges from 60% to 80%, even after explicitly attempting to control for the noise contamination [2]. According to my estimates, a monetary policy contraction that raises the Dollar-Euro exchange rate by 1 [3] leads to a 0.89% reduction in industrial production, compared to a 0.70% reduction when using the same specification in a single-stage approach.

The structure of the paper is as follows: In Section 2, I derive a framework for evaluating the likelihood of LP models. Section 3 presents the Gibbs sampler for the posterior distribution of impulse responses. Section 4 provides two empirical applications: first, I identify the impact of marginal income tax shocks, and second, I identify the impact of monetary policy shocks. Finally, Section 5 concludes.

Related Literature This paper contributes to the long-standing literature that focuses on identifying and estimating structural impulse responses in macroeconomic time-series data. Bayesian methods are extensively utilized in this field, both for the imposition of informative priors that help to manage the high dimensionality of the models, and for the computational techniques applicable to Bayesian posterior sampling. Notably, methods such as Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) are commonly employed for their flexibility and effectiveness in handling complex model estimation and inference. Canova (2007) provides a comprehensive textbook treatment of Bayesian methods as applied to the case of Vector Autoregressions (VARs). Similarly, DeJong and Dave (2012) offers detailed guidance on the computational techniques widely used in this field, including this paper.

[2] For example, by using measures of economic news releases as in Bauer and Swanson (2023) or teal-book forecast revisions as in Miranda-Agrippino and Ricco (2021).

[3] Equivalent to a 60-90 basis point change in 1-year nominal Treasury bond yields, according to uncovered interest rate parity.

The local projection literature, initiated by Jordà (2005), introduced an estimator for impulse responses that remains valid under certain forms of misspecification, such as omitted control variables or incorrect lag orders. In this framework, however, residuals tend to exhibit serial correlation, which complicates the estimation process as the likelihood function does not have a closed-form solution, making Bayesian approaches more challenging. Several studies have addressed this issue. Plagborg-Møller (2019) suggest utilizing a general moving average representation of stationary time series. Alternatively, Ferreira et al. (2023) proposes correcting the misspecified covariance matrix of the residuals using a "sandwich" estimator, a technique for misspecified Bayesian models introduced by Müller (2013). This paper builds on the approach proposed by Plagborg-Møller (2019) approach, though the two-stage regression introduced here is specifically designed to address the issue of noise contamination in external instruments.

2 Identification of Structural Impulse Responses with Noisy Instruments

I begin by reviewing local projections and their motivation as a direct method for estimating and inferring the causal effects of random, unpredictable shocks that drive business cycle fluctuations. Following Stock and Watson (2018) and Plagborg-Møller (2019), I assume macroeconomic data can has Vector Moving Average (VMA) representation. I demonstrate that, under additional assumptions, the likelihood of local projection residuals is well-defined and their parameters can identify structural impulse responses.

Notation Before proceeding with the discussion I introduce some notation that will be used throughout this paper. A vector's subscript stands for its period. A superscripts stands for a vector's respective entry e.g. x_t^k is the $k-th$ entry of the vector x at time t. The same applies to matrices. For example $A^{i,j}$ is the (i,j)-th entry of A. As is standard in the time series literature, the subscript $s:t$ aggregates the history between s and t in a $(t+1-s)\times K$ matrix i.e. $x_{s:t} = (x_s', x_{s+1}', ..., x_t')'$. The researcher's set of observations at time t is denoted by $\mathcal{D}_t$ [4]. Identity matrices are denoted by I_i where i is its dimension. The "curly" epsilon ε_t always denotes Gaussian white noise with unit variance, that is $\varepsilon_t \sim^{iid} N(0,1)$.

The technique of writing regression models conditionally on known parameters is extensively throughout the derivation of the sampler. Each transformation will be defined in its subsection and transformed data is denoted by $\tilde{y}$. Transformations are unique to each subsection.

[4] Keep in mind this is different from the information set Ω_t, as the latter usually includes non-observables, such as projection residuals and latent states, as well.

2.1 Local Projection Estimation of Impulse Responses

Suppose researcher observes a sample of macroeconomic time series $\mathcal{D}_N = \{y_{1:N}\}$, where y_t is a K-dimensional vector. For simplicity, suppose data has been transformed so that a manner that y_t is co-variance stationary and all deterministic factors, including intercepts, trends, and seasonality have been accounted for. These observations are potentially correlated among themselves both contemporaneously and across time. Researchers and policymakers are often interested in understanding the causal impact of economic policy or external events, such as oil supply shortages or military conflicts, in these aggregates. Once one assumes y_t is stationary, representation theorems such as Wold's can be invoked as way to give the time series a certain amount of structure without making explicit assumptions about its distribution. I formalize this idea in the following assumption:

Assumption 1. *Let y_t be time series observations whose joint distribution is covariance stationary. It admits the following representation:*

$$y_t = A_0\epsilon_t + A_1\epsilon_{t-1} + A_2\epsilon_{t-2} + ..., \tag{1}$$

where $\{A_l\}_{l=0}^{\infty}$ are $K \times Q$, absolutely summable matrices, $\sum_{l=0}^{\infty} |A_l^{i,j}| < \infty \;\forall\; i = 1,...,K$ and $\forall\; j = 1,...,Q$, and ϵ_t follows Q-dimensional Gaussian white noise with diagonal covariance matrix,

$$\Sigma^{i,j} = \begin{cases} \sigma_i^2 \textit{ case } i = j \\ 0 \textit{ otherwise.} \end{cases} \tag{2}$$

Furthermore, assume that not all $A_{l\geqslant 1} = 0_{K\times Q}$, to exclude the trivial case where y_t is white noise.

Q can be equal, larger, or smaller than K. The representation is an extension of Wold's theorem to the case where the researcher observes measures of the true, unknown stochastic process, but not the process itself. The appeal of this representation is its broadness, as all linear, stationary state-space models can be represented according to Assumption (2.1) [5]. Such models includes, for example, log-linear approximations of first-order rational expectations DSGE model (Sims (2002)), VARs (when $K = Q$), and dynamic factor models ($K > Q$).

One of the original motivations of local projections, as proposed by Jordà (2005), is consistency and unbiasedness of impulse response estimates when the VMA representation is non-invertible ($Q \neq K$). On

[5] To see this just solve the measurement equation by backwardly substituting the state-transition. Recall I'm neglecting deterministic and non-dynamic terms in this representation.

an intuitive level, this allows researchers to contemplate structural shocks without a one-to-one mapping between them and observed time series. For example, consider income tax shocks (my first application). Income taxation has several dozen possible measures: average and marginal income tax rates at different income levels, size of the tax base, deduction possibilities, and taxation of different income categories such as wages and capital gains, etc. All these measures could be part of the set of observations $\mathcal{D}$. The same legislative change may perturb several of those measures simultaneously, precluding the plausibility of orthogonal structural shocks for each measure. Defining tax shocks simply as policy changes that are orthogonal to the remaining structural shocks as in Romer and Romer (2010) is more convincing. Tax shocks having a different dimension than policy measures do not present a conceptual challenge to the proposed representation (2.1).

Let ϵ_t^q represent the particular policy or event the researcher is interested in, such as marginal income tax shocks and monetary policy shocks as in the case of this paper. Its causal impact on aggregate outcome k, h periods after its arrival, is commonly defined as the impulse response (3), a simple expected counterfactual variation. The objective of local projection analysis is to estimate and conduct inference over those impulse responses:

$$\beta_h^k \equiv E[y_{t+h}^k | \epsilon_t^q = 1] - E[y_{t+h}^k | \epsilon_t^q = 0]. \tag{3}$$

Now, consider the linear projection of some outcome k over the shock of interest:

$$y_{t+h}^k = \rho \epsilon_t^q + u_{t+h}^k \text{ such that } (y_{t+h}^k - \beta_{k,h}\epsilon_t^q) \perp \epsilon_t^q. \tag{4}$$

According to (1), the projection error u_{t+h}^k is a linear combination of all shocks from $t+h$ to the infinity past except for ϵ_t^q, $u_{t+h}^k = \sum_{l=0}^{\infty} \alpha_l^{k'} \epsilon_{t+h}$. The coefficient vector α_l^k is the k-th line of matrix A_l, except for α_h^k whose q-ith entry is exactly zero. By taking conditional expectations over (4), one can verify ρ identifies the impulse response β_h^k.

Surprisingly even though a general dynamic model was assumed, *mean-identification* of impulse response does not require an explicit estimation of transmission channels between macroeconomic time series or consideration over its dynamic properties. The OLS estimator of y_{t+h}^k over ϵ_t^q is an unbiased and consistent estimator of the impulse responses. In practice, such a regression is not feasible unless observations of the structural shock of interest ϵ_t^q are available. Identification of the impulse responses must be carried out with an external instrument z_t [6]. In section (2.1.2) I will detail how to carry out this form of

[6] Alternatively, one can impose restrictions on the matrices A_l as Plagborg-Møller (2019). However, this paper focuses on the former

identification. For now, I will discuss a second issue.

In addition to the fact ϵ_t^q is generally not observable, Bayesian inference over β_h^k is infeasible, as it is, for a second reason. The projection residuals u_{t+h}^k are serially correlated through the common shock terms $\epsilon_{t+h}^{k\neq q}$, making estimation of their covariance non-trivial. To see this, consider their autocovariance function $s = 0, 1, 2, ...$

$$\gamma^k(s) = \sum_{l=0}^{\infty} \sum_{r=0}^{\infty} \alpha_l^{k\prime} \Sigma^* \alpha_r^k \tag{5}$$

$$\Sigma^* = \begin{cases} \Sigma & \text{if } l = r + s \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note from (5) that: (i) u_{t+h}^k is autocorrelated and of an arbitrarily high order; (ii) the right-hand-side does not depend on t, so that e_{t+h}^k is covariance stationary. Those two observations imply $E(u^k u^{k\prime})$ is a symmetric Toeplitz matrix with unique serial correlation terms equal to the sample size $N - h$. Even if one could derive a posterior sampler for such a matrix, a daunting task in itself, this posterior wouldn't have desirable large sample properties.

For those reasons, evaluating the likelihood through (4) is not practical. However, there is a simpler alternative: evaluate errors of forecasting models of y_t. As I show in the next section, those forecasting errors have a known covariance structure (which is exploited to construct a filter) and the projection of ϵ_t^q over those forecasting errors identify the impulse responses. This is the parametric equivalent of the semi-parametric LPs. The trade-off between the parametric and semi-parametric methods is the standard one - parametric models are more efficient when correctly specified, but less robust to distributional assumptions.

2.1.1 Representation when ϵ_t^q is known.

In this section assume the researcher observe the shock of interest ϵ_t^p in addition to the macroeconomic time series, $\mathcal{D}_N = \{y_{1:N}, \epsilon_{1:N}^q\}$. The goal is to conduct Bayesian inference over the impulse response, β_h^k. Although this assumption is quite strong in general, this exercise is useful for multiple reasons. First, it showcases the issues with LPs parametric representation that are tangential to the identification of β_h^k. Second, this is a necessary step in the derivation for the case of ϵ_t^q is identified with an instrument z_t. Third, there are cases where a first stage is not necessary, for example, when using a shock . In those cases, the following posterior sampler could be used to simulate posteriors of impulse responses.

First of all, let $\Omega_t = \{\epsilon_s\}_{s=-\infty}^{t}$ be an information set at time t which spans the space of ϵ_t [7]. Let the optimal linear h-step ahead forecasts of each individual time series be given by $y^k_{t+h,t-1} = \text{Proj}(y^k_{t+h}|\Omega_{t-1})$ and its forecast error $u^k_{t+h,t-1} = y^k_t - y^k_{t+h,t-1}$. The following theorem yields a parametrization of local projections that can be used in maximum likelihood and Bayesian methods:

Theorem 1. *Let y_t be a stochastic process that follows (1) and let ϵ^q_t be the structural shock series of interest. The parameter of the projection of forecasting errors be $u^k_{t+h,t-1}$ over ϵ^q_t identify the h-step ahead impulse response. That is, let ψ be such that*

$$u^k_{t+h,t-1} = \psi\epsilon^q_t + v^k_{t+h,t-1} \text{ such that } \epsilon^q_t \perp v^k_{t+h,t-1} \tag{7}$$

then, $\psi = E(y^k_{t+h}|\epsilon^q_t = 1) - E(y^k_{t+h}|\epsilon^q_t = 0)$.

Proof. By construction $\psi = E(u^k_{t+1,t-1}|\epsilon^q_t = 1) - E(u^k_{t+1,t-1}|\epsilon^q_t = 0)$. One only needs to show $E(u^k_{t+h,t-1}|\epsilon^q_t) = E(y^k_{t+h}|\epsilon^q_t)$. In the Gaussian case, $\text{Proj}(y_{t+h}|\Omega_{t-1}) = E(y^k_{t+h,t-1})$, which, from (1), is a linear combination of shocks ϵ_t within the information set Ω_{t-1}. Since $\epsilon^q_t \notin \Omega_{t-1}$, $\epsilon^q_t \perp y_{t+h,t-1}$. As result, $E(y^k_{t+h,t-1}|\epsilon^q_t) = E(y_{t+h,t-1}) = 0$. Taking expectations conditional on ϵ^q_t over $y_{t+h} = y_{t+h,t-1} + u_{t+h,t-1}$ concludes the proof. □

At first glance, the parametrization (7) may not seem very useful, as the projection error $v^k_{t+h,t-1}$ is still serially correlated for $h \geqslant 1$. However, the following result establishes $v^k_{t+h,t-1}$ has MA(h) representation hence usual techniques used in likelihood-based and Bayesian analysis of ARMA process can be used.

Corollary 2. *Define $v^k_{t+h,t-1}$ as in Theorem 1. There exist $\phi_{1:h} \neq 0$ such that:*

$$v^k_{t+h,t-1} = e^k_{t+h} + \phi_1 e^k_{t+h-1} + ... + \phi_h e^k_t \tag{8}$$

$$e^k_{t+h} \sim \mathcal{N}(0, \sigma^2_k) \tag{9}$$

Proof. See Appendix. □

Theorem (1) and Corollary (2) together motivate the parametric representation of LPs that I use for the remainder of this work. They imply that if research includes enough controls such that they are able to predict y^k_t to a sufficient degree, the residuals of the LP regression will have exact MA(h) representation. Henceforth, I consider LPs of the form:

[7] This is not to be confused with observation set $\mathcal{D}_t$.

$$y^k_{t+h} = \beta^k_h \epsilon^q_t + \gamma^k w_{t-1} + e^k_{t+h} + \phi_1 e^k_{t+h-1} + ... + \phi_h e^k_t, \quad (10)$$

where w_t is a vector of controls rich enough such that the forecasts $\gamma^k w_{t-1}$ approximate the population projection of the target outcome variable. To formalize this notion, I make it explicit as additional assumption:

Assumption 2. *The vector of controls w_t is such that $\gamma^{k\prime} w_t \approx Proj(y_{t+h}|\Omega_{t-1})$.*

About the representation (10), several remarks are in order:

1. It is valid even when $Q \neq K$ (y_t is non-invertible). As I show in the appendix, $u_{t+h,t-1}$ has a vector moving average (VMA) representation, because it is covariance stationary. However, one can verify e_{t+h} is *not* a linear transformation of ϵ_t. The structural shocks cannot be retrieved from e_{t+h}.

2. Note that a univariate parametrization for $v^k_{t+h,t-1}$ can be achieved though y_t is multivariate. This is indicative that cross-correlations between k and $j \neq k$, as well their auto-correlations, are nuisance parameters, as first argued by Jordà (2005).

3. The fact that $\gamma^k w_{t-1}$ needs to approximate $\text{Proj}(y_{t+h}|\Omega_{t-1})$ in order for the parametrization (10) to be valid motivates my emphasis on usage of models with good forecasting performance, such as medium to large size time series models, models with time-varying parameters, dynamic factor models, etc.

4. The term $\gamma^k w_{t-1}$ is fairly flexible and one can interpret it as the measurement of a linear state-space model. It follows that all the approaches cited above can be easily implemented by taking (10) as the measurement equation.

5. Although forecasting and causal inference are fundamentally different objectives, the former is a prerequisite for the latter when using such parametrization of LPs. Researchers using (10) should not dismiss a model's forecasting performance. While this consideration may seem important only in the likelihood and Bayesian contexts, it is well-known that an equivalent condition is required to relax the (often strong) lag-exogeneity condition of the instrument.

6. Under the Gaussian distributional assumption, the required approximation becomes $\gamma^k w_t \approx E(y_{t+h}|\Omega_{t-1})$. In that case, the validity of the parametrization only relies on *accuracy* of forecasts, not their *precision*. For this reason, traditional model selection criteria, which typically minimize RMSE, cannot be used in this context. For the purpose of identifying impulse response functions with LPs, all emphasis should be put on forecast bias reduction, which suits LPs.

Inference of (10) can be done using any method that allows for moving-averaging residuals. There are several such methods, including textbook applications of Kalman smoothing as in Hamilton (2020). In the Bayesian case with conforming priors, Chib and Greenberg (1994) algorithm can be applied without any modifications. We next consider the case where ϵ_t^q is unobserved but an external instrument z_t for it is available.

2.1.2 Representation with Instruments

The procedure described above is infeasible as the shock of interest ϵ_t^q cannot be observed. However, many structural shocks can be traced back to concrete policy changes or events, and information about such events is available to researchers. In my marginal income tax setting, legislative records and congressional reports provide ample information about changes to tax policy and their motivation. In the case of monetary policy, researchers exploit the fact that policy changes to interest rates must be first announced by the Federal Open Market Committee (FOMC). In both cases, the external information provides the dates on which shocks *do not* arrive and a dummy variable can be constructed to capture this movement. In practice, proxies are constructed to capture as much of the assumed exogenous variation as possible using both narrative accounts and "external data", data that are not included as controls w_t.

Previously, some researchers used to assume those constructed shocks stood as perfect proxies of structural shocks, such as in Romer and Romer (1989). The algorithm described in the last session could be used to implement such an approach by simply assuming $z_t = \epsilon_t^q$. While economic theory can provide good justification for why a proxy approximates the desired structural shock, it cannot ensure the proxy captures the full extent of the structural shock variation, or do so without measurement errors. For this reason, the assumption $z_t = \epsilon_t^q$ is too strong. Since Mertens and Ravn (2012), many researchers have instead treated the proxies constructed from narrative accounts and external data as instrumental variables / external instruments.

Let be z_t be the instrumental variable. Consider the coefficient of $\tilde{y}_{t+h}^k = y_{t+h}^k - \gamma' w_t$ over z_t:

$$\delta_h^k = \frac{E(\tilde{y}_{t+h}^k z_t)}{E(z_t^2)} \tag{11}$$

$$\delta_h^k = \beta_h^k \frac{E(\epsilon_t^q z_t)}{E(z_t^2)} + \frac{E(v_{t+k,t-1}^k z_t)}{E(z_t^2)}. \tag{12}$$

The coefficient δ_h^k does not identify β_h^k without additional identifying assumptions. The first set of identifying assumptions relates to the second term on the right-hand-side:

Assumption 3. *There is a instrument $z_t \in \mathcal{D}$ that it is lead and contemporaneously exogenous:*

1. $E(z_t \epsilon_t^{p \neq q}) = 0$

2. $E(z_t \epsilon_{t+1:h}) = 0.$

Thus, $E(u_{t+k \perp t}^k z_t) = 0$ *for every* $k = 1, ..., K$.

Assumption 3 are the exogeneity conditions for instrumental variable regressions in the context of local projections, though the lag-exogeneity condition is unnecessary. Both conditions must be met through the construction of the instrument. Under Assumption 3 δ_h^k simplifies to:

$$\delta_h^k = \beta_h^k \pi, \tag{13}$$

where $\pi = \frac{E(\epsilon_t^q z_t)}{E(z_t^2)}$. That is, the projection coefficient of a suitable instrument over $y_{t+h \perp t}^k$ identifies the product of the desired impulse response times the projection coefficient of the instrument over the unobserved structural shock. Substituting a suitable instrument z_t for the shock in the algorithm described in subsection 2.1.1 identifies β_h^k, but with an unknown scale. Bayesian inference of structural impulse responses requires joint inference of both β_h^k and π. To do so, I normalize the impulse responses as follows:

Assumption 4. *Let y_t^1 be the first variable in the vector y_t. Its impulse response at impact-period is normalized to one, $\beta_0^1 = 1$, such that*

$$y_t^1 = \pi z_t + \gamma_0^{1\prime} w_t + e_t^1. \tag{14}$$

This is a real assumption and not just normalization because it requires $\beta_0^1 \neq 0$. That is, we are assuming ϵ_t^q impulse responses are not zero for the policy variable. Note normalization does *not* need to be done with respect to y_t^q. For example, suppose a small trivariate VAR of government revenues ($k = 1$), government spending ($k = 2$), and some measure of real activity ($k = 3$). We could potentially normalize β_0^1 even if when using an instrument for spending. In that case, we interpret impulse responses as changes caused by spending shocks that raise revenues by one unit. This makes our normalization scheme exactly the same as that for proxy/instrumental variables VARs. One practical consequence of this feature is that one can draw inferences with respect to ϵ_t^q even when y_t^q is not available.

To perform Bayesian inference over β_h^k using an instrument, one needs to estimate π as well. Evaluating equation (14) at $t+h$ yields:

$$y_t^1 = \pi z_t + \gamma_0^{1\prime} w_t + e_t^1 \tag{15}$$

$$y_{t+h}^k = \beta_h^k \pi z_t + \gamma_h^{k\prime} w_t + e_{t+h}^k + \phi_1^k e_{t+h-1}^k + ... + \phi_h^k e_t^k \tag{16}$$

$$\begin{pmatrix} e_{t+h}^1 \\ e_{t+h}^k \end{pmatrix} \sim iidN(\mathbf{0}, \Sigma_h^k) \tag{17}$$

Equation (15) is the first stage. Equation (16) is the second stage. Together with the distributional assumption (17), they form the empirical model.

First, notice, that due the exogeneity conditions, π is the only common parameter across both linear projections. This local projection can be parametrized as a restricted reduced form instrumental variable regression with unusual dynamic structure. This parameterization, as opposed to estimating δ_h^k and dividing by π, allows to implement priors that are approachable for weak instrument inference: (i) the model directly identifies the impulse response β_h^k, allowing dispersed priors over those; (ii) the restricted reduced form also avoids the "divide by zero" problem that arises when the instrument is not valid, $\pi = 0$, or close to zero [8].

Second, notice the correlation between e_{t+h}^1 and e_{t+h}^k is *not* what gives rise to endogeneity in our model, in contrast with traditional formulations of two-stage regressions. That is, even if y_t^1 and y_t^k have no shocks in common other than the structural shock and Σ_h^k is a diagonal matrix, the projection coefficient of y_t^k on y_t^1 wouldn't identify β_h^k unless $e_t^1 = 0$ and ϵ_t^q is the only structural shock responsible for y_t^1 variation. This identification scheme would be akin to recursive identification.

Finally, I do not assume $\pi \neq 0$ *a priori*. A weak instrument does not contradict the empirical model (15)-(17), although it reduces the identifying variation used to update the posterior of β_h^k.

[8] See Kleibergen and Zivot (2003) for details of alternative Bayesian parametrizations of IV regressions. Notice my priors are differ from theirs Bayesian two-stages, although both models are in restricted reduced form.

3 Bayesian Inference

In this section I describe the posterior sampler of equations (15)-(17) parameters. The residuals with moving average dynamics $e_t^k, ..., e_{t+h-1}^k$ will be treated as latent states, s_{t+h}. Let $\phi = (\phi_1, ..., \phi_h)'$, $s_{t+h} = (e_t^k, ..., e_{t+h-1}^k)'$, $S = s_{T:(T+h)}$, $\gamma = (\gamma_0^1, \gamma_h^k)'$. To simplify notation I will also omit k and h from β and Σ. The target posterior distribution is given by $p(\pi, \beta, \gamma, \phi, \Sigma, S|\mathcal{D})$.

The moving average dynamics and the non-linearity of the RRF makes it difficult to derive an expression of (15)-(17) likelihood and a direct sampler for the posterior distribution. I instead use Gibb's sampler. The partitioning of the parameter space and the respective conditional posteriors are

$$S|\pi, \beta, \gamma, \phi, \Sigma, \mathcal{D} \tag{18}$$

$$\Sigma|\pi, \beta, \gamma, \phi, S, \mathcal{D} \tag{19}$$

$$\phi|\pi, \beta, \gamma, \Sigma, S, \mathcal{D} \tag{20}$$

$$\beta, \gamma|\pi, \phi, \Sigma, S, \mathcal{D} \tag{21}$$

$$\pi|\beta, \gamma, \phi, \Sigma, S, \mathcal{D}. \tag{22}$$

Now I proceed to derive each conditional posterior distribution and as briefly discuss the choice of prior distributions.

3.0.1 Step 1: $p(S|\pi, \beta, \gamma, \phi, \Sigma, \mathcal{D})$

Given the conditionals, evaluate moving average residuals $u_{t+1}^1 = e_{t+h}^1$ and $u_{t+h}^k = e_{t+h}^k + \phi_1 e_{t+h-1}^k + ... + \phi_h e_t^k$. They form a bivariate restricted VMA(h) model where one of the series is white noise. Kalman smoother provides draws from the posterior $p(S|\pi, \beta, \gamma, \phi, \Sigma, \mathcal{D})$. Assume a dispersed, conforming prior for initial conditions, $S_0 \sim \mathcal{N}_h(0, I_h a_0^{S\,2})$, where $a_0^{S\,2} \to \infty$. See Durbin and Koopman (2012) for details.

3.0.2 Step 2: $p(\Sigma|\pi, \beta, \gamma, \phi, S, \mathcal{D})$

Conditional on the state variables, the iid reduced form shocks can be evaluated:

$$e_{t+h}^1 = y_{t+h}^1 - \pi z_{t+h} - \gamma_0^1 w_{t+h} \tag{23}$$

$$e_{t+h}^k = y_{t+h}^k - \pi z_{t+h} - \gamma_h^k w_t - \rho_1 e_{t+h-1}^k - ... - \rho_h e_t^k \tag{24}$$

Under the assumption of Gaussian errors and with Inverse Wishart priors, the conditional posteriors has closed form solution:

$$p(\Sigma) \sim \mathcal{IW}(V_0, N_0) \tag{25}$$

$$p(\Sigma|\pi, \beta, \gamma, \phi, S, \mathcal{D}) \sim \mathcal{IW}(V_0 + e'e, N_0 + T - h), \tag{26}$$

where $e = (e^1_{h:(t+h)}, e^k_{h:(t+h)})'$.

3.0.3 Step 3: $p(\phi|\pi, \beta, \gamma, S, \Sigma, \mathcal{D})$

Under this conditioning set, evaluate the model as

$$y^1_{t+h} - \pi z_{t+h} - \gamma^1_0 w_{t+h} = e^1_{t+h} \tag{27}$$

$$y^k_{t+h} - \pi \beta^k_h z_t - \gamma^k_h w_t = \phi s_{t+h} + e^k_{t+h} \tag{28}$$

Since e^1_{t+h} is conditionally observed through equation (27), I rewrite the model in a single equation conditional on e^1_{t+h} realization:

$$y^k_{t+h} - \pi \beta^k_h z_t - \gamma^k_h w_t = \phi s_{t+h} + e^k_{t+h}|e^1_{t+h}, \tag{29}$$

where $e^k_{t+h}|e^1_{t+h} \sim \mathcal{N}(\mu_{k|1}, \sigma^2_{k|1})$, $\mu_{k|1} = \Sigma^{1,2}\Sigma^{2,2^{-1}}$ and $\Sigma^{k|1} = \Sigma^{1,1} - \Sigma^{1,2}\Sigma^{2,2^{-1}}$. Standardizing equation (29) yields

$$(y^k_{t+h} - \pi \beta^k_h z_t - \gamma^k_h w_t - \mu_{k|1})\sigma^{-1}_{k|1} = \phi(s_{t+h}\sigma^{-1}_{k|1}) + \varepsilon \tag{30}$$

$$\tilde{y}_{t+h} = \phi \tilde{s}_{t+h} + \varepsilon. \tag{31}$$

Equation (31) is Bayesian linear regression with unit variance. Since I explicitly assumed non-invertibility, ϕ is not uniquely identified in the entire parameter space. To avoid computational problems, the support needs to be truncated to invertible region with an indicator prior distribution (32), as suggested by Chib and Greenberg (1994).

$$p(\phi) \sim N(\phi_0, A_0^{-1})\mathbb{1}(\phi \in R_\phi) \tag{32}$$

$$p(\phi|\pi,\beta,\gamma\Sigma,S,\mathcal{D}) \propto p(\tilde{y}_{1:T+h}|\phi,\pi,\beta,\gamma\Sigma,S,\mathcal{D}-y_{1:T+h})N(\phi_0, A_\phi^{-1})\mathbb{1}(\phi \in R_\phi), \tag{33}$$

where $\mathbb{1}$ stands for an indicator function and R_ϕ is the invertible region. The resulting posterior distribution given by (33) is non-standard and draws must be generated using the Metropolis-Hasting algorithm. I use the posterior of regression coefficients of (31) when $\phi \in R_\phi$ as the proposal density, $N((\tilde{S}'\tilde{S} + A_\phi)^{-1}\tilde{S}'\tilde{y}, (\tilde{S}'\tilde{S} + A_\phi\phi_0)^{-1})\mathbb{1}(\phi \in R_\phi)$.

3.0.4 Step 4: $p(\beta,\gamma|\pi,\phi,\beta,\Sigma,S,\mathcal{D})$.

First adjust the model to the conditional set:

$$y_{t+h}^1 - \pi z_t = \gamma_0^1 w_{t+h} + e_{t+h}^1 \tag{34}$$

$$y_{t+h}^k - \phi s_{t+h} = \beta_h^k(\pi z_t) + \gamma_h^k w_t + e_{t+h}^k, \tag{35}$$

equation (34) is simply a system of seemingly unrelated regressions with known covariance Σ_h^k. Such models have conforming, Gaussian priors given by equation (36). The posterior immediately follows from the model standardized in matrix form

$$p(\beta,\gamma) \sim N(\mu_0, A_\mu^{-1}\mu_0) \tag{36}$$

$$p(\beta,\gamma|\pi,\phi,\beta,\Sigma,S,\mathcal{D}) \sim N((\tilde{X}'\tilde{X} + A_\mu)^{-1}(\tilde{X}'\tilde{y} + A_\mu\mu_0)), \tag{37}$$

where

$$D = (\Sigma_h^k \otimes I_{T-h})^{-1/2} \tag{38}$$

$$\tilde{y} = D\begin{pmatrix} y_{1:T+h}^1 \\ y_{1:T+h}^k \end{pmatrix} \tag{39}$$

$$\tilde{X} = D\begin{pmatrix} w_{1:T+h} & 0 & 0 \\ 0 & w_{1:T} & \pi z_{1:T}, \end{pmatrix} \tag{40}$$

3.0.5 Step 5: $p(\pi|\beta,\phi,\gamma\Sigma,S,\mathcal{D})$

For draws of the first stage parameter π, one can use a similar device as in step 4. However, the system of equations is not seemingly unrelated since both regressions, by assumption, share the same parameter.

$$y^1_{t+h} - \gamma^1_0 w_{t+h} = \pi z_{t+h} + e^1_{t+h} \tag{41}$$

$$y^k_{t+h} - \gamma^h_h w_t - \phi s_{t+h} = \pi(\beta^k_h z_t) + e^k_{t+h}. \tag{42}$$

The system formed by equations (41) and (42) can be rewritten as single linear regression by stacking z_{t+h} and $\beta^k_h z_t$:

$$\tilde{y} = \tilde{Z}\pi + \tilde{e}, \tag{43}$$

where,

$$D = (\Sigma^k_h \otimes I_{T-h})^{-1/2} \tag{44}$$

$$\tilde{y} = D \begin{pmatrix} y^1_{1:T+h} - \gamma^1_0 w_{t+h} \\ y^k_{1:T+h} - \gamma^h_h w_t - \phi s_{t+h} \end{pmatrix} \tag{45}$$

$$\tilde{Z} = D \begin{pmatrix} z_{1:T+h} \\ \beta^k_h z_{1:T}. \end{pmatrix} \tag{46}$$

The simplicity of equation (43) is what gives almost total flexibility in the choice of prior distribution $p(\pi)$ as it is always straightforward to derive a sampler with non-conforming priors for univariate, linear regressions. For example, priors with bounded support can be used to impose sign restrictions on this first-stage parameter. One of the central arguments of this paper is that economic theory that justifies instrument validity can also be used to obtain likely ranges for π. For this reason I leave π prior unspecified, since it ultimately depends on the context of the application. Next, I discuss some common cases.

3.1 Connecting Instrument Designs to First Stage Parameters π

Consider the example of narrative identification with nothing but sign indicators:

$$
z_t = \begin{cases} -1 \text{ if } g(\epsilon_t^q + m_t) < \underline{a} \\ 0 \text{ if } \underline{a} \leqslant g(\epsilon_t^q + m_t) \leqslant \bar{a} \\ 1 \text{ if } g(\epsilon_t^q + m_t) > \bar{a}, \end{cases} \tag{47}
$$

where the constructed external instrument z_t is meant to identify the shock ϵ_t. Similar examples have been proposed by Plagborg-Møller and Wolf (2021) and Boer and Lütkepohl (2021). Sufficiently large shocks are connected to events recorded by the researcher, such as oil supply disruptions, war-time spending or tax reforms. If the function $g()$ is increasing, researchers guess the sign of the shocks correctly, on average. The function $g()$ and the bands $(\underline{a}, \bar{a})$ determine how often shocks are recorded, as well as possible skewness in the recording [9]. Finally, the recording might be contaminated by measurement error m_t. Under (47) it can be shown the first stage parameter is

$$
\pi = \frac{E(\epsilon_t | g(\epsilon_t^q + m_t) > \bar{a}) - E(\epsilon_t^q | g(\epsilon_t^q + m_t) < \underline{a})}{p(g(\epsilon_t^q + m_t) < \underline{a}) + p(g(\epsilon_t^q + m_t) > \bar{a}) - (p(g(\epsilon_t^q + m_t) < \underline{a}) - p(g(\epsilon_t^q + m_t) > \bar{a}))^2} \geqslant 0, \tag{48}
$$

where $p()$ is the joint probability density function of (ϵ_t^q, m_t). It is positive than zero since both probabilities in the denominator are between 0 and 1, while the numerator is strictly positive because $E(\epsilon_t^q | g(\epsilon_t^q + m_t) \geqslant \bar{a}) > E(\epsilon_t^q | g(\epsilon_t^q + m_t) \leqslant \underline{a})$. Instrument weakness in this context can be characterized by $(\underline{a}, \bar{a})$ being so wide that few shocks are recorded or measurement error being so high that $E(\epsilon_t^q | g(\epsilon_t^q + m_t) \geqslant \bar{a}) \approx E(\epsilon_t^q | g(\epsilon_t^q + m_t) \leqslant \underline{a})$.

It is also possible to find an upper bound for π. The highest performance sign-based external instrument for ϵ_t^q is when its sign is always recorded and done so correctly. In other words, $\underline{a} = \bar{a} = 0$ and $m_t = 0$. In that case, the first stage parameter simplifies to

$$
\pi = E|\epsilon_t^q|,
$$

which is strictly positive, finite under mild assumptions and proportional to ϵ_t in scale. For example, if structural shock is Gaussian, $|\epsilon_t^q|$ is half-normal distributed and $\pi \propto \sigma$. Unfortunately, because $E|\epsilon_t^q|$ is not observable, it's not possible to impose any more restrictions on π parameter space from (47) alone. This is a direct consequence of using instruments that are not of the same scale as ϵ_t^q.

[9] For example, oil supply shocks would be mostly negative, so that $\bar{a} \to \infty$.

Recall the unit-scale normalization fixes the scale of the structural shock ϵ_t^q as the scale of y_t^1. In many applications, instruments can be constructed to be in that same scale as well. In that case the instrument design is given by:

$$z_t = \begin{cases} 0 \text{ if } \underline{a} \leqslant g(\epsilon_t^q + m_t) \leqslant \bar{a} \\ e_t + m_t \text{ otherwise} . \end{cases} \tag{49}$$

That is, for small shock realizations, no narrative is recorded. When large shocks realize, the instrument is constructed to capture its full effect, but can still be contaminated by measurement error m_t. If instruments are given by (49) the first stage parameter is

$$\pi = \frac{E({\epsilon_t^q}^2|G(\epsilon_t^q, m_t)) + E(\epsilon_t^q m_t|G(\epsilon_t^q, m_t))}{E({\epsilon_t}^2|G(\epsilon_t, m_t)) + 2E(\epsilon_t m_t|G(\epsilon_t, m_t)) + E(m_t^2|G(\epsilon_t, m_t))}, \tag{50}$$

where $G(\epsilon_t, m_t)) = \{g(\epsilon_t, m_t) \notin [\underline{a}, \bar{a}]\}$. Since instrument is more likely to record when ϵ_t^q and m_t have the same sign, $E(\epsilon_t^q m_t|G(\epsilon_t^q, m_t)) > 0$, henceforth $\pi < 1$.

In all these examples, the sign of π is derived from the theoretical assumptions. Any choice of prior distribution with support in the real line incorporates this information. Uniform priors in particular are convenient as the posterior distribution of π has a closed-form solution

$$p(\pi) \sim U(0, \pi_0) \tag{51}$$

$$p(\pi|\beta, \gamma, \phi, \Sigma, \mathcal{D}) \sim N_{[0,\pi_0]}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{y}, \tilde{e}'\tilde{e}), \tag{52}$$

where $N[0, \pi_0](\mu, \sigma^2)$ is the Gaussian distribution truncated at $[0, \pi_0]$.

4 Empirical Applications

To illustrate the Bayesian LP-IV, I use it to estimate impulse responses of economic policy in two different applications. First application I identify marginal income tax shocks in US economy, as in Mertens and Ravn (2012) and Mertens and Montiel Olea (2018). Second, I estimate the effect of monetary policy shocks, also in the United States economy, as in Gertler and Karadi (2015) using the BLP-IV.

4.1 Marginal Income Tax Shocks

A central question in the fiscal policy literature is to what extent marginal income tax rates influence individual decisions to work, invest, and innovate. Governments that find themselves with high debt and slow growth often raise taxes to boost tax revenue, even though raising taxes could hurt activity in the short run. Income tax cuts are also a common counter-cyclical policy, such as the American Tax Cuts and Jobs Acts of 2017.

The empirical literature studying US individual tax returns [10] finds that changes in marginal income only have a modest impact on aggregate income. This is puzzling, as the empirical macro literature often finds that average marginal tax rates (AMTR) are an important factor in explaining fluctuations in economic activity and unemployment. Mertens and Montiel Olea (2018) made a significant contribution by bringing individual tax return data into dynamic macro models and their evidence largely supports the findings of macroeconomics literature — marginal tax shocks do affect activity.

In this application, I extend the analysis of AMTR shocks in several ways. To better explain the contributions of the Bayesian LP-IV, let me first introduce the instruments used.

4.1.1 AMTR Instrument

The instrumental variable/external instrument z_t used to capture exogenous variation in income tax policy is given by

$$z_t = d_t \times \frac{1}{M_t} \sum_{i=1}^{M_t} (\tau_t(\text{income}_{t-1,i}) - \tau_{t-1}(\text{income}_{t-1,i})), \tag{53}$$

where $income_{t-1,i}$ is payroll taxes plus individual taxable income, defined as all sources of income excluding capital gains and government transfers.[11] The function τ_t income tax schedule at period t and d_t is a dummy variable. This function maps the declared individual taxable income to its marginal income tax during tax year t. The quantity $(\text{income}_{t-1,i}) - \tau_{t-1}(\text{income}_{t-1,i}))$ measures the marginal income tax variation caused by only changes in the schedule between periods $t-1$ and t. This quantity was measured for a sample of M_t individuals. The total average of those measures is the statutory variation (SV) in tax policy which captures mechanical variation in AMTR during policy changes, discounting the effects of fluctuations in the tax base. These measures are provided in Mertens and Montiel Olea (2018) replication files.

[10] See Saez et al. (2012) for a survey.
[11] This is the definition most commonly used in the taxable income literature. See Piketty and Saez (2003).

Using the SV as instrument faces two potential identification problems. First, some tax reforms may not work as an appropriate basis to construct tax shocks if the reforms are anticipated or are done in response to contemporaneous events. The literature has long used Romer and Romer (2009) narrative of US tax reforms as guidance to select suitable reforms, which is done through the d_t interaction term.[12] Unfortunately, a consensus regarding which reforms should be included is hard to come by and as I show, results are sensitive to this choice. For this reason, I run the models with three different instrumental variables:[13]

- IV1: Includes all statutory variations. That is, $d_t = 1$ for all sample.
- IV2: Includes all tax shocks classified by Romer and Romer (2009) as exogenous changes.
- IV3: A subset of IV2, include only reforms classified as exogenous by Mertens and Montiel Olea (2018).

A second identification issue comes from instruments being constructed from lagged taxable income which is expected to be strongly correlated with other lagged shocks. Without a suitable set of controls and an adequate number of distributed lags, the instrument violates the lag-exogeneity condition. A rich set of controls is a challenge in this context, as income taxation has a yearly frequency, resulting in a relatively small sample ($N = 65$). It is also well known in the micro panel literature that just controlling for lags of income does not correct these biases. [14] The reason is that several macro level shocks that are correlated with the regression outcome y^k_{t+h} may only affect aggregate income with a delay. For example, the effect of productivity shocks on aggregate income tend to "build over time". Since the contemporaneous correlation between the two is not strong, lags of taxable income do not account for this. To address this concern I use a relatively large set of controls together with the shrinkage priors described in the previous session. An additional advantage of this larger model approach is that we are also able to identify a broader range of effects.

4.1.2 Instrument Validation

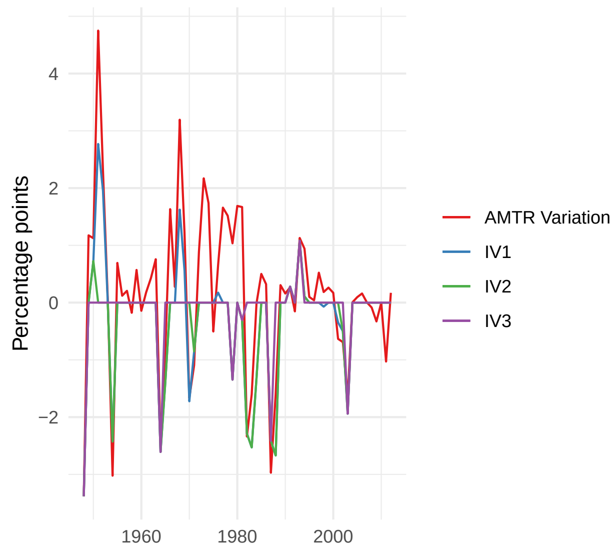
To showcase the heuristic presented in section 3.1, I propose an explicit model for (z_t, ϵ^q_t). Consider again how the instrument is constructed: whenever the narrative, modeled through the dummy d_t, identifies a revenue act or episode of tax reform, the external instrument is exactly equal to the statutory variation. In that case, differences between ϵ^q_t and z_t must come from either:

[12] They attribute four possible justifications to each tax reform: (a) response to current or future adjustments to government spending; (b) offsetting cyclical fluctuations; (c) addressing long term debt growth; (d) stimulate investment and/or long-term growth. The literature considers the last two as candidates for tax shocks.

[13] See the appendix for a table detailing all the included tax reforms and revenue acts.

[14] See Weber (2014) for alternative exposition of this point.

Figure 1: AMTR and Instruments



Time series plot of U.S. Average Marginal Tax Rate (All statutory variations) and each of the three instruments.

1. An endogenous reform is accidentally included in the dummy.

2. A shock arrives but is not included in the dummy

3. The statutory variation mismeasures the structural shock due sampling error.

The second source is unlikely to be important, as all exogenous changes in AMTR must do due either through tax reforms or revenue acts — and a comprehensive list of those changes are readily available. The third source, in contrast, is expected as both SV measures provided by Mertens and Montiel Olea (2018) only use a sample of the population. Additionally, tax calculators used to compute $\tau_t(income_{t-1})$ might be imprecise. Such measurement errors muddy identification if z_t were used directly but not in a two-stage regression.[15] The first source of discrepancy though violates the identification assumption. The instrument validity hinges upon no endogenous reform being included in the dummy.

A good AMTR instrument would be subjected only to the third possibility. In that case, the relation between shocks and instruments is given by:

[15] If they are assumed to be orthogonal to structural shocks and samples are large.

$$z_t = \epsilon_t^q d_t + m_t \tag{54}$$

$$m_t \sim N(0, \sigma_m^2), \tag{55}$$

where m_t is measurement error. Since the first stage parameter is the projection parameter of $\text{Proj}(\epsilon_t^q | z_t)$. It is easy to see that:

$$\pi = \frac{\sigma_q^2}{\sigma_q^2 + \sigma_m^2} \in [0, 1]. \tag{56}$$

If the tax instrument is constructed correctly, the first stage parameter has to be less than one and it can be interpreted the proportion of the variance of the instrument explained by the structural shock. This provides a *necessary* condition to validate an instrument based on posterior distribution of π, as probability mass on regions $\pi > 1$ indicates an endogenous reform was included. That is the posterior distribution of the first-stage is identifying not just π but, in addition, the bias term in (12). Note that the sign of this bias term has to be positive, as SV is always positively correlated with z_t.

Figure 2: First Stage Parameter Distributions

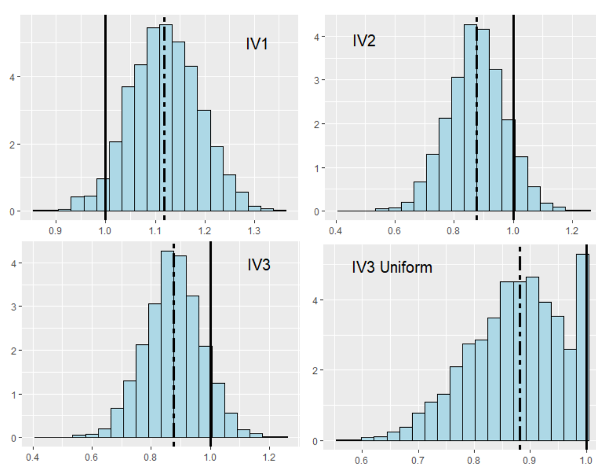


Figure 3: First stage parameter posterior distribution, obtained by running the first stage with each instrument separately, with all control variables (as described in the following section). Top left and right and bottom left all have the same disperse prior distributions for π, bottom right has $U(0,1)$ prior.

The graphs in Figure (2) show the posterior distribution of the first-stage parameter π for each of the three different instruments given the same dispersed uniform prior over $\pi \sim U(0, 10)$ for all the cases, except for the bottom right where I use a $U(0,1)$ prior. According to the posterior distributions, the probability of $\pi > 1$ when z_t is the IV1 is 88%. The evidence points the first instrument is endogenous. Some narrative criteria to select the right reforms are needed. The posteriors for the remaining two instruments are close. This is expected as (56) is invariant wtih respect to the choice of reforms. While the stability of these posteriors is evidence the instruments are valid (otherwise endogenous reforms included in IV2 but not IV3 would shift the posterior towards 1), they do not amount to a sufficient condition, as the evidence is still consistent with bias together with a higher level of signal-to-noise ratio.

Regardless, this analysis is still informative of the quality of the instruments. The point estimates π for both IV2 and IV3 are approximately 0.83 and it is unlikely the sampling error on the construction of SVs measures could amount to more.

Alternatively, since the theoretical restriction is that $\pi \in [0,1]$, the more informative prior $U[0,1]$ could be used to eliminate the probability mass on the region $\pi > 1$. I advise against such practice, as the role of π in the econometric modeling is simply to give the second-stage estimates adequate scale. For this reason, as long as posterior π has enough mass far from zero, specific values of π shouldn't matter. That said, impulse responses for that case were estimated and provided in the online appendix.

4.1.3 Data and Model Specification

I run two BLP-IVs for each of the three instruments. The sample has an annual frequency and covers the period between 1948 to 2012.

I focus on two specifications. The first is a small four-variable system that includes the outcomes: aggregate taxable income, gross domestic product, effective federal funds rate, and unemployment rate. The policy measure is the AMTR time series. Controls include two lags of each outcome[16].

The second specification is medium-scale BLP-IV that includes all of the previous four outcomes as well as: household consumption, investment, and CPI inflation. The control variables include three lags of each outcome and additionally three lags of each: (1) federal government debt, real stock prices, government spending, Gertler and Karadi (2015) monetary policy surprise shocks, Ramey (2011) fiscal news shocks, Arezki et al. (2017) international oil supply news shock and a dummy with all NBER recessions. Posteriors distributions were simulated from 20,000 iterations of the Gibb's sampler algorithm described in the previous session.

4.1.4 Prior Distributions

Priors of the impulse responses β_h^k are assumed dispersed $N(0,100)$. The prior covariance matrix is the usual $IW^{-1}(\Sigma_0, 3)$ where Σ_0 is a diagonal matrix containing the variances of y_t^k, y_t^1. Priors over π, as de-

[16] All stock variables have been normalized so that IRs are in %.

scribed above, are $U(0,10)$. Finally, priors over the remaining auto-regressive, controls and moving average parameters are built assuming that each time series follows an AR $y = \mu_0 + 0.85Ly + e_t$ process, with μ_0 being the long-run mean of the respective time series. The tightness of each of those priors ($\alpha_0 = 0.89$) was chosen to reduce the root mean squared errors compared to classical LP-IV.

4.1.5 Main Results

The panels of the left of Figures (4) and (5) showcase the impulse responses using IV2 and the right panels present the results using IV1, using the medium-scale BLP-IV in both cases. Results for the model with IV3 were omitted as they don't differ substantially from the model with IV2. All figures display the median of the posterior IRF distribution and their 90% equal tail credible sets. My preferred estimate is given by the model with IV2.

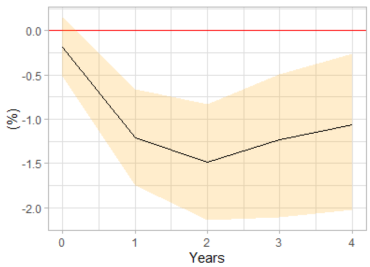
4.1.6 Results from Model with IV1

Across all seven outcomes, the effects in the same year of the reform implementation are small and are only statistically different from zero for GDP (-0.63%). This finding is consistent with the micro-level literature on marginal income tax changes who generally find small and non-significant effects on aggregate income. These results are usually attributed to either identification issues, difficulty in finding good model specifications, or myopic behavior by low and middle-income households, which may delay their reaction to changes in tax incentives only after they pay their taxes at least once. My results corroborate the last possibility.

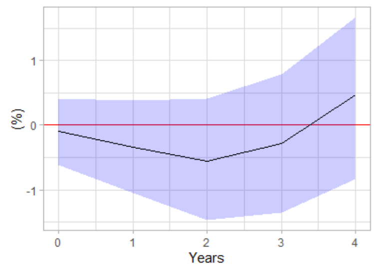
The bulk of the effects happen between the first and third years after the tax changes. Aggregate income drops by 1.25% in the first year, peaks at a 1.49% drop in the second year, and slowly reverts to the trend in the long run. The high persistence of AMTR shocks on aggregate income is expected since tax reforms permanently change household incentives. In contrast, the effects on GDP and Consumption peak in the first year (-1.21% and -0.75% respectively) and quickly return to their trends by the third year.

The effect on consumption is puzzling, as one would expect the long-run effect on taxable income to show up on consumption. The discrepancy between the impact on GDP and income is puzzling as well. Both can be attributed to a substitution effect that the BLP-IV is able to capture. Since U.S. income tax schedules incidence over wages is higher than over alternative sources of income, a raise in AMTR induces substitution of labor for those alternatives in the aggregate level. Such an effect is consistent with all estimates: the AMTR shock causes a large temporary increase in investment (2.11% in the first year and 3.76% in the second year) and a decrease in interest rates (-0.62% in the second year). As a result, economic

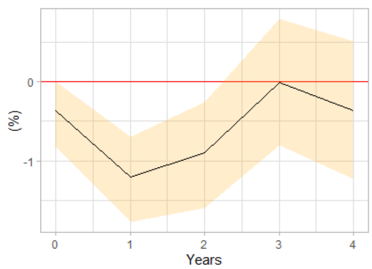
Figure 4: BLP-IV Responses to 1 p.p. AMTR shock.



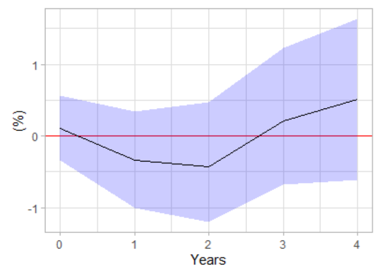
IV1 - Taxable Income



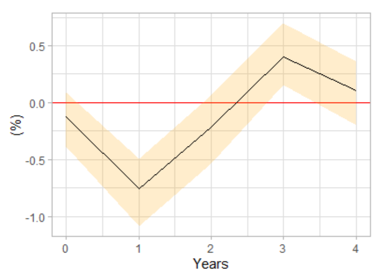
IV3 - Taxable Income



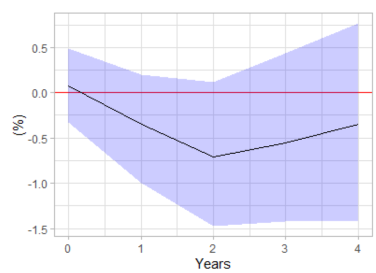
IV1 - Real GDP



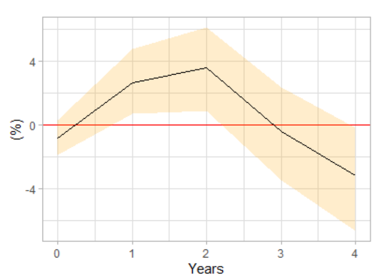
IV3 - Real GDP



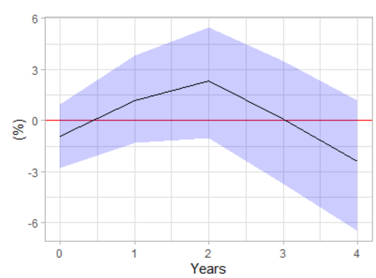
IV1 - Consumption



IV3 - Consumption



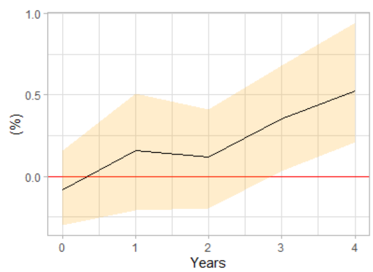
IV1 - Investment



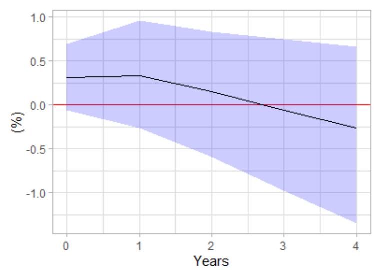
IV3 - Investment

Note: Figure shows the impulse response functions to a 1 p.p. AMTR shock identified with all statutory variations (right column) and the statutory variation of selected tax reforms (left column). The shaded areas represent 90% (light) credible bands and the solid black line represents the median of the posterior distribution.

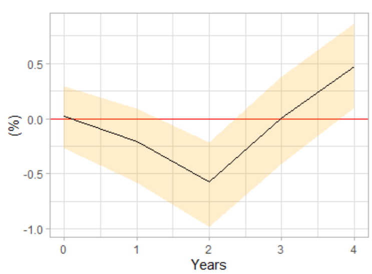
Figure 5: BLP-IV Responses to 1 p.p. AMTR shock (Continuing).



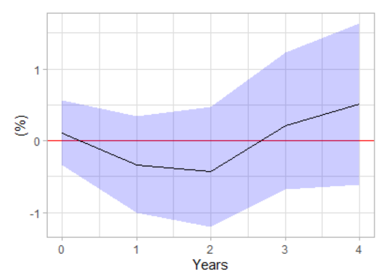
IV1 - CPI (All Urban)



IV3 - CPI (All Urban)



IV1 - 1-Year T-Bond Yield



IV3 - 1-Year T-Bond Yield

Note: Figure shows the impulse response functions to a 1 p.p. AMTR shock identified with all statutory variations (right column) and the statutory variation of selected tax reforms (left column). The shaded areas represent 90% (light) credible bands and the solid black line represents the median of the posterior distribution.

activity and consumption recover faster than taxable income. This substitution effect is also identified in alternative specifications of BLP-IV using other labor market outcomes such as wages and labor force participation. [17]

In contrast with all the other effects discussed above, the impact of AMTR shocks on the consumer price index (CPI) is highly persistent and only occurs in the long run. A 1% increase in AMTR causes 0.5% inflation after four years. This finding is consistent with fiscal new-Keynesian DSGE models such as Bhattarai and Trzeciakiewicz (2017). A rise in labor income taxes reduces both aggregate supply, by raising the cost of labor input, and aggregate demand through the income effect channel. The effect on prices is ambiguous in the short run, but takes over in the long run because the income effect is temporary and the supply restriction is not.

4.1.7 Results from Model with IV1

In Section 4.1.2 I argued that the posterior density of π when estimated with the first IV1 implied that the dummy d_t had to include a few endogenous reforms. To further investigate whether that is the case, I estimate the model using IV1 to verify that the estimates are biased in the expected direction. Results are shown in Figures (**??** and 5).

Endogenous tax reforms are expected to be pro-cyclical. For this reason, the AMTR instrument that includes all reforms may capture the effect of other shocks that positively impact economic activity, creating a bias in the opposite direction of estimates. The estimates for all seven outcomes confirm this intuition, although only in the case of CPI is the bias large enough to reverse the direction of the impulse responses.

4.1.8 Comparison between Small-Scale and Medium-Scale Models

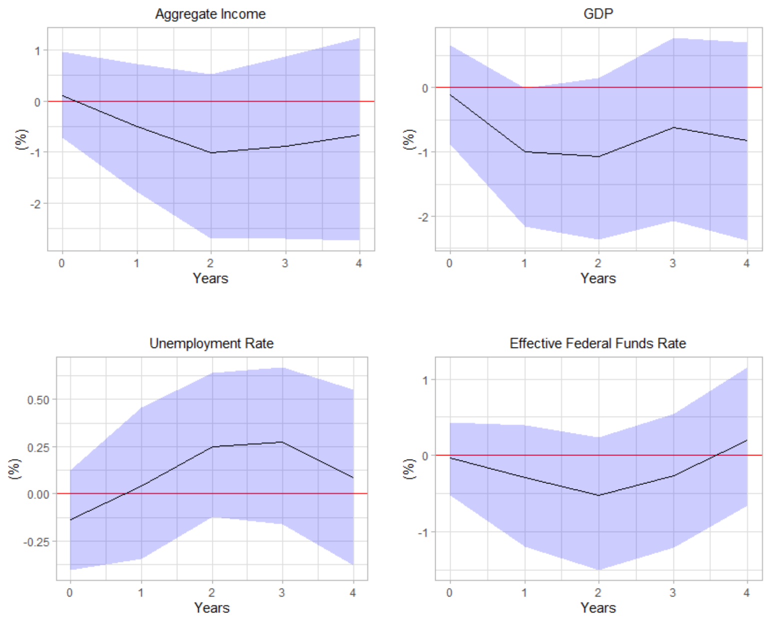
To showcase the merits of including a robust set of controls, I provide estimates of the small-scale model in Figure 6, utilizing IV3 and the same priors as the medium-scale models. The BLP-IV with dispersed priors over the impulse response is unable to identify even the first-order effects on aggregate taxable income.

4.2 Monetary Policy Surprises

n the previous empirical application, the BLP-IV method was used to estimate impulse responses when dealing with small samples and when the instrument faced two potential identification threats: the violation of lag exogeneity and the possible inclusion of endogenous policy variables. In this section, I address the issue of instrumental variables contaminated by measurement errors and/or noise. To illustrate this, I

[17] Results for this version are available on request.

Figure 6: BLP-IV Responses to 1 p.p. AMTR shock (Small Scale Model).



Note: Figure shows the impulse response functions to a 1 p.p. AMTR shock identified with the statutory variation of selected tax reforms (left column). The shaded areas represent 90% (light) credible bands and the solid black line represents the median of the posterior distribution.

revisit the analysis by Bauer and Swanson (2023) on monetary policy surprises, which attempts to identify the causal effects of monetary policy shocks on macroeconomic outcomes such as industrial production and inflation. The instruments in this study are constructed using a series of high-frequency event studies that capture the impact of surprise announcements during FOMC meetings on various financial variables, such as Euro-Dollar futures, stock indices like the SP 500, and the prices of traded securities.

Since each event study captures a single shock without cross-sectional variation across the broader economy, these studies pick up not only policy shocks but also random fluctuations in financial markets during the event window. As a result, impulse response estimates from single-equation regressions are subject to attenuation bias, similar to the cases discussed earlier in this paper. To address this issue, researchers such as Stock and Watson (2012) have proposed using two-stage regressions. However, the reliance on one-stage regressions remains widespread, likely due to the popularity of reduced-form or one-stage approaches in recent local projection studies, such as Barnichon and Brownlees (2019) and Ferreira et al. (2023), or perhaps due to a prevailing belief that noise in high-frequency instruments is minimal or adequately accounted for by instrument refinements or sign-restricted decompositions, as in Jarociński and Karadi (2020). In this section, I show that this belief is unjustified: in my main exercise, the attenuation parameter is significantly different from 1, leading to a substantially larger estimated impact of monetary policy shocks.

4.2.1 Instrument Description and Validation

To identify monetary policy shocks, I use the monetary policy surprises estimated by Bauer and Swanson (2023), which are publicly available in their replication files. These surprises are measured by taking the principal component of fluctuations in Euro-Dollar futures (1, 2, 3, and 4 months) within a 30-minute window around FOMC announcements. I refer to this instrument as MPS (monetary policy surprises), and it is normalized to match the scale of the 4-month Euro-Dollar future.

The literature has long suspected that MPS captures more than pure monetary policy shocks. Many studies have sought to address identification issues by controlling for potential distortions in the instrument, such as central bank information effects. To test whether the common refinements used in the literature adequately address all the noise in MPS, I employ a second version of the instrument. This version is the residual from regressions that account for news releases, as suggested by Bauer and Swanson (2023), as well as teal-book forecasts, nowcasts, and forecast revisions. The latter is often used as a proxy for the central bank's information set and is intended to control for information effects stemming from the central bank's informational advantage (Miranda-Agrippino and Ricco (2021)). I refer to this second instrument as MPS_{ORTH}.

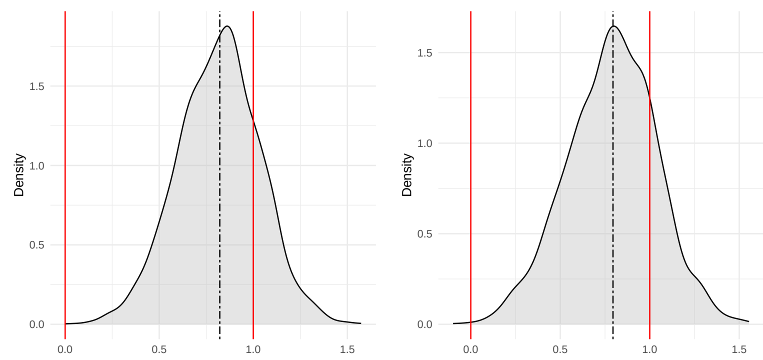
Figure 7 depicts the posterior distribution of the parameter π when the 4-month Euro-Dollar future is used as the dependent variable in the first-stage equation. Three key conclusions emerge from this exercise: first, explicitly controlling for news or central bank information effects does not fully eliminate the excess noise in monetary policy surprises; second, the amount of noise is substantial, with an estimated 20.54% of the variation in MPS_{ORTH} attributed to noise; and finally, even without these refinements, MPS still satisfies the necessary condition for the exclusion restriction.

4.2.2 Data and Model Specification

I use the refined instrument MPS_{ORTH} to estimate the impact of monetary policy shocks on the U.S. economy, focusing on industrial production and CPI inflation as the outcome variables. The regressions are conducted at a monthly frequency, controlling for 12 lags of each outcome variable, the instrument, the unemployment rate, and 1-year Treasury bond yields. To compare the results, I run the analysis twice: first using a one-stage approach, and then applying my two-stage Bayesian local projections.

For the impact period, I use the priors outlined in the previous section of the paper: the Minnesota prior for the nuisance parameters, an uninformative normal prior for the impulse responses, and a uniform unit interval prior for π. For $h > 0$ periods, the impulse response from period $h - 1$ serves as the prior for β.

Figure 7: Posterior π comparisons between MPS and MPS_{ORTH}.



Note: Figure depicts the posterior distribution of the first stage parameter π for each of the alternative instrument MPS and MPS_{ORTH} using the Euro-Dollar Future (4 months) as the endogenous policy measure. In that case, we can interpret these as the posterior distribution of the attenuation bias.

This interactive prior scheme is employed to smooth the results, following the approach of Barnichon and Brownlees (2019).

4.2.3 Results

Figure 8 illustrates the effects of a monetary policy contraction, with the shock normalized to increase the Euro-Dollar future by 1, equivalent to approximately a 90 basis point increase in the 1-year Treasury yield.

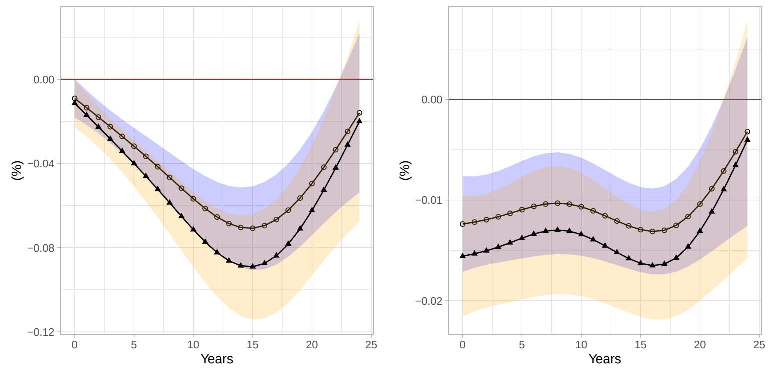
The impulse responses estimated using the two-stage regressions are larger in magnitude compared to the one-stage results. For industrial production, the most significant difference occurs at the peak of the effect, 16 months after the shock: the two-stage estimate shows a decline of -0.89%, compared to -0.70% in the one-stage estimate. For CPI inflation, the differences are evident in both the short run and mid-term, with the two-stage estimates being approximately 25% larger. In both cases, the bias diminishes in the long-run estimates.

5 Concluding Remarks

I demonstrate that a parametric Local Projection (LP) model can be derived when researchers have access to a sufficiently rich set of controls to propose a forecasting model for the outcome variables. In this framework, structural impulse responses can be identified using instrumental variables. To estimate the dynamic, causal effects of average marginal tax rate (AMTR) shocks and monetary policy shocks, I propose a Gibbs sampler that simulates the posterior distributions of these structural impulse responses.

The primary advantage of the Bayesian framework—its capacity to incorporate prior information into

Figure 8: BLP-IV Responses to 1 p.p. Euro-Dollar Future (4 Months) monetary policy shock.



Note: Figure shows the impulse response functions to a 1 p.p. monetary policy shock identified with high-frequency monetary policy surprises orthogonalized to new data releases as in Bauer and Swanson (2023). The blue-shaded areas represent 90% credible bands when using the one-stage regression. The orange shaded area represents the 90% credible bands when adjusting for the attenuation bias, using the two-stage BLP.

econometric modeling—is employed to tackle three key identification challenges. First, the typically stringent lag-exogeneity condition for instrument validity can be relaxed when LPs are defined over forecasting errors. Additionally, shrinkage priors on nuisance auto-regressive and moving-average coefficients allow finite-sample models to incorporate a broader set of controls, thus enhancing robustness.

Second, I demonstrate that explicit assumptions about the joint distribution of the target shock and the instrument influence the values of the attenuation bias coefficients. In many cases, posterior inference on these coefficients can provide valuable insights into instrument quality, offering researchers an additional robustness check. Moreover, incorporating theoretically justified prior information about these parameters, such as their sign, can refine inference. Weak instrument concerns are mitigated by using dispersed priors over impulse responses, which, while potentially leading to imprecise estimates, allow for informative priors on other parameters to produce qualitatively meaningful results. These techniques are applied in two empirical analyses to address open research questions.

First, I estimate the impulse responses of AMTR shocks using a medium-scale Bayesian Local Projection model. The results indicate that marginal income tax shocks are contractionary, with effects on economic activity persisting for two years, despite their longer-lasting impact on taxable income. This is attributed to capital-labor displacement effects, which I estimate. Additionally, I validate commonly used instruments in the tax literature, finding that statutory tax variations—without narrative-based tax reform selections—are endogenous, while other tax instruments satisfy the necessary exclusion restrictions.

Second, I estimate the impulse responses of monetary policy shocks using a refined version of Bauer and Swanson (2023), accounting for both the economic news channel and central bank information advantage in the spirit of Laumer and Santos (2024). In the first stage, I confirm that the contamination channels suggested in the literature do not fully explain the excess noise in monetary policy surprises. I also demonstrate that significant attenuation bias arises when monetary policy surprises are used in a single-stage framework, underscoring the importance of employing a two-stage approach for accurate estimation.

References

Andrews, I. and T. B. Armstrong (2017). Unbiased instrumental variables estimation under known first-stage sign. *Quantitative Economics 8*(2), 479–503.

Arezki, R., V. A. Ramey, and L. Sheng (2017). News shocks in open economies: Evidence from giant oil discoveries. *The quarterly journal of economics 132*(1), 103–155.

Barnichon, R. and C. Brownlees (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics 101*(3), 522–530.

Bauer, M. D. and E. T. Swanson (2023). A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual 37*(1), 87–155.

Bhattarai, K. and D. Trzeciakiewicz (2017). Macroeconomic impacts of fiscal policy shocks in the uk: A dsge analysis. *Economic Modelling 61*, 321–338.

Boer, L. and H. Lütkepohl (2021). Qualitative versus quantitative external information for proxy vector autoregressive analysis. *Journal of Economic Dynamics and Control 127*, 104118.

Canova, F. (2007). *Methods for applied macroeconomic research*, Volume 13. Princeton university press.

Chib, S. and E. Greenberg (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics 64*(1-2), 183–206.

DeJong, D. N. and C. Dave (2012). *Structural macroeconometrics*. Princeton University Press.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford university press.

Ferreira, L. N., S. Miranda-Agrippino, and G. Ricco (2023). Bayesian local projections. *Review of Economics and Statistics*, 1–45.

Gertler, M. and P. Karadi (2015, January). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Hirano, K. and J. R. Porter (2015). Location properties of point estimators in linear instrumental variables and related models. *Econometric Reviews 34*(6-10), 720–733.

Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics 12*(2), 1–43.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review 95*(1), 161–182.

Jordà, Ò., M. Schularick, and A. M. Taylor (2015). Betting the house. *Journal of International Economics 96*, S2–S18.

Kleibergen, F. and E. Zivot (2003). Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics 114*(1), 29–72.

Laumer, S. and I. M. Santos (2024). The impact of monetary policy shocks—do not rule out central bank information effects or economic news. *Economics Letters 237*, 111634.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2022). Local projections vs. vars: Lessons from thousands of dgps. Technical report, National Bureau of Economic Research.

Lusompa, A. (2021). Local projections, autocorrelation, and efficiency. *Federal Reserve Bank of Kansas City Working Paper* (21-01).

Mertens, K. and J. L. Montiel Olea (2018). Marginal tax rates and income: New time series evidence. *The Quarterly Journal of Economics 133*(4), 1803–1884.

Mertens, K. and M. O. Ravn (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy 4*(2), 145–81.

Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics 13*(3), 74–107.

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*(4), 1789–1823.

Müller, U. K. (2013). Risk of bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica 81*(5), 1805–1849.

Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913–1998. *The Quarterly journal of economics 118*(1), 1–41.

Plagborg-Møller, M. (2019). Bayesian inference on structural impulse response functions. *Quantitative Economics 10*(1), 145–184.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and vars estimate the same impulse responses. *Econometrica 89*(2), 955–980.

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics 126*(1), 1–50.

Romer, C. D. and D. H. Romer (1989). Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual 4*, 121–170.

Romer, C. D. and D. H. Romer (2009). A narrative analysis of postwar tax changes. *Unpublished paper, University of California, Berkeley (June)*.

Romer, C. D. and D. 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.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of economic literature 50*(1), 3–50.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Sims, C. A. (2002). Solving linear rational expectations models. *Computational economics 20*(1-2), 1.

Sims, E. R. (2012). News, non-invertibility, and structural vars. In *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*. Emerald Group Publishing Limited.

Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007-2009 recession. Technical report, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics 117*, 90–103.

A Appendix - Proof of Corollary 2

First, notice $e_{t+h,t-1}$ a is covariance stationary process. I will this assertion directly, by deriving its moments. To do this, I will introduce additional notation specific of this appendix. I also re-scale the shocks ϵ_t to one standard deviation, in order to simplify calculations.

Definition 3. *Let A be a $m \times n$ matrix. The term $\{A\}_\alpha$ refers to a linear combination of all columns of A, that is $\{A\}_\alpha \in span(A)$. The term α is a n-dimensional vector containing the scalars.*

From the representation (1):

$$e_{t+h,t-1} = \{\epsilon_{t:t+h}\}_\alpha - \beta_h^k \epsilon_t^q.$$

I proceed to prove its covariance stationarity by deriving its moments. First the mean:

$$\begin{aligned} E(e_{t+h,t-1}) &= E(\{\epsilon_{t:t+h}\}_\alpha - \beta_h \epsilon_t^q) \\ E(e_{t+h,t-1}) &= 0. \end{aligned}$$

Now all the auto-correlation coefficients:

$$\begin{aligned} E(e_{t+h,t-1} e'_{t+h+s,t-1}) &= E[(\{\epsilon_{t:t+h}\}_\alpha - \beta_h \epsilon_t^q)(\{\epsilon_{t:t+h}\}_\alpha - \beta_h \epsilon_t^q)'] \\ E(e_{t+h,t-1} e'_{t+h+s,t-1}) &= \begin{cases} \{I_{h+s}\}_{\alpha^2} + \beta_h \beta_h' \text{ for all } |s| = 0, 1, 2, ..., h-1. \\ 0 \text{ otherwise} \end{cases} \end{aligned}$$

Since $I_{h+s}\}_{\alpha^2} + \beta_h \beta_h' < \infty$ and it does not depends on t, $e_{t+h,t-1}$ is covariance stationary and admits Wold representation. That is, $\exists\ C(L) = \sum_{t=1}^{\infty} C_{-i} L^i$ such that $e_{t+h,t-1} = C(L)\zeta_t$ where ζ_t is white noise. Since $e_{t+h,t-1}$ auto-correlations are zero for $|s|$ equal or higher than h, the lag order of $c(L)$ is exactly h.

Does $e_{t+h,t-1}$ following a VMA(h) process imply that each individual projection error $e^k_{t+h,t-1}$ has MA(h) representation ? The answer is yes. To see this, write the projection errors as $e^k_{t+h,t-1} = re_{t+h,t-1}$ where r is vector of zeros with a single 1 on its k-ith entry. Since any continuous function of a covariance stationary process is also covariance stationary, $e^k_{t+h,t-1}$ admits its own univariate Wold representation and the lag order evidently cannot exceed h (or else $E(e_{t+h+s,t-1} e'_{t+h,t-1}) \neq 0$ for $s \geqslant h$).