Causal mechanism and mediation analysis for macroeconomics dynamics

Jean-Marie Dufour[1] and Endong Wang[2]

December 28, 2023

[1]William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

[2]Department of Economics, McGill University, Montréal, Québec H3A2T7, Canada. TEL: 514 772 7078; Email: endong.wang@mail.mcgill.ca.

Abstract

In macroeconomics, a central focus is placed on unraveling the roles of specific variables in the transmission of exogenous shocks to a target variable across varying time periods. One influential study in this topic is Bernanke's (1995) exploration of the credit channel's role in transmission of momentary policy shocks to real output. This paper introduces an innovative index designed to provide a quantitative measure of a mediator's impact as either an amplifier or an attenuator within a dynamic system during the transmission of a specific shock over time. The index offers researchers a clearer perspective on the underlying transmission mechanisms. Our research is rooted in the Structural VAR model. Firstly, we present the concept of 'impulse response decomposition', which illustrates that, for a given impulse response horizon and time of decomposition, the impulse response is contributed to the movements of variables triggered by the initial intervention. Secondly, we engage in a hypothetical scenario at the time of evaluation where the mediator of interest exerts no causal influence on the target variable, holding all other factors constant. The index is computed based on the ramifications of this counterfactual analysis on the target variable. The causal relationship between the mediator and the target variable, while keeping all other factors constant, is assessed through Granger causality over multiple time horizons, as formalized by Dufour and Renault (1998). This pivotal counterfactual analysis is akin to assuming that the target variable responds to all variables except for the mediator of interest. Lastly, we apply our channel index to quantify the role of the credit channel in transmitting monetary policy shocks, shedding light on the intricate dynamics of macroeconomic responses to such shocks.

Keywords. Causal channel, mediation analysis, dynamic causal effect, Granger-causality

Contents

1 Introduction

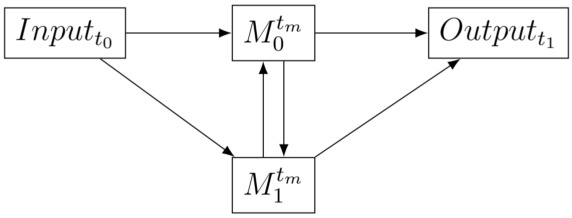
The examination of the dynamic causal effects of external interventions in macroeconomics has been a focal point of extensive research, primarily to inform policy decisions. The foundational work in intervention analysis can be traced back to Box and Tiao (1975), who introduced various polynomial representations to capture the dynamic configuration of causal effects based on prior economic knowledge. Subsequently, Sims (1980) made significant contributions by emphasizing impulse response functions to capture dynamic causal effects within the context of Structural Vector Autoregression (SVAR) models. Impulse responses as functions of VAR coefficients are employed to quantify the causal effects of exogenous one-time orthogonal shocks. The use of VAR models made the study of dynamic effects more accessible, leading to various empirical results and insights into this field.

While impulse response functions provide a direct and interpretable view of dynamic causal effects, researchers have increasingly sought to understand the intricate mechanisms underlying these effects. This quest led Bernanke and Gertler (1995) to question the "black box" nature of empirical research in the field of monetary policy transmission, highlighting the importance of elucidating the intermediate processes. This focus on the role of intermediary variables, often referred to as mediators, has brought forth the concept of causal channels. These channels elucidate how specific variables mediate the transmission of causal effects, shedding light on the dynamics of a monetary policy shock's causality.

Despite numerous studies addressing causal channels for specific dynamic causal effects, there has been a notable lack of a unified econometric theory or a quantitative measure for these channels. Moreover, existing literature has been largely silent on the dynamic nature of causal channels within a dynamic system. This omission is crucial because certain channels might play different roles during the transmission of a shock, providing an opportunity to delve deeper into the "black box" of causality. To address this gap, we propose a methodology for measuring the magnitude, timing, and composition of an output variable's response to an exogenous one-time shock concerning specific variables representing the causal channels of interest. Our approach, known as "impulse response decomposition," unveils the contribution of each variable and their causal interconnections during the decomposition period. This process not only reveals the significance of particular variables during the shock transmission but also constructs a data-driven causal chain.

Our methodology is rooted in a fundamental causal philosophy that the response of an output variable to an exogenous shock is not realized unless each intermediary variable responds endogenously. This aligns with mediation analysis, where the intermediary variables are often referred to as mediators. To illustrate our research objectives and the causal philosophy, we present a heuristic paradigm. At time t_0, an exogenous shock occurs, while the output at time t_1 represents the movement of the output variable. We select an intermediate time, denoted as t_m, for decomposition or evaluation. At this time, we distinguish between two sets of mediators: $M^{t_m} = \{M_0^{t_m}, M_1^{t_m}\}$, where M_1 represents the causal channel of interest, and M_0 comprises the remaining variables. Our measurement of the significance of specific variables involves a counterfactual analysis, where we assess how the output variable would respond if all variables acted as they did in the standard case, except for the mediator of interest, M_0.

Crucially, we allow the time of decomposition, t_m, to take any value between the initial



time, t_0, and the realization time, t_1. This setup acknowledges that a given variable or mediator may have varying contributions to the output response over time. This approach reveals the underlying philosophy of a causal chain for a specific causality in dynamic macroeconomics. Moreover, deliberately suppressing a specific variable's contribution to the output variable response aligns with the principles of Granger-type causality, focusing on the contribution of one variable at the time of decomposition to the output variable's response at some point in the future. Consequently, the coefficients of Granger-type causality and the endogenous responses of each mediator are essential for computing the index that reveals the significance of specific channels.

Our research bridges the gap between Granger causality and impulse response functions, offering a quantitative measure for the transmission mechanisms of exogenous shocks in macroeconomics. By employing mediation analysis, we shed light on how different variables contribute to the total effect of the output response. Our approach involves decomposing impulse responses over various time horizons through a multi-horizon VAR model and its coefficients, known as Generalized Impulse Response Coefficients (Dufour and Renault (1998)).

Links with literature: Our research is primarily founded on the SVAR model, as understanding the contribution of each variable to an impulse response necessitates a comprehensive grasp of the model's structure. Although this paper is conceptual, the identification, estimation, and inference for SVAR models are crucial, as discussed in Montiel Olea et al. (2021). However, it is reasonable to suspect structural breaks or regime switches in stable VAR models, given that the economic environment is subject to change over time. Researchers have explored non-linear VAR models to account for dynamic causal effects and parametric assumptions (Kilian and Vigfusson (2011); Gonçalves et al. (2021)). Consequently, we propose a general framework that accommodates both linear and non-linear cases before delving into the specifics of the linear VAR model.

The study of causal channels draws inspiration from classic mediation analysis, which also considers moderators when the causal effect depends on the state of the economy (Baron and Kenny (1986)). In a linear model like VAR, the causal effect, as measured by VAR coefficients, remains constant regardless of economic conditions or variable values. However, in non-linear VAR models, the causal effect varies over time. This variation introduces the possibility of impulse response decomposition being a stochastic or path-dependent function, aligning with the criticism put forth by Lucas Jr (1976). We discuss non-linearity, moderation, and Lucas' criticism, acknowledging that while non-linearity may better reflect reality. It does pose challenges in model specification, estimation, inference, and interpretation, however. The choice between linear and non-linear models is ultimately left to researchers from a practical perspective.

The counterfactual experiments conducted in this paper, which involve assessing how the output variable responds to all mediators except the one of interest, are grounded in the 'policy intervention' approach employed by Bernanke et al. (1997), Sims and Zha (2006), Hamilton and Herrera (2004), and Kilian and Lewis (2011). Unlike existing literature, which typically examines the role of specific variables in a single hypothetical scenario, our innovative approach provides a more comprehensive view of causal channels, offering insights beyond a binary 'significant' or 'insignificant' classification. Importantly, our paper provides a quantitative measure of each variable's contribution to the output response over time.

The concept of impulse response decomposition, or the selective suppression of certain variables' contributions at various horizons, shares common ground with the application of Granger causality over multiple horizons. Our approach aligns with the idea proposed by Dufour and Renault (1998), highlighting that Granger causality at one horizon does not necessarily imply causality at higher horizons. It implies that a specific variable could exert a significant impact on the output response in the short term but diminish in influence as time elapses.

Outline: In Section 2, we expound upon our motivation, emphasizing the limitations of impulse response analysis in fully exploring the 'black box' of dynamic causal effects. Section 3 provides a brief overview of the concepts of Granger-type causality and impulse response functions. Section 4 outlines our main mediation analysis philosophy and presents the general econometric model framework. Section 5 delves into the linear SVAR model, conducts impulse response decomposition, and formalizes the causal interpretation for Generalized Impulse Response Functions (GIRs). Section 6 introduces the contribution index that reveals the significance level of each causal channel. Finally, Section 7 provides the conclusion of our research.

2 Motivation

2.1 Impulse responses do not tell everything

Impulse responses provide insights into the causal or total effect of an exogenous intervention, but they do not reveal the complete picture of how this effect is mediated. To illustrate this, we introduce a trivariate VAR(4) process that is deliberately structured with the first five-horizon impulse responses set to zero. A zero impulse response does not signify an absence of the causal effects from the shock; rather, it suggests that the null response is the outcome of indirect effects that counterbalance each other. The Generalized Impulse Response Functions (GIRs), incorporating higher-order coefficients, are the elements that hold this crucial information. Consequently, this illustrative example underscores the significance of delving deeper into Generalized Impulse Response Functions (GIRs) beyond the conventional impulse response functions when researchers aim to explore the intricacies of a causal mechanism.

This example serves to emphasize two critical considerations: First, the absence of a nonzero impulse response does not necessarily imply an absence of causality. Second, impulse responses can be analyzed across multiple time horizons.

$$
\begin{aligned}
y_t = & \begin{bmatrix} 0.6 & 0 & 0.2 \\ 0.2 & 0.6 & 0 \\ -0.2 & 0.4 & 0.7 \end{bmatrix} y_{t-1} + \begin{bmatrix} -0.4 & -0.08 & 0.36 \\ 0 & -0.2 & 0.1 \\ 0.1 & 0 & -0.5 \end{bmatrix} y_{t-2} \\
& + \begin{bmatrix} 0.1 & -0.2 & 0 \\ 0.1 & 0.2 & 0 \\ 0.1 & 0 & -0.2 \end{bmatrix} y_{t-3} + \begin{bmatrix} 0.3 & -0.1 & 0.19 \\ 0 & 0.2 & 0 \\ 0 & 0.05 & 0.15 \end{bmatrix} y_{t-4} \\
& + \begin{bmatrix} 0 & -0.04 & -0.1 \\ 0 & 0.08 & 0.03 \\ 0 & 0 & -0.02 \end{bmatrix} y_{t-5} + \begin{bmatrix} -0.1 & 0.01 & 0.03 \\ -0.08 & 0.03 & 0.06 \\ 0 & 0 & 0 \end{bmatrix} y_{t-6} + u_t.
\end{aligned}
\tag{2.1}
$$

where u_t is a white noise process with identity covariance matrix. This particular setup rules out the potential complexity of contemporaneous causal effects. By employing this trivariate VAR(4) model, our goal is to address the limitations associated with exclusively examining impulse response functions. While a zero impulse response might imply an immediate lack of effect between variables, it does not eliminate the possibility of Granger causality or causality manifesting at longer time horizons. GIRs offer a more comprehensive approach to capture the intricate dynamics and unveil hidden causal mechanisms that may not be evident in standard impulse response analysis.

Table 1

Table 1
A numerical example of impulse responses and GIRs

horizon h	$\pi^{(h)}_{12,1}$	$\pi^{(h)}_{12,2}$	$\pi^{(h)}_{12,3}$	$\pi^{(h)}_{12,4}$	$\pi^{(h)}_{12,5}$	$\pi^{(h)}_{12,6}$
1	0	-0.080	-0.200	-0.100	-0.040	0.010
2	0	-0.248	-0.220	-0.090	-0.014	0.006
3	0	-0.214	-0.074	0.025	0.009	-0.001
4	0	-0.051	0.083	0.065	0.011	-0.003
5	0	0.068	0.027	-0.002	-0.011	0.002
6	-0.062	-0.013	-0.101	-0.076	-0.018	0.005
7	-0.124	-0.104	-0.128	-0.059	-0.008	0.000
8	-0.074	-0.087	-0.044	0.001	-0.002	-0.006
9	0.044	-0.012	0.029	0.027	-0.003	-0.004
10	0.054	0.016	0.026	0.006	-0.003	0.002

Moreover, we scrutinize the impulse response at the fourth time horizon and perform decomposition across various horizons to gain deeper insights into the evolving contributions of variables over time. This decomposition affords a holistic comprehension of the dynamic nature of the investigated causal relationships. GIRs provide valuable insights into the time-varying behavior of these relationships.

2.2 Causal channel and causal chain in macro dynamics

Recent literature has made substantial strides in enhancing our understanding of the impulse response function from diverse angles, including estimation models, identification methods, estimation techniques (both frequentist and Bayesian), linear and nonlinear models, and parametric and nonparametric approaches (Montiel Olea and Plagborg-Møller (2021), Plagborg-Møller and Wolf (2021)). These contributions have yielded valuable tools for identifying and estimating the impulse response function, a central component in mediation analysis, often referred to as the total effect. However, when our objective is to dive deeper into mediation analysis and unravel the intricacies within the dynamic model governing a specific sequence of impulse response functions, how can we proceed?

The exploration of what occurs within a dynamic model for a specific sequence of impulse response functions can be traced back to seminal works like that of Bernanke and Gertler (1995), which examined the causal channel of monetary policy transmission. At the heart of their investigation was a common macroeconomic phenomenon:

$$\text{Monetary policy shock} \xrightarrow{h} \text{Real output,}$$

As corroborated by Friedman and Schwartz (2008), with the advancement of impulse response function analysis over the past two decades, we can conduct an exercise to estimate the dynamic causal effect from GDP to a monetary policy shock. Typically, we observe a U-shaped curve, suggesting that a positive monetary policy shock initially hampers economic growth, but the negative effect eventually dissipates.

However, research into this fundamental economic phenomenon doesn't stop there, as our goal is to comprehend the mechanisms that lie between the input and output. Bernanke and Gertler (1995) offered a narrative explanation for this phenomenon, rooted in widely accepted economic models like the Neoclassical Keynesian framework. This narrative posits that monetary policymakers use their leverage over short-term interest rates to influence the cost of capital, and consequently, spending on durable goods, such as fixed investment, housing, inventories, and consumer durables. These changes in aggregate demand, in turn, affect the level of production:

$$\text{Monetary policy} \rightarrow \text{Cost of capital} \rightarrow \text{Durable goods} \rightarrow \text{Real output.}$$

While this narrative is built upon established economic models, it has three notable limitations:

1. These economic models are constructed within a simplified economic world, making them inadequate for interpreting certain economic phenomena.
2. The proposed causality may overlook variables that play a significant role in this transmission, such as the credit channel.
3. The causality implies that variables participate in this transmission with varying timing and extent as the shock propagates, but it lacks quantitative measures.

This described phenomenon holds broad relevance for many dynamic models, especially in macroeconomics. For instance, the impact of oil price shocks on output has been a subject of extensive research, including studies by Bernanke et al. (1997), Kilian and Lewis (2011), Kilian and Vigfusson (2013), and Hamilton (2011).

3 Econometrics framework and causal literature review

3.1 Short-run and long-run Granger causality

The concepts of Granger causality Granger (1969) is usually interpreted as predictability from one variable to another given potentially some auxiliary variable. The original work proposed by Granger (1969) was to deal with the time series data under a bi-variate framework and horizon one. The criteria of Granger causality back then depends on conditional mean. Later on, Granger (1980) and Granger and Newbold (1986) 'upgraded' the mean condition to conditional distribution. Moreover, the horizon one causality was extended to multiple-horizon causality under the presence of auxiliary variables; see Lütkepohl (1993) and Dufour and Renault (1998). The extension reveals an interesting as well as practical meaningful phenomenon called 'indirect effect' that two variables could be Granger non-causal at horizon one (short run) but causal at multiple horizon (long run) accomplished through an auxiliary variable. In general, the Granger non-causality from variable x to variable y at horizon h with the presence of variable z, $x \overset{h}{\nrightarrow} y|z$ can be written as

$$F_{\text{full}}(y_{t+h}|I_{xyz}(t)) = F_{\text{sub}}(y_{t+h}|I_{yz}(t)), \tag{3.1}$$

where $F_{\text{full}}(y_{t+h}|I_{xyz}(t))$ denotes the conditional distribution of variable y_{t+h} given a 'full' information set (σ-algebra) of variable x, y, z up to time t; and $F(y_{t+h}|I_{yz}(t))$ denotes that of y_{t+h} on a 'sub' information set containing only variable y and z, $I_{yz}(t)$. Note that we explicitly point out the distribution is upon sub-sample (with only variable y, z) when the variable x is removed from the information set. Under the null hypothesis of Granger non-causality, it makes no difference because the information in variable x is redundant. However, if variable x does help predict variable y at h period ahead, from the perspecitive of sensitivity analysis (e.g., Dorn and Guo (2022) and reference therein) and causality measure, it does matter the distribution is on sub-sample or full-sample, that is, $F_{\text{sub}}(y_{t+h}|I_{yz}(t)) \neq F_{\text{full}}(y_{t+h}|I_{yz}(t))$.

As illustrated, if the condition of (3.1) does not hold that variable x contains some particular information to improve the forecasibility of variable y at h period ahead, it is naturally to raise a question that how much variable x could improve the forecasting. Following the work of Geweke (1984), Dufour and Taamouti (2010) proposed an index called 'mean-square causality measure'. In a simple case that variable y is a scalar, it is defined as

$$C_L(x \overset{h}{\to} y \mid z) := \ln \left[\frac{\sigma^2(y_{t+h} \mid I_{yz}(t))}{\sigma^2(y_{t+h} \mid I_{xyz}(t))} \right]. \tag{3.2}$$

where $\sigma^2(y_{t+h} \mid I(t))$ is denoted as the variance of forecast residual of y_{t+h} on the information set $I(t)$. In this formula to compute $\sigma^2(y_{t+h} \mid I_{yz}(t))$, it conducts a linear projection of y_{t+h} onto the space $I_{yz}(t)$. With linearity assumption, this implicitly consider a sub-sample distribution, that is

$$P_L(y_{t+h} \mid I_{yz}(t)) = \int y_{t+h} dF_{\text{sub}}(y_{t+h}|I_{yz}(t)). \tag{3.3}$$

The issue of using sub-sample distribution is raised when variable x is highly correlated with some other variable in the information set, it induces that removal x does not affect too much the mean-square of the residual before and after the removal. That could probability has some misleading results, at least from the economic interpretation perspective, as usually the common critics toward Granger causality.

3.2 Impulse response function: treatment effect in time series

The impulse response function, following the seminal work of Sims (1980), is based on the foundation that a vector of multivariate variables $y_t := (x_t', y_t', z_t')'$ follows a VAR model, and an implicitly restricted assumption, albeit widely accepted, that the VAR residual spans the same space as the structural shocks e_t, that is

$$\Theta_0 e_t = u_t, \tag{3.4}$$

where u_t is the VAR projection residual $u_t = P(y_t \mid I_w(t-1))$ defined on some probability space $(\Omega, \mathbb{P}, \mathcal{F})$, e_t is a random variable with diagonal covariance matrix, and Θ_0 is a non-singular matrix with ones on the main diagonal. Because of the orthogonalization procedure of the residual, the residual e_t is usually called as exogenous shocks with the interpretation of driving the dynamic system evolution. Generally, the orthogonalization is processed under the help of economic interpretation and mild and possibly fragile assumption toward statistical identification of the rotation matrix Θ_0; see Kilian and Lütkepohl (2017). After all, the impulse response function, which is the Wold decomposition coefficients with the orthogonalized 'shocks', is defined with the so-called causal interpretation. More precisely, it is analogous to treatment effect that reveals the difference for the outcome between situation with and without an exogenous shock. It induces the generic definition for impulse response function,

$$\mathrm{IRF}(t,h) := \mathbb{E}[y_{t+h} \mid e_{1,t} = 1, I_w(t-1)] - \mathbb{E}[y_{t+h} \mid e_{1,t} = 0, I_w(t-1)], \tag{3.5}$$

where the shock of interest is defined with no loss of generosity to be the first one in the vector of e_t. Under the linear VAR framework, the impulse response is independent with the state of economy, $I_w(t-1)$, the size/sign of the shock, or the future path of the economy. Nevertheless, the intrinsic spirit of impulse response is with no doubt compare the potential outcome between two scenario, one with the case of shock and one without. Bojinov and Shephard (2019) and Rambachan and Shephard (2021) extends the definition of impulse response to non-parametric framework and propose a definition for potential outcome time series, which close the bond with the potential outcome and causal inference rooted in Neyman (1923) and Rubin (1974). Under linear VAR and stationarity assumption (or a slightly more general assumption: linear (non)-invertible SVMA, as adapted and discussed in the literature among Stock and Watson (2018), Plagborg-Møller and Wolf (2022)), the impulse response function can be equivalently defined as the difference between two potential outcomes:

$$\mathrm{IRF}(t,h) = y_{t+h}^{(1)} - y_{t+h}^{(0)}, \tag{3.6}$$

where $y_{t+h}^{(1)}$ is defined as the potential outcome with the shock (treatment) and $y_{t+h}^{(0)}$ is the one without the shock, both are determined with the same sequence of shocks (or called assignment process) but the one of interest, $y_{t+h}^{(1)} = y_{t+h}(e_{t+h:t+1}, e_{1,t} = 1, e_{2:k,t}, e_{t-1:-\infty})$ and $y_{t+h}^{(0)} = y_{t+h}(e_{t+h:t+1}, e_{1,t} = 0, e_{2:k,t}, e_{t-1:-\infty})$. Because of the linearity and stationarity of the VAR process, and contemporaneous and serial correlated feature of the shock e_t,

$$y_{t+h}^{(1)} - y_{t+h}^{(0)} = \mathbb{E}[y_{t+h}^{(1)} - y_{t+h}^{(0)}] = \mathbb{E}[y_{t+h}^{(1)} - y_{t+h}^{(0)} \mid \mathcal{F}_{t-1}], \tag{3.7}$$

where the two other effect is named by Rambachan and Shephard (2021) as average treatment effect and filtered treatment effect, respectively. Nevertheless, if the impulse response is path-dependent, that the path of future shock (assignment) will affect the causal effect of the shock of interest, or the causal effect is dependent on the state of economy ($\mathcal{F}_{t-1}$), then generally the equality above does not hold. (see nonlinear VAR or Local Projection in Kilian and Lütkepohl (2017)).

4 Analyzing causal channel in macroeconomic dynamics

In macroeconomics, the elucidation of cause-and-effect relationships often takes the form of narrative causal chains found in numerous textbooks. These causal chains lay out a sequence of events, maintaining a chronological order, with the primary objective of revealing the underlying mechanisms that account for the observed performance of a specific outcome variable in response to external interventions.

For instance, when faced with an unexpected positive monetary policy shock, such as a sudden 50 basis point increase in short-term interest rates, it is a reasonable expectation that GDP might exhibit a decline, perhaps occurring in six months or more down the line. Textbooks expound upon this phenomenon within a causal chain, attributing it to the increase in the cost of capital resulting from higher-than-expected inflation. This rise in the cost of capital, according to the narrative, triggers a decrease in aggregate demand, ultimately resulting in a reduction in overall production. Notably, this narrative implies a sequential progression of events and variable performance within the causal chain, underscoring the specific role played by each variable in this chain of causation.

Moreover, this causal chain implicitly introduces a counterfactual notion: if an external intervention, whether it be government policies or actions taken by an economic authority, deliberately interferes with one of the causal channels, it can exert a significant impact on the ultimate outcome.

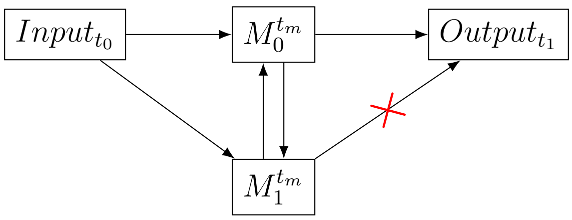


Figure 1: The counterfactual transmission with no contribution of M_1

In light of this backdrop, we introduce a structured framework for measuring causal channels:

1. **The Focal Causal Effect**: Originates from a one-time shock intervention at time t_0 and extends to the output variable observed at time t_1.
2. **The Evaluation Time**: A specific causal channel's measurement occurs at time t_m, $t_0 \leq t_m \leq t_1$. This designated moment is named the "time of evaluation."
3. **The Contribution Assessment**: To ascertain the contribution of a causal channel to the output at the time of evaluation, a counterfactual analysis is employed. This analysis considers the output variable's response to the movements of variables up to time t_m, excluding the specific variable representing the channel of interest.

This structured framework suggests that the response of future output to a current exogenous intervention is mediated through various components within a dynamic system. To measure this causal pathway, we adhere to the fundamental principles of causal philosophy commonly employed in social science research. Here, investigators isolate the influence of a specific variable on the output variable, hypothetically keeping all other factors constant. As our primary focus revolves around the impact of a specific causal pathway, which is represented by the changes in a particular variable in the presence of an exogenous shock, we conduct a "thought experiment" to eliminate the effects of the variable or mediator of interest and evaluate the resulting consequences.

Significantly, we grant the evaluation time, denoted as t_m, the flexibility to span any period between the intervention time t_0 and the outcome time t_1. This flexibility implies that the influence of a specific channel can exhibit varying effects on the final outcome as the shock propagates. For example, in the context of a monetary policy shock, the role of the cost of capital exerts significant influence in the early time window. However, it subsequently transfers its impact to the durable goods variable, ultimately leading to a decline in output. This dynamic role suggests that the significance of each mediator can change as the shock is transmitted over time. Moreover, it provides a quantitative measure for each mediator at different time points, ultimately enabling a quantitative approach to depict the causality chain. This quantitative perspective complements the traditional narrative explanations often found in textbooks regarding causality.

5 SVAR and impulse respons decomposition

In this section, we consider a K-variate structural vector autoregressive model

$$A(L)y_t = \epsilon_t \tag{5.1}$$

where y_t denotes a $K \times 1$ random variable presumed to have zero mean and be detrended for illustrative purposes. The matrix $A(L)$ is defined as $A(L) = A_0 - \sum_{i=1}^{\infty} A_i L^i$, where L signifies the lag operator, and A_0 is a non-singular constant matrix with ones on the main diagonal, such that $a_{ii,0} = 1$ for all i. The vector ϵ_t comprises mutually uncorrelated structural shocks, with a normalized covariance matrix given by $\epsilon_t \sim (0, I_K)$. Importantly, the matrix A_0 cannot be uniquely identified based solely on the second moments of variable y_t. In practice, identification relies on methods such as recursive identification, sign restrictions, higher moments, and the introduction of external instrumental variables.

From this structural model, we derive a reduced-form VAR, as expressed in the form:

$$\pi(L)y_t = A_0^{-1}\epsilon_t \tag{5.2}$$

where $\pi(L)$ is defined as $\pi(L) = A_0^{-1}A(L)$. Subsequently, we generate a multi-horizon VAR, a method frequently employed for examining Granger-causality over multiple periods, as introduced by Dufour and Renault (1998):

$$y_{t+h} = \pi_1^{(h)} y_t + \pi_2^{(h)} y_{t-1} + \cdots + e_{t,h} \tag{5.3}$$

The explanatory variables in this regression encompass observables up to time t, representing the information set at the current period. We address two scenarios, one where the horizon h is a positive integer and the other where it is zero. In the case of a non-zero horizon ($h > 0$), the model is in a reduced-form and can be identified naturally through the second moments of observables. However, when the horizon is zero, the model is in structural form and requires additional information for identification. This distinction in horizon impacts the interpretation, and we consider both cases in this paper, with particular focus on the zero horizon scenario, which involves contemporaneous causal effects.

Granger-type causality analysis revolves around examining the causality or predictability from one variable to another, considering auxiliary variables. The assessment of causality relies on multiple coefficient matrices within Equation (5.3). The economic interpretation is centered on understanding how predictability is enhanced or impaired by adding or removing specific variables from the information set. This interpretation aligns more closely with forecasting than treatment effects. Notably, we find that Granger causality can serve as an interpretation for causal channels.

In cases where the horizon h is greater than zero, the formula for each coefficient matrix in Equation (5.3) can be derived using methods outlined in (3.17) of Dufour and Renault (1998). These coefficients result from non-linear transformations of the underlying VAR parameters, and the residuals $e_{t,h}$ take on the form of a moving average (MA) process of order $h-1$. We define a polynomial function as follows:

$$\pi_i(L) = I + \sum_{h=1}^{\infty} \pi_i^{(h)} L^{h^*}, \text{ for all } i \in \mathbb{Z}^+ \tag{5.4}$$

where $h^* = h + i - 1$. In numerous instances found in the literature, the structural impulse response function is defined when $i = 1$, yielding $\Theta(L) = \pi_1(L)\Theta_0$. Here, Θ_0 is the structural matrix identified after scaling the matrix A_0^{-1}, with the scaling reflecting the standard error of the structural shocks. We also explore scenarios when $i > 1$, which pertain to higher-order impulse response coefficients. These coefficients encapsulate information regarding Granger-type causality between two variables at a given horizon, considering a set of control variables. The presence of non-causality implies that the removal of a particular variable does not hinder the predictability of another variable, provided that the same control variables are retained. This concept resonates with the ideas put forward by Diebold and Yılmaz (2014), who examined the decomposition of forecast errors using Pesaran-Shin-type impulse responses and established relationships between variables over multiple periods. In this paper, we bridge the gap between these higher-order coefficients and causal/treatment interpretations, presenting potential applications for these coefficients in the context of a more general type of intervention. We employ these coefficients to quantify the portion of the impulse response attributed to a specific variable at a given horizon.

It is worth noting that the inverse of the matrix A_0 induces the structural impulse response at horizon zero, with parameters identified based on the scaling of the standard error of the structural shocks. Although we impose certain restrictions on A_0, including it being invertible with ones on the main diagonal, the inverse of matrix A_0 typically does not feature ones on the main diagonal, unless A_0 possesses a specific structure, such as a lower triangular matrix. To facilitate the interpretation that a unit-sized shock has a one-unit impact on the corresponding contemporaneous variable, we normalize the structural matrix by constraining the elements on the main diagonal to be one, subsequently scaling the standard error of the structural shocks accordingly. Thus, we have:

$$\Theta_0 = A_0^{-1} \Sigma_e^{-1/2} \tag{5.5}$$

In this context, the structural matrix $\Theta_0 = \partial y_t / \partial e_t'$, and each shock induces a unit impact effect on its corresponding variable ($\partial y_{i,t} / \partial e_{i,t} = 1$). The covariance matrix of the scaled shocks is defined as $\Sigma_e = \text{diag}[\sigma_i^2]_{i=1}^K$, where σ_i represents the i-th element on the main diagonal of matrix A_0^{-1}.

5.1 Intervention interpretation of GIRs

The existing literature on multiple-horizon Vector Autoregression (VAR) models, also known as Local Projections, has primarily concentrated on the first coefficient matrix due to its direct relation to impulse responses when multiplied by a structural vector. There has been limited attention given to the interpretation of individual higher-order Generalized Impulse Responses (GIRs). In this study, we aim to bridge this gap by extending the principles of intervention analysis to the multivariate case, providing an interpretation for all GIRs. Ultimately, we will illustrate the application of GIRs in counterfactual analysis.

We begin by considering an indicator function:

$$p_t^{(t_0)} = \begin{cases} 1 \text{ if } t = t_0 \\ 0 \text{ if } t \neq t_0, \end{cases} \tag{5.6}$$

We introduce a one-time pulse, representing an intervention, at time t:

$$I_{1,t}^{(t_0)}(\omega) = \omega p_t^{(t_0)}, \tag{5.7}$$

where ω is a scalar defined as $\omega = \tilde{y}_t - y_t$. Here, y_t is an observable in reality, and $\tilde{y}_t$ is a hypothetical observable in a counterfactual scenario. Typically, ω is set to be a column of the structural matrix, $\omega = \theta_0$, representing the shock of interest. The causal response to a one-time pulse $I_{1,t}^{(t_0)}$ is defined as follows:

$$y_t^{(1)}(I_{1,t}^{(t_0)}) = y_t^{(0)} + \pi_1(L) I_{1,t}^{(t_0)} \tag{5.8}$$

In the context of the linear VAR framework, where shocks are unpredictable, and observables are influenced by these shocks over time, the time series observables can be likened to potential outcomes in microeconomics, as thoroughly discussed by Rambachan and Shephard (2021). The difference in the expected value of a variable under a unit-size shock versus no shock, known as the impulse response, holds a causal effect interpretation similar to binary treatment models. Importantly, under the linear VAR framework, the shock (intervention) is randomly assigned, and its causal effect (coefficient) is constant and independent of past information. Therefore, the definition of causal response remains consistent, whether or not the past information, specifically the information set at time t-1, is known. This result does not hold under a general nonlinear VAR framework, such as state-dependent VAR and threshold VAR.

Next, we define each higher-order GIR. The second GIR measures the causal effect of a future observable in a hypothetical scenario in which a shock occurs at time $t_0 - 1$ while the observable at time t_0 remains unaffected. In the multiple-horizon VAR framework, it treats y_{t-1} as "super exogenous" to y_t (see Definition 2.9 of Engle et al. (1983)). In the context of impulse response, this is equivalent to the endogenous response of y_t to the shock at time t-1 being turned off. As illustrated by Bernanke et al. (1997) and Sims and Zha (2006), this procedure is equivalent to combining the initial shock with a hypothetical shock just sufficient to offset the response at time t. As a result, the "compound" shock is defined as follows:

$$I_{2,t} = I_{1,t}^{(t_0)} - \psi_1 I_{1,t-1} \tag{5.9}$$

Here, the portion of $\psi_1 I_{1,t-1}$ is used to offset the endogenous response of the variable at time $t_0 + 1$ to the one-period previous intervention. In other words, the observables one period after the shock are arbitrarily set to be orthogonal to the shock:

$$y_{t+1} \perp I_{1,t}^{(t_0)} \tag{5.10}$$

This adjustment induces potential output changes accordingly:

$$y_t^{(1)}(I_{2,t}) = y_t^{(0)} + \pi_1(L) I_{2,t} = y_t^{(0)} + \pi_2(L) I_{1,t}^{(t_0)} \tag{5.11}$$

This means that the causal effect of the compound shock $I_{2,t}$ can be measured equivalently by the polynomial constructed using the second coefficients in the multi-horizon VAR model.

Formally, for any higher-order GIR at horizon h, it assesses impulse responses at various horizon while holding certain subsequent observations fixed.

$$I_{n,t} = I_{1,t}^{(t_0)} - (\psi_1 I_{n-1,t} + \psi_2 I_{n-2,t} + \cdots + \psi_{n-1} I_{1,t-n+1}). \tag{5.12}$$

This process is equivalent to setting the n-period observable at the shock to be orthogonal to the shock:

$$y_{t+1:t+n} \perp I_{1,t}^{(t_0)} \tag{5.13}$$

The consequence of this restriction is equivalent to that caused by a sequence of hypothetical shocks offsetting or neutralizing the causal effect of that particular pulse on the lagged variable. This interpretation aligns with the formula of GIRs. Then, we define "compound" pulses for the GIR at the order (n+1):

$$y_t^{(1)}(I_{n,t}) = y_t^{(0)} + \pi_1(L) I_{n,t} = y_t^{(0)} + \pi_n(L) I_{1,t}^{(t_0)} \tag{5.14}$$

Similar to the interpretation of $I_{2,t}$, the compound shock $I_{n,t}$ is equivalent to assessing the causal effect of the shock at time t-n and treating the observable y_t as super exogenous to $y_{t+1}, \cdots, y_{t+n}$. This describes a hypothetical situation in which variable y_t changes while the subsequent n-period observables remain fixed.

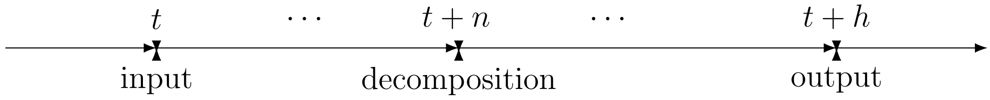
The concept of GIRs presents an alternative perspective for analyzing interventions in multivariate time series data. Researchers have the flexibility to construct counterfactual series of observables for specific hypothetical scenarios and utilize GIRs to forecast potential future paths at different horizons. This versatility of GIRs can be highly valuable in various contexts, such as policy analysis in political economics, stress testing in financial investments, and scenario analysis in risk management.

In the subsequent subsections, we will demonstrate how GIRs can be employed to represent more generalized cases of interventions. Additionally, the decomposition of Equation (5.12) facilitates the use of GIRs to break down an impulse response (total effect) into direct and indirect effects within a dynamic system. This decomposition provides a clearer perspective on how external interventions impact observables over time.

5.2 Impulse response decomposition

The concepts of mediation analysis are widely used in the analysis of treatment effects in microeconometrics or other social sciences. Drawing parallels with macroeconomic dynamic systems, we can consider an external shock as a form of "treatment," while the variables influencing the dynamic system represent the series of observable variables following the shock. In a dynamic system, the total effect of a treatment is evaluated using the standard impulse response, while the direct and indirect effects are quantified using GIRs.

In contrast to causal mediation analysis in static models, the decomposition of the impulse response (total effect) can be performed at different time periods following the occurrence of the shock. This decomposition aims to investigate the contribution of each variable to the impulse response, as highlighted in Kilian and Lewis (2011). However, our approach goes beyond examining the decomposition solely at the time when the impulse response is realized. We demonstrate that this decomposition can be conducted at any period after the shock occurrence, denoted as "n", where "n" is less than or equal to the desired horizon of the impulse response, denoted as "h". To begin, we consider the case of decomposition at horizon zero, where "n"



equals zero. In this scenario, the impulse response represents the direct transmission from time "t" to "t+h" without the involvement of any confounding variables. Consequently, the direct effect is equivalent to the total effect and is denoted as:

$$\theta_h = \mathrm{P}(y_{t+h} \mid y_t^{(1)}(I_{1,t}^{(t_0)}), \mathcal{F}_{t-1}) - \mathrm{P}(y_{t+h} \mid y_t^{(0)}, \mathcal{F}_{t-1}).$$

Under the linear VAR assumption, this can be expressed as:

$$\theta_h = \pi_1^{(h)}\theta_0, \text{ for all } h > 0. \tag{5.15}$$

where θ_0 denotes the contemporaneous effect triggered by the structural shock $e_{1,t}$ and $\pi_1^{(h)}$ is the coefficient matrix attached to y_t in a horizon h autoregressive model. This equality allows for the standard procedure to compute structural impulse responses in local projection literature, as seen in Plagborg-Møller and Wolf (2021), where the structural vector is obtained through various identification approaches such as recursive methods, sign restrictions, and higher moments on the innovations.

Moving on to decomposition at horizon one, $n = 1$, the shock at time t is now transmitted in the system for one period. Therefore, when decomposing the impulse response at time $t+1$, we must account for how the variables at time t and $t+1$ respond to the shock. These response of observables at time t and $t+1$ are quantified by the standard impulse responses θ_1 and θ_0, respectively. Subsequently, we employ the projection method to map these changes to the time period $t+h$, resulting in the equation:

$$\theta_h = \mathrm{P}(y_{t+h} \mid y_{t+1}^{(1)}(I_{1,t}^{(t_0)}), y_t^{(1)}(I_{1,t}^{(t_0)}), \mathcal{F}_{t-1}) - \mathrm{P}(y_{t+h} \mid y_{t:t+1}^{(0)}, \mathcal{F}_{t-1}).$$

This equality represents the concept of 'causal equality.' The left-hand side is the causal effect induced by the independent shocks, while the right-hand side is the causal effect resulting from the potential outcome of the observables at the time of decomposition of period one. Under the linear VAR assumption, it can be expressed as:

$$\theta_h = \pi_1^{(h-1)}\theta_1 + \pi_2^{(h-1)}\theta_0, \text{ for all } h > 1. \tag{5.16}$$

where $\pi_1^{(h-1)}$ and $\pi_2^{(h-1)}$ are coefficient matrices respectively attached to y_{t+1} and y_t in a horizon $h-1$ autoregressive model.

Thus, for any desired impulse response at horizon h, there exist up to $(h-1)$ types of decomposition. Generally, let the decomposition occur at time $t+n$, where $0 \leq n < h$. It yields:

$$\theta_h = \mathrm{P}(y_{t+h} \mid y_{t:t+n}^{(1)}(I_{1,t}^{(t_0)}), \mathcal{F}_{t-1}) - \mathrm{P}(y_{t+h} \mid y_{t:t+n}^{(0)}, \mathcal{F}_{t-1}).$$

We calculate the response of observables from time t to $t+n$ using the standard impulse responses. Next, we employ GIRs at horizon $h-n$ to map the changes in these observables to the time period $t+h$. This leads to the following equation:

$$\theta_h = \pi_1^{(h-n)}\theta_n + \pi_2^{(h-n)}\theta_{n-1} + \cdots + \pi_{n+1}^{(h-n)}\theta_0, \text{ for all } h > n. \tag{5.17}$$

where the coefficient matrices $\pi_i^{(h-n)}$ is the one from horizon $(h-n)$ autoregressive model.This decomposition allows us to assess the contribution of each variable to an impulse response across various time windows. The magnitude of each variable's contribution depends on two crucial factors: (1) whether a variable has experienced the impact of a shock up to a specific time point and (2) the extent of multiple-horizon causality exhibited by that variable toward the target variable. By gaining insights from this decomposition, we can enhance our understanding of how dynamic causal effects are transmitted within the system. Policymakers can also benefit from this analysis by obtaining valuable information regarding optimal timing and policy instruments to mitigate specific impulse responses, such as the adverse impact of increasing inflation on gross output.

Lastly, as implemented by Kilian and Lewis (2011), the decomposition can also occur at time $t+h$. We named it as simultaneous decomposition, since the decomposition rely on the coefficients from a simultaneous equation model (SEM):

$$y_t = \pi_1^{(0)}y_t + \pi_2^{(0)}y_{t-1} + \cdots + \epsilon_t \tag{5.18}$$

where

$$\pi_1^{(0)} = I - A_0, \quad \pi_i^{(0)} = A_{i-1} \text{ for } i \geq 2. \tag{5.19}$$

Notice since the normalization of the matrix A_0 that elements on the main diagonal are all ones, it is readily to check the regressor set does not contain the regresand for each equation. The impulse response therefore can be decomposed as

$$\theta_h = \sum_{i=1}^{h} \pi_i^{(0)}\theta_{h+1-i}. \tag{5.20}$$

In summary, we show that for any impulse response at horizon h, we can decompos it from the time of impact and the time of realization. The decomposition can be summarized in the follow table.

Table 2

Causal mediation analysis - one time intervention
Direct and indirect effect on target variable Y_{t+h} interpreted through multiple-horizon VAR

Decomposition	direct effect	indirect effect					total effect
	Y_t	Y_{t+1}	Y_{t+2}	$\cdots$	Y_{t+h-1}	Y_{t+h}	Y_{t+h}
t	$\pi_1^{(h)}\theta_0$	0	0	0	0	0	θ_h
$t+1$	$\pi_2^{(h-1)}\theta_0$	$\pi_1^{(h-1)}\theta_1$	0	0	0	0	θ_h
$t+2$	$\pi_3^{(h-2)}\theta_0$	$\pi_2^{(h-2)}\theta_1$	$\pi_1^{(h-2)}\theta_2$	0	0	0	θ_h
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t+h-1$	$\pi_h^{(1)}\theta_0$	$\pi_{h-1}^{(1)}\theta_1$	$\pi_{h-2}^{(1)}\theta_2$	$\cdots$	$\pi_1^{(1)}\theta_{h-1}$	0	θ_h
$t+h$	$\pi_h^{(0)}\theta_0$	$\pi_{h-1}^{(0)}\theta_1$	$\pi_{h-2}^{(0)}\theta_2$	$\cdots$	$\pi_2^{(0)}\theta_{h-1}$	$\pi_1^{(0)}\theta_h$	θ_h

Table 2 presents a comprehensive econometric analysis of the impulse response decomposition at specific time points and desired horizons. This decomposition encompasses two essential aspects: the evaluation of individual variable contributions and the assessment of multiple variables' contributions over time. In the time domain, the impulse response propagates through sequential variables. The direct effect captures the causal impact originating from the treated variable at the initial stage, while the indirect effect is determined by subsequent observables, commonly known as post-treatment confounders, spanning from the time of shock occurrence to the time of decomposition.

Notably, it is crucial to emphasize that the coefficients (GIRs) associated to each observable differ across the time interval between the decomposition time point and the desired horizon of the impulse response. These coefficients quantify the specific causal effects on the variable of interest at the designated horizon. Consequently, the contribution of each variable to the targeted impulse response exhibits temporal variations, underscoring the dynamic nature of the causal relationships.

6 Measurment of causal channel

6.1 Variable contribution measure

The preceding discussion has unveiled the potential for decomposing impulse responses concerning the horizon "h" and the time of decomposition "n". This decomposition carries economic significance as it empowers us to assess the individual contribution of each variable to the total effect that materializes in the causal effect at a specific horizon.

It's worth noting that the choice of the time of decomposition has a substantial impact on the evaluation of each variable's contribution. For example, in the initial stages following a monetary shock, it is reasonable to expect that the monetary instrument variable primarily drives the causal effect in the future. This is because, immediately after the shock impacts the economic system, there is a limited time window during which the shock has not yet fully propagated, and individuals (firms or people) have not fully responded to it.

Conversely, with more time having passed since the shock, as often depicted in macroeconomic textbooks, variables like consumption, investment, and durable goods may come into play and contribute to a decline in GDP. Therefore, when the observation time slot occurs later, it is reasonable to anticipate that these variables play a more substantial role in explaining the decline in GDP.

Now, consider the target variable of interest, denoted as the j-th variable at horizon "h" and at the time of decomposition "n". The contribution of the i-th variable to the causal effect is denoted as $c_{i\to j}^{(h,n)}$. The analytical definition is as follows:

$$c_{i\to j}^{(h,n)} = \theta_{j,h} - \theta_{j,h}^{\perp(i,n)} \tag{6.1}$$

where $\theta_{j,h}^{\perp(i,n)}$ represents the causal effect under a counterfactual scenario in which variable i has no contribution at the time of decomposition:

$$\theta_{j,h}^{\perp(i,n)} = P(y_{j,t+h} \mid y_{\backslash i,t:t+n}^{(1)}, y_{i,t:t+n}^{(0)}, \mathcal{F}_{t-1}) - P(y_{j,t+h} \mid y_{t:t+n}^{(0)}, \mathcal{F}_{t-1}) \tag{6.2}$$

The value of $\theta_{h}^{\perp(i,n)}$ is equivalent to computing when variable i is not Granger-causal to variable j at horizon $(h-n)$.

In a linear VAR model, we deduce the contribution of variable i to the impulse response at horizon h:

$$\theta_{j,h} = \sum_{i=1}^{K} c_{i\to j}^{(h,n)}, \text{ for all } 0 \leq n < h, h > 0. \tag{6.3}$$

where $c_{i\to j}^{(h,n)}$ is characterized by the explicit formulation:

$$c_{i\to j}^{(h,n)} = \sum_{l=1}^{h} \pi_{ji,l}^{(h-n)} \theta_{i1,n+1-l}, \tag{6.4}$$

and $\theta_{j1,h} = \psi_{j\bullet,h}\theta_{\bullet 1,0}$.

Our analysis makes use of the dataset introduced in Smets and Wouters (2007), as also employed by Antolin-Diaz et al. (2021). Due to limited access to real wage data, our sample is confined to the period spanning from 1976 Q3 to 2019 Q4, encompassing a total of 172 observations. This dataset encompasses crucial economic indicators, including Gross Domestic Product (GDP), Consumption, Investment, Real Wage, Working Hours, Inflation, and the Federal Funds Rate (FFR). Our analysis centers on a seven-variable Vector Autoregressive (VAR) system, adhering specifically to a VAR(4) model. The collected data is predominantly in levels, with the exception of the Federal Funds Rate. We employ a 4-lag VAR model to ensure robustness and align with the quarterly nature of the data. The choice of lag order is determined using R (command: VARselect, result: AIC(15) HQ(2) SC(2) FPE(2)). The VAR coefficients are estimated through the least squares method.

Our primary investigation is centered around evaluating the impulse response of GDP in response to an inflationary shock, as revealed through a recursive methodology. Subsequently, we dissect this impulse response across various post-intervention time intervals.

The following enumeration delineates the temporal evolution of causal attributions following the intervention. In the immediate period post-intervention, inflation itself emerges as the primary driver, substantiated by its sustained persistence. As the analysis progresses to the second post-intervention period, a distinct shift becomes evident, with the Federal Funds Rate (FFR) gradually shouldering a substantial portion of the causal effect. This transformation underscores the Central Bank's proactive response, characterized by a strategic increase in interest rates to counter heightened inflationary pressures.

In the period immediately preceding the realization of effects, the influence of the intervention becomes notably evident in both GDP and Consumption. Remarkably, inflation no longer maintains the same degree of prominence as a driving force behind the overall impact, while the influence of the Federal Funds Rate (FFR) similarly wanes.

Our empirical findings are consistent with the propositions advanced by Bernanke et al. (1997), where the exogenous response of the FFR to oil price shocks is found to be entirely quelled. Our results make a valuable contribution to the ongoing discourse by emphasizing the crucial role of the FFR as a substantial causal channel that triggers economic downturns in the wake of mounting inflationary pressures. Within the context of oil price shocks, which invariably kindle inflationary pressures, our analysis underscores the consequential ripple effect culminating in economic recession.

6.2 Weighted variable contribution

In the preceding subsection, our analysis has unveiled the capability to disentangle the contributions of individual variables at various stages of decomposition for a given impulse response at a specific horizon. This newfound insight is drawn from the numerical values encapsulated within $c_{i\to j}^{(h,n)}$. The magnitudes of these contributions furnish essential information, shedding light on the evolving roles that each variable assumes over consecutive time periods. The variability in these contributions underscores the intricate causal mechanisms embedded within the model's coefficient parameters.

To delve deeper into this analysis, it is paramount to acknowledge a fundamental tenet: elevated inflation often begets adverse consequences for an economy. This economic interpretation, often articulated in a narrative framework, posits that heightened inflation triggers a responsive intervention by the Central Bank. This intervention, guided by the Bank's commitment to maintaining inflation within a controlled range, aims to avert both excessive inflation that hampers economic activity and insufficient inflation that may usher in deflationary pressures. The commitment of the Central Bank, coupled with its historical responses to elevated inflation, gives rise to the systematic reaction of the Federal Funds Rate (FFR) to high inflation rates. The elevation of the interest rate seeks to mitigate exuberant economic activity emanating from the demand side, ultimately leading to an economic deceleration. This causal chain, tracing from mounting inflation to a subsequent economic slowdown, is widely understood and accepted by both the general public and the academic community. Nevertheless, while this nar-

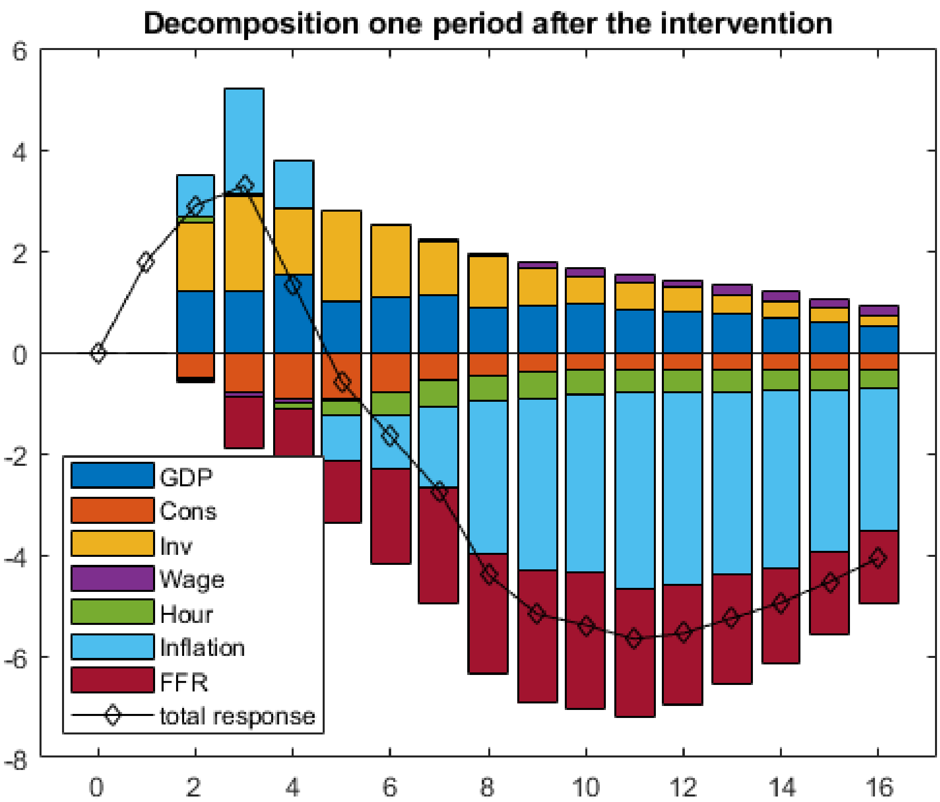


Figure 2: The "total response" signifies the reaction of GDP to an inflation shock. In our analysis, we adopt the recursive identification method, specifically the Cholesky decomposition, where the inflation shock is found to exert a contemporaneous causal effect only on the Federal Funds Rate (FFR), with a magnitude of 0.1373 for a one-unit inflation shock. Our examination focuses on the period one step after the intervention. The presented stacked bar chart illustrates the values of contributions, denoted as $c_{i \to j}^{(h,n)}$, for the time of decomposition where $n = 1$, and the horizons range from 2 to 16. For a view of the decomposition without recursive identification (orthogonality), readers are referred to the Appendix (see Figure A).

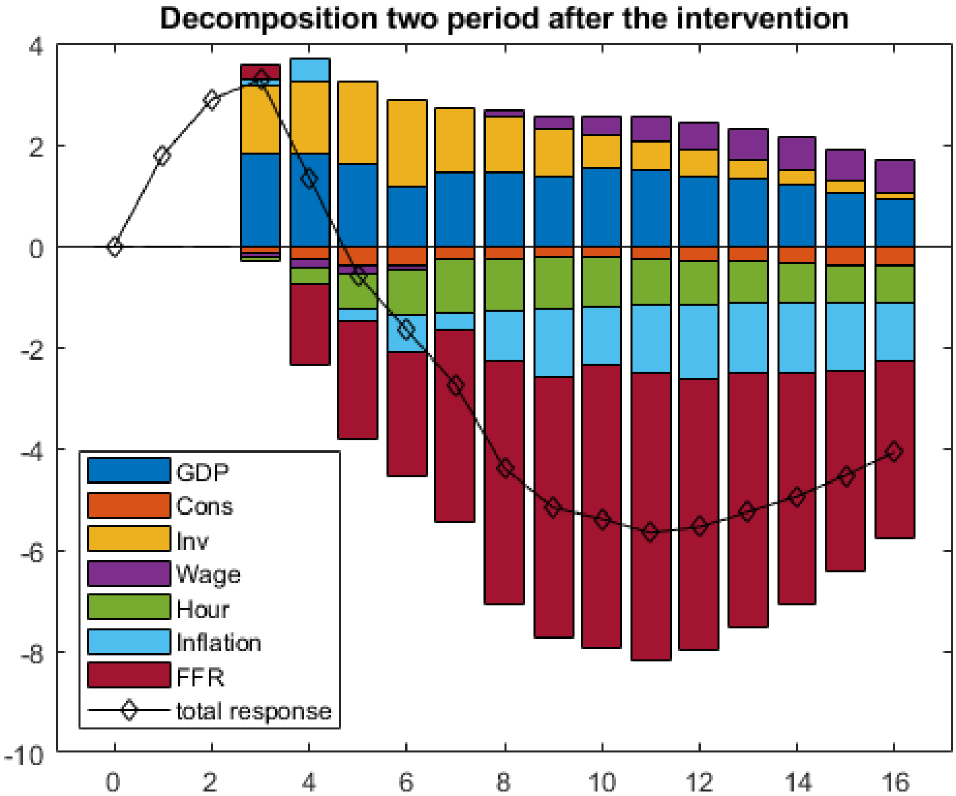


Figure 3: The "total response" characterizes the impact of an inflation shock on GDP. In our analysis, we employ the recursive identification method, specifically Cholesky decomposition, revealing that the inflation shock induces a contemporaneous causal effect solely on the Federal Funds Rate (FFR), with a magnitude of 0.1373 for a one-unit inflation shock. Our examination focuses on the period two steps after the intervention. The presented stacked bar chart illustrates the values of contributions, denoted as $c_{i\to j}^{(h,n)}$, for the time of decomposition when $n = 2$, and the horizons span from 3 to 16. For a view of the decomposition without recursive identification (orthogonality), readers are referred to the Appendix (see Figure A).

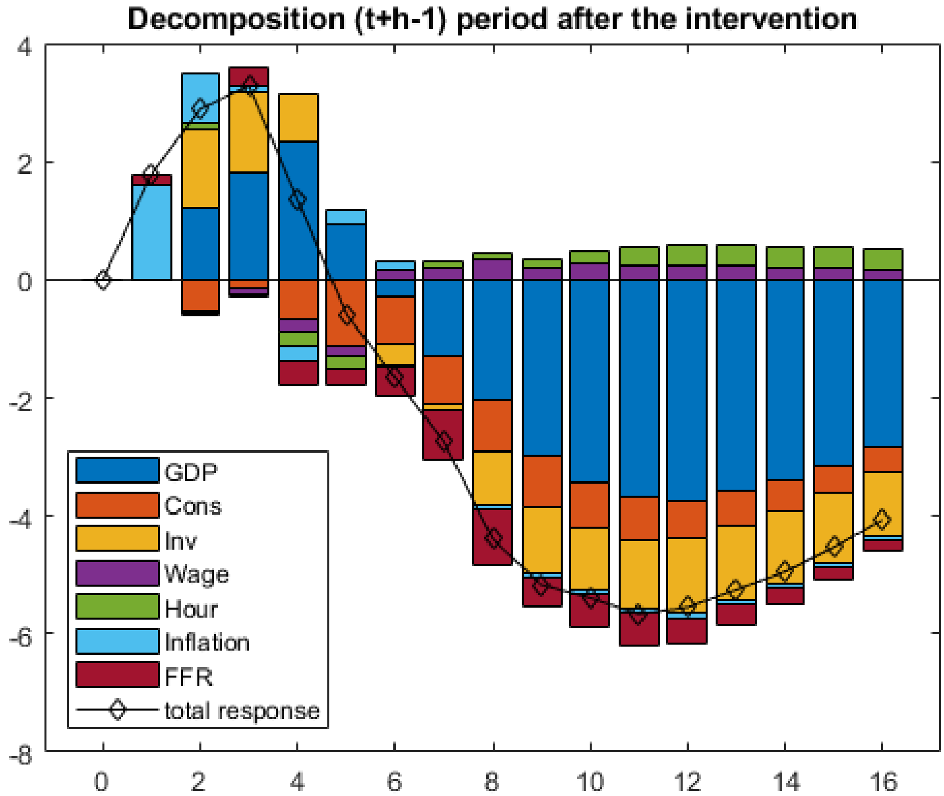


Figure 4: The "total response" signifies the reaction of GDP to an inflation shock. We employ the recursive identification method, specifically Cholesky decomposition, which reveals that the inflation shock elicits a contemporaneous causal effect solely on the Federal Funds Rate (FFR), with a magnitude of 0.1373 for a one-unit inflation shock. In our analysis, we consider the time of decomposition to be $(t + h - 1)$ periods after the intervention. This time point is effectively one period before the impulse response manifests, a concept we refer to as "rolling-decomposition." The stacked bar chart provided illustrates the contributions denoted as $c_{i \to j}^{(h,n)}$ for the time of decomposition when $n = t + h - 1$, spanning horizons from 1 to 16. For a view of the decomposition without recursive identification (orthogonality), readers are referred to the Appendix (see Figure A).

rative is well-grasped, it often lacks precise numerical evidence to substantiate the intricacies of its operational dynamics.

In this context, we introduce a weighted breakdown of each variable's influence at distinct decomposition time points for a specific impulse response function at a given horizon. Since our focus is on a sequence of output: impulse response functions across multiple horizons, the influence of one variable at a particular time of decomposition is determined by an arithmetic average of its contribution to a series of output responses. Specifically:

$$\overline{\omega}_i(h,n) := \frac{1}{H-n} \sum_{h=n+1}^{H} \omega_i(h,n) \tag{6.5}$$

where $\omega i(h,n)$ denotes the proportional contribution of variable i to the impulse response at horizon h given the time of decomposition n. The computation of $\omega_i(h,n)$ bears similarity to the decomposition of major economic indices, such as Gross Domestic Product (GDP), into primary manufacturing, secondary manufacturing, and service sectors. However, it introduces a challenge due to potential variable contributions having both positive and negative signs—some variables may amplify the causal effect, while others may mitigate it. To address this, we propose a weighted computation employing absolute values[1]:

$$\omega_i(H,h) := |c_{i\to j}^{(h,n)}| / \sum_{l=1}^{k} |c_{l\to j}^{(h,n)}| \tag{6.6}$$

When all contributions are positive, this formula aligns with the standard procedure for computing proportional contributions, akin to sector contributions to GDP. The chosen approach to setting weights presents several advantages. First, the arithmetic summation of $\omega_i(H,h)$ equates to one, offering a straightforward interpretation of each variable's contribution. Second, the use of absolute values accommodates situations where variable contributions $c_{i\to j}^{(h,n)}$ exhibit opposite signs. For instance, in a bivariate system, suppose an impulse response can be decomposed into two parts of equal magnitude but opposite sign. Even though the impulse response yields a net zero value, our formula duly provides a 50/50 contribution from each variable, aligning with our causal interpretation. Finally, the arithmetic mean used in (6.5) implies that a sequence of contributions to impulse responses over time periods is treated equally. This may raise concerns from two perspectives: (1) some impulse responses may be of greater interest than others; (2) longer-horizon impulse responses may exert less impact on the present, at least from a utility perspective. However, these concerns can be alleviated by applying a sequence of weights to the formula. The choice of weights can be tailored by researchers based on their primary research objectives, and this article adheres to the arithmetic mean as the chosen approach.

We present an example of the caual effect from GDP to inflation and compute the contribution weights of each variable over peeriods. At time zero, following the Cholesky decomposition, the affected variables are limited to inflation and the Federal Funds Rate (FFR). The proportional contributions of these variables stand at approximately 70% and 30%, respectively. As time progresses, a noticeable trend emerges: the weight of inflation significantly diminishes, while the weights of the FFR and other variables increase. This pattern signifies the gradual propagation of the inflation shock to other variables in the system. The increasing prominence

[1] We set $\omega_i(H,h) = 0$ if $\sum_{l=1}^{k} |c_{l\to j}^{(h,n)}| = 0$, following the convention $0/0 = 0$.

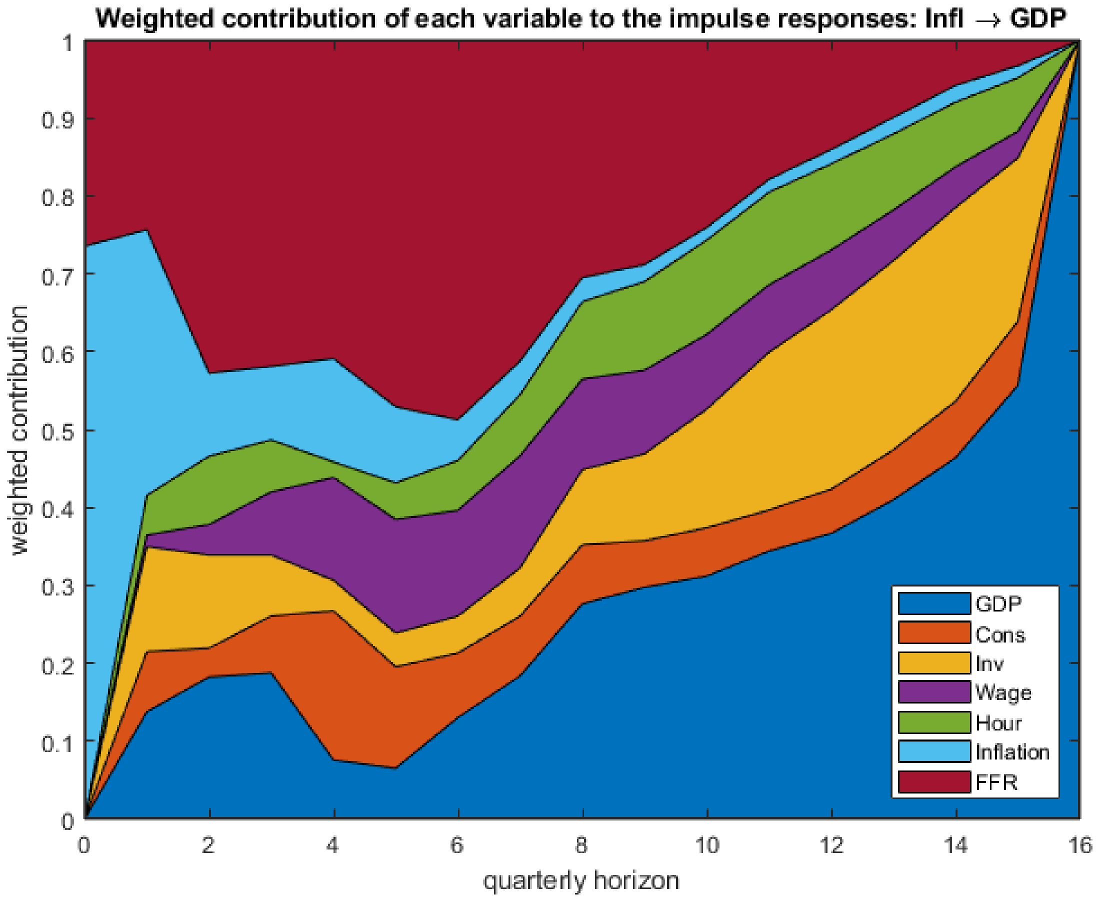


Figure 5: The total effect is the response of GDP to inflation shock at various horizon up to sixteen periods. The weighted contribution is computed through the formula presented above.

of the FFR aligns with the narrative description that attributes the decline in GDP to the endogenous response of interest rates. After a span of six periods, the weights shift from the FFR to other variables, such as investment and wages. Moreover, the increasing weights of GDP itself indicate that the causal effect eventually transmits to GDP, and the decline in GDP persists, affecting itself in subsequent periods. This observation underscores the sustained impact of the inflation shock on GDP and its reverberating effects into the future.

7 Conclusion

This paper presents an innovative approach aimed at elucidating the causal mechanisms governing macroeconomic dynamics. The methodology employed draws inspiration from classic mediation analysis, with a specific focus on counterfactual analysis to suppress responses from particular channels and assess their consequences. The econometric model applied in this study combines Sims' impulse response functions with a Granger-type of causality. It is crucial to underscore that impulse responses provide insights into the total effect of a given output variable in response to an exogenous shock, while Granger-type causality elucidates the causal contributions of each variable or channel across multiple time periods.

Our contribution lies in the proposal of a contribution index, which quantifies the extent to which each variable contributes to the impulse response over time. Additionally, we introduce a weighted contribution index that gauges the influence of each variable on future impulse responses. Through visual representations of these contribution indices, we systematically explore the causal mechanisms at play, providing quantitative insights into the causal channels and data-driven causal chains. It is our aspiration that our method will prove valuable to policymakers and researchers, offering an enhanced understanding of dynamic causality within the realm of macroeconomics.

A Appendix A: variable contribution without orthogonality

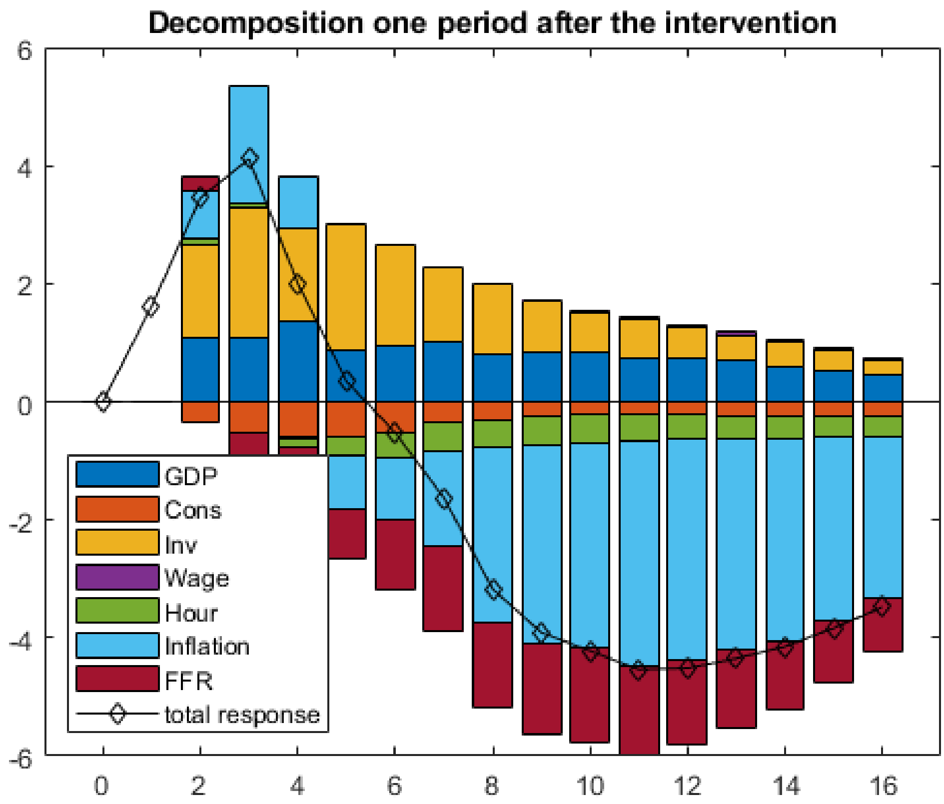


Figure 6: The "total response" is the response of GDP to shock of inflation. Only change the value of variable inflation and remain other contemporaneous variable constant: we are not applying orthogonality of the residual here. The time of decomposition is ONE period after the intervention. The stacked bar represent the value of contribution, $c_{i \to j}^{(h,n)}$, for time of decomposition $n = 1$ and $h = 2, 3, \cdots, 16$.

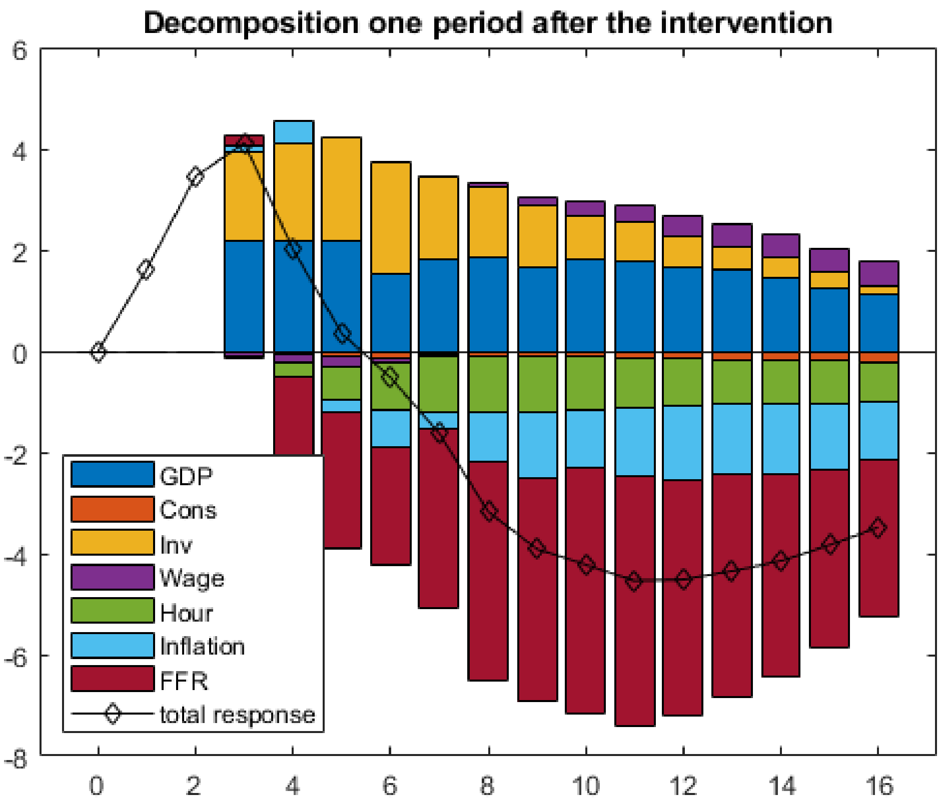


Figure 7: The "total response" is the response of GDP to shock of inflation. Only change the value of variable inflation and remain other contemporaneous variable constant: we are not applying orthogonality of the residual here. The time of decomposition is TWO period after the intervention. The stacked bar represent the value of contribution, $c_{i \to j}^{(h,n)}$, for time of decomposition $n = 2$ and $h = 3, 4, \cdots, 16$.

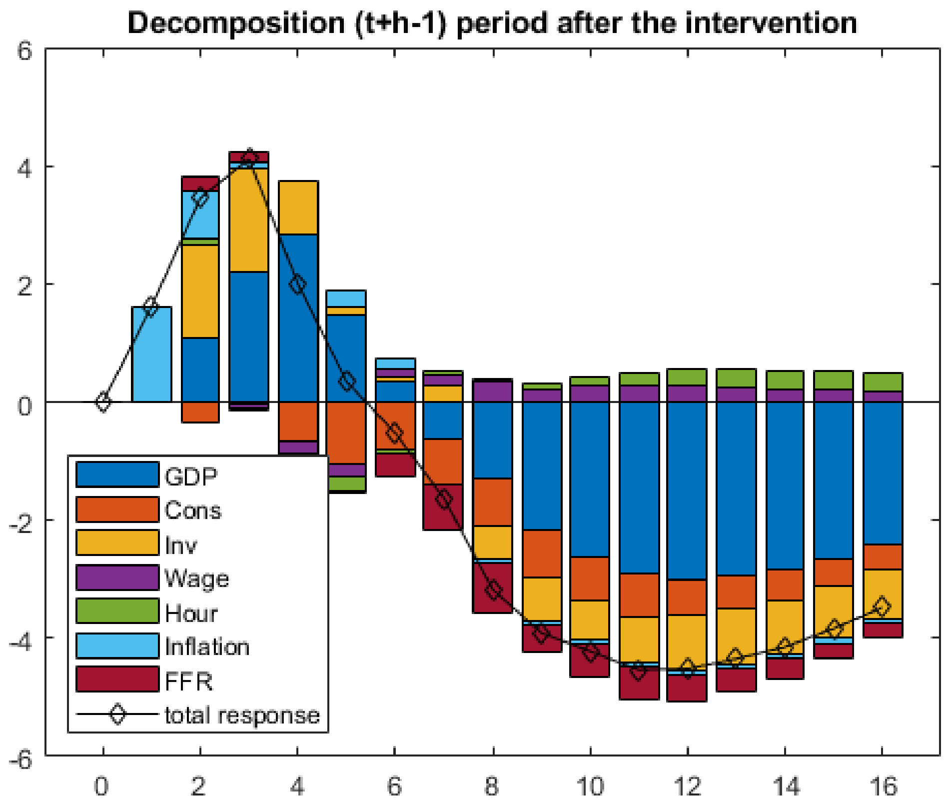


Figure 8: The "total response" is the response of GDP to shock of inflation. Only change the value of variable inflation and remain other contemporaneous variable constant: we are not applying orthogonality of the residual here. The time of decomposition is $(t + h - 1)$ period after the intervention. It is equivalently to One period before the impulse response is realized, so it can always been called as "rolling-decomposition". The stacked bar represent the value of contribution, $c_{i \to j}^{(h,n)}$, for time of decomposition $n = t + h - 1$ and $h = 1, 2, 3, \cdots, 16$.

References

J. Antolin-Diaz, I. Petrella, and J. F. Rubio-Ramirez. Structural scenario analysis with svars. *Journal of Monetary Economics*, 117:798–815, 2021.

R. M. Baron and D. A. Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

B. S. Bernanke and M. Gertler. Inside the black box: the credit channel of monetary policy transmission. *Journal of Economic perspectives*, 9(4):27–48, 1995.

B. S. Bernanke, M. Gertler, M. Watson, C. A. Sims, and B. M. Friedman. Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity*, 1997(1): 91–157, 1997.

I. Bojinov and N. Shephard. Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528):1665–1682, 2019.

G. E. Box and G. C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79, 1975.

F. X. Diebold and K. Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of econometrics*, 182(1):119–134, 2014.

J. Dorn and K. Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, pages 1–13, 2022.

J.-M. Dufour and E. Renault. Short run and long run causality in time series: theory. *Econometrica*, pages 1099–1125, 1998.

J.-M. Dufour and A. Taamouti. Short and long run causality measures: Theory and inference. *Journal of Econometrics*, 154(1):42–58, 2010.

R. F. Engle, D. F. Hendry, and J.-F. Richard. exogeneity. *Econometrica: Journal of the Econometric Society*, pages 277–304, 1983.

M. Friedman and A. J. Schwartz. *A monetary history of the United States, 1867-1960*, volume 9. Princeton University Press, 2008.

J. Geweke. Inference and causality in economic time series models. *Handbook of econometrics*, 2:1101–1144, 1984.

S. Gonçalves, A. M. Herrera, L. Kilian, and E. Pesavento. Impulse response analysis for structural dynamic models with nonlinear regressors. *Journal of Econometrics*, 225(1):107–130, 2021.

C. Granger and P. Newbold. Forecasting economic time series. Technical report, Elsevier, 1986.

C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

C. W. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352, 1980.

J. D. Hamilton. Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic dynamics*, 15(S3):364–378, 2011.

J. D. Hamilton and A. M. Herrera. Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy. *Journal of Money, credit and Banking*, pages 265–286, 2004.

L. Kilian and L. T. Lewis. Does the fed respond to oil price shocks? *The Economic Journal*, 121(555):1047–1072, 2011.

L. Kilian and H. Lütkepohl. *Structural vector autoregressive analysis.* Cambridge University Press, 2017.

L. Kilian and R. J. Vigfusson. Nonlinearities in the oil price–output relationship. *Macroeconomic Dynamics*, 15(S3):337–363, 2011.

L. Kilian and R. J. Vigfusson. Do oil prices help forecast us real gdp? the role of nonlinearities and asymmetries. *Journal of Business & Economic Statistics*, 31(1):78–93, 2013.

R. E. Lucas Jr. Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46. North-Holland, 1976.

H. Lütkepohl. Testing for causation between two variables in higher-dimensional var models. In *Studies in applied econometrics*, pages 75–91. Springer, 1993.

J. L. Montiel Olea and M. Plagborg-Møller. Local projection inference is simpler and more robust than you think. *Econometrica*, 89(4):1789–1823, 2021.

J. L. Montiel Olea, J. H. Stock, M. W. Watson, et al. Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, 225(1):74–87, 2021.

J. Neyman. On the application of probability theory to agricultural experiments. essay on principles. *Ann. Agricultural Sciences*, pages 1–51, 1923.

M. Plagborg-Møller and C. K. Wolf. Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980, 2021.

M. Plagborg-Møller and C. K. Wolf. Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy*, 130(8):2164–2202, 2022.

A. Rambachan and N. Shephard. When do common time series estimands have nonparametric causal meaning. *URL: https://scholar. harvard. edu/shephard/publications/nonparametric-dynamic-causal-modelmacroeconometrics*, 2021.

D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

C. A. Sims. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48, 1980.

C. A. Sims and T. Zha. Does monetary policy generate recessions? *Macroeconomic Dynamics*, 10(2):231–272, 2006.

F. Smets and R. Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American economic review*, 97(3):586–606, 2007.

J. H. Stock and M. W. Watson. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948, 2018.