Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy*

Kurt G. Lunsford†
Federal Reserve Bank of Cleveland

October 2016

Abstract

This paper provides a simple estimator to identify one structural shock in a vector autoregression (VAR). This estimator uses a proxy variable that is correlated with the structural shock of interest but uncorrelated with the other structural shocks. When the covariance of the proxy variable and the structural shock of interest is local-to-zero, then the proxy variable is weak and the estimator is inconsistent. To test for the presence of a weak proxy variable, I use the F statistic from the regression of the proxy variable onto the VAR errors, and I provide critical values for this statistic. This F statistic and its critical values are distinct from those in the weak instrumental variables literature.

Keywords: F Statistic, Productivity Shocks, Proxy Variable, Structural Vector Autoregression, Weak IV

JEL Codes: C12, C13, C32, C36, O47

*I am grateful to Todd E. Clark, Joachim Freyberger, Bruce E. Hansen, Lutz Kilian, José Luis Montiel Olea, Tatevik Sekhposyan, Xiaoxia Shi, Kenneth D. West, Sarah Zubairy, seminar participants at Texas A&M University, the University of California – Riverside, the University of Wisconsin – Madison, Vanderbilt University and the Board of Governors of the Federal Reserve, and participants at the 42nd Annual Eastern Economic Association Conference, the 24th Symposium of the Society of Nonlinear Dynamics and Econometrics, the Spring 2016 Federal Reserve Bank of St. Louis Econometrics Workshop, and the International Association of Applied Econometrics 2016 Annual Conference for helpful comments and discussions. All errors are my own. The views expressed in this paper are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†email: kurt.lunsford@clev.frb.org

1 Introduction

Since Sims (1980), identifying structural shocks in vector autoregressions (VARs) has been important for research in macroeconomics. Specifically, consider the $n \times 1$ vector of time series variables, denoted by Y_t, that follows

$$Y_t' = A_0 + Y_{t-1}' A_1 + \cdots + Y_{t-p}' A_p + u_t', \tag{1}$$

where u_t is the $n \times 1$ vector of VAR innovations. Let v_t be the $n \times 1$ vector of structural shocks, which are related to the VAR innovations by

$$u_t = B v_t. \tag{2}$$

Then, the objective for much of the structural VAR literature is to estimate the column of B that corresponds to the structural shock of interest. For ease of exposition, assume that the relevant column of B is the first, B_1.

In this paper, I study the proxy variable approach to estimating B_1, developed by Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013). This approach assumes that there exists a variable that is external from the VAR in Equation (1) that acts as a proxy for the structural shock of interest. Specifically, the proxy variable is correlated with the structural shock of interest and uncorrelated with the other structural shocks.[1] This proxy variable approach can estimate B_1 without placing any restrictions on the values of the elements of B, such as zero or sign restrictions.

Proxy variables can come from a variety of sources and identify many different shocks. Mertens and Ravn (2013) use the narrative approach of Romer and Romer (2009) to construct proxy variables for tax shocks, Gertler and Karadi (2015) use the high frequency approach of Gürkaynak, Sack, and Swanson (2005) to construct proxy variables for monetary policy shocks, and Stock and Watson (2012) use 18 different proxies for identifying shocks to oil, monetary policy, productivity, uncertainty, liquidity and financial risk, and fiscal policy. While having a large number of proxies shows the applicability of the proxy approach, a problem is that many proxies may only be weakly relevant for identification, similar to the problem of weak instrumental variables (IV) (Staiger and Stock, 1997). Stock and Watson (2012) find that only 4 of their 18 proxies yield F statistics greater than 10. In contrast, Gertler and Karadi (2015) find multiple proxies for monetary policy shocks that yield F statistics greater than 10. However, a problem when comparing these two papers is that their F statistics come from different regressions. Stock and Watson (2012) regress their proxy variable on the VAR errors, but Gertler and Karadi (2015) regress one VAR error on their proxy variable. To test for weak proxies, Montiel Olea, Stock, and Watson (2012)

[1] These assumptions are similar to those made in the instrumental variables literature. Because of this, the exogenous variable in this paper might also be referred to as an "external instrument." However, to avoid confusion with the instrumental variables literature, I follow the terminology in Mertens and Ravn (2013) and refer to the exogenous variable as a "proxy" or a "proxy variable."

consider F statistics from both of these regressions and find that they can yield very different values. Given the potential prevalence of weak proxies and the potentially conflicting tests for them, this paper answers the following questions. What are the effects of a weakly relevant proxy? How can researchers test for a weak proxy?

To answer these questions, this paper proceeds in three steps. First, it provides a simple estimator for B_1 when the proxy is not weak. Previous papers that have used proxy variables to estimate structural VARs have relied on tedious matrix algebra to construct their estimators. See, for example, the Appendix of Mertens and Ravn (2013) or Appendix D of Lunsford (2015). In contrast, I show that B_1 can be estimated as a direct function of the VAR errors and the proxy variable. In addition to making the proxy variable approach easier to use, this simple estimator makes the analysis of weak proxy variables tractable.

The second step of this paper characterizes the asymptotic limit of the estimate of B_1 when the the proxy variable is weak. To model a weak proxy, I use a local-to-zero assumption as in Staiger and Stock (1997) and assume that the covariance of the proxy and the structural shock of interest goes to zero at a rate of square root of the sample size. With this assumption, the estimate of B_1 is not consistent; rather, it converges to a distribution. This asymptotic distribution is a function of the signal for the structural shock that the proxy provides relative to the noise in the proxy. As this signal-to-noise ratio increases, the asymptotic distribution collapses to B_1. Thus with a large signal-to-noise ratio, a proxy variable can provide a close estimate of B_1 even when it is weak. To measure how close this estimate is, I follow the weak IV literature and use the asymptotic bias of the estimate of B_1 (Stock, Wright, and Yogo, 2002; Stock and Yogo, 2005). This bias decreases as the signal-to-noise increases, and weak proxy sets, defined as the set of proxy variables with signal-to-noise ratios below a given threshold, can be produced based on a researcher's tolerance for bias. Using simulation, I provide these signal-to-noise thresholds for asymptotic biases of 20%, 10%, 5% and 1%, and for VAR dimension from 2 to 20.[2]

The third step of this paper provides a test for a weak proxy variable based on an F statistic. This F statistic is from the regression of the proxy variable on the VAR errors, providing theoretical support for this regression in Stock and Watson (2012). This contrasts with the weak IV literature, which Gertler and Karadi (2015) use, where the F statistic is from a regression where the instrument or proxy is an independent variable. For this paper's F statistic, n times F converges to a non-central χ_n^2 distribution, where the signal-to-noise ratio is the non-centrality parameter. Thus, if the F statistic is sufficiently large, then the probability that the true signal-to-noise ratio is below the weak proxy set threshold is small and the null hypothesis that the proxy variable is in the weak proxy set can be rejected. I provide critical F values for asymptotic biases of 20%, 10%, 5% and 1%, for VAR dimensions from 2 to 20, and for levels of significance of 0.10, 0.05 and 0.01. As in the weak IV literature, the critical F statistics are large. For conventional VAR dimensions, F statistics between 7 and 9 are needed to reject the null hypothesis that the proxy variable yields greater than

[2] In this paper, I do not address the optimal level of bias tolerance. Rather, throughout this paper I follow Stock, Wright, and Yogo (2002) and use 10% bias when discussing applications.

10% asymptotic bias at a 5% level of significance.

An appealing feature of the F statistic in this paper is that its limiting distribution does not depend on B. In contrast, the limiting distribution of the F statistic from the weak IV literature, which is used in Montiel Olea, Stock, and Watson (2012), Gertler and Karadi (2015) and Lunsford (2015), does depend on B. This is problematic for two reasons. First, the critical values of this F statistic will vary with the elements of B and will not be the same as those from the weak IV literature (Stock and Yogo, 2005). Hence, the weak IV rule of thumb that requires an F statistic to be greater than 10 will lead to tests that are mis-sized. The second problem is that estimates of some of the elements of B are needed to compute the correct critical values. However, when the proxy variable is weak, these estimates will be inconsistent and the estimated critical values will be unreliable. Hence, while the econometric theory in this paper is closely related to the weak IV literature, the results from the weak IV literature cannot be applied to the proxy structural VAR framework, and the theory in this paper is needed to test for weak proxies.

As an application, I study the dynamic effects of productivity shocks, and I use Fernald's (2014) measures of utilization-adjusted total factor productivity (TFP) for the consumption and investment sectors as the basis for my proxy variables. For the F statistic proposed in this paper, I find values greater than 9 for both the consumption TFP and investment TFP proxies, and I reject the null hypothesis that these proxies are in the weak proxy set. In contrast, I show that the F statistic from the weak IV literature yields values from 0 to 66 depending on the VAR ordering, giving ambiguous results about proxy strength.

In addition, this application demonstrates the usefulness of the proxy approach more broadly. To do this, I compare the structural impulse response functions (IRFs) identified with a proxy variable to IRFs from two other identification schemes. The first other scheme treats the proxy as the structural shock itself and identifies the structural IRFs by least squares. The second other scheme puts the proxy in the VAR as an endogenous variable and uses Choleski identification for the IRFs. Because the proxy is a noisy measure of the structural shock of interest, these alternative methods bias the estimates of B_1 toward zero. For the consumption TFP shock, the Choleski estimate's bias of B_1 is severe enough to cause the employment IRF to flip signs. Hence, while variables that contain dummies or missing observations are commonly used as proxies, the proxy approach can be used more generally and is especially useful with noisy measures of structural shocks.[3]

This paper's weak proxy test guards against inconsistency of the point estimate of B_1. This complements related work by Montiel Olea, Stock, and Watson (2016), which provides confidence intervals for IRFs that are robust to weak proxies, in three ways. First, economic meaning is often applied to the point estimates of statistics from structural VARs. For example, Mertens and Ravn (2013, 2014) use IRFs to measure tax multipliers. Hence, in

[3]Similarly, Carriero et al. (2015) argue that the proxy approach has less measurement error bias when estimating the impact of Bloom's (2009) uncertainty shocks compared to putting the shocks directly into the VAR. Further, Mumtaz, Pinter, and Theodoridis (2015) show that the proxy approach better matches the effects of credit supply shocks from a general equilibrium model than a Choleski decomposition.

addition wanting to know that their estimates are statistically significant, researchers may also want to know that their estimates are close to the true parameters and will need to test for proxy weakness. Second and simlarly, point estimates of IRFs are often used to estimate structural parameters in economic models. For example, see Christiano, Eichenbauma, and Evans (2005). Again, this indicates a need to guard against inconsistent estimates by testing for weak proxies. Third, Montiel Olea, Stock, and Watson's (2016) confidence intervals only apply to IRFs. Thus, for other statistics from structural VARs, like forecast error variance decompositions, researchers will need the weak proxy testing proposed in this paper.

The remainder of the paper is as follows. Section 2 describes the VAR and proxy variables, provides the estimator for B_1, and establishes the consistency of this estimator. Section 3 defines a weak proxy variable, shows that the estimator of B_1 is inconsistent with a weak proxy variable, describes the test for a weak proxy variable, and provides the critical F statistics. Section 4 studies the F statistic from the weak IV literature and shows that standard weak IV critical values cannot be applied. Section 5 studies Fernald's (2014) TFP proxy variables and the dynamic effects of TFP shocks. Section 6 concludes.

2 The Model

2.1 The Structural VAR

The structural VAR follows Equations (1) and (2). Without loss of generality, I order v_t so that its first element is the structural shock of interest. Then, Equation (2) is

$$u_t = \begin{bmatrix} \underset{(n \times 1)}{B_1} & \underset{(n \times n-1)}{B_2} \end{bmatrix} \begin{bmatrix} \underset{(1 \times 1)}{v_{1,t}} \\ \underset{(n-1 \times 1)}{v_{2,t}} \end{bmatrix} \tag{3}$$

so that $v_{1,t}$ is the shock of interest, and $v_{2,t}$ contains the other structural shocks. Here, the vector B_1 determines how $v_{1,t}$ impacts Y_t, and estimating this vector is the focus of this paper. With this in mind, I make the following common assumptions about the properties of the structural VAR model.

Assumption 1:

a) The lag order p is known and the VAR is stationary.

b) B is invertible.

c) $\mathbb{E}(v_t) = 0$.

d) $\mathbb{E}(v_t v_t') = \Sigma_v$, where Σ_v is positive-definite. Further,

$$\Sigma_v = \begin{bmatrix} \sigma_{v_1}^2 & 0 \\ 0 & \Sigma_{v_2} \end{bmatrix}, \tag{4}$$

where $\mathbb{E}(v_{1,t}^2) = \sigma_{v_1}^2$ and $\mathbb{E}(v_{2,t}v_{2,t}') = \Sigma_{v_2}$ so that $v_{1,t}$ is uncorrelated with $v_{2,t}$.

e) $\mathbb{E}(v_t v_s') = 0$ for $t \neq s$.

It is useful to note that the stationarity assumption implies that the VAR has a Wold decomposition, $Y_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j}$, where $\Psi_0 = I_n$ and I_n is the $n \times n$ identity matrix. Then, Equation (2) and Assumption 1.e imply $\mathbb{E}(v_t Y_{t-j}') = 0$ for $j \geq 1$.

Given Assumption 1.d, if $v_{1,t}$ were observable, then B_1 could be consistently estimated by simply including $v_{1,t}$ in a least-squares estimation of Equation (1). However, because $v_{1,t}$ is not directly observable, the structural VAR literature has turned to a variety of alternative methods of estimating B_1. In the next section, I show that having an exogenous proxy variable can identify B_1.

2.2 The Proxy Variable and Identification

There exists a time series variable, denoted by z_t, that can be used as a proxy for $v_{1,t}$. Specifically, I make the following assumptions.

Assumption 2:

a) z_t has finite mean, denoted by μ_z.

b) z_t has a finite variance, denoted by σ_z^2.

c) z_t is a *relevant* proxy for $v_{1,t}$,

$$\mathbb{E}(v_{1,t} z_t) = \phi \neq 0, \text{ with } \phi \text{ finite.} \tag{5}$$

d) z_t is *exogenous* from the structural shocks $v_{2,t}$,

$$\mathbb{E}(v_{2,t} z_t) = 0. \tag{6}$$

Assumptions 2.a and 2.b ensure that the proxy is well behaved; however, it is assumptions 2.c and 2.d that make it useful for identification. These relevance and exogeneity assumptions mirror those used in the IV literature. However, it is important to note that the proxy variable is solving a different identification problem than an IV. In this model, the econometric problem is not that B_1 cannot be identified because $v_{1,t}$ is correlated with $v_{2,t}$. This has been ruled out by Equation (4) in Assumption 1.d. Rather the econometric problem is that B_1 cannot be identified because $v_{1,t}$ is not observable. Thus, although the relevance and exogeneity assumptions are similar to the assumptions for instruments, they are being used for a different identification problem.

Equations (5) and (6), along with the partition in Equation (3), imply

$$\mathbb{E}(u_t z_t) = B_1 \phi. \tag{7}$$

Here, the population covariance between the proxy variable and the VAR innovations gives B_1 up to the scalar ϕ. In order to estimate ϕ, note that

$$\begin{aligned}\mathbb{E}(z_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t z_t) &= \phi B_1'(B\Sigma_v B')^{-1} B_1 \phi \\ &= \phi e_1' \Sigma_v^{-1} e_1 \phi \\ &= \phi^2 \sigma_{v_1}^{-2},\end{aligned}$$

where the first line applies Equations (2), (4) and (7), the second line applies $B^{-1}B_1 = e_1$ where $e_1 = [1, 0, \ldots, 0]'$, and the third line applies $e_1'\Sigma_v^{-1}e_1 = \sigma_{v_1}^{-2}$ from Equation (4). I summarize this list of equations with

$$\phi^2 = \sigma_{v_1}^2 \mathbb{E}(z_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t z_t). \tag{8}$$

Thus, given the variance of the structural shock of interest, the covarainces of the VAR innovations and the proxy variable can recover ϕ^2. However, because $v_{1,t}$ is unobservable, there is insufficient information to separately identify ϕ^2 from the scalar $\sigma_{v_1}^2$. Because of this, I follow Mertens and Ravn (2013) and use the normalization

$$\sigma_{v_1}^2 = 1. \tag{9}$$

This normalization has no impact on the IRFs from a one standard deviation shock to $v_{1,t}$ nor on the variance contribution of $v_{1,t}$ to Y_t. Following Stock and Watson (2008), we can interpret this normalization as simply assigning a unit of measurement to $v_{1,t}$ so that its variance is equal to 1. Given the normalization in Equation (9), Equations (7) and (8) imply

$$B_1 = \pm \mathbb{E}(u_t z_t)\{\mathbb{E}(z_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t z_t)\}^{-1/2}. \tag{10}$$

This shows that B_1 can be computed as a simple function of the covariances of the VAR innovations and the proxy variable. The plus or minus in Equation (10) is a result of the square root of ϕ^2. Because ϕ is the covariance of z_t and $v_{1,t}$, it is up to the researcher to determine whether their proxy is positively or negatively correlated with the structural shock of interest. If the proxy is intended to be positively correlated, then the positive sign in Equation (10) should be applied. If the proxy is intended to be negatively correlated, then the negative sign should be applied.[4]

[4]Of course, one can always multiply the proxy variable by -1 and use the other sign.

To recover the structural shock of interest, I follow Stock and Watson (2012) and identify $v_{1,t}$ as the predicted value in the population regression of z_t on u_t. That is,

$$\pi = \left[\mathbb{E}(u_t u_t')\right]^{-1} \mathbb{E}(u_t z_t) \tag{11}$$

are the regression coefficients of z_t on u_t. Because z_t is correlated with $v_{1,t}$ but not with $v_{2,t}$, this regression returns the component of u_t that is driven by $v_{1,t}$ without influence from $v_{2,t}$. That is, using u_t as a best linear predictor of z_t yields an estimate of $v_{1,t}$ up to the scalar ϕ:

$$\begin{aligned}\mathbb{E}(z_t|u_t) &= \mu_z + u_t'\pi \\ &= \mu_z + u_t' \left[\mathbb{E}(u_t u_t')\right]^{-1} \mathbb{E}(u_t z_t) \\ &= \mu_z + v_t' B'(B\Sigma_v B')^{-1} B_1 \phi \\ &= \mu_z + v_{1,t}\phi.\end{aligned}$$

This list of equations shows that the expectation of z_t conditional on u_t is equivalent to what the expectation of z_t conditional on $v_{1,t}$ would be if $v_{1,t}$ could be used as a regressor. I summarize the above list of equations as

$$u_t'\pi = v_{1,t}\phi. \tag{12}$$

Rewriting Equation (12) as $v_{1,t} = u_t'\pi\phi^{-1}$ gives an estimate of the structural shock of interest where ϕ and π can be computed from Equations (8) and (11) above, along with the normalization in Equation (9).

2.3 Estimation and Consistency

To estimate B_1, I first define $A = [A_0', A_1', \ldots, A_p']'$ and $X_t = [1, Y_{t-1}', \ldots, Y_{t-p}']'$ so that Equation (1) can be written as $Y_t' = X_t'A + u_t'$. Then, I define the following matrices

$$\underset{(T\times n)}{Y} = \begin{bmatrix} Y_1 & \cdots & Y_T \end{bmatrix}' \quad \underset{(T\times np+1)}{X} = \begin{bmatrix} X_1 & \cdots & X_T \end{bmatrix}' \quad \underset{(T\times 1)}{Z} = \begin{bmatrix} z_1 & \cdots & z_T \end{bmatrix}'$$
$$\underset{(T\times n)}{U} = \begin{bmatrix} u_1 & \cdots & u_T \end{bmatrix}' \quad \underset{(T\times n)}{\hat{U}} = \begin{bmatrix} \hat{u}_1 & \cdots & \hat{u}_T \end{bmatrix}',$$

where $\hat{u}_t$ denotes the estimate of u_t. I estimate the VAR coefficients by least squares

$$\hat{A} = (X'X)^{-1}X'Y, \tag{13}$$

with the VAR errors given by

$$\hat{U} = Y - X\hat{A}, \tag{14}$$

and the small sample estimate of π is given by

$$\hat{\pi} = (T^{-1}\hat{U}'\hat{U})^{-1}(T^{-1}\hat{U}'Z). \tag{15}$$

Next, the estimators of the moments in Equations (7), (8) and (9) can be written as

$$\widehat{B_1\phi} = T^{-1}\hat{U}'Z \tag{16}$$

and

$$\widehat{\phi^2} = (T^{-1}Z'\hat{U})(T^{-1}\hat{U}'\hat{U})^{-1}(T^{-1}\hat{U}'Z). \tag{17}$$

Then, the estimator for B_1 from Equation (10) is

$$\hat{B}_1 = \pm(T^{-1}\hat{U}'Z)[(T^{-1}Z'\hat{U})(T^{-1}\hat{U}'\hat{U})^{-1}(T^{-1}\hat{U}'Z)]^{-1/2}, \tag{18}$$

which can be computed directly from $\hat{U}$ and Z.

To establish consistency of $\hat{B}_1$, I make the following assumption.

Assumption 3:

a) $T^{-1}X'X \xrightarrow{p} \mathbb{E}(X_tX_t')$ where $\mathbb{E}(X_tX_t')$ is positive definite.

b) $T^{-1}X'U \xrightarrow{p} \mathbb{E}(X_tu_t')$.

c) $T^{-1}X'Z \xrightarrow{p} \mathbb{E}(X_tz_t)$ where $\mathbb{E}(X_tz_t)$ is finite.

d) $T^{-1}U'U \xrightarrow{p} \mathbb{E}(u_tu_t')$.

e) $T^{-1}U'Z \xrightarrow{p} \mathbb{E}(u_tz_t)$.

More primitive assumptions that yield these law of large number results can be found in Jentsch and Lunsford (2016). Next, note that Equation (16) can be rewritten as

$$\widehat{B_1\phi} = T^{-1}(U - X'(X'X)^{-1}X'U)'Z. \tag{19}$$

Then, Assumptions 1, 2, and 3 and the continuous mapping theorem imply $\widehat{B_1\phi} \xrightarrow{p} \mathbb{E}(u_tz_t) = B_1\phi$. Similarly, Assumptions 1, 2, and 3 and the continuous mapping theorem imply $\hat{\phi}^2 \xrightarrow{p} \mathbb{E}(z_tu_t')[\mathbb{E}(u_tu_t')]^{-1}\mathbb{E}(u_tz_t) = \phi^2$. Finally, from these results and the continuous mapping theorem, it is the case that $\hat{B}_1 \xrightarrow{p} \pm B_1\phi/|\phi|$, where $|\cdot|$ denotes absolute value. As discussed above, when researchers intend for their proxy variable to be positively correlated ($\phi > 0$) with the shock of interest then they choose the positive sign in Equation (18), and when they intend for this correlation to be negative ($\phi < 0$) then they choose the negative sign. In either case, Equation (18) provides a consistent estimator for B_1.

After estimating ϕ and applying the appropriate sign, the structural shock of interest, $v_{1,t}$, can be estimated from Equation (12) with

$$\hat{v}_{1,t} = \hat{u}_t' \hat{\pi} \hat{\phi}^{-1}. \tag{20}$$

Note that Assumptions 1, 2, and 3 along with the continuous mapping theorem imply $\hat{\pi} \xrightarrow{p} [\mathbb{E}(u_t u_t')]^{-1} \mathbb{E}(u_t z_t) = \pi$.

3 Testing for a Weak Proxy Variable

I center the analysis of a weak proxy variable on the regression of z_t on u_t that was used to recover $v_{1,t}$. Define $\epsilon_t = z_t - \mu_z - u_t'\pi$ to be the regression error, where π is given in Equation (11). Then using Equation (12), it is the case that

$$\begin{aligned} z_t &= \mu_z + u_t'\pi + \epsilon_t \\ &= \mu_z + v_{1,t}\phi + \epsilon_t. \end{aligned} \tag{21}$$

As in the weak IV literature, I model a weakly relevant proxy variable with a local-to-zero assumption (Staiger and Stock, 1997). Here, the covariance of a weak proxy and the structural shock of interest is local-to-zero:

$$\phi_T = C/\sqrt{T}. \tag{22}$$

From the assumptions in Section 2, it is true that $\pi = (B\Sigma_v B')^{-1} B_1 \phi$. Then, I use Equation (22) to define $\pi_T = (B\Sigma_v B')^{-1} B_1 \phi_T$, and I rewrite the equations in (21) as

$$\begin{aligned} z_t &= \mu_z + u_t'\pi_T + \epsilon_t \\ &= \mu_z + v_{1,t} C/\sqrt{T} + \epsilon_t. \end{aligned} \tag{23}$$

Hence, when z_t is weak, its signal for $v_{1,t}$ goes to zero as the sample size grows. In the next three subsections, I characterize the limiting distribution of $\hat{B}_1$ given the equations in (23). Further, I discuss the bias of this distribution and show that it is small when C is large relative to the standard deviation of ϵ_t. Finally, I show that the F statistic on the null hypothesis that $\pi_T = 0$ in Equation (23) provides a test that this parameter is large.

3.1 Inconsistency of the B_1 Estimator

To characterize the limit of $\hat{B}_1$ in Equation (18) with the weak proxy model in (23), I first note that the construction of ϵ_t implies $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{E}(u_t \epsilon_t) = 0$ with a strong proxy. I assume that these properties continue to hold with a weak proxy. Further, I define $E = [\epsilon_1, \ldots, \epsilon_T]'$ and place the following assumptions on the model in (23):

Assumption 4:

a) $\mu_z = 0$.

b) $T^{-1}X'E \xrightarrow{p} \mathbb{E}(X_t\epsilon_t) = 0$.

c) $T^{-1/2}U'E \xrightarrow{d} N(0, B\Omega B')$ where Ω is the long-run variance of $v_t\epsilon_t$ and is positive-definite.

Assumption 4.a is a restriction on Assumption 2.a. Next, assumption 4.b. implies $T^{-1}X'Z \xrightarrow{d} 0$ so that z_t is uncorrelated with X_t asymptotically, further restricting Assumption 3.c. However, the requirements that z_t be mean zero and uncorrelated with X_t are not very restrictive in practice. To fulfill these requirements, we can define $\hat{z}_t$ as the elements of $\hat{Z} = Z - X(X'X)^{-1}X'Z$ and use $\hat{z}_t$ as the proxy in place of z_t. From the construction of $\hat{U}$, it is true that $\hat{U}'\hat{Z} = \hat{U}'Z$, and all results from Section 2 will continue to hold.

Next, I combine Equations (19) and (23) to get

$$\widehat{B_1\phi} = T^{-1}U'U\pi_T + T^{-1}U'E + (T^{-1}U'X)(T^{-1}X'X)^{-1}(T^{-1}X'Z). \tag{24}$$

Given $\pi_T = (B\Sigma_v B')^{-1}B_1C/\sqrt{T}$, it is the case that

$$\sqrt{T}\widehat{B_1\phi} \xrightarrow{d} B\theta, \tag{25}$$

where

$$\theta \sim e_1C + N(0, \Omega). \tag{26}$$

This follows from Assumptions 1, 3 and 4, the continuous mapping theorem and Slutsky's theorem. Then, from Equations (16) and (18), $\hat{B}_1$ can be written as

$$\hat{B}_1 = \pm\sqrt{T}\widehat{B_1\phi}\left[(\sqrt{T}\widehat{B_1\phi})'(T^{-1}\hat{U}'\hat{U})^{-1}(\sqrt{T}\widehat{B_1\phi})\right]^{-1/2},$$

with the following result

$$\hat{B}_1 \xrightarrow{d} \pm B_1^* \quad \text{where} \quad B_1^* = B\theta(\theta'\Sigma_v^{-1}\theta)^{-1/2}. \tag{27}$$

This follows from Equation (25), $T^{-1}\hat{U}'\hat{U} \xrightarrow{p} B\Sigma_v B'$, the continuous mapping theorem and Slutsky's theorem. Thus, when z_t is weak, $\hat{B}_1$ is no longer a consistent estimator for B_1. Rather, $\hat{B}_1$ converges in distribution to B_1^*,[5] which is a function of the random vector θ.

[5] For ease of notation, I use B_1^* in place of $\pm B_1^*$ with the understanding that the appropriate sign is chosen. Application of the appropriate sign on $\hat{B}_1$ follows as in the strong proxy case. When the proxy is intended to be positively correlated with the structural shock, then $C > 0$ and the positive sign is used in Equation (27). When the proxy is intended to be negatively correlated with the structural shock, then $C < 0$ and the negative sign is used in Equation (27).

While $\hat{B}_1$ is inconsistent under the weak proxy assumption, its limiting distribution, B_1^*, may be close to B_1, implying that $\hat{B}_1$ still can be useful for approximating B_1. To see this, first re-write B_1^* as $BC^{-1}\theta(C^{-1}\theta'\Sigma_v^{-1}C^{-1}\theta)^{-1/2}$. Then, as $C^2 \to \infty$ holding Ω fixed, it is the case that $C^{-1}\theta \xrightarrow{p} e_1$ so that $B_1^* \xrightarrow{p} B_1$. Similarly, as the elements of $\Omega \to 0$ holding C^2 fixed, it is the case that $C^{-1}\theta \xrightarrow{p} e_1$ so that $B_1^* \xrightarrow{p} B_1$. Hence, as C^2 becomes large relative to Ω, it is the case that B_1^* collapses to B_1, and $\hat{B}_1$ becomes an increasingly accurate estimator.

3.2 The Asymptotic Bias of the B_1 Estimator

To measure how close B_1^* is to B_1, I follow the weak IV literature and use the bias of B_1^* (Stock, Wright, and Yogo, 2002; Stock and Yogo, 2005). Correspondingly, I make the following assumption about the long-run variance of $v_t\epsilon_t$:

Assumption 5: $\Omega = \Sigma_v\sigma_\epsilon^2$, where $\mathbb{E}(\epsilon_t^2) = \sigma_\epsilon^2$.

This assumption parallels part (c) of Assumption M in Staiger and Stock (1997) and in Stock and Yogo (2005). One way to satisfy Assumption 5 is to assume that ϵ_t is serially uncorrelated and homoskedastic: $\mathbb{E}(\epsilon_t\epsilon_s) = 0$ for $t \neq s$ and $\mathbb{E}(\epsilon_t^2|v_t) = \mathbb{E}(\epsilon_t^2)$. Given assumption 1.e, another way to satisfy Assumption 5 is to assume a more general homoskedastic form: $\mathbb{E}(\epsilon_t\epsilon_s|v_t, v_s) = \mathbb{E}(\epsilon_t\epsilon_s)$ for all t and s. Thus, Assumption 5 does not require ϵ_t to be serially uncorrelated if its autocovariances are independent of the structural shocks. While Assumption 5 is not desirable, it is necessary because $v_{2,t}$ cannot be estimated, preventing the estimation of an unrestricted Ω. Hence, the approach in Montiel Olea and Pflueger (2013), which relaxes part (c) of Assumption M in Staiger and Stock (1997) and Stock and Yogo (2005) but requires consistent estimates of long-run variances, cannot be applied here.

To characterize the bias of B_1^*, I use the following lemma.

Lemma 1 *Define the $n \times 1$ random vector $\tilde{\theta} = \theta(\theta'\Sigma_v^{-1}\theta)^{-1/2}$ with elements $\tilde{\theta}_j$ for $j = 1, \ldots, n$. Then, $\mathbb{E}(\tilde{\theta}_j) = 0$ for $j \geq 2$.*

The proof of this lemma is provided in the appendix. Given the definition of $\tilde{\theta}$ in Lemma 1, (27) implies that $B_1^* = B\tilde{\theta}$. Define $b = \mathbb{E}(\tilde{\theta}_1)$. Then, Lemma 1 implies $\mathbb{E}(B_1^*) = B_1 b$ so that $\mathbb{E}(\tilde{\theta}_1)$ characterizes the asymptotic bias of $\hat{B}_1$. Next, note that Equations (4) and (9) and Assumption 5 imply

$$\tilde{\theta} = \Sigma_v^{1/2}(e_1 C/\sigma_\epsilon + x)[(e_1 C/\sigma_\epsilon + x)'(e_1 C/\sigma_\epsilon + x)]^{-1/2}, \tag{28}$$

where x is an $n \times 1$ vector of independent standard normal random variables. $\tilde{\theta}$ is a function of Σ_v, C/σ_ϵ and n standard normal random variables. Further, given Equations (4) and (9), $\tilde{\theta}$ converges in probability to e_1 as $C^2/\sigma_\epsilon^2 \to \infty$. Thus, the signal-to-noise ratio of the proxy for the structural shock interest corresponds to the concentration parameter in the weak

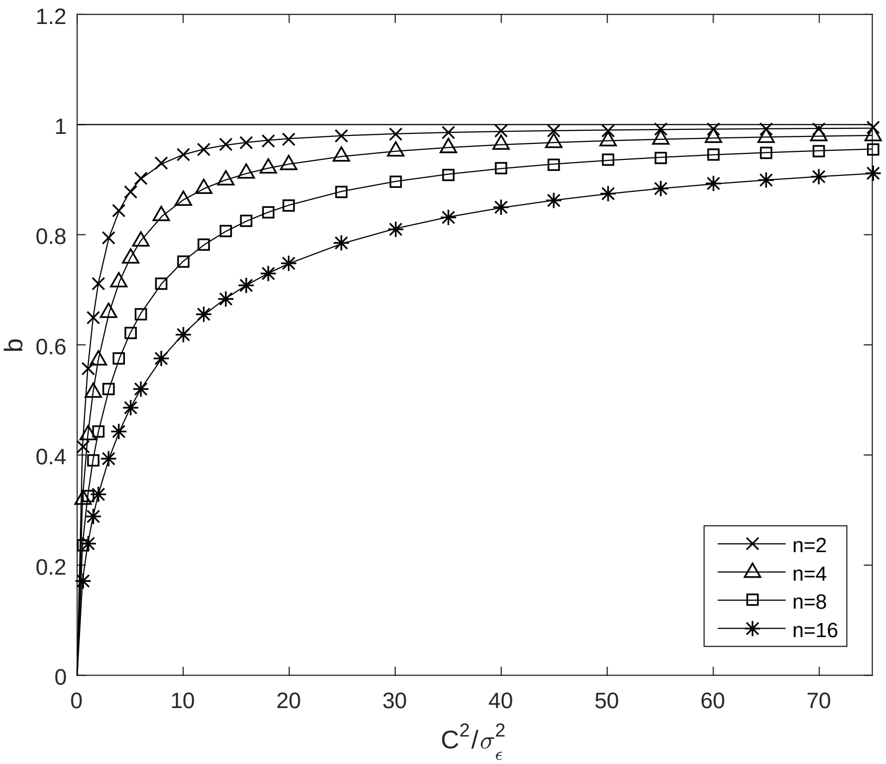

Figure 1: b as a function of C^2/σ_ϵ^2 and n.

IV literature (Stock, Wright, and Yogo, 2002; Stock and Yogo, 2005) because B_1^* becomes increasingly concentrated around B_1 as C^2/σ_ϵ^2 increases.

I characterize b with simulation of Equation (28). First, I fix a value of C^2/σ_ϵ^2. Second, I draw 10,000 observations of the random vector x. Third, b is the average of the first element of $(e_1 C/\sigma_\epsilon + x)[(e_1 C/\sigma_\epsilon + x)'(e_1 C/\sigma_\epsilon + x)]^{-1/2}$ over the 10,000 observations. Figure 1 shows the results of these simulations for multiple choices of C^2/σ_ϵ^2 and n. In this figure, I use $C/\sigma_\epsilon = +\sqrt{C^2/\sigma_\epsilon^2}$. Results for $C/\sigma_\epsilon = -\sqrt{C^2/\sigma_\epsilon^2}$ simply flip this image so that b is negative.

Figure 1 shows that b is bounded between 0 and 1, implying that B_1^* is biased toward zero. When $C^2/\sigma_\epsilon^2 = 0$, then Lemma 1 applies to $\tilde{\theta}_1$ so that $b = 0$. As C^2/σ_ϵ^2 increases, b increases and asymptotes to 1. When n is small, b converges quickly to 1 and the asymptotic bias becomes small. However, as n increases, b goes to 1 more slowly and larger values of C^2/σ_ϵ^2 are needed to produce a small bias.

Figure 1 implies that for a given bias tolerance, $1-b$, and for a given VAR dimension n, a researcher can find a minimum value of C^2/σ_ϵ^2 so that the the bias tolerance is not exceeded. I denote this minimum value by $(C^2/\sigma_\epsilon^2)^*$, and I use it to define the *weak proxy set*. This set is the collection of all proxy variables where $C^2/\sigma_\epsilon^2 \in [0, (C^2/\sigma_\epsilon^2)^*]$. These are the proxy variables with small signal-to-noise ratios, implying that they generate biases larger than what is tolerated. Given b and n, I compute the values of $(C^2/\sigma_\epsilon^2)^*$ by using simulation and the method of bisection. My algorithm proceeds as follows

1. Fix b and n.
2. Set a minimum bound on C/σ_ϵ of $min = 0$ and a maximum bound of $max = 40$.[6]
3. Compute a guess of $C/\sigma_\epsilon = (min + max)/2$.
4. Given the guess of C/σ_ϵ, draw 100,000 observations of the random vector $e_1C/\sigma_\epsilon + x$.
5. Use these draws to compute 100,000 observations of $(e_1C/\sigma_\epsilon+x)[(e_1C/\sigma_\epsilon+x)'(e_1C/\sigma_\epsilon+x)]^{-1/2}$ and take the average of the first element to be $\bar{b}$.
6. If $|b - \bar{b}| < 1 \times 10^{-8}$, then stop the algorithm and set $(C^2/\sigma_\epsilon^2)^*$ to be the square of the guess of C/σ_ϵ. If not, proceed to the next step.
7. If $b < \bar{b}$, then set the maximum bound equal to the guess of C/σ_ϵ. If not, then set the minimum bound equal to the guess of C/σ_ϵ. Return to step 3.

I repeated this process for b equal to 0.80, 0.90, 0.95 and 0.99 and for $n = 2, \ldots, 20$. The results are presented in Table 1. They indicate that the threshold for the weak proxy set increases with b and n. However, because C^2/σ_ϵ^2 cannot be consistently estimated, this table is not directly applicable for determining whether a proxy variable is in the weak proxy set or not. Because of this, I use the testing procedure described in the next subsection.

Before turning to that test however, I first note that the biases in Figure 1 and the weak proxy sets in Table 1 apply to the estimator of B_1. This parameter is useful for producing IRFs to shocks of one standard deviation, as in Gertler and Karadi (2015). However, some researchers may compute IRFs where an element of the initial response has been normalized to a specific value. For example, researchers may normalize a monetary shock so that the initial response of the federal funds rate is 0.25% or they may normalize a fiscal shock so that government spending increases by 1%. In this case, the estimated initial response is $s\hat{B}_1/(e_i'\hat{B}_1)$, where s is the scale of the normalization and e_i is a vector that selects the ith element of B_1 to be normalized.[7] This can be re-written as $s\hat{B}_1/(e_i'\hat{B}_1) = s\sqrt{T}\widehat{B_1\phi}/(e_i'\sqrt{T}\widehat{B_1\phi})$, which converges in distribution to $sB\theta/(e_i'B\theta)$. Thus, as with $\hat{B}_1$, $s\hat{B}_1/(e_i'\hat{B}_1)$ will be inconsistent. Further, $s\hat{B}_1/(e_i'\hat{B}_1)$ will also be biased. Details of the form of this bias are provided in the appendix. As with $\hat{B}_1$, the bias of $s\hat{B}_1/(e_i'\hat{B}_1)$ goes to zero as $(C^2/\sigma_\epsilon^2) \to \infty$. Hence, a large signal-to-noise ratio for the proxy variable will also ensure that normalized IRFs are close to the true parameters.

[6]This implicitly assumes that $C > 0$. This algorithm can also be run for $C < 0$ by setting $min = -40$ and $max = 0$. Identical results down to a small simulation error will be achieved.

[7]That is, e_i is an $n \times 1$ with a value of 1 in the ith position and zeros everywhere else.

Table 1: Values of $(C^2/\sigma_\epsilon^2)^*$ given b and n

n	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$
2	3.12	6.03	11.05	51.05
3	4.77	10.02	20.07	100.29
4	6.48	14.18	29.26	149.55
5	8.21	18.40	38.52	198.99
6	9.98	22.68	47.84	248.60
7	11.74	26.93	57.07	297.74
8	13.51	31.19	66.35	347.11
9	15.27	35.42	75.54	396.05
10	17.04	39.68	84.79	445.27
11	18.81	43.93	94.03	494.39
12	20.60	48.23	103.37	544.16
13	22.36	52.47	112.58	593.15
14	24.14	56.73	121.83	642.39
15	25.93	61.02	131.16	692.03
16	27.69	65.25	140.34	740.87
17	29.48	69.54	149.67	790.54
18	31.26	73.82	158.96	839.99
19	33.04	78.10	168.24	889.38
20	34.81	82.35	177.48	938.55

Notes: $(C^2/\sigma_\epsilon^2)^*$ is the upper bound on the weak proxy set, n is the VAR dimension, and $1 - b$ is the asymptotic bias.

3.3 A Test for a Weak Instrument

This subsection provides a test for whether or not z_t is in the weak proxy set. Formally, the null and alternative hypotheses are

$$\mathbb{H}_0 : C^2/\sigma_\epsilon^2 \in [0, (C^2/\sigma_\epsilon^2)^*] \quad \text{vs.} \quad \mathbb{H}_1 : C^2/\sigma_\epsilon^2 \in ((C^2/\sigma_\epsilon^2)^*, \infty).$$

To test these hypotheses, I regress z_t on u_t and test if the regression coefficients are zero. That is, I am testing $\pi_T = 0$ in Equation (23). The F statistic is given by

$$\begin{aligned} F &= \left(\frac{T-n}{n}\right) \frac{Z'Z - (Z - \hat{U}\hat{\pi})'(Z - \hat{U}\hat{\pi})}{(Z - \hat{U}\hat{\pi})'(Z - \hat{U}\hat{\pi})} \\ &= \frac{1}{n}\left(\frac{T-n}{T}\right) \frac{(T^{-1/2}Z'\hat{U})(T^{-1}\hat{U}'\hat{U})^{-1}(T^{-1/2}\hat{U}'Z)}{T^{-1}\hat{E}'\hat{E}}, \end{aligned} \tag{29}$$

where $\hat{E} = Z - \hat{U}\hat{\pi}$. Then, with the weak proxy model in (23),

$$F \xrightarrow{d} n^{-1}(e_1 C/\sigma_\epsilon + x)'(e_1 C/\sigma_\epsilon + x), \tag{30}$$

which follows from Equation (25), Assumption 5, $T^{-1}\hat{U}'\hat{U} \xrightarrow{p} B\Sigma_v B'$, $T^{-1}\hat{E}'\hat{E} \xrightarrow{p} \sigma_\epsilon^2$, the continuous mapping theorem and Slutsky's theorem. Equation (30) shows that nF converges to a non-central χ_n^2 distribution with a non-centrality parameter of C^2/σ_ϵ^2. Then, the testing procedure is as follows.

1. Choose a bias tolerance $1 - b$ and a level of significance α.
2. Using the choice of b along with n, find the weak proxy set threshold $(C^2/\sigma_\epsilon^2)^*$.
3. Compute F in Equation (29).
4. Compute $\Pr(X \leq nF)$ where X is a random variable from the non-central χ_n^2 distribution with a non-centrality parameter of $(C^2/\sigma_\epsilon^2)^*$.
5. If $1 - \Pr(X \leq nF) < \alpha$ reject $\mathbb{H}_0$; otherwise, fail to reject $\mathbb{H}_0$.

The idea behind this testing procedure is that if F is sufficiently large, then the probability that it comes from a non-central χ_n^2 distribution with a non-centrality parameter at or below $(C^2/\sigma_\epsilon^2)^*$ is sufficiently small to indicate that the true value of C^2/σ_ϵ^2 is above $(C^2/\sigma_\epsilon^2)^*$. Using the values in Table 1, I compute the corresponding F statistics for α equal to 0.90, 0.95 and 0.99 and report them in Table 2. As in the weak IV literature, the threshold F statistics are large. For example, if one wants to reject the threshold signal-to-noise ratio that corresponds to a 10% asymptotic bias at a 5% level of significance, then the critical F statistics are between 7 and 9 for conventional VAR dimensions. For comparison purposes,

Table 2: Critical Values for the F Statistic

n	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$
2	5.31	7.61	11.20	36.04	6.51	9.06	12.96	39.18	9.12	12.13	16.61	45.42
3	4.85	7.38	11.82	43.24	5.77	8.53	13.28	46.03	7.73	10.92	16.26	51.50
4	4.58	7.24	12.03	46.42	5.35	8.22	13.30	48.92	6.96	10.23	15.85	53.78
5	4.40	7.12	12.09	48.19	5.07	7.98	13.23	50.47	6.46	9.74	15.50	54.87
6	4.27	7.03	12.11	49.31	4.87	7.81	13.14	51.41	6.10	9.39	15.20	55.45
7	4.17	6.94	12.08	49.98	4.71	7.65	13.04	51.93	5.83	9.10	14.93	55.69
8	4.09	6.86	12.05	50.47	4.59	7.53	12.94	52.30	5.61	8.86	14.70	55.83
9	4.01	6.79	12.00	50.77	4.48	7.41	12.84	52.51	5.43	8.66	14.49	55.83
10	3.95	6.73	11.96	51.03	4.39	7.32	12.76	52.67	5.28	8.49	14.31	55.83
11	3.90	6.68	11.92	51.21	4.31	7.24	12.68	52.78	5.15	8.34	14.15	55.80
12	3.86	6.64	11.90	51.41	4.25	7.17	12.61	52.92	5.04	8.22	14.02	55.80
13	3.82	6.59	11.86	51.50	4.19	7.10	12.54	52.95	4.95	8.10	13.89	55.73
14	3.78	6.55	11.82	51.59	4.14	7.04	12.48	52.99	4.86	8.00	13.78	55.66
15	3.75	6.52	11.79	51.70	4.09	6.99	12.43	53.05	4.78	7.91	13.68	55.63
16	3.72	6.48	11.76	51.73	4.05	6.93	12.38	53.04	4.71	7.83	13.58	55.53
17	3.69	6.46	11.73	51.80	4.01	6.89	12.33	53.07	4.65	7.75	13.50	55.50
18	3.67	6.43	11.71	51.85	3.98	6.85	12.29	53.08	4.60	7.68	13.42	55.44
19	3.65	6.40	11.68	51.89	3.95	6.81	12.25	53.09	4.54	7.62	13.34	55.38
20	3.62	6.38	11.66	51.91	3.92	6.78	12.21	53.08	4.50	7.56	13.27	55.31

Notes: The F statistic is from the regression of the proxy variable on the VAR errors. n is the dimension of the VAR, $1 - b$ is the level of asymptotic bias, and α is the level of statistical significance.

if $(C^2/\sigma_\epsilon^2)^*$ was zero so that the asymptotic distribution of nF was a central χ^2_n, then the critical F values would be between 1.5 and 3 for the 5% level of significance.

4 An Analysis of Alternative Identification Methods and Weak Proxy Tests

Previous papers that have used proxy variables to identify structural VAR shocks have followed a slightly different identification procedure than the one proposed in Section 2 above. Gertler and Karadi (2015) and Lunsford (2015) further partition Equation (3) into

$$\begin{bmatrix} u_{1,t} \\ (1\times 1) \\ u_{2,t} \\ (n-1\times 1) \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ (1\times 1) & (1\times n-1) \\ b_{21} & b_{22} \\ (n-1\times 1) & (n-1\times n-1) \end{bmatrix} \begin{bmatrix} v_{1,t} \\ (1\times 1) \\ v_{2,t} \\ (n-1\times 1) \end{bmatrix}.$$

Mertens and Ravn (2013) use a similar partition except that they have two shocks of interest so that that $u_{1,t}$ and $v_{1,t}$ are 2×1. Montiel Olea, Stock, and Watson (2012) also use a similar partition, but make a different normalization. Instead of setting $\sigma^2_{v_1} = 1$ as in Equation (9), they use a unit shock normalization where $b_{11} = 1$, implying that a one unit shock in $v_{1,t}$ produces a one unit shock to $u_{1,t}$. However, whether the normalization used is $\sigma^2_{v_1} = 1$ or $b_{11} = 1$, the above partition along with the relevance and endogeneity assumptions in Equations (5) and (6) imply

$$\mathbb{E}(u_{1,t}z_t) = b_{11}\phi$$

and

$$\mathbb{E}(u_{2,t}z_t) = b_{21}\phi,$$

which combine to yield

$$b_{21}b_{11}^{-1} = \mathbb{E}(u_{2,t}z_t)[\mathbb{E}(u_{1,t}z_t)]^{-1}.$$

In Montiel Olea, Stock, and Watson (2012) this completes the estimation of B_1 because of the normalization $b_{11} = 1$. In Mertens and Ravn (2013), Gertler and Karadi (2015) and Lunsford (2015), once $b_{21}b_{11}^{-1}$ is estimated, b_{11} and b_{21} can then be separately estimated using $\mathbb{E}(u_t u_t') = B\Sigma_v B'$ and the normalization in Equation (9). See the Appendix of Mertens and Ravn (2013) or Appendix D of Lunsford (2015).

The above equation indicates that $b_{21}b_{11}^{-1}$ can be estimated by IV where z_t is an instrument for $u_{1,t}$. The corresponding structural and reduced form equations are

$$u_{2,t} = b_{21}b_{11}^{-1}u_{1,t} + \eta_{1,t}$$

and

$$u_{1,t} = \gamma z_t + \eta_{2,t}.$$

Thus, one may attempt to identify a weak proxy by following the weak IV literature and using the first-stage F statistic that tests $\gamma = 0$. This statistic is given by

$$\begin{aligned} F_{IV} &= (T-1)\frac{\hat{U}_1'\hat{U}_1 - (\hat{U}_1 - Z\hat{\gamma})'(\hat{U}_1 - Z\hat{\gamma})}{(\hat{U}_1 - Z\hat{\gamma})'(\hat{U}_1 - Z\hat{\gamma})} \\ &= \left(\frac{T-1}{T}\right)\frac{(T^{-1/2}\hat{U}_1'Z)(T^{-1}Z'Z)^{-1}(T^{-1/2}Z'\hat{U}_1)}{T^{-1}\hat{U}_1'U_1 - (T^{-1}\hat{U}_1'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z\hat{U}_1)} \end{aligned} \tag{31}$$

where $\hat{U}_1 = [\hat{u}_{1,1}, \ldots, \hat{u}_{1,T}]'$, $\hat{\gamma} = (Z'Z)^{-1}(Z'\hat{U}_1)$, and the notation F_{IV} denotes that this F statistic follows from the weak IV literature. This IV approach is taken by Montiel Olea, Stock, and Watson (2012), Gertler and Karadi (2015) and Lunsford (2015) who then compare F_{IV} to critical values from the weak IV literature to test for proxy strength.

However, this weak IV approach is flawed. Comparing F_{IV} to critical values from the weak IV literature is not useful for testing the strength of the covariance between z_t and $v_{1,t}$. This is because the asymptotic distribution of F_{IV} depends on the matrix B, which will be different in each application. Thus, the standard critical values from the weak IV literature will not apply because they cannot take B into account. To hightlight the problems with using F_{IV} as a test of proxy strength, I present a simple Mont Carlo experiment before turning to the asymptotic properties of F_{IV} in the following two subsections.

4.1 A Simple Monte Carlo Example

To hightlight the problem with the F_{IV} statistic, I first run a simple Monte Carlo experiment. In the experiment, I run two simulations where only one element of B changes between the two simulations. Then, I compare how frequently F_{IV} rejects the null of a weak proxy variable. In these simulations, I put aside the reduced-form VAR in Equation (1) and assume that the VAR innovations u_t are directly observable. This will allow for direct study of the weak proxy testing without any confounding problems that may arise from estimating u_t.

In both simulations, u_t is 2×1 and follows Equation (3) where $v_{1,t}$ and $v_{2,t}$ are both standard normal random variables. The data generating processes (DGPs) of each simulation are differentiated by B. DGP1 has

$$B = \begin{bmatrix} 1 & 10 \\ 1 & 1 \end{bmatrix},$$

and DGP2 has

$$B = \begin{bmatrix} 1 & 0.1 \\ 1 & 1 \end{bmatrix},$$

so that the only differences between the GDPs are the value of b_{12}. For both DGPs, the proxy variable follows the local-to-zero form in Equation (23):

$$z_t = v_{1,t}(2.456/\sqrt{T}) + \epsilon_t,$$

where ϵ_t is a standard normal random variable and T is the sample size. I choose $C = 2.456$ so that the signal-to-noise ratio is 6.03, which is from the $b = 0.90$ column of Table 1. Hence, $\hat{B}_1$ has an asymptotic bias of 10%.

For both DGPs, I run 10,000 simulations with sample size $T = 200{,}000$. This sample size is large so that the testing statistics behave similarly to their asymptotic distributions. I then compute F_{IV} in Equation (31) and compute the percentage of simulations where it exceeds the value 10. This value of 10 is a common rule of thumb for testing proxy strength and was used as a threshold by Stock and Watson (2012) and Gertler and Karadi (2015). For comparison purposes, I also compute F in Equation (29) and compute the percentage of simulations where it exceeds the value 9.06, which is the 5% critical value for $n = 2$ and $b = 0.90$ in Table 2. That is, this value of F is testing $\mathbb{H}_0 : C^2/\sigma_\epsilon^2 \in [0, 6.03]$ versus $\mathbb{H}_1 : C^2/\sigma_\epsilon^2 \in (6.03, \infty)$ at the 5% level.

In DGP1, F_{IV} exceeds 10 in only 0.3% of the simulations, implying that DGP1 would almost never yield a test that rejects z_t in the weak proxy set. However, in DGP2, F_{IV} exceeds 10 in 23.3% of the simulations, implying that DGP2 would yield a test that rejects z_t in the weak proxy set nearly a quarter of the time. These results are problematic given that F_{IV} is intended to be a test of the signal-to-noise ratio and that the signal-to-noise ratio is the same in both DGPs. This suggests that the standard critical values from the weak IV literature (Stock and Yogo, 2005) may not apply in this context and that the statistical size of this test can fluctuate depending on the elements of B. Further, the size can depend on a parameter, b_{12}, that is associated with structural shocks other than the shock of interest.

For comparison, F exceeds 9.06 in 5.3% of the simulations for both DGPs, which is very close to the 5% critical value that 9.06 represents for this DGP. Further, this shows that for large sample sizes, B does not impact F as implied by Equation (30).

4.2 Analysis of the F_{IV} Statistic

To understand why the two DGPs in the Monte Carlo experiment above gave different reject rates for the F_{IV} statistic, I first note that

$$F_{IV} \xrightarrow{d} (e_1' B \Sigma_v B' e_1)^{-1} [e_1' B \Sigma_v^{1/2} (e_1 C/\sigma_\epsilon + x)]^2, \tag{32}$$

where $e_1 = [1, 0, \ldots, 0]'$ and x is a $n \times 1$ vector of independent standard normal random variables. This result follows from the second line of Equation (31), $\hat{u}_{1,t} = e_1' \hat{u}_t$, Equation (25), Assumption 5, $T^{-1}\hat{U}'\hat{U} \xrightarrow{p} B\Sigma_v B'$, $T^{-1}Z'Z \xrightarrow{p} \sigma_\epsilon^2$, the continuous mapping theorem and Slutsky's theorem. The limiting distribution in Equation (32) can be rewritten as

$$\left(\frac{b_{11}C/\sigma_\epsilon}{\sqrt{b_{11}^2 + b_{12}\Sigma_{v_2}b_{12}'}} + w\right)^2,$$

where w is a standard normal random variable. Hence, F_{IV} converges in distribution to a non-central χ_1^2 with a non-centrality parameter of

$$\frac{b_{11}^2 C^2/\sigma_\epsilon^2}{b_{11}^2 + b_{12}\Sigma_{v_2}b_{12}'}.$$[8]

Here, we see that unlike F, F_{IV} does not converge to a distribution that is independent of B and Σ_v. Rather, with F_{IV}, the signal-to-noise ratio is now scaled by $b_{11}^2/(b_{11}^2 + b_{12}\Sigma_{v_2}b_{12}')$ in the asymptotic distribution. This scaling is the ratio of the variance of $u_{1,t}$ that is attributable to $v_{1,t}$ relative to the total variance of $u_{1,t}$. Thus, if the variance of $u_{1,t}$ is largely driven by $v_{2,t}$ as it was in DGP1 above, then the non-centrality parameter is near 0 and an F_{IV} statistic larger than 10 will rarely be drawn. This is consistent with the simulation results of DGP1. If the variance of $u_{1,t}$ is largely driven by $v_{1,t}$ as it was in DGP2 above, then the non-centrality parameter is close C^2/σ_ϵ^2. Because the signal-to-noise ratio in DGP2 was 6.03, this DGP will draw $F_{IV} > 10$ at a much higher rate. These results show that the weak IV critical values (Stock and Yogo, 2005) will no longer apply, and the rule of thumb of $F_{IV} > 10$ will not be a good guide for proxy strength.

With knowledge of the variance decomposition of $u_{1,t}$, it is still possible to test whether or not C^2/σ_ϵ^2 is in the weak proxy set with F_{IV}. To do this, the threshold signal-to-noise ratio needs to be scaled by the fraction of the variance of $u_{1,t}$ that is attributable to $v_{1,t}$ before finding a critical value from the corresponding non-central χ_1^2 distribution. For DGP1, the appropriate variance fraction is 1/101. Multiplying this by the threshold $(C^2/\sigma_\epsilon^2)^* = 6.03$ yields a non-centrality parameter of 0.060 and 5% critical value of 4.07, which is less than half of the rule of thumb critical value. When $F_{IV} > 4.07$ is used in DGP1 rather than $F_{IV} > 10$, then the null hypothesis is rejected in 4.8% of the simulations. The corresponding critical value for DGP2 is 16.71, and when $F_{IV} > 16.71$ is used in DGP2 rather than $F_{IV} > 10$, then the null hypothesis is rejected in 5.3% of the simulations. This shows that using F_{IV} may require big differences in critical values that will depend on the variance decomposition of $u_{1,t}$. These critical values will be different for each empirical application and will also vary depending on how researchers make their choice of $u_{1,t}$. That is, different orderings of the VAR will also imply different critical values for F_{IV}.

This process of scaling the threshold signal-to-noise ratio by the fraction of the variance of $u_{1,t}$ that is attributable to $v_{1,t}$ before computing critical values does not make F_{IV} useful in practice, however. This is because b_{11} is not known to researchers and has to be estimated. With a weak proxy, the estimate of b_{11} will be inconsistent and lead to an incorrectly

[8] This same result can be found in Montiel Olea, Stock, and Watson (2012).

estimated variance decomposition, yielding an incorrect critical value. As an example of this problem, I return to the Monte Carlo experiment above. For both DGPs, I estimate b_{11} with the first element of $\hat{B}_1$, and I estimate $b_{11}^2 + b_{12}\Sigma_{v_2}b_{12}'$ with $T^{-1}\sum_{t=1}^{T} u_{1,t}^2$. Then, I scale the threshold signal-to-noise ratio, 6.03, with the estimate of $b_{11}^2/(b_{11}^2 + b_{12}\Sigma_{v_2}b_{12}')$ to estimate the relevant non-centrality parameter and compute the 5% critical value. With DGP1, this estimated critical value rejects the null hypothesis in 0.1% of the simulations. With DGP2, this estimated critical value rejects the null hypothesis in 6.1% of the simulations. Thus, using F instead of F_{IV} gives better statistical size for both DGPs. Further, using F allows researchers to use the same critical values from Table 2 for all applications and regardless of the VAR ordering.

5 The Dynamic Effects of Productivity Shocks

As an application of the estimation and weak proxy testing in this paper, I study the dynamic effects of productivity shocks. In doing so, I highlight the usefulness of the F statistic proposed in Section 3 relative to the F_{IV} statistic studied in Section 4. I also highlight the benefits of using the proxy approach relative to treating z_t as the structural shock itself or to putting z_t into a Choleski VAR.

To study the dynamic effects of productivity shocks, I use Fernald's (2014) two measures of utilization-adjusted total factor productivity (TFP). The first is a measure of TFP in the consumption sector, which excludes durable goods. The second is a measure of TFP in the sector for durable goods and equipment investment. For simplicity, I refer to these as consumption TFP and investment TFP.

The VAR has five variables. The first three variables are annualized GDP growth, private employment growth and inflation in percent.[9] The last two variables are the combined annualized percent growth in non-durable goods and services consumption and the combined annualized percent growth in equipment investment and durable goods consumption.[10] These two variables correspond to Fernald's (2014) consumption and investment TFP series. The data are quarterly from 1947:Q2 to 2015:Q4. The VAR has a constant and 4 lags.

I construct the proxy variables as the residuals from the regression of Fernald's (2014) consumption and investment TFP series on the right-hand side variables of the VAR. This ensures that the proxies are mean zero and uncorrelated with the VAR's right-hand side variables as implied by Assumption 4. I denote the consumption and investment TFP

[9] Annualized percent growth in GDP is from the National Income and Product Accounts (NIPA) Table 1.1.1. For the other series, I use $g_t = (x_t/x_{t-1} - 1) \times 400$ to compute annualized percentage growth. Inflation is computed as an annualized growth percent from the price index for GDP from NIPA Table 1.1.4. Employment is defined as all employees in total private industries, and I use a quarterly average of monthly data from the FRED database before computing annualized percent growth.

[10] To combine these series, I start with annualized percent growth from NIPA Table 1.1.1. Then, I compute a weighted average based on the relative size of the sectors, using nominal series from NIPA Table 1.1.5 to compute the relative size.

Table 3: Correlations of proxies, structural shocks, and regression errors

$\text{corr}(z_t^C, z_t^I) =$	0.28 $(0.15, 0.40)$	$\text{corr}(\hat{v}_{1,t}^C, \hat{v}_{1,t}^I) =$	-0.01 $(-0.16, 0.13)$
$\text{corr}(z_t^C, \hat{v}_{1,t}^C) =$	0.59 $(0.50, 0.67)$	$\text{corr}(z_t^I, \hat{v}_{1,t}^I) =$	0.39 $(0.28, 0.50)$
$\text{corr}(z_t^C, \hat{v}_{1,t}^I) =$	-0.01 $(-0.15, 0.14)$	$\text{corr}(z_t^I, \hat{v}_{1,t}^C) =$	0.00 $(-0.13, 0.12)$
$\text{corr}(\hat{\epsilon}_t^C, \hat{\epsilon}_{t-1}^C) =$	0.02 $(-0.12, 0.11)$	$\text{corr}(\hat{\epsilon}_t^I, \hat{\epsilon}_{t-1}^I) =$	-0.08 $(-0.12, 0.11)$

Notes: z_t^C and z_t^I are the proxy variables for consumption and investment TFP. $\hat{v}_t^C$ and $\hat{v}_t^I$ are the estimated consumption and investment TFP structural shocks. $\hat{\epsilon}_t^C$ and $\hat{\epsilon}_t^I$ are the errors from the regressions of z_t^C and z_t^I on $\hat{u}_t$. 95% confidence intervals in parentheses are percentile intervals from an i.i.d. bootstrap with 10,000 replications.

proxies as z_t^C and z_t^I, respectively.

Before testing for proxy weakness or estimating IRFs, I first check that the assumptions of the proxy structural VAR model are satisfied. Many of these checks are summarized by the correlations in Table 3. The first column of the first row of Table 3 shows the correlation between the proxy variables is 0.28. Because the VAR has four lags, this statistic is computed on the 1948:Q2 to 2015:Q4 sample. In addition, Table 3 gives the 95% confidence interval in parentheses, which is the percentile intervals from an i.i.d. bootstrap with 10,000 replications. This confidence interval shows that the positive correlation of the proxy variables is statistically distinct from zero. However, this is not necessarily a problem because there is no assumption that different proxy variables have to be uncorrelated. Because of this, I follow Stock and Watson (2012) and estimate the structural shocks one at a time. Table 3 shows that the proxy variables, the estimated structural shocks, and the errors from the regressions of z_t^C and z_t^I on $\hat{u}_t$ are all consistent with the assumptions of the model. The second column of the first row of Table 3 shows that the estimated structural are uncorrelated with each other, satisfying Assumption 1.d. The second row of Table 3 shows that the proxy variables are correlated with their structural shocks, satisfying the relevance assumption. The third row of Table 3 shows that the proxies are not correlated with the other structural shock, satisfying the the exogeneity assumption. The fourth row of Table 3 shows that the first autocorrelation of the regression errors is not statistically significant. In addition, White (1980) tests for heteroskedasticity fail to reject that $\hat{\epsilon}_t^C$ and $\hat{\epsilon}_t^I$ are homoskedastic. Taken together, these last two results imply that Assumption 5 is satisfied.

Table 4: F and F_{IV} statistics for both proxy variables

Proxy	F	F_{IV}				
		GDP	Employment	Inflation	Consumption	Investment
z_t^C	28.1	66.0	0.0	7.9	13.2	17.2
z_t^I	9.7	19.3	49.4	1.4	4.9	8.7

Notes: The F statistic is given in Equation (29), and the F_{IV} statistic is given in Equation (31). The five columns under F_{IV} indicate the values of the F_{IV} statistic when the different VAR variables are ordered first.

Given the satisfaction of the model's assumptions, I test the strength of the proxy variables by computing the F statistic in Equation (29). In addition, I compute five different F_{IV} statistics from Equation (31) – one for each variable of the VAR ordered first. Table 4 displays the results for both proxies. I follow the weak IV literature and define my weak proxy set as containing any proxy variable that produces an asymptotic bias larger than 10% (Stock, Wright, and Yogo, 2002). From Table 1, this implies that the threshold signal-to-noise ratio is 18.40. From Table 2, the 10% critical F statistic is 7.12 and the 5% critical F statistic is 7.98. Table 4 shows that the F statistics for both of the proxies exceed these critical values. Thus, I reject the null hypothesis that the proxies are in the weak proxy set.

If the F_{IV} statistic were used, then inference for proxy strength would be ambiguous. For the consumption TFP proxy, F_{IV} ranges from 0 to 66, depending on the ordering of the VAR. The F_{IV} statistic does not vary as much for the investment TFP proxy, but still ranges from 1.4 to over 49. Given these results with the rule of thumb critical value of 10, then it is possible to either accept or reject the null hypothesis that the proxies are in the weak proxy set, depending on which variable is ordered first.

One way to get around this problem of ordering is to put the variable that most closely aligns with the shock of interest first in the VAR. For example, Gertler and Karadi (2015) order interest rates first when computing the F_{IV} statistic to test the strength of a monetary policy proxy. In the case of TFP shocks, when consumption is ordered first, then F_{IV} statistic is 13.2, and the null hypothesis of a weak proxy can be rejected using the rule of thumb critical value. However, when investment is ordered first, the F_{IV} statistic is only 8.7, which does not exceed the rule of thumb critical value. Hence, z_t^I would have been viewed as a weak proxy if the standard weak IV test had been applied.

To highlight the benefits of using the proxy approach, I consider two other methods for identifying structural shocks in VARs. First, I treat the proxy variables as the structural shocks themselves and estimate B_1 with least squares by including z_t as a right-hand-side variable. I call this the direct estimate of B_1. Second, I include Fernald's (2014) TFP measures as endogenous variables in the VAR and estimate B_1 with a Cholski decomposition of the covariance matrix of the VAR errors. I call this the Choleski estimate of B_1.[11]

[11] I assume Σ_v is the identity matrix with the Choleski estimate.

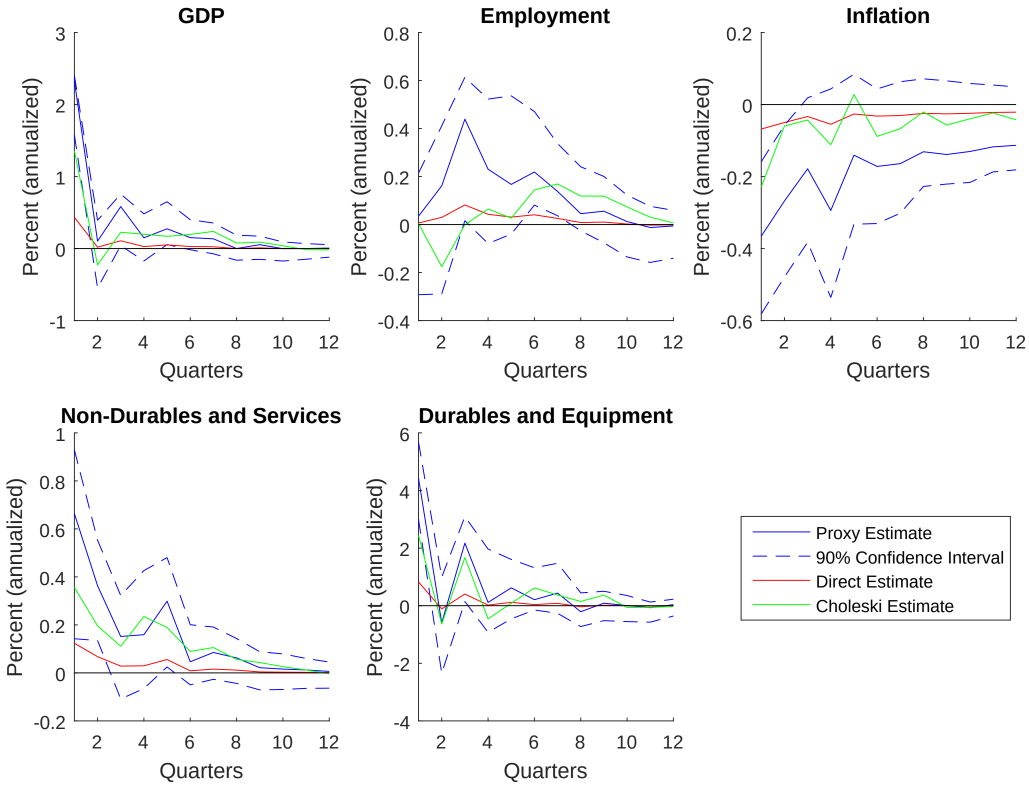

Figure 2: Dynamic effect of a positive TFP shock to consumption (excluding durables).

Figure 2 shows the dynamic effect of a positive one standard deviation TFP shock to consumption. The 90% confidence intervals are the percentile intervals from a residual-based moving block bootstrap algorithm described in Jentsch and Lunsford (2016) with 10,000 bootstrap replications. Figure 2 shows that a positive TFP shock to consumption causes an increase in GDP growth, a delayed increase in employment growth, and a decrease in inflation. Further, this shock causes an jump in growth in non-durables and services consumption and an increase durables consumption and equipment investment. These effects are theoretically consistent with a positive supply shock because they generate opposite movements in quantities and prices.

Figure 3 shows the dynamic effect of a positive one standard deviation TFP shock to investment. As in Figure 2, the 90% confidence intervals are percentile intervals from a residual-based moving block bootstrap algorithm with 10,000 bootstrap replications. Figure 3 shows that a positive TFP shock to investment causes immediate decreases in GDP growth, employment growth, and durables consumption and equipment investment growth. These drops are then followed by smaller booms in growth after about 4 quarters. Further, an increase in investment TFP causes a decrease in inflation. These effects are theoretically inconsistent with a standard supply shock because quantities and prices move in the same direction. Rather, Figure 3 looks like a negative demand shock. However, these results are consistent with the empirical findings in Basu et al. (2013). Basu, Fernald, and Liu (2012)

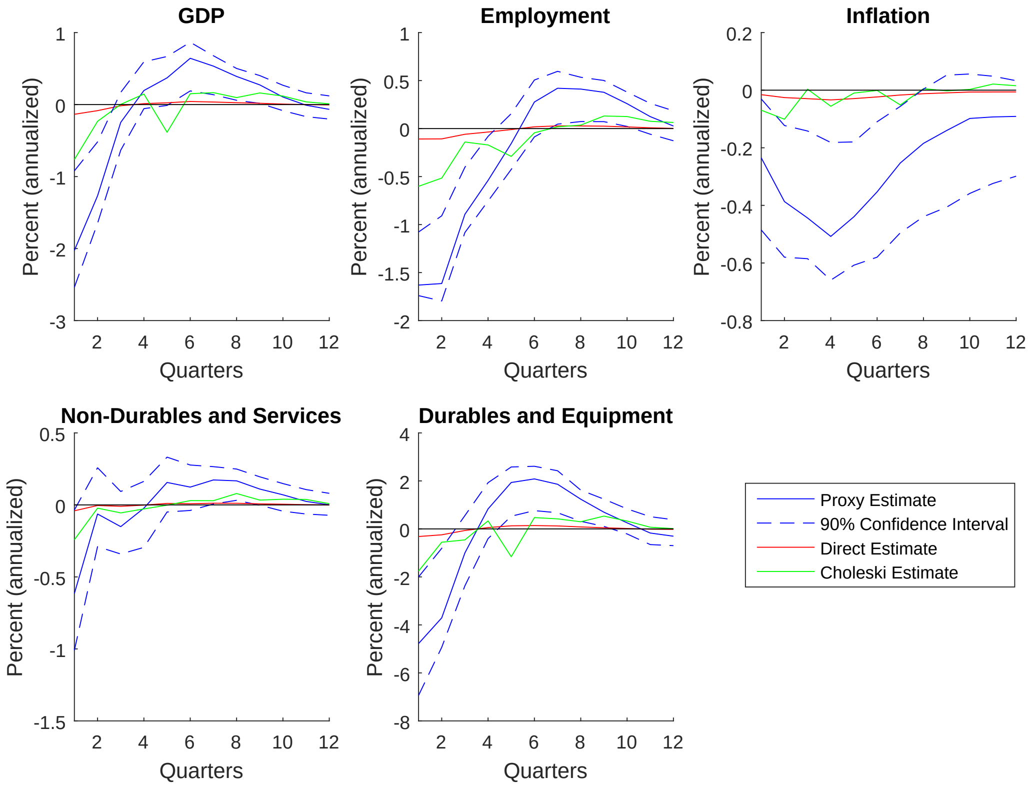

Figure 3: Dynamic effect of a positive TFP shock to durable consumption and equipment investment.

argue that sticky prices for investment goods in a dynamic stochastic general equilibrium model generate this negative demand response from an increase in investment TFP.[12]

In addition to showing the IRFs from the proxy variable approach. Figures 2 and 3 also show that the IRFs to positive one standard deviation TFP shocks when B_1 is estimated directly or with a Choleski decomposition. For both shocks, the estimate of B_1 is closer to zero with the direct and the Choleski estimates than with proxy estimate, yielding attenuated IRFs. For the direct estimate, the bias toward zero is easy to identify. Because z_t is uncorrelated with the other right-hand side variables in the VAR, the estimate of B_1 is $\mathbb{E}(u_t z_t)[\mathbb{E}(z_t^2)]^{-1}$. Equations (7), (9) and (21) imply that this estimate is $B_1\phi(\phi^2 + \sigma_\epsilon^2)^{-1}$. Thus, when $\phi > 1$ and the variance of ϵ_t is non-trivial, as it is for both proxies in this application, the direct estimate of B_1 will be biased toward zero.

For the Choleski estimate, the source of the downward bias of B_1 is more complicated, and I provide details in the appendix. The intuition is that including Fernald's (2014) TFP measures as endogenous variables introduces measurement error into the VAR. This means

[12]The intuition is that an increase in productivity in the investment sector causes an increase in the markup of investment goods with sticky prices. This causes investment goods to be expensive in the current period relative to future periods, suppressing current demand for investment goods.

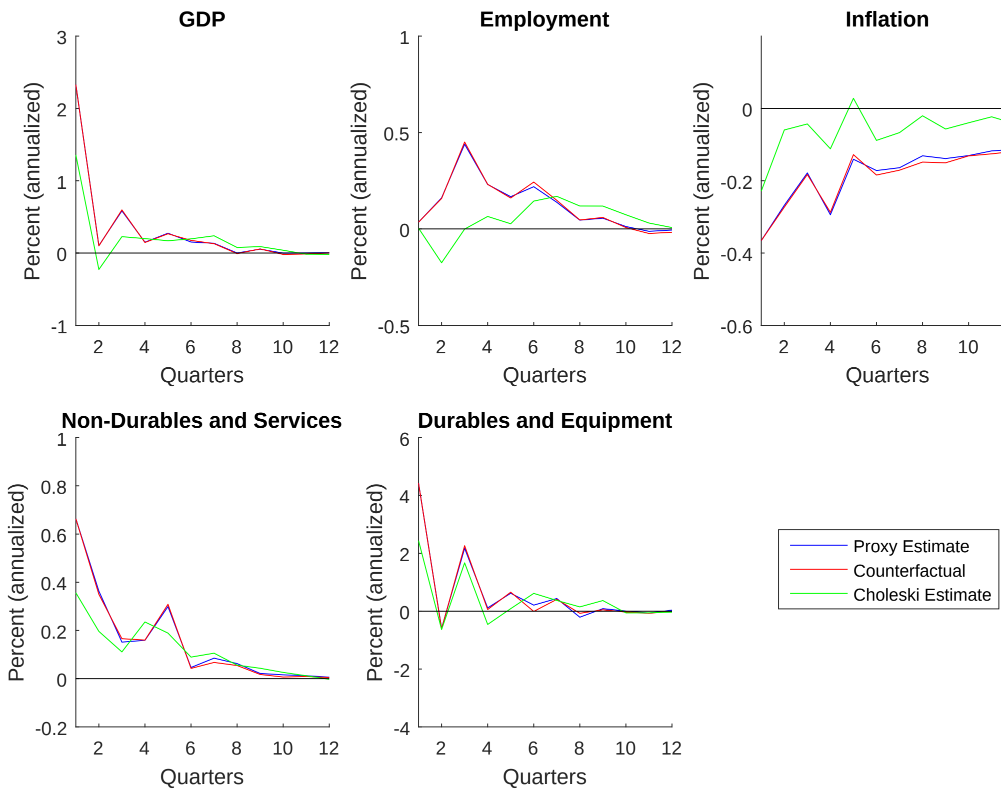

Figure 4: Dynamic effect of a positive TFP shock to consumption (excluding durables).

that the estimated VAR errors will be functions of ϵ_t in addition to v_t. Hence, a Choleski decomposition of the covariance matrix of the errors will not recover the parameters of interest. Figure 2 shows that the attenuation of $\hat{B}_1$ is particularly problematic for consumption TFP shocks. An ongoing debate in the study of TFP shocks is whether or not they initially increase or decrease hours worked.[13] With the proxy approach, Figure 2 shows that a positive consumption TFP shock increases employment growth. In contrast, the Choleski estimate indicates that a positive consumption TFP shock initially decreases employment growth and that employment growth only turns positive with a delay. Thus, these two approaches yield conflicting evidence on how TFP shocks impact the labor market.

This conflicting evidence is entirely driven by the different estimates of B_1 between the proxy and Choleski approaches. To see this, I compute counterfactual IRFs with the proxy variable estimate of B_1 and the Choleski VAR estimate of A. This yields IRFs that include Fernald's (2014) TFP measures as endogenous variables but do not rely on a Choleski decomposition to estimate B_1. Figure 4 displays these conuterfactual IRFs along with the proxy and Choleski estimates. It shows that these counterfactual IRFs lie on top of the proxy estimates for all five variables. Hence, the difference between the proxy and Choleski IRFs is not due to the VAR coefficients. Rather, the difference is driven by the Choleski

[13]For example, see section 5 of Ramey (2016).

VAR's inability to identify B_1.[14]

6 Conclusion

A proxy variable can identify a structural shock in a VAR, where a proxy variable is defined as being external from the VAR, correlated with the structural shock of interest, and uncorrelated with all other structural shocks. I provide a simple estimator for the impact of the structural shock of interest and show that this estimator is consistent when the proxy variable is strong. Next, I study the case of a weak proxy variable by assuming that the covariance between the proxy and the structural shock of interest is local-to-zero. With this assumption, the estimator for the impact of the structural shock of interest is inconsistent and converges in distribution to a function of normal random variables. Finally, I propose a test for a weak proxy based on the F statistic from the regression of the proxy variable onto the VAR errors. I give critical F values that depend on the level of statistical significance, asymptotic bias tolerance, and VAR dimension.

An important feature of F the statistic used in this paper is that its asymptotic distribution does not depend on parameters that need to be estimated. This contrasts with the F statistic from the weak IV literature, which has also been used to test for weak proxy variables. The asymptotic distribution of the F statistic from the weak IV literature is a function of the variance decomposition of the relevant VAR error. Because different empirical applications and different choices of the relevant VAR error will yield different variance decompositions, the limiting distribution of the F statistic from the weak IV literature will be different from application to application and for different VAR orderings. Thus, the critical values for this statistic cannot be taken from Stock and Yogo (2005); rather, they need to be computed for each application. Finally, the problem with computing these critical values is that a weak proxy makes them inconsistent and leads to mis-sized statistical tests.

To highlight the weak proxy test and the proxy variable approach more broadly, I estimate the effects of productivity shocks, using Fernald's (2014) measures of consumption TFP and investment TFP as the basis for my proxy variables. I find that both of these proxies yield structural shocks that satisfy the relevance and exogeneity assumptions. Using the F statistic from the regression of the proxy variables on the VAR errors, I show that both proxies are strong. In contrast, the F statistic from the weak IV literature gives ambiguous results.

This application also compares IRFs identified with the proxy approach to IRFs where the proxy is treated as the structural shock of interest and to IRFs where the proxy is included as an endogenous variable in a Choleski VAR. Both treating the proxy as the structural shocks of interest and including it in a Choleski VAR lead to biased IRFs. In the case of consumption TFP shocks, the Choleski VAR implies that a positive shock is initially contractionary for the labor market. This result contrasts with the proxy estimate, which shows that consumption TFP shocks are expansionary for the labor market. This

[14] This exercise for the investment TFP proxy yields similar results.

conflicting result is driven by different estimates for B_1, and it is the attenuation in the Choleski VAR's estimate that causes it to display a negative employment response. Hence, the proxy approach is not just for variables that contain dummies or missing observations. Rather, it should be used anytime there is a noisy measure of a structural shock.

This paper focuses on the case where one proxy variable is used to identify one structural shock. Future research should extend this analysis to study cases where multiple proxies exist for one structural shock. It should also study cases where researchers want to identify multiple structural shocks at once with multiple proxy variables as in Mertens and Ravn (2013) or Lakdawala (2016). Finally, future research should relax Assumption 5 to allow for heteroskedasticity when regressing the proxy variable on the VAR errors.

Appendix

Proof of Lemma 1 Equations (4) and (9) and the definitions of θ and $\tilde{\theta}$ imply

$$\tilde{\theta} = \Sigma_v^{1/2}\xi \quad \text{where} \quad \xi = (e_1 C/\sigma_\epsilon + x)[(e_1 C/\sigma_\epsilon + x)'(e_1 C/\sigma_\epsilon + x)]^{-1/2}.$$

Hence, $\mathbb{E}(\tilde{\theta}) = \Sigma_v^{1/2}\mathbb{E}(\xi)$. For $j \geq 2$, the jth element of ξ can be written as

$$\xi_j = \frac{x_j}{\sqrt{(C/\sigma + x_1)^2 + x_2^2 + \cdots + x_n^2}},$$

where x_j is the $j$the element of x. Because each element of x is an independent standard normal random variable, I re-write ξ_j as a function of x_j and λ,

$$\xi_j(x_j, \lambda) = x_j/\sqrt{x_j^2 + \lambda},$$

where λ is a random variable from the non-central χ^2_{n-1} with a non-centrality parameter of C^2/σ_ϵ^2, and x_j and λ are independent. This implies that the expectation of ξ_j is

$$\mathbb{E}(\xi_j) = \int_0^\infty \left[\int_{-\infty}^\infty \xi_j(x_j, \lambda) f(x_j) dx_j\right] g(\lambda) d\lambda, \tag{A.1}$$

where f is the probability density function (pdf) for a standard normal random variable, and g is the pdf for a non-central χ^2_{n-1} distribution with a non-centrality parameter of C^2/σ_ϵ^2. I show that the interior integral in Equation (A.1) is zero for any $\lambda \geq 0$, implying $\mathbb{E}(\xi_j) = 0$ for $j \geq 2$. This result and the form of Σ_v in Equation (4) imply $\mathbb{E}(\tilde{\theta}_j) = 0$ for $j \geq 2$.

First, fix some $\lambda > 0$. Then, $0 \leq \xi_j(x_j, \lambda) \leq 1$ for $x_j \geq 0$, and $-1 \leq \xi_j(x_j, \lambda) \leq 0$ for $x_j \leq 0$. These inequalities imply

$$0 \leq \int_0^\kappa \xi_j(x_j, \lambda) f(x_j) dx_j \leq \int_0^\kappa f(x_j) dx_j,$$

and

$$0 \geq \int_{-\kappa}^0 \xi_j(x_j, \lambda) dx_j \geq \int_{-\kappa}^0 -f(x_j) dx_j.$$

Because $\lim_{\kappa\to\infty} \int_0^\kappa f(x_j)dx_j = 1/2$ and $\lim_{\kappa\to\infty} \int_{-\kappa}^0 -f(x_j)dx_j = -1/2$, it is the case that

$$0 \leq \int_0^\infty \xi_j(x_j, \lambda) f(x_j) dx_j \leq \frac{1}{2} \quad \text{and} \quad 0 \geq \int_{-\infty}^0 \xi_j(x_j, \lambda) f(x_j) dx_j \geq -\frac{1}{2}.$$

Thus, the interior integral of Equation (A.1) is finite for all $\lambda > 0$. Then,

$$\begin{aligned}\int_{-\infty}^{\infty} \xi_j(x_j,\lambda) f(x_j) dx_j &= \int_{0}^{\infty} \xi_j(x_j,\lambda) f(x_j) dx_j + \int_{-\infty}^{0} \xi_j(x_j,\lambda) f(x_j) dx_j \\ &= \int_{0}^{\infty} \xi_j(x_j,\lambda) f(x_j) dx_j + \int_{0}^{\infty} \xi_j(-x_j,\lambda) f(-x_j) dx_j \\ &= \int_{0}^{\infty} \frac{x_j}{\sqrt{x_j^2+\lambda}} f(x_j) dx_j + \int_{0}^{\infty} \frac{-x_j}{\sqrt{x_j^2+\lambda}} f(x_j) dx_j = \int_{0}^{\infty} 0 \cdot f(x_j) dx_j = 0.\end{aligned}$$

In the event that $\lambda = 0$, $\xi_j(x_j,\lambda) = 1$ for $x_j > 0$, and $\xi_j(x_j,\lambda) = -1$ for $x_j < 0$. When $\lambda = 0$ and $x_j = 0$, then $\xi_j(x_j,\lambda)$ is undefined. However, I can apply

$$\lim_{\kappa \to 0^+} \int_{\kappa}^{\infty} \xi_j(x_j,0) f(x_j) dx_j = \lim_{\kappa \to 0^+} \int_{\kappa}^{\infty} f(x_j) dx_j = \frac{1}{2}$$

and

$$\lim_{\kappa \to 0^-} \int_{-\infty}^{\kappa} \xi_j(x_j,0) f(x_j) dx_j = \lim_{\kappa \to 0^-} \int_{-\infty}^{\kappa} -f(x_j) dx_j = -\frac{1}{2}.$$

Thus, the interior integral of Equation (A.1) is zero for all $\lambda \geq 0$.

Bias of Normalized Impulse Response Functions Consider the parameter $sB_1/(e_i'B_1)$. This is the initial response of an IRF that is normalized so that one of its elements is s. The estimate of this parameter is given by $s\hat{B}_1/(e_i'\hat{B}_1)$ which converges in distribution to $sB\theta/(e_i'B\theta)$. Without loss of generality, I assume that e_1 is used in the denominator so that the first element of the IRF is normalized to s. Then, partition B so that

$$B = \begin{bmatrix} \underset{(1\times 1)}{b_{11}} & \underset{(1\times n-1)}{b_{12}} \\ \underset{(n-1\times 1)}{b_{21}} & \underset{(n-1\times n-1)}{b_{22}} \end{bmatrix}.$$

In addition, for simplicity, I assume that $\Sigma_v = I_n$, which is standard in the structural VAR literature. Then, Assumption 5 implies

$$sB\theta/(e_1'B\theta) = s \begin{bmatrix} b_{11}C/\sigma_\epsilon + b_{11}x_1 + b_{12}x_2 \\ b_{21}C/\sigma_\epsilon + b_{21}x_1 + b_{22}x_2 \end{bmatrix} (b_{11}C/\sigma_\epsilon + b_{11}x_1 + b_{12}x_2)^{-1},$$

where x_1 is a standard normal random variable and x_2 is a $n-1\times 1$ vector of independent standard normal random variables. The first row of $sB\theta/(e_1'B\theta)$ reduces to s by construction.

However, the bottom $n-1$ rows of $sB\theta/(e_1'B\theta)$ can be written as

$$sb_{21}b_{11}^{-1} + s[(b_{11}b_{22} - b_{21}b_{12})x_2][b_{11}(b_{11}C/\sigma_\epsilon + b_{11}x_1 + b_{12}x_2)]^{-1}.$$

The first term of this expression is the parameter of interest. Then, the expectation of the second term is the bias of the bottom $n-1$ rows of $sB\theta/(e_1'B\theta)$. As with the bias of $\hat{B}_1$, the bias of $s\hat{B}_1/(e_1'\hat{B}_1)$ goes to zero as $C^2/\sigma_\epsilon^2 \to \infty$. However, as $C^2/\sigma_\epsilon^2 \to 0$ then the second term of the above expression converges to a vector of Cauchy random variables, and the bias of the bottom $n-1$ rows of $s\hat{B}_1/(e_1'\hat{B}_1)$ is undefined. This suggests that normalized IRFs can be very poorly behaved in the case of weak proxies.

The size of the bias of $s\hat{B}_1/(e_1'\hat{B}_1)$ will also depend on the elements of B. When $(b_{11}b_{22} - b_{21}b_{12})$ is near zero, then the bias might not be very severe. However, in case of $n=2$, $(b_{11}b_{22} - b_{21}b_{12})$ is the determinant of B, which is guaranteed not be zero by the assumption that B is invertible. Hence, at least for the case of $n=2$, bias will always be present.

Finally, the process of normalizing of IRFs requires the assumption that $b_{11} \neq 0$. Hence, researchers need to be careful about their choice of $u_{1,t}$. This is because if b_{11} is near zero, then a large signal-to-noise ratio of the proxy can be negated, and large biases in $s\hat{B}_1/(e_1'\hat{B}_1)$ can exist even for large values of C^2/σ_ϵ^2.

Attenuation Bias of a Choleski VAR To identify the source of the attenuation bias in a Choleski VAR, I consider a simple version of the proxy structural VAR model presented in this paper. First, the proxy variable follows $z_t = \phi v_{1,t} + \epsilon_t$. Next, the other variables follow $Y_t = A'Y_{t-1} + h_{21}v_{1,t} + h_{22}v_{2,t}$. Combined, the model is

$$\begin{bmatrix} z_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & A' \end{bmatrix} \begin{bmatrix} z_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} h_{11} & 0 \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t,$$

where $h_{11} = \phi$. This implies that the VAR innovations are given by

$$\zeta_t = \begin{bmatrix} h_{11} & 0 \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t.$$

When $\mathbb{E}(v_t v_t') = I_n$, then

$$\mathbb{E}(\zeta_t \zeta_t') = \begin{bmatrix} h_{11}^2 & h_{11}h_{21}' \\ h_{21}h_{11} & h_{22}h_{22}' \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \sigma_\epsilon^2$$

because $\mathbb{E}(v_t\epsilon_t) = 0$. With a Choleski decomposition, the estimate of h_{11} is given by $\tilde{h}_{11} = [\mathbb{E}(\zeta_{1,t}^2)]^{1/2}$, and the estimate of h_{21} is $\tilde{h}_{21} = \mathbb{E}(\zeta_{2,t}\zeta_{1,t})[\mathbb{E}(\zeta_{1,t}^2)]^{-1/2}$, where $\zeta_{1,t}$ is the first element of ζ and $\zeta_{2,t}$ are the remaining N elements of ζ. In this model, $\tilde{h}_{11} = (h_{11}^2 + \sigma_\epsilon^2)^{1/2}$, implying that the estimated value of h_{11} is too large. Conversely, $\tilde{h}_{21} = h_{21}h_{11}(h_{11}^2 + \sigma_\epsilon^2)^{-1/2}$, implying that the estimated value of h_{21} is too small. Because h_{21} identifies the effect of the structural shock of interest on Y_t, it is this small value of $\tilde{h}_{21}$ that yields the attenuation bias in the Choleski VAR.

This simple model provides a good empirical explanation for the attenuation of the Choleski estimates in this paper. Consider the case of consumption TFP shocks. With the proxy identification, $\hat{h}_{11} = \hat{\phi} = 1.9$ and $\hat{\sigma}_\epsilon^2 = 6.5$, yielding $(\hat{h}_{11}^2 + \hat{\sigma}_\epsilon^2)^{1/2} = 3.2$. Consistent with this, the Choleski identification yields $\tilde{h}_{11} = 3.1$. Thus, the model suggests that the Choleski identification over states the value of h_{11} by over 60% because of the introduction of ϵ_t into the VAR innovations. Further, with the proxy identification, $\hat{h}_{21} = [2.3, 0.0, -0.4, 0.7, 4.4]'$ so that $\hat{h}_{21}\hat{h}_{11}(\hat{h}_{11}^2 + \hat{\sigma}_\epsilon^2)^{-1/2} = [1.4, 0.0, -0.2, 0.4, 2.6]'$. Consistent with this, the Choleski identification yields $\tilde{h}_{21} = [1.4, 0.0, -0.2, 0.4, 2.4]'$.

Next, consider the case of the investment TFP shocks. With the proxy identification, $\hat{h}_{11} = \hat{\phi} = 2.3$ and $\hat{\sigma}_\epsilon^2 = 29.6$, yielding $(\hat{h}_{11}^2 + \hat{\sigma}_\epsilon^2)^{1/2} = 5.9$. Consistent with this, the Choleski identification yields $\tilde{h}_{11} = 5.8$. Thus, the model suggests that the Choleski identification over states the value of h_{11} by nearly 150% because of the introduction of ϵ_t into the VAR innovations. Further, with the proxy identification, $\hat{h}_{21} = [-2.0, -1.6, -0.2, -0.6, -4.8]'$ so that $\hat{h}_{21}\hat{h}_{11}(\hat{h}_{11}^2 + \hat{\sigma}_\epsilon^2)^{-1/2} = [-0.8, -0.6, -0.1, -0.2, -1.9]'$. Consistent with this, the Choleski identification yields $\tilde{h}_{21} = [-0.8, -0.6, -0.1, -0.2, -1.8]'$.

References

Basu, Susanto, John Fernald, Jonas Fisher, and Miles Kimball. 2013. "Sector-Specific Technical Change." Working Paper, Federal Reserve Bank of San Francisco.

Basu, Susanto, John Fernald, and Zheng Liu. 2012. "Technology Shocks in a Two-Sector DSGE Model." Society for Economic Dynamics, 2012 Meeting Paper 1017.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3):623–685.

Carriero, Andrea, Haroon Mumtaz, Konstantinos Theodoridis, and Angeliki Theophilopoulou. 2015. "The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach." *Journal of Money, Credit and Banking* 47 (6):1223–1238.

Christiano, Lawrence J., Martin Eichenbauma, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1):1–45.

Fernald, John. 2014. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." Working Paper, Federal Reserve Bank of San Francisco.

Gertler, Mark and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1):44–76.

Gürkaynak, Refet S, Brian Sack, and Eric T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1):55–93.

Jentsch, Carsten and Kurt G. Lunsford. 2016. "Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States." Federal Reserve Bank of Cleveland, Working Paper no. 16-19.

Lakdawala, Aeimit K. 2016. "Decomposing the Effects of Monetary Policy using an External Instruments SVAR." Working Paper, Michigan State University.

Lunsford, Kurt G. 2015. "Housing Search and Fluctuations in Residential Construction." Job Market Paper, University of Wisconsin – Madison.

Mertens, Karel and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103 (4):1212–1247.

———. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68 (Supplement):S1–S19.

Montiel Olea, José Luis and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3):358–369.

Montiel Olea, José Luis, James H. Stock, and Mark W. Watson. 2012. "Inference in Structural VARs with External Instruments." Presentation Slides, Harvard University.

———. 2016. "Uniform Inference in SVARs with External Instruments." Working Paper, New York University.

Mumtaz, Haroon, Gabor Pinter, and Konstantinos Theodoridis. 2015. "What Do VARs Tell Us About the Impact of a Credit Supply Shock?" Working Paper No. 739 Queen Mary University of London.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propogation." NBER Working Paper No. 21978.

Romer, Christina D. and David H. Romer. 2009. "A Narrative Analysis of Postwar Tax Changes." Unpublished manuscript, University of California, Berkeley.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1):1–48.

Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3):557–586.

Stock, James H. and Mark W. Watson. 2008. "Lecture 7 - Recent Developments in Structural VAR Modeling." Presentd at the NBER Summer Institute Econometrics Lectures: What's New in Economics - Time Series, July 15.

———. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity* Spring:81–135.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20 (4):518–529.

Stock, James H. and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models*, edited by Donald W.K. Andrews. Cambridge University Press, 80–108.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4):817–838.