Measuring Dynamic Transmission using Pass-Through Impulse Response Functions

Giorgi Nikolaishvili*
University of Oregon & Wake Forest University

(Preliminary & Incomplete)
This Version: May 13, 2024
(Click Here for the Most Recent Version)

Abstract

An impulse response function (IRF) captures the net dynamic effect of a shock, but offers no insight regarding the composition of the effect. Instead, the quantification of shock transmission channels is commonly done using structural parameter counterfactual exercises. In this paper, I formulate the concept of a pass-through IRF (PT-IRF) and propose it as a reduced-form alternative to measuring transmission channel dynamics. In essence, a PT-IRF quantifies the propagation of a shock through the Granger causality of a specified set of endogenous media within a dynamical system. I present convenient parametric and semi-parametric methods of conducting inference on PT-IRFs, and discuss their practical advantages over alternative methods of transmission channel quantification. I also demonstrate the empirical flexibility of PT-IRFs by applying them to study the transmission of oil price shocks to inflation and output via monetary policy, and monetary policy shocks to output via credit supply. Finally, I present two example cases in which PT-IRFs have a direct structural interpretation.

JEL Classifications: C10; C32; C50; E52

Keywords: Dynamic transmission; impulse response; vector autoregression

*This paper was formerly circulated as "Pass-Through Impulse Response Functions (PT-IRFs)". I am grateful to Jeremy Piger, David Evans, George Evans, Aeimit Lakdawala, Bruce McGough, Jose Carreno, and participants at the 2023 Meeting of the Midwest Econometrics Group, Fall 2023 Midwest Macroeconomics Meeting, and University of Oregon Macro Group for many helpful comments and suggestions. All errors are my own. E-mail: nikolag@wfu.edu.

1 Introduction

Often in macroeconomics and finance we are interested in studying the dynamic effects of a particular shock on the economy. For this purpose, impulse response functions (IRFs) are used to capture the effects of a disturbance on a set of response variables over some specified time horizon (Ramey, 2016). Empirical estimates of IRFs produced either using vector autoregressions (Stock and Watson, 2016) or local projections (Jordà, 2005) allow for inference on the dynamic effects of economic shocks. However, a key feature of an IRF is that it captures the *net* dynamic effect of a shock – it offers no insight into the quantitative nature of the channels contributing to the net effect of a shock. For instance, an unanticipated change in monetary policy invokes dynamic responses in key macroeconomic variables through the monetary transmission mechanism, which is composed of a variety of channels. If one wants to somehow test the strength or compare the contributions of monetary transmission channels to the net effect of a monetary policy shock, they would be unable to do so using standard IRF analyses.

To address this challenge, Sims and Zha (2006) develop a method of quantifying the contribution of a secondary variable to the dynamic effect of an economic shock on an endogenous response variable. Their approach involves holding the secondary "medium" variable constant over a set horizon by simulating a path of its respective structural shock(s) accordingly in the face of a shock of interest. The most prominent example of an application of the Sims-Zha methodology is presented in Bernanke et al. (1997), in which the authors quantify the contribution of the systematic portion of monetary policy to the net dynamic effect of oil price shocks on key macroeconomic variables. Other applications of variations of the Sims-Zha methodology can be found in Kilian and Lewis (2011) and Bachmann and Sims (2012). A recent refinement of the Sims-Zha approach can be found in McKay and Wolf (2023).

In a similar vein to Sims and Zha, I formulate a new target statistic called the pass-through impulse response function (PT-IRF), which allows for the quantification of transmission channels of a shock within a dynamical system. Given a dynamical system expressed in the form of a VAR, a PT-IRF captures the effect of a structural shock k on an endogenous variable i through some other "medium" variable j. PT-IRFs can be conveniently estimated using the same information and procedures required to estimate

IRFs in the context of VARs, which holds true for inference as well. As explained later in the paper, the PT-IRF approach to measuring transmission channels differs fundamentally from the Sims-Zha methodology, since the former hinges on the Granger causality of the medium to quantify transmission while the latter relies on keeping the medium fixed over time using a sequence of structural shocks. In other words, the two approaches interpret the notion of a contribution to dynamic shock propagation differently. However, in terms of the practical comparison between the two methods, it is worth mentioning that PT-IRFs (1) do not require the identification of multiple structural shocks, while also (2) offering straight-forward ways of conducting inference.

Section 2 of the paper builds up from a simple VAR(1) case to a general formulation of the PT-IRF. Section 3 presents and discusses methods of estimating and conducting inference on PT-IRFs using parametric and semi-parametric methods. Section 4 illustrates the empirical flexibility of PT-IRFs by estimating the pass-through of various macroeconomic shocks via channels examined in the literature using small-scale VARs. Section 5 shows special cases of structural models – specifically, a cobweb model and a standard macroeconomic model – in which PT-IRFs have a clear and intuitive structural interpretation. The final section concludes the paper with a brief discussion of additional potential applications of PT-IRFs.

2 Target Statistic

In this section I define the concept of a PT-IRF starting with the simple case of a linear VAR(1), proceeding to the more general case of a linear VAR(p), and finally generalizing to a stationary Markov process. I also describe how existing methods for estimating VAR IRFs can be used to produce point-estimates and confidence intervals for PT-IRFs.

2.1 Linear VAR(1)

Consider the following VAR(1) process:

$$Y_{t+1} = \alpha + AY_t + B\varepsilon_{t+1}\,, \tag{1}$$

where $Y_t = (y_{1t}, \ldots, y_{Nt})'$ is a vector of N endogenous variables, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ is a vector of K structural shocks, α is an intercept vector, and $A \in \mathbb{R}^{N \times N}$ and $B \in \mathbb{R}^{N \times K}$ are the lag coefficient and contemporaneous impact matrices, respectively. Our goal is to interpret the given linear VAR(1) as a directed weighted graph through which shock impulses travel over time, use this alternative interpretation of a linear VAR(1) to reinterpret the familiar IRF, and finally define the PT-IRF within the given context.

Firstly, notice that the ik-th entry of B represents the contemporaneous effect of the k-th structural shock on the i-th endogenous variable. Refer to Figure 1 for an illustration of the special case of a 3-dimensional VAR(1).

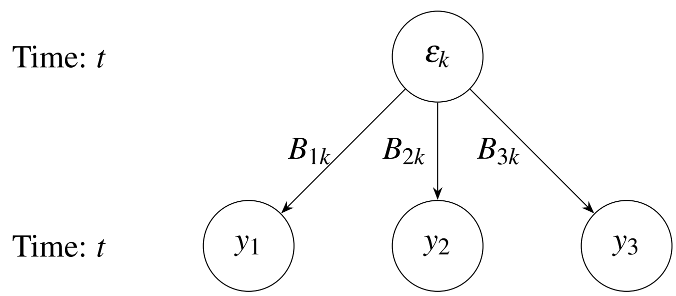

Figure 1: Contemporaneous effects of a structural shock ε_k on a 3-dimensional VAR as a weighted directed graph. ***Note:*** Notice that for each B_{ik}, i indexes the affected variable, while k indexes the shock of origin.

Next, notice that A_{ij} represents the one-period-ahead effect of the j-th variable on the i-th variable. If we think of A as the adjacency matrix in the context of a directed weighted graph, where each endogenous variable at a given point in time is a vertex, then A_{ij} may also be interpreted as the intensity of the travel path of a signal from variable j at time t to variable i at time $t+1$. Once again, for an illustration of the above in the special case of a 3-dimensional VAR(1), refer to Figure 2.

Lastly, we can simply put the above interpretations of B and A together to formulate a VAR(1) as a directed weighted graph, which allows us to trace the propagation of a shock through the system and assess its impact on some variable of interest at a future point in time. Each possible path of a given shock ε_k has a corresponding weight equal to the product of the weights of each of its edges, determined by the contemporaneous impact and lag coefficient matrices. An IRF is simply the sum of the weights of all paths that ultimately reach a destination node corresponding to a variable of interest y_i at a given horizon h. Refer to Figure 3 for an illustration of the one-period-ahead propagation of a shock through a 3-dimensional VAR(1).

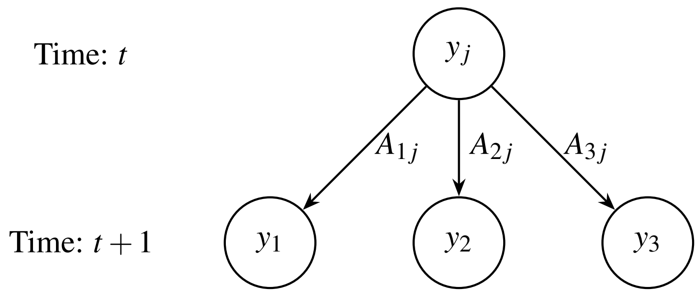

Figure 2: One-period-ahead effects of a change in the variable y_j of a 3-dimensional VAR as a weighted directed graph. ***Note:*** Notice that for each A_{ij}, i indexes the next period's destination variable, while j indexes the variable of origin.

A PT-IRF is the sum of weights associated with the subset of the above-mentioned paths that pass at least once through some medium of interest y_j – if a given path never passes through y_j, then it is irrelevant in gauging the role of y_j as a medium for a shock in the system. For example, the one-period-ahead pass-through response of y_i with respect to y_1 as a medium to some shock ε_k in the case illustrated by Figure 3 is equal to $A_{i1}B_{1k}$ – the weight of the only path that allows for the shock to pass through y_1 at least once before reaching its destination. If we were interested in the union of y_1 and y_2 as a medium for ε_k, then the PT-IR would be $A_{i1}B_{1k} + A_{i2}B_{2k}$ – the sum of the weights of the two paths that allow the shock to pass through either y_1 or y_2 at least once before reaching its destination. The same logic can be extended to h-period-ahead impulse responses, with h being strictly greater than 1.

It can be shown that in the case of a VAR(1), the h-period-ahead impulse response (IR) with respect to some vector of shocks $\overline{\varepsilon}$ may be expressed as

$$\mathrm{IR}(h, \overline{\varepsilon}) = A^h B \overline{\varepsilon}. \tag{2}$$

It can also be shown that for $h > 0$, the corresponding pass-through impulse response (PT-IR) with pass-through/medium variable y_j is algebraically equivalent to

$$\text{PT-IR}(h, j, \overline{\varepsilon}) \equiv \left(A^h - \widetilde{A}^h\right) B \overline{\varepsilon}, \tag{3}$$

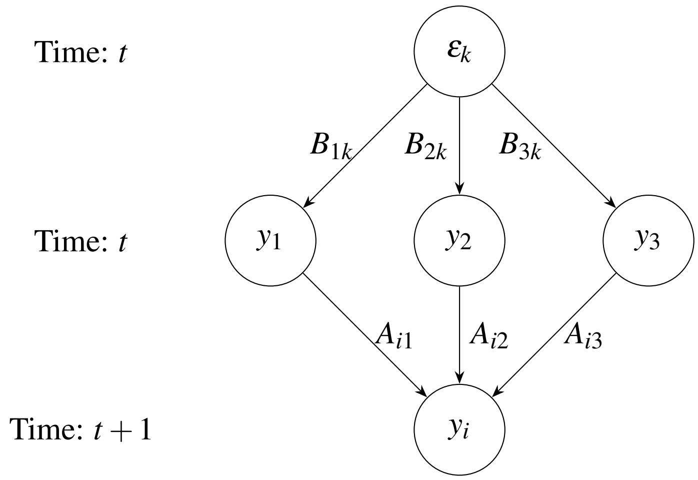

Figure 3: The propagation of an impulse originating at the k-th shock with the i-th variable as its destination, one period ahead. ***Note:*** The one-period-ahead impulse response of y_i with respect to a unit shock to ε_k equals the sum of the weights of all three paths leading to y_{it+1}: $A_{i1}B_{1k} + A_{i2}B_{2k} + A_{i3}B_{3k}$.

where $\widetilde{A}$ is identical to A across all but the j-th column, which is set equal to the zero vector. In the case that $h = 0$, the PT-IR always equals to zero due to the fact that contemporaneous pass-through of any given shock occurs purely through the impact matrix:

$$\text{PT-IR}(0, j, \overline{\varepsilon}) \equiv 0\,. \tag{4}$$

The above equations completely define the PT-IRF in the context of a VAR(1).

2.2 Linear VAR(p)

Consider the following VAR(p) process:

$$Y_t = \alpha + A(L)Y_t + B\varepsilon_t\,, \tag{5}$$

where all familiar objects are defined as before, and $A(L)$ is a lag polynomial of the form

$$A(L) = \sum_{i=1}^{p} A_i L^i\,, \tag{6}$$

such that each A_i is a lag coefficient matrix corresponding to Y_{t-i}. Suppose we aim to derive PT-IR$(h, j, \overline{\varepsilon})$ for this system. The goal is once again to sum the weights associated only with those paths that originate at the shock of interest, pass through y_j at least once over the given horizon, and end at the response variable of interest h periods ahead.

Suppose we represent a linear VAR(p) in state-space form as a VAR(1) with companion matrix Φ and augmented contemporaneous impact matrix $\Gamma = \begin{bmatrix} B' & \mathbf{0} \end{bmatrix}'$. Then for $h \geq 0$ the corresponding PT-IR with pass-through/medium variable y_j may be expressed as

$$\text{PT-IR}(h, j, \overline{\varepsilon}) \equiv \left(\Phi^h - \widetilde{\Phi}^h \right) \Gamma \overline{\varepsilon}\,, \tag{7}$$

where $\widetilde{\Phi}$ is the companion matrix of a modified version of the process described in Eq. (5) with the i-th lag coefficient matrix restricted to equaling

$$\widetilde{A}_i \equiv \begin{bmatrix} \vec{a}_1 & \dots & \vec{a}_{j-1} & \vec{0} & \vec{a}_{j+1} & \dots & \vec{a}_N \end{bmatrix}, \tag{8}$$

where $\vec{a}_m$ denotes the m-th column of A_i. Notice that $\widetilde{\Phi}^h \Gamma \overline{\varepsilon}$ captures the impulse response of a shock for a restricted version of the given linear VAR(p) in which the Granger causality of the j-th endogenous variable is completely removed (Kilian and Lütkepohl, 2017) – all paths passing through the j-th variable are assigned a weight of zero. Therefore, PT-IR$(\cdot)$ sums the weights of only those paths that pass through the j-th variable, which can be interpreted as the impulse response of the system attributable to the Granger-causality of the j-th endogenous variable.

2.3 General Formulation

Let $Y_t = (y_{1t}, \dots, y_{Nt})' \in \mathbb{R}^N$ and $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Kt})' \in \mathbb{R}^K$ such that Y_t is determined by the stationary Markov process

$$Y_t = G(\varepsilon_t, Y_{t-1}; \theta)\,, \tag{9}$$

where $t \in \mathbb{N}^+$ denotes time, $G : \mathbb{R}^K \times \mathbb{R}^N \to \mathbb{R}^N$ is a mapping conditioned on a set of parameters θ, and ε_t is a vector of zero-mean i.i.d. shocks. We may define the h-step impulse response of the given system with respect to some shock vector $\overline{\varepsilon} \in \mathbb{R}^K$ as the following difference between two forecasts $\forall h \in \mathbb{N}$:

$$\mathrm{IR}(t,h,\bar{\varepsilon}) \equiv \mathbb{E}\left[Y_{t+h} \,|\, \varepsilon_t = \bar{\varepsilon},\, Y_{t-1},\, \theta\right] - \mathbb{E}\left[Y_{t+h} \,|\, \varepsilon_t = 0,\, Y_{t-1},\, \theta\right], \tag{10}$$

where the conditional expectation operator $\mathbb{E}[\cdot\,|\,\cdot]$ represents the best mean squared error predictor. The pass-through impulse response of the system with respect to the same shock is once again defined as

$$\text{PT-IR}(t,h,j,\bar{\varepsilon}) \equiv \mathrm{IR}(t,h,\bar{\varepsilon}) - \widetilde{\mathrm{IR}}(t,h,j,\bar{\varepsilon}), \tag{11}$$

where $\widetilde{\mathrm{IR}}$ denotes an object similar to that in Eq. (10), but applied to a transformed version of the process expressed in Eq. (9) in which the Granger causality of the j-th variable in the system is removed. More specifically, we may define

$$\widetilde{\mathrm{IR}}(t,h,j,\bar{\varepsilon}) \equiv \mathbb{E}\left[\widetilde{Y}_{t+h} \,|\, \varepsilon_t = \bar{\varepsilon},\, Y_{t-1},\, \theta\right] - \mathbb{E}\left[\widetilde{Y}_{t+h} \,|\, \varepsilon_t = 0,\, Y_{t-1},\, \theta\right], \tag{12}$$

where $\widetilde{Y}_t \equiv G(\varepsilon_t, \widetilde{I}_j Y_{t-1}; \theta)$, such that $\widetilde{I}_j$ is the identity matrix with the j-th diagonal entry set equal to zero. In other words, $\widetilde{Y}_t$ is generated by the same process as Y_t, but with the influence of the lags of the j-th variable removed from the data generating process (DGP).[1]

Notice that $\widetilde{I}_j$ removes the Granger causality of the j-th variable in the system, thus allowing for $\widetilde{\mathrm{IR}}(\cdot)$ to isolate the impulse response to a shock without accounting for its transmission through y_{jt}. Therefore, subtracting $\widetilde{\mathrm{IR}}(\cdot)$ from $\mathrm{IR}(\cdot)$ yields the impulse response associated with the transmission of a shock through y_{jt}.

3 Estimation and Inference

PT-IRF point estimates, much like those of a VAR-based IRF, can be obtained by estimating a VAR and mapping its parameters to the PT-IRF.[2] A model that explicitly approximates

[1]The PT-IRF does not account for the contribution of the transmission medium to the contemporaneous effect of the shock at a given frequency. This is true due to the fact that Granger causality relates to the lag structure of a reduced-form autoregressive process. Concerns of hidden contemporaneous transmission mechanisms may be addressed in an application by sufficiently increasing the frequency of the data, if possible.

[2]This approach applies to nonlinear VAR PT-IRFs as well. For example, in the case of time-varying parameter (TVP) VARs, PT-IRF mappings can be done at different sample periods of interest.

the dynamic structure and Granger causality of the transmission media, such as a VAR, is necessary to estimate PT-IRFs. Non-parametric or semi-parametric alternatives like local projections (LPs) cannot be used to generate point estimates, unless realizations of the counterfactual DGP in which the Granger causality of the transmission media is shut off can be either directly observed or perfectly inferred using available data.

On the other hand, confidence intervals for a PT-IRF may be obtained using both parametric methods like VARs and semi-parametric ones like LPs. Frequentist VAR-based PT-IRF inference can be done via bootstrapping, such that an accompanying PT-IRF is generated using the synthetically-estimated VAR parameters at each step of the bootstrap.[3] Refer to Algorithm 1 for a standard VAR-based recursive-design PT-IRF bootstrapping procedure based on Runkle (1987). Bias adjustment and wild bootstrap modifications (Gonçalves and Kilian, 2004) for the given procedure to account for possible biases and heteroskedasticity in the errors are straight-forward.

Algorithm 1 Bootstrap Procedure for PT-IRF VAR-based Confidence Intervals

Estimate a VAR and save its parameters and residuals.
Set the initial conditions using the first p observations, where p is the number of lags.
for each bootstrap iteration **do**
    Generate new series by bootstrapping:
    for each replication **do**
        Select a residual at random (from the original series of residuals).
        Use the selected residual and past observations to generate a new observation.
        Repeat to form a complete bootstrapped series.
    end for
    Re-estimate the VAR parameters using the bootstrapped series.
    Compute $\mathrm{IR} - \widetilde{\mathrm{IR}}$ using the bootstrapped VAR parameters and store the result.
end for
Compute the empirical confidence intervals from a sufficiently large sample of stored PT-IRFs.

LP-based PT-IRF confidence intervals may also be bootstrapped using synthetic samples simulated using a VAR fitted to real data. At each iteration of the bootstrap, LPs are used to estimate the net response, as well as the response of the counterfactual

[3]Similarly, Bayesian VAR-based PT-IRF posterior distributions can be obtained by mapping the VAR parameter estimates to the PT-IRF at each step of the sampler. Once again, this is possible since PT-IRFs are nonlinear mappings of VAR parameters.

DGP in which the Granger causality of the transmission media is shut off. Refer to Algorithm 2 for a standard LP-based recursive-design PT-IRF bootstrapping procedure that combines elements of Runkle (1987) and the lag-augmented LP framework of Montiel Olea and Plagborg-Møller (2021). Once again, various modifications to account for bias and deviations from the identical and independent distribution of the errors are easy to make. Analyses of the statistical properties and a comparison of the performance of VAR- and LP-based inference methods for PT-IRFs are left for future research.

Algorithm 2 Bootstrap Procedure for PT-IRF LP-based Confidence Intervals

Estimate a VAR and save its parameters and residuals.
Set the initial conditions using the first p observations, where p is the number of lags.
for each bootstrap iteration **do**
    Generate new series by bootstrapping:
    for each replication **do**
        Select a residual at random (from the original series of residuals).
        Generate a new obs. for both Y and $\widetilde{Y}$ using the selected residual and past obs.
        Repeat to form complete bootstrapped series of Y and $\widetilde{Y}$.
    end for
    Estimate IR using Y and $\widetilde{\text{IR}}$ using $\widetilde{Y}$.
    Compute $\text{IR} - \widetilde{\text{IR}}$ and store the result.
end for
Compute the empirical confidence intervals from a sufficiently large sample of stored PT-IRFs.

4 Illustrative Applications

In this section, I demonstrate the empirical flexibility of the PT-IRF by applying it to study the transmission of oil price shocks to inflation and output via monetary policy, as well as the transmission of monetary policy shocks to output via credit supply.

4.1 Oil Price Shock Transmission to Output via Monetary Policy

Bernanke et al. (1997) use the counterfactual analysis methodology developed by Sims and Zha (2006) to study how the systematic component of monetary policy affects business cycles in response to oil price shocks. They construct a sequence of monetary policy shocks that keep the federal funds rate unchanged in the face of an oil price shock – this is equivalent to shutting down the systematic monetary policy response. They then compare the dynamic responses of key macroeconomic variables under scenarios with and without systematic monetary responses to conclude how much it contributes to the propagation of oil price shocks. The papers finds that a significant portion of the economic downturns following oil price shocks can be attributed to the way monetary policy respond to these shocks rather than the direct impact of the shocks themselves.

Kilian and Lewis (2011) study the same question using a modified counterfactual framework. They propose constructing a sequence of monetary policy shocks that offset the contemporaneous and lagged effects of including the real price of oil in the policy reaction function.[4] The paper argues that this alternative exercise yields a more accurate picture of the contributions of systematic monetary policy to oil price shock propagation. Their results challenge the notion that systematic monetary policy responses to oil price shocks have been a major source of aggregate fluctuations in the U.S. economy.

I replicate the VAR estimated by Kilian and Lewis (2011) to study the transmission of oil price shocks to output and inflation via monetary policy using PT-IRFs. The model consists of the following five endogenous variables, constructed in precisely the same way as in Kilian and Lewis (2011): (1) the percentage change in the real price of imported commodities; (2) the percentage change in the real price of imported crude oil; (3) the Chicago Fed National Activity Index (CFNAI) measure of US real activity; (4) the U.S. CPI inflation rate; and (5) the federal funds rate (FFR). The sample runs from May 1967 until July 1987. All data is obtained from the replication package for Chen (2023), which successfully replicates the results of Kilian and Lewis (2011).

The oil price shock in this VAR is recursively identified as the shock corresponding to the real price of imported crude oil variable. This identification scheme exploits the conventional assumption that oil prices are predetermined with respect to domestic macroeconomic aggregates. Figure 4 shows the net dynamic responses of output and inflation (black lines), matching those of Kilian and Lewis (2011). Furthermore, the same Figure shows the PT-IRFs of output and inflation to oil price shocks via the FFR (dashed

[4] Chen (2023) introduces a direct method to implement the Kilian and Lewis (2011) counterfactual analysis without the need for constructing sequences of hypothetical shocks. This approach utilizes recursive algebra directly in the VAR model framework, making it easier to apply and interpret.

lines) – these capture the transmission of oil price shocks to business cycles via the Granger causality of monetary policy.

The directions of point estimates of both PT-IRFs shown in Figure 4 generally match the directions of their respective net impulse responses at most horizons.[5] This suggests that monetary policy amplifies the effects of oil price shocks on the macroeconomy, as argued by Bernanke et al. (1997) – although, PT-IRFs do not measure transmission in the same manner as the Sims and Zha (2006) counterfactual method. Furthermore, it is worth noting that only some of the inflation response PT-IRF point estimates are statistically significant at the 68% level, whereas none of the output responses are significant. If tested jointly, however, both PT-IRFs may be statistically significant over certain horizons in which point estimates are consistently above or below zero.

4.2 Monetary Transmission to Output via Credit Supply

The credit channel of monetary transmission refers to the mechanism through which changes in monetary policy affect the macroeconomy by influencing the supply and demand for credit. The credit channel consists of the bank lending and balance sheet (sub-)channels (Bernanke and Gertler, 1995). The bank lending channel (BLC) describes the influence of monetary policy on banks' ability to supply loans, which affects macroeconomic activity beyond the direct effects of monetary policy on interest rates (Bernanke and Blinder, 1988). The balance sheet channel (BSC) describes changes in credit demand brought on by changes in creditworthiness and borrowing costs due to monetary policy actions. Like the BLC, the BSC produces indirect effects on macroeconomic activity beyond the direct interest rate channel.

In this illustrative application, I study the credit channel by applying PT-IRFs to the VAR used by Bu et al. (2021). Gilchrist and Zakrajšek (2012) develop the excess bond premium (EBP) index, an increase in which is associated with a contraction in the supply of credit. Bu et al. (2021) include the EBP as an endogenous variable in a monthly VAR, which they use to examine the effect of (externally-identified) monetary policy shocks on key macroeconomic variables. Their justification for including the EBP in the VAR

[5]Theoretically, this does not need to be the case – PT-IRFs can have arbitrarily different shapes relative to their corresponding net dynamic effects.

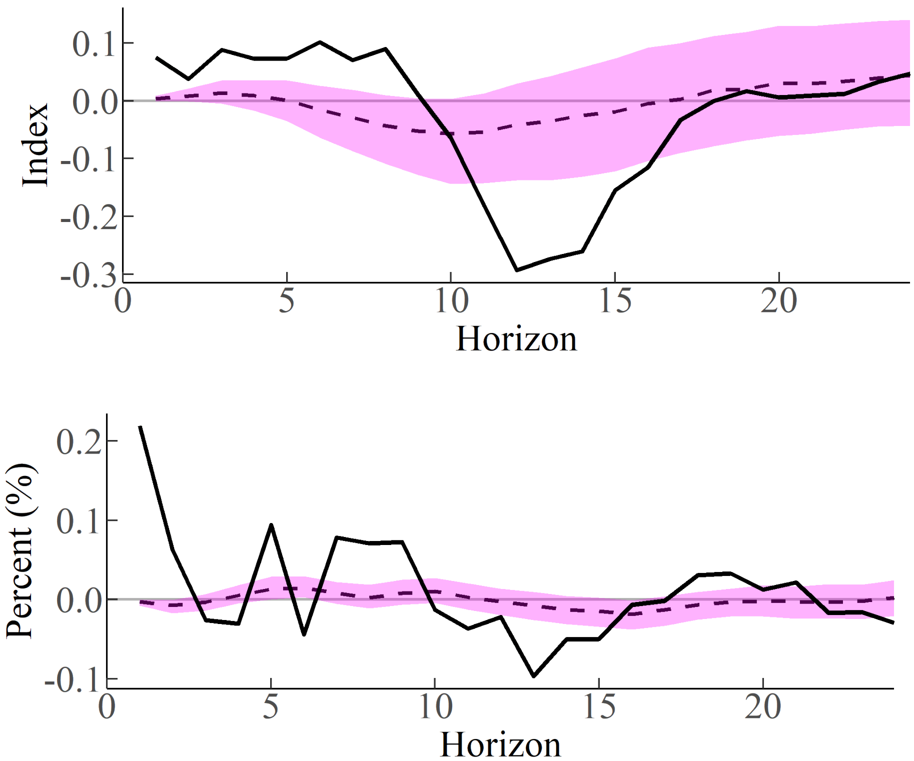

Figure 4: The above plots show the responses of output and inflation to a positive real oil price shock, respectively. The solid black lines represent IRs, while the dashed lines are the PT-IRs of the two response variables with the federal funds rate as the transmission medium. The magenta bands are bootstrapped 68% PT-IRF confidence using 1,000 runs. Contemporaneous responses are omitted, since all PT-IRFs equal to zero instantaneously.

focused largely on its strong predictive power for future macroeconomic activity. I replicate the VAR and monetary policy shock identification scheme used by Bu et al. (2021), and estimate the transmission of unanticipated changes in monetary policy to output via changes in credit supply using the EBP as the proxy medium of transmission.[6]

The model consists of the following five endogenous variables: (1) the cumulated Bu-Rogers-Wu (BRW) externally-identified monetary policy shock series; (2) log-transformed industrial production; (3) log-transformed consumer price index for all items in the U.S; (4) log-transformed producer price index for all commodities; and (5) the EBP. The sample

[6]Nikolaishvili (2024) applies PT-IRFs to study the transmission of monetary policy shocks to output via community versus noncommunity bank lending in a factor-augmented VAR with hierarchical bank lending factors.

runs from January 1994 until December 2019. The monetary policy shock is identified recursively as having a contemporaneous effect on all endogenous variables in the system – in other words, it is the innovation to the BRW series (ordered first in the system).

The resulting PT-IRF point estimates and confidence intervals are presented in Figure 5. Notice that the PT-IRF point estimates have the same sign as the IRF of a monetary tightening on output at all but two periods over a 30-month horizon. Furthermore, notice that a couple of these point estimates are statistically significant at the 68% level when tested separately. Therefore, if the EBP is indeed a good proxy of variation in credit supply, then these findings show aggregate evidence of monetary transmission through the credit channel amplifying the effect of monetary policy on economic activity. This finding is in line with the theoretical foundation of the supply-side credit channel, which hypothesizes that micro-level frictions in lenders' ability to issue credit amplify the effects of monetary shocks.[7]

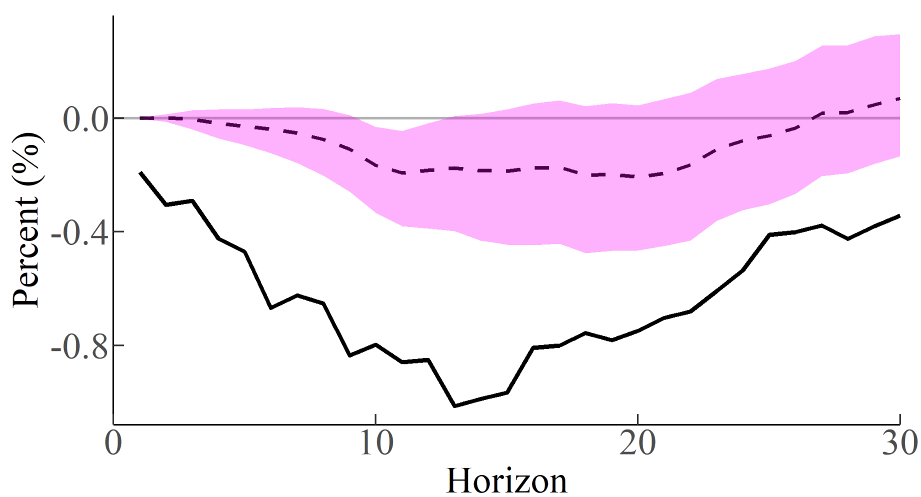

Figure 5: The response of industrial production to an unexpected monetary tightening. The solid black line is the IR, the dashed line is a PT-IR representing the transmission of the monetary shock via the EBP (proxy for credit supply), and the magenta bands are bootstrapped 68% PT-IRF confidence using 1,000 runs. Contemporaneous responses are omitted, since all PT-IRFs equal to zero instantaneously.

[7]See Bernanke et al. (1996) for a seminal treatment of the topic of macrofinancial amplification.

5 Structural Interpretation

The structural interpretation of a PT-IRF depends entirely on the mapping between the primitive parameters of a model and Granger causality of the transmission medium of interest. $\widetilde{IR}(\cdot)$ can be obtained as a standard impulse response after restricting the structural parameters such that the Granger causality of the medium is eliminated from the model dynamics. The PT-IRF can then be obtained by subtracting this counterfactual impulse response from the "true" impulse response to a given shock.

The above-mentioned parameter restrictions can potentially be over-identifying and convoluted – however, in some cases, these restrictions may be easily interpretable. I propose two such cases in the following subsections. The first example is of the one-dimensional cobweb model, in which the contribution of prices to the dynamic transmission of supply and demand shocks can be attributed to future price expectations. The second example is of a small-scale dynamic stochastic general equilibrium (DSGE) model, in which the PT-IRF of all endogenous variables to demand, cost-push, and monetary policy shocks are shown to be determined by the stickiness of interest rates in the Taylor rule.

5.1 Cobweb Model

The cobweb model was used by Muth (1961) to develop rational expectations. Consider a competitive market for a perishable good with linear demand, produced by a continuum of firms. Production is delayed by one period, so that the planned supply of firm i, $s_t^*(i)$, depends on $p_t^e(i)$, its period $t-1$ expectation of p_t. The cost structure is such that supply is linear in $p_t^e(i)$. Demand is expressed as

$$d_t = A - Bp_t + v_{1t}\,, \tag{13}$$

where $B > 0$ and v_{1t} is zero mean *iid* demand shock. The cost of producing for firm i is

$$\text{Cost}(s_t^*(i)) = G_1 s_t^*(i) + \frac{1}{2}G_2(s_t^*(i))^2 - H' w_{t-1} s_t^*(i)\,, \tag{14}$$

where w_{t-1} is a vector of (possibly non-*iid*) observable exogenous shocks and $G_1, G_2 > 0$. The period-t realized quantity supplied by firm i, $s_t(i)$, is

$$\text{i.e. } s_t(i) = s_t^*(i) + v_{2t}\,, \tag{15}$$

where v_{2t} is a zero mean *iid* supply shock that affects all firms.

Firms maximize expected profits with respect to $s_t^*(i)$, so that

$$s_t(i) = K + Cp_t^e(i) + F'w_{t-1} + v_{2t}\,, \tag{16}$$

where $C = G_2^{-1}$, $K = G_2^{-1}G_1$ and $F = G_2^{-1}H$. The expectation $p_t^e(i)$ is formed at the end of period $t-1$ using the information available up to that point in time. In other words, we may express $p_t^e(i) = E_{t-1}^* p_t(i)$, where E_{t-1}^* is based on an ambiguous forecasting rule. Aggregate supply is

$$s_t = \int_0^1 s_t(i)di = K + C\int_0^1 p_t^e(i)di + F'w_{t-1} + v_{2t}\,. \tag{17}$$

Assuming homogeneous expectations across firms, we may express $E_{t-1}^* p_t(i) = E_{t-1}^* p_t = p_t^e$, so that

$$s_t = K + Cp_t^e + F'w_{t-1} + v_{2t}\,. \tag{18}$$

The market clearing condition, $d_t = s_t$, gives the following expression for the temporary equilibrium (TE) of the model:

$$p_t = \mu + \alpha p_t^e + \delta' w_{t-1} + \eta_t\,, \tag{19}$$

where $\mu = B^{-1}(A-K)$, $\alpha = -B^{-1}C$, $\delta = -B^{-1}F$, and $\eta_t = B^{-1}(v_{1t} - v_{2t})$. According to Hommes (2013), economic agents often appear to use simple forecasting rules, such as the "naive" rule: $p_t^e = p_{t-1}$. Substituting this forecasting rule into the TE expression yields the following autoregressive distributed lag (ARDL) model:

$$p_t = \mu + \alpha p_{t-1} + \delta' w_{t-1} + \eta_t\,. \tag{20}$$

In the cobweb model described by Eq. (20), p is the only endogenous variable – for this reason, it is also the only available transmission medium for the structural shocks contained in η. The h-period-ahead effect of a positive demand and supply shock on the price of the perishable good is $\alpha^h B^{-1} v_1$ and $-\alpha^h B^{-1} v_2$, respectively. These dynamic effects are driven entirely by the Granger causality of p, which is determined by the reduced-form parameter $\alpha = -B^{-1}C$.

Setting α equal to zero shuts off the Granger causality of p and yields h-period-ahead responses equal to 0 for both structural shocks across all $h \geq 1$. In other words, the counterfactual in which p does not Granger-cause the ARDL process yields no dynamic effects of demand and supply shocks. Therefore, the PT-IRF of the price of the good with respect to shocks to demand and supply equals the net effect of those shocks. Since the only restriction that allows for $\alpha = 0$ is $C = 0$, and C loads on firms' expectations of the future price of the good, then the PT-IRF implies that the dynamic transmission of exogenous shifts in demand and supply of the perishable good is made possible by firms' expectations.

5.2 DSGE Model

An and Schorfheide (2007) study the estimation of DSGE models using Bayesian methods. They specify a canonical New Keynesian DSGE model that abstracts away from wage rigidities and capital accumulation. The model economy contains a representative household, a single final goods-producing firm, a continuum of intermediate goods producing firms, as well as monetary and fiscal authorities. The equilibrium of their model – linearized at the steady state – consists of the following six equations:

$$y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}\left(r_t - E_t \pi_{t+1} - E_t z_{t+1}\right) \tag{21}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t) \tag{22}$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)\left(\psi_\pi \pi_t + \psi_y (y_t - g_t)\right) + \varepsilon_{rt} \tag{23}$$

$$c_t = y_t - g_t \tag{24}$$

$$z_t = \rho_z z_{t-1} + \varepsilon_{zt} \tag{25}$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{gt}\,. \tag{26}$$

The above are the Euler equation, Phillips curve, Taylor rule, aggregate accounting equality, law of motion for TFP, and law of motion for government spending, respectively. Here y_t, π_t, r_t, c_t, z_t, and g_t denote output, inflation, interest rate, consumption, TFP, and government spending, respectively. All but one coefficient is a model primitive – κ is determined by the following mapping:

$$\kappa = \frac{\tau(1-\nu)}{\nu\phi\overline{\pi}^2}\,, \tag{27}$$

where τ is the constant relative risk aversion (CRRA) parameter, ν is the inverse elasticity of demand, ϕ is the level of price stickiness, and $\overline{\pi}$ is the level of steady-state inflation. The vector of innovations $\varepsilon_t = (\varepsilon_{zt}, \varepsilon_{gt}, \varepsilon_{zt})'$ is mean-zero *iid*.

Morris (2014) simplifies the An-Schorfheide model and solves it analytically to yield a reduced-form VAR. Consider the case in which TFP is *iid*, $z_t = \varepsilon_{zt}$, and government spending is zero at all times. In this restricted case, the model has the following three equations:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\left(\psi_\pi \pi_t + \psi_y(y_t - g_t)\right) + \varepsilon_{rt} \tag{28}$$

$$y_t = E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1} - \varepsilon_{zt}) \tag{29}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - g_t)\,. \tag{30}$$

The solution of this model suggests the following functional form for the interest rate equation:

$$r_t = \phi_{rr} r_{t-1} + d_{rz}\varepsilon_{zt} + d_{rg}\varepsilon_{gt} + d_{rr}\varepsilon_{rt}\,. \tag{31}$$

We may define $\rho_r = \phi_{rr}$ and let $\varepsilon_{rt} = d_{rz}\varepsilon_{zt} + d_{rg}\varepsilon_{gt} + d_{rr}\varepsilon_{rt}$. Adding a cost-push shock, $\varepsilon_{\pi t}$, to the Phillips curve and defining $\varepsilon_{yt} = (1/\tau)\varepsilon_{zt}$ allows for the following expression for the TE of the three-equation model:

$$r_t = \rho_r r_{t-1} + \varepsilon_{rt} \tag{32}$$

$$y_t = E_t y_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1}) + \varepsilon_{yt} \tag{33}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t}\,. \tag{34}$$

By the method of undetermined coefficients, assuming $E_t y_{t+1} = \phi_{yr} r_t$ and $E_t \pi_{t+1} = \phi_{\pi r} r_t$, we obtain the following VAR(1) representation, $Y_t = \Phi Y_{t-1} + DU_t$:

$$\underbrace{\begin{bmatrix} r_t \\ y_t \\ \pi_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \rho_r & 0 & 0 \\ \phi_{yr} & 0 & 0 \\ \phi_{\pi r} & 0 & 0 \end{bmatrix}}_{\Phi} \begin{bmatrix} r_{t-1} \\ y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \frac{\phi_{yr}}{\rho_r} & 1 & 0 \\ \frac{\phi_{\pi r}}{\rho_r} & \kappa & 1 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \varepsilon_{rt} \\ \varepsilon_{yt} \\ \varepsilon_{\pi t} \end{bmatrix}}_{U_t}, \tag{35}$$

where

$$\phi_{yr} = \left(\frac{1 - \rho_r \beta}{\kappa} \right) \phi_{\pi r} \tag{36}$$

$$\phi_{\pi r} = -\left(\frac{1}{\tau} \cdot \frac{\rho_r}{1 - \rho_r} \right) \left(\frac{1 - \rho_r \beta}{\kappa} - \left[\frac{1}{\tau} \cdot \frac{\rho_r}{1 - \rho_r} \right] \right)^{-1}. \tag{37}$$

Notice that y and π do not Granger-cause the reduced-form model presented in Eq. (35) – PT-IRFs that condition on these two endogenous variables as transmission media must equal to zero at all horizons. r is the only variable allowing for the dynamic transmission of shocks via its Granger causality similar to p in the cobweb model – therefore, PT-IRFs that condition on r as the transmission medium equal their corresponding IRFs beyond the initial period.

The Granger causality of r is determined by the structural parameter ρ_r and reduced-form parameters ϕ_{yr} and $\phi_{\pi r}$, both of which are scaled by ρ_r. If the interest rate stickiness in Eq. (23) is removed by setting $\rho_r = 0$, then the Granger-causality of r is also shut off and all dynamic effects in the model disappear. So, the PT-IRF of an endogenous variable in response to a shock with r as the medium can be obtained by estimating the response under the counterfactual $\rho_r = 0$, and subtracting it from the true IRF over a set horizon. Therefore, the dynamic transmission of shocks via the interest rate occurs due to the stickiness of the Taylor rule.[8]

To demonstrate the relationship between the theoretical model and empirics, I calibrate the simplified model – parameter values are presented in Table 1. I compare the empirical

[8] One caveat worth mentioning is that since ρ_r also scales the contemporaneous effects of monetary policy shocks, the instantaneous effect of a monetary shock in the counterfactual model will not match that of the true DGP. Notice that the contemporaneous effect of ε_r on π and y equals to $-\kappa/\tau$ and $-1/\tau$, respectively, given $\rho_r = 0$.

and theoretical PT-IRFs of all three endogenous variables in response to an innovation in government spending (which affects the system through the monetary policy shock ε_r), with the interest rate as the medium of transmission. The theoretical PT-IRFs are generated using the above-mentioned counterfactual exercise involving ρ_r, whereas the empirical PT-IRFs are obtained by estimating a VAR(2) on 1,000 observations simulated using the structural model. Figure 6 presents the results of this exercise.

Parameter	Description	Value
β	Discount factor	0.9975
ρ_r	Taylor rule stickiness	0.75
τ	CRRA	2
κ	Composite parameter	0.33

Table 1: Simplified An-Schorfheide model parameter calibration following that of Morris (2014).

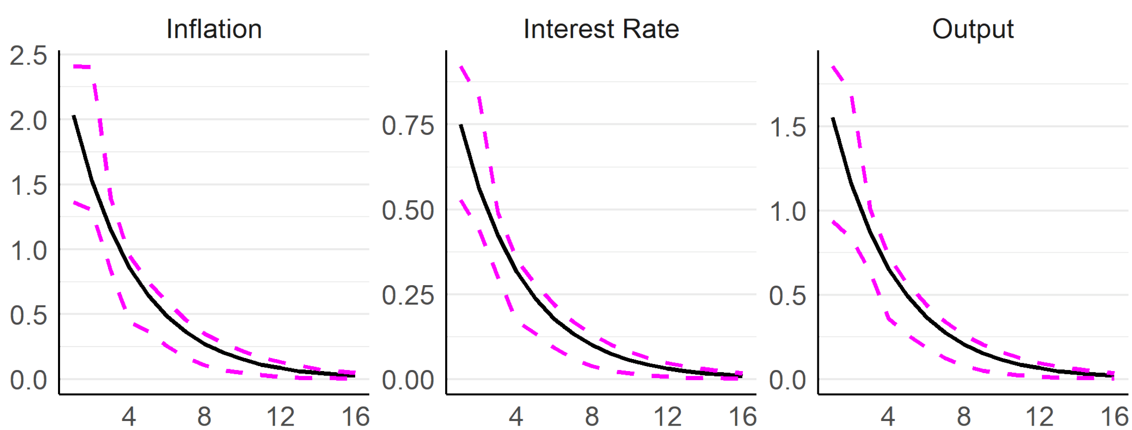

Figure 6: Structural vs. empirical PT-IRFs of all endogenous variables in the simplified An-Schorfheide model in response to a government spending shock, with the interest rate as the transmission medium. The solid black line is the theoretical PT-IRF, produced using the $\rho_r = 0$ counterfactual exercise described earlier in Section 5.2. The dashed magenta lines are the bootstrapped 95% confidence intervals for matching PT-IRFs obtained using a VAR(2) estimated on 1,000 observations generated by the structural model. Contemporaneous responses are omitted, since all PT-IRFs equal to zero instantaneously.

6 Concluding Remarks

PT-IRFs can be a useful tool for quantifying transmission channels of a wide variety of economic shocks. Although my applications in this paper involve linear VARs, PT-IRFs can easily be estimated in arbitrarily nonlinear settings as well – for example, an extension to TVP-VARs would be straight-forward. Furthermore, the ability of PT-IRFs to accommodate multiple transmission media allows for the simultaneous quantification of potentially overlapping contributions of multiple transmission channels.[9] Future studies can explore and compare the properties of VAR vs. LP methods of conducting inference on PT-IRFs – does one tend to be more efficient than the other at certain horizons and/or given certain levels of persistence? Additionally, future work can explore ways of conducting joint inference across PT-IRF point estimates. Finally, PT-IRFs can be used to study dynamic transmission mechanisms both empirically and theoretically in fields outside of macroeconomics and macrofinance.

[9]One instance of this is done by Nikolaishvili (2024) in assessing the contributions of local vs. geographically-diversified banks' lending to the transmission of monetary policy.

References

An, S. and Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, 26(2-4):113–172.

Bachmann, R. and Sims, E. (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics*, 59(3):235–249.

Bernanke, B., Gertler, M., and Gilchrist, S. (1996). The financial accelerator and the flight to quality. *The Review of Economics and Statistics*, 78(1):1–15.

Bernanke, B., Gertler, M., and Watson, M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, 28(1):91–157.

Bernanke, B. S. and Blinder, A. S. (1988). Credit, money, and aggregate demand. *The American Economic Review*, 78(2):435–439.

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

Bu, C., Rogers, J., and Wu, W. (2021). A unified measure of fed monetary policy shocks. *Journal of Monetary Economics*, 118:331–349.

Chen, S.-S. (2023). A direct approach to kilian–lewis style counterfactual analysis in vector autoregression models. *Journal of Applied Econometrics*, 38(7):1068–1076.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120.

Hommes, C. (2013). *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge University Press.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

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

Kilian, L. and Lütkepohl, H. (2017). *Vector Autoregressive Models*, page 19–74. Themes in Modern Econometrics. Cambridge University Press.

McKay, A. and Wolf, C. (2023). What Can Time-Series Regressions Tell Us About Policy Counterfactuals? Staff Report 642, Federal Reserve Bank of Minneapolis.

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

Morris, S. (2014). The Statistical Implications of Common Identifying Restrictions for DSGE Models. 2014 Meeting Papers 738, Society for Economic Dynamics.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335.

Nikolaishvili, G. (2024). Commercial Bank Heterogeneity and the Transmission of Monetary Policy Through Bank Lending. Working Paper.

Ramey, V. (2016). Chapter 2 - macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier.

Runkle, D. E. (1987). Vector autoregressions and reality. *Journal of Business & Economic Statistics*, 5(4):437–442.

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

Stock, J. and Watson, M. (2016). Chapter 8 - dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.