Local Projections vs. VARs for Structural Parameter Estimation

Juan Castellanos

European University Institute

July 2024

Abstract: This paper conducts a Monte Carlo study to examine the small sample performance of IRF matching and Indirect Inference estimators that target impulse responses (IRFs) that have been estimated with Local Projections (LP) or Vector Autoregressions (VAR). The analysis considers various identification schemes for the shocks and several variants of LP and VAR estimators. Results show that the lower bias from LP responses is a big advantage when it comes to IRF matching, while the lower variance from VAR is desirable for Indirect Inference applications as it is robust to the higher bias of VAR-IRFs. Overall I recommend the use of Indirect Inference over IRF matching when estimating DSGE models as the former is robust to potential misspecification coming from invalid identification assumptions, small sample issues or incorrect lag selection.

Keywords: DSGE Estimation, Impulse Responses, Indirect Inference, Local Projection, Vector Autoregression, Monte Carlo Analysis.

JEL classification: C13, C15, E00.

---

juan.castellanos@eui.eu

I am highly indebted to my advisor Russell Cooper for his continuous guidance and valuable inputs during the development of this project. I would also like to thank Jesus Bueren, Wouter Den Haan, Andrea Gazzani, Marc Henry, Jeremy Majerovitz, Lukas Nord, Giorgio Primiceri, Sebastian Rast, Barbara Rossi, Valerio Scalone, three anonymous referees, as well as conference and seminar participants at European Central Bank, European University Institute, 2022 Spanish Economic Association Meeting, 2023 BSE Summer Forum and 2023 Theory and Methods in Macro for useful comments and suggestions.

A previous version of this paper was circulated under the title “Indirect Inference: A Local Projection Approach”.

1. Introduction

The Local Projections (LP) approach to understanding the dynamic effects of exogenous shocks, originated in Jordà (2005), has become a common and alternative tool to the traditional Vector Autoregression (VAR) approach. In light of the theoretical result in Plagborg-Møller and Wolf (2021), i.e. VARs and LPs estimating the same impulse responses in population, one may think that using either VAR or LP should not matter. However, the finite sample properties of these two estimators differ. In particular, when p lags of the data are included in the VAR and as controls in the LP, IRFs approximately agree out to horizon p, but at longer horizons $h > p$ there is a bias-variance trade-off (Li, Plagborg-Møller, and Wolf 2023).

This paper explores the implications of using one of these two econometric models for summarizing key features of the data, such as the dynamic response to exogenous shocks, in order to estimate the structural parameters of a DSGE model. Given their different small sample properties, targeting impulse responses (IRFs) estimated by LP or VAR will lead to different structural parameter estimates. Hence, in practice, using LP- or VAR-IRFs for DSGE estimation may lead to different outcomes and potentially different quantitative predictions from the structural model.

I carry out a Monte Carlo analysis to investigate the consequences of targeting LP vs. VAR estimated responses in a minimum distance estimation. I consider two estimators within this class: impulse response matching and indirect inference, and use the Smets and Wouters' (2007) model as the Data Generating Process (DGP). Further, and as a benchmark, I assume that the econometrician observes the true shock, which guarantees correct identification. Nevertheless, estimated responses will vary depending on the econometric model used for estimation as well as on the sample size and the number of lags. In general, targeting LP responses which have a lower bias than VARs is a great idea if resorting to *IRF matching*. On the other hand, when estimating the structural parameters via *Ind. Inf.*, using the VAR as the auxiliary model outperforms LPs because *Ind. Inf.* is robust to misspecification and VARs have a lower variance.

These results are better understood in conjunction with the choice of p, the lag length. Note that LP responses are independent of the lag length when the shock is observed, while VAR responses become more similar to LP's as the lag length increases. Actually, the reduction of the bias in VAR responses as p increases comes also at the cost of a larger variance (Olea, Plagborg-Møller, Qian, and Wolf 2024). Consequently, when p is large, there are little differences between targeting LP- or VAR-IRFs. On the other hand, when p is small, the LP approach is significantly better than VARs for *IRF matching* due to its smaller bias, while using a small p VAR as the auxiliary model for *Ind. Inf.* is the superior choice due to its smaller variance.

The sample size used to estimate these responses also has an impact on the structural parameter estimation. The larger the small sample bias, the worse the estimation outcome. However, such deterioration in the performance of the estimation is more important for *IRF matching* than for *Ind. Inf.* to the point that the latter is preferred regardless of the econometric model used to estimate IRFs. Moreover, I also show that using the bias corrected version of LPs and VARs improves the estimation outcome in an *IRF matching* exercise, while it is not so relevant for *Ind. Inf.* applications.

On a second set of Monte Carlo simulations I relax the observed shock assumption and consider a scenario in which the econometrician does not observe the shock at all and has to infer it from recursive assumptions. Here I show that if assumptions are correct, e.g. by assuming that TFP does not affect other endogenous variables at time 0 as it is the case in the Smets and Wouters' model, the results from the observed shock scenario still hold. On the other hand, when these recursive assumptions are incorrect, e.g. if I assume that the policy rate has no contemporaneous impact on real variables, a common assumption for monetary policy shocks that doesn't hold in the Smets and Wouters' model, then *IRF matching* estimates are significantly worse relative to the observed shock identification due to the larger bias in IRFs, while *Ind. Inf.* estimates are surprisingly better than the observed shock case because of the lower variance in IRFs, specially at shorter horizons.

In the last set of Monte Carlo simulations I consider an intermediate scenario in which the econometrician observes a proxy for the shock that is contaminated with measurement error, which can or cannot be correlated with other shocks. In either case, the estimation performance worsens for both (auxiliary) econometric models (LP & VAR) and estimation approaches (*IRF matching* & *Ind. Inf.*). However, an improvement can be attained if applying the unit effect normalization of Stock and Watson (2018), which corrects for the bias in the estimated IRFs, and consequently, improves the structural estimation outcome for both approaches, but specially for *IRF matching*.

Overall, these findings provide a novel perspective on DSGE estimation setups that target estimated impulse responses and shed light on how the bias-variance trade off between LPs and VARs translate to the structural parameters of the economic model. The main lesson is that *Ind. Inf.* is robust to misspecification, which is more common among VARs, and benefits from more tightly estimated IRFs. The opposite is true for the *IRF matching* approach. Thus, researchers should rely more often on *Ind. Inf.* to estimate their DSGE models as it is robust to potential misspecification coming from invalid identification assumptions, small sample bias or incorrect lag selection.

Related Literature. The Monte Carlo study in this paper is inspired by the seminal work of Smith (1993) on the use of VAR models as the binding function in an indirect inference exercise that estimates the structural parameters of a DSGE model. Unlike Smith (1993), who uses all the coefficients in the VAR, I only select those coefficients that identify the impulse responses to a given shock. Hence, my paper is also related to the literature that relies on IRF matching for DSGE estimation (Rotemberg and Woodford 1997). In fact, throughout the paper, I compare the performance of these two estimators, *Ind. Inf.* and *IRF matching*, when targeting responses to various shocks under different IRF estimation methods and identification strategies. Consequently, my paper belongs to the broader literature that studies the small sample properties of minimum distance, simulation based, partial information estimators. Examples include: *(i)* Jordà and Kozicki (2011) who propose an estimator in which the economic model restrictions are based on its impulse response representation; *(ii)* Creel and Kristensen (2011) who propose an Indirect Likelihood Estimator as an alternative to Simulated Method of Moments or Indirect Inference; *(iii)* Scalone (2018) who advocates for the use of Bayesian Method of Moments for the estimation of non-linear economic models; or *(iv)* Ruge-Murcia (2007, 2012, 2020) who studies the small sample properties of minimum distance estimators in linear and non-linear environments as well as with linear and non-linear binding functions for the indirect inference applications. Unlike these papers, my Monte Carlo study aims to analyze the small sample properties of the two most common minimum distance estimators used in macroeconomic applications, *IRF matching* and *Ind. Inf.*, under various identification assumptions for the estimated responses that act as targets. Moreover, I consider LPs, in addition to VARs, as the auxiliary econometric model adopted for estimation.

Given my interest in the performance of LP and VARs as the source of (data) moments or as the auxiliary model for indirect inference, my paper is also related to the literature that studies the performance of these two methods in the context of IRF estimation. Plagborg-Møller and Wolf (2021) have proven that these econometric models estimate the same IRFs in population and that LPs can impose the same amount of identification restrictions used in SVARs after appropriately choosing the set of controls. However, the small sample properties of these two estimators differ. In fact, there is a bias-variance trade-off beyond horizon p as shown in Li, Plagborg-Møller, and Wolf (2023). My paper complements their results as it confirms the bias-variance trade-off under a different DGP, but more importantly, investigates its implications for the purposes of uncovering structural DSGE parameters.

Overview. The rest of the paper is organized as follows. Section 2 describes and justifies the choice of the DSGE model used to generate the data. Section 3 describes the estimation methodology, the auxiliary models employed to estimate IRFs, and the various identification strategies within the context of the DSGE model used as DGP. Section 4 outlines the Monte-Carlo design and discusses the metrics used to evaluate the results, which are then presented in Section 5. Finally, Section 6 concludes.

2. The Data Generating Process

This section describes the model used to generate the data for the Monte Carlo study in which I compare the impulse response function matching (*IRF matching*) and the indirect inference (*Ind. Inf.*) estimation strategies as a way to infer the structural parameters of a DSGE model. Many models could have fulfilled this purpose, nonetheless, I have chosen the Smets and Wouters (2007) model for several reasons. First, it is a well-understood and widely used model in academia as well as in policy circles. Second, the vast majority of existing applications that estimate their model economies by matching impulse responses concern linearized models, see for example Rotemberg and Woodford (1997), Christiano, Eichenbaum, and Evans (2005), Iacoviello (2005), or Jordà and Kozicki (2011). It is true, however, that the theoretical foundations of indirect inference were grounded on the estimation of nonlinear models (Gourieroux, Monfort, and Renault 1993). I acknowledge this limitation associated to the chosen DGP, nonetheless, how to choose between LPs and VARs within these two estimation set-ups is still an open question in linearized, and hence, simpler settings. [1] And third, the model is sufficiently rich to allow us to explore different types of shocks and identification strategies. As discussed in Ramey (2016), monetary, fiscal and technology shocks are the most widely studied in empirical applications and hence responses to these shocks are potentially also being used as data moments/targets for structural estimation. Importantly, the Smets and Wouters model is able to generate reasonable responses to all these three shocks.

[1] Ruge-Murcia (2020) studies the performance of non-linear auxiliary models in non-linear settings, but he only focuses on local projections and indirect inference estimation.

2.1. The Smets and Wouters Model

The model is based on Christiano, Eichenbaum, and Evans (2005) who added various frictions to a basic New Keynesian DSGE in order to capture the dynamic response to a monetary policy shock as measured by a structural vector autoregression (SVAR). When price and wage stickiness are paired with adjustment costs for investment, capacity utilization costs, habit formation in consumption, partial indexation of prices and wages as well as autocorrelated disturbance terms, the model is able to generate a rich autocorrelation structure. These elements are key for capturing the joint dynamics of output, consumption, investment, hours worked, wages, inflation and the interest rate in the Euro Area (Smets and Wouters 2003). The 2007 version of the model, which I use in this paper, is a minor modification of the 2003 Smets and Wouters model in order to fit the US data. Given the importance of the Smets and Wouters model in the DSGE literature, I do not describe their economy in this paper. Nonetheless, the log linearized equilibrium conditions are reproduced in Appendix A.

3. Estimation Methods

3.1. Indirect Inference

Any economic model, including the Smets and Wouters model, can be represented as a function, $M(\cdot)$, that for a given vector of parameters Θ maps a sequence of endogenous states $\{y_{t-1}\}$, exogenous variables $\{x_t\}$ and shocks $\{\varepsilon_t\}$, into a sequence of endogenous variables $\{y_t\}$. That is,

$$y_t = M(y_{t-1}, x_t, \varepsilon_t; \Theta) \tag{1}$$

for $t = 1, \ldots, T$. Therefore, using this mapping it is possible to generate infinite data sequences $\{y_t\}_{t=1}^{T}$, given an initial value of the endogenous state y_0 and a sequence of the shocks $\{\varepsilon_t\}_{t=1}^{T}$. Model simulation is in fact the basis for the class of minimum distance estimators that seek to find an ex-ante unknown parameter vector that minimizes the distance between data and simulated moments. The most common among this class are the Simulated Method of Moments (*SMM*) and the Indirect Inference (*Ind. Inf.*) estimators. The only difference between these two is that *SMM* uses unconditional moments, while in an *Ind. Inf.* exercise these come from an auxiliary, typically econometric, model. The auxiliary model that summarizes the key features of the data into a tractable vector of parameters is often referred to as the *binding function*.

The *Ind. Inf.* approach was popularized in macroeconomics by Smith (1993) who used a VAR to summarize the key features from the data that he wanted to replicate with his economic model. Formally, the *Ind. Inf.* estimates arise from solving the following minimization problem

$$J^{smm} = \min_{\Theta} \left(\beta - \beta(\Theta)\right)' W \left(\beta - \beta(\Theta)\right) \tag{2}$$

where β and $\beta(\Theta)$ are the estimated coefficients of an auxiliary (econometric) model and W is a weighting matrix. In this paper, these estimated coefficients are those identifying the dynamic response to an aggregate shock. As shown in Sections 3.3 and 3.4, there are various approaches to identify and estimate IRFs. Consequently, the main objective of this paper is to study how the choice of these particular binding functions $\beta(\cdot)$ affect the structural parameter estimates $\hat{\Theta}$.

3.2. Impulse Response Function Matching

An alternative to *Ind. Inf.*, that is used frequently in DSGE estimation, is impulse response function matching (*IRF matching*). It is also a minimum distance estimator as it minimizes the distance between data targets (estimated IRFs) and its model counterparts (structural IRFs). In fact, this approach is more similar to calibration than it is to estimation. Nonetheless, it provides a natural benchmark to compare the *Ind. Inf.* estimation results as bias-variance trade-offs, small sample biases or incorrect identification strategies associated to the estimated dynamic responses will only affect the data moments/targets. Hence, it is more likely that the properties of the estimated responses are inherited by the structural parameters when using this approach.

Formally, the *IRF matching* estimates are obtained after solving the following minimization problem

$$J^{irf} = \min_{\Theta} \left(\beta - IRF(\Theta)\right)' W \left(\beta - IRF(\Theta)\right) \tag{3}$$

where the only difference with respect to *Ind. Inf.* is on how IRFs are computed when the candidate vector of parameters Θ is updated in search of a minimum. Notice that in (3), the model counterpart, $IRF(\cdot)$, is the structural IRFs and hence they do not require a simulated dataset because they are directly computed from the ABCD representation of the model (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson 2007).

3.3. The (Auxiliary) Econometric Models

Assume that I observe data $w_t = (r_t', \tilde{x}_t, \tilde{y}_t, q_t')$ where $\tilde{x}_t$ and $\tilde{y}_t$ are scalar time series and r_t' and q_t' are $n_r \times 1$ and $n_q \times 1$ vectors of time series including contemporaneous and lagged controls, respectively. I am interested in the dynamic response of $\tilde{y}_t$ after an impulse in $\tilde{x}_t$ as a way of summarizing some features of the data that I would like to replicate with my structural macroeconomic model. The most common approaches to estimate these impulse responses in the data involve the use of VAR or LP. The choice between these two econometric models is important because, despite estimating the same responses in population (Plagborg-Møller and Wolf 2021), their small sample performance is characterized by a bias-variance trade off (Li, Plagborg-Møller, and Wolf 2023). Hence, I am interested in how these small sample properties may affect the structural parameters when VARs or LPs are used to summarize the data in a minimum distance estimation.

3.3.1. VAR approaches

Least Squares VAR. I consider a recursive VAR specification in w_t

$$w_t = c + \sum_{\ell=1}^{p} A_\ell w_{t-\ell} + u_t \tag{4}$$

where u_t is the projection residual and $(c, \{A_\ell\}_{\ell=1}^p)$ are the projection coefficients. These coefficients are estimated by least-squares and the residual covariance matrix, $\hat{\Sigma}_u = T^{-1} \sum_{t=2}^{T} \hat{u}_t \hat{u}_t'$, is factorized using a lower triangular Cholesky factor $\hat{B}$, such that $\hat{B}\hat{B}' = \hat{\Sigma}_u$. Define the lag polynomial $\sum_{\ell=0}^{p} C_\ell L^\ell = C(L) \equiv A(L)^{-1}$. Noting that $\tilde{x}_t$ and $\tilde{y}_t$ are the (n_r+1)-th and the (n_r+2)-th elements in w_t, I can now define the VAR impulse response function of $\tilde{y}_t$ with respect to an impulse in $\tilde{x}_t$ as $\{\Lambda_h\}_{h\geq 0}$ where

$$\Lambda_h \equiv C_{n_r+2,\bullet,h} B_{\bullet,n_r+1} \tag{5}$$

and $B_{\bullet,n_r+1}$ is the (n_r+1)-th column of B and $C_{n_r+2,\bullet}$ refers to the (n_r+2)-th row of C_h.

Bias corrected VAR. The impulse responses are estimated as above, but I use the modification proposed by Kilian (1998) that applies the formula in Pope (1990) to analytically correct the bias of the reduced-form coefficients caused by persistent data.

3.3.2. Local projection approaches

Least Squares LP. The least-squares local projection estimator β_h is obtained from the OLS regression

$$\tilde{y}_{t+h} = \mu_h + \beta_h \tilde{x}_t + \gamma_h' r_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' w_{t-\ell} + \xi_{h,t} \tag{6}$$

where $\tilde{y}_{t+h}$ is the response variable, $\tilde{x}_t$ is the impulse variable, and r_t are contemporaneous controls, $\{w_{t-\ell}\}_{\ell=1}^{p}$ controls for p lags of all data series included in the regression, and $\xi_{h,t}$ is the projection residual.

Bias Corrected LP. I use the version proposed by Herbst and Johannsen (2023) where they partially remove the bias associated to high persistence in the data. This bias, although asymptotically negligible relative to the standard deviation, can be sizable in small samples.

3.3.3. Lag length selection

A key element to understand the differences in the estimated IRFs when using Local Projections or SVARs is the lag length, p. Recall one of the Plagborg-Møller and Wolf's (2021) results: *Local Projections with p lags as controls and VAR(p) estimators approximately agree at impulse response horizons $h \leq p$.* Consequently, using longer lag lengths given a fixed horizon H will certainly deliver more similar targeted responses across the two econometric models. As a result, the estimated economic parameters should also be more similar when comparing across VARs and LPs as the source of moments/targets. To test this hypothesis, we will consider estimation setups with various lag lengths for the (auxiliary) econometric models, i.e. I let $p \in \{2, 4, 8, 12\}$. Alternatively, I could have opted for using information criteria such as AIC or BIC, however, these tend to select very short lag lengths which are not consistent with the typical choices in applied work. In fact, Li, Plagborg-Møller, and Wolf (2023) use the following lag length rule, $p = \max\{\hat{p}_{AIC}, 4\}$, which for my DGP will have always resulted in picking $p = 4$.

3.4. Impulse Response Estimands & Identification

I follow Li, Plagborg-Møller, and Wolf (2023) in considering three types of structural impulse response estimands to mimic as closely as possible the schemes used in applied macroeconometrics to identify impulse responses in the data. Recall that these responses are simply a way of summarizing the data for our structural estimation, and not the main focus of our analysis.

3.4.1. Observed innovation / observed shock identification

I assume that the econometrician observes the endogenous variables $\bar{w}_t$ and the true structural shock ε_t or equivalently its innovation η_t. For the VAR approaches, I order the shock as the first variable in the VAR system with r_t' being empty. Equivalently for the LP approaches, the impulse variable $\tilde{x}_t$ is the shock (or the innovation) itself. Consequently, no controls are needed in the OLS regression (6) to mope out any measurement error or serial correlation in the shocks, as typically done in many empirical applications (Ramey 2016, Stock and Watson 2018). As a result, the observed data vector $\bar{w}$ includes the shock itself as well as the macroeconomic variables of interest for both econometric models. The latter include: *(i)* output, *(ii)* consumption, *(iii)* investment and *(iv)* hours worked.

I estimate the dynamic response of each of these variables to one of the three most common aggregate shocks: *(i)* monetary, *(ii)* fiscal and *(iii)* technology shocks. For monetary and technology shocks using the innovation η_t or the shock ε_t will lead to identical estimated responses, however, this is not the case for the fiscal policy shock. Recall that in the Smets and Wouters model government spending is completely exogenous but it is affected by the technology shock as follows:

$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \rho_{ga}\eta_t^a + \eta_t^g \tag{7}$$

where ρ_{ga} captures the contemporaneous correlation between the two shocks. As a result, if $\rho_{ga} \neq 0$, using the shock ε_t^g without controlling for TFP will lead to incorrect responses. To circumvent this issue I will initially use the innovation rather than the shock itself as our impulse variable for all the shocks, i.e. $\tilde{x}_t = \eta_t^i$ for $i \in \{m, g, a\}$. Nonetheless, I will still explore the differences between using the innovation or the correlated fiscal policy shock as explained in Section 3.4.3.

3.4.2. Recursive identification

On the other extreme, I assume that the econometrician only observes the endogenous variables with no direct measure of the shock. Consistent with the large literature in recursive shock identification in VARs (e.g. see Christiano, Eichenbaum, and Evans, 1999), the shock of interest is the orthogonalized innovation to a policy variable i_t included in the vector of endogenous variables $\bar{w}_t$.

There are two common identification assumptions to impose recursive zero restrictions on contemporaneous coefficients (Ramey 2016). First, *the policy variable does not respond within the period to other endogenous variables*. For example, Blanchard and Perotti (2002) impose this constraint in the context of government spending shocks. And second, *other endogenous variables do not respond to the policy variable within the period*. Bernanke and Blinder (1992) were the first to identify monetary policy shocks in this way, but they have been followed by others like Christiano, Eichenbaum, and Evans (2005).

Consistent with this literature, we follow the second approach for monetary shocks and order the policy rate last as this restricts other variables in the VAR to not respond contemporaneously to the monetary innovations. Among the other macro variables in the VAR we include: *(i)* output, *(ii)* consumption, *(iii)* investment, *(iv)* hours worked, *(v)* wages, and *(vi)* inflation. On the other hand, we follow the first approach for the fiscal and technology shocks and use government expenditures or productivity as the first series in the VAR, respectively. For both shocks, we include: *(i)* output, *(ii)* consumption, *(iii)* investment, and *(iv)* hours worked as the other variables in the VAR.

Interestingly, in the context of the Smets and Wouters model these recursive assumptions will only be correct in the case of technology shocks as TFP is purely exogenous. Government expenditures, despite being exogenous, are correlated with the productivity shock while real variables and prices respond contemporaneously to monetary innovations despite price and wage rigidities as shown in Figure 6 of Smets and Wouters' (2007) paper. These invalid identification assumptions for fiscal and monetary policy shocks make our estimation exercises more interesting as it will allow us to test to what extend *Ind. Inf.* is robust to this type of misspecification, i.e. one in which the recursively orthogonalized innovations do not equal the structural shocks in the Smets and Wouters model.

Turning to the LP responses, we know that any SVAR identification scheme can be also implemented using LP methods (Plagborg-Møller and Wolf 2021). In fact, for the identification strategy used for the technology and fiscal policy shocks, this only requires to set the impulse variable $\tilde{x}_t$ to the policy variable i_t; while for the monetary policy identification scheme, we also need to control for the contemporaneous variables that are ordered before the policy variable in the VAR system.

3.4.3. Noisy direct measures of the shocks of interest

In between these two extremes, there is a growing and very popular strand of the literature that relies on external information to construct a direct measure of the shock of interest. These directly measured shocks often capture only part of shock or are measured with error (Stock and Watson 2018). For example, Romer and Romer (2004) use narrative methods to construct a monetary policy shock in which Greenbook forecasts are used to separate the Fed's superior information from the exogenous shock. Nonetheless, they still use additional recursive assumptions when studying the responses of output and prices as they do not view their shock as pure (Ramey 2016). [2] Consequently, I consider a third identification strategy in which the observed innovation / shock is contaminated with measurement error. In particular, I assume that the econometrician observes a proxy for the innovation of the shock:

$$\eta_t^{obs} = \eta_t + \sigma_\nu \nu_t \tag{8}$$

where ν_t is an iid innovation with zero mean and a standard deviation of one. The IRF estimation approach is identical to the observed innovation case in Section 3.4.1 but replacing η_t by equation (8) and assuming that $\sigma_\nu = 0.5$.

In addition to the classical measurement error scenario, I also consider the possibility that the measured shock is correlated with other shocks, which would violate the exogeneity condition. Recall that this is the case for the government spending shock within the Smets and Wouters (2007) model – see equation (7). Hence, I compare the estimation results from targeting responses to fiscal policy which have been estimated using information about the innovation η_t^g versus those that rely on the actual correlated shock ε_t^g. Again, the estimation procedure is identical to the observed shock case, but with a different information set.

Finally, I consider the case in which the IRFs have been normalized using the unit effect of Stock and Watson (2018) as they show that fixing the shock units via normalization allows to capture the dynamic causal effect even in the presence of measurement error.

[2] Other examples of this approach include the fiscal policy shock measure in Ramey (2011) which uses Business Week's articles.

4. Design, Implementation & Evaluation

This section describes how I set up the Monte Carlo study to analyze the small sample properties of the *IRF matching* and the *Ind. Inf.* estimators that use VARs or LPs as the source of data moments or as the binding function, respectively. The analysis is based on the economic model described in Section 2 under the hypothesis that the DGP and the estimated model are the same. [3] That is, the log linearized version of the Smets and Wouters (2007) model is used to generate time series of macroeconomic variables as well as time series for the innovation of the shocks. These series are then used to estimate impulse response functions using the econometric models described in Section 3.3 and under the different identification schemes explained in Section 3.4. Finally, these estimated responses, which summarize the dynamics of the model/data, are used as moments/targets in estimation to pin down the structural parameters.

The Smets and Wouters (2007) model has 36 structural parameters but to reduce the computational burden I focus on 8 of these: the intertemporal elasticity of substitution $\{\sigma_c\}$, the consumption habit parameter $\{h_c\}$, the elasticity of labor supply $\{\sigma_l\}$, the investment adjustment cost parameter $\{\varphi\}$, and the non-adjustment probabilities and indexation parameters for wages $\{\xi_w, \iota_w\}$ and prices $\{\xi_p, \iota_p\}$. The "true" values of these structural parameters are listed in Table 1, while the remaining ones are set and fixed at the mean estimated values by Smets and Wouters (2007) – see Table 1A & 1B in their paper.

Using these parameter values, the "true" model is simulated $S = 100$ times for $T = 300$ periods. [4] This artificial dataset is used to estimate the dynamic response of four macro aggregates: output, consumption, investment and hours worked to either monetary, fiscal or technology shocks over $H = 20$ quarters. Hence, for each Monte Carlo draw and each estimation setup we target $84 = 21 \times 4$ moments. Note that the Monte Carlo distribution of these targets/data moments is identical for both *IRF matching* and *Ind. Inf.* exercises – it represents β in problems (2) and (3). Recall that for *Ind. Inf.* approach the estimated IRFs are also computed based on the model simulated data at each candidate parameter vector, $\beta(\Theta)$. In that case, the sample size is inflated by a factor $\tau = 10$. In theory, we know that the asymptotic distribution of the estimates depends on this choice as simulation uncertainty decreases when the length of the simulated series to the sample size increases. However, in practice, having very long simulated series

[3] We do not consider the alternative that the model is misspecified because this has been already studied by Ruge-Murcia (2007) in the context of the simulated method of moments.

[4] A sufficiently long burning sample is used to get rid of the initial conditions.

TABLE 1. True values of structural parameters

Parameter	**Value**	**Interpretation**
σ_c	1.38	Intertemporal elasticity of substitution
h_c	0.71	Habit parameter
σ_l	1.83	Elasticity of labor supply
φ	5.74	Investment adjustment cost parameter
ξ_w	0.70	Probability of non-adjustment (wages)
ξ_p	0.66	Probability of non-adjustment (prices)
ι_w	0.58	Wage indexation parameter
ι_p	0.24	Price indexation parameter

NOTE. This table depicts the true value of the estimated parameters from the Smets and Wouters model. Their values coincide with the mean estimates from their 2007 paper.

increases the computational cost and is not needed to obtain accurate estimates. Ruge-Murcia (2012) shows how this choice affects the parameter estimates in the context of DSGE models estimated by SMM. Consequently, I do not explore this dimension and simply set this hyper-parameter to a common value used in practice. Nonetheless, I consider the case in which the data moments/targets are estimated on a sample with just $T = 100$ periods to study the small sample bias in LPs documented in Herbst and Johannsen (2023). For simplicity, this robustness test is performed only on the context of the observed innovation scheme.

Finally, I use the identity matrix as the weighting matrix $W = I$ since it is one of the most widely used in empirical work. Nevertheless, I also consider: *(i)* the inverse of the variance-covariance matrix of the data moments (VCM) as it is the optimal weighting matrix, and *(ii)* a diagonal matrix whose entries are the inverse of the IRFs horizon $1/h$. The latter tries to address the possible identification problem arising from the little and noisy information contained in impulse responses at long horizons (Canova and Sala 2009).

4.1. Performance Metrics

To evaluate the performance of a given estimator $\hat{\Theta}$ of Θ, we consider different metrics that can be classified into two groups: *(i) overall performance* metrics that speak about the structural estimation as a whole and consequently inform us about how the estimated model fits the DGP, and *(ii) parameter-by-parameter* metrics that look at each estimated parameter individually. Most of the literature focus only on the latter and assesses the performance of the estimation based on the bias and standard deviation of each estimated parameter and even sometimes on the sum of the two squared: the Root Mean Squared Error (Smith 1993, Ruge-Murcia 2007, 2012, 2020, Scalone 2018). Equally important is the overall fit, and hence, I stress the importance of these metrics in the discussion of our results as they sometimes draw a different picture.

4.1.1. Overall performance

The most natural metric that speaks about the overall performance of the estimation is the value of the objective function that one is trying to minimize, that is J^{smm} and J^{irf} in equations (2) and (3). These are often refer to as the J-statistic. A recurrent problem with this statistic is that it depends on the units of the weighting matrix W. Consequently, when reporting the value of the J-statistic for the different estimation setups we will fix the weighting matrix to the identity independently of which weights have been used during the optimization stage.

The J-statistic is frequently used in practice because it is easy to compute, however, it only gives an approximate sense of how well the estimated model is able to capture the dynamic responses to various shocks. Given that we control the DGP, one can do better by looking at the distance between the structural IRFs at the true parameter vector Θ^* and at the estimated one $\hat{\Theta}$ as shown below

$$J^* = \left(IRF(\Theta^*) - IRF(\hat{\Theta})\right)' \left(IRF(\Theta^*) - IRF(\hat{\Theta})\right) \tag{9}$$

Equation (9) can be computed for the targeted responses, but also for untargeted ones, e.g. output response to a technology shock when targeting monetary policy responses.

4.1.2. Parameter-by-parameter performance

The literature on DSGE estimation has looked at bias and standard deviations of the estimated parameters when evaluating different methods for obvious reasons. I also look at these metrics but with a small twist motivated by the loss function in Li, Plagborg-Møller, and Wolf (2023). Given the bias-variance trade off in estimated IRFs, and also, the heterogenous researcher's preferences about biases and noise in their parameter estimates, I consider a linear combination of bias and variance with different bias weights as shown below

$$\mathcal{L}_\omega(\hat{\Theta}_i, \Theta_i^*) = \omega \times \underbrace{\left(\mathbb{E}\left[\hat{\Theta}_i\right] - \Theta_i^*\right)^2}_{\text{bias}} + (1-\omega) \times \underbrace{\text{Var}(\hat{\Theta}_i)}_{\text{variance}} \tag{10}$$

Note that for $\omega = 1$, the researcher is only concerned about bias. For $\omega \in (0.5, 1)$ the researcher is more concerned about (squared) bias than variance, while for equal weights $\omega = 0.5$, this metric is proportional to the mean squared error (MSE).

Then, when comparing two different approaches, for example one that uses VARs and other that uses LPs, I will compute the difference between the two loss functions for different bias weights and as a fraction of the true structural parameter value to make deviations comparable across parameters. That is, my preferred measure of parameter-by-parameter performance has the following form

$$z_i \equiv \frac{\left(\mathcal{L}_\omega(\hat{\Theta}_i^{VAR}, \Theta_i^*) - \mathcal{L}_\omega(\hat{\Theta}_i^{LP}, \Theta_i^*)\right)}{\Theta_i} \tag{11}$$

where $\hat{\Theta}_i^{VAR}$ and $\hat{\Theta}_i^{LP}$ denote the parameters estimated when using VAR or LP as the (auxiliary) econometric model, respectively.

5. Results

5.1. The Best Case Scenario: Observed Innovations as Benchmark

I start by discussing the Monte-Carlo results under the assumption that the econometrician observes the true innovation. This is a situation that would never occur in practice, however, it is a good benchmark to initially test the properties of the *Ind. Inf.* and *IRF matching* estimators. The targeted estimated responses under such assumption are depicted in Appendix B.1, where I show that there is a bias-variance trade off between LPs and SVARs in the context of the Smets and Wouters model. But, what are the implications of this trade off for the estimated structural parameters?

Table 2 shows the overall performance metrics for the two estimation strategies and econometric models while averaging across the three sources of variation and the four lag lengths considered. The sample size is set to $T = 300$ observations. Focusing only on the top block that relies on the identity as the weighting matrix for now, one sees that using LP responses as targets in an *IRF matching* exercise is a better idea (lower J^*)

TABLE 2. Overall performance using the observed innovation

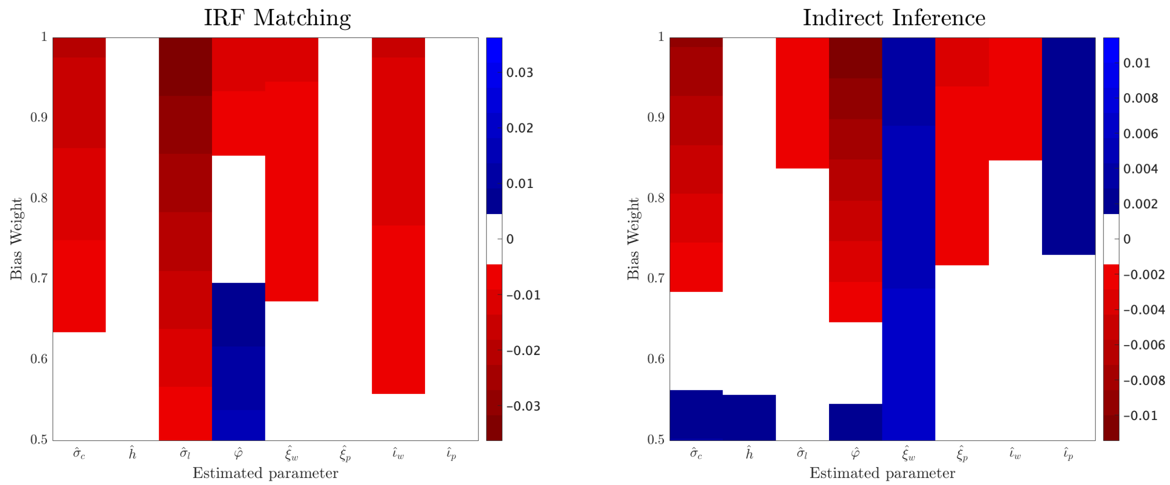
	IRF matching				Indirect Inference			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
				Identity Matrix				
Local Projection	35.10	0.27	3.49 min	18.70	32.54	0.39	42.88 min	17.91
Structural VAR	35.23	0.41	3.93 min	17.93	33.87	0.33	14.47 min	18.39
				Diagonal Matrix				
Local Projection	34.44	0.22	3.61 min	18.87	32.82	0.35	40.56 min	18.22
Structural VAR	34.87	0.27	3.85 min	18.20	34.17	0.31	11.55 min	18.62
				Optimal Weighting Matrix				
Local Projection	33.63	0.04	3.07 min	21.56	32.69	0.06	35.56 min	21.41
Structural VAR	34.17	0.05	3.20 min	20.80	34.26	0.08	10.69 min	20.90

NOTE. This table shows the overall performance metrics and the average computing time for *IRF matching* and *Ind. Inf.* exercises that use either the identity, the diagonal or the optimal weighting matrix.

because their smaller bias. However, this is no longer true in an *Ind. Inf.* exercise where the SVAR approach is slightly better given that SVAR responses have lower variance and their larger bias is irrelevant for the *Ind. Inf.* approach as it is robust to this type of misspecification.

In terms of parameter by parameter performance, what seems to drive these differences between targeting the LP versus the SVAR estimated responses in an *IRF matching* exercise is the lower bias obtained for the inter-temporal and intra-temporal elasticities of substitution $\{\hat{\sigma}_c, \hat{\sigma}_l\}$, as shown in panel A of Figure 1 by the darker red color at $\omega \approx 1$. On the other hand, the better overall performance of the SVAR approach in the *Ind. Inf.* exercise is driven by the lower variance of the intra-temporal elasticity, the habit parameter, the investment adjustment cost and specially the Calvo (1983) probability of wage adjustment $\{\hat{\sigma}_c, \hat{h}_c, \hat{\varphi}, \hat{\xi}_w\}$, as shown by the blue bars in panel B of Figure 1. More generally, it is interesting to observe that for most estimated parameters the LP approach tends to do better when the researcher gives a lot weight to the bias, while the SVAR approach is more desirable under low bias weights.

Lag Length. Understanding these previous results requires to dig deeper into what drives the differences in the estimated IRFs. The lag length is a natural choice as the trade off between LPs and VARs depends on it. As shown in Figure A7 in Appendix C,



A. IRF matching B. Indirect Inference

FIGURE 1. Parameter-by-parameter performance

NOTE. This figure show our preferred measure of parameter-by-parameter performance, equation (11), for both *IRF matching* and *Ind. Inf.* estimation approaches under the identity weighting matrix. Here, for each parameter consider in the estimation, a red color indicates that the LP outperforms the SVAR approach, while the blue color highlights the opposite situation: SVAR better than LP.

LP responses are independent of the lag length and SVAR responses approximately agree with them up to horizon p. It is only beyond horizon $h > p$ where they disagree substantially. In fact, as discussed in Li, Plagborg-Møller, and Wolf (2023), it is the more restrictive way in which SVAR extrapolate long run responses from the first p sample auto-covariances that yields the lower variance at a higher bias. Nonetheless, when increasing the lag length the confidence intervals of the SVAR responses increase and become more alike to those of the LP, which is consistent with the latest “no free lunch” result in Olea et al. (2024). So again, what are the implications of these results on LP and SVAR estimates for the structural parameters when LP/SVAR estimated responses are used as the source of moments in a partial information DSGE estimation?

Table A1 breaks down the overall performance of the *IRF matching* and *Ind. Inf.* approaches by the choice of the lag length. A couple interesting observations arise. First, for the *IRF matching* the J^* from the SVAR gets closer and closer to the LP counterpart as the lag length increases. This is mostly driven by the REDUCTION of J^* associated to the lower bias of the SVAR responses at long horizons. In fact, median and confidence intervals of the targeted IRFs are almost identical when $p = 12$, and consequently, estimated parameters and J^* are very similar too. And second, for the *Ind. Inf.* exercise, which recall is robust to misspecification, the gap in J^* is also decreasing but because that J^*

in the SVAR approach INCREASES as the confidence intervals of the SVAR responses get wider. Overall, it seems that the DSGE modeler will be better off by matching tightly estimated responses, independently of their bias, while using an *Ind. Inf.* approach. However, this comes at higher computational cost as it requires model simulation and IRF estimation at each iteration. Consequently, for some models *IRF matching* may be more suitable and therefore targeting a well estimated response with low bias becomes crucial.

Sample Size. The presence of small sample bias can become an issue for *IRF matching* for obvious reasons, but it can also affect *Ind. Inf.* as long as it also affects the variance of the responses. Consequently, the choice of the sample size used to generate the data moments / targets is another relevant dimension to understand the differences between the estimation approaches studied in this paper. Hence, I repeat the Monte Carlo estimations using a smaller sample of $T = 100$ observations since this is the typical sample length encounter in most macroeconomic applications (Herbst and Johannsen 2023). As shown in Figure A8, small sample bias in LP responses is also present in the baseline sample with $T = 300$ observations, however, it becomes larger when I reduce the sample size. [5] Hence, I consider two avenues: *(i)* I investigate whether *Ind. Inf.* improves upon *IRF matching* when the small sample bias is more severe in the LP approach, and *(ii)* I study whether correcting for bias in the data moments / targets using bias correction terms improves the overall performance of the estimation.

Table A2 addresses these two questions. First, by comparing the LP approach under the two sample sizes one sees that *Ind. Inf.* improves upon *IRF matching* when the small sample bias becomes very large at $T = 100$. Nonetheless, the performance of the estimation under both approaches is worse as the variance of the targets / data moments increases, which can be seen by comparing Figures A1 and A9. The bias in the SVAR is not related to the sample size, but smaller samples also increase the variance. As result, the overall performance when using SVAR responses with $T = 100$ is also worse than when $T = 300$ observations are employed. And second, when I repeat the estimation using the bias corrected versions of the LP and SVAR, discussed in Section 3.3, one can see that correcting for small sample bias is very effective when estimating the model via *IRF matching*. In fact, the J^* is around 1.5 times smaller when bias correction terms are employed to generate the targets. Finally, bias correction in the auxiliary models is not as relevant for *Ind. Inf.* estimation.

[5] Recall that Plagborg-Møller and Wolf's (2021) result about $LP(p)$ exactly agreeing with the structural responses is a population result, i.e. for very large T. Panel C & D in Figure A8 illustrate this point.

Weighting Matrices. All the previous discussions were based on the identity matrix which is a common choice in practice given its simplicity. However, I also explore the choice of two alternative weighting matrices. First, a diagonal matrix that has $1/h$ as its diagonal elements and hence gives a lower weight to the responses at longer horizons. And second, the optimal weighing matrix, which is known to be the inverse of the VCM of the moments.

To understand how this particular choice affects the overall performance of the estimation, start by looking at the second block of Table 2, which shows the J^* under the diagonal matrix. It is not surprising that the differences in J^*'s between using LPs or SVARs shrinks (relative to the identity matrix). Recall that at short horizons SVARs and LP responses approximately agree and consequently putting more weight on these coefficients imply more similar outcomes for the estimation. In fact, this is particularly strong for the SVAR targets in the *IRF matching* approach as it discounts the importance of matching the biased long-run responses. Another interesting observation is that J^*'s are generally lower for both estimation approaches and econometric models. Turning now to the last block of Table 2, which shows the overall performance metrics when the optimal weighting matrix is used, J^*'s are significantly smaller and remarkably close to 0, which is the best possible outcome. Moreover, the use of these more efficient weighting matrices reduces the computational time, with the optimal weighting matrix being the best option among the three.

The parameter by parameter performance under each of these three alternative weighting matrices can be seen in Appendix C.3. The main takeaway is that the mean estimates improve substantially when a more efficient weighting matrix is used. Surprisingly, the improvement in terms of standard deviations is not as large as initially expected.

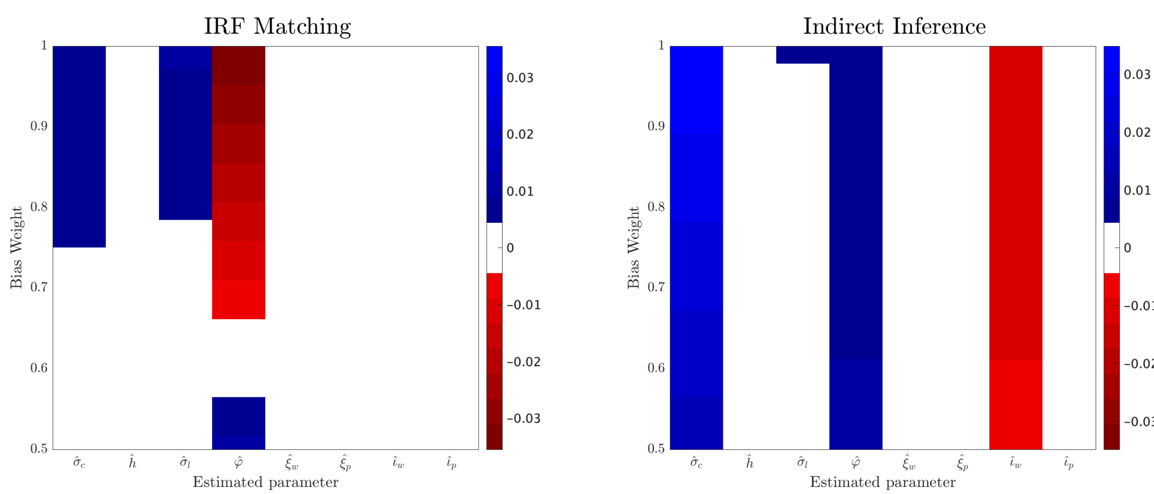
5.2. The Good Old-Fashioned Days: Recursive Identification

I now turn to discuss the estimation set ups that assume that the econometrician does not observe the shock, but it is able to recover it using restrictions based on economic theory. The most widely used approach is to impose zero restrictions on contemporaneous coefficients. As discussed in Section 3.4.2, I will identify technology and fiscal policy shocks by assuming that the policy variable, TFP or government spending, does not respond within the period to other exogenous variables. Importantly, this assumption will not hold for the fiscal policy shock in the Smets and Wouters model because government spending is contemporaneously correlated with the technology shock.

Hence, I will postpone that discussion to Section 5.3 where I address the problem of identifying shocks subject to measurement error and its implications for the structural parameters. The results from targeting technology shocks can be found below in Section 5.2.1. Regarding the monetary policy shock, I instead assume that the policy variable, the interest rate, does not affect other endogenous variables within the period. This assumption does not hold in the Smets and Wouters model either and so I explore what are the consequences of targeting responses to misspecified VAR/LP models in Section 5.2.2 below.

5.2.1. Technology shock

The responses to a technology shock recursively identified within the Smets and Wouters (2007) model are identical to those obtained by assuming that the econometrician observes the innovation/shock. Obviously, the recursive assumption is correct and hence one can recover the true shock via a Cholesky decomposition. Hence, the estimation results using the minimum distance approach will be identical under the two assumptions. The first block of Table 3 shows the overall performance metrics where one sees that the main lessons from Section 5.1 still apply when focusing only on technology shocks. Another interesting observation concerns how the model captures the dynamic



A. IRF matching

B. Indirect Inference

FIGURE 2. Breakdown of Figure 1 by targeted shock – Technology

NOTE. This figure show our preferred measure of parameter-by-parameter performance, equation (11), for both *IRF matching* and *Ind. Inf.* estimation approaches under the identity weighting matrix. Here, for each parameter consider in the estimation, a red color indicates that the LP outperforms the SVAR approach, while the blue color highlights the opposite situation: SVAR better than LP.

response to other shocks, which is measured by J^*_{unt}. Its relatively large values across estimation approaches and econometric models indicate that targeting the response to technology shocks is not a great idea as the estimated model will miss the dynamic responses to fiscal and monetary policy at the optimal parameter vector. Further, Figure 2 shows how LP and SVAR compare when individually focusing on each estimated parameter. Such comparison is informative about the contribution of each estimated parameter to the overall outcome. Actually, one can confirm by looking at panel A that the superior performance from targeting LP-IRFs in the *IRF matching* estimation comes from a more accurate estimation of the investment adjustment cost parameter $\{\hat{\varphi}\}$. Similarly, the SVAR approach to *Ind. Inf.* is also better than the LP approach because it does a better job in pinning down φ. Notice that even though the SVAR approach to *Ind. Inf.* is also better at identifying other parameters, such as the intra-temporal elasticity of substitution $\{\hat{\sigma}_c\}$, these are not so relevant for shaping the responses to technology innovations. In fact, σ_c is better identified when targeting the SVAR-IRFs in a *IRF matching* exercise despite its overall performance is worse than when targeting LP-IRFs.

TABLE 3. Decomposition by the targeted shock

	IRF matching				**Indirect Inference**			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
				Technology shocks				
Local Projection	1.05	0.67	2.87 min	37.30	0.70	0.84	42.41 min	35.92
Structural VAR	2.53	1.07	3.11 min	35.74	0.97	0.66	14.34 min	37.31
				Observed monetary policy shock				
Local Projection	50.65	0.07	3.46 min	9.36	48.46	0.31	41.39 min	9.40
Structural VAR	54.07	0.11	4.38 min	9.26	53.60	0.30	14.65 min	9.44
				Recursive monetary policy shock				
Local Projection	48.11	0.29	3.34 min	9.60	56.91	0.18	78.57 min	9.34
Structural VAR	47.09	0.34	3.78 min	9.31	58.70	0.12	11.44 min	9.34

NOTE. This table shows the overall performance metrics for *IRF matching* and *Ind. Inf.* when estimated responses to technology shocks (top block), observed monetary policy shocks (middle block) or recursive monetary policy shocks (bottom block) are being targeted. In all set-ups we are averaging the results across different lag lengths.

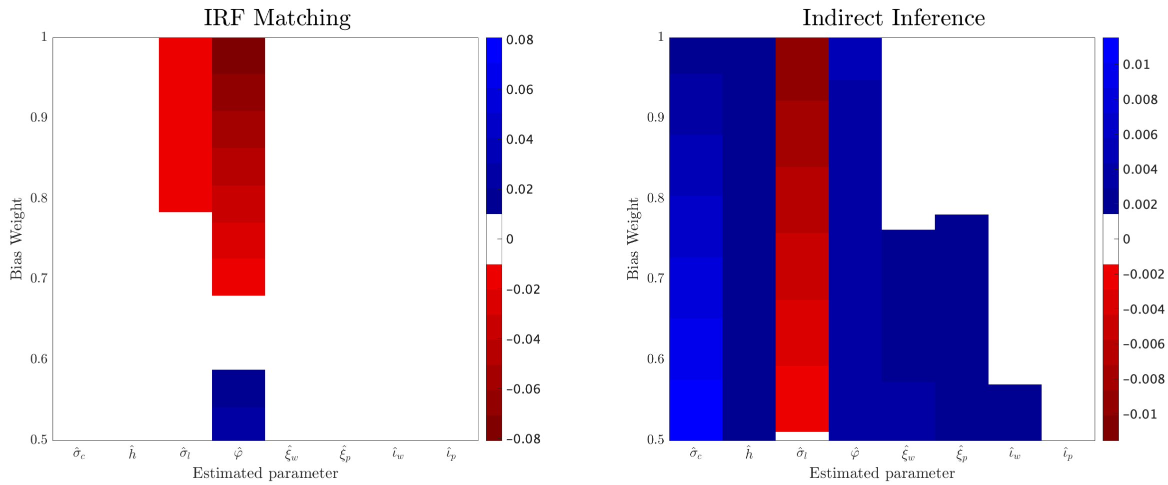
5.2.2. Monetary policy shock

In the Smets and Wouters model a negative monetary policy shock has a positive impact in real activity at time $t = 0$ as shown by the dashed lines in Figure A1 or A3. On the other hand, ordering the policy rate last in the VAR and recovering the responses through a Cholesky decomposition implicitly assumes that monetary policy does not have a contemporaneous impact on other endogenous variables. Consequently, all the estimated responses, either via LP or SVAR, start at 0 when the monetary policy shock has been identified in this way. Obviously, this assumption is at odds with the model. Thus, differently from the technology shock, I now investigate what are the implications for the structural parameters of targeting these misspecified responses.

The middle and bottom blocks of Table 3 show the overall performance metrics when targeting responses to the observed or the recursively identified shocks, respectively. Focusing first on the observed shock, one sees that in line with the previous results, targeting LP-IRFs is a better approach when relying on *IRF matching*, while using a SVAR is better than LP as a binding function for *Ind. Inf.*, even though only by a small margin in this case. Additionally, and differently from the technology shock, targeting the responses to monetary policy shocks are a good idea in the context of the Smets and Wouters model as one would also be able to capture the dynamics of technology and fiscal policy fairly well, as shown by the lower J^*_{unt} (relative to the results obtained by targeting the technology shock).

Turning now to the estimation set-ups that targets the misspecified responses to the recursive monetary policy shock, one can see that when *IRF matching* is the estimation approach, overall performance gets worse as the larger bias of estimated responses relative to the true structural IRFs gets reflected into the estimated structural parameters, independently of the econometric model employed. On the other hand, *Ind. Inf.* is robust to misspecification and in fact improves upon the observed shock case: J^* is lower in the bottom block than in the middle block. This may seem surprising initially, but it is explained by the lower variance of the responses to recursive shocks. Imposing a zero contemporaneous response reduces the bands of the estimated IRFs that are used as data moments, and hence, structural parameters are more tightly estimated. Further, within the recursive shock, the LP approach outperforms the SVAR approach in a *IRF matching* exercise while the opposite is true in the *Ind. Inf.* approach. But what parameters are responsible for these overall estimation outcomes?

Figure 3 compares the difference in parameter-by-parameter losses for various bias weights as shown in equation (11) when a estimated responses to a recursive monetary



A. IRF matching

B. Indirect Inference

FIGURE 3. Parameter-by-Parameter Performance - Recursive Monetary Policy Shock

NOTE. This figure shows our preferred measure of parameter-by-parameter performance, equation (11), for both *IRF matching* and *Ind. Inf.* estimation approaches that target responses to a recursive monetary policy shock under the identity weighting matrix.

policy shock are used as targets / data moments. Starting by panel A in which the *IRF matching* is considered, one sees that again φ plays an important role and is responsible for explaining the better performance of the LP approach. Note that the intra-temporal elasticity of substitution σ_l is also better pinned down by the LP in the *Ind. Inf.* approach but still LP underperforms. In fact, as shown in panel B, almost all the other structural parameters are better estimated with the SVAR as the auxiliary model, independently of the bias weight.

5.3. Direct Proxies for the Shocks: Measurement Error & Unit Effect Normalization

Finally, I present the results of the Monte Carlo analysis that assumes that the econometrician does not observe the true shock but a proxy for it. I distinguish three cases: *(i)* the proxy is contaminated with a white noise error and the econometrician is not aware of it, *(ii)* the proxy is contaminated with a term that is correlated with other shocks in the system / model and the econometrician also does not correct for it in any way or form, *(iii)* the proxy is contaminated with white noise error but the responses are normalized such that the error cancels out by means of the Stock and Watson (2018) unit effect normalization. The targeted moments used in the structural parameter estimation under each of these variants are depicted in Appendix B.3.

5.3.1. Classical measurement error in the innovation

Figure A4 shows that the presence of measurement error in the innovation leads to attenuation bias. It arises from the variance term in the denominator of the least squares estimator and hence it is common to both econometric models, LPs and VARs. Since neither of these two IRF estimators are robust to the presence of measurement error, then using LP or SVAR estimated responses during the structural estimation stage won't affect the results in any different way than in the observed shock case. Nonetheless, the bias associated to the presence of measurement error will still worsen the structural estimation outcome for both *IRF matching* and *Ind. Inf.* estimators. Because targeted responses are now biased towards zero, then those parameters that dampen the IRFs are selected as optimal. Note that the econometrician is not aware of the measurement error and hence uses the true innovation in the model for updating the model counterpart of the IRFs for each parameter vector considered. As a result, the simulated / structural IRFs at each candidate vector do not suffer from attenuation bias. Then, a potential solution that may improve the estimation outcome will be to estimate the variance of the white noise error that contaminates the innovation. As a result, the attenuation bias in the model moments can be introduced through this parameter rather than by driving the structural parameters away from their true values. This extension is left for future work.

TABLE 4. Shock proxies and classical measurement error

	IRF matching				Indirect Inference			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
	A technology shock proxy ($\eta_t^{a,obs}$)							
Local Projection	1.79	1.25	3.05 min	34.30	1.35	1.40	40.23 min	33.31
Structural VAR	3.41	1.70	2.80 min	33.47	1.70	1.18	13.74 min	34.39
	A monetary policy proxy ($\eta_t^{m,obs}$)							
Local Projection	46.81	0.33	3.46 min	9.73	45.99	0.61	40.57 min	9.70
Structural VAR	48.07	0.35	3.72 min	9.43	49.42	0.71	12.51 min	9.77
	A fiscal policy proxy ($\eta_t^{g,obs}$)							
Local Projection	48.05	0.05	4.23 min	8.21	42.52	0.19	45.60 min	7.47
Structural VAR	44.04	0.19	4.07 min	7.80	41.41	0.14	13.73 min	7.62

NOTE. This table shows the overall performance metrics for *IRF matching* and *Ind. Inf.* when the shock used to estimate IRFs has been contaminated with classical measurement error.

The top block of Table 4 shows that the overall performance metrics of the structural estimation that targets estimated responses to a technology shock that features uncorrelated measurement error and confirms the above intuition. As shown by the J^*, the estimation outcome is significantly worse for both LPs and VARs as well as for the *IRF matching* and *Ind. Inf.* estimators relative to the observed shock case. Nevertheless, the lessons from the observed shock case still apply. That is, targeting LP-IRFs is a good idea when resorting to *IRF matching* given their low bias, while using VARs for *Ind. Inf.* is a better choice given their lower variance.

These findings also apply to other sources of variation such as monetary or fiscal policy shocks in which the innovation is also observed with classical measurement error. These results are shown in the middle and bottom block of Table 4. One can see there how the J^* is significantly larger relative to the observed shock case for both *IRF matching* and *Ind. Inf.* estimators.

5.3.2. Correlated measurement error: government spending and its correlation with technology

Now I consider the case in which the observed shock is correlated with other shocks. Recall that in the Smets and Wouters model this is the case of government spending. Differently from the previous scenario here I assume that this correlation is known during the optimization stage. That is, ρ_{ga} is neither set to zero nor estimated, but instead fixed to its true value when updating the model moments during the estimation.

TABLE 5. Correlated shocks & unit normalization

	IRF matching				**Indirect Inference**			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
				A correlated fiscal policy proxy ($\varepsilon_t^{g,obs}$)				
Local Projection	30.82	0.34	4.09 min	7.80	39.05	0.35	46.13 min	10.15
Structural VAR	31.45	0.34	4.19 min	7.78	42.42	0.40	14.20 min	10.53
				A 1% increase in r_0 (Stock and Watson (2018) normalization)				
Local Projection	50.77	0.08	3.83 min	19.34	49.49	0.52	49.84 min	17.85
Structural VAR	53.41	0.32	4.04 min	18.86	51.23	0.42	12.49 min	17.93

NOTE. This table shows the overall performance metrics for *IRF matching* and *Ind. Inf.* when the shock used to estimate IRFs has been contaminated with measurement error.

The targeted responses to a government spending shock are shown in Figure A5.[6] The estimation results from targeting these responses are shown in the top block of Table 5. The J^* is again much larger than in the observed shock case or in the proxy measure with classical measurement error, and for both estimation approaches. Hence, as expected, correlated errors are a bigger issue than uncorrelated ones for structural parameter estimation. Surprisingly, *Ind. Inf.* is not more robust to this type of biases than *IRF matching*. Thus, differently from (misspecified) recursive shocks, there is not an advantage in using *Ind. Inf.* over the traditional *IRF matching* approach when IRFs are estimated using proxies of the shocks.

5.3.3. Unit normalization: a 1% increase in the policy rate

Stock and Watson (2018) has shown a way of dealing with measurement error in the proxy variables by estimating relative rather than absolute impulse responses. The unit effect normalization is shown at work in Figure A6 where I plot the responses to monetary policy shock estimated with a contaminated proxy but whose responses have been normalized such that the policy rate increases by 1% upon impact. The first thing to notice is that the population (dotted line) and the structural (dashed line) responses coincide at all horizons and for all variables. Nonetheless, there are still some discrepancies in finite samples as it was the case for the observed shock identification scheme. But how does this rescaling of the IRFs affect the structural parameters and the overall performance of the structural estimation?

The bottom block of Table 5 shows that the J^* is still larger than in the observed shock case, but the improvement upon the unnormalized responses is substantial. For example, the J^* coming from the *IRF matching* that targets LP-IRFs is 0.07 and 0.08 in the observed shock and normalized responses, respectively; while it equals 0.33 when using the responses to the monetary policy shock contaminated with classical measurement error. As the unit normalization helps in correcting the bias in estimated responses, it is very effective when employed in an *IRF matching* exercise. For *Ind. Inf.* the bias is less relevant and consequently the unit effect normalization is not as effective. In fact, the J^* for the VAR is 0.42 when using the normalization, 0.71 without normalization and 0.30 in the observed shock case. Finally, here the main lesson from the observed shock still applies and using LP-IRFs is better for *IRF matching* while SVAR-IRFs are more effective in *Ind. Inf.* estimations.

[6] These responses are identical to those obtained when ordering government spending first in the VAR and inferring the shock recursively.

6. Conclusion

This paper conducts a Monte Carlo analysis to examine the small sample performance of *IRF matching* and *Ind. Inf.* estimators that target IRFs that have been estimated with LP or VAR models. I drew the following five conclusions:

1. The bias-variance trade off between LP and SVAR estimated IRFs affects the estimated structural parameters obtained with minimum distance estimators such as *Ind. Inf.* and *IRF matching* estimators. Nonetheless, it affects them differently. *IRF matching* is more sensitive to bias in targeted responses and hence using LP-IRFs is preferable, while *Ind. Inf.* is robust to misspecification and hence benefits from the lower variance of VAR-IRFs.

2. The number of lags used in the VAR or as controls in the LP is crucial in understanding not only the differences between estimated IRFs but also in the estimated structural parameters. When the lag length p is large, then IRFs and estimated parameters are similar independently of the econometric model used. On the other hand, when p is small LP-IRFs are less biased and hence better for *IRF matching*, while SVAR-IRFs have a larger bias but lower variance which helps when estimating the parameters via *Ind. Inf.* as the later is robust to these type of biases in estimated responses.

3. The small sample bias of LPs, documented by Herbst and Johannsen (2023), worsens the performance of the structural estimation, specially in the case of *IRF matching*. Using their bias correction term for the targeted moments improves the estimation outcome of the *IRF matching* estimators, while it is irrelevant for *Ind. Inf.* applications.

4. Incorrect recursive identification for the target moments are not an issue for the estimation of structural parameters as long as *Ind. Inf.* is employed. However, it is problematic for *IRF matching*.

5. The presence of measurement error in the proxies used to estimate IRFs worsens the structural estimation outcome for both estimation methods and econometric models. Using the unit effect normalization of Stock and Watson (2018) help ameliorating this problem.

These findings are applicable to a wide range of estimation set-ups and economic models as the Smets and Wouters (2007) contains many ingredients that are still used in many macro models. However, results may vary in the context of a fully non-linear or state dependent model. Thus, a fruitful line of research will be expanding this analysis by using a solution method for this or another economic model that allows to capture non-linear and state-dependent responses. A good starting point is the work of Ruge-Murcia (2020), which already considers non-linear solution and estimation methods but falls short in investigating the trade-offs between non-linear LPs and SVARs as well as different identification schemes for the shocks.

References

Bernanke, Ben S., and Alan S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *The American Economic Review* 82 (4): 901–921. Publisher: American Economic Association.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics* 117 (4): 1329–1368.

Calvo, Guillermo A. 1983. "Staggered prices in a utility-maximizing framework." *Journal of Monetary Economics* 12 (3): 383–398.

Canova, Fabio, and Luca Sala. 2009. "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics* 56 (4): 431–449.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" In *Handbook of Macroeconomics*, edited by John B. Taylor and M. Woodford, vol. 1A, chap. 2, 65–148: Elsevier.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.

Creel, Michael, and Dennis Kristensen. 2011. "Indirect Likelihood Inference." Barcelona GSE Working Paper Series, No. 558.

Fernández-Villaverde, Jesús, Juan F Rubio-Ramírez, Thomas J Sargent, and Mark W Watson. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97 (3): 1021–1026.

Gourieroux, C., A. Monfort, and E. Renault. 1993. "Indirect Inference." *Journal of Applied Econometrics* 8: S85–S118.

Herbst, Edward P., and Benjamin K. Johannsen. 2023. "Bias in Local Projections." Manuscript.

Iacoviello, Matteo. 2005. "House prices, borrowing constraints, and monetary policy in the business cycle." *American Economic Review* 95 (3): 739–764.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *The American Economic Review* 95 (1): 161–182.

Jordà, Òscar, and Sharon Kozicki. 2011. “Estimation and Inference by the Method of Projection Minimum Distance: An Application to the New Keynesian Hybrid Phillips Curve*.” *International Economic Review* 52 (2): 461–487.

Kilian, Lutz. 1998. “Small-sample Confidence Intervals for Impulse Response Functions.” *Review of Economics and Statistics* 80 (2): 218–230.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2023. “Local Projections vs. VARs: Lessons From Thousands of DGPs.” NBER Working Paper No. W30207.

Olea, José Luis Montiel, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf. 2024. “Double Robustness of Local Projections and Some Unpleasant VARithmetic.” NBER Working Paper No. 32495.

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

Pope, Alun Lloyd. 1990. “Biases of estimators in multivariate non-Gaussian autoregressions.” *Journal of Time Series Analysis* 11 (3): 249–258.

Ramey, V. A. 2016. “Macroeconomic Shocks and Their Propagation.” In *Handbook of Macroeconomics,* edited by John B. Taylor and Harald Uhlig, chap. 2, 71–162: Elsevier.

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

Romer, Christina D., and David H. Romer. 2004. “A New Measure of Monetary Shocks: Derivation and Implications.” *The American Economic Review* 94 (4): 1055–1084.

Rotemberg, Julio, and Michael Woodford. 1997. “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy.” In *NBER Macroeconomics Annual,* edited by Julio Rootemberg and Ben S. Bernanke, chap. 12, 297–361: MIT Press.

Ruge-Murcia, Francisco. 2012. “Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles.” *Journal of Economic Dynamics and Control* 36 (6): 914–938.

Ruge-Murcia, Francisco. 2020. “Estimating nonlinear dynamic equilibrium models by matching impulse responses.” *Economics Letters* 197 (109624).

Ruge-Murcia, Francisco J. 2007. "Methods to estimate dynamic stochastic general equilibrium models." *Journal of Economic Dynamics and Control* 31 (8): 2599–2636.

Scalone, Valerio. 2018. "Estimating Non-Linear DSGEs with the Approximate Bayesian Computation: an application to the Zero Lower Bound." Banque de France Working Paper, Number: 688.

Smets, Frank, and Rafael Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–1175.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606.

Smith, A. A. 1993. "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions." *Journal of Applied Econometrics* 8: S63–S84.

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.

Appendix A. The Smets-Wouters Model

The log-linearized equilibrium conditions of the Smets and Wouters (2007) model take the following form:

(A1) $$\hat{y}_t = c_y \hat{c}_t + i_y \hat{i}_t + z_y \hat{z}_t + \varepsilon_t^g$$

(A2) $$\hat{c}_t = \frac{h/\gamma}{1+h/\gamma}\hat{c}_{t-1} + \frac{1}{1+h/\gamma}\mathbb{E}_t\hat{c}_{t+1} + \frac{wl_c(\sigma_c-1)}{\sigma_c(1+h/\gamma)}\left(\hat{l}_t - \mathbb{E}_t\hat{l}_{t+1}\right) + \\ - \frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}\left(\hat{r}_t - \mathbb{E}_t\hat{\pi}_{t+1}\right) - \frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}\varepsilon_t^b$$

(A3) $$\hat{i}_t = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\hat{i}_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}}\mathbb{E}_t\hat{i}_{t+1} + \frac{1}{\varphi\gamma^2\left(1+\beta\gamma^{(1-\sigma_c)}\right)}\hat{q}_t + \varepsilon_t^i$$

(A4) $$\hat{q}_t = \beta(1-\delta)\gamma^{-\sigma_c}\mathbb{E}_t\hat{q}_{t+1} - \hat{r}_t + \mathbb{E}_t\hat{\pi}_{t+1} + \left(1-\beta(1-\delta)\gamma^{-\sigma_c}\right)\mathbb{E}_t\hat{r}_{t+1}^k - \varepsilon_t^b$$

(A5) $$\hat{y}_t = \Phi\left(\alpha\hat{k}_t^s + (1-\alpha)\hat{l}_t + \varepsilon_t^a\right)$$

(A6) $$\hat{k}_t^s = \hat{k}_{t-1} + \hat{z}_t$$

(A7) $$\hat{z}_t = \frac{1-\psi}{\psi}\hat{r}_t^k$$

(A8) $$\hat{k}_t = \frac{(1-\delta)}{\gamma}\hat{k}_{t-1} + (1-(1-\delta)/\gamma)\hat{i}_t + (1-(1-\delta)/\gamma)\varphi\gamma^2\left(1+\beta\gamma^{(1-\sigma_c)}\right)\varepsilon_t^i$$

(A9) $$\hat{\mu}_t^p = \alpha\left(\hat{k}_t^s - \hat{l}_t\right) - \hat{w}_t + \varepsilon_t^a$$

(A10) $$\hat{\pi}_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\iota_p\beta\gamma^{(1-\sigma_c)}}\mathbb{E}_t\hat{\pi}_{t+1} + \frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}}\hat{\pi}_{t-1} + \\ - \frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_p\right)\left(1-\xi_p\right)}{\left(1+\iota_p\beta\gamma^{(1-\sigma_c)}\right)\left(1+(\Phi-1)\varepsilon_p\right)\xi_p}\hat{\mu}_t^p + \varepsilon_t^p$$

(A11) $$\hat{r}_t^k = \hat{l}_t + \hat{w}_t - \hat{k}_t^s$$

(A12) $$\hat{\mu}_t^w = \hat{w}_t - \sigma_l\hat{l}_t - \frac{1}{1-h/\gamma}\left(\hat{c}_t - h/\gamma\hat{c}_{t-1}\right)$$

(A13) $$\hat{w}_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}}\left(\mathbb{E}_t\hat{w}_{t+1} + \mathbb{E}_t\hat{\pi}_{t+1}\right) + \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\left(\hat{w}_{t-1} - \iota_w\hat{\pi}_{t-1}\right) + \\ - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\hat{\pi}_t - \frac{\left(1-\beta\gamma^{(1-\sigma_c)}\xi_w\right)\left(1-\xi_w\right)}{\left(1+\beta\gamma^{(1-\sigma_c)}\right)\left(1+(\lambda_w-1)\epsilon_w\right)\xi_w}\hat{\mu}_t^w + \varepsilon_t^u$$

(A14) $$\hat{r}_t = \rho\hat{r}_{t-1} + (1-\rho)\left(r_\pi\hat{\pi}_t + r_y\left(\hat{y}_t - \hat{y}_t^*\right)\right) + r_{\Delta y}\left(\left(\hat{y}_t - \hat{y}_t^*\right) - \left(\hat{y}_{t-1} - \hat{y}_{t-1}^*\right)\right) + \varepsilon_t^r$$

And the (seven) exogenous shocks evolve according to:

$$\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a \tag{A15}$$

$$\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \eta_t^b \tag{A16}$$

$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \rho_{ga} \eta_t^a + \eta_t^g \tag{A17}$$

$$\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \eta_t^i \tag{A18}$$

$$\varepsilon_t^r = \rho \varepsilon_{t-1}^r + \eta_t^r \tag{A19}$$

$$\varepsilon_t^p = \rho \varepsilon_{t-1}^p + \eta_t^p \tag{A20}$$

$$\varepsilon_t^w = \rho \varepsilon_{t-1}^w + \eta_t^w \tag{A21}$$

Appendix B. Data Moments / Targets

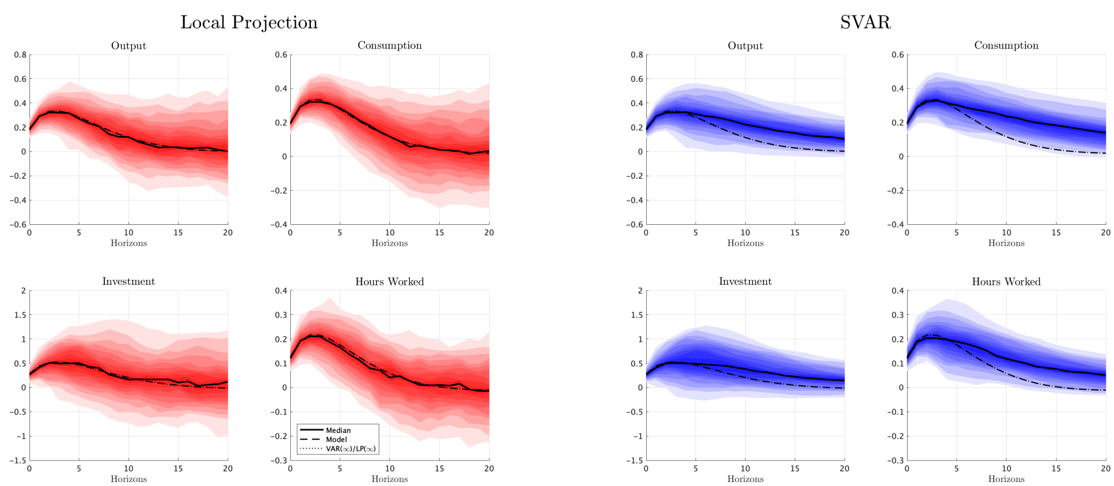
A selection of the impulse responses used as targets or data moments in estimation are depicted below. These are obtained after simulating the model at the true parameter vector Θ^* for each of the $S = 100$ draws of the shocks under different identification strategies and using either LPs or SVAR methods for estimation.

B.1. Observed Innovation

B.1.1. The bias-variance trade off

Figure A1 depicts the response of output, consumption, investment and hours worked to one standard deviation of the monetary policy shock. The dashed line in both panels is the structural IRF that one aims to estimate using either the LP (panel A) or the SVAR (panel B) models. These are depicted with a fan chart to capture the distribution over the different draws of the shock. The median response is plotted with a solid line. In both cases, the sample size is $T = 300$ and the lag length is set to $p = 4$.

From this simple plotting exercise, one learns that the median LP estimated response (solid line in panel A) is very similar to the structural IRF, while the the median estimated



A. Local Projections

B. SVARs

FIGURE A1. Responses to an observed monetary innovation

NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a monetary innovation that have been estimated using either a LP (panel A) or SVAR (panel B) approach and $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF.

SVAR response (solid line in panel B) differs substantially at long horizons. Moreover, the distribution of the LP responses is wider than that of SVAR responses as the latter tend to die out at long horizons. In other words, LP has lower bias than SVAR responses, but it comes at the cost of having also a larger variance than SVARs. This result is consistent with the findings in Li, Plagborg-Møller, and Wolf (2023) and it also present in the response to other shocks within the Smets and Wouters model, such as technology or fiscal policy shocks.

B.1.2. Observed innovation vs. observed shock

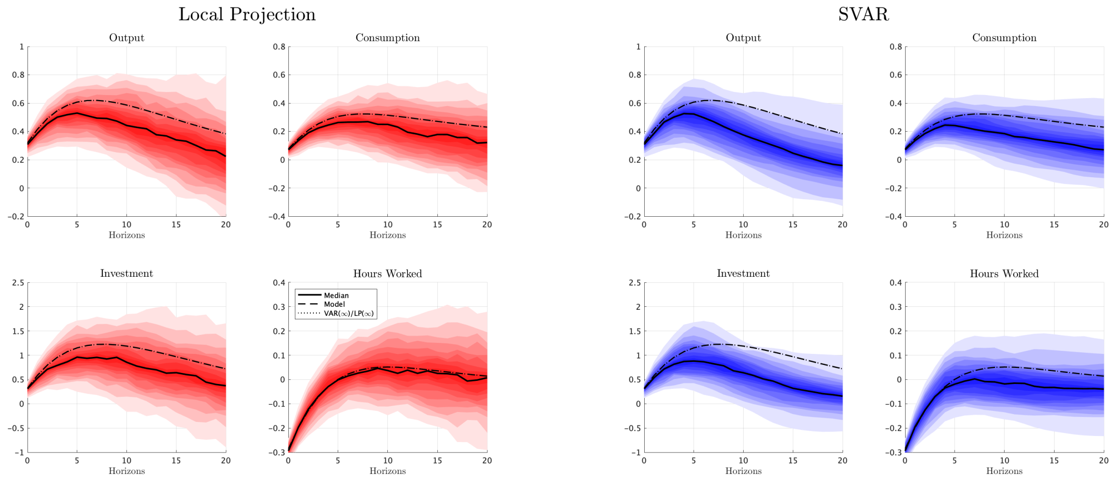
It is common knowledge that using the innovation or the shock itself gives the same impulse responses as long as the shocks are independent and identically distributed. Hence, in the context of the Smets and Wouters (2007) model estimating the responses to technology and monetary policy using the shock, i.e. by setting $\tilde{x}_t$ to ε_t^m or ε_t^a, will give the same answer as to using the innovation itself, i.e. setting $\tilde{x}_t$ to η_t^m or η_t^a. Therefore, the results in Section 5.1 can be also interpreted as if the econometrician were to observe the shock, but with one caveat. Government spending is correlated with the technology shock, as shown in equation (A17), and hence the response to the innovation is not identical to the response of the shock. I will explore the difference between the innovation and the shock when I study the case in which the econometrician observes a noisy measure of the shock of interest – see Section 3.4.3 for a discussion and Section 5.3 for the results.

B.2. Recursive Identification

As discussed in Section 3.4.2, there are two widely used alternatives to identify the shocks through imposing recursive zero restrictions on contemporaneous coefficients. The first one assumes that the policy variable does not respond within the period to other exogenous variables, while the second one imposes that other endogenous variables do not respond to the policy shock within the period. See Ramey (2016) for details.

B.2.1. Technology shock

The technology shock governs by the evolution of TFP in the Smets and Wouters (2007) model. The TFP process follows an AR(1) in logs and it is completely exogenous, as shown in equation (A15). Hence, it is reasonable to assume that the policy variable, TFP, does not respond to other exogenous variables within the period. In fact, that is the



A. Local Projections B. SVARs

FIGURE A2. Responses to a recursive technology shock

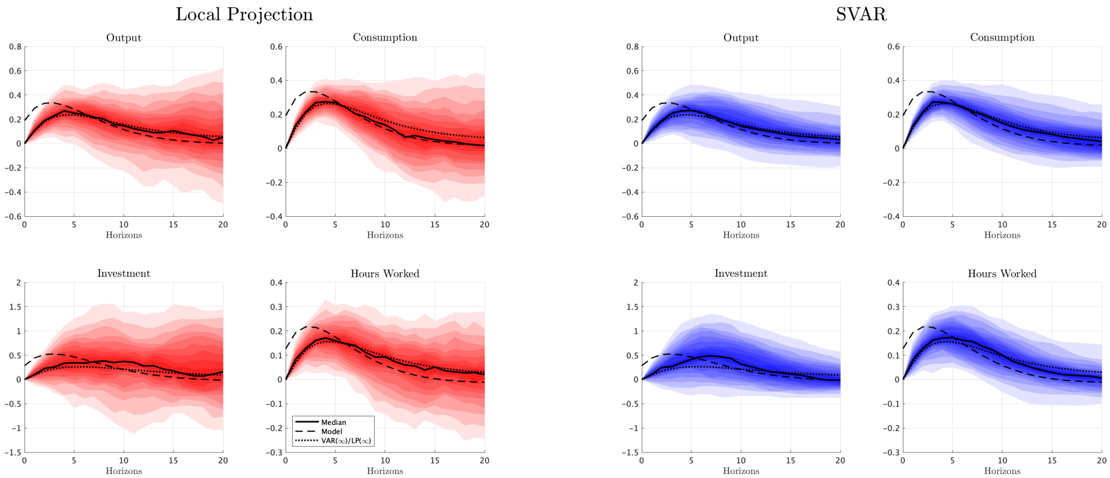
NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a technology shock identified recursively and that has been estimated using either a LP (panel A) or SVAR (panel B) approach with $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF and the dotted line is the population LP/SVAR response with infinite lags.

correct assumption as illustrated by the fact that the population response (dotted line) and the structural IRF (dash line) coincide at all horizons. Therefore, the estimated SVAR-IRFs, which rely on a VAR where I order TFP as the first variable, as well as the LP-IRFs, that set $\tilde{x}_t = \varepsilon_t^a$, coincide with the estimated IRFs under the observed innovation assumption. Figure A2 depicts the distribution of the output, consumption, investment and hours worked estimated responses using the aforementioned recursive identification strategy with $T = 300$ and $p = 4$, and in fact, they are identical to the distribution of responses estimated under the observed shock assumption.

Finally, note that the bias-variance trade off is also present here as well as the small sample bias. These issues concern the estimation approach and are independent of the identification strategy.

B.2.2. Monetary policy shock

For the monetary policy shock I assume instead that other endogenous variables do not respond to the policy shock within the period as it commonly assumed in the literature, see for example Bernanke and Blinder (1992) or Christiano, Eichenbaum, and Evans (2005). Differently from the technology shock, this assumption does not hold within the



A. Local Projections

B. SVARs

FIGURE A3. Responses to a recursive monetary policy shock

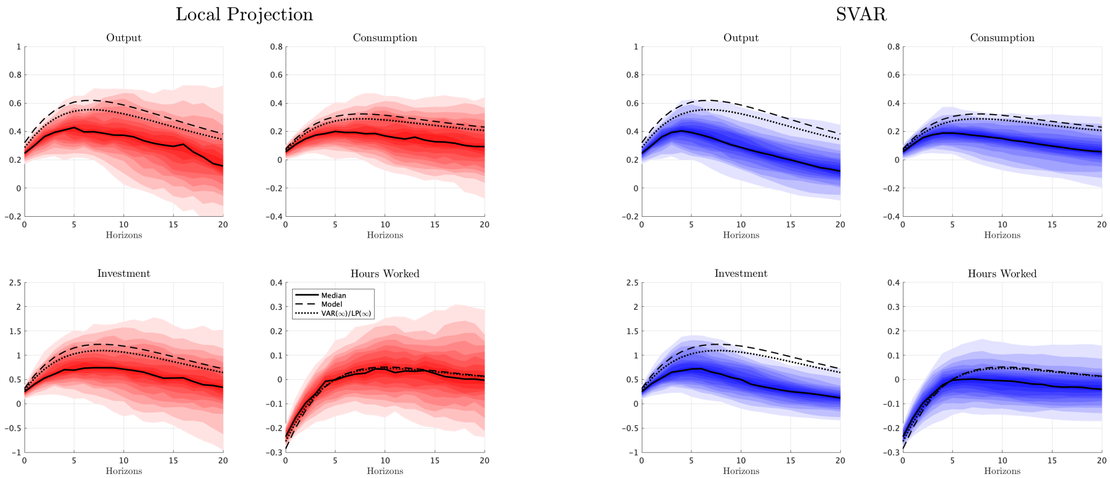
NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a monetary policy shock identified recursively and that has been estimated using either a LP (panel A) or SVAR (panel B) approach with $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF and the dotted line is the population LP/SVAR response with infinite lags.

Smets and Wouters (2007) model. This can be seen graphically in Figure A3 in which the population response (dotted line) disagrees with structural IRF (dash line). In fact, real variables respond contemporaneously to a monetary policy shock in the Smets and Wouters (2007) model, which is ruled out by our identification assumption. The distribution of these estimated responses by either LP or SVAR is also plotted in this figure and features the usual bias variance trade off with respect to the population responses.

B.3. Direct measures of the shocks of interest

B.3.1. Uncorrelated external proxies

Figure A4 plots the distribution of estimated responses to a technology shock under the assumption that the econometrician observes a proxy for the shock and that the noise in the proxy is uncorrelated with other shocks (classical measurement error). As shown from the difference between the dash and the dotted lines, the presence of uncorrelated measurement error lead to attenuation bias. The presence of measurement error increase the variance term in the denominator of the least square estimator and



A. Local Projections

B. SVARs

FIGURE A4. Responses to a mismeasured technology shock

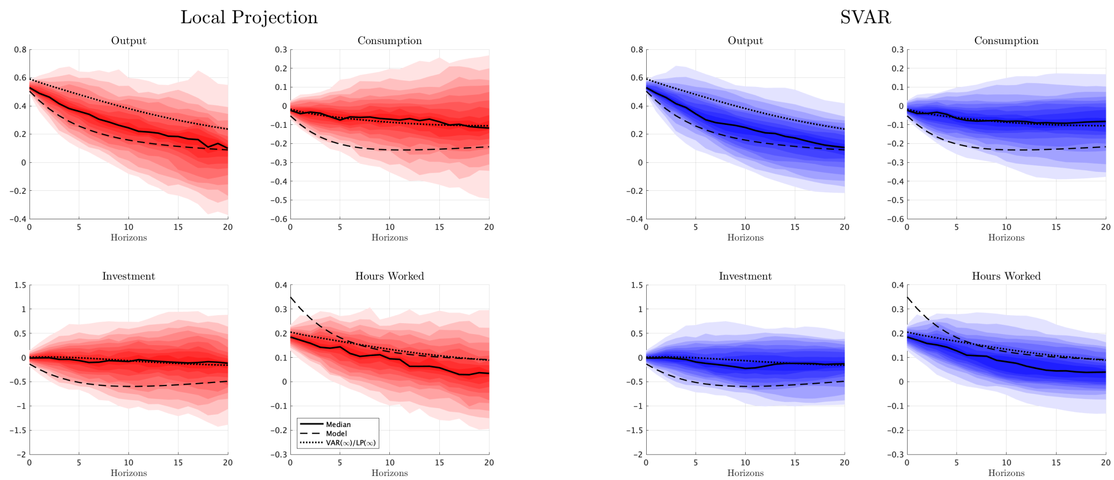
NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a technology shock that is subject to classical measurement error and that has been estimated using either a LP (panel A) or SVAR (panel B) approach with $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF and the dotted line is the population LP/SVAR response with infinite lags.

consequently biases the population response from the VAR(∞)/LP(∞) towards zero. This effect is even more pronounced on the estimated IRFs in a finite sample and using finite lags, as shown by the distribution of LP- and SVAR-IRFs. Importantly, notice that attenuation bias is a problem regarding identification and hence it is common to both estimation approaches.

B.3.2. The correlated government spending shock

Figure A5 show the estimated responses to a fiscal policy shock that uses the correlated government shock, rather than the innovation, and without controlling for TFP. Hence, they can be interpreted as the responses to an identified shock that is subject to correlated measurement error and hence that breaks the exogeneity assumption. As result, the structural IRFs (dash lines) and the population responses (dotted lines) do not agree. Therefore, similarly to the recursive monetary policy shock and the uncorrelated proxies, the estimated IRFs are also misspecified.

Moreover, these responses coincide with those that one would have obtained by assuming that government spending is exogenous and therefore does not affect other endogenous variables contemporaneously. In other words, the recursive identified gov-



A. Local Projections

B. SVARs

FIGURE A5. Responses to a measured fiscal policy shock

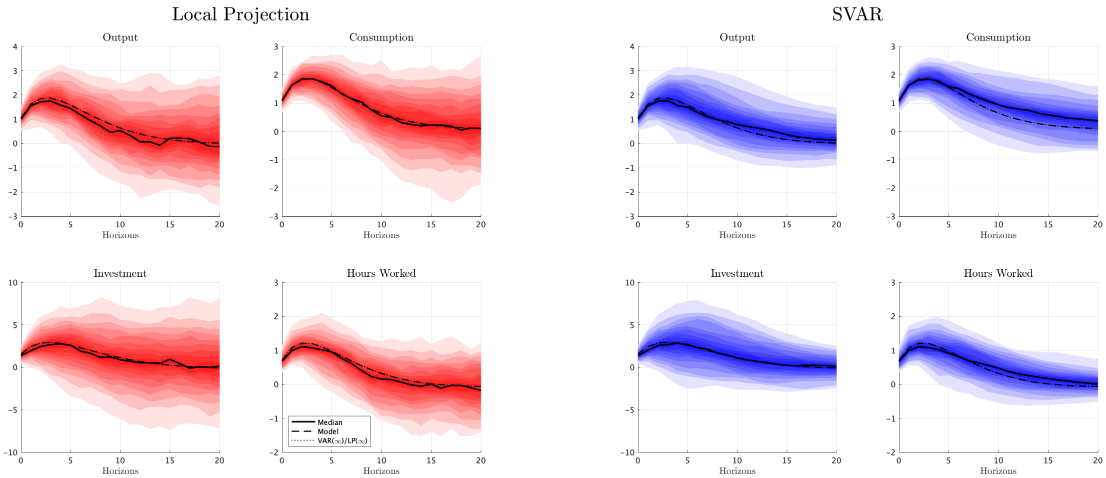
NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a fiscal policy shock that have been estimated using either a LP (panel A) or SVAR (panel B) approach and $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF. Note that these responses are identical to the recursive identified fiscal policy shock that orders government spending first in the VAR.

ernment spending shock leads to the same IRFs as the external proxy that is correlated with technology. Hence, the estimation results from targeting these estimated responses are identical.

B.3.3. Unit normalization with uncorrelated external proxies

Figure A6 shows the responses of output, consumption, investment and hours worked to a 1 percentage point increase in the policy rate. These responses have been obtained after implementing the unit effect normalization of Stock and Watson (2018). That is, the size of the shock has been normalized to unity using the initial impact of the shock on the policy variable, i.e. the policy rate in the case of monetary policy.

As shown by the aforementioned figure, the unit effect normalization helps eliminating the attenuation bias incurred in the estimation that uses proxies that have uncorrelated measurement error. In fact, one sees how the structural IRFs (dashed lines) and the population responses (dotted lines) agree at all horizons and for all variables. Note that these responses are identical to those obtained when employing the true innovation of the shock – see Figure A1 – if they were rescaled by a constant factor that captures the size of the shock.



A. Local Projections

B. SVARs

FIGURE A6. Unit normalized responses to a measured monetary policy shock

NOTE. This figure shows the distribution of the estimated responses of output, consumption, investment and hours worked to a 1% increase in the real interest rate that have been estimated using either a LP (panel A) or SVAR (panel B) approach and $p = 4$ lags. The solid line is the median response, while the dash line is the structural IRF.

The estimated responses also present the bias variance trade-off which is not affected by the normalization of the size of the shock.

Appendix C. Hyperparameter Choices

C.1. Lag Length

The number of lags used in the VAR or as controls in the LP is a fundamental choice that may shape the dynamic response to shocks. Hence, given its relevance for the targeted responses used in a minimum distance estimator, such as the ones consider in this paper, it is also crucial for understanding the structural estimated parameters and the performance of the estimation as a whole.

To shed light on this issue, I plotted the response of output to a monetary policy shock estimated by LP and SVAR models under four different choices of the lag length $p \in \{2, 4, 8, 12\}$ in Figure A7. It shows that: *(i)* impulse responses estimated with the observed innovation and using LPs are independent of the lag length, and consequently, the median and the confidence intervals are similar across the four panels; *(ii)* SVAR-IRFs approximately agree with the LP-IRFs up to horizon $h \leq p$ as shown in Plagborg-Møller and Wolf (2021); and *(iii)* the SVAR confidence intervals converge to those of the

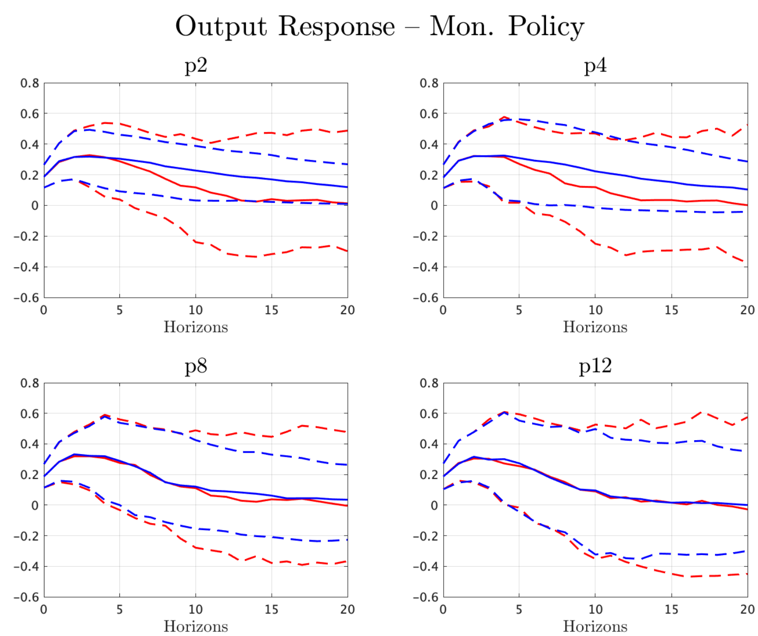


FIGURE A7. Output responses to an observed monetary innovation

NOTE. This figure plots the response of output to a monetary policy shock when it is estimated using either LPs (red) or SVAR (blue) under different choices of the lag length $p \in \{2, 4, 8.12\}$. The solid line is the median response while the dash lines are the 5th and 95th percentiles coming from the different draws of the shock.

TABLE A1. Overall performance & lag length

	IRF matching				Indirect Inference			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
				$p = 2$				
Local Projection	35.75	0.24	3.30 min	18.97	25.47	0.34	18.93 min	18.02
Structural VAR	34.61	0.61	4.32 min	17.00	26.25	0.16	11.88 min	19.32
				$p = 4$				
Local Projection	35.68	0.25	3.40 min	18.74	30.26	0.37	28.99 min	17.95
Structural VAR	36.01	0.39	3.89 min	17.75	31.49	0.26	15.35 min	18.26
				$p = 8$				
Local Projection	34.69	0.28	3.83 min	18.47	35.91	0.44	45.06 min	17.69
Structural VAR	34.92	0.34	3.85 min	18.36	37.26	0.49	13.35 min	18.01
				$p = 12$				
Local Projection	34.27	0.29	3.44 min	18.63	38.52	0.41	78.53 min	17.98
Structural VAR	35.39	0.30	3.67 min	18.61	40.47	0.41	17.29 min	17.98

NOTE. This table breaks down the overall performance of the two econometric models in the two estimation strategies by the lag length.

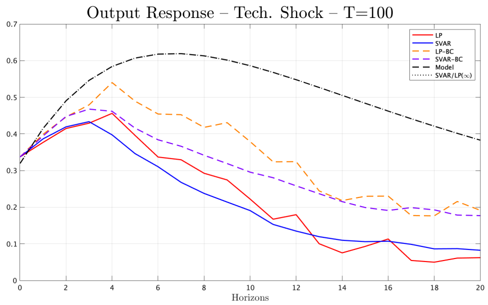
LP as suggested by the theoretical results in Olea et al. (2024). In fact, they also show that increasing the lag length ameliorates the VAR coverage, but at the cost of delivering intervals as wide as those of LP.

These properties are also present when analyzing the responses to other variables as well as other shocks. Hence, they are general enough to help us understand the role of p in the minimum distance estimation that uses IRFs as targets or data moments. Table A1 breaks down by lag length the metrics presented in Table 2 in the main text, where recall I was averaging across different sources of variation as well. As mentioned in Section 5.1, the J^*, the preferred measure of overall performance, is very similar across both econometric models and in both estimation approaches when the lag length is big enough $p = 12$. However, the explanation on why J^* gets closer between LPs and SVARs is very different depending on DSGE estimation method. For the *IRF matching* approach, it is the reduction of the bias in the SVAR-IRFs as p gets large that reduces the value of J^* and consequently the bias of the estimated parameters; while for the *Ind. Inf.* approach is the increase in the variance of the SVAR-IRFs as p gets large that explains the increase in J^* until it converges to the level of the J^* associated with the LP approach. From this table, one can also learn that the estimation time is independent of the lag length in the *IRF matching* approach because the model counterpart of the targeted IRFs are the structural responses which are independent of p. However, for the *Ind. Inf.* approach, the computation time is increasing in p as it requires to estimate more coefficients in each iteration of the minimization problem. This issue is even more acute in the LP approach as its flexibility is associated in part to the larger number of estimated coefficients. Overall, these results seem to call for estimating DSGE models by *Ind. Inf.* and using a VAR with small p as the auxiliary model. Nonetheless, if computational time turns to be a problem, resorting to *IRF matching* while targeting LP-IRFs becomes the second best.

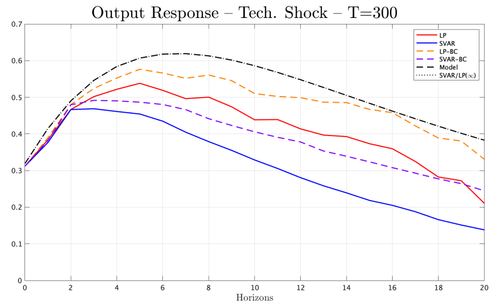
C.2. Sample Size

Herbst and Johannsen (2023) have shown that LP can be severely biased in small samples and proposed an approach to correct for it. I investigate the consequences of this finding, as well as their proposed solution, in the context of DSGE estimation that uses estimated IRFs as targets / data moments in a minimum distance optimization. To shed light on the issue I plot in Figure A8 the estimated output response to a technology shock using LP and SVARs as well as their bias corrected counterparts for different sample sizes. In all scenarios, the simulated sample comes from the Smets and Wouters model at the true parameter vector and the lag length is set to $p = 2$. Focusing initially on the Least Squares LP (solid red line), one sees that the smaller T is, the larger the small sample bias is, and it is only at very large Ts when the estimated response follows closely the structural IRF at all horizons. Moreover, the bias correction LP model of Herbst and Johannsen (2023), depicted by the dashed orange line, partially corrects for the bias in the estimated responses and are closer to the true structural IRF at all horizons and all sample sizes, which validates their Monte Carlo results for a different DGP. Moving into the SVAR-IRFs, one sees that increasing the sample size does not decrease the higher bias relative to the LP. In fact, the SVAR-IRF is very similar across all samples. Nonetheless, the bias correction term from Pope (1990) reduces the bias of the response and brings it closer to the structural IRF.

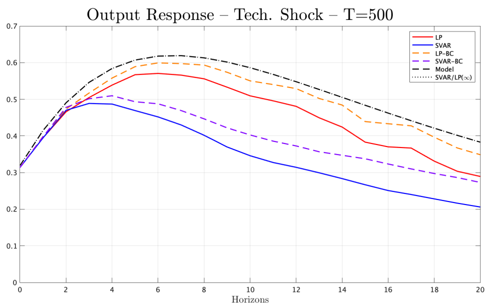
Small sample uncertainty is not only concerning in terms of bias, but also in terms of variance. As shown in Figure A9, the fan chart that depict the distribution of output, consumption, investment and hours worked responses to a monetary policy shock are wider relative to those in Figure A1, which were estimated on a sample with $T = 300$ observations. Hence, the lower sample size can potentially impact the outcomes of both



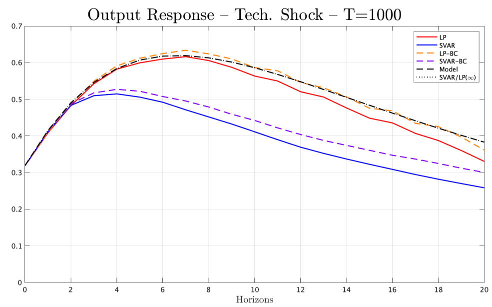
A. $T = 100$



B. $T = 300$



C. $T = 500$



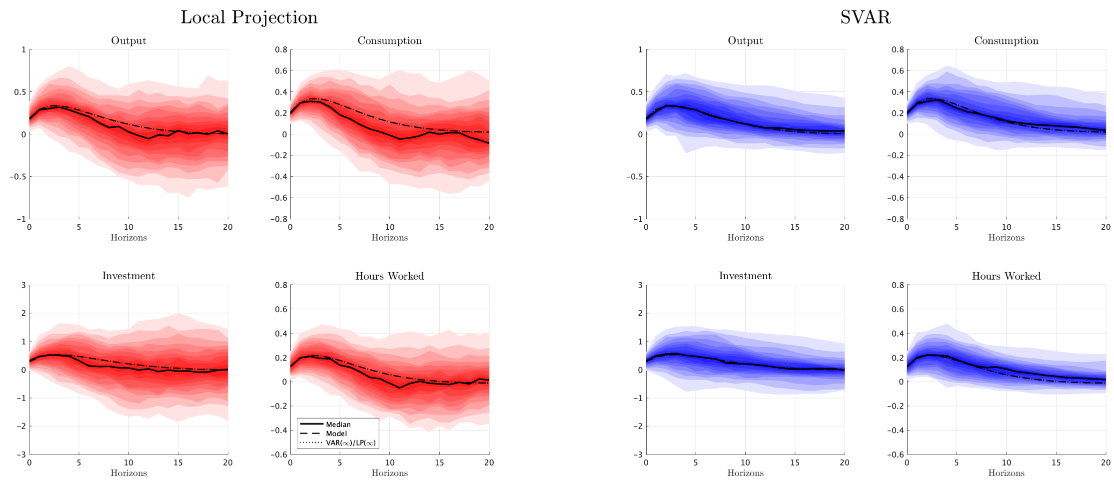
D. $T = 1000$

FIGURE A8. Small sample size & bias correction

NOTE. This figure plots the response of output to a technology shock. The black dash line is the structural IRF at the true parameter vector Θ^*. The other IRFs are estimated using $T = 300$ observations (panel A) or $T = 100$ observations (panel B) by means of Least Squares LP (solid red), Bias Corrected LP (dashed orange), Least Squares VAR (solid blue) or Bias Corrected VAR (dashed purple). To give context to the role of sample size, panels C and D also plot these IRFs for $T = 500$ and $T = 1000$, respectively.

estimation approaches considered in this paper. Intuitively, the increased bias has a larger bite in the *IRF matching* approach, while the increase uncertainty affects the *Ind. Inf.* more as this approach is robust to misspecification in the auxiliary econometric model.

The implications of these results for the overall performance of the estimation are shown in Table A2. As already discussed in Section 5.1, I distinguish between the role of sample size when implementing or not the bias correction in the econometric models. If bias correction is not used, i.e. least squares still being used to estimate LP and VAR coefficients, then the smaller sample worsens the performance of the estimation for both auxiliary econometric models. This can be seen by the larger J^* when comparing the 3rd and 4th row to the 1st and 2nd row in that table. Interestingly, larger bias of targeted responses affects differently *IRF matching* and *Ind. Inf.* approaches. Recall that the later is robust to misspecification. Hence, *Ind. Inf.* outperforms *IRF matching* when the small sample bias in LPs is sufficiently large as shown by the smaller J^* in the 3rd



A. Local Projections B. SVARs

FIGURE A9. Counterpart of Figure A1 with $T = 100$ observations

NOTE. This figure is the counterpart of Figure A1 when using $T = 100$ observations, instead of $T = 300$, to estimate the impulse responses to a monetary policy shock.

row. In other words, if the DSGE modeler suspects that her targeted IRFs can suffer from small sample issues, she will be better off by estimating her model using *Ind. Inf.* techniques.

TABLE A2. Overall performance & sample size

	IRF matching				**Indirect Inference**			
	J_{irf}	J^*	Time	J^*_{unt}	J_{smm}	J^*	Time	J^*_{unt}
				$T = 300$				
Local Projection	35.10	0.27	3.49 min	18.70	32.54	0.39	42.88 min	17.91
Structural VAR	35.23	0.41	3.93 min	17.93	33.87	0.33	14.47 min	18.39
				$T = 100$				
Local Projection	29.71	0.53	3.56 min	18.13	22.00	0.46	18.46 min	19.03
Structural VAR	31.62	0.47	3.33 min	17.98	25.16	0.36	9.78 min	19.50
Bias Corrected LP	31.55	0.32	3.26 min	19.18	23.29	0.35	20.48 min	19.50
Bias Corrected SVAR	33.48	0.32	3.42 min	18.65	26.06	0.33	11.02 min	20.11

NOTE. This table show the overall performance of the estimation when using two different sample sizes to generate the data moments / targets as well as the role of bias correction terms in the estimation of IRFs and its implications for the estimation outcomes.

Regarding the use of bias correction terms such as those proposed by Herbst and Johannsen (2023), the second block of Table A2 shows that they can be very useful in the context of *IRF matching*. In fact, the J^* is significantly lower when using bias corrected responses as targets. For *Ind. Inf.* bias correction seems not to be very relevant as the overall outcome, specially for VARs, is similar to that obtained without bias correction terms. This is a puzzling result as bias correction comes at the cost of higher variance, but this increase in IRF uncertainty doesn't seem to reflect on the structural parameters.

C.3. Weighting Matrices

The selection of the weighting matrix may have a substantial impact on the estimation outcome. Differently from the previous hyper-parameter choices, this choice affects *IRF matching* and *Ind. Inf.* in the same way. Therefore, in Table A3 I only report the mean and standard deviation of the each estimated parameter via the *Ind. Inf.* approach, as the same lessons apply to the *IRF matching* estimates.

First, we have seen that J^*'s decreased as we move away from the identity matrix. This is reflected in the lower bias in key parameters for capturing the dynamic responses to the targeted shocks. In fact, the bias of the inter- e intra-temporal elasticities of substitution and the habit parameter $\{\sigma_c, \sigma_l, h_c\}$ present when using the identity matrix almost disappear when using the optimal weighting matrix. In fact, these three parameters are crucial for capturing the dynamic response to aggregate shocks in the Smets and

TABLE A3. Indirect Inference Estimated Parameters

Parameter	Truth	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
		Identity Matrix		Diagonal Matrix		Optimal Matrix	
$\hat{\sigma}_c$	1.38	1.24	0.38	1.25	0.38	1.37	0.38
$\hat{h}$	0.71	0.77	0.16	0.78	0.12	0.73	0.15
$\hat{\sigma}_l$	1.83	1.88	0.59	1.89	0.59	1.84	0.57
$\hat{\varphi}$	5.74	5.44	1.82	5.43	1.78	5.11	1.77
$\hat{\xi}_w$	0.70	0.62	0.20	0.62	0.20	0.62	0.19
$\hat{\xi}_p$	0.66	0.66	0.20	0.65	0.19	0.64	0.20
$\hat{\iota}_w$	0.58	0.55	0.19	0.56	0.19	0.57	0.19
$\hat{\iota}_p$	0.24	0.23	0.08	0.23	0.08	0.23	0.08

NOTE. This table depicts the true value of the estimated parameters from the Smets and Wouters model. It also displays the mean and standard deviation of each parameter under the three analyzed weighting matrices. The values of the mean and standard deviation are the average and the maximum across the different sources of variation and lag lengths considered, respectively.

Wouters model. And second, it seems that the standard deviations of these parameters are not affected by the choice of the weighting matrix. The optimal weighting matrix slightly improves the efficiency of the estimation but not as much as initially expected.