On the Likely Success of a Method for Trying to Recover a Counterfactual Response*

Adrian Pagan† and Ippei Fujiwara‡

January 22, 2025

Contents

1 Introduction **2**

2 A Different Derivation of MW's Result **5**

3 A Simple Baseline Two Variable Structural Model **7**

4 An Extended Baseline Two Variable Structural Model **10**

5 Finding the Impulse Responses **12**
5.1 Finding Responses for the Simple Two Variable Model 13
5.2 Finding Responses for MW's Empirical Example 15

6 Conclusion **18**

7 References **20**

Abstract

Counterfactual analysis is generally performed with a DSGE model incorporating expectations and dynamics. McKay and Wolf (2004)

*We would like to thank Ben Schumann for comments on this paper.
†University of Sydney, adrian.pagan@sydney.edu.au
‡Keio University and the Australian National University, ippei.fujiwara@keio.jp

(MW) describe a method for finding counterfactuals which requires the estimation of the Structural Moving Average (SMA) from that DSGE model and its use to find impulse responses to shocks. A key feature of their methodology is the presence of news shocks in the SMA. They show that with enough news shocks one can recover the counterfactual. Since there are only a few observable series on such shocks they go on to propose the construction of a "robust counterfactual" with this smaller number. Using simple models we show that the robust and actual counterfactuals can be quite different, and it seems premature to claim that one has found the counterfactual, as many of those using the MW methodology do. We also deal with the related issue of estimating the SMA. This is often done by assuming that there are observable proxies for the shocks. When the proxy and the shocks deviate due to measurement errors we observe that one does not recover the counterfactual but some function of it that depends on the importance of the measurement error. Measuring the latter is difficult.

JEL Classifications: C3 E3

Keywords: Counterfactual, News Shcoks, Shock Recovery, Local Projection

1 Introduction

Counterfactual analysis has long been a feature of macroeconomics. Questions such as what would have been the effects on inflation or output during the Depression if monetary, fiscal or wages policy had been different were addressed by using early macroeconometric models. Initially this looked relatively easy, as the variables to be varied were treated as exogenous, and so one formulated a baseline model of the economy and then asked how the solutions for output and inflation would vary under different values for these exogenous variables. Complications emerged when the variables to be modified were actually endogenous. For example, changing government expenditure will mean either a change in the money supply or an interest rate, depending on how it is financed, and then the issue is whether any effects were due to the variation in the latter rather than the former variable. This led to a lot of applied work in institutions such as central banks that involved "fixes" for endogenous variables. It took some time to work out how to do this. Then forward looking expectations started to become important in

modelling the economy and this added a further layer of complication to how one might determine the impact of considering some counterfactual outcome. Solutions emerged and are well summarized by Burgess et. al. (2013).

When the model being used to represent history is a DSGE like model, that is one which incorporates dynamics and expectations, it is relatively easy to perform a counterfactual study.[1] An objection that has been made by McKay and Wolf (2023) to this procedure is that ".. the researcher needs to commit to a particular parametric model, thus introducing concerns about model mis-specification". This is true, but we are trying to compare an historical outcome with a suggested alternative history, and so some model is needed. Suppose for example that a SVAR had been chosen and one wanted to look at a counterfactual in that context. Then there is also no certainty that the SVAR model used to describe history is correctly specified.

Because of this concern McKay and Wolf (2023) describe an approach to constructing counterfactual outcomes which they summarize as "..We show that, in a general family of linearized structural macroeconomic models, knowledge of the empirically estimable causal effects of contemporaneous and news shocks to the prevailing policy rule is sufficient to construct counterfactuals under alternative policy rules"(p 1696). Section 2 provides a different way of seeing how this is done. Essentially MW's methodology finds weights for the required impulse responses from a known baseline model (that incorporates the original policy rule) and these are then used to reproduce the counterfactual.

Section 3 looks at this is the context of a simple two variable DSGE model that is a simplified NK policy model. The experiment is to move from an interest rate rule to a counterfactual that involves a constant rate. This model involves a shock of interest, a policy shock and a news shock. We assume that the shocks and their impulse responses are known from the DSGE model. Then the two shocks that are the policy and news shocks are enough to reproduce the counterfactual. The reason is that the counterfactual impulse horizon for the shock of interest (the period after which responses are zero) is one, and so there are two elements to be determined. This is done through the policy and news shock. If the counterfactual to be computed has responses

[1] We will refer to these models as DSGE in the remainder of the paper but such models may not come from a formal optimizing framework. They may just involve a set of structural equations that are often involved with New Keynesian macro models. The key feature is that they involve expectations, as that is the complicating factor in looking at counterfactual policies.

that are zero after a horizon of m periods then we would need m news shocks. As MW observe this could be a large number.

That consideration led MW to propose a solution termed the "robust counterfactual" (RCF) which tries to get as close as possible to the actual counterfactual using some norm when there are too few news shocks. The obvious query is whether this alternative gives responses that are close enough to the counterfactual as to make it worthwhile adopting in place of the clearly defined method with a structural model featuring expectations and dynamics. Therefore, in Section 3 we look at a simple re-specification of the Phillips curve in the DSGE model mentioned above, so that the horizon of the impulse response of interest is now two, and two news shocks are therefore needed. When the model is augmented by only one news shock there are significant departures of MW's RCF estimate from the actual counterfactual value CF, even in this simple context. It does not seem appropriate to say that the RCF captures the counterfactual. It may point towards what the counterfactual would be like but, as we see in examples later, it may even fail in that regard.

Section 4 moves on to an extended two variable model where the Phillips curve has more inertia in order to investigate whether this failure of the RCF to match the actual counterfactual is a common issue. In this experiment the counterfactual impulses have a long horizon rather than just one or two periods, as in the baseline models of section 3. Consequently, many news shocks are needed to capture the counterfactual. In this context we see again that there are substantial divergences between the RCF and the actual counterfactual solution if one uses only three policy shocks, a common choice made in the literature. To us this suggests that one should be cautious about claiming that the RCF is likely to give you the counterfactual solution, as is becoming increasingly popular e.g. in studies by Georgiadis et. al (2024), Ider et al. (2024) and Castelnuovo et al. (2024).

Now in the sections above it was assumed that the DSGE model generating the data and all the impulse responses are known. In practice one may need to compute the impulse responses to the policy and non-policy shocks from data without such knowledge. The DGP of the data described by the DSGE model can be represented by a Structural Moving Average (SMA). Many of the studies looking at constructing a counterfactual assume that the SMA is invertible i.e. can be expressed as a SVAR. A problem with that assumption arises from the fact that, once news shocks are in the DGP process, one will generally have more shocks than observables, and so the SMA is not invertible. In the studies mentioned above (and in MW)

extra information is assumed to be available in the form of proxy variables that capture the shocks of interest, as well as policy and news shocks. This assumption raises the number of observables. To get invertibility however one must not add on extra shocks along with the proxy variables. Often the constructed proxy is said to be an imperfect measure of the shock, so measurement errors are likely, and that fact creates extra shocks, and so a lack of invertibility again.

In this instance the SMA must be estimated utilizing the proxies. Currently this is done by treating the proxy as the shock and then fitting a set of equations i.e. the proxy is treated as a regressor. Because there will be a correlation between the proxies and the error terms of those equations, there is a large sample bias in estimates of impulse responses. In turn this means that applying MW's method gives only a function of the counterfactual, and the extent to which the counterfactual cannot be recovered depends on the importance of the measurement error. Section 5 details all of this, first with the simple two variable model and then with an application used by MW. How one is to determine the importance of the measurement error is a critical issue. Section 6 concludes.

2 A Different Derivation of MW's Result

We will work with a simple two variable system that essentially consists of inflation π_t and an interest rate i_t. The DGP will be the NK model with two shocks - a monetary policy shock v_t and some other shock ε_t. The latter may be a productivity shock. It has the form

$$\begin{aligned} \pi_t &= \beta E_t(\pi_{t+1}) - i_t + \varepsilon_t + \varepsilon_{t-1} && (1) \\ i_t &= \delta\pi_t + v_t + n_{t-1}, && (2) \end{aligned}$$

where n_t is a news shock announced at time $t-1$ about the interest rate at time t. It is assumed that the parameters of this DSGE model are known and a counterfactual of $i_t = 0$ is to be imposed, following which the impulse response of π_t to ε_t are to be found under this counterfactual policy rule.

If we know the DSGE model above this is a simple task. The Structural Moving Average (SMA) associated with it will be

$$\pi_t = A^*(L)v_t + A^{**}(L)n_t + B(L)\varepsilon_t \quad (3)$$
$$i_t = C^*(L)v_t + C^{**}(L)n_t + D(L)\varepsilon_t. \quad (4)$$

where the impulse responses of variables to shocks are captured by the polynomials in the lag operator L - so $A^*(L)$ shows the responses of inflation to ε_t etc. The coefficients attached to the L^j are the $j'th$ period impulse responses, so we know the latter once we have $A^*(L)$ etc. Our objective is to then recover the impulse responses of π_t to ε_t under the new policy rule that $i_t = 0$ i.e. we want to find the counterfactual responses[2]. This can of course be done by putting $i_t = 0$ into $(1) - (2)$ and solving for the resulting equation for π_t.

Assume that the impulse responses of π_t to ε_t under the counterfactual is $CF(L)$. MW proceeded by finding weights s_1 and s_2 for the baseline responses $A^*(L)$ and $A^{**}(L)$ so that the estimated counterfactual response of π_t to ε_t was

$$\widehat{CF(L)} = B(L) + A^*(L)s_1 + A^{**}(L)s_2. \quad (5)$$

They then described how to get these weights s_1 and s_2 by solving the equation that comes from setting $i_t = 0$ in (4)

$$0 = D(L) + C^*(L)s_1 + C^{**}(L)s_2. \quad (6)$$

Defining d as a vector containing the L^j terms in $D(L)$, c^* and c^{**} doing the same for $C^*(L)$ and $C^{**}(L)$, MW proposed that s_1 and s_2 be found by regressing $-d$ on c^* and c^{**}. If there is a perfect fit then one might expect that the estimated counterfactual will be recovered To measure the fit we look at $\tau = \frac{(C^*(L)\hat{s}_1 + C^{**}(L)\hat{s}_2)^2.}{D(L)^2}$. This is not the same as the R^2 as computing that requires an intercept, and there isn't one in the regression.

To see how this strategy extends to a move to other counterfactuals consider what happens if a change was made from a baseline rule $i_t = \delta\pi_t + v_t + n_{t-1}$ to a counterfactual $i_t = \psi\pi_t$. Here we are adopting the formulation in MW's equation (3). The equivalent of (6) is now found by imposing $i_t = \psi\pi_t$ on (3)-(4) to get

$$C^*(L)v_t + C^{**}(L)n_t + D(L)\varepsilon_t = \psi(A^*(L)v_t + A^{**}(L)n_t + B(L))\varepsilon_t,$$

[2] To begin we take the alternative policy rule as stabilizing the interest rate i_t to a value $\bar{\imath}$ (one of their examples). Nothing depends on this value so we put $\bar{\imath} = 0$, as MW do.

and so we would want s_1 and s_2 to satisfy

$$[C^*(L) - \psi A^*(L)]s_1 + [C^{**}(L) - \psi A^{**}(L)]s_2 = \psi B(L) - D(L). \quad (7)$$

Then, just as for the target interest rate case, a regression would be run, now of $(\psi b - d)$ against $(c^* - \psi a^*)$ and $(c^{**} - \psi a^{**})$ to get s_1 and s_2. In all cases of changing rules there will be some regression to get s_1 and s_2, but with different regressors, depending on the counterfactual.

3 A Simple Baseline Two Variable Structural Model

Now assign some parameter values to the simple two variable model above. These will be $\beta = .99, \delta = 2$. This structure is meant to represent a simplified model that follows MW's equations (1)-(3). The SMA solution for inflation can be found analytically as follows

$$\begin{aligned}
\pi_t &= \beta E_t(\pi_{t+1}) - \delta\pi_t - v_t + \varepsilon_t + \varepsilon_{t-1} - n_{t-1} \\
(1+\delta)\pi_t &= \beta E_t(\pi_{t+1}) + \varepsilon_t + \varepsilon_{t-1} - v_t - n_{t-1} \\
\pi_t &= \phi E_t(\pi_{t+1}) + a(\varepsilon_t + \varepsilon_{t-1} - v_t - n_{t-1}) \\
\pi_t &= a(1+\phi)\varepsilon_t + a\varepsilon_{t-1} - av_t - an_{t-1} - a\phi n_t \\
&= A^*(L)v_t + A^{**}(L)n_t + B(L)\varepsilon_t,
\end{aligned}$$

where $\phi = \frac{\beta}{1+\delta}, a = \frac{1}{1+\delta}$. The interest rate equation is (2)

$$\begin{aligned}
i_t &= \delta\pi_t + v_t + n_{t-1} && (8) \\
&= \delta B(L)\varepsilon_t + (\delta A^*(L) + 1)v_t + (\delta A^{**}(L) + L)n_t && (9) \\
&= D(L)\varepsilon_t + C^*(L)v_t + C^{**}(L)n_t. && (10)
\end{aligned}$$

Thus we have found an SMA like in (3)-(4). A key factor in our analysis of this simple model is that the baseline model and its interest rate rule is captured with the trio of responses $\{A^*(L), A^{**}(L), B(L)\}$. Only a triple is needed for this case as $D(L), C^*(L), C^{**}(L)$ are constructed from the triple.

One wants to use the triple to recover the impulse responses when the interest rate rule is replaced with some counterfactual. Here, the latter is the

interest rate target $i_t = 0$. This counterfactual sets $i_t = 0$ in (2)

$$\pi_t = \beta E_t(\pi_{t+1}) + \varepsilon_t + \varepsilon_{t-1}$$

and will be

$$\pi_t = (1+\beta)\varepsilon_t + \varepsilon_{t-1} = CF(L)\varepsilon_t.$$

We then need to use the information in $D(L)$ etc. to see if we can get $CF(L) = 1 + \beta + L = 1.99 + L$

To get s_1 and s_2 we solve

$$C^*(L)s_1 + C^{**}(L)s_2 = -D(L), \tag{11}$$

That is done by matching powers of L.[3]Hence we first compute

$$B(L) = 0.443 + 0.33L, \;\; A^*(L) = -.333, A^{**}(L) = -0.11 - .333L, \tag{12}$$

while $C^*(L), C^{**}(L)$ and $D(L)$ are found from these using

$$D(L) = \delta B(L), C^*(L) = \delta A^*(L) + 1, C^{**}(L) = \delta A^{**}(L) + L. \tag{13}$$

Matching of powers of L in (11) gives $s_1 = -3.98, s_2 = -2.0$ and then

$$\begin{aligned} \widehat{CF(L)} &= B(L) - 3.98A^*(L) - 2A^{**}(L) \\ &= 1.99 + L, \end{aligned}$$

which replicates the counterfactual.

It is worth looking at this example in more detail. d is a 2×1 vector as there are only two non-zero elements in $D(L)$. Consequently, putting

$$d = -Hs \tag{14}$$

where $H = [\; c^* \quad c^{**} \;]$ and $s' = [\; s_1 \quad s_2 \;]$ gives a solution for s as $-H^{-1}d$ i.e. the estimates of s come from the regression of $-d$ on c^* and c^{**}, and there is a perfect fit. This regression will be referred to as the "weights regression".

[3]One can do this analytically but it is easier to recover the responses from Dynare output.

Does one need to have news shocks in the interest rate rule? Suppose that $i_t = \delta\pi_t + v_t + v_{t-1}$. Then $\bar{C}^*(L) = c_0 + c_1 L$ and we would be regressing $-d$ on $\bar{c}^*$ to get s_1. Because both d and $\bar{c}^*$ are now 2×1 vectors it would be rare that there is a unique s_1 that satisfies $-d = \bar{c}^* s_1$. One needs two *different shocks* in the regression. *As MW say news shocks are the key to their methodology and they must be present in the model producing the data.*

Suppose now that the order of $D(L)$ is m. Then d has $m+1$ non-zero elements and it would be necessary to have $m+1$ starred $c's$ i.e. $m+1$ policy shocks, of which m will be news shocks, if the regression was to perfectly predict d. If there were less than $m+1$ shocks one cannot find all the necessary weights, and this led to MW's proposal that one get the weights by a "robust" procedure such as least squares i.e. they are chosen to minimize the squared distance between d and the predictions of it from what the regression with c^* and c^{**} would give - this minimizes the Euclidean distance τ of $D(L)$ from the regression prediction using the C^* and C^{**} elements.

To look at this consider a modification of our simple baseline model which adds an extra lag to the shock of interest ε_t to (1) but uses the same counterfactual policy rule of a constant interest rate. This produces

$$\pi_t = \beta E_t(\pi_{t+1}) - i_t + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}.$$

Then

$$CF(L) = 1 + \beta + \beta^2 + (1+\beta)L + L^2 = 2.97 + 1.99L + L^2.$$

If two news shocks are used this counterfactual is replicated. If only one is used the "robust counterfactual" is

$$\widehat{RCF}(L) = 2.3 + 1.33L + .33L^2,$$

so there is quite a difference. It is worth noting the τ from the regression of -$D(L)$ on $C^*(L)$ and $C^{**}(L)$ is .79 and is comparable with MW's value of .78 used in their empirical work. It seems likely that how close τ is to one would be taken to be some sort of signal that one is likely to recover the counterfactual of inflation to ε_t, since it relates to whether one has managed to set $i_t = 0$. Accordingly, a value like .79 might give one confidence that it has been found, so we will need to look more closely at that later.

4 An Extended Baseline Two Variable Structural Model

Suppose we extend the baseline model to something that is a little closer to traditional NK models. This adds in output and allows for inertia in the Phillips curve

$$\begin{aligned} y_t &= \gamma(i_t - \pi_t) && (15)\\ \pi_t &= \beta E_t(\pi_{t+1}) + \alpha\pi_{t-1} + \psi y_t + \varepsilon_t + \varepsilon_{t-1} && (16)\\ i_t &= \delta\pi_t + v_t + \sum_{j=1}^{7} n_{j,t-j} && (17)\\ \beta &= .3, \alpha = .4, \gamma = -.4, \psi = .5, \delta = 2. && (18) \end{aligned}$$

Once again the counterfactual will be $i_t = 0$. This looks like a standard NK model. The counterfactual responses of π_t to ε_t essentially become zero at a horizon of about thirty periods.

Consider trying to recover the counterfactual up to an horizon of seven periods. The correct values of $CF(L)$ are

$$\begin{aligned} CF(L) &= 2.5 + 3.33L + 2.22L^2 + 1.48L^3 + .99L^4 + .66L^5 \\ &\quad +.44L^6 + .29L^7. \end{aligned}$$

Using seven news shocks one would compute the weights $s_1, .., s_8$ by regressing the elements of $-D(L)$ up to the seventh order upon the eight policy shock responses (seven of which are news). The τ from this regression is 1 and the estimated counterfactual is

$$\begin{aligned} \widehat{CF(L)} &= 2.5 + 3.33L + 2.22L^2 + 1.48L^3 + .98L^4 + .65L^5 \\ &\quad +.41L^6 + .24L^7. \end{aligned}$$

Looking at this one sees the need to have more shocks in order to capture the counterfactual perfectly, even at the assumed horizon of seven periods.

Now suppose we look at MW's RCF when there are three policy shocks. In that case τ is .92 and

$$\begin{aligned}\widehat{RCF}(L) &= 2.3 + 2.95L + 1.6L^2 + .59L^3 + .22L^4 + .08L^5 \\ &\quad +.03L^6 + .01L^7,\end{aligned}$$

producing quite major departures from the actual counterfactual. It is odd then that researchers present the results as if they have the counterfactual. The RCF may give the best estimate - in the Euclidean norm sense - whenever there are insufficient shocks, but it does not mean that it either produces the counterfactual or even impulse responses that are close to it, as the example above shows.

A question that arises is whether the problem of recovering a counterfactual is due to it being a fixed interest rate rule in this example. This issue arises since MW say in their on-line appendix about their empirical example that "Compared with the policy counterfactuals discussed in Section 4, we see that a nominal rate peg is a counterfactual policy that is not spanned particularly well by our available monetary shock evidence" (p.15) i.e. when they are using some numerical estimates of the shocks. This is the example where ρ was .78. Thus we are led to ask if one might recover the counterfactual if one adopts a new rule.

MW seemed to feel that it could be recovered if the counterfactual was a Taylor-type rule of the form

$$i_t = .5i_{t-1} + .5(1.5\pi_t + y_t).$$

Using the model in (15)-(16)

$$\begin{aligned}i_t &= .5i_{t-1} + .75\pi_t + .5\gamma(i_t - \pi_t) \\ &= \frac{.5}{1 - .5\gamma} i_{t-1} + \frac{(.75 - .5\gamma)}{1 - .5\gamma}\pi_t \\ &= .4167i_{t-1} + .7917\pi_t\end{aligned}$$

and so we must find s_j to solve

$$\begin{aligned}D(L) + C^*(L)s_1 + \sum_{k=1}^{7} C^{**}(L)s_{k+1} &= .4167L[D(L) + C^*(L)s_1 + \sum_{k=1}^{7} C^{**}(L)s_{k+1})] \\ &\quad +.7917(B(L) + A^*(L)s_1 + \sum_{k=1}^{7} C^{**}(L)s_{k+1})\end{aligned}$$

Let d be the elements in $D(L), d1$ those in $LD(L)$ etc., then s_j can be found from a regression of $(d - .4167d1 - .7917b)$ on $-c^* + .4167c1^* + .7917a^*$ and -$c_j^{**} + .4167c1_j^{**} + .7917a^*$.

The same model as before is used as before. Only the counterfactual policy rule differs. Now the CF for an horizon of 7 periods is

$$\begin{aligned} CF(L) &= 1.61 + 1.80L + .63L^2 + .15L^3 - .01L^4 - .03L^5 \\ &\quad -.03L^6 - .01L^7 \end{aligned}$$

When three policy shocks are used

$$\begin{aligned} \widehat{RCF}(L) &= 1.62 + 1.83L + .69L^2 + .25L^3 + .09L^4 + .03L^5 \\ &\quad +.01L^6 + .00L^7. \end{aligned}$$

With this counterfactual the RCF estimates are closer to CF than with the stable interest rate rule but there are still noticeable differences. The gaps get larger as the shocks get more persistent. Thus, when $\alpha = .65$ we would get

$$\begin{aligned} CF(L) &= 1.96 + 2.94L + 2.23L^2 + 1.59L^3 + 1.08L^4 + .71L^5 \\ &\quad +.45L^6 + .28L^7 \end{aligned}$$

$$\begin{aligned} \widehat{RCF}(L) &= 1.92 + 2.82L + 2.03L^2 + .1.31L^3 + .85L^4 + .55L^5 \\ &\quad +.35L^6 + .23L^7. \end{aligned}$$

5 Finding the Impulse Responses

A key issue in the MW methodology is how one finds the impulse responses to policy shocks including those for news. Suppose we don't have the DSGE model. We do know that there will be a SMA solution to it and so can take this as the general baseline model generating the data i.e. the DGP (Data Generating Process). To simplify the following discussion we initially take the simplest two variable model of section 3 in (1)-(2) and work with it and then move onto the empirical application in MW.

5.1 Finding Responses for the Simple Two Variable Model

The SMA solution to the model was

$$\pi_t = .443\varepsilon_t + .333\varepsilon_{t-1} - .333v_t - .11n_t - .333n_{t-1-} \quad (19)$$
$$i_t = .886\varepsilon_t + .666\varepsilon_{t-1} + .333v_t - .22n_t + .333n_{t-1}, \quad (20)$$

and this gave $B(L) = .443 + .333L$ etc. in (12). The counterfactual was $1.99 + L$.

Now suppose that ε_t is not known but the other shocks are. We might try to estimate the SMA unrestrictedly. As Fujiwara and Pagan (2024) point out this is not simple. Without any restrictions there are 12 parameters to estimate. There are three in the covariance matrix of π_t, i_t and four in the correlation between these variables and their lags, so not enough information to find the parameters without some restrictions.

To overcome this problem the literature has assumed that there is a proxy w_t for the shock of interest ε_t. For example the proxy for investment shocks used by MW comes from Ben Zeev and Khan (2015) while monetary policy and news shocks proxies are from Gertler and Karadi (2015) and Romer and Romer(2004). Suppose that $w_t = \varepsilon_t$. Then the system of three equations for w_t, π_t and i_t after replacing ε_t with w_t can be shown to be a VAR(2). Since w_t is placed first in this VAR the impulse responses to w_t will be those for ε_t. This approach is often used i.e. assume a VAR driven by the structural shocks with the VAR incorporating a proxy for the shock of interest in the counterfactual. Such a VAR exists when the SMA is invertible. A necessary condition for invertibility is that the number of shocks equals the number of observables, so this is satisfied when using proxy variables that equal the shocks.

Now assume that the proxy is not equal to ε_t but differs from it by a measurement error me_t i.e. $w_t = \varepsilon_t + me_t$. Then the system consisting of w_t, π_t and i_t is not invertible as there are three observables and four shocks since the system of equations for the observables w_t, π_t and i_t has measurement error included in their residual terms. This means that w_t will be correlated with those residuals and that leads to large sample bias in parameter estimates for the system. To look more closely at this assume that the measurement error has a variance of unity. Then OLS applied to each of the equations (19)-

(20) treating w_t as known (the "internal instrument" approach of Plagborg-Muller and Wolf (2021) (PMW)) produces

$$\begin{aligned} \pi_t &= .221w_t + .167w_{t-1} - .333v_t - .11n_t - .333n_{t-} \quad (21) \\ i_t &= .442w_t + .334w_{t-1} + .333v_t - .22n_t + .333n_{t-1}. \quad (22) \end{aligned}$$

It is clear that the impulse responses to v_t and n_t are unchanged, while those to w_t differ from the known values for ε_t by a factor of .5. These internal instrument responses in (21)-(22) will be termed $B^I(L), D^I(L)$ etc., and so $B^I(L) = .5B(L), D^I(L) = .5D(L)$, while $A^{*I}(L)$ etc equal $A(L)$ etc. The difference between them stems from the fact that the regressor w_t is correlated with the measurement error in the error terms, but the bias only affects responses to ε_t. This is basically a Linear Projection (LP) approach to recover the SMA.

Now apply the MW methodology to the system (21)-(22).The weights regression would now involve matching $-D^I(L)$ with $C^*(L)$ and $C^{**}(L)$ rather than $D(L)$ with these responses. That will mean the weights are now $.5s_1$ and $.5s_2$. Then the estimated counterfactual will be

$$.5B(L) + .5(s_1C^*(L) + s_2C^{**}(L)) = .5(1.99 + L),$$

and so the estimated counterfactual is one half of the actual one.

In the example above we see that $w_t = \varpi\varepsilon_t$ where $\varpi = 2$. The factor ϖ will depend on how important the measurement error is. As it declines ϖ will approach unity. Those using MW's methodology in this case go a further step before applying the approach. They work with new shocks $\eta_t^* = \phi w_t$ where ϕ is chosen so that the impact of η_t^* is prescribed to some value, say unity. In the example above $\phi = .221$ would be set. So there is a new system

$$\begin{aligned} \pi_t &= B^*(L)\eta_t^* + A^*(L)v_t + A^{**}(L)n_t \quad (23) \\ i_t &= D^*(L)\eta_t^* + C^*(L)v_t + C^{**}(L)n_t \quad (24) \end{aligned}$$

where $B^*(L) = B^I(L)/\phi, D^*(L) = D^I(L)/\phi$ and the other polynomials do not change. So we have $\eta_t^* = \phi\varpi\varepsilon_t$. The counterfactual response of π_t to η_t^* can be found but to get the counterfactual to ε_t from this one needs to know ϖ.

Therefore, to find the counterfactual to the DSGE model shock we need to be able to measure the importance of the measurement error. One way to

do that may be to estimate the system in π_t and i_t augmented by the extra equation $w_t = \sigma_\varepsilon \varepsilon_t + \sigma_{me} me_t$ using a program such as Dynare. For the simple two variable model more would be needed. When all shocks are measured with error one would have five observables, but only the covariance matrix would be available, as there are no dynamics. That means a maximum of fifteen parameters can be estimated in the system. There are sixteen in the system if all shocks are observed with error. Even in this case information has been used that would not normally be available, namely that v_{t-1} does not appear in the SMA. The extended two variable model does possibly allow the determination of the size of the measurement error as there are now lagged covariances as well. Of course it may be that those providing the proxies can give the necessary information.

The analysis above assumes that one is trying to recover the counterfactual from the DSGE model. What is really being recovered is the counterfactual if the SMA parameters were those in (21)-(22) rather than (19)-(20). In some applications one gets the impression that this is what researchers are focussed on. They fit an SVAR using w_t, π_t and i_t and impulse responses are reported under one policy rule that is in the data. Then results are found under a different policy rule, and a comparison is made of these latter results with the former. So the focus is on the system in (21)-(22). This does not measure the true counterfactual response of inflation to ε_t. It seems more of an exercise in looking at what changes one would get in impulse responses from a given SVAR results under a counterfactual. The parameter estimates of the SVAR are not those for the DSGE model whenever there is measurement error in the proxy for the shocks.

There are some extra issues about the estimation of responses coming up in the extended two variable model. In that case the inflation equation has π_{t-1} as a regressor as well as the shocks in (19), and so there is a large sample bias in the estimator of the coefficient of π_{t-1} due to its being correlated with the news shocks component. The large sample bias seemed small for the extended two variable model, although it might be larger for other DGPs. Consequently, we simply point out the issue.

5.2 Finding Responses for MW's Empirical Example

We consider MW's empirical example at the end of their paper. Just as in the previous sub-section the counterfactual involves stabilizing the interest rate. There are three shocks of interest and for which are proxies are used.

One is an investment specific technology shock drawn from Ben Zeev and Khan (2015), which we will call BZK_t. The others are monetary policy shocks from Gertler and Karadi (2015) and Romer and Romer (2004). It will be assumed that the BZK shock is measured with error and that the *actual* investment specific shock is ε_t. The other two proxies are taken to be exact measures of the policy shock v_t and n_t. Then we wish to study the consequences of measurement error for computing the response of inflation π_t to ε_t under the counterfactual policy. For this case one can only compute MW's "robust counterfactual" as there are only two policy shocks, so we do that in the same way as they do.

The relation between the proxy and actual investment specific shock is

$$BZK_t = \sigma\varepsilon_t + k\sigma me_t,$$

where ε_t and me_t are $n.i.d(0,1)$, σ is the standard deviation of the actual technology shock and k measures the importance of measurement error.[4] When $k = 0$ the BZK_t shock is the actual shock $\sigma\varepsilon_t$. For a given k, σ will be $\sqrt{var(BZK_t)/(1+k^2)}$, where $var(BZK_t)$ is fixed at the value from the sample.

This formulation says there is no serial correlation in BZK_t. Regressing this against 3 lags of MW's data on BZK_t, the output gap, inflation and the FFR we find that the F test for all influences being zero is .71 and the R^2 is .057, so it seems a reasonable assumption. It is not really needed but it is much simpler to understand why measurement error is important for computing counterfactuals.

Now to apply MW's methodology impulse responses are needed of observables to ε_t, v_t and n_t. In section 4.2 of their paper MW set up a recursive SVAR(3) in BZK_t, the output gap, inflation, and FFR. We do the same, so that the actual SVAR is taken to be a recursive SVAR(3) in ε_t, the output gap, inflation, and FFR.[5] Once this SVAR is estimated we may hope to

[4] Do MW allow a measurement error in the proxy shock? A comment we received on Fujiwara and Pagan (2024) was "I will here—and as in the applications of McKay and Wolf—assume that the researcher has access to valid IVs z_t given as

$$z_t = \alpha v_t + \sigma v_t + \sum_{q=0}^{\infty} \Psi_j z_{t-q} + \sum_{q=0}^{\infty} \Lambda_q y_{t-q}$$

where y_t denotes the vector of macroeconomic observables, and v_t is a measurement error".

[5] Under the assumptions ε_t is a strongly exogenous variable that is latent.

get the responses of observables to the actual technology specific investment shock ε_t. This is the same as MW's SVAR when there is no measurement error, because then $BZK_t = \varepsilon_t$.

Consider the equation for the output gap y_t in this SVAR(3). Because of recursivity it has the format

$$y_t = \sigma \sum_{j=0}^{3} a_j \varepsilon_{t-j} + lags(y_t, \pi_t, FFR_t) + u_t,$$

where u_t is some structural shock uncorrelated with ε_t. To estimate this we replace $\sigma\varepsilon_t$ with the observable BZK_t so it becomes

$$y_t = \sum_{j=0}^{3} a_j BZK_{t-j} + lags(y_t, \pi_t, FFR_t) + u(t) - k\sigma \sum_{j=0}^{3} a_j me_{t-j}.$$

The same is true of the other SVAR equations. Then MW estimate these equations by OLS and get estimates of a_j etc. to compute impulse responses. The estimate of a_j is not a consistent estimator of a_j since $corr(BZK_{t-j}, me_{t-j})$ is not zero. Hence the estimates of the impulse responses of variables to ε_t one gets from the proxy variable regression are not correct, unless there is no measurement error.

We apply MW's methodology to compute their robust counterfactual policy of a constant policy rate. Because we assume that the monetary policy proxies have no measurement error the impulse responses to the shocks v_t and n_t can be computed exactly as MW did i.e. from a recursive $VAR(4)$ in the proxies, the output gap, inflation, commodity prices and FFR, with the ordering the same as theirs. These are then used along with the estimates of the responses of π_t and FFR_t to ε_t from the SVAR(3) to compute the weights to be applied to get the robust counterfactual. This is the same as the method used in the preceding sub-section.

With MW's methodology we compute the counterfactual response of inflation to ε_t as k varies. MW use a contractionary investment specific shock. Below we give the "robust counterfactual" for $k = 0$ and $k = 1$.[6] The latter

[6]These relate to the middle graph in MW's figure 5. MW employ a scaling factor in their presentation. We also differ from them in using an LP method to compute the responses of inflation to ε_t, so ours are not as smooth. We aim here to show the role of measurement error so it is the comparison of responses as k varies that is important.

value is when measurement error has the same variance as the shock ε_t and the former is when there is no measurement error. Clearly they are different, just as for the simple model of the previous sub-section. Here the "robust counterfactual" responses computed with no measurement error are around 48% higher than when an allowance is made for such error. Consequently, in order to recover the actual counterfactual one needs to know how important measurement error is, and that is difficult to find but it cannot be ignored.

$$\begin{aligned}\widehat{RCF}(L) &= .241 + .256L + .166L^2 + .210L^3 + .095L^4 - .064L^5 \\ &\quad -.023L^6 - .038L^7 + .156L^8 - .103L^9, k = 0\end{aligned}$$

$$\begin{aligned}\widehat{RCF}(L) &= .163 + .174L + .112L^2 + .143L^3 + .064L^4 - .044L^5 \\ &\quad -.016L^6 - .026L^7 + .0702L^8 - .069L^9, k = 1\end{aligned}$$

One is running regressions to get the impulse responses, not doing instrumental variable estimation. Because measurement error means more shocks than observables there is no external instrument solution available to recover the initial responses.

6 Conclusion

Counterfactual analysis has long been a feature of macroeconomics. Recently it has re-surfaced with a greater emphasis being placed on scenario analysis as a vital part of the macroeconomic policy process. For example, the recent Bernanke Review of the Bank of England pointed to more scenarios being performed. Scenarios almost always involve counterfactual analysis. Because of this it seems important to examine a method set out by MW to capture counterfactual responses by just using impulse responses found from some baseline structural model. The method seems very promising for this task and has been enthusiastically adopted by a number of studies e.g. Georgiadis et. al (2024), Ider et al. (2024) and Castelnuovo et al. (2024). MW's claim that one does not need to know the actual structural model that

generated the data is very appealing, and the computer program made available which uses impulse responses to construct the counterfactual responses of interest has been widely used. Their methodology tries to avoid assuming that one knows the structural (DSGE) model incorporating expectations on the grounds that adopting any such model specification has the risk of being subject to specification errors. Our paper has firstly questioned if one can be sure that one is getting the counterfactual responses with their method, even if one knows the DGP, and we give simple examples where the estimated and actual counterfactuals are quite far apart. At the very least, it does not seem accurate to say that one has recovered the counterfactual when using MW's method. Many papers using the MW methodology make such a claim. One needs some criterion to tell the researcher how close the recovered counterfactual is to the actual one. It is unclear how this would be done unless one knew the counterfactual. MW's methodology is meant to do as well as possible using some norm - hence their term "robust counterfactual" - but whether it is good enough seems to be problematic.

On a separate but vitally connected issue the MW methodology also requires one to compute the impulse responses from a baseline model that may or may not be exactly the DGP. As Fujiwara and Pagan (2024) pointed out it is extremely unlikely that the SMA for the DGP can be found. To overcome this problem extra information (proxies) have been suggested by researchers - such as MW - as the extra information available to capture the shocks. Given this information researchers have assumed that the SMA can be captured via a SVAR. That requires invertibility. If the proxy is the shock then this is correct. But if there is measurement error the estimated impulse responses differ from those needed by a factor that depends on the importance of the measurement error in the proxy. Estimating this factor is a challenge and it is not currently done. Given that one needs many assumptions relating to the estimation of impulse responses in the MW methodology, and there is no certainty that we would recover the counterfactual since the number of proxy variables is limited, we are led to ask whether the cure is worse than the disease. Building a suitable DSGE model to perform counterfactual analysis seems to demand far less than all the adjustments and assumptions one would need to make in order to capture a counterfactual with MW's methodology and proxies.

7 References

Ben Zeev, N. and H Khan (2015). "Investment-Specific News Shocks and US Business Cycles," *Journal of Money, Credit and Banking*, 47, 1443–1464.

Buncic, D, A. Pagan and T. Robinson (2024) "Recovering Stars in Macroeconomics" *Available at http://dx.doi.org/10.2139/ssrn.4762983*

Burgess, S., E. Fernandez-Corugedo, C. Groth, R. Harrison, F. Monti, K. Theodoridis and M. Waldron "The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models", *Bank of England Working Paper No. 471.*

Castelnuovo, E., L. Mori and G. Peersman (2024) "Commodity Price Shocks and Global Cycles: Monetary Policy Matters", *CAMA Discussion Paper 36/2024*

Fujiwara, I and A. Pagan, (2024), "Re-Examining What We Can Learn About Counterfactual Results from Time Series Regression", *CAMA Discussion Paper* 44/2024.

Gertler,M. and P. Karadi (2015), "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Georgiadis, G., G. J. Müller and B. Schumann, (2024) "Global risk and the dollar", *Journal of Monetary Economics* 144 (2024).

Ider, G., A. Kriwoluzky, F. Kurcz and B. Schumann (2024), "Friend, Not Foe – Energy Prices and European Monetary Policy", Discussion Paper 2089, DIW Berlin.

McKay, A. and C.K. Wolf (2023), "What Can Time Series Regressions Tell us About Policy Counterfactuals", *Econometrica*, 91, 1695–1725

Pagan, A. R. and T. Robinson (2022), "Excess Shocks Can Limit the Interpretation," *European Economic Review*, 145, 104-120.

Plagborg-Muller, M and C.K. Wolf (2021), "Local Projections and VARS Estimate the Same Impulse Responses", *Econometrica*, 89, 955–980

Romer,C.D, and D. H. Romer (2004), "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.