working paper

15 28

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

Kurt G. Lunsford

FEDERAL RESERVE BANK OF CLEVELAND

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Banks of Cleveland or Boston or of the Board of Governors of the Federal Reserve System.

Working papers are available on the Cleveland Fed's website at:

www.clevelandfed.org.

Working Paper 15-28 December 2015

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

Kurt G. Lunsford

This paper develops a simple estimator to identify structural shocks in vector autoregressions (VARs) by using a proxy variable that is correlated with the structural shock of interest but uncorrelated with other structural shocks. When the proxy variable is weak, modeled as local to zero, the estimator is inconsistent and converges to a distribution. This limiting distribution is characterized, and the estimator is shown to be asymptotically biased when the proxy variable is weak. The F statistic from the projection of the proxy variable onto the VAR errors can be used to test for a weak proxy variable, and the critical values for different VAR dimensions, levels of asymptotic bias, and levels of statistical significance are provided. An important feature of this F statistic is that its asymptotic distribution does not depend on parameters that need to be estimated.

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

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

Suggested citation: Lunsford, Kurt G., 2015. "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy," Federal Reserve Bank of Cleveland, Working Paper no. 15-28.

Kurt G. Lunsford is at the Federal Reserve Bank of Cleveland and can be reached at kurt.lunsford@clev.frb.org. The author thanks Todd E. Clark, Joachim Freyberger, Bruce E. Hansen, Lutz Kilian, Tatevik Sekhposyan, Xiaoxia Shi, Kenneth D. West, Sarah Zubairy, and seminar participants at Texas A&M University and the University of Wisconsin–Madison for helpful comments and discussions.

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.

Some popular approaches for structural identification, such as zero and sign restrictions, estimate B_1 by reducing the set of allowable values of the elements of B. In this paper, I take an alternative approach. Following Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013), I use a variable external from the VAR in Equation (1) as a proxy for the structural shock of interest. Specifically, I assume that the proxy variable is correlated with the structural shock of interest and uncorrelated with the other structural shocks.[1] This proxy variable allows me to estimate B_1 without placing any restrictions on the values of the elements of B.

Recent research suggests that this proxy variable approach has good empirical properties. Carriero et al. (2015) argue that it has less downward bias due to measurement error when estimating the impact of Bloom's (2009) uncertainty shocks compared to putting the shocks directly into the VAR, and Mumtaz, Pinter, and Theodoridis (2015) show that it does better in matching the effects credit supply shocks from a dynamic stochastic general equilibrium model than a Choleski decomposition. Further, the proxy variable approach can be used to identify a wide variety structural shocks and proxies can come from a wide variety of sources. As examples, Mertens and Ravn (2013) follow the narrative approach of Romer and Romer (2009) to construct proxy variables for tax shocks, Gertler and Karadi (2015) follow the high frequency approach of Gürkaynak, Sack, and Swanson (2005) to construct proxy variables for monetary policy shocks, and Montiel Olea, Stock, and Watson (2012) and Stock and Watson (2012) use 18 different proxies to identify shocks to oil, monetary policy, productivity, uncertainty, liquidity and financial risk, and fiscal policy. Finally, Mertens and Ravn (2014) show that it can be used to reconcile the differences between structural VAR

[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 "instrumental variable." 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."

and narrative estimates of tax multipliers; however, Kliem and Kriwoluzky (2013) argue that it is not able to reconcile structural VAR and narrative estimates of monetary policy shocks.

Given this proliferation in the use proxy variables, it is important to ask how relevant these proxies are for estimating the effects of the structural shocks of interest. Thus, the objectives of this paper are to answer the following questions. Can a weakly relevant proxy reliably estimate B_1? What statistic can be used to test the weakness of a 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 is important because it will make the analysis of weak proxy variables tractable.

The second step of this paper is to characterize the asymptotic limit of the estimate of B_1 under the assumption that the proxy variable is weak. To do this, I follow Staiger and Stock (1997) and model a weak proxy as local to zero by assuming that the covariance between the proxy and the structural shock of interest goes to zero at a rate of square root of the sample size. Given this assumption, the estimate of B_1 is not consistent. Rather, it converges in distribution to a function of normal random variables. While this limiting distribution is not the same as that in the weak instrumental variables (IV) literature, many characteristics of the limiting distribution can be summarized by a concentration parameter similar to the concentration parameter in Stock, Wright, and Yogo (2002), and Stock and Yogo (2005). As the concentration parameter increases, the asymptotic distribution collapses to B_1. This implies that even when the local to zero assumption holds, a proxy variable can provide a close estimate of B_1 as long as the concentration parameter is sufficiently large. To get a measure of how close this estimate is, I follow the weak IV literature and use the asymptotic bias of the estimate of B_1. This asymptotic bias decreases as the concentration parameter increases, and weak proxy sets, defined as the set of proxy variables with concentration parameters below a given threshold, can be produced based on a researcher's tolerance for bias. Using simulation, I provide concentration parameter thresholds for asymptotic biases of 20%, 10%, 5% and 1%, and for VARs of dimensions between 2 and 20.[2]

Because the concentration parameter cannot be directly estimated, the third step of this paper is to provide a test for a weak proxy variable, which is based on an F statistic. An important point of distinction from the weak IV literature is which equation the F statistic is derived from. The weak IV literature uses the F statistic from the first-stage or the reduced form equation where the endogenous variables of interest are linearly projected on the instruments. In contrast, the F statistic in this paper comes from the linear projection

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

of the proxy on the VAR errors. That is, the proxy variable is the dependent variable when computing the F statistic. I show that n times this F statistic converges in distribution to a non-central χ^2_n distribution, where the concentration parameter is the non-centrality parameter. Thus, if the F statistic is sufficiently large, then there is a low probability that the true concentration parameter is below the weak proxy set threshold, and the null hypothesis that the proxy variable is in the weak proxy set can be rejected. Based on the computed concentration parameter thresholds noted above, I provide critical F statistic values for asymptotic biases of 20%, 10%, 5% and 1%, for VARs of dimensions between 2 and 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 10% asymptotic bias at a 5% level of significance.

An appealing feature of this F statistic is that its limiting distribution does not depend on parameters of the model that have to be estimated. This feature is important because previous tests of proxy strength have followed the weak IV literature and been based on an F statistic where one of the VAR errors is projected onto the proxy variable (Montiel Olea, Stock, and Watson, 2012; Gertler and Karadi, 2015; Lunsford, 2015). I show that the limiting distribution this F statistic is a function of 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 derived in the weak IV literature (Stock and Yogo, 2005). Because of this, rules of thumb from the weak IV literature, such as requiring that an F statistic be grater 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 this framework and the theory in this paper is needed to test for weak proxies.

The F statistic studied in the paper has been applied by Montiel Olea, Stock, and Watson (2012) and Stock and Watson (2012). They find that only five of their 18 proxy variables have F statistics that exceed 7.81, which is the 5% critical value for 10% asymptotic bias.[3] This result suggests that weak proxy variables are prevalent and that the theory for testing proxy weakness in this paper will be important in future applied work.

As an application of the theory developed in this paper, I study the dynamic effects of productivity shocks. To do this, I use Fernald's (2014) measures of utilization-adjusted total factor productivity (TFP) for the consumption and investment sectors as proxy variables. The F statistics for the consumption and investment TFP shocks are 26.26 and 9.01, respectively. Thus, I reject the null hypothesis that these are weak proxy variables. Further, I find that these proxies yield structural shocks that are serially uncorrelated, mutually uncorrelated, correlated with their corresponding proxy variable, and uncorrelated with the

[3]This critical value can be found in Table 2 below and is based on the fact that Montiel Olea, Stock, and Watson (2012) and Stock and Watson (2012) use a VAR with 6 variables.

other TFP proxy variable. Hence, Fernald's (2014) measures of consumption and investment TFP yield structural shocks that are consistent with the econometric theory below. Given this, I show that an increase in consumption TFP causes an immediate increase in output growth, an immediate decrease in inflation, and a delayed increase in private employment growth. These responses are consistent with standard economic theory that a supply-side shock causes quantities and prices to move in opposite directions. In contrast, a positive shock to investment TFP causes an immediate drop in output growth and private employment growth along with a hump-shaped decline in inflation. While these empirical responses are similar to those found in Basu et al. (2013), they are not consistent with a positive supply-side shock. Rather, they are consistent with a negative demand-side shock. To provide a theory for these puzzling responses, Basu, Fernald, and Liu (2012) solve a dynamic stochastic general equilibrium model where investment goods have sticky prices. They show that a positive TFP shock to investment production increases the mark-up of investment goods. This reduces current demand for investment, suppressing both output and inflation in the short-run. Finally, I show that the two TFP shocks are very important for aggregate fluctuations. From 1948:Q1 to 2015:Q2, they contributed nearly 80% of the variance of GDP growth. In addition, the investment TFP shock alone contributed over 70% of the variance in private employment growth.

The remainder of the paper is as follows. Section 2 lays out the assumptions for the VAR and the proxy variable, 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 that has been used in the literature up to this point 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 the structural shocks so that the first element of v_t is the shock of interest. Then, Equation (2) is

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

so that $v_{1,t}$ is the shock of interest, and $v_{2,t}$ contains the other $n-1 \times 1$ 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 finite, symmetric, positive-definite and invertible. 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 $t \neq s$.

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 a finite mean $\mathbb{E}(z_t) = \mu_z$.

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

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

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

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

These relevance and exogeneity assumptions mirror those assumptions used in the IV literature. However, it is important to note that the proxy here is serving a different purpose than an IV. In this model, the econometric problem is not that B_1 cannot be consistently estimated 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 estimated 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 towards a different end. Equations (5) and (6), along with the partition in Equation (3), imply

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

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

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

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

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

Thus, given the variance of the structural shock of interest, the covaraincs of the VAR innovations and the proxy variable can be used to 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

$$\mathbb{E}(v_{1,t}^2) = \sigma_{v_1}^2 = 1. \tag{9}$$

This normalization has no impact on the impulse response functions 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

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

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]

To recover the structural shock of interest, I follow Stock and Watson (2012) and linearly project z_t onto u_t and a constant. Because z_t is correlated with $v_{1,t}$ but not with $v_{2,t}$, this projection will return the component of u_t that is driven by $v_{1,t}$ without influence from $v_{2,t}$. Specifically, I use

$$z_t = \mu_z + u_t'\pi + \epsilon_t, \tag{11}$$

where π is an $n \times 1$ vector. I make the following assumptions about ϵ_t.

Assumption 3:

a) $\mathbb{E}(\epsilon_t) = 0$.

b) $\mathbb{E}(\epsilon_t^2) = \sigma_\epsilon^2$, where σ_ϵ^2 is finite.

c) $\mathbb{E}(\epsilon_t \epsilon_s) = 0$ for $t \neq s$.

d) ϵ_t is independent of lags of Y_t so that $\mathbb{E}(Y_{t-j}\epsilon_t) = 0$ for $j \geq 1$.

e) ϵ_t is independent of v_t.

Then, the population projection of $z_t - \mu_z$ on u_t is given by

$$\pi = \left[\mathbb{E}(u_t u_t')\right]^{-1} \mathbb{E}[u_t(z_t - \mu_z)]. \tag{12}$$

Using this projection, the expectation of z_t conditional on u_t from Equation (12) yields an estimate of $v_{1,t}$ up to a constant and a scalar. That is,

$$\begin{aligned}
\mathbb{E}(z_t|u_t) &= \mu_z + u_t'\pi \\
&= \mu_z + u_t' \left[\mathbb{E}(u_t u_t')\right]^{-1} \mathbb{E}[u_t(z_t - \mu_z)] \\
&= \mu_z + v_t' B'(B\Sigma_v B')^{-1} B_1 \phi \\
&= \mu_z + v_t' \Sigma_v^{-1} e_1 \phi \\
&= \mu_z + v_{1,t}\phi,
\end{aligned}$$

where $e_1 = [1, 0, \ldots, 0]'$. 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 we could linearly project z_t on $v_{1,t}$. I summarize the above list of equations as

$$u_t'\pi = v_{1,t}\phi. \tag{13}$$

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

Rewriting Equation (13) 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 (12) 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 $X = [1, Y_{t-1}', \ldots, Y_{t-p}']'$ so that Equation (1) can be written as $Y_t' = X_t'A + u_t'$. Next, I define $\bar{z} = \sum_{t=1}^{T} z_t$ to be the sample average of the proxy variable. Then, I define the following matrices

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

$$\underset{(T\times 1)}{Z} = \begin{bmatrix} z_1 \\ \vdots \\ z_T \end{bmatrix} \quad \underset{(T\times 1)}{\bar{Z}} = \begin{bmatrix} \bar{z} \\ \vdots \\ \bar{z} \end{bmatrix} \quad \underset{(T\times 1)}{M_z} = \begin{bmatrix} \mu_z \\ \vdots \\ \mu_z \end{bmatrix} \quad \underset{(T\times 1)}{E} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} \quad \underset{(T\times 1)}{\hat{E}} = \begin{bmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_T \end{bmatrix},$$

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

$$\hat{A} = (X'X)^{-1}X'Y, \tag{14}$$

with VAR errors given by

$$\hat{U} = Y - X\hat{A}. \tag{15}$$

Further, I estimate π by least squares so that

$$\hat{\pi} = (T^{-1}\hat{U}'\hat{U})^{-1}[T^{-1}\hat{U}'(Z - \bar{Z})] \tag{16}$$

and

$$\hat{E} = (Z - \bar{Z}) - \hat{U}'\hat{\pi}. \tag{17}$$

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

$$\widehat{B_1\phi} = T^{-1}\hat{U}'(Z - \bar{Z}) \tag{18}$$

and

$$\hat{\phi}^2 = [T^{-1}(Z - \bar{Z})'\hat{U}](T^{-1}\hat{U}'\hat{U})^{-1}[T^{-1}\hat{U}'(Z - \bar{Z})]. \tag{19}$$

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

$$\hat{B}_1 = \pm[T^{-1}\hat{U}'(Z - \bar{Z})]\{[T^{-1}(Z - \bar{Z})'\hat{U}](T^{-1}\hat{U}'\hat{U})^{-1}[T^{-1}\hat{U}'(Z - \bar{Z})]\}^{-1/2}, \tag{20}$$

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

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

Assumption 4:

a) $T^{-1}X'X \xrightarrow{p} \mathbb{E}(X_t X_t')$.

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

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

d) $T^{-1}\sum_{t=1}^{T} u_t \xrightarrow{p} \mathbb{E}(u_t)$.

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

f) $T^{-1}\sum_{t=1}^{T} \epsilon_t \xrightarrow{p} \mathbb{E}(\epsilon_t)$.

Next, note that Equation (18) can be rewritten as

$$\begin{aligned}\widehat{B_1\phi} = T^{-1}U'(M_z - \bar{Z}) + (T^{-1}U'X)(T^{-1}X'X)^{-1}[T^{-1}X'(M_z - \bar{Z})] \\ + T^{-1}U'U\pi + (T^{-1}U'X)(T^{-1}X'X)^{-1}(T^{-1}X'U)\pi \\ + T^{-1}U'E + (T^{-1}U'X)(T^{-1}X'X)^{-1}(T^{-1}X'E).\end{aligned} \tag{21}$$

Then, Assumptions 1, 2, 3, and 4 along with Equation (12) and the continuous mapping theorem imply $\widehat{B_1\phi} \xrightarrow{p} \mathbb{E}[u_t(z_t - \mu_z)] = B_1\phi$. Similarly, Assumptions 1, 2, 3, and 4, Equation (12) and the continuous mapping theorem imply $\hat{\phi}^2 \xrightarrow{p} \mathbb{E}[(z_t - \mu_z)u_t'][\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}[u_t(z_t - \mu_z)] = \phi^2$. Finally, from these results, the continuous mapping theorem, and Equations (10) and (20), 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 the researcher chooses the positive sign in Equation (20), and when they intend for this correlation to be negative ($\phi < 0$) then they choose the negative sign. In either case, Equation (20) provides a consistent estimator for B_1.

After estimating ϕ and applying the appropriate sign, I can also estimate the structural shock of interest, $v_{1,t}$. To do this, I use

$$\hat{v}_{1,t} = \hat{u}_t'\hat{\pi}\hat{\phi}^{-1} \tag{22}$$

from Equation (13), noting that Assumptions 1, 2, 3, and 4 along with the continuous mapping theorem imply $\hat{\pi} \xrightarrow{p} [\mathbb{E}(u_t u_t')]^{-1}\,\mathbb{E}[u_t(z_t - \mu_z)] = \pi$.

3 Testing for a Weak Proxy Variable

I model a weak proxy variable comparably to a weak instrumental variable by following Staiger and Stock (1997) and treating ϕ as being local to zero:

$$\phi = \phi_T = C/\sqrt{T}. \tag{23}$$

Then, the analysis of a weak proxy variable is based on the projection of the proxy on the VAR innovations in Equation (11). In addition, I make one additional assumption.

Assumption 5:

1. $T^{-1/2}U'E \xrightarrow{d} N(0, \mathbb{E}(u_t\epsilon_t^2 u_t'))$.

Given Equation (23) and this assumption, the following subsections show that $\hat{B}_1$ is no longer a consistent estimator for B_1. Rather it converges in distribution to a function of normal random variables. Further, I discuss the bias of this distribution and show that it is small when a parameter similar to the concentration parameter in the weak IV literature is large. Finally, I show that the F statistic on the null hypothesis that $\pi = 0$ in Equation (11) can provide a test that this parameter is large.

3.1 Inconsistency of the B_1 Estimator

Given the weak proxy assumption in Equation (23), it is the case that Equation (12) implies

$$\pi = (B\Sigma_v B')^{-1} B_1 C/\sqrt{T}. \tag{24}$$

Then, Assumptions 1, 2, 3, 4 and 5, Equations (23) and (24), the continuous mapping theorem and Slutsky's theorem imply

$$\sqrt{T}\widehat{B_1\phi} \xrightarrow{d} B\Sigma_v^{1/2}\sigma_\epsilon\theta \tag{25}$$

where $\widehat{B_1\phi}$ is taken from Equation (21),

$$\theta \sim e_1\frac{C}{\sigma_\epsilon} + N(0, I_n), \tag{26}$$

and I_n is the $n \times n$ identity matrix. Then, from Equations (18) and (20), $\hat{B}_1$ 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

$$\hat{B}_1 \xrightarrow{d} \pm B\Sigma_v^{1/2}\theta(\theta'\theta)^{-1/2}. \tag{27}$$

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 a weak instrument, it is the case that $\hat{B}_1$ is no longer a consistent estimator for B_1. Rather, $\hat{B}_1$ converges in distribution to a linear transformation of the random vector θ divided by the Euclidean norm of θ.

Given the definition of θ in Equation (26), it is the case that $\theta(\theta'\theta)^{-1/2} \to \pm e_1$ as $C^2/\sigma_\epsilon^2 \to \infty$. Applying this result to Equation (27) along with the appropriate sign on $\hat{B}_1$ implies that the asymptotic distribution of $\hat{B}_1$ collapses to B_1 as $C^2/\sigma_\epsilon^2 \to \infty$.[5] Thus, $\hat{B}_1$ becomes an increasingly accurate estimator as C^2/σ_ϵ^2 becomes large, and C^2/σ_ϵ^2 is comparable to the concentration parameter in the weak IV literature (Stock, Wright, and Yogo, 2002; Stock and Yogo, 2005). Because of this, I refer to C^2/σ_ϵ^2 as the concentration parameter and use it to construct the weak proxy set. First, I define $(C^2/\sigma_\epsilon^2)^*$ to be a threshold value. Then, all proxy variables with concentration parameters in the interval $[0, (C^2/\sigma_\epsilon^2)^*]$ compose the weak proxy set. Conversely, all proxy variables with concentration parameters in the interval $((C^2/\sigma_\epsilon^2)^*, \infty)$ are deemed strong. Following the weak IV literature, I set the threshold value $(C^2/\sigma_\epsilon^2)^*$ so that proxies that are not in the weak proxy set have small asymptotic biases. I discuss the asymptotic bias of $\hat{B}_1$ and threshold concentration parameters further in the next subsection.

3.2 Asymptotic Bias of B_1

To characterize the asymptotic bias of $\hat{B}_1$ with a weak proxy, I use the following lemma.

Lemma 1 *Define the $n \times 1$ random vector $\tilde{\theta} = \theta(\theta'\theta)^{-1/2}$ 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 this definition of $\tilde{\theta}$, Equation (27) implies that $\hat{B}_1$ converges to the random vector $\pm B\Sigma_v^{1/2}\tilde{\theta}$ as $T \to \infty$. Define $b = \mathbb{E}(\tilde{\theta}_1)$. Then, Lemma 1 implies that

$$\mathbb{E}(B\Sigma_v^{1/2}\tilde{\theta}) = B_1 b.$$

Thus, the expectation of the first element of $\tilde{\theta}$ characterizes the asymptotic bias of $\hat{B}_1$. This element is a function of C/σ_ϵ and n standard normal random variables, and so its expectation can be fully characterized by the parameters C^2/σ_ϵ^2 and n.

To characterize b, I use a simulation. First, I fix a value of C^2/σ_ϵ^2. Second, I draw 10,000 observations of the random vector θ. Third, I compute $\tilde{\theta}_1$ and average over the 10,000 observations, which I take to be $\mathbb{E}(\tilde{\theta}_1)$. 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.

[5] 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).

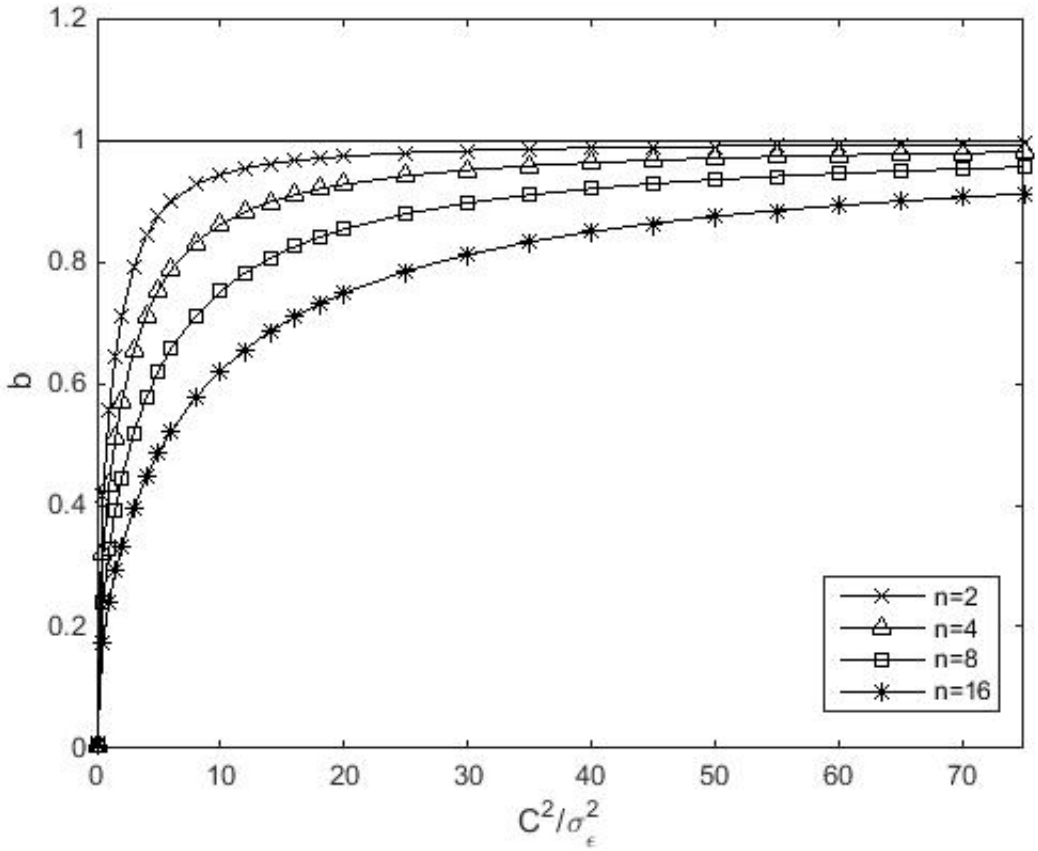

Figure 1: b as a function of C^2/σ_ϵ^2 and n.

Figure 1 shows that b is bounded between 0 and 1, implying that the asymptotic bias from a weak proxy biases $\hat{B}_1$ 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 to 1 quickly 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 asymptotic bias.

Figure 1 implies that for a given asymptotic bias tolerance, $1 - b$, and for a given VAR dimension n, a researcher can find a minimum value of the concentration parameter so that the the asymptotic bias tolerance is not exceeded. This minimum value of C^2/σ_ϵ^2 is given by $(C^2/\sigma_\epsilon^2)^*$ and characterizes the weak proxy set discussed in the previous subsection. To compute different values of $(C^2/\sigma_\epsilon^2)^*$ for different values of b and n, I use 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 θ.

5. Use the draws of θ to compute 100,000 observations of $\tilde{\theta}_1$ and compute $\bar{b}$ to be the average of the 100,000 observations of $\tilde{\theta}_1$.

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 minimum bound equal to the guess of C/σ_ϵ. If not, then set the maximum 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 instrument set increases with b and n. However, because C^2/σ_ϵ^2 is not directly observable, 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.

3.3 A Test for a Weak Instrument

The objective of this subsection is to provide test for whether or not z_t is in the weak proxy set. Formally, I test

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

To do this, I use the F statistic for the null hypothesis that $\pi = 0$ in Equation (11). This is given by

$$\begin{aligned} F &= \left(\frac{T-n}{n}\right) \frac{(Z-\bar{Z})'(Z-\bar{Z}) - [(Z-\bar{Z}) - \hat{U}\hat{\pi}]'[(Z-\bar{Z}) - \hat{U}\hat{\pi}]}{[(Z-\bar{Z}) - \hat{U}\hat{\pi}]'[(Z-\bar{Z}) - \hat{U}\hat{\pi}]} \\ &= \frac{1}{n}\left(\frac{T-n}{T}\right) \frac{[T^{-1/2}(Z-\bar{Z})'\hat{U}](T^{-1}\hat{U}'\hat{U})^{-1}[T^{-1/2}\hat{U}'(Z-\bar{Z})]}{T^{-1}\hat{E}'\hat{E}}. \end{aligned} \tag{28}$$

Then under the weak proxy assumption in Equation (23),

$$F \xrightarrow{d} n^{-1}\theta'\theta, \tag{29}$$

[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.

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

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

which follows from Equation (25), $T^{-1}\hat{U}'\hat{U} \xrightarrow{p} B\Sigma_v B'$, $T^{-1}\hat{E}'\hat{E} \xrightarrow{p} \sigma_\epsilon^2$, the continuous mapping theorem and Slutsky's theorem. Equation (29) 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 (28).
4. Compute the $\Pr(X \leq nF)$ 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 computed threshold values in Table 1, I compute the corresponding F statistics for α equal to 0.90, 0.95 and 0.99. These F statistics are reported in Table 2.

As in the IV literature, the threshold F statistics are large. For example, if one wants to reject the threshold concentration parameter 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, if the threshold concentration parameter was zero so that the asymptotic distribution of nF was a central χ_n^2, 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. For example, 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 (2014) 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 $\sigma_{v_1}^2 = 1$ as in Equation

Table 2: Critical Values for the F Statistic

	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
n	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$	$b = 0.80$	$b = 0.90$	$b = 0.95$	$b = 0.99$
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 projection of the de-meaned proxy variable onto 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.

(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

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

and

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

which combine to yield

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

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). For example, 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 instrumental variables where $z_t - \mu_z$ is an instrument for $u_{1,t}$. The corresponding structural and reduced form equations for this estimate are

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

and

$$u_{1,t} = \gamma(z_t - \mu_z) + \eta_{2,t}.$$

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

$$\begin{aligned} F_{IV} &= (T-1)\frac{\hat{U}_1'\hat{U}_1 - [\hat{U}_1 - (Z-\bar{Z})\hat{\gamma}]'[\hat{U}_1 - (Z-\bar{Z})\hat{\gamma}]}{[\hat{U}_1 - (Z-\bar{Z})\hat{\gamma}]'[\hat{U}_1 - (Z-\bar{Z})\hat{\gamma}]} \\ &= \left(\frac{T-1}{T}\right)\frac{[T^{-1/2}\hat{U}_1'(Z-\bar{Z})][T^{-1}(Z-\bar{Z})'(Z-\bar{Z})]^{-1}[T^{-1/2}(Z-\bar{Z})'\hat{U}_1]}{T^{-1}\hat{U}_1'U_1 - [T^{-1}\hat{U}_1'(Z-\bar{Z})][T^{-1}(Z-\bar{Z})'(Z-\bar{Z})]^{-1}[T^{-1}(Z-\bar{Z})'\hat{U}_1]} \end{aligned} \tag{30}$$

where $\hat{U}_1 = [\hat{u}_{1,1}, \ldots, \hat{u}_{1,T}]'$, $\hat{\gamma} = [(Z-\bar{Z})'(Z-\bar{Z})]^{-1}(Z-\bar{Z})'\hat{U}_1$, 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

$$z_t = 2 + v_{1,t}(2.456/\sqrt{T}) + \epsilon_t,$$

where ϵ_t is a standard normal random variable and T is the sample size. Here, z_t follows the local to zero assumption in Equation (23). The value $C = 2.456$ is chosen so that the concentration parameter is 6.03, which is from the $b = 0.90$ column of Table 1. Hence, $\hat{B}_1$ has an asymptotic bias of 10%. The value of $\mu_z = 2$ is chosen arbitrarily.

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 (30) and compute the percentage of simulations where it exceeds the value 10. This value of 10 a is 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 (28) 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. Effectively, 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 concentration parameter and that the concentration parameter 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 (29).

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

$$F_{IV} \xrightarrow{d} (e_1' B \Sigma_v B' e_1)^{-1} (e_1' B \Sigma_v^{1/2} \theta)^2, \tag{31}$$

where $e_1 = [1, 0, \ldots, 0]'$ and θ is defined in Equation (26). This result follows from the second line of Equation (30), $\hat{u}_{1,t} = e_1' \hat{u}_t$, Equation (25), $T^{-1}\hat{U}'\hat{U} \xrightarrow{p} B\Sigma_v B'$, $T^{-1}(Z - \bar{Z})'(Z - \bar{Z}) \xrightarrow{p} \sigma_\epsilon^2$, the continuous mapping theorem and Slutsky's theorem. The limiting distribution in Equation (31) 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

$$\frac{B_{11}^2 C^2 / \sigma_\epsilon^2}{B_{11}^2 + B_{12} \Sigma_{v_2} B_{12}'}.$$ [7]

Here, we see that unlike F, which converges to a distribution that is independent of B and Σ_v, F_{IV} does not. This is problematic because the weak IV critical values (Stock and Yogo, 2005) will no longer apply, and as highlighted by the Monte Carlo experiment above, the rule of thumb of $F_{IV} > 10$ will not be a good guide for proxy strength. With F_{IV}, the concentration parameter C^2/σ_ϵ^2 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

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

DGP1 above, then the non-centrality parameter is 0 and an F_{IV} static 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 concentration parameter in DGP2 was 6.03, this DGP will draw $F_{IV} > 10$ at a much higher rate.

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 concentration parameter 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 concentration parameter 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 instrument, the estimate of B_{11} will be inconsistent and lead to an incorrectly 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 concentration parameter, 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 laid out in the previous sections, I study the dynamic effects of productivity shocks. To do this, I use Fernald's (2014) measures of utilization-adjusted total factor productivity (TFP) as my proxy variables. Fernald (2014) constructs two measures of 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 that I use includes GDP growth, private employment growth and inflation. All variables are in annualized percentage terms.[8] In addition, I include the combined annualized percent growth in non-durable goods and services consumption to correspond to Fernald's (2014) consumption TFP series as well as the combined annualized percent growth in equipment investment and durable goods consumption to correspond to Fernald's (2014) investment TFP series.[9] Thus, the VAR dimension is $n = 5$. The data is quarterly with a sample of 1947:Q2 to 2015:Q2, and I estimate the VAR with 3 lags.

Before testing for proxy weakness or estimating the impulse response functions (IRFs) and variance contributions of the TFP shocks, I first estimate the correlation of the proxy variables. The first row of the first column of Table 3 shows that this correlation 0.29. Because the VAR has three lags, this statistic is computed from the 1948:Q1 to 2015:Q2 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. Because of this, it may be appropriate to treat z_t as a 2×1 vector where $z_t = [z_t^C, z_t^I]'$ and to identify both structural shocks simultaneously as in Mertens and Ravn (2013). However, because the econometric theory in this paper is based on z_t being a scalar, I estimate the structural TFP shocks one at a time. To ensure that estimating the structural shocks one at a time does not produce structural shocks that violate the assumptions of the model, Table 3 also gives the correlation of the estimated structural shocks, the first autocorrelation of the proxies and the structural shocks, the correlations between the proxies and their corresponding structural shocks, and the correlations between the proxies and the other structural shock. This table shows that the structural shocks have a correlation 0.08, which is much lower than the correlation between the proxies. Further, this correlation is not statistically distinct from zero. Thus, I treat the estimated structural shocks as uncorrelated and consistent with Equation (4) in Assumption 1. In addition, Table 3 shows that both of the proxies and the estimated structural shocks have small and statistically insignificant autocorrelations, which is consistent with Assumptions 1 and 3. Next, the correlations between the proxies and their corresponding structural shocks are both positive and statistically significant, implying that proxies satisfy the relevance assumption. Finally, the correlation of the consumption TFP proxy and the investment TFP structural

[8] Annualized percent growth in GDP is from the National Income and Product Accounts (NIPA) Table 1.1.1. For the other series, I use the formula $g_t = (x_t/x_{t-1} - 1) \times 400$ to compute annualized growth percents. 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.

[9] 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 and structural shocks

$\text{corr}(z_t^C, z_t^I) =$	0.29	$\text{corr}(\hat{v}_{1,t}^C, \hat{v}_{1,t}^I) =$	0.08
	$(0.16, 0.42)$		$(-0.05, 0.22)$
$\text{corr}(z_t^C, z_{t-1}^C) =$	0.04	$\text{corr}(\hat{v}_{1,t}^C, \hat{v}_{1,t-1}^C) =$	-0.01
	$(-0.12, 0.12)$		$(-0.12, 0.11)$
$\text{corr}(z_t^I, z_{t-1}^I) =$	-0.01	$\text{corr}(\hat{v}_{1,t}^I, \hat{v}_{1,t-1}^I) =$	0.01
	$(-0.12, 0.11)$		$(-0.12, 0.12)$
$\text{corr}(z_t^C, \hat{v}_{1,t}^C) =$	0.58	$\text{corr}(z_t^I, \hat{v}_{1,t}^I) =$	0.38
	$(0.48, 0.66)$		$(0.28, 0.48)$
$\text{corr}(z_t^C, \hat{v}_{1,t}^I) =$	0.05	$\text{corr}(z_t^I, \hat{v}_{1,t}^C) =$	0.03
	$(-0.09, 0.19)$		$(-0.11, 0.17)$

Notes: z_t^C and z_t^I are the proxy variables for consumption TFP and investment TFP, respectively. $\hat{v}_t^C$ and $\hat{v}_t^I$ are the estimated structural shocks for consumption TFP and investment TFP, respectively. The sample is 1948:Q1 to 2015:Q2. 95% confidence intervals are presented in parentheses, and they are percentile intervals that are computed from an i.i.d. bootstrap with 10,000 replications.

shock is small and statistically insignificant. The same is true for the correlation of the investment TFP proxy and the consumption TFP structural shock. Thus, Fernald's (2014) TFP proxies do not violate the exogeneity assumption with themselves. Taken together, the correlations presented in Table 3 imply that the estimated structural shocks do not violate the basic assumptions of the model even though the proxies are positively correlated and the structural shocks are identified one at a time.

Next, I test the strength of Fernald's (2014) TFP proxies. To do this, 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 concentration parameter is 18.40. Then from Table 2, the 10% critical F statistic is 7.12 and the 5% critical F statistic is 7.98. For the consumption TFP proxy, the estimated F statistic is 26.26, which well exceeds both of these critical values. For the investment TFP proxy, the estimated F statistic is 9.02, which also exceeds both critical values. Thus, I reject the null hypothesis that each proxy is in the weak proxy set.

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 modified version of the residual-based moving block bootstrap algorithm described in Brüggemann, Jentsch, and Trenkler (2014) with 10,000 bootstrap replications. This algorithm is modified

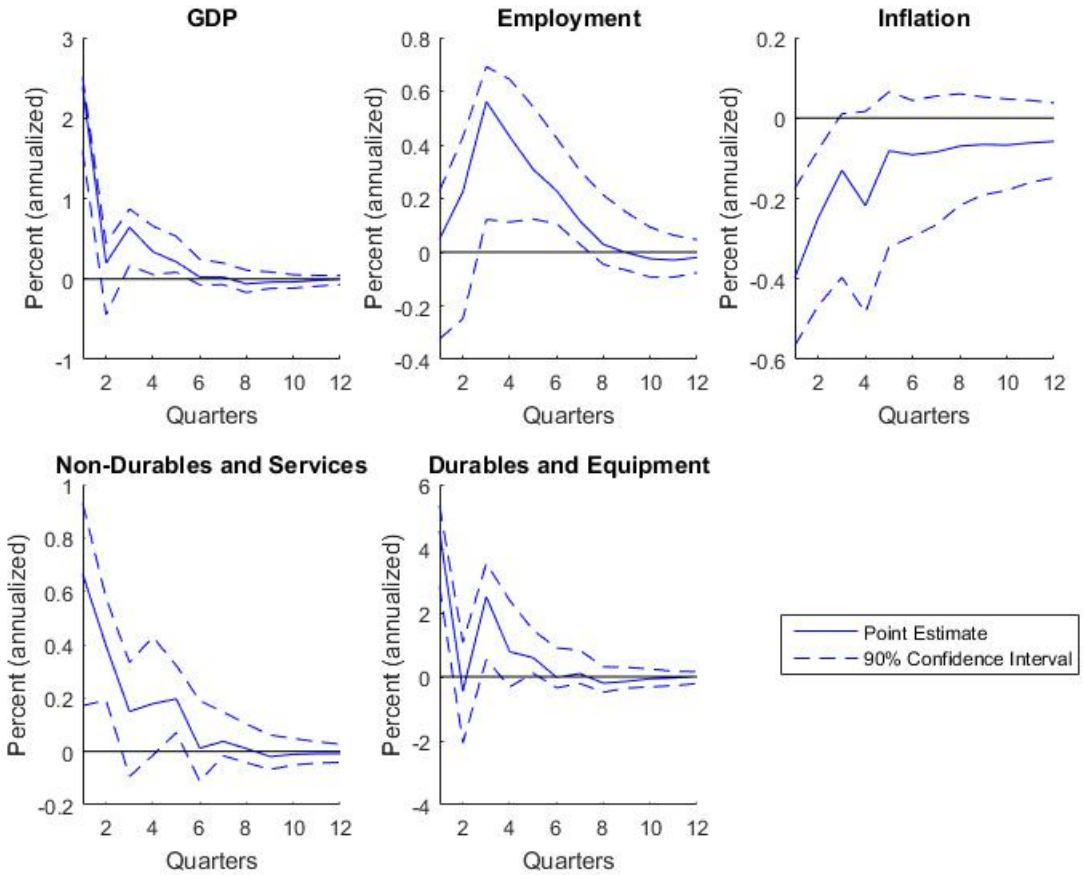

Figure 2: Dynamic effect of a TFP shock to consumption (excluding durables).

so that blocks of the proxy variable that correspond to the blocks of the VAR errors are also centered and re-sampled. This allows for re-estimation of $\hat{B}_1$ in every bootstrap loop to account for the uncertainty in estimating $\hat{B}_1$. 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 modified 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

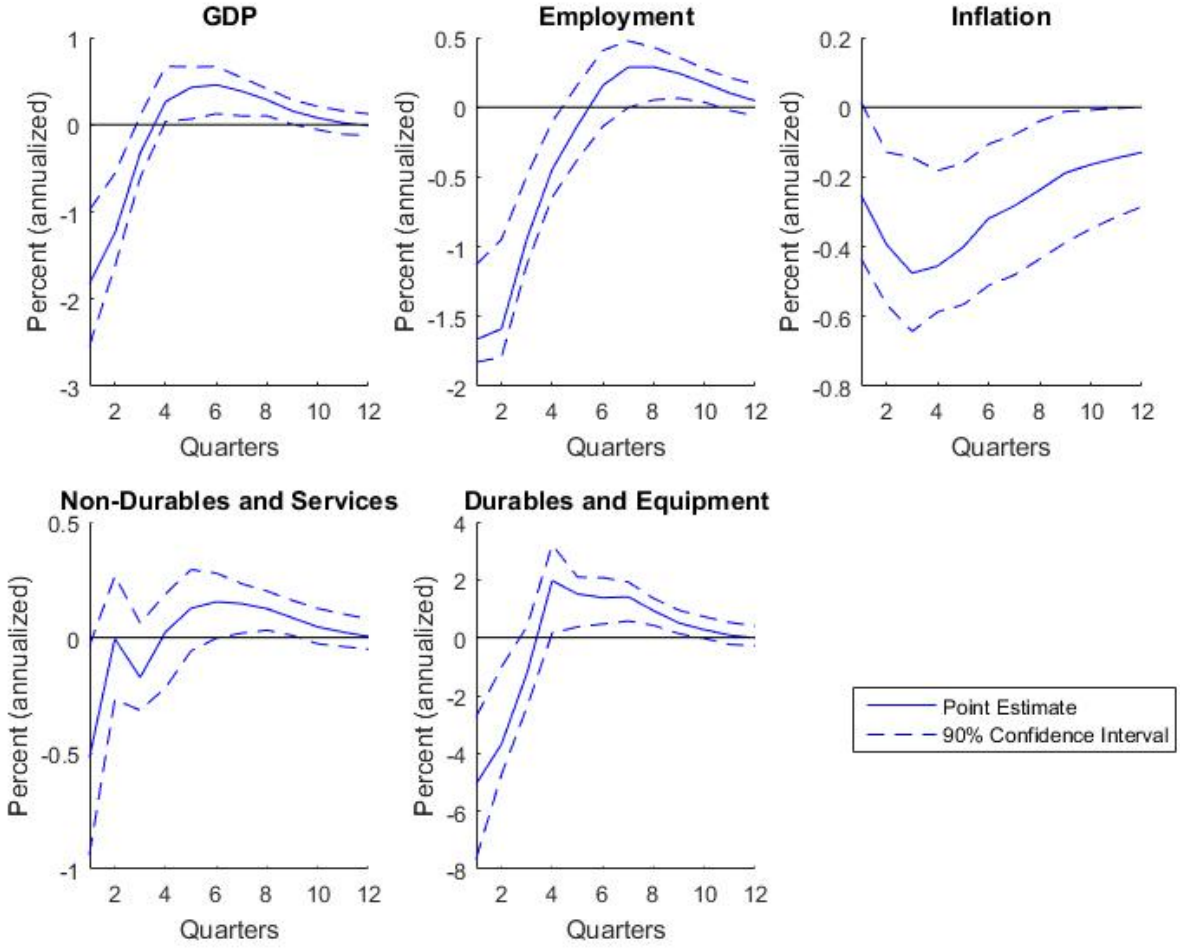

Figure 3: Dynamic effect of a TFP shock to durable consumption and equipment investment.

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). To make sense of these results, Basu, Fernald, and Liu (2012) solve a dynamic stochastic general equilibrium model where consumption and investment are produced in different sectors and both sectors have sticky prices. In their model, an increase in productivity in the investment sector causes a drop in both quantities and prices similar to Figure 3. The intuition is that an increase in productivity in the investment sector causes an increase in the mark-up of investment goods with sticky prices. This causes investment goods to be expensive in the current period

Table 4: Variance Contributions of Structural Shocks from 1948:Q1 to 2015:Q2

Structural Shock	GDP	Private Employment	Inflation	Non-Durables & Services	Durables & Equipment
Consumption TFP	0.39	0.05	0.07	0.17	0.14
Investment TFP	0.39	0.74	0.23	0.10	0.29

relative to future periods, suppressing current demand for investment goods. Thus, with sticky investment prices, positive TFP shocks can act like negative demand shocks.

Finally, I compute the variance contribution of each of the structural shocks. To do this, I first estimate

$$\hat{\zeta}_t = Y_t - \hat{A}_0' - \hat{A}_1' Y_{t-1} - \cdots - \hat{A}_p' Y_{t-p} - \hat{B}_1 \hat{v}_{1,t},$$

where $\hat{\zeta}_t$ is the estimate of $B_2 v_{2,t}$. I then simulate

$$\tilde{Y}_t = \hat{A}_0' + \hat{A}_1' \tilde{Y}_{t-1} + \cdots - \hat{A}_p' \tilde{Y}_{t-p} + \hat{\zeta}_t,$$

where the initial condition is given by $[\tilde{Y}_0, \ldots, \tilde{Y}_{-p+1}] = [Y_0, \ldots, Y_{-p+1}]$. This generates the variables in the VAR without the contribution of the structural shocks of interest. Finally, for each element in the VAR, indexed by j, I compute $1 - \text{var}(\tilde{Y}_{j,t})/\text{var}(Y_{j,t})$ to be the variance contribution from the structural shock of interest. Because the initial condition is the same for both Y_t and $\tilde{Y}_t$ by construction, I drop the initial condition from this calculation and use the sample 1948:Q1 to 2015:Q2. Table 4 presents these contributions for each variable in the VAR and for both structural shocks. The most striking result is that both shocks contribute 39% to the variance of GDP growth, implying that productivity shocks explain nearly 80% of the variance in output growth. Next, the investment TFP shock contributes 74% to the variance of growth in private employment, indicating that the majority of the fluctuations in employment growth comes from this one shock. The investment TFP shock also has a large impact on inflation and growth in durables and equipment investment, contributing 23% and 29% of those variances, respectively. Finally, the consumption TFP shock has only modest variance contributions to employment growth, inflation, growth in non-durables and services, and growth in durables and equipment.

6 Conclusion

In this paper, I show that 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 a weak proxy is local to zero as in Staiger and Stock (1997). Given 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 projection 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 been used to test for weak proxy variables up to this point. 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 value for this statistic cannot simply be read from the weak IV literature (Stock and Yogo, 2005); rather, it will need to be computed for each application. Finally, the problem with computing these critical values is that the presence of a weak proxy makes them inconsistent and leads to mis-sized statistical tests.

I use Fernald's (2014) measures of consumption TFP and investment TFP as proxy variables to study the dynamic effects of productivity shocks. I find that both of these proxies are strong, that they yield structural shocks that satisfy the relevance and exogeneity assumptions, and that these structural shocks are mutually and serially uncorrelated. A positive shock to consumption TFP produces an immediate increase in output growth and an immediate decrease in inflation, which is consistent with the standard theory of a supply shock. A positive shock to investment TFP causes an immediate decrease in output growth along with a hump-shaped decrease in inflation, resembling a negative demand shock rather than a positive supply shock. However, this result is consistent with the empirical findings of Basu et al. (2013) and the theoretical model of Basu, Fernald, and Liu (2012). Finally, I find that the consumption and investment TFP shocks combine to contribute nearly 80% of the variance in GDP growth and that the investment TFP shock alone contributes over 70% of the variance in private employment growth.

This paper focuses on the case where one proxy variable is used to identify one structural shock. Thus, future research should extend this analysis to study the cases where multiple proxies exist for one structural shock. Further, this analysis should be extended for when researchers want to identify multiple structural shocks at once with multiple proxy variables as in Mertens and Ravn (2013) or with Fernald's (2014) TFP shocks.

Appendix

Proof of Lemma 1 The jth element of $\tilde{\theta}_j$ can be written as

$$\tilde{\theta}_j = \frac{\theta_j}{\sqrt{\theta_1^2 + \cdots + \theta_n^2}},$$

where θ_j is the jth element of θ. The elements of θ as defined in Equation (26) are uncorrelated normal random variables, implying that they are also independent. Next, define a $n-1 \times 1$ random vector λ, containing all elements of θ except for θ_j for any $j \geq 2$. Then,

$$\tilde{\theta}_j = \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}}.$$

Here $\lambda'\lambda$ is a random scalar from the non-central χ^2_{n-1} distribution with a non-centrality parameter of C^2/σ_ϵ^2. Because the elements of θ are independent, it is also the case that θ_j and $\lambda'\lambda$ are independent. This implies that the expectation of θ_j can be written as

$$\mathbb{E}(\tilde{\theta}_j) = \int_0^\infty \left[\int_{-\infty}^\infty \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}} f(\theta_j) d\theta_j \right] g(\lambda'\lambda) d\lambda'\lambda, \tag{A.1}$$

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.

The proof proceeds in two steps. First, I establish that the interior integral in Equation (A.1) is finite for any $\lambda'\lambda \geq 0$ so that it can be computed for each $\lambda'\lambda \geq 0$. Second, I establish that the interior integral is equal to zero for all $\lambda'\lambda \geq 0$ implying that $\mathbb{E}(\tilde{\theta}_j) = 0$.

First, fix some $\lambda'\lambda > 0$. Then, for $\theta_j \geq 0$ and for $\theta_j \leq 0$

$$0 \leq \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}} \leq 1 \quad \text{and} \quad -1 \leq \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}} \leq 0,$$

respectively. Then,

$$0 \leq \int_0^\kappa \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}} f(\theta_j) d\theta_j \leq \int_0^\kappa f(\theta_j) d\theta_j,$$

and

$$0 \geq \int_{-\kappa}^0 \frac{\theta_j}{\sqrt{\theta_j^2 + \lambda'\lambda}} f(\theta_j) d\theta_j \geq \int_{-\kappa}^0 -f(\theta_j) d\theta_j.$$

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

$$0 \leq \int_0^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j \leq \frac{1}{2} \quad \text{and} \quad 0 \geq \int_{-\infty}^0 \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j \geq -\frac{1}{2}.$$

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

Second, fix some $\lambda'\lambda > 0$. Then,

$$\begin{aligned}
\int_{-\infty}^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j &= \int_0^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j + \int_{-\infty}^0 \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j \\
&= \int_0^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j + \int_0^\infty \frac{-\theta_j}{\sqrt{(-\theta_j)^2+\lambda'\lambda}} f(-\theta_j)d\theta_j \\
&= \int_0^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j + \int_0^\infty \frac{-\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j \\
&= \int_0^\infty \frac{\theta_j - \theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j = \int_0^\infty 0 \cdot f(\theta_j)d\theta_j = 0.
\end{aligned}$$

In the event that $\lambda'\lambda = 0$,

$$\frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} = 1 \quad \text{and} \quad \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} = -1$$

for $\theta_j > 0$ and for $\theta_j < 0$, respectively. When $\lambda'\lambda = 0$, then $\theta_j/\sqrt{\theta_j^2+\lambda'\lambda}$ is undefined when $\theta_j = 0$. However, I can apply

$$\lim_{\kappa\to 0^+}\int_\kappa^\infty \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j = \lim_{\kappa\to 0^+}\int_\kappa^\infty f(\theta_j)d\theta_j = \frac{1}{2}$$

and

$$\lim_{\kappa\to 0^-}\int_{-\infty}^\kappa \frac{\theta_j}{\sqrt{\theta_j^2+\lambda'\lambda}} f(\theta_j)d\theta_j = \lim_{\kappa\to 0^-}\int_{-\infty}^\kappa -f(\theta_j)d\theta_j = -\frac{1}{2}.$$

Thus, the interior integral of Equation (A.1) is zero for all $\lambda'\lambda \geq 0$, and Equation (A.1) can be re-written as

$$\mathbb{E}(\tilde{\theta}_j) = \int_0^\infty 0 \cdot g(\lambda'\lambda)d\lambda'\lambda = 0.$$

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.

Brüggemann, Ralf, Carsten Jentsch, and Carsten Trenkler. 2014. "Inference in VARs with Conditional Heteroskedasticity of Unknown Form." University of Konstanz Working Paper 2014-13.

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.

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.

Kliem, Martin and Alexander Kriwoluzky. 2013. "Reconciling Narrative Monetary Policy Disturbances with Structural VAR Model Shocks?" *Economics Letters* 121 (2):247–251.

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

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.