Crawford School of Public Policy

# CAMA

**Centre for Applied Macroeconomic Analysis**

---

# Re-Examining What We Can Learn About Counterfactual Results from Time Series Regression

---


## CAMA Working Paper 44/2024
## June 2024



**Ippei Fujiwara**
Keio University
Centre for Applied Macroeconomic Analysis, ANU

**Adrian Pagan**
University of Sydney
Centre for Applied Macroeconomic Analysis, ANU



## Abstract

McKay and Wolf (2023) describe a method for finding counterfactuals which only requires that one know the impulse responses of shocks from a baseline structural model generating the data. A key feature in their work is the use of news shocks. This an elegant piece of theory and they indicate it can be applied empirically. We argue that one cannot recover the impulse responses from data generated by the structural model when there are news shocks as there are more shocks than observables in that case. We investigate an alternative proposal whereby some off model variables are used to find the requisite impulse responses and find that there are issues with doing that. Their theoretical result also relies upon the baseline structural model only having monetary policy operating via the interest rate channel so it excludes models that might be thought relevant for capturing data.



**Keywords**

counterfactual, news shocks, shock recovery, local projection

**JEL Classification**

C3, E3

**Address for correspondence:**

(E) cama.admin@anu.edu.au

**ISSN 2206-0332**

# Re-Examining What We Can Learn About Counterfactual Results from Time Series Regression*

Ippei Fujiwara† and Adrian Pagan‡

June 18, 2024

**Abstract**

McKay and Wolf (2023) describe a method for finding counterfactuals which only requires that one know the impulse responses of shocks from a baseline structural model generating the data. A key feature in their work is the use of news shocks. This an elegant piece of theory and they indicate it can be applied empirically. We argue that one cannot recover the impulse responses from data generated by the structural model when there are news shocks as there are more shocks than observables in that case. We investigate an alternative proposal whereby some off model variables are used to find the requisite impulse responses and find that there are issues with doing that. Their theoretical result also relies upon the baseline structural model only having monetary policy operating via the interest rate channel so it excludes models that might be thought relevant for capturing data.

**JEL Classifications: C3 E3**

*We have benefited from discussions with Alisdair McKay and Christian Wolf. In particular, we would like to thank Alisdair McKay for a great deal of assistance in enabling us to understand the central idea of McKay and Wolf (2023) and helping with the execution of their program code. Sophocles Mavroeidis and Qazi Harque provided important comments as well. When this project was started, Pagan was the Keio University Guest Professor (Global). Fujiwara acknowledges financial support from JSPS KAKENHI Grant-in-Aid for Scientific Research (A) No. 22H00058.

†Keio University and the Australian National University, ippei.fujiwara@keio.jp

‡University of Sydney, adrian.pagan@sydney.edu.au

**Keywords:** Counterfactual, News Shocks, 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 using models with forward looking expectations. A solution which emerged was to enforce the "fix" with the aid of anticipated shocks. When the fix was applied to an endogenous variable, inversion algorithms were required to translate these into values for the anticipated shocks. Burgess et. al. (2013) have an excellent discussion of this. Anticipated shocks are the equivalent of "news shocks" in academic models, being shocks that are known contemporaneously having been announced in the past.

Of course the approach above requires a fully specified model of the endogenous variables and any expectations. It is interesting therefore that McKay and Wolf (2023) (MW) seem to have circumvented that need by describing a method for 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). There are essentially two strands to this method. One defines a general family of models - we will refer to this family as the MW family (MWF) - that counterfactuals are to be performed on, and the other is that one can recover the counterfactual when one knows the casual effects of contemporaneous and news shocks from data. This MWF class is captured by their equation (6) and it stipulates that monetary policies only operate via the interest rate channel i.e. there are no lagged policy shocks in the system describing the economy. They point out that there are standard DSGE models that fall into this class. However, there are also models that are not in the class. Whether the data being analyzed has been generated from a model in this class becomes an issue that arises a number of times in the following sections.

Given this family of models MW distinguish between policy and non-policy shocks. If the DGP of the data is being captured by those models, they argue that any change in the policy rule which results in changes in the impulse responses to a selected non-policy shock can be found by combining the "causal effects of contemporaneous and news shocks to the prevailing policy rule". This insight is novel and unexpected. We provide an alternative derivation of how to find these combination weights. This utilizes lag operators and is perhaps a little clearer than their derivation. Section 2 describes it.

Essentially their method finds the weights by using the impulse responses from the Structural Moving Average (SMA) that is associated with the member of the MWF that generated the data. The latter is often referred to as a baseline structural model. The SMA is driven by the structural shocks of this baseline model. Section 3 sets out a simple two variable NK model in inflation and an interest rate and treats it as the baseline model. This model is a simplified version of what MW provide as their motivating example. There is an associated SMA which provides impulse responses for the structural policy and non-policy shocks. MW's proposition that possession of this information will enable the impulse responses for a non-policy shock upon variables to be recovered under some counterfactual monetary rule is found to hold.

The question we then ask is what are the implications of this theoretical insight for empirical work? MW seem to suggest that their method is not just of theoretical interest but can also be used for empirical work, and they provide an example of this. This raises the issue of whether we can recover the counterfactual without knowing the structural model that has future expectations in it. To do so it is necessary to recover the impulse responses to the structural shocks. This has to be done without knowing the form of the baseline structural model. On this MW comment (p 1696) that "Using standard time-series methods, she can estimate the causal effects of these policy shocks". Can this really be done? Using the data one can certainly determine the autocovariance matrices for the model variables under the baseline, and this will indicate the order of the Wold Moving Average (WMA) process that would characterize the data generated from it. But is this enough information to recover the SMA? The latter is needed in order to get the impulse responses to the structural shocks so as to capture the counterfactual.

"Empirically estimable" in their statement seems to involve not attempting to estimate the SMA but to utilize some extra (off-model) information and an alternative Impulse Recovery Process (IRP). One such process would be Local Projections (LP). To perform the latter it is necessary to recover the shocks whose impulse responses are required. Instead MW fit a SVAR to all the observables and recover some shocks and impulse responses from the SVAR. So there are two different IRP's here and one will most likely get different responses. Consequently, which should be chosen and how likely is it that either IRP can recover the SMA shocks?

Section 3 looks at these questions with our simple two variable model and asks if there is an IRP that will recover the SMA impulse responses when the baseline structural model is not known. We first observe that in the baseline model there are more shocks than observables. Specifically, there are 3 baseline model shocks and two observables. The latter fact means that there are only two Wold innovations, $e_t$. Consequently, the three baseline model shocks cannot be recovered from a knowledge of just two WMA innovations, unless one imposes some restrictions upon the SMA, and that requires knowledge of the baseline structural model. SVARs would not deliver the required shocks, one reason being that the baseline structural model has a VARMA solution. The latter form arises due to MW having "news" shocks in the baseline model that is the DGP generating the data. This section also asks whether one can recover the baseline shocks by extending the data set beyond that generated by the baseline model. Under some strong assumptions the answer would be in the affirmative, provided one used an IRP like LP. But we argue these are questionable assumptions.

News shocks are of fundamental importance to MW's method for calculating counterfactuals as they are needed for MW's theorems about the ability to generate the counterfactual to hold. To see this we construct a model where we replace the counterfactual policy of a stable interest rate in our two variable model with an interest rate rule, in the same way as in one of MW's examples. This produces impulse responses of sixth rather than first order and, in order to ensure that one can recover the counterfactual impulse response of interest, one needs five news shocks. The need for that many news shocks arises from the fact that the order of the polynomials in the lag operator describing the impulse responses is high, and there can be no lags in the monetary policy shock in the structural model so as to keep it in the MWF.

MW look at the issue of what one does if a large number of news shocks are needed and there are not enough observables to allow them to be determined. They propose a "robust counterfactual solution" that uses least squares to find a "best" predictor of the counterfactual. In our two variable model there are *more shocks than one needs.* Consequently, there is no need for a robust solution, and all one needs to accurately construct the counterfactual *is knowledge of the SMA impulse responses.* The problem is we cannot plausibly find the latter, so a "robust" solution, while it deals with robustness to the *number* of shocks, fails to answer the question of what one does if one cannot recover the SMA impulse responses. If we don't know the latter we cannot recover the counterfactual, even if we have more than enough shocks.

Section 4 examines some empirical work MW provide which exploits a range of monetary policy shocks that have been constructed in the literature. They place these in a recursive SVAR model, treating that as the IRP. We look at whether this is likely to recover the shocks from the structural baseline model that is the DGP, and find it seems very unlikely. Certainly the SVAR solution they offer is most likely bettered by an LP approach.

Finally, in section 5 we move away from trying to find the impulse responses for the SMA and return to the question of whether the MW family of models is broad enough to capture the responses that characterize the data. To do this we work with a model that is a simple extension of the earlier one. It features effects of monetary policy on variables that do not operate via the interest rate channel. In the MWF there are only effects via that channel. If this broader family produces the data then one can no longer recover the counterfactual responses with their method. The impulse responses coming from the non-interest rate channel seem plausible, and so they are being excluded by the MWF. This is quite separate from the problem described above of trying to recover the impulse responses when the baseline model is in the MWF. Section 6 concludes.

## 2 A Different Derivation

We will work with a simple two variable system consisting of inflation $\pi_t$ and an interest rate $i_t$. In terms of our later discussion nothing is different for more than two variables. This will be like their "illustrative example", where there is a monetary shock $v_t$, a news shock $n_t$ and an oil price shock $\varepsilon_t$. Following MW the first two shocks will be termed "policy" shocks. We will use the illustrative example throughout this paper. The relation between inflation, the interest rate and the three shocks is captured by an SMA

$$\pi_t = A^*(L)v_t + A^{**}(L)n_t + B(L)\varepsilon_t \quad (1)$$

$$i_t = C^*(L)v_t + C^{**}(L)n_t + D(L)\varepsilon_t. \quad (2)$$

The impulse responses are captured by the polynomials in the lag operator $L, A^*(L)$ etc. The coefficients attached to the $L^j$ are the $j'th$ period ahead impulse responses.

The SMA representation above comes from combining some baseline structural model with a monetary rule. Following that it is envisaged that there is a change to a different policy rule i.e. a counterfactual experiment. 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 they do. Now assume that the impulse responses of $\pi_t$ to $\varepsilon_t$ under the counterfactual is $CF(L)$. We want to put weights $s_1$ and $s_2$ on the baseline responses $A^*(L)$ and $A^{**}(L)$ so that

$$CF(L) = B(L) + A^*(L)s_1 + A^{**}(L)s_2. \quad (3)$$

To get $s_1$ and $s_2$ MW impose the counterfactual $i_t = 0$ on (2) and so

$$0 = D(L) + C^*(L)s_1 + C^{**}(L)s_2. \quad (4)$$

Defining $d$ as a vector containing the $L^j$ terms in $D(L)$, $c^*$ and $c^{**}$ doing the same for $C^*(L)$ and $C^{**}(L)$, they propose getting $s_1$ and $s_2$ by regressing $-d$ on $c^*$ and $c^{**}$.

To see how this strategy extends to a move to other counterfactuals consider what happens if we moved from a baseline rule $i_t = \delta\pi_t + v_t + n_{t-1}$ to a counterfactual $i_t = \phi\pi_t$. Here we are adopting the formulation in MW's equation (3), where $v_t$ is a contemporaneous shock and $n_{t-1}$ is a news shock about $t$ that is announced in period $t-1$. The equivalent of (4) is now found by imposing $i_t = \phi\pi_t$ to get

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

and so we would want $s_1$ and $s_2$ to satisfy

$$[C^*(L) - \phi A^*(L)]s_1 + [C^{**}(L) - \phi A^{**}(L)]s_2 = \phi B(L) - D(L). \tag{5}$$

Then, just as for the target interest rate case, we would run a regression, now of $(\phi b - d)$ against $(c^* - \phi a^*)$ and $(c^{**} - \phi a^{**})$ 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.

Two questions arise over this approach. The first is whether, given knowledge of the sextuplet $\mathcal{F} = \{B(L), A^*(L), A^{**}(L), D(L), C^*(L), C^{**}(L)\}$, there are values for $s_1$ and $s_2$ that produce an estimated $CF(L)$ from (3) that is the counterfactual. MW deal with that in their Proposition 1. We will assume, as they do in examples, that if one knows $\mathcal{F}$ there are values for $s_1$ and $s_2$ which produce the counterfactual. The second question is whether one can estimate $\mathcal{F}$ in practice *given data* available from the baseline model, so as to get $s_1$ and $s_2$ via a regression like (5). To investigate this it is necessary to specify some baseline structural model that generates the data and then to look at methods for recovering $\mathcal{F}$ from that data.

# 3 Example: From a Monetary Rule to an Interest Rate Target

## 3.1 A Simple Baseline Two Variable Structural Model

Consider a simple two equation baseline structural model with an interest rate rule

$$\pi_t = \beta E_t(\pi_{t+1}) - i_t + \varepsilon_t + \varepsilon_{t-1} \tag{6}$$
$$i_t = \delta\pi_t + v_t + n_{t-1} \tag{7}$$

This structure is meant to represent a simplified model that follows their equations (1)-(3). The SMA solution can be found 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 solution for the interest rate is

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

Thus we have found an SMA like in (1)-(2). 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 will be an interest rate target $i_t = 0$. This counterfactual sets $i_t = 0$ in (6)

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

and will be

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

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

To get $s_1$ and $s_2$ we solve $C^*(L)s_1 + C^{**}(L)s_2 = -D(L)$. That is done by matching powers of $L$. One needs some parameter values, and so we put these to $\delta = 1.1, \beta = .99$, thereby producing responses[1]

$$B(L) = 0.7007 + 0.4762L,\ \ A^*(L) = -.4762, A^{**}(L) = -0.2245 - .4762L \quad (11)$$

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

$C^*(L), C^{**}(L)$ and $D(L)$ are found from these. Because $D(L) = \delta B(L), C^*(L) = A^*(L) + 1, C^{**}(L) = A^{**}(L) + L$ the matching of powers of $L$ gives $s_1 = 2.189, s_2 = -1.1$ and then

$$\begin{aligned} \widehat{CF(L)} &= B(L) + s_1 A^*(L) + s_2 A^{**}(L) \\ &= 1.99 + L, \end{aligned}$$

which replicates the counterfactual.

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

$$\begin{aligned} d &= -[\ c^* \quad c^{**}\ ] \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ &= -Hs. \end{aligned} \tag{12}$$

A solution for $s$ is -$H^{-1}d$ i.e. the estimates of $s$ come from the regression of $-d$ on $c^*$ and $c^{**}$, and there is a perfect fit. 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 $\bar{c}^* s_1$. Because both $d$ and $\bar{c}^*$ are now $2 \times 1$ 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. If $d$ had $m$ non-zero elements it would be necessary to have $m$ starred $c's$ i.e. $m$ shocks, if the regression was to perfectly predict $d$. If there were less than $m$ shocks (say two) one cannot find all the necessary weights, and this led to MW's proposal that one get the two 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 maximizes the $R^2$ from this regression.

## 3.2 The Issue with Shock Recovery in the Simple Model

Now the key element in the solution above is being able to recover $A^*(L), A^{**}(L)$ and $B(L)$ from the SMA using data generated by the baseline model. Potentially, this can be done by recovering the shocks $v_t$ and $\eta_t$ and then regressing $\pi_t$ against contemporaneous and lagged values of $v_t$, $n_t$ and $\varepsilon_t$. This brings up the question of whether we can recover these shocks, particularly the one capturing "news", as that is a key factor in MW's results. The basic issue is that there are three shocks $\varepsilon_t, v_t$ and $n_t$ but only two observables $\pi_t$ and

$i_t$, and the recoverability literature shows one cannot recover all three shocks from fewer observables e.g. Pagan and Robinson (2022), *inter alia*. Often this set-up is referred to as a "short" system.

To understand the problem that a short system raises consider the case where there is just one observable $y_t$ driven by two structural policy shocks $\eta_{1t}$ and $\eta_{2t}$, where $\eta_{2t}$ is a news shock

$$y_t = a\eta_{1t} + b\eta_{2t-1}. \tag{13}$$

One can estimate this system and get parameter estimates $a, b$ provided one knows the structural model is that in (13). If we all we have is the data generated by (13) then the empirical covariances for $y_t$ will show it is an MA(1) and therefore can be written as

$$y_t = \omega_t + \alpha\omega_{t-1},$$

where, like $\eta_{1t}$ and $\eta_{2t}$, the error $\omega_t$ is white noise. We can estimate $\alpha$ and $var(\omega_t)$ from the data on $y_t$ and find the impulse response of $y_t$ to $\omega_t$. However $\omega_t$ is neither $\eta_{1t}$ nor $\eta_{2t}$. Indeed it is a linear combination of all $\{\eta_{1,t-j}\}, \{\eta_{2,t-j}\}$−see Nelson (1975). The $\omega_t$ are the Kalman filter prediction errors when the equation is placed into a state space form, and so there is only one of these, but there are two structural errors. In the same way, for our simple two variable model there are three structural shocks and two Kalman filter prediction errors, so it is not possible to recover all the three shocks.

Suppose we knew $\varepsilon_t$. Then we could get $B(L)$ and $C(L)$ by regressing $\pi_t$ and $i_t$ against $\varepsilon_t$ and its lags. This is because the other determinants of these variables are combinations of $v_{t-j}$ and $n_{t-j}$ and these are uncorrelated with $B(L)\varepsilon_t$ and $C(L)\varepsilon_t$, producing two observables $z_{1t} = \pi_t - B(L)\varepsilon_t$ and $z_{2t} = i_t - D(L)\varepsilon_t$. So now one would not have a short system, as there are two observables and two shocks.

In order to recover the shocks $v_t$ and $n_t$ in this case - the system is not short - one needs to know the SMA impulse responses, so we need to ask if these can be recovered from the data. Because there are two non-zero autocovariances for $z_t$, $z_t$ follows an SMA of first order, meaning it has the form

$$z_{1t} = a_{11}^0 v_t + a_{12}^0 n_t + \alpha_{11}^1 v_{t-1} + \alpha_{12}^1 n_{t-1} \tag{14}$$
$$z_{2t} = a_{21}^0 v_t + a_{22}^0 n_t + \alpha_{21}^1 v_{t-1} + \alpha_{22}^1 n_{t-1}. \tag{15}$$

There are eight parameters which need to be determined in the SMA. The contemporaneous covariance matrix of $z_t$ , $\Gamma_0$, has three parameters, and the lagged covariance matrix, $\Gamma_1$, has four, so there are only seven parameters which summarize the data. Hence a restriction needs to be placed on this SMA if we are to recover the impulse responses. One restriction that might be applied is that $v_t$ only has contemporaneous effects on $z_{1t}$ i.e. $a_{21}^0 = 0$, which involves some knowledge of the structural system. This does not work when there are three shocks to be found from the two observables, as there are still only seven parameters in the covariance matrices, while the SMA would now have twelve parameters. Accordingly, setting just $a_{21}^0$ to zero leaves too many unknowns at eleven.

What other IRP might be used? One which only estimates seven parameters is a recursive SVAR. Defining $\phi_t = \begin{bmatrix} \phi_{1t} \\ \phi_{2t} \end{bmatrix}$ the recursive SVAR in $z_t$ will be $B(L)z_t = \phi_t$, with $B(L)$ being triangular. The estimated shocks $\phi_t$ will have covariance matrix $I_2$. The model shocks $v_t$ and $n_t$ also have this. Let these be $\eta_t$. Then $\phi_t = Q\eta_t$, where $Q$ is orthonormal. The problem is that without extra information $Q$ is not unique, and so the impulse responses for $v_t$ and $n_t$ cannot be computed.[2] The responses found from the SVAR for $\phi_t$ are not the $A^*(L), A^{**}(L)$ from the baseline model SMA. Indeed, treating them as if they were, we would find that $\widehat{CF(L)} = 0.301 + .297L$, and not $1.99 + L$. To understand the failure of this approach we observe that in an SVAR one of the variables $z_{1t}$ or $z_{2t}$ must have a shock that is either $v_t$ or $n_t$, but this is not the case for (14)-(15). So imposition of a SVAR implies a mis-specified SMA.

Is there any other way around this dilemma? One needs to either eliminate shocks as above or look for other observables. If one just adds in a new observable it would be stochastic and so introduce a new shock, keeping the system short. To see that consider a simple extension of our two variable system that is very close to what MW use in their equations (1)-(3).

[2] There are some questions that can be answered by recovering $\phi_t$ i.e. one does not need $\eta_t$ but some combination of them that involves an orthonormal matrix $Q$. One of these is the recovery of the variance of a target variable under the counterfactual. The situation arose in sign restriction studies using a $Q$ and was pointed out in Fry and Pagan (2011, p 955). This has been exploited by Caravello et al. (2013).

$$y_t = \gamma(i_t - \pi_t) + \omega_t \quad (16)$$
$$\pi_t = \beta E_t(\pi_{t+1}) + \psi y_t + \varepsilon_t + \varepsilon_{t-1} \quad (17)$$
$$i_t = \delta \pi_t + v_t + n_{t-1}. \quad (18)$$

This system has three variables but still more (four) shocks. One cannot recover all shocks, but perhaps one can recover those of interest - $\varepsilon_t, v_t$ and $n_t$. To do this we examine the covariance matrix of the difference between the estimated and actual shocks, $P^*_{t|t}$, given in Pagan and Robinson (2022). The $P^*_{t|t}$ comes from the Kalman filter information. For a system to be invertible $P^*_{t|t}$ must equal zero, as then one can recover the actual shocks via a regression of them on the estimates. The diagonal elements of $P^*_{t|t}$ tell whether a particular shock is recoverable. Designating the $j'th$ element of $diag(P^*_{t|t})$ as $\xi$, this scalar index lies between zero and 1, as it equals $1 - R^2$, where $R^2$ is from the regression of the actual shock against the estimated filter shocks - see Buncic et. al. (2024). It has to be zero for invertibility to hold for that shock.

Moreover, the index provides information about the degree of recoverability (in an infinite sample). With $\gamma = -.5, \psi = .6, \beta = .99$ and $\delta = 2$ we find that $diag\{P^*_{t|t}\} = \begin{bmatrix} 0 & .02 & .34 & .91 \end{bmatrix}$, so that the only shock which can be recovered is $\omega_t$.[3] The index is small for $\varepsilon_t$ but the other two shocks cannot be recovered, in particular the news shock.[4] This is an issue with having news shocks in the baseline model. They cannot be omitted from the baseline model. If they are then one cannot capture the counterfactual.

## 3.3 A Recovery Method Consistent with the SMA?

There are some approaches in the literature in which the baseline model is augmented with extra data, and the resulting observables are treated as noisy

[3]To see why this is so we note that $(i_t - \pi_t)$ is autoregressive and can be instrumented with $(i_{t-1} - \pi_{t-1})$ to recover $\gamma$ and $\omega_t$. It would then be possible to proceed in the same way to get $u_t = v_t + n_{t-1}$. But this is an MA(1) so $v_t$ and $n_{t-1}$ cannot be recovered from that.

[4]For recovery of shocks from a SVAR representation one needs to look at $P^*_{t|t}$ since invertibility is needed. Another way of capturing the baseline model shocks would be to use all the data rather than just current and lagged values, i.e. estimate them via the Kalman smoother rather than the filter. Then it is necessary to examine $P^*_{t|T}$ . The conclusions about recovery of impulse responses remain the same.

signals of the shocks. Thus Gertler and Karadi (2015) assumed that there was an observable based on futures, $f_t$, and it was used as an external instrument to identify a monetary shock. However, if one is to follow that approach one has to be careful that no extra shocks are introduced, as one would then still have a short system. Instead, *assume* there are three observables $\zeta_{1t}, \zeta_{2t}$ and $\zeta_{3t}$ which relate to the policy shocks in the following way

$$\begin{aligned}\zeta_{1t} &= g_{1t-1} + \varepsilon_t &(19)\\ \zeta_{2t} &= g_{2t-1} + v_t &(20)\\ \zeta_{3t} &= \gamma\zeta_{2t} + g_{3t-1} + n_t. &(21)\end{aligned}$$

Here $g_{jt-1}$ summarize the contribution from past observables. The past observables must either be in the baseline model or be strictly exogenous.[5] If $g_{jt-1} = 0$ then the observables $\zeta_{jt}$ would be the baseline structural model shocks. Of course the specifications above are an extra set of assumptions not present in the baseline model.

Now the two variable system can be written as

$$\begin{aligned}\pi_t &= A^*(L)v_t + A^{**}(L)n_t + B(L)\varepsilon_t\\ i_t &= \delta\pi_t + v_t + n_{t-1},\end{aligned}$$

and, after substituting for the policy shocks from (19)-(20), it becomes

$$\begin{aligned}\zeta_{1t} &= g_{1t-1} + \varepsilon_t &(22)\\ \zeta_{2t} &= g_{2t-1} + v_t. &(23)\\ \zeta_{3t} &= \gamma\zeta_{2t} + g_{3t-1} + n_t &(24)\\ \pi_t &= A^*(L)(\zeta_{2t} - g_{2t-1}) + A^{**}(L)(\zeta_{3t} - \gamma\zeta_{2t} - g_{3t-1}) &(25)\\ &\quad + B(L)(\zeta_{1t} - g_{1t-1})\\ i_t &= \delta\pi_t + (\zeta_{2t} - g_{2t-1}) + (\zeta_{3t-1} - \gamma\zeta_{2t-1} - g_{3t-2}). &(26)\end{aligned}$$

Looking at the system in the five observables it is clear that it is recursive, although $\zeta_{3t}$ is missing from the interest rate rule and there is no shock in that equation once the three model shocks are removed. Perhaps of most significance is that $\zeta_{3t}$ would have to precede $\pi_t$ and $i_t$ in any recursive ordering.

$B(L)$, $A^*(L)$ and $A^{**}(L)$ could be found by estimating (22) – (24) to get $\varepsilon_t, v_t$ and $n_t$ and then using LP to get them. We will refer to this strategy

[5] Thus one cannot allow $\zeta_{3t}$ to depend on inflation as it depends on the shock $n_t$.

as adjusted MW (AMW). In order to proceed in this way one needs some observable measures $\zeta_{jt}$ to work with and we return to a more concrete analysis of it when discussing MW's empirical example in section 4. In that example MW do not use the LP strategy but rather an IRP that is a recursive SVAR fitted to all observables. So it is like the system (22)-(26) above. It is therefore worth discussing some conceptual issues with this IRP using the simple system above. The first step is to recover shocks by estimating the equations (22)-(24). In the second step the recursive SVAR is used to produce impulse responses to the shocks. Now the covariance matrix of the observables is singular, since there are only three shocks in the system, and that would mean identities link $A^{*}(L)v_t, A^{**}(L)n_t$ and $B(L)\varepsilon_t$. A singular covariance matrix is unlikely to be the case for the data. To overcome this one might assume that the model variables $\pi_t$ and $i_t$ are measured with white noise error. If they are not white noise then extra dynamics are introduced that are not in the structural model. Thus observed inflation $\pi_t^D$ is different to $\pi_t$ and the inflation equation has the measurement error as its shock. One cannot replace $\pi_t$ with $\pi_t^D$ in the interest rate rule as that would mean that the error terms in (25) and (26) are correlated, so the rule must incorporate just the latent inflation $\pi_t$. It is therefore clear that using an SVAR as an IRP is problematic in this simple model. Things would worsen if the interest rate rule had a term like $i_{t-1}$ in it, since that would mean extra shocks with serial correlation. The AMW approach seems a much more robust way of capturing $B(L)$, $A^{*}(L)$ and $A^{**}(L)$, although it uses an assumption about the additional data that is not part of the baseline structural model. Moreover, there is no certainty that the error term in the regression will be the shocks which one wants. It might be that the equation for $\zeta_{2t}$ has an error term that is either a MA(1) in $v_t$ or even a combination of $v_t$ and $n_t$.What the connection between the constructed SVAR and $SMA$ shocks was never clarified by MW. *In order for MW's theoretical results on replication of counterfactuals to be used one has to have the shocks obtained from the selected IRP being the SMA shocks.*

Another comment on how the MW material is being applied is worth making. In the MW approach one needs to begin with a baseline model that incorporates both monetary and news shocks. Some applications of MW's method start with a recursive SVAR in observables and then suggest that a counterfactual can be recovered using some extra observables such as $\zeta_{jt}$. This is an incorrect use of MW's result. Because $CF(L)$ will have an order greater than zero the baseline model must have a news shock in it and so it will not be a VAR process. Indeed, if the counterfactual is that an interest rate is held constant, imposing that on the initial SVAR would lead to an unstable model.

### 3.4 The Role of News and Shock Recovery in A Modified Example

Let us look more closely at the role of news shocks. We return to a three variable model in (16)-(17), now setting $\psi = .5$ and $\delta = 1.5$ for the counterfactual rule. Solving this gives the counterfactual response of $\pi_t$ to $\varepsilon_t$, and it emerges that $CF(L)$ is essentially a sixth order polynomial - the first six impulse responses captured by $CF(L)$ are non-zero and they are very very small after that. To replicate $CF(L)$ six "policy" shocks will now be needed. The weights for the policies will be found from matching the polynomial terms in

$$[C^*(L) - \phi A^*(L)]s_1 + \sum_{j=1}^{5}[C_j^{**}(L) - \phi A_j^{**}(L)]s_{j+1} = \phi B(L) - D(L). \quad (27)$$

So the baseline interest rate rule needs to incorporate six shocks. Just as anticipated shocks were used to replicate fixes in macroeconometric models, here one needs to use five news shocks $n_{jt}$, $j = 1, .., 5$. The baseline interest rate rule then has the form

$$i_t = \delta\pi_t + v_t + n_{1,t-1} + n_{2,t-1} + n_{3,t-1} + n_{4,t-1} + n_{5,t-1},$$

where $\delta$ is 1.1. Using this and the impulse responses for the six shocks from the baseline structural model (16)-(17), we confirm MW's theoretical result that the counterfactual response of $\pi_t$ to $\varepsilon_t$ can indeed be recovered, and it is

$$CF(L) = .4904 + .2284L - .0619L^2 + .0168L^3 - .0045L^4 + .0012L^5. \quad (28)$$

One might also ask what would happen if we only used two policy shocks and the baseline model was in the MWF? Then the regression being run of $(\phi b - d)$ against $(c^* - \phi a^*)$ and $(c^{**} - \phi a^{**})$ does not have an $R^2 = 1$ (unlike the one with six policy shocks). The predicted counterfactual - MW's "robust counterfactual" - would be

$$\widehat{CF(L)} = 0.2356 - .2186L + .0378L^2 - .0088L^3 + .0021L^4 - .0005L^5.$$

Even if one just looked at the first two terms there is no match to the counterfactual. Notice that in order for us to quantify the extent to which the robust counterfactual fails to capture $CF(L)$ we needed to know what the latter is. The "robust counterfactual" fails to capture $CF(L)$, so shouldn't one just conclude that? Unless $CF(L)$ is known one cannot know how far short the estimate is of the correct value but it does fail to capture the actual counterfactual. Isn't it better to use the structural model to perform counterfactual analysis than some short-cut that fails to produce the correct counterfactual?

One important point to note is that the "robust counterfactual" is a solution that aims to overcome the problem of having too few recoverable shocks to replicate $D(L)$. When there are more than enough, as in our two variable model, there is no need for a robust solution, but there is still the other crucial factor that one has to exactly recover $\mathcal{F}$. The inability to do this comes from having *too many shocks*. The "robust counterfactual" does not address that concern.

Looking at this problem through the lens of autocovariances $\Gamma_j$ there are three observables in (16)-(17). Consequently, this means six parameters in $\Gamma_0$ and nine in each of the $\Gamma_j$ $(j = 1, .., 6)$. So a total of sixty parameters. In contrast the unrestricted SMA is a Vector MA process of order six, and each of the MA matrices has thirty elements, giving two hundred and ten parameters to estimate. To do that with the available information means one hundred and fifty restrictions are needed on the SMA.

# 4 An Application to Exploiting Existing Studies of the Impact of Monetary Shocks

What do we conclude for empirical work from the analysis of section 3? Firstly, it is hard to see how one can recover the SMA without a great deal of knowledge of the structural model, particularly when the baseline structural model has news shocks. There are more shocks than observables, and it has to be very rare that one can recover the impulse responses that one needs to recover in order to replicate the counterfactual. MW's theoretical work that one can recover the counterfactual if one knows the impulse responses from the SMA is not in dispute; it is the ability to recover the latter impulse responses that is.

It is hard to know what to make of the application that has probably attracted most attention to the MW paper. It features their proposal to use impulse responses found from a variety of measured monetary policy shocks in order to construct "robust counterfactual" responses. One has to interpret the data on monetary shocks as being able to capture, to some degree, the shocks embodied in the baseline model SMA which generated the data. In MW's example (Figure C.2 of their paper) one wants to find the response of inflation to an investment shock $\varepsilon_t$ under a counterfactual monetary rule. So, as in the examples of the previous section, one needs to get $\mathcal{F}_t$.

Following the discussion in section 3.3 MW utilize data on three constructed variables $\zeta_{jt}$ to capture the shocks of interest $\varepsilon_t, v_t$ and $n_t$ - investment shocks coming from Ben Zeev et. al. (2015) (BZ), monetary policy shocks from Gertler and Karadi (2015) (GK) and Romer and Romer (2004) (RR). A strong interpretation of these observables would be that *they are* $\varepsilon_t, v_t$ and $n_t$. However, as we observed when discussing the two variable model in section 3.3, it seems more general to allow the observable shocks to differ from the baseline structural shocks via a wedge of past observables. Then one might regress the BZ, GK and RR observables against past observables and recover the shocks from those regressions, treating them as $\varepsilon_t, v_t, n_t$. We referred to those estimated shocks as being AMW. Because MW fit a SVAR(4) as their IRP, with the observable GK ordered first, the AMW estimate of $v_t$ from the GK regression is the residual from the first equation of their SVAR. The local projection of inflation upon the AWM shocks give $A^*(L)$. Their estimate comes from the SVAR(4) in all the observables.[6] Figure 1 shows these for inflation and one sees that MW's estimated responses are a smoothed version of the LP values.

However RR is different. MW find $n_t$ by regressing RR against all the variables of the SVAR except the contemporaneous interest rate (lags of this are included). They order the RR variable *after* inflation and just before the final variable that is the interest rate. The analysis in section 3.3 points to a problem with doing that - the additional variables had to precede inflation. Therefore we construct another estimate of $n_t$ by utilizing this re-ordered

[6] As well as BZ,GK and RR there are variables for an output gap, inflation, commodity prices and the interest rate.

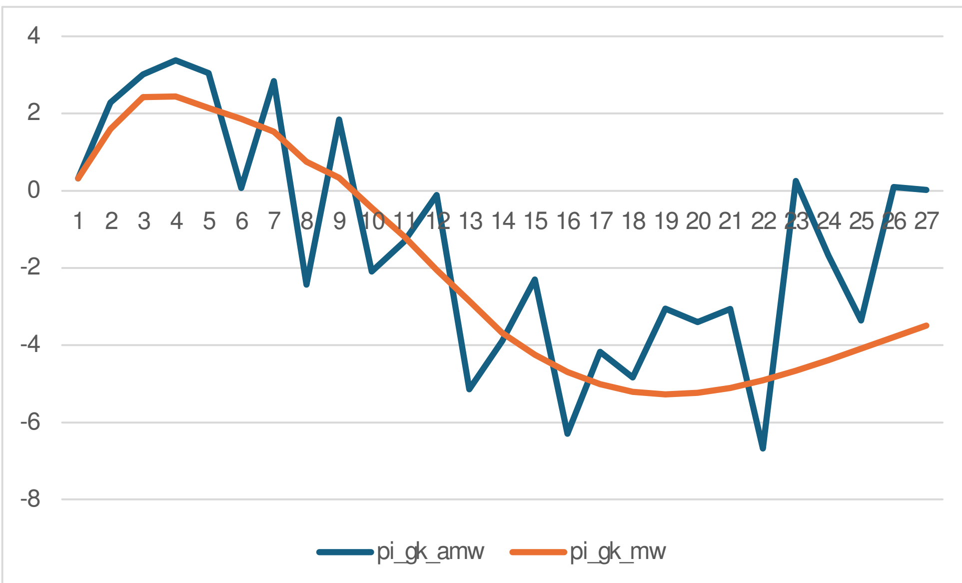


Figure 1: Responses of Inflation to estimates of $v_t$ from LP (AWM) and MW's SVAR(4)

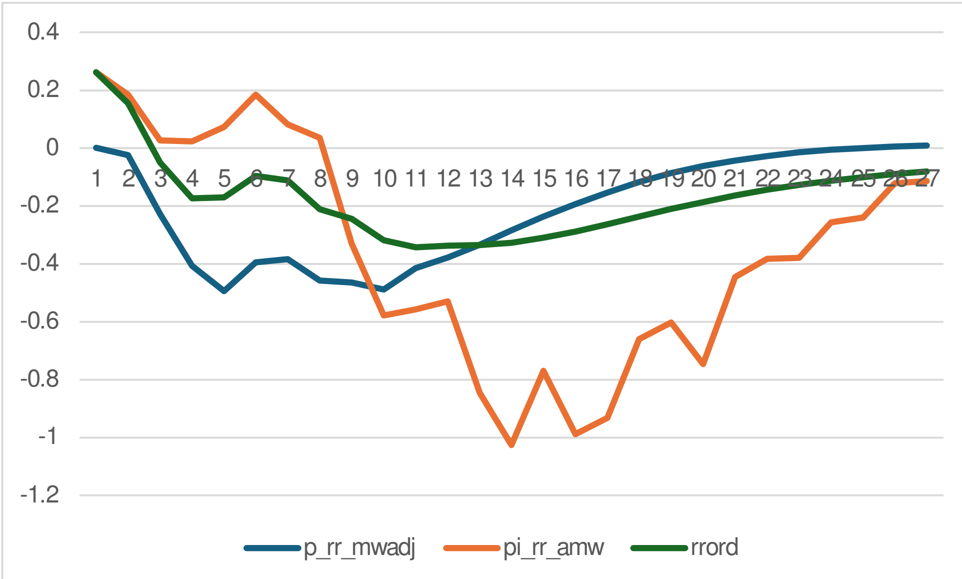


Figure 2: Impulse Responses of Inflation to a Unit News Shock from the MW SVAR(4) (mwadj), the Adjusted MW (AMW) shocks, and Reordered SVAR(3) shocks (rrord)

SVAR and get the responses from that reordered SVAR. Impulse responses to the AMW shocks using RR are found by local projection. Figure 2 shows the inflation responses from the three different IRPs. They produce quite different estimates of what would be $A^{**}(L)$ and so different counterfactual solutions. Given the analysis of the two variable model in section 3.3 it would seem that the AMW estimates would better capture the SMA shocks.

# 5 A Broader Two Variable Model

Now, as seen from the value of $A^{*}(L)$ in the simple two variable model of section 3.1 there is only a contemporaneous effect of the $v_t$ shock on inflation. Why shouldn't there be a lagged effect? To look at that we assume that the baseline model generating the data and used to compute responses is

$$\pi_t = \beta E_t(\pi_{t+1}) - i_t + \varepsilon_t + \varepsilon_{t-1} - \mu v_{t-1} \quad (29)$$
$$i_t = \delta \pi_t + v_t + n_{t-1}. \quad (30)$$

Using the same parameters as for the model of section 3.1 and $\mu = 1.2$, $A^*(L) = -.7456 - .5714L$, while $B(L), A^{**}(L)$ are the same. The responses in $A^*(L)$ don't seem implausible. Solving for $s_1$ and $s_2$ and computing $\widehat{CF(L)}$ it is $-6.4863 - 7.5619L$, and this does not recover the counterfactual.

Why don't we manage to produce the counterfactual with the the $\mathcal{F}$ above? The answer lies in the fact that the baseline model in (29) and (30) does not fit into the "general family" of baseline structural equations that MW set up (their equation (6)). In the context of this simple two variable case the MWF allows only for an interest rate channel and so it precludes a *direct* effect of $v_{t-1}$ upon inflation, allowing only an indirect one that comes via the lagged interest rate i.e. from an equation like

$$\pi_t = \beta E_t(\pi_{t+1}) + \varepsilon_t + \varepsilon_{t-1} - \alpha_1 i_t - \alpha_2 i_{t-1}.$$

Thus one can recover the counterfactual if the baseline model is from the MWF but one would not recover the counterfactual with this broader class of model where the effects of monetary actions comes from channels other than the interest rate. It might be that either announced news or the unanticipated shock could have an effect upon sentiment or credit, and this could affect output and inflation, even when there is a minor response of these variables to the movement in the interest rate. This means that it is necessary to ask how we know that the data has been generated by a member of the MWF i.e. has the data produced impulse responses that do not come from a member of the MWF family, as in the example above?[7]

# 6 Conclusion

The simple method set out by MW to capture counterfactual responses by just finding impulse responses from some baseline structural model seems

[7]It might be asked why one can not find a MWF that will generate the same impulse responses as the broader baseline model just used. Suppose one could. Then one would be getting a trio of responses that fail to produce the counterfactual and that would reject MW's result that the MWF can reproduce the counterfactual.

very promising. If one does have the complete set of responses $\mathcal{F}$ from the SMA the method works as promised. So, as a piece of theoretical work, it is fine. But to be of empirical use one has to find $\mathcal{F}$. It is necessary to find that *exactly,* if one is to recover the counterfactual impulse responses of interest. The problem with MW's procedure is that it works with baseline models that have many news shocks, and that results in more shocks than observables. In such a context one cannot expect to recover all the shocks of the SMA and their impulse responses. Very strong assumptions on the SMA would be needed to do that. It is unlikely that structural models with forward looking expectations will deliver SMAs that have simple zero restrictions, as the whole point of DSGE models is that they involve cross equation restrictions. However, MW maintain that knowledge of the baseline structural model is not needed. Adopting simple models that are close to those they use for motivating their theory we saw that a great deal of information about the structural model is needed if you are to recover the SMA. MW move away from estimating the SMA and propose alternative Impulse Recovery Processes that are meant to emulate the SMA. We look at an SVAR that they use as an IRP and point out that there are conceptual issues with such a choice. We first do this in the context of a simple model and then in connection with their empirical work.

We also raise a question about the utility of their theoretical result. The family of models that they adopt only has policy working through an interest rate channel. This restricts the range of impulse responses to policy shocks. To us there is nothing which says that the *DGP of the data* must be captured by a member of the MWF. We show that such models exclude plausible impulse responses. It may be that the data rejects such impulse responses but one needs to show that.

# 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 (2013) "The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models", *Bank of*

*England Working Paper No. 471.*

Caravello, T.E., A. McKay and C K. Wolf (2023) "Evaluating Policy Counterfactuals: A “VAR-Plus” Approach" (mimeo)

Fry, R. and A. Pagan, (2011). "Sign Restrictions in Structural Vector Autoregressions: A Critical Review," *Journal of Economic Literature*, 49, 938-960.

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

Harrison, R., K. Nikolov, M. Quinn, G. Ramsay, A. Scott, and R. Thomas (2005), *The Bank of England Quarterly Model*, London: Bank of England

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

Nelson, C. R. (1975), “Rational Expectations and the Predictive Efficiency of Economic Models,” *Journal of Business*, 43, 331-343.

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

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