How Big is the Bias in Estimated Impulse Responses? A Horse Race between VAR and Local Projection Methods Preliminary - comments welcome

André Meier*

First version June 2005,
this version September 2005

Abstract

Impulse response functions are one of the major analytic concepts in modern macroeconomics. However, there is some controversy over how well they are actually estimated by standard VAR procedures. Thus, Jorda (2005) has proposed an alternative estimation approach based on local projections, which is supposedly more robust to dynamic misspecification. The goal of this paper is to evaluate the importance of approximation bias for VAR-based impulse responses and to compare with the performance of local projections. As a data-generating process, we choose Smets and Wouters' (2003) estimated model for the Euro area. We evaluate the performance of impulse response estimators for a wide range of specifications, varying sample size, lag length and the number of variables. Three main results stand out from our analysis. First, across various structural shocks that we consider, estimated impulse responses tend to have the right dynamic shape, conditional upon correct identification. Quantitatively, however, most estimates deviate quite clearly from the true impulse responses in that they are biased towards zero. Second, we do find this bias to be largely confined to small samples. Estimates for a hypothetical large sample of 5,000 observations are mostly quite accurate, questioning the view that small-order VARs are inherently unable to capture the dynamics of a more complex DGP. Third, we do not find evidence that local projections perform any better than standard VAR methods. If anything, our results indicate that they suffer a somewhat greater bias and variance.

JEL classification: C32, E10, E32

Keywords: Impulse Responses, VAR, Local Projections, DSGE, Monte Carlo study

*European University Institute, Via della Piazzuola 43, 50133 Firenze. E-mail: andre.meier@iue.it. I wish to thank Oscar Jorda and Frank Smets for making available computer codes that were used as input for the simulations reported in this paper. All errors are, of course, my own responsibility.

1 Introduction

Impulse responses are one of the major analytic concepts in modern macroeconomics. They are routinely used in theoretical and empirical work to characterize the propagation of shocks in the (model) economy. Apart from serving as a descriptive statistic, impulse responses (henceforth IRs) have also been suggested as input into more involved estimation exercises such as Christiano, Eichenbaum and Evans (2005), who obtain estimates for the structural parameters of their model by minimizing the distance between theoretical and empirical IRs. For these uses to be insightful and reliable, it is important that empirical IRs be estimated sufficiently well. Because IRs are usually obtained from a preceding vector autoregression (VAR) analysis, this means that the (structural) VAR must give a sufficiently good description of the dynamics in the data.

Even abstracting from the important issue of identification, however, there are doubts as to whether likely data-generating processes (DGPs) are well approximated by the low-order VARs that are typically considered in the literature. At a theoretical level, it has been known for some time that many macroeconomic models actually imply vector-autoregressive-moving-average (VARMA) rather than simple VAR structures, at least for the subset of variables usually examined in empirical analyses. As a consequence, standard VAR models are often misspecified, which may give rise to more or less severe biases affecting coefficient estimates as well as the implied IRs. A discussion of the theoretical links between dynamic stochastic general equilibrium (DSGE) models and VAR representations is provided by Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005). Ultimately though, the importance of dynamic misspecification bias is an empirical matter, and a crucial one given the ubiquitous use of IR functions and their common interpretation as "stylized facts".

A few recent studies have addressed the issue in specific modeling contexts, mostly focussing on the identification of a technology shock through long-run restrictions. Chari, Kehoe and McGrattan (2005), for example, conclude from their Monte Carlo study that omission of the variable 'capital' from the VAR introduces a strong bias in the estimated IRs. Yet their analysis does not afford a clear distinction between problems of poor dynamic approximation and those associated exclusively with their specific identification scheme. In addition, their findings have been criticized as pertaining to a (largely irrelevant) special case by Christiano, Eichenbaum and Vigfusson (2005) and Erceg, Guerrieri and Gust (2005), who consider a very similar setup but report much less daunting results. However, even these other papers confine attention to relatively simple models with few structural shocks. Another paper that reports only a limited bias in IR estimates is due to Altig et al. (2005). Although these authors examine a relatively rich specification with three shocks, their model still does not generate sufficient stochastic variation to allow estimation of their ten-variable VAR on the simulated data. As a way around this problem, the authors actually import into the model residual variance from the empirical VAR, whose performance is meant to be scrutinized. The importance of this trick for the findings of Altig et al. (2005) is hard to assess. Moreover, some of the graphs contained in their paper (figure 7) actually show more than just a mild bias. In effect, the incongruity of the particular approaches used in the previous literature may account for the extent of ongoing controversy.

Against this background, it seems desirable to obtain additional evidence on dynamic approximation bias in an improved setup that allows for several structural shocks, clearly nets out problems of identification and, globally, puts the VAR methodology to a realistic test. Above all, this implies that IR estimates should be obtained for data that have been generated from a sufficiently rich and plausible model. A promising candidate, in our view, is the estimated Euro-area model of Smets and Wouters (2003). On the one hand, this model is firmly rooted in the microfoundation paradigm, representing an extended variant of the New Keynesian DSGE framework. Consequently, results obtained for this model can be easily linked to a plethora of ongoing work in macroeconomic theory, which is important not least in view of theory-driven estimation exercises like the one by Christiano, Eichenbaum and Evans (2005). On the other hand, the model allows for a rich dynamics that matches Euro-area macroeconomic data along many dimensions. One important aspect, indeed, is the specification of several structural shocks that ensure sufficient stochastic variation to fit real-world data. Even in terms of forecasting performance, Smets and Wouters (2003) observe that their estimated model is able to compete with more standard, unrestricted time series models such as VARs. The good properties of the model are also confirmed by its use as one of the ECB's four macroeconomic models for the Euro area.

The first step of our analysis, therefore, is to systematically explore the performance of VAR-based IR estimates. This contribution of our paper complements selective prior work by the above-mentioned authors as well as Del Negro et al. (2004), who also consider a version of the Smets-Wouters model and find some bias in conventional IR estimates. Their setup is different from ours in that they add stochastic trends to the model giving rise to issues of cointegration. Furthermore, they consider a more limited range of VAR models and leave out the question of small-sample vs. large-sample bias. Yet this latter aspect strikes us as quite significant and thus will be stressed in our own study, given that bias stemming from dynamic misrepresentation should be clearly distinguished from standard small-sample problems that are known to arise in the estimation of autoregressive processes. Finally, another related paper is by Kapetanios, Pagan and Scott (2005), who experiment with a large-scale macroeconomic model and find severe and persistent biases. Although a very insightful study, this paper is less closely related to most of the current work in theoretical macroeconomics, because it considers a large central bank model rather than a typical medium-scale model all components of which are strictly microfounded. Overall though, the evidence from the more elaborate simulation exercises points towards problems with conventional IR estimates, even if the precise nature, proportion and determinants of the bias are still unclear.

Thus, the second, and principal, objective of this paper is to confront the evidence on bias in VAR-based IR estimates with alternative estimates obtained from local projections (LPs). The idea of considering such "model-free" IR estimates was recently suggested by Jorda (2005), notably in order to address the shortcomings associated with standard VAR methods. Among other advantages, Jorda (2005) stresses that LPs may help to contain the bias arising from dynamic misspecification, especially at long horizons. This is possible because LPs mimic the virtues of multi-step forecasts, whereas VAR-based IR estimates are basically computed as iterated one-step-ahead forecasts, thus compounding misspecification error at increasingly distant horizons. Jorda's (2005) paper actually presents some simulation evidence confirming the greater robustness of LP-based IR estimates. However, the setup, i.e. data generation from an empirical VAR(12), is rather special, and it is doubtful if and how the insights from this exercise extend to DGPs in the standard DSGE framework. Consequently, the present paper aims at a rigorous comparison - across several structural shocks, sample sizes and empirical specifications - between the performance of common VAR methods and Jorda's (2005) LP alternative. In doing so, we abstract throughout from identification issues, specifically by assuming that the correct identification scheme is known. Obviously, this is not meant to imply that identification is straightforward or innocuous in practice, but it seems helpful to separate problems of identification from those concerning poor dynamic approximation. In fact, there is little doubt that wrong identifying assumptions will invalidate the empirical exercise, while the practical relevance of dynamic misspecification is far less clear.

The subsequent exposition is structured as follows. Section 2 discusses the potential shortcomings of VAR-based IR estimates and contrasts them with the suggestion of Jorda (2005) to use LPs instead. The following section, 3, sketches the Smets-Wouters model for the Euro area, from which we are simulating our experimental data. The precise setup of the simulation exercise is detailed in section 4. The results of the exercise are discussed in section 5, and section 6 concludes. All tables and figures are relegated to the appendix.

2 Pitfalls in the Estimation of Impulse Responses

Following Sims' (1980) seminal contribution, VARs have become the most prominent tool of empirical macroeconomic research. Studies investigating the dynamics of a macroeconomy's variables by means of a VAR are ubiquitous in both academic and more applied policy work. Perhaps an even greater proof of their predominance, VARs are routinely used as an auxiliary tool even where they are of no particular interest per se. For instance, researchers often estimate VARs as a means of obtaining IR estimates in order to characterize the transmission of shocks in the data.

However, there is no obvious reason to expect that a set of macroeconomic variables are well described by a low-order VAR. As already Zellner and Palm (1974) and Wallis (1977) pointed out, even when the DGP for the full system has a VAR structure, this may not be true for the subset of variables observed by the researcher. At a fundamental level, the dynamics of modern macroeconomic models often have VARMA representations that are incompatible with finite-order VARs. This may invalidate standard empirical procedures for one of two reasons. First, the model may have a non-invertible moving-average representation and thus no autoregressive representation at all, implying that it is not possible to recover the fundamental shocks from a VAR of any lag length. Lippi and Reichlin (1993, 1994), Hansen and Sargent (1980, 2005) and Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005) have analyzed this scenario. In our subsequent analysis, however, we verify that such invertibility problems do not arise, because our interest lies elsewhere. In fact, the second and, for our purposes, more relevant issue concerns the case where a model has an invertible moving-average representation, but the associated autoregressive representation may not be well approximated with a small number of lags. This problem was emphasized by Braun and Mittnik (1993) and Cooley and Dwyer (1998) and features prominently in the current debate on the merits and pitfalls of standard IR analysis.

The inherent misspecification of parsimonious VAR models has the potential to cause a severe bias that carries over from coefficient estimates to IRs. A particular example of this problem has been highlighted in a recent paper by Chari, Kehoe and McGrattan (2005). The authors show that the omission of an important state variable, i.e. capital in their case, leads to seriously mistaken inference on the economy's response to a technology shock. Although the precise findings of this paper have been called into doubt, e.g. by Christiano, Eichenbaum and Vigfusson (2005), it is clear that a satisfactory accuracy of finite-order VAR approximations cannot be taken for granted. Neither, of course, should the theoretical considerations above be judged sufficient to discard VARs and the implied IR estimates as useful analytic tools. The reliability of standard VAR methods - also compared to available alternatives - is ultimately an empirical question. Unfortunately, the set of existing studies have not provided a conclusive picture, and many stated results appear to be tied to specific expository examples. Overall, it seems fair to say that the studies based on small-scale models with few shocks have provided mixed results on the performance of VAR-based IR estimators, while the larger-scale simulations of Del Negro et al. (2004) and Kapetanios, Pagan and Scott (2005) both indicate considerable bias problems. Even then, however, the effects of changes in the VAR order, the number of variables, or the sample size do not always become clear.

Importantly, problems of dynamic misspecification should be neatly separated from other problems that arise in the context of IR estimation. One such problem is the well-known small-sample bias in estimates of autoregressive processes that has been highlighted by Shaman and Stine (1988), for instance. In fact, there is evidence provided by Erceg, Guerrieri and Gust (2005) that much of the apparent bias in IR estimates is actually confined to small samples, pointing to causes other than fundamental misspecification. Accordingly, our own study will contrast results for (realistic) small sample sizes with those for a hypothetical large sample in order to discriminate between the two sources of bias.

Another great challenge facing empirical researchers concerns identification. While our own strategy will be detailed in section 4.3 below, it should be noted in the current context that a mistaken identification scheme will, of course, tend to exacerbate the aforementioned estimation problems. However, the case of strictly wrong identifying assumptions may not actually be very interesting to analyze.[1] In our view, a sensible Monte Carlo study will instead attempt to reveal the extent of bias conditional upon correct identification. After all, our goal is to isolate and quantify the impact of poor dynamic approximation, which may by itself impair the quality of VAR-based IR estimates quite severely.

2.1 Estimation by local projections

Given the above-mentioned pitfalls of conventional IR estimation, Jorda (2005) proposes to estimate IRs in an alternative fashion, i.e. without prior recourse to an auxiliary VAR model.[2] As a matter of fact, IRs are important statistics in their own right. In particular, they can be estimated for a given vector time series based on LPs that do not require specification and estimation of the unknown true multivariate dynamic system itself. This is an important insight if, indeed, the VAR is thought to be a poor representation of the true DGP. Certainly, IR estimation based on LPs is also affected by misspecification, but Jorda (2005) provides some evidence that the resulting bias can be significantly reduced. There is thus reason to hope that the performance of IR estimators can be improved relative to the standard VAR methodology in cases where a finite-order VAR is necessarily misspecified, e.g. when it is used to approximate a VARMA model.

The potential advantages of LPs stem from their close similarities with multi-step forecasts. Basically, estimation of a VAR on a given sample represents a linear global approximation to the DGP and is optimally designed for one-step-ahead forecasts. However, IRs are functions of forecasts at increasingly distant horizons, so misspecification errors that may appear mild for one-step-ahead purposes tend to aggravate over longer horizons. More generally, the possible dynamics of the IR function deduced from a parsimonious AR model are very limited.

Against this backdrop, Jorda (2005) proposes to rely instead on a collection of projections that are local to each forecast horizon. Specifically, LPs are based on sequential regressions of the endogenous variable shifted several steps ahead. The potential advantage of such direct multi-step forecasts over

[1] This was also the criticism raised by Christiano (2004) against an earlier version of the paper by Chari, Kehoe and McGrattan (2005), which found severely biased IR estimates in a setup where one of the (explicit) identifying assumptions was allegedly wrong. In contrast, the present paper, and most others in the literature, focus on the case with correct identification but a wrong *auxiliary* assumption, i.e. regarding the accuracy of a low-order VAR approximation.

[2] Another model-free method of estimating impulse responses has recently been suggested by Chang and Sakata (2004). However, the authors present their estimator for the univariate case only.

iterated forecasts for autoregressive models whose lag length is too short had already been observed by Bhansali (1996) and Ing (2003). Jorda (2005) applies this insight from the forecasting literature to the related issue of estimating IRs. In the following, we will borrow from the exposition in Jorda (2005) to briefly sketch his econometric approach.

To begin with, note that IRs can be defined without reference to the unknown DGP, namely as a simple difference between two forecasts:

$$IR(t, s, \mathbf{d}_i) = E(y_{t+s}|v_t = \mathbf{d}_i; X_t) - E(y_{t+s}|v_t = \underline{\mathbf{0}}; X_t), \quad s = 0, 1, 2, ..., \tag{1}$$

where $E(\cdot|\cdot)$ denotes the best mean-squared-error predictor, y_t is an $n \times 1$ random vector, $X_t \equiv (y_{t-1}, y_{t-2}, ...)'$, v_t is is the $n \times 1$ vector of reduced-form disturbances, and D is an $n \times n$ matrix whose columns $\mathbf{d}_i$ contain the relevant experimental shocks. The setup proposed by Jorda (2005) accommodates any choice of experiment D without loss of generality. Basically, the procedure focusses on dynamic causation while remaining agnostic about contemporaneous causal relations, or identifying assumptions that help organize the triangular factorization of the reduced-form, residual covariance matrix, $\Omega = PP'$. As a consequence, any identification strategy used in a VAR context can be mimicked by simply choosing the experimental matrix $D = P^{-1}$, so that $\mathbf{d}_i$ represents the "structural shock" to the i-th element in y_t.

The statistical objective is to obtain the best mean-squared-error multi-step predictions. These can be calculated by recursively iterating on an estimated VAR model, which is optimized to capture the dependence structure of successive observations. If the postulated model does not coincide with the true DGP, however, better multi-step predictions can often be found with direct forecasting models that are reestimated for each horizon. Thus consider projecting y_{t+s} onto the linear space spanned by $(y_{t-1}, y_{t-2}, ..., y_{t-p})$ such that

$$y_{t+s} = \alpha^s + B_1^{s+1} y_{t-1} + B_2^{s+1} y_{t-2} + ... + B_p^{s+1} y_{t-p} + u_{t+s}^s, \quad s = 0, 1, 2, ..., h, \tag{2}$$

where α^s is a vector of constants and the B_i^{s+1} are matrices of coefficients for each lag i and horizons $s + 1$. The collection of h regressions in (2) is referred to as LPs. Then, according to definition (1), the IRs implied by the local-linear projections in (2) are

$$\widehat{IR}(t, s, \mathbf{d}_i) = \widehat{B}_1^s \mathbf{d}_i \quad s = 0, 1, 2, ..., h$$

with the obvious normalization $B_1^0 = I$. In practice, Jorda (2005) proposes to estimate all h regressions in (2) jointly so as to improve efficiency.

The main advantage of IR estimates based on LPs for the purposes of this paper is their supposed robustness to misspecification of the unknown DGP. Jorda (2005) provides some evidence that LP-based estimates are indeed superior to VAR-based estimates if the empirical model is dynamically misspecified. However, the setup of his exercise is quite specific and should thus be confronted with a comprehensive evaluation using a prominent macroeconomic model as the DGP. Our choice, i.e. the estimated Euro-area model of Smets and Wouters (2003), will be introduced in the next section. Before that, it is worth emphasizing that Jorda's (2005) approach presents additional merits. For one thing, LPs allow for simple least-squares estimation and inference that does not require asymptotic delta-method approximations or bootstrapping. In addition, they easily accommodate nonlinear specifications provided that sufficient amounts of data are available. Since our DGP is loglinear, however, such potentially useful generalizations will not be taken into account here.

3 The Model

In this section we briefly describe Smets and Wouters' (2003) Euro-area model, from which our data will be simulated. The model is an extended version of the standard New Keynesian DSGE model with monopolistic competition and nominal rigidities. Apart from sticky prices and wages, it features a relatively rich specification of preferences, technologies and real frictions.

The model distinguishes households, firms, and a central bank whose monetary policy is characterized by an interest feedback rule. Households are infinitely-lived and choose their optimal consumption path over differentiated goods provided by firms. Preferences over consumption are characterized by external habit formation, or a motive of "catching up with the Joneses". Further, households have some monopolistic power in the labor market, because they provide differentiated labor services to firms. The model assumes that wages are set in a staggered fashion according to the Calvo (1983) model. Households also rent capital services to firms and decide upon investment given certain capital adjustment costs. Existing capital can be used with varying intensity, subject to some utilization cost schedule. Firms combine labor and capital to produce differentiated goods; they face downward-sloping demand functions and also set prices à la Calvo.

In the following, we will present the log-linearized version of the model. For a detailed description of all equations in levels, see Smets and Wouters (2003).[3]

Intertemporal consumption choice is governed by the linearized Euler equation

$$c_t = \frac{\gamma}{1+\gamma} c_{t-1} + \frac{1}{1+\gamma} E_t c_{t+1} - \frac{1-\gamma}{(1+\gamma)\sigma} E_t \left[r_t - \pi_{t+1} \right] + \frac{1-\gamma}{(1+\gamma)\sigma} E_t \left[\varepsilon_t^b - \varepsilon_{t+1}^b \right], \tag{3}$$

where c_t denotes consumption, r_t is the nominal interest rate between t and $t+1$, and π_{t+1} stands for the corresponding inflation rate. The variable ε_t^b denotes a temporary, but persistent shock to the generic consumer's rate of time preference. Specifically, a positive realization $\varepsilon_t^b > 0$ implies a sudden

[3] Note that the subsequent version of the model reduces the number of structural shocks from ten to seven. While this implies a slight departure from the original setup in Smets and Wouters (2003), it mimics the modeling choice taken in follow-on work by the same authors, i.e. Del Negro et al. (2004). Specifically, we do not consider shocks to the inflation objective, the equity premium and the wage markup. Our motivation for considering no more than seven shocks is to match a VAR setup in seven variables and ensure the invertibility of the model's moving average representation as discussed in Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005). However, it is important to note that all the qualitative results of our paper have been found robust to using the ten-shock version of the model.

unexpected decrease in the discount rate, or greater impatience. The parameter γ indexes the degree of external habit formation. In the absence of consumption habits, i.e. for $\gamma = 0$, (3) simplifies to a standard forward-looking Euler equation, where σ denotes the inverse of the intertemporal elasticity of substitution. For $0 < \gamma \leq 1$, current consumption depends on a weighted average of lagged and future consumption.

Next, investment demand is characterized by

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{\varphi}{1+\beta} q_t - \frac{1}{1+\beta} E_t \left[\beta \varepsilon^i_{t+1} - \varepsilon^i_t \right], \tag{4}$$

where i_t is investment, q_t is the price of capital, β is the household's (mean) time preference rate and φ denotes the inverse elasticity of the function that governs capital adjustment costs. Following Christiano, Eichenbaum and Evans (2005), Smets and Wouters (2003) model these adjustment costs as a function of the change in investment rather than its level. This assumption helps to generate richer dynamic responses to a given shock. Note that a positive investment shock $\varepsilon^i_t > 0$ corresponds to a negative shock to the capital adjustment cost function.

The price of capital evolves according to

$$q_t = -E_t \left[r_t - \pi_{t+1} \right] + \beta \left(1 - \delta\right) E_t q_{t+1} + \left(1 - \beta \left(1 - \delta\right)\right) E_t r^k_{t+1}, \tag{5}$$

where δ is the steady-state depreciation rate, and r^k_{t+1} denotes the rental rate on capital. The current value of the capital stock depends positively on its expected future value and the expected rental rate, and negatively on the ex-ante real interest rate.

The optimal degree of capital utilization is determined by the first-order condition

$$z_t = \psi r^k_t, \tag{6}$$

where z_t denotes capital utilization and ψ is the inverse of the elasticity of the utilization cost function. Intuitively, as the rental rate on capital goes up, existing capital will be used more intensively.

The investment side of the model is completed by the following standard law of motion for capital:

$$k_t = \left(1 - \delta\right) k_{t-1} + \delta i_{t-1}, \tag{7}$$

where k_t denotes the current stock of capital.

Turning to price setting, the generic firm's optimal behavior leads to the following version of the New Keynesian Phillips Curve:

$$\pi_t = \frac{\beta}{1+\beta\kappa_p} E_t \pi_{t+1} + \frac{\kappa_p}{1+\beta\kappa_p} \pi_{t-1} + \frac{\left(1 - \beta\theta_p\right)\left(1 - \theta_p\right)}{\left(1 + \beta\kappa_p\right)\theta_p} \left[\alpha r^k_t + \left(1 - \alpha\right) w^r_t - \varepsilon^a_t + \eta^p_t \right], \tag{8}$$

From (8), inflation can be seen to depend on past and expected future inflation as well as current marginal costs, which, in turn, are driven by the rental rate on capital (weighted by α, the share of capital in the production of goods), the real wage, w^r_t, and the stochastic productivity parameter ε^a_t.

In addition, the model allows for a temporary shock to the markup, denoted by η_t^p. The dependence of inflation on lagged inflation arises from the assumption of indexation: firms that do not reoptimize their price in a given period simply adjust in line with last period's inflation rate. The degree of indexation is captured by κ_p, so the case of no inflation persistence is nested in (8) for $\kappa_p = 0$. Further, the elasticity of inflation with respect to marginal costs depends primarily on the degree of price stickiness; the relevant Calvo parameter θ_p denotes the probability that a firm will not reoptimize its price in a given period. Accordingly, a high θ_p will imply strong rigidities and thus a low pass-through of variation in marginal costs into inflation.

Similarly to (8), an equation for wage inflation can be obtained from the households' optimal wage-setting problem:

$$\begin{aligned} w_t^r &= \frac{\beta}{1+\beta} E_t \left[w_{t+1}^r + \pi_{t+1} \right] + \frac{1}{1+\beta} w_{t-1}^r - \frac{1+\beta\kappa^w}{1+\beta} \pi_t + \frac{\kappa^w}{1+\beta} \pi_{t-1} \\ & -\frac{1}{1+\beta} \frac{(1-\beta\theta^w)(1-\theta^w)}{\left(1+\frac{(1+\lambda^w)\sigma}{\lambda^w}\right)\theta^w} \left[w_t^r - \nu l_t - \frac{\sigma}{1-\gamma}(c_t - \gamma c_{t-1}) - \varepsilon_t^l \right], \end{aligned} \tag{9}$$

where l_t is labor. In analogy to the corresponding parameters for price setting above, κ^w captures the degree of indexation in wage contracts, while θ^w is the Calvo parameter. Thus, the real wage depends on expected and past real wages and the expected, current and past inflation rate, with relative weights determined by the degree of indexation of non-optimized wages. Moreover, the real wage responds to the relative utility cost of working. Specifically, the last term in (9) captures the deviation of the current real wage from what it would be in a fully flexible labor market. Note that ν stands for the inverse of the labor supply elasticity, so a high ν implies a steep labor supply curve. Further, ε_t^l is a persistent labour supply shock. The overall effect of variation in the term in squared brackets will be greater, the higher the elasticity of demand for labor, λ^w, and the smaller the degree of wage stickiness, θ^w.

Next, labor demand reads as

$$l_t = -w_t^r + r_t^k + z_t + k_{t-1}. \tag{10}$$

For a given installed capital stock, labour demand depends negatively on the real wage and positively on the rental rate of capital and on the degree of capital utilization.

The linearized resource constraint for the economy is given by

$$y_t = \frac{C}{Y} c_t + \delta \frac{K}{Y} i_t + \frac{G}{Y} \varepsilon_t^g + (1/\beta + \delta - 1) \frac{K}{Y} z_t, \tag{11}$$

where capital letters denote steady-state levels. Naturally, deviations of output from its steady state are associated with deviations of consumption, investment or government spending (ε_t^g denotes a persistent shock to government spending) from their respective steady-state levels, with weights given by the respective output shares of these components. The last term in (11) captures the resources used up in capital utilization.

The economy-wide production function reads as

$$y_t = \phi\alpha\left(k_{t-1} + z_t\right) + \phi\left(1-\alpha\right)l_t + \phi\varepsilon_t^a, \tag{12}$$

where ϕ denotes one plus the share of the fixed cost in production.

Lastly, the model is closed by adding the following empirical monetary policy reaction function:

$$\begin{aligned} r_t &= \rho r_{t-1} + (1-\rho)\left[\overline{\pi} + r_\pi\left(\pi_{t-1} - \overline{\pi}\right) + r_y\left(y_{t-1} - y_{t-1}^p\right)\right] \\ &\quad + r_{\Delta\pi}\left(\pi_t - \pi_{t-1}\right) + r_{\Delta y}\left[\left(y_t - y_t^p\right) - \left(y_{t-1} - y_{t-1}^p\right)\right] + \eta_t^r \end{aligned} \tag{13}$$

According to (13), the monetary policy-makers gradually respond to deviations of lagged inflation from their inflation objective, $\overline{\pi}$, and to the lagged output gap, defined as the difference between actual and potential output, y_{t-1}^p. Consistent with the DSGE model, potential output is defined as the level of output that would prevail under flexible prices and wages in the absence of the 'cost-push' shock η_t^p. The parameter ρ captures the degree of interest rate smoothing. In addition, there is also a short-run feedback from the current change in inflation and the output gap. Finally, Smets and Wouters (2003) allow for a temporary interest rate shock, η_t^r, which will also be denoted as a monetary policy shock.

Equations (3) to (13) determine the key endogenous variables $\left(\pi_t; w_t^r; k_{t-1}; q_t; y_t; i_t; c_t; z_t; l_t; r_t; r_t^k\right)$ of the model. The stochastic behavior of the system of linear rational expectations equations is driven by seven exogenous shock variables: five shocks arising from technology and preferences $\left(\epsilon_t^a; \epsilon_t^b; \epsilon_t^g; \epsilon_t^l; \epsilon_t^i\right)$, one 'cost-push' shock (η_t^p), and one monetary policy shock (η_t^r). The first set of shock variables (ε's) are assumed to follow independent first-order autoregressive processes with normal innovations, e.g. $\epsilon_t^a = \rho^a \epsilon_{t-1}^a + \upsilon_t^a$, whereas the remaining two shocks (η's) are assumed to be IID normal processes.

4 Simulation

4.1 Parameterization of the model

We simulate the model laid out in the previous section in order to generate artificial data which are then used to estimate IRs. In order to ensure simulating from a plausible DGP, we set all parameter values to the estimates reported in Angeloni, Coenen and Smets (2003). The numbers are reproduced in table 1 in the appendix. The first panel lists the parameters characterizing preferences and technology, the second panel refers to the monetary policy rule (13), and the third panel reports autocorrelation parameters and standard errors for the model's shock processes. All parameter values correspond to the modes of the original Bayesian estimation exercise undertaken in Smets and Wouters (2003). As the authors show in that paper, the data generated from the estimated model match actual Euro-area data along many dimensions. In particular, forecasts from the model are able to compete with standard, unrestricted forecasting tools such as VARs, which is a rather rare achievement for micro-founded macroeconomic models. The good match between model economy and actual data recommends the Smets-Wouters model for our Monte-Carlo exercise, in that the simulation will put both the VAR and the LP methodologies to a sufficiently realistic test. At the same time, the model is firmly rooted in macroeconomic theory, so our study will be informative regarding prominent recent uses of IR analysis in the estimation of DSGE models, notably the minimum distance approaches advocated by Christiano, Eichenbaum, Evans (2005) and others. Moreover, we verified, using the formula developed in Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005, section III), that the above parameterization of the model does not imply invertibility problems for the moving-average representation of our baseline set of seven observable variables. This ensures that our subsequent analysis can safely presuppose a meaningful match between the model's structural shocks and shocks in the VAR, ruling out the sort of fundamental representation issues examined by, for instance, Lippi and Reichlin (1993).

4.2 Steps of the analysis

The first 10,000 points of each simulated data set are discarded in order to avoid starting values having any effect on our results. Accordingly, our samples run from observation 10,001 onwards. The length of samples is motivated by the goal to mimic typical situations facing empirical researchers, who generally do not have more than 20 up to, say, 40 years of data available. We thus consider 80 quarterly observations as our "short sample", and 160 observations as our "long sample". Because we would also like to distinguish between persistent bias, possibly related to omitted state variables, and problems that are confined to small samples, we also consider a hypothetical large-sample data series of length 5,000.

Each of the simulated data sets is used as an input for two types of estimation exercises. First, we estimate a VAR(4) and calculate the corresponding IRs. The issue of identification is discussed below. We will refer to these IR estimates as VAR-based IRs. Second, we obtain alternative IR estimates using the LP approach advocated by Jorda (2005). These IR estimates will be referred to as LP-based IRs below. For comparability, we choose the same lag length of four for the LPs. As a robustness check for our results, we also consider VAR- and LP-based IR estimates based on a lag length of eight. Although a lag length of four seems much more common in current macroeconomic work, a VAR(8) still seems like an implementable option where a sufficient number of observations is available. As another robustness check, we also repeat our exercise allowing the lag length to be chosen based on information criteria. Thus, for every simulated data set, we determine the lag lengths suggested by the Akaike and Schwartz information criteria and then conduct both the VAR and the LP estimation using the respective number of lags.

Apart from differences in sample size and lag length, we also consider differently large sets of variables across both the VAR and the LP approaches. Our baseline set of variables includes output, inflation, the nominal interest rate, consumption, investment, the real wage and hours ('7 variables'). All of these seven variables can arguably be observed quite well in reality and have, therefore, been considered in many empirical studies. Note, however, that even this rather extensive set does not comprise all state variables from the model, e.g. the capital stock. This imitates a realistic practical constraint insofar as empirical researchers rarely have all the variables available that they would like to include in their analysis. As an alternative, we consider an even tighter set of variables that comprises only output, inflation and the nominal interest rate ('3 variables'). Although this setup may seem overly parsimonious, it is not uncommon in the empirical literature either; see, for example, Rotemberg and Woodford (1997). Clearly, if missing state variables constitute an important problem for estimating IRs, we would suspect the three-variable setup to be particularly prone to biases. At the same time, the inclusion of more variables also presents a downside in a sample of given size, if the lower degrees of freedom turn out to worsen existing small-sample problems. A similar trade-off characterizes the choice of the lag length discussed above. Thus, comparing different specifications along these lines will be an interesting side aspect of our study.

Throughout, we choose to consider an horizon of 16 quarters for all IRs. Lastly, we have to fix the number of repetitions for each experiment, i.e. the number of random draws, on which the above-mentioned estimation exercises are conducted. We settle for 1,000, which seems a large enough number to establish the relevant statistical properties of our IR estimates.

4.3 Identification of structural shocks

One important issue in empirical work using VARs is how to identify underlying structural shocks. Basically, identification requires an assumption on the covariance matrix of the reduced-form error terms in different equations of the VAR. A common approach, for example in the literature on monetary policy shocks, is to assume a recursive structure that allows identification based on a Cholesky decomposition of the VAR residual matrix. One recent example of this identification based on short-run restrictions is provided in Christiano, Eichenbaum and Evans (2005). The main alternative strategy consists of exploiting long-run restrictions motivated from theoretical priors, as has been done primarily in the analysis of technology shocks, for instance in Gali (1999).

The most adequate way to proceed, however, is often quite controversial, and choosing the right identification may be a formidable challenge in practice. Basically, two different problems can be distinguished. First, there is the obvious risk of imposing the wrong economic structure, e.g. by ruling out certain contemporaneous responses under a recursive scheme. The situation is more difficult still in the case of long-run restrictions, where even the correct identification strategy may give rise to additional practical problems. Specifically, the identification scheme proposed by Shapiro and Watson (1988) is known to suffer from the prevalence of weak instruments. This point has been observed by Sarte (1997), Cooley and Dwyer (1998) and Pagan and Robertson (1998). Similarly, the identification strategy of Blanchard and Quah (1999) may be badly affected by the poor estimates of long-run effects in finite samples, as has been noted by Faust and Leeper (1997). In both cases, there is the potential for strong biases directly related to applying a long-run identification strategy in a small sample.

Notwithstanding the importance of these considerations, they are quite distinct from the issue this paper is focussing on. Specifically, our goal is to find out if - even conditional upon accurate identification - IR estimates may still be unreliable just because the approximating model does not capture the true dynamics in the data sufficiently well. In order to isolate this aspect of dynamic misspecification, it should not be confounded by an additional practical problem, i.e. the issue of identification. Thus, we abstract throughout from the difficulty of making the right identifying assumptions. Practically, we do this by imposing the true structure of contemporaneous causal linkages as implied by the theoretical model. While this entails that the estimated impact IRs necessarily coincide with the true ones, the IRs at all subsequent horizons are also affected by the quality of the VAR or LP estimates. Accordingly, to the extent that the VAR or the LP miss or misrepresent part of the actual dynamics in the data, IR estimates at horizons two and higher will systematically differ from the true IRs. In this regard, our analytical strategy is identical to the one proposed by Kapetanios, Pagan and Scott (2005) and, in effect, very close to the one by Del Negro et al. (2004), although these authors operate in a Bayesian framework.[4] Naturally, the extent of inference problems unveiled by our exercise should be seen as a lower bound, given that practitioners are faced with the additional intricacies of accurate identification.

Based on the aforesaid, the relevant metric of success is the extent of systematic differences between true and estimated IRs at all but the first horizon. In terms of structural shocks, we choose to focus attention on three shocks that seem very important according to at least one of two criteria. First, the shock qualifies if it is an important source of variation in the estimated model of Smets and Wouters (2003). This is true particularly for the monetary policy shock and the labor supply shock, both of which are shown to have a strong impact on the overall fluctuations of output, inflation and other key macroeconomic variables. A second criterion concerns the prominence of certain benchmark exercises in the empirical literature. Indeed, apart from monetary policy shocks, particular attention has been given to technology shocks. Against this background, we consider this shock in our analysis, as well.

[4] In contrast, all other authors in the literature have imitated, in their Monte Carlo studies, the practical difficulty of identification, i.e. they have actually estimated the relevant matrix transforming reduced-form into structural shocks. While this certainly makes the exercise more realistic, authors have generally been unable, as a consequence, to discriminate between different sources of bias. This is unfortunate, especially since most studies have actually focussed on the most intricate case of identification through long-run restrictions. One notable exception is the recent paper by Erceg, Guerrieri and Gust (2005), which also considers long-run restrictions but proposes a smart way of isolating and quantifying the bias directly associated with that empirical identification strategy.

5 Results

5.1 Baseline findings for monetary policy shock

5.1.1 Median bias

In the following, we will discuss the IR estimates for each of the three structural shocks, in turn. Beginning with the prominent case of a monetary policy shock, note that the solid black line in figure 1 represents the true IR implied by our model evaluated at the parameter values listed in table 1. All other six lines depict the median estimates of IRs for different empirical specifications.[5] Specifically, the red lines represent VAR-based estimates for samples of size 80 (dotted line), 160 (dash-dotted line), and 5,000 (dashed line), while the blue lines show the corresponding LP-based estimates. All estimates were obtained using the '7-variable' setup and four lags. Lastly, the shaded areas indicate 95 % confidence intervals obtained from the VAR-based estimates for a sample of size 80. Thus the upper/lower bound of the shaded region represents the 97.5 % (2.5 %) quantile of the empirical distribution of VAR-based IR estimates, based on 1,000 random simulations.

Several interesting observations stand out from figure 1. First, for all seven series, the estimated IRs tend to trace the rough general shape of the true responses. Remember in this context, however, that the exact coincidence of true and estimated IRs upon impact is imposed by our identification scheme, while all subsequent dynamics of the empirical IRs are driven by the estimated VAR or LP coefficients, respectively. Quantitatively, the performance of the estimators is rather weak. In particular, the estimated IRs tend to show considerably less persistence than their true counterparts. Even in the case of the federal funds rate, which still offers the best fit among all series, the estimated responses return to the zero line more quickly than the true IRs. The same sort of bias is visible across all series, although it is particularly conspicuous for the real wage response, where some estimates suggest a peak decline of less than one third of the actual peak decline. Correspondingly, the true IR of the real wage is on the lower boundary of the 95 % confidence interval for the VAR-based IR estimates for sample size 80. Based on these estimates, a researcher would miss the dynamics of the Smets-Wouters model quite clearly.

Nonetheless, a second observation is important, which concerns the relative performance of IR estimates for samples of different size. In fact, while the estimates based on 80 quarterly observations are outright disappointing, the results look somewhat better for the case of 160 observations. Throughout all series, the estimated curves move closer to their true counterparts with the doubling of the sample size. More strikingly still, the bias completely disappears (and the two dashed lines virtually coincide with the solid black line) when we consider estimates for a large sample of 5,000 observations. Of course, this insight may provide limited comfort to empirical researchers, who will find sample sizes of 80 or 160 much more realistic. Still, the large-sample case serves as an important

[5] We also considered the means of estimated IRFs and obtained nearly identical results.

reference point in that it highlights the nature of the observed bias as a small-sample problem. This is clearly at odds with the notion that omitted state variables fundamentally preclude an accurate approximation of complex DGPs. On the contrary, even our simple low-order VAR seems, in principle, able to capture the dynamics of the Smets-Wouters model. Note in this context that the lag length was kept constant at four through the increase in sample size. Accordingly, the improved properties of the estimates come exclusively from a greater number of observations available. In this respect, our results align well with the evidence reported by Erceg, Guerrieri and Gust (2005), who also find bias problems primarily associated with small samples. Conversely, our results caution against the strong dismissal of (VAR-based) IR estimation expressed in Chari, Kehoe and McGrattan (2005). Even though finite-sample bias constitutes a serious problem, the omission of state variables does not, in our case, turn out to be as universally destructive as these authors argue.

All previous statements have applied to VAR and LP results indiscriminately. As a matter of fact, the median estimates for both approaches are generally quite close. Still, our analysis provides an informative comparison between their performances. Based on Jorda's (2005) own Monte Carlo evidence, one should expect the LP methodology to provide more accurate IR estimates than the standard VAR approach. Perhaps surprisingly though, our results point in the opposite direction. For all variables and both sample sizes 80 and 160, the LP estimates are actually a little worse than the corresponding VAR results. In figure 1, this can be seen from the fact that the blue curves, which represent LP estimates, tend to be even further from the true IRs than the red curves, which depict the VAR-based estimates. In order to formalize this comparison, we have calculated the weighted squared deviations of the shown median estimates from the true IRs. The results are given in table 2, which also contains the corresponding results for other specifications discussed below. The numbers are obtained from computing the difference between median estimates and true IRs for periods 2 through 16 (on impact, the responses necessarily coincide) across all seven series. These differences are then squared and weighted by the inverse of the point-wise variance at each estimated response, where variance estimates are obtained from simulations of the relevant VAR. Accordingly, IRs that are estimated with little precision will be downweighted, and the final statistic, desirably, informs about squared distance in terms of standard errors. As can be seen from the relevant cells of panel a), the optical impression is neatly confirmed: VAR-based estimates deviate less from the true IRs than those estimates based on LPs à la Jorda (2005).

5.1.2 Variance

From the viewpoint of an empirical researcher, however, median bias is not the only relevant consideration. In order to also take the variance of estimates into account, table 3 reports the average (over 1,000 random simulations in each case) weighted squared deviation of IR estimates from the true responses. In contrast to the statistics in table 2, we now calculate the weighted distance for every single estimation and report the average thereof. Thus, apart from taking bias into account, the statistics in table 3 also reflect the relative variance of estimates. For instance, the numbers in table 3 may be big not only if the estimates consistently miss the true IRs but also if they fluctuate excessively around a good mean estimate. As the numbers in panel a) show, the LP approach again does not look like a promising alternative to the standard VAR methodology: the difference between VAR and LP results for a given specification is fairly large throughout.

Lastly, note that tables 2 and 3 also confirm our previous statements regarding improvements with growing sample size. To be sure, the comparison of numbers across different sample sizes should be done with some caution, since the weights used in the computation are constant only within neighboring VAR and LP cells for the exact same specification and sample size. Still, we may at least infer that the distance between estimates and true IRs *in terms of standard deviations* continuously diminishes as the sample size increases.[6]

All things considered, our simulations indicate the presence of considerable small-sample bias in IR estimates for a monetary policy shock, with VAR-based estimates even slightly outperforming the supposedly superior LP methodology.

5.2 Alternative empirical specifications

Before repeating the above exercise for other shocks, we would like to check if our previous results can be confirmed for different specifications considering the same monetary shock. For this purpose, figure 2 contrasts our baseline findings (seven variables, four lags) with two alternative cases, i.e. a) estimation of VARs and LPs on a tighter set of variables (only inflation, output and the nominal interest rate) with again four lags, and b) estimation on the same 7-variable set but with eight lags. For ease of exposition, figure 2 depicts the IRs of only three variables even where the empirical exercise involves all seven variables. In addition, we confine attention to a sample size of 160, not least because the smaller sample would not seem sufficiently big for a sensible estimation exercise using eight lags.

The upper panel of figure 2 suggests that moving to a smaller set of variables actually improves our estimates for a given sample size. Throughout, the dashed lines (for the 3-variable case) are closer to the true IRs than the dotted lines imported from the 7-variable case in figure 1. This result could not be easily foreseen, because there are two effects operating in opposite directions. On the one hand, the smaller set of variables means that even more relevant state variables are omitted from the empirical specification, potentially worsening the bias in IR estimates. On the other hand, a more parsimonious specification will mean more degrees of freedom for estimation in a given sample, so the small-sample problem noted above may simply be less potent. Clearly, our findings indicate that the former effect is more than offset by the latter, so there may be some gain from keeping specifications small despite

[6] Note that comparisons across different shocks require even more caution, because small numbers may be caused by good overall fit as well as by high variance of the respective VAR estimates, which tends to push down the point-wise weights.

the implied omission of perhaps important state variables. As a matter of fact, the remaining bias is actually quite contained for the 3-variable case. This insight may, of course, not be exploitable if the responses of additional variables are of direct interest to the researcher. Finally, as far as the relative performance of VAR and LP methods is concerned, the picture is roughly the same as before: the average weighted squared deviations from the true IRs are larger for LP-based estimates (see table 3), although the difference in terms of median bias becomes rather small for a sample size of 160 and even changes sign for the 5000-observation sample, where there is almost no bias left and the LP-based estimates look even a little better than the VAR benchmark (see table 2).

The lower panel of figure 2 depicts the true IRs of the three key macroeconomic variables along with estimates obtained using four and eight lags on a sample of 160 observations, respectively. The trade-off is similar to the one discussed above. More lags may allow a better approximation of the true DGP, but in a sample of given size they are costly in terms of degrees of freedom. As before, the latter effect turns out to be more important, so the baseline specifications with only four lags actually outperform those with eight lags. This is true for both the VAR and the LP methodology, while the relative ranking of the two remains unchanged.

Finally, we explore the possibility of specifying the lag length based on information criteria rather than imposing four or eight lags a priori. For each of the 1,000 randomized data sets, we thus determine the number of lags suggested, respectively, by the Akaike information criterion (AIC) and the Schwartz information criterion (SIC) and then re-estimate VARs and LPs using the relevant lag length. We conduct this exercise for both the 7-variable and the 3-variable case, considering, in turn, 80 and 160 observations. Where only 80 observations are available, we set the maximum lag length to six, while it is fixed at eight for 160 observations. As the bottom panel of table 4 indicates, the criteria suggest a relatively low number of lags throughout. The lag length implied by the Schwartz criterion, for instance, never exceeds three across all specifications and replications. The Akaike criterion, while naturally suggesting a higher number of lags, also keeps specifications rather parsimonious. The only exception is the 7-variable setup for 80 observations, where the median suggested lag length over 1,000 replications is six. Turning to the upper panel of table 4, we can examine what the criteria-based choice of lag length implies for the relative performance of our IR estimators. Throughout, the findings look very familiar. The median bias is considerable, although it decreases in the size of the sample. Also as before, the economy's response to a monetary policy shock tends to be traced with greater accuracy when IR estimates are based on a VAR than when they are taken from LPs. Only in the 3-variable setup with 160 observations do we find one instance, where the LP estimates actually outperform the VAR-based estimates in terms of median bias. Nonetheless, the LP estimates remain consistently inferior if variance is also taken into account. This can be seen from table 5, where, in analogy to table 3, we report average weighted squared deviations for the different estimates.

In sum, our basic conclusions are confirmed across a number of alternative specifications.

5.3 Results for technology shock

In a next step, we repeat all previous exercises for a different structural shock, i.e. a shock to productivity. The relevant results are presented in figures 3 and 4 as well as in panels b) of tables 2 through 5. Note first from figure 3 that the bias in IR estimates looks somewhat more moderate now than for the monetary policy shock analyzed in the previous section. Although the estimates for a sample of 80 observations are still clearly off the true curves for most variables, the estimates for the larger 160-observation sample are already quite good. For 5,000 observations, as before, the median estimates basically coincide with the true IRs. In terms of relative performance, it is visible from both the figures and the tables that Jorda's LP approach continues to underperform the VAR benchmark for all realistic sample sizes. Likewise, panel a) of figure 4 indicates that a more parsimonious set of variables works again towards reducing the small-sample bias in IR estimates. In fact, the estimates obtained for the "small" VAR and LP setups are very good: there is hardly any bias left even in a sample of 160 observations.[7] Relatedly, panel b) of figure 4 confirms our previous conclusion that increasing the number of lags to eight actually worsens the quality of the IR estimates.

Overall, the results obtained for a technology shock are more comforting than the ones discussed in the previous section. The small-sample bias of IR estimates is still visible but more limited than in figures 1 and 2. Especially the results obtained for the small 3-variable set are basically as good as one could have hoped for. This statement naturally has to be qualified to the extent that, in practical work, the estimation of IRs for a technology shock may also be troubled by formidable identification problems, which we have sidestepped throughout. Irrespectively, we still do not find any evidence that would suggest LP-based estimates as a superior alternative to standard VAR methods.

5.4 Results for labor supply shock

The third and final structural shock we wish to consider is a shock to labor supply. The results pertaining to this set of exercises are reported in figures 5 and 6 as well as panels c) of tables 2 through 5. Generally, the findings are rather similar to the results obtained for the monetary policy shock. In particular, the small-sample bias is quite pronounced, diminishes somewhat as the sample size is increased from 80 to 160 and disappears completely for a sample of size 5,000. As figure 6 indicates, a more parsimonious set of variables is again conducive to reducing small-sample bias in a given sample, while considering more lags slightly worsens the results. In addition, the VAR-based estimates appear again superior with one notable exception, i.e. the case of three variables and a sample size of 160. Here, Jorda's (2005) LP-based estimates have a slightly smaller median bias, provided that the lag

[7] However, note from the relevant entries in table 2 that the 3-variable case differs from previous results insofar as the median bias (in terms of standard deviations) now fails to improve monotonically with the sample size. To be sure, the optical fit for the case of 5,000 observations is still very good, but it does not reach the accuracy observed for other specifications. Due to the very small variance of estimates for this sample size, the remaining bias shows up through fairly high numbers in tables 2 and 3.

length is relatively short, i.e. fixed at four (table 2) or chosen based on information criteria (table 4). At the same time, however, tables 3 and 5 also suggest that this marginal advantage is offset by a greater variance of the LP estimates. Thus, all of our previous conclusions go through.

6 Conclusion

Impulse responses (IRs) are one of the major analytic concepts in modern macroeconomics. They are routinely used in theoretical and empirical work to characterize the propagation of shocks in the economy. Apart from serving as a descriptive statistic, IRs have also been suggested as input into more involved econometric exercises such as minimum distance estimation. For these uses to be insightful and reliable, it is important that empirical IRs be estimated with sufficient accuracy. Because IRs are usually obtained from a preceding VAR analysis, this means that the VAR must give a sufficiently good description of the dynamics in the data. However, previous theoretical and empirical work has raised concerns that poor dynamic approximation may often give rise to biases in VAR-based estimates of IRs. Although the quantitative importance of such biases is a topic of ongoing debate, there is clearly some demand for alternative methods that are more robust to dynamic misspecification. In this vein, Jorda (2005) proposes to use local projection methods to estimate IRs.

Our first objective, against this background, is to establish the importance of approximation bias for VAR-based IRs, using a plausible micro-founded model to generate artificial data. We choose the estimated model of Smets and Wouters (2003), because of its superior properties as a data description for the Euro-area macroeconomy. At the same time, the model belongs to the class of New Keynesian DSGE models, which feature prominently in the current macroeconomic literature. This ensures that our analysis keeps a close link with prominent recent applications estimating and matching IRs such as Christiano, Eichenbaum and Evans (2005). The Smets-Wouters model describes a rich dynamic structure that is not nested in a standard low-order VAR in a few observable variables. Hence, we can fruitfully examine the accuracy of the VAR approximation and its impact on IR estimates, depending on sample size, lag length and the number of variables in the VAR.

The second, and principal, goal of our paper is to confront the VAR-based results with the alternative local projection approach proposed by Jorda (2005). Accordingly, we compare the performance of both approaches in terms of producing accurate IR estimates for a wide range of specifications.

Three main results stand out from our analysis. First, across the various shocks we consider, it turns out that estimated IRs generally have the right dynamic shape, conditional upon the (accurate) contemporaneous identification scheme that we impose. However, the estimated curves deviate quite clearly from the true IRs in that median estimates tend to be biased towards zero. This observation is true for realistic sample sizes of 80 and 160 quarters, although the bias diminishes somewhat for the larger sample. There are certain quantitative differences across the structural shocks we consider.

Specifically, the bias appears more severe for monetary policy and labor supply shocks than for a technology shock. Generally though, this first set of results suggests treating common IR estimates with some caution.

Second, however, the observed bias vanishes almost completely when we consider a hypothetical large sample of 5,000 observations. Although this may provide little comfort for empirical researchers, who generally have to do with samples of much smaller size, it is an interesting insight nevertheless. Specifically, it highlights the nature of the bias as a small-sample issue, quite at odds with the view that the omission of state variables from the VAR fundamentally precludes a good approximation of the true economic model. Even a low-order VAR seems, in principle, able to capture very well the dynamics of our fairly complex data-generating process.

Third, we do not find evidence that local projection methods perform any better than standard VAR methods. If anything, our results indicate that they suffer a somewhat greater bias and variance, thus failing to show their supposedly greater robustness to dynamic misspecification. It should be noted, however, that our exercise has not touched upon other seeming advantages of local projection methods, notably their flexibility to accommodate nonlinear dynamics.

The main conclusion to be drawn for practical work is that complex data-generating processes are likely to cause dynamics that are poorly approximated by standard methods at least in small samples. To the extent that small-sample performance is the relevant criterion, researchers should thus be careful not to trust estimated IRs too blindly. Especially in applications of minimum distance estimation, it might seem desirable for internal consistency to work with models whose state-space solution is actually nested in an appropriately specified VAR. An example of this idea can be found in Bilbiie, Meier and Mueller (2005).

References

[1] Altig, David, Lawrence J. Christiano, Martin S. Eichenbaum, and Jesper Lindé (2005): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle." NBER Working Paper 11034.

[2] Angeloni, Ignazio, Guenter Coenen, and Frank Smets (2003): "Persistence, the Transmission Mechanism and Robust Monetary Policy." Scottish Journal of Political Economy 50, 527-549.

[3] Bhansali, Rajendra J. (1996): "Asymptotically Efficient Autoregressive Model Selection for Multistep Prediction." Annals of the Institute of Statistical Mathematics 48, 577-602.

[4] Bilbiie, Florin O., Andre Meier, and Gernot Mueller (2005): "What Accounts for the Change in U.S. Fiscal Policy Transmission?" Mimeo.

[5] Blanchard, Olivier J. and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances." American Economic Review 79, 655-673.

[6] Braun, Phillip A. and Stefan Mittnik (1993): "Misspecifications in Vector Autoregressions and Their Effects on Impulse Responses and Variance Decompositions." Journal of Econometrics 59, 319-341.

[7] Calvo, Guillermo A. (1983): "Staggered Prices in a Utility-Maximizing Framework." Journal of Monetary Economics 13, 383-398.

[8] Chang, Pao-Li and Shinichi Sakata (2004): "A Misspecification-Robust Impulse Response Estimator." Mimeo.

[9] Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan (2005): "A Critique of Structural VARs Using Real Business Cycle Theory." Federal Reserve Bank of Minneapolis Working Paper 631.

[10] Christiano, Lawrence J. (2004): "Discussion of 'Are Structural VARs Useful Guides for Developing Business Cycle Theories' by Chari, Kehoe, McGrattan". Slides presented at '3rd Banco de Portugal Conference on Monetary Policy', Lisbon, June 2004.

[11] 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, 1-45.

[12] Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson (2005): "Assessing Structural VARs." Mimeo dated May 24, 2005.

[13] Cooley, Thomas F. and Mark Dwyer (1998): "Business Cycle Analysis without much Theory: A Look at Structural VARs." Journal of Econometrics 83, 57-88.

[14] Del Negro, Marco, Frank Schorfheide, Frank Smets, and Raf Wouters (2004): "On the Fit and Forecasting Performance of New-Keynesian Models." Federal Reserve Bank of Atlanta Working Paper No. 37.

[15] Erceg, Christopher J., Luca Guerrieri, and Christopher Gust (2005): "Can Long-Run Restrictions Identify Technology Shocks?" Board of Governors International Finance Discussion Paper No. 792 (updated March 2005).

[16] Faust, Jon and Eric Leeper (1997): "When Do Long-Run Identifying Restrictions Give Reliable Results?" Journal of Business and Economic Statistics 15, 345-353.

[17] Fernandez-Villaverde, Jesus, Juan Rubio-Ramirez, and Thomas J. Sargent (2005): "The A's, B's, C's (and D)'s for Understanding VARs." NBER Technical Working Paper No. 308.

[18] Gali, Jordi (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" American Economic Review 89, 249–271.

[19] Hansen, Lars P. and Thomas J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models." Journal of Economic Dynamics and Control 2, 1-46.

[20] Hansen, Lars P. and Thomas J. Sargent (2005): "Recursive Models of Dynamic Linear Economies." Mimeo.

[21] Ing, Ching-Kang (2003): "Multistep Prediction in Autoregressive Processes." Econometric Theory 19, 254-279.

[22] Jorda, Oscar (2005): "Estimation and Inference of Impulse Responses by Local Projections." American Economic Review 95, 161-182.

[23] Kapetanios, George, Adrian Pagan, and Alasdair Scott (2005): "Making a Match: Combining Theory and Evidence in Policy-oriented Macroeconomic Modeling." Mimeo.

[24] Lippi, Marco and Lucrezia Reichlin (1993): "The Dynamic Effects of Aggregate Supply and Demand Disturbances: Comment." American Economic Review 83, 644–652.

[25] Lippi, Marco and Lucrezia Reichlin (1994): "VAR Analysis, Nonfundamental Representations, Blaschke Matrices." Journal of Econometrics 63, 307-325.

[26] Pagan, Adrian and John Robertson (1998): "Structural Models of the Liquidity Effect." Review of Economics and Statistics 80, 202-217.

[27] Rotemberg, Julio R. and Michael Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In NBER Macroeconomics Annual, pp. 297-346. Cambridge, MA.: MIT Press.

[28] Sarte, Pierre-Daniel G. (1997): "On the Identification of Structural Vector Autoregressions." Federal Reserve Bank of Richmond Economic Quarterly 83, 45-67.

[29] Shaman, Paul and Robert A. Stine (1988): "The Bias of Autoregressive Coefficient Estimators." Journal of the American Statistical Association 83, 842-848.

[30] Shapiro, Matthew and Mark Watson (1988): "Sources of Business Cycle Fluctuations." In: NBER Macroeconomics Annual, pp. 111-156. Cambridge, MA.: MIT Press.

[31] Sims, Christopher A. (1980): "Macroeconomics and Reality." Econometrica 48, 1-48.

[32] Smets, Frank and Raf Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." Journal of the European Economic Association 1, 1123-1175.

[33] Wallis, Kenneth F. (1977): "Multiple Time Series and the Final Form of Econometric Models." Econometrica 45, 1481-1497.

[34] Zellner, Arnold and Franz Palm (1974): "Time Series Analysis and Simultaneous Equation Econometric Models." Journal of Econometrics 2, 17-54.

Appendix: Tables and Figures

Table 1: Parameterization of the Smets-Wouters Euro area model

Parameter	Meaning	Value
α	capital share	0.3000
δ	depreciation rate	0.0250
λ^w	elasticity of labor demand	0.5000
β	discount rate	0.9900
C/Y	consumption share	0.6000
I/Y	investment share	0.2200
φ	inverse elast. of capital adjustment cost fct.	0.1477
σ	consumption utility parameter	1.3530
γ	consumption habit parameter	0.5730
ν	inverse of labor supply elasticity	2.4000
ϕ	one plus share of fixed cost in production	1.4080
ψ	inverse elast. of utilization cost fct.	5.9172
θ^w	Calvo parameter wages	0.7370
θ^p	Calvo parameter prices	0.9080
κ^w	wage indexation	0.7630
κ^p	price indexation	0.4690
ρ	interest rate smoothing	0.9610
r_π	monetary policy rule parameter - inflation	1.6840
r_y	monetary policy rule parameter - output	0.0990
$r_{\Delta\pi}$	monetary policy rule parameter - Δ inflation	0.1400
$r_{\Delta y}$	monetary policy rule parameter - Δ output	0.1590
	AR parameter productivity shock	0.8230
	AR parameter consumer preference shock	0.8550
	AR parameter government spending shock	0.9490
	AR parameter labor supply shock	0.8890
	AR parameter investment shock	0.9270
	Standard deviation productivity shock	0.5980
	Standard deviation consumer preference shock	0.3360
	Standard deviation government spending shock	0.3250
	Standard deviation labor supply shock	3.5200
	Standard deviation investment shock	0.0850
	Standard deviation interest rate shock	0.0810
	Standard deviation price markup shock	0.1600

Table 2: Weighted squared deviations of median estimates from true impulse responses

Shock	Number of observations	Specification: 7 variables, 4 lags		7 variables, 8 lags		3 variables, 4 lags		3 variables, 8 lags	
		VAR	LP	VAR	LP	VAR	LP	VAR	LP
a) Monetary policy	80	257.24	413.22	----	----	22.82	38.26	23.07	49.12
	160	112.20	186.89	140.02	232.88	10.09	11.89	9.70	13.64
	5,000	3.02	3.39	2.94	3.22	10.67	0.68	3.89	0.60
b) Technology	80	72.76	124.40	----	----	2.79	14.39	4.76	16.10
	160	39.92	62.68	44.50	68.45	1.26	5.77	2.08	4.96
	5,000	1.01	0.97	0.90	0.85	201.80	199.61	127.02	130.01
c) Labor Supply	80	207.40	336.18	----	----	25.03	38.19	20.82	40.24
	160	99.31	159.63	118.06	194.39	18.91	14.16	12.40	13.92
	5,000	4.50	4.71	4.16	4.23	153.67	118.38	85.99	80.54

Note: The reported numbers represent the weighted squared distance between true impulse responses and the respective median estimates (over 1,000 random replications). Weights are constructed from the inverse of the point-wise variance of impulse response estimates as obtained from the respective VAR. Accordingly, VAR and LP numbers can be compared within each specification to judge the relative median bias of impulse response estimates.

Table 3: Average weighted squared deviations of estimates from true impulse responses

Shock	Number of observations	Specification: 7 variables, 4 lags		7 variables, 8 lags		3 variables, 4 lags		3 variables, 8 lags	
		VAR	LP	VAR	LP	VAR	LP	VAR	LP
a) Monetary policy	80	347.41	557.94	----	----	65.82	111.56	65.87	112.19
	160	213.56	323.82	239.73	367.21	55.18	76.93	54.43	72.52
	5,000	107.41	120.85	107.49	113.75	55.34	59.41	48.84	51.37
b) Technology	80	175.11	293.55	----	----	47.29	112.87	49.30	94.68
	160	143.86	217.23	148.77	213.64	46.00	95.49	46.85	70.23
	5,000	105.91	122.19	105.86	111.58	244.84	273.96	170.37	178.76
c) Labor Supply	80	305.30	487.28	----	----	68.87	121.86	64.57	115.38
	160	200.56	302.32	218.52	332.98	63.65	87.62	56.57	73.19
	5,000	109.07	124.40	108.86	114.86	199.25	183.41	130.26	131.20

Note: The reported numbers represent averages, taken over 1,000 random replications respectively, of the weighted squared distance between true and estimated impulse responses, where weights are constructed from the inverse of the point-wise variance of impulse response estimates as obtained from the respective VAR. Accordingly, VAR and LP numbers can be compared within each specification to judge the relative performance of impulse response estimates, taking bias and variance into account.

Table 4: Weighted squared deviations of median estimates from true impulse responses (lag length determined by information criteria)

Shock	Number of observations	7 variables, AIC		7 variables, SIC		3 variables, AIC		3 variables, SIC	
		VAR	LP	VAR	LP	VAR	LP	VAR	LP
a) Monetary policy	80	260.95	399.19	200.97	343.00	24.37	31.30	33.15	65.94
	160	104.37	175.89	86.44	144.63	12.94	6.84	13.85	6.89
b) Technology	80	75.55	123.63	61.65	116.89	2.03	12.08	4.98	15.34
	160	37.70	62.65	33.40	54.77	2.71	9.09	3.57	9.62
c) Labor Supply	80	216.35	332.01	192.73	324.12	33.36	43.41	63.54	92.71
	160	93.63	151.56	81.72	119.99	28.78	14.86	34.18	19.30

Note: The reported numbers represent the weighted squared distance between true impulse responses and the respective median estimates (over 1,000 random replications). For each replication, the lag length for the VAR/LP estimation is determined by either of two criteria: the Akaike Information Criterion (AIC) or the Schwartz Information Criterion (SIC).
Max. lag length considered for 80/160 observations: 6/8. Weights are constructed from the inverse of the point-wise variance of impulse response estimates as obtained from the respective VAR. Accordingly, VAR and LP numbers can be compared within each specification to judge the relative median bias of impulse response estimates.
Descriptive statistics regarding lag length suggested by different criteria:

Lag length sugested	Number of observations	7 variables, AIC	7 variables, SIC	3 variables, AIC	3 variables, SIC
Mean	80	5.62	1.87	2.22	1.41
Median		6.00	2.00	2.00	1.00
Minimum		2.00	1.00	1.00	1.00
Maximum		6.00	2.00	6.00	2.00
Mean	160	3.13	2.03	2.24	1.94
Median		3.00	2.00	2.00	2.00
Minimum		2.00	2.00	2.00	1.00
Maximum		8.00	3.00	7.00	2.00

Table 5: Average weighted squared deviations of estimates from true impulse responses (lag length determined by information criteria)

Shock	Number of observations	Specification: 7 variables, AIC		7 variables, SIC		3 variables, AIC		3 variables, SIC	
		VAR	LP	VAR	LP	VAR	LP	VAR	LP
a) Monetary policy	80	351.58	557.18	296.80	495.49	67.22	125.98	74.37	157.65
	160	207.16	317.83	190.01	304.20	58.21	95.20	59.46	103.45
b) Technology	80	178.94	288.52	163.85	318.70	46.48	136.18	48.68	138.89
	160	141.36	220.42	137.11	220.36	47.46	126.21	48.44	137.26
c) Labor Supply	80	313.47	493.66	289.21	491.69	76.68	150.72	100.37	172.31
	160	194.94	294.91	182.74	301.28	73.43	118.57	78.87	132.68

Note: The reported numbers represent averages, taken over 1,000 random replications respectively, of the weighted squared distance between true and estimated impulse responses. For each replication, the lag length for the VAR/LP estimation is determined by either of two criteria: the Akaike Information Criterion (AIC) or the Schwartz Information Criterion (SIC).
Max. lag length considered for 80/160 observations: 6/8. Weights are constructed from the inverse of the point-wise variance of impulse response estimates as obtained from the respective VAR. Accordingly, VAR and LP numbers can be compared within each specification to judge the relative performance of impulse response estimates.
Descriptive statistics regarding lag length suggested by different criteria:

Lag length sugested	Number of observations	Specification: 7 variables, AIC	7 variables, SIC	3 variables, AIC	3 variables, SIC
Mean	80	5.62	1.87	2.22	1.41
Median		6.00	2.00	2.00	1.00
Minimum		2.00	1.00	1.00	1.00
Maximum		6.00	2.00	6.00	2.00
Mean	160	3.13	2.03	2.24	1.94
Median		3.00	2.00	2.00	2.00
Minimum		2.00	2.00	2.00	1.00
Maximum		8.00	3.00	7.00	2.00

Figure 1: Monetary Policy Shock - True impulse responses vs. median estimates from 7-variable VAR(4) and LP(4)

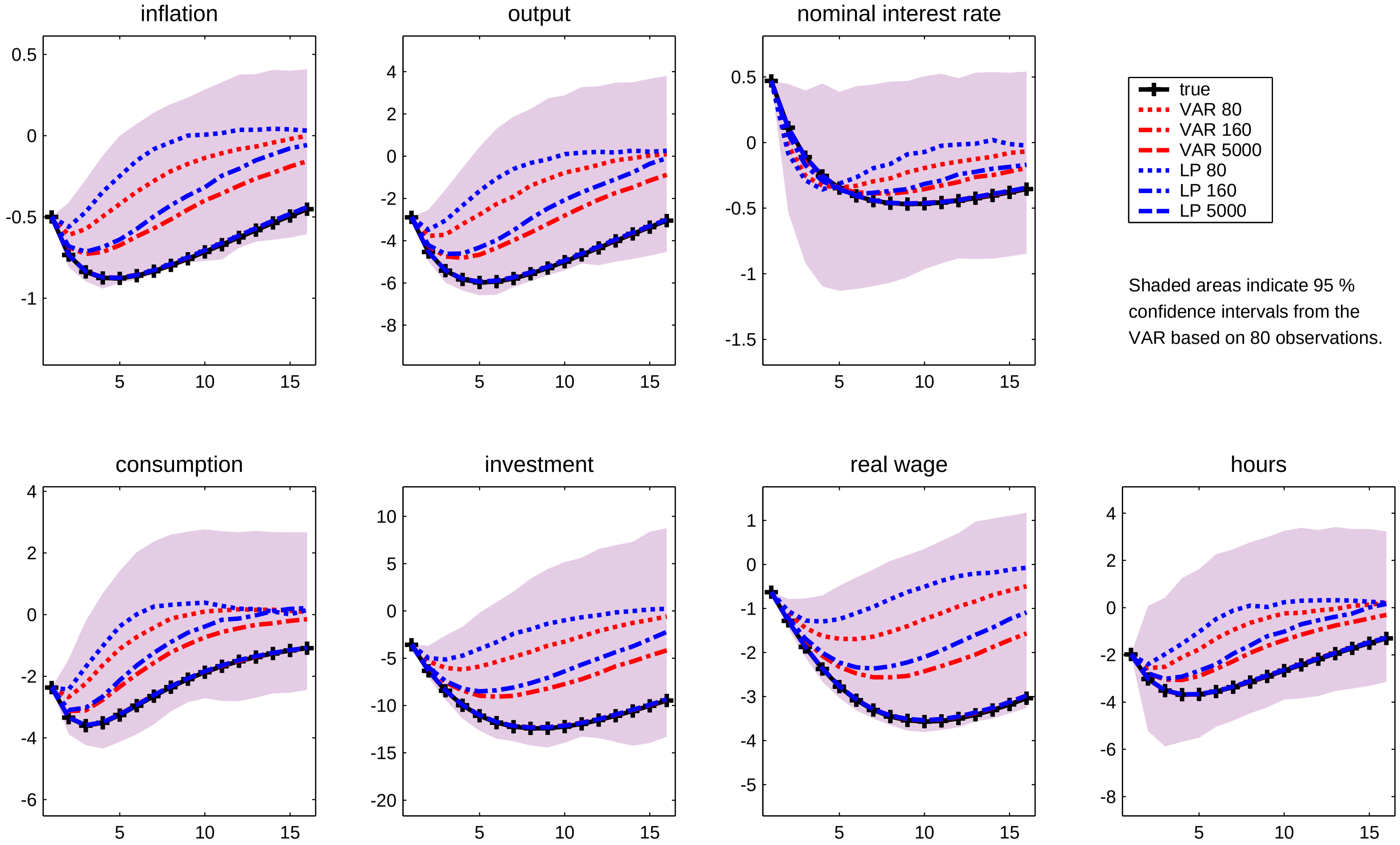
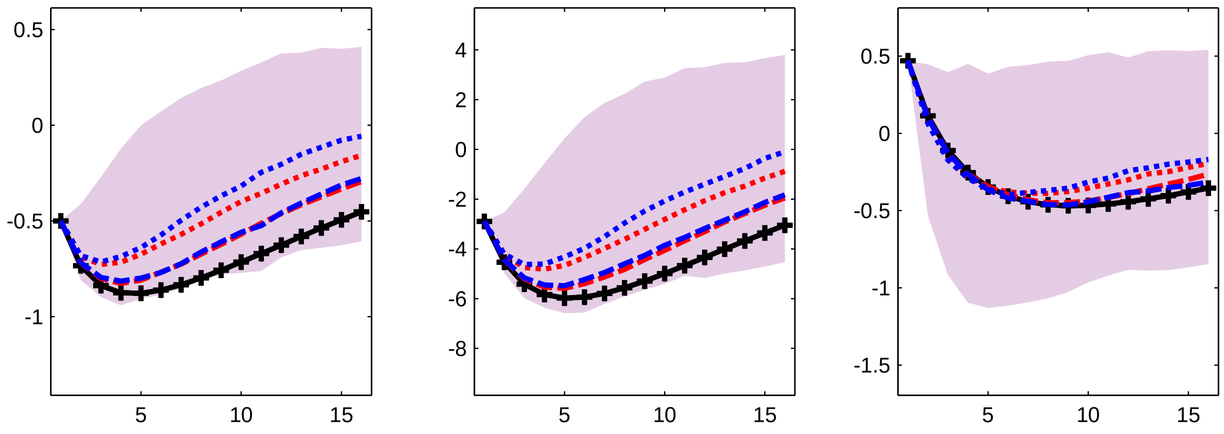
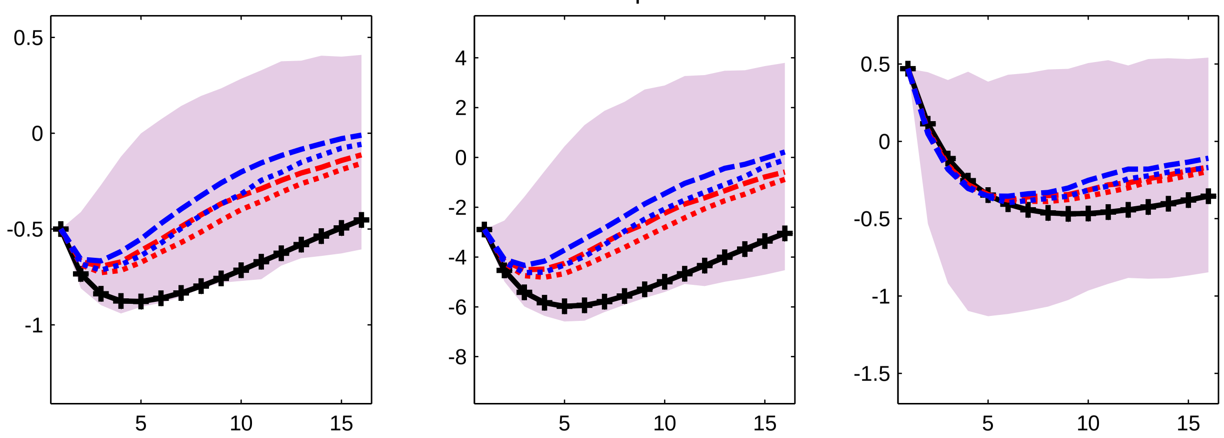


Figure 2: Monetary Policy Shock - Baseline estimates vs. alternative specifications

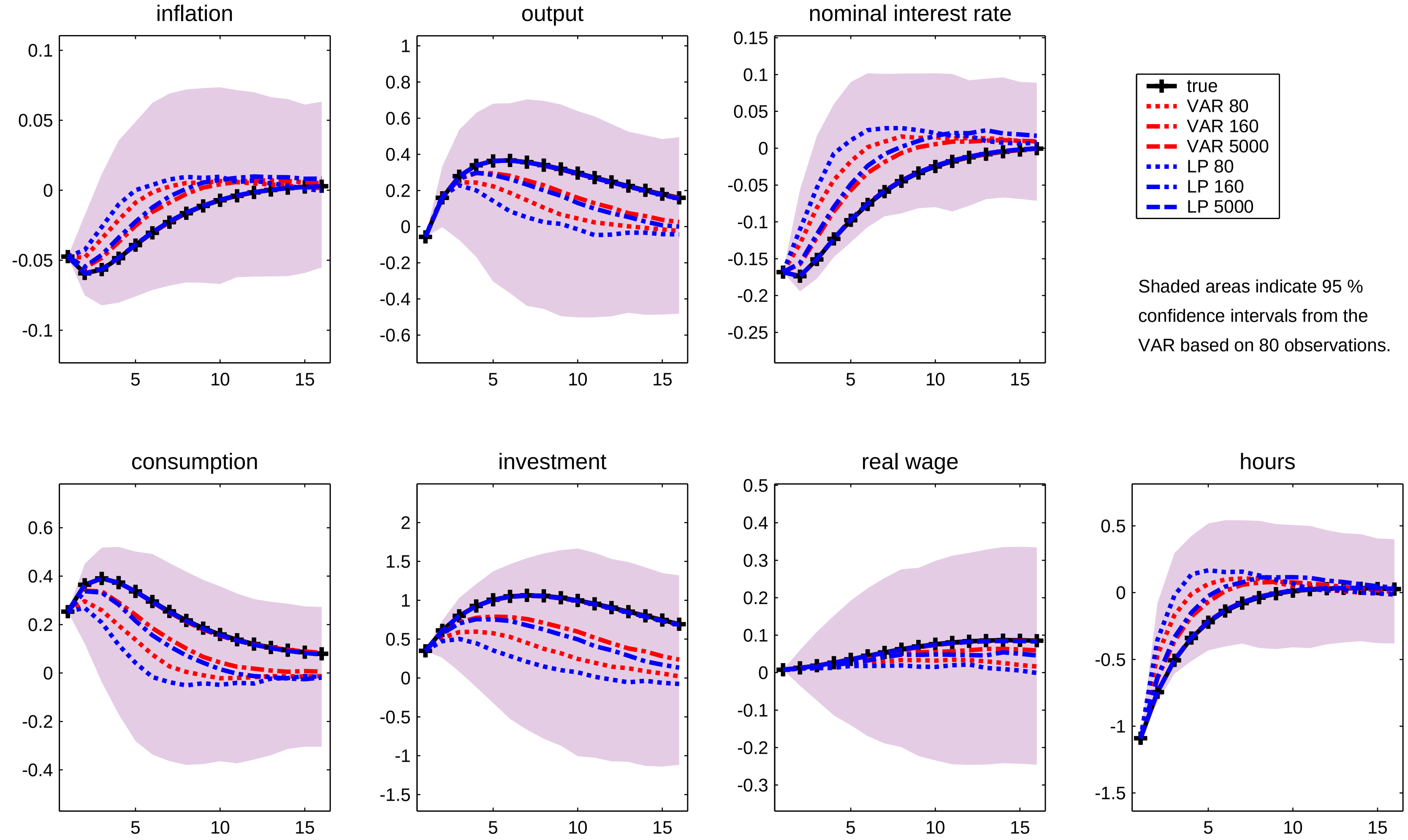


a) Variation in the no. of variables:
Baseline refers to VAR/LP in 7 variables as in figure 1, while "small" denotes VAR/LP in only 3 variables: inflation, output, interest rate.
Specification with 4 lags throughout.



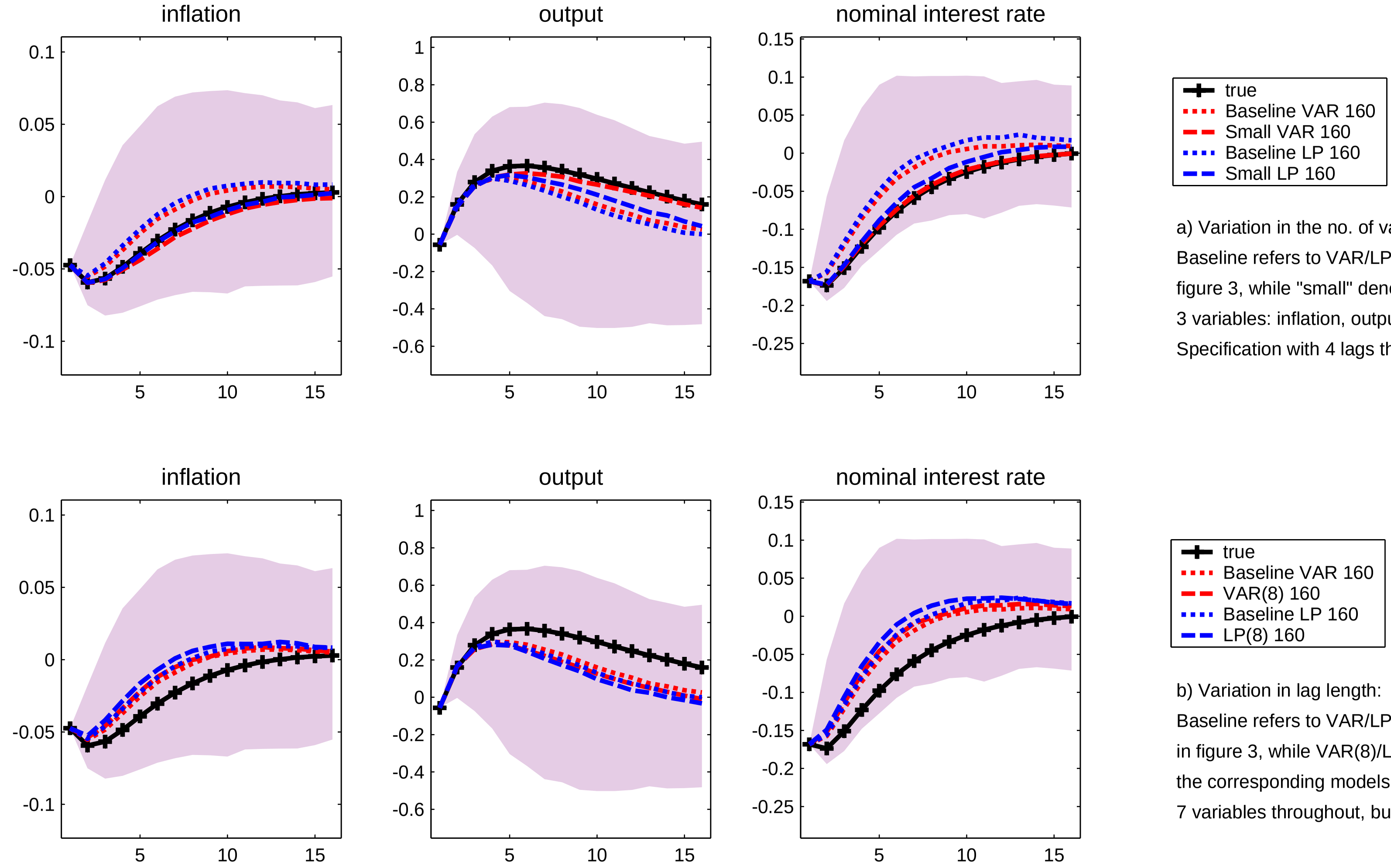
b) Variation in lag length:
Baseline refers to VAR/LP with 4 lags as in figure 1, while VAR(8)/LP(8) denote the corresponding models with 8 lags.
7 variables throughout, but only 3 depicted.

Figure 3: Technology Shock - True impulse responses vs. median estimates from 7-variable VAR(4) and LP(4)



Shaded areas indicate 95 % confidence intervals from the VAR based on 80 observations.

Figure 4: Technology Shock - Baseline estimates vs. alternative specifications



a) Variation in the no. of variables:
Baseline refers to VAR/LP in 7 variables as in figure 3, while "small" denotes VAR/LP in only 3 variables: inflation, output, interest rate. Specification with 4 lags throughout.

b) Variation in lag length:
Baseline refers to VAR/LP with 4 lags as in figure 3, while VAR(8)/LP(8) denote the corresponding models with 8 lags. 7 variables throughout, but only 3 depicted.

Figure 5: Labor Supply Shock - True impulse responses vs. median estimates from 7-variable VAR(4) and LP(4)

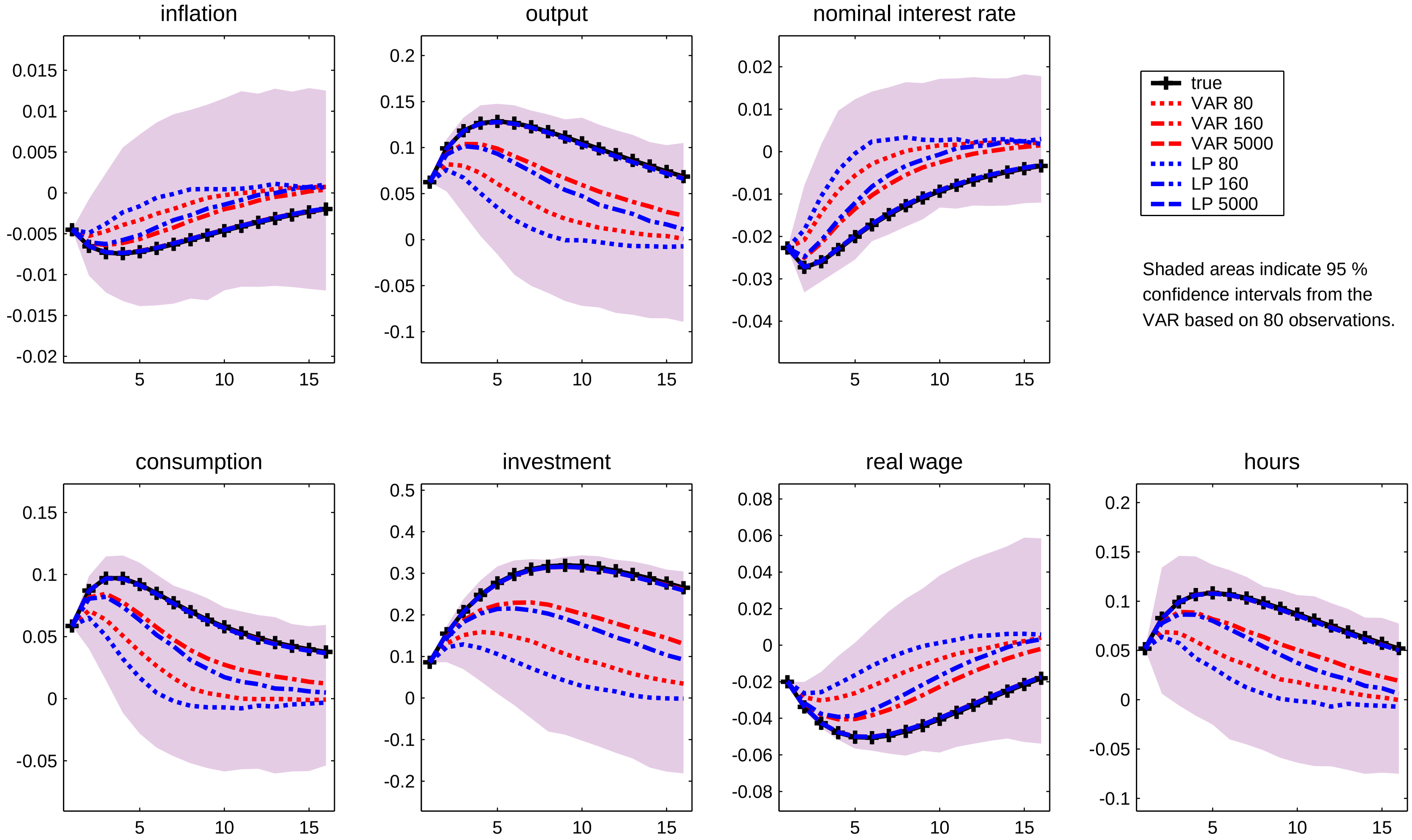
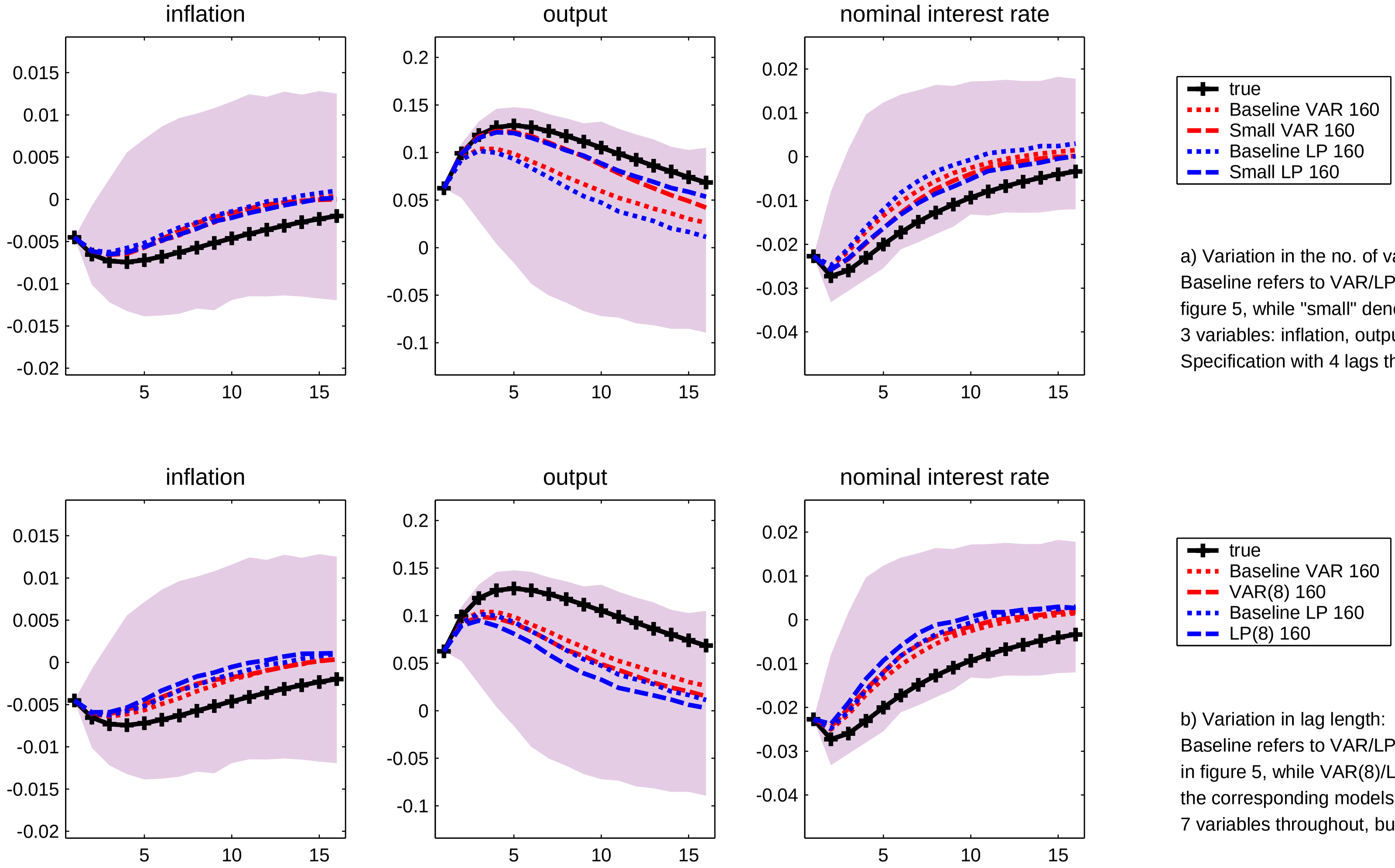


Figure 6: Labor Supply Shock - Baseline estimates vs. alternative specifications



a) Variation in the no. of variables:
Baseline refers to VAR/LP in 7 variables as in figure 5, while "small" denotes VAR/LP in only 3 variables: inflation, output, interest rate. Specification with 4 lags throughout.

b) Variation in lag length:
Baseline refers to VAR/LP with 4 lags as in figure 5, while VAR(8)/LP(8) denote the corresponding models with 8 lags. 7 variables throughout, but only 3 depicted.