Referee Report

Quantitative Economics

MS-2406

"Indirect Inference: a Local Projection approach"

by Castellanos & Cooper

This paper proposes an indirect inference approach to estimate DSGE model parameters based on local projection impulse response function estimates.

The paper contributes to a long literature that estimates DSGE models via impulse-response matching (e.g., Rotemberg and Woodford, 1997; Christiano et al., 2005). Differently from those papers, the authors propose to recover the impulse response function (IRF) targets through a local projection (LP) rather than a VAR. A key finding of the paper is that, with the structural model of Smets and Wouters (2007) treated as an artificial data-generating laboratory, model estimation based on LP IRF targets appears to perform better than the conventional VAR-based approach. In the second part of the paper, the authors then use empirically estimated policy and non-policy shock IRFs—for their applications also recovered via LPs—to re-estimate the parameters of the Smets and Wouters model.

Main Points

DSGE model estimation via impulse-response matching is popular, and researchers usually—though not exclusively (e.g., see Auclert et al., 2020)—use VARs as the auxiliary model. Recent work has aimed to compare LPs and VARs as tools for IRF estimation (Li et al., 2022), so I think it's a useful contribution to also compare their performance as inputs for model estimation. In my view this is the core contribution of the paper, and my comments 1. - 2. are supposed to help further sharpen that part.

My third comment is instead about the second part of the paper—on re-estimating the model of Smets and Wouters with LP-estimated IRFs as estimation targets. In my view at least this part of the analysis detracts from the paper's core contribution.

1. **LPs vs. VARs as auxiliary models.** In my view, the headline finding of the paper is that DSGE model estimation works better when LPs are used as the auxiliary model. I would encourage the authors to dig deeper into that result, along the following margins.

   - *Lags.* As the number of lags is increased, VARs and LPs should perform similarly as auxiliary models (Plagborg-Møller and Wolf, 2021). This is visible in Table B.6, and I would promote that discussion to the main text. I would furthermore welcome more of a discussion of lag length selection. For example, for $p = 12$, the two methods seem to perform similarly. Why is $p = 4$ a sensible choice as the headline number? In particular, what lag length would standard information criteria select?
   - *Shock identification.* Currently the authors consider an idealized case where the researcher directly observes a measure of the shock. I would like to know how robust the authors' findings are likely to be to the choice of identification scheme. In particular, much of the applied macroeconometrics literature identifies shocks by first assuming invertibility and then imposing additional restrictions (e.g., a recursive ordering, long-run restrictions, or sign restrictions).

     For example, the authors could modify the timing assumptions in the model so that monetary shocks can be identified through a recursive ordering, as in Christiano et al. (2005). They could then implement the recursive identification scheme using either LPs or VARs, repeat the model estimation, and report new results.
   - *VAR dimensionality and SVAR-IV.* The authors consider relatively short-lag, bivariate VARs (p.6). Why not instead include the shock together with *all* outcomes of interest in a single recursive VAR? Such a larger-dimensional system would give richer dynamics and also be closer to actual empirical IRF matching practice (e.g., see Christiano et al., 2005).

     In fact, given the authors' assumed identification scheme, they could even consider a SVAR-IV implementation, where the measured shock is just used to rotate the Wold IRFs from a reduced-form VAR in the macro observables alone (Gertler and Karadi, 2015). I would—given the results in Li et al. (2022)—expect such a procedure to actually perform rather well (in spite of the fact that it is generally going to be inconsistent, given the likely lack of invertibility).

- *Shocks with measurement error.* I don't really follow the discussion in Section 4.4.1. Why should that estimation indicate more price and wage flexibility than the baseline? Measurement error just results in *absolute* IRFs being biased downward, while relative IRFs are still correctly identified (Stock and Watson, 2018). In particular, in population, a shock IV with measurement error suffices to correctly identify the response of output or inflation (say) to a monetary shock that moves nominal rates by a certain percentage amount. Thus, in population, there should be no effect at all on the best-fitting parameter values. What am I missing?
- *Other DGPs.* The structural model of Smets and Wouters (2007) is a very natural laboratory for the kind of exercise that the authors are considering. That being said, I would find it encouraging to know that similar conclusions would also apply in other popular structural environments, preferably those that have previously been used for model estimation through impulse response matching. Two notable examples that come to my mind are the DSGE models in Rotemberg and Woodford (1997) and Christiano et al. (2005). Maybe such exercises could be added to an appendix.

Finally, I would like the authors to relate their findings a bit more to those of Li et al., who investigate the LP-VAR bias-variance trade-off in IRF estimation (rather than model parameter estimation), and conclude that VARs on average perform *better*. I would like to understand those differences a bit more. One simple strategy would be the following: take the model of Smets and Wouters (2007) as the data-generating process, and then for this DGP study the LP vs. VAR trade-off in IRF estimation—i.e., do for Smets and Wouters (2007) what Li et al. do for their dynamic factor model. If the results from this exercise are similar to Li et al., then we know that the differences between the present paper and Li et al. are not due to the underlying DGP, but instead are about IRF estimation vs. model parameter estimation. That would be a useful insight.

2. **Extension to other estimation techniques.** I think the present paper makes a valuable point—structural model estimation via IRF matching is popular, and LPs may well offer advantages over VARs in terms of IRF estimation and thus indirectly also for model parameter estimation. But then why stop at just the VAR vs. LP dichotomy? There are many intermediate estimation techniques, like Bayesian VARs, Bayesian LPs, or penalized LPs. It would be interesting to know how those alternative techniques would perform in terms of the authors' key J^* statistic.

3. **Do we need an empirical application?** In Section 5 the authors use LP-estimated

shock IRFs to re-estimate the Smets and Wouters (2007) model. I find those results interesting, but I am not sure that they actually fit well with the rest of the paper. I would thus actually encourage the authors to remove this part and perhaps elaborate on it in a different paper.

- As I discuss above, an important message of the paper is that structural model estimation through IRF matching may actually perform better when based on LPs (rather than VARs, as mostly done so far). If that conclusion turns out to be robust, then that is a quite useful finding.
- I don't understand, however, why that clean conceptual point needs to be supplemented with an application to model (re-)estimation. The core message of the paper is about LP vs. VARs as different techniques to generate estimation targets—the question of whether standard structural models can then generate such estimated IRFs is also an important one, but clearly a logically separate one.
- If there is a strong preference to keep that last part in the paper, then I would at least like to see it be tied back better to the first part of the paper. What if instead a researcher would estimate the IRFs of Section 5 with a VAR, and then re-estimate the model? If the conclusions are very different, then Section 5 could have the benefit of illustrating that the LP vs. VAR choice has very real practical bite, at least in one particular application.

Minor Issues

4. **The importance of technology shocks.** In Section 5 the authors (among other shocks) focus on technology shocks, arguing that they are "the most important type of non-policy shock". I think that statement needs to be qualified. In terms of history of thought I of course agree, going back all the way to Kydland and Prescott (1982). The recent structural and semi-structural literature, however, has emphasized much more the importance of investment-specific technology shocks (Justiniano et al., 2010; Ramey, 2016).

5. **Typos.**

   - "Start by consider any economic model" (p.5)
   - "Thus, one fruitfully line of research is ..." (p.28)

References

Auclert, A., M. Rognlie, and L. Straub (2020): "Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model," Tech. rep., National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113, 1–45.

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

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2010): "Investment shocks and business cycles," *Journal of Monetary Economics*, 57, 132–145.

Kydland, F. E. and E. C. Prescott (1982): "Time to build and aggregate fluctuations," *Econometrica: Journal of the Econometric Society*, 1345–1370.

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

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

Ramey, V. A. (2016): "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2, 71–162.

Rotemberg, J. J. and M. Woodford (1997): "An optimization-based econometric framework for the evaluation of monetary policy," *NBER macroeconomics annual*, 12, 297–346.

Smets, F. and R. Wouters (2007): "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American economic review*, 97, 586–606.

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