Lewis, Daniel J.

Working Paper

Robust inference in models identified via heteroskedasticity

Staff Report, No. 876

Provided in Cooperation with:
Federal Reserve Bank of New York

Suggested Citation: Lewis, Daniel J. (2018) : Robust inference in models identified via heteroskedasticity, Staff Report, No. 876, Federal Reserve Bank of New York, New York, NY

Federal Reserve Bank of New York
Staff Reports

Robust Inference in Models Identified via Heteroskedasticity

Daniel J. Lewis

Staff Report No. 876
December 2018

Robust Inference in Models Identified via Heteroskedasticity

Daniel J. Lewis

Federal Reserve Bank of New York Staff Reports, no. 876
December 2018
JEL classification: C12, C32, E43

Abstract

Identification via heteroskedasticity exploits differences in variances across regimes to identify parameters in simultaneous equations. I study weak identification in such models, which arises when variances change very little or the variances of multiple shocks change close to proportionally. I show that this causes standard inference to become unreliable, propose two tests to detect weak identification, and develop nonconservative methods for robust inference on a subset of the parameter vector. I apply these tools to monetary policy shocks, identified using heteroskedasticity in high frequency data. I detect weak identification in daily data, causing standard inference methods to be invalid. However, using intraday data instead allows the shocks to be strongly identified.

Key words: heteroskedasticity, weak identification, robust inference, pretesting, monetary policy, impulse response function

---

Lewis: Federal Reserve Bank of New York (email: daniel.lewis@ny.frb.org). The author thanks James Stock, Isaiah Andrews, Gabriel Chodorow-Reich, Richard Crump, Domenico Giannone, Adam McCloskey, José Luis Montiel Olea, Emi Nakamura, Mikkel Plagborg-Möller, Neil Shephard, and Jòn Steinsson for helpful feedback. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

To view the author's disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr876.html.

1 Introduction

Unobserved structural shocks, like those in the structural vector auto-regressions (SVARs) of Sims (1980), are ubiquitous in economic models across fields, where observed innovations are related to structural shocks by a linear combination matrix. Economists frequently study the effects of such structural shocks to identify causal relationships. A variety of identification approaches to recover the structural shocks exist, but identification via heteroskedasticity, which does not require the researcher to impose assumptions on the responses themselves, has grown in popularity in empirical work. Holding constant contemporaneous responses, this methodology compares differences in innovation covariances across regimes to identify those constant parameters as coefficients on the changing variances of the structural shocks. This identification scheme is most popular in macro-financial contexts, but has also been adopted in fields including public finance, growth, trade, political economy, environmental economics, agriculture, energy, education, marketing, and even fertility studies.[1] However, no work has addressed the possibility of weak identification in these studies.

The identifying variation is the difference in covariances across regimes. If the structural variances are in fact the same across regimes, then so too are the reduced-form covariances, and there is no identifying variation. More subtly, if the structural variances all change by the same factor across regimes, there is no new identifying information, as the covariance matrices are just scalar multiples. This presents two cases that may lead to weak identification – if the variances change by too little, or if they change (perhaps substantially) by too similar a factor. The latter means that even if ample heteroskedasticity is present, identification is not guaranteed. The effects are akin to the more familiar weak instruments (IV) context – where an instrument that offers little information about an endogenous regressor leads to poor identification of the parameter of interest. As a result, multiple sets of parameters may be almost observationally equivalent, causing the asymptotic distribution of estimators to be non-standard. Standard inference methods will be unreliable, as will any empirical conclusions based on them. If not properly detected and accounted for, this can undermine the credibility of empirical work.

I provide a framework for inference in models identified via heteroskedasticity when weak

[1] The macrofinance core includes Rigobon (2003), Rigobon & Sack (2003, 2004), Boyarchenko et al (2017), Craine & Martin (2008), Ehrmann & Fratzscher (2016), Eichengreen & Panizza (2016), Hébert & Schreger (2017), Nakamura & Steinsson (2018), and Wright (2012). Examples in other fields include public finance (Jahn & Weber (2016)), growth (Islam et al (2017)), trade (Lin et al (2016), Feenstra & Weinstein (2017)), political economy (Rigobon & Rodrik (2005) Khalid (2016)), environmental economics (Millimet & Roy (2016) Gong et al (2017)), agriculture and energy (Fernandez-Perez et al (2016)), education (Hogan & Rigobon (2009), Klein & Vella (2009)), marketing (Zaefarian et al (2017)), and even fertility studies (Mönkediek & Bras (2016)).

identification causes standard methods to provide a poor approximation of the asymptotic distribution. I present two tests for the presence of weak identification. In an empirically common simple case, where only one variance changes, the model can be written as IV, where the instrument is ± 1 times one of the innovations, with the sign depending on the regime. I propose a rule of thumb of $F > 23$ for the first-stage $F-$statistic. In the general case, where all variances change, I propose the GMM approach of Andrews (2017), which compares the size of standard and robust confidence sets. This test would be (perhaps prohibitively) conservative using robust sets computed using the only previously available option, projection inference, which I show to have very low size and power. However, I extend robust inference results for the full parameter vector (e.g. Stock & Wright (2000), Kleibergen (2005), Magnusson & Mavroeidis (2014)) to prove the validity of non-conservative robust inference on subsets of the parameter vector. The resulting confidence sets can be used with Andrews' (2017) approach to detect weak identification and then to conduct inference if necessary. I show that this subset inference result can also be applied to impulse response functions (IRFs), treating the data as estimated residuals.

I demonstrate, both in data and empirically calibrated simulations, that weak identification does in fact cause standard inference approaches to perform poorly. I consider the application of Nakamura & Steinsson (2018), who attempt to exploit higher variance in monetary policy shocks around monetary policy announcements, compared to ordinary days, to identify monetary policy shocks. I find that the shocks are weakly identified in daily data, while intraday data provides strong identification. In daily data, using robust confidence intervals reverses a surprising finding that treasury yields respond significantly to noise in forward rates; in intraday data, conclusions are unchanged. In simulations based on the daily data, estimates of the effect of monetary policy are not well approximated by a normal distribution. Accordingly, standard tests for the full parameter vector, as well as for this parameter in particular, exhibit dramatic size distortions. Conversely, robust projection-type tests for this parameter of interest are severely under-sized. In contrast, the subset tests that I propose are consistently well-sized and have desirable power properties, making them the first viable option for this problem.

With the tools I propose, applied research using heteroskedasticity for identification can address concerns of weak identification head-on. I hope that it can become best practice to verify the strength identification using these methods, much like it has for IV following the work of Staiger & Stock (1997).

The paper is organized as follows. Section 2 presents the intuition of identification via heteroskedasticity and how weak identification can arise, characterizes a simple case of weak identification analytically, and presents evidence of weak identification in empirically cali-

brated simulations. Section 3 develops the model more formally and describes how weak identification can arise. Section 4 presents existing robust inference approaches for the full parameter vector and extends them to allow for non-conservative inference on a subset of the parameter vector; the performance of tests based on these results is illustrated in simulations. Section 5 presents two tests to detect weak identification. Section 6 extends the subset results to inference on IRFs. Section 7 applies the full collection of methods to the data of Nakamura & Steinsson (2018). Section 8 concludes.

2 An illustration of the problem

In this section, I sketch the intuition behind identification via heteroskedasticity and how weak identification may arise. I then derive the asymptotic distribution of parameter estimates in an empirically common simple case. Finally, I show evidence of weak identification in empirically-calibrated simulations.

2.1 Identification and when it might fail

The phenomenon of weak identification in models identified via heteroskedasticity is well-captured in a simple bivariate model. I present the model here, informally, to clearly exposit the issues I aim to address, before providing a fully general formal treatment in the subsequent section. Consider a 2×1 vector of non-serially correlated mean-zero innovations η_t. These could be residuals from a VAR, daily changes in asset prices, or otherwise obtained. These innovations are related to a 2×1 vector of structural shocks, ε_t, by a time-invariant invertible matrix H:[2]

$$\eta_t = \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{bmatrix} 1 & H_{12} \\ H_{21} & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = H\varepsilon_t. \tag{1}$$

As a normalization, the diagonal of H is all ones.[3] There are two regimes, R_1, R_2. For each regime,

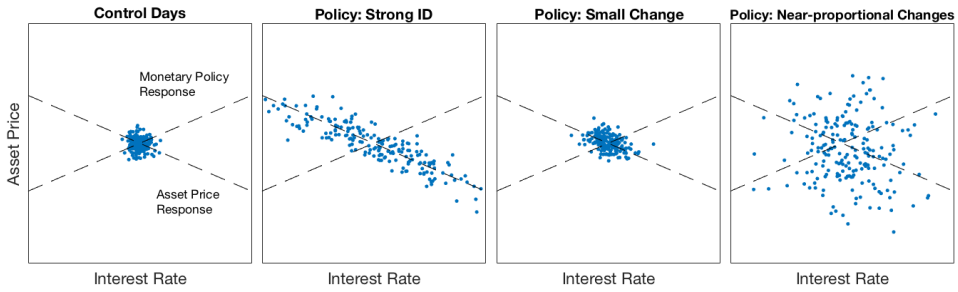
$$E\left[\varepsilon_t | t \in R_r\right] = 0,\ E\left[\varepsilon_t \varepsilon_t' | t \in R_r\right] = \Sigma_{\varepsilon,r}, r = 1, 2 \tag{2}$$

where $\Sigma_{\varepsilon,1}, \Sigma_{\varepsilon,2}$ are diagonal. This is the standard orthogonality assumption on structural shocks. Correspondingly, $\Sigma_{\eta,1} = E\left[\eta_t \eta_t' | t \in R_1\right] = H\Sigma_{\varepsilon,1}H'$ and similarly $\Sigma_{\eta,2} = H\Sigma_{\varepsilon,2}H'$.

[2]Note that Rigobon & Sack (2003, 2004) consider bivariate models with 3 structural shocks. These models can only be identified with additional application-specific structural assumptions. Instead, I focus on models where identification follow exclusively from the heteroskedasticity.

[3]This is without loss of generality, simply imposing a scale on the columns of the H matrix.

Figure 1: Distinguishing simultaneous responses using heteroskedasticity



Observations are simulated to replicate the setup in Figures 1 and 2 of Rigobon & Sack (2004).

Identification via heteroskedasticity requires that

$$\Sigma_{\varepsilon,2} \neq a\Sigma_{\varepsilon,1} \tag{3}$$

for any scalar a. This means that the two variances cannot change by the same factor across regimes (or, consequently, both be stable across regimes). In practice, regimes are frequently chosen using external information about volatility, like monetary policy announcement dates, and are thus known ex ante.

Figure 1 presents the intuition of the identification approach. The first two panels follow the example from Rigobon & Sack (2004), who identify the response of asset prices to monetary policy via variance changes on policy announcement days; data is simulated to crystalize the argument. The first panel plots η_t, asset price changes against interest rate changes, on "control" days – those with no monetary policy announcement. Due to the simultaneity of the problem, with two structural shocks impacting η_t contemporaneously, the asset price response cannot be identified. The second panel plots what might happen on days with monetary policy announcements if the variance of the policy shock increases dramatically (and the asset price shock variance changes only slightly). Now – due to the increase in volatility in the monetary policy shock – the data begin trace out the asset price response. Since there is still non-negligible volatility in the second structural shock, the response cannot be identified from the second regime alone, but it can be identified by contrasting the information contained in both regimes.

More formally, in a model with a single variance regime, (1) furnishes three equations, but there are four free parameters to estimate: two off-diagonal elements in H, and the two variances of the structural shocks. Adding a second variance regime doubles the number of equations to six, while adding an additional two parameters (the extra structural variances), so the model may now be just identified. Rigobon (2003) establishes that these equations have a unique solution provided (3) holds. In practice, these equations are often estimated

via GMM, with the moment equations corresponding to $\Sigma_{\eta,1} = H\Sigma_{\varepsilon,1}H'$ and $\Sigma = H\Sigma_{\varepsilon,2}H'$. Inference then proceeds using either the standard GMM asymptotic distribution or a bootstrap approach; see Rigobon (2003) for details. Brunnermeier et al (2017) offer an elegant argument for identification, showing that H can be recovered (up to column order and normalization) as the eigenvectors of $\Sigma_{\eta,1}^{-1}\Sigma_{\eta,2}$. This simple estimator yields the GMM solution in this just-identified context.

The condition in (3) surely holds in a literal sense in finite samples due to sampling variability. But what happens if it is *close* to failing? First, the variances might not change much at all across regimes. For example, if much of the information contained in monetary policy announcements is anticipated, the volatility may not increase much on announcement days over its average level. This would make the two variance regimes close to identical. The third panel of Figure 1 depicts this concern; the variance of the monetary policy shock increases but the cloud of data does not clearly trace out the asset price response curve. The policy regime offers little additional identifying information over the control sample. Second, all variances could change together. In the Great Moderation, many volatilities decreased simultaneously, while during the Financial Crisis, many volatilities increased together. On announcement days, there may be increased volatility in more than one shock, in particular if there are multiple dimensions of monetary policy shocks. In such episodes, the variances of structural shocks may be moving together. The closer the comovement, the less identifying information the second variance regime provides about H. The final panel of Figure 1 depicts this concern. There is a large increase in volatility in both dimensions, and the increase in the variance of the policy shock does not dwarf that of the other shock. Again, the policy regime offers little additional identifying information.

2.2 The weak identification distribution in a simple case

I now discuss the asymptotic behaviour of estimators of H in a simple case under the additional assumption that only one structural shock variance changes; the general case is discussed in Section 3. The assumption of a single variance change is common in practice (e.g. Nakamura & Steinsson (2018), Rigobon & Sack (2004), Hébert & Schreger (2017), Wright (2012)) and admits closed-form estimators for the parameter of interest that parallel the instrumental variables model. This similarity allows for an intuitive illustration of the problem of weak identification. In particular, I fix the first shock's variance, with identification following from the change in variance of the second. Weak identification arises if this variance does not change dramatically, representing the scenario where there is essentially no heteroskedasticity. Throughout the paper, I refer to this model as the "*simple case*". The

identifying equations are given by

$$\Sigma_{\eta,1} = \begin{bmatrix} \sigma^2_{\varepsilon_1} + H^2_{12}\sigma^2_{\varepsilon_2,1} & - \\ H_{21}\sigma^2_{\varepsilon_1} + H_{12}\sigma^2_{\varepsilon_2,1} & H^2_{21}\sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_2,1} \end{bmatrix}, \tag{4}$$
$$\Sigma_{\eta,2} = \begin{bmatrix} \sigma^2_{\varepsilon_1} + H^2_{12}\sigma^2_{\varepsilon_2,2} & - \\ H_{21}\sigma^2_{\varepsilon_1} + H_{12}\sigma^2_{\varepsilon_2,2} & H^2_{21}\sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_2,2} \end{bmatrix},$$

where $\sigma^2_{\varepsilon_1,1} = \sigma^2_{\varepsilon_1,2} \equiv \sigma^2_{\varepsilon_1}$ by assumption. H_{12} is the parameter of interest. This off-diagonal element measures the impact of a unit structural shock (say a policy shock) on a variable in the system. For instance, in Nakamura & Steinsson (2018), the policy shock (ε_{2t}) alone is assumed to exhibit heteroskedasticity on policy announcement days, so H_{12} represents the impact of monetary policy shocks on instantaneous Treasury forward rates (the "dependent" variable). As before, the equations (4) can be estimated via GMM. However, in this setting, H_{12} can be identified in closed form from (4), which admits two new estimation approaches.

First, following Rigobon & Sack (2004), writing the ratio of differences in second moments across regimes yields

$$\frac{\sigma_{\eta_1\eta_2,2} - \sigma_{\eta_1\eta_2,1}}{\sigma^2_{\eta_2,2} - \sigma^2_{\eta_2,1}} = \frac{H_{12}\left(\sigma^2_{\varepsilon_2,2} - \sigma^2_{\varepsilon_2,1}\right)}{\left(\sigma^2_{\varepsilon_2,2} - \sigma^2_{\varepsilon_2,1}\right)} = \frac{H_{12}\Delta\left(\sigma^2_{\varepsilon_2}\right)}{\Delta\left(\sigma^2_{\varepsilon_2}\right)} = H_{12}, \tag{5}$$

where $\sigma_{\eta_1\eta_2,j}, \sigma^2_{\eta_2,j}$ are elements of $\Sigma_{\eta,j}$ and the $\Delta(\cdot)$ operator takes the difference in the argument between regimes 2 and 1.[4] A simple estimator uses the sample analogs of (5), as in Nakamura & Steinsson (2018). Note that if the assumption that $\sigma^2_{\varepsilon_1,1} = \sigma^2_{\varepsilon_1,2}$ fails, then H_{12} will be misidentified by (5), which will instead yield

$$\frac{H_{12}\Delta\left(\sigma^2_{\varepsilon_2}\right) + H_{21}\Delta\left(\sigma^2_{\varepsilon_1}\right)}{\Delta\left(\sigma^2_{\varepsilon_2}\right) + H^2_{21}\Delta\left(\sigma^2_{\varepsilon_1}\right)}, \tag{6}$$

so this assumption should only be made carefully in practice.

I now move from H_{12}, identified in population, to estimators, $\hat{H}_{12}$. The sample analogue

[4]This is in fact just one of three possible closed-form expressions for H_{12} implied by (4) (as the problem is over-identified under the assumption $\sigma^2_{\varepsilon_1,1} = \sigma^2_{\varepsilon_1,2} \equiv \sigma^2_{\varepsilon_1}$). The following development of weak identification applies equally to either of the alternative identifying equations.

of (5) is equivalent to an instrumental variables problem (Rigobon & Sack (2004)):

$$\hat{H}_{12} = \frac{\Delta\left(\hat{\sigma}_{\eta_1\eta_2}\right)}{\Delta\left(\hat{\sigma}^2_{\eta_2}\right)} = \frac{\frac{T}{T_2}\sum_{t\in R_2}\eta_{1t}\eta_{2t} - \frac{T}{T_1}\sum_{t\in R_1}\eta_{1t}\eta_{2t}}{\frac{T}{T_2}\sum_{t\in R_2}\eta^2_{2t} - \frac{T}{T_1}\sum_{t\in R_1}\eta^2_{2t}}$$
$$= \frac{\sum_{t=1}^{T}\eta_{1t}Z_t}{\sum_{t=1}^{T}\eta_{2t}Z_t}, \tag{7}$$

where $T_r = |R_r|$, (with T_r/T fixed), and

$$Z_t = \left[\mathbf{1}\left(t \in R_2\right) \times \frac{T}{T_2} - \mathbf{1}\left(t \in R_1\right) \times \frac{T}{T_1}\right]\eta_{2t}.$$

Thus, $\hat{H}_{12}$ can be estimated via TSLS, as suggested in Rigobon & Sack (2004), using

$$\text{first stage:} \quad \underbrace{\eta_{2t}}_{X} = \underbrace{\Pi Z_t}_{\Pi Z} + v_t$$
$$\text{second stage:} \quad \underbrace{\eta_{1t}}_{Y} = \underbrace{H_{12}\eta_{2t}}_{\beta X} + u_t,$$

where standard IV notation is indicated below the terms. If Z_t is strongly correlated with the innovation η_{2t} (exogeneity follows from (1) and (2)), standard asymptotic results for TSLS apply. First,

$$\hat{H}_{12} = \frac{\frac{1}{T}\sum_{t=1}^{T}\eta_{1t}Z_t}{\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t} \xrightarrow{p} \frac{E\left[\eta_{1t}Z_t\right]}{E\left[\eta_{2t}Z_t\right]} = H_{12}, \tag{8}$$

as long as the denominator, $\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t$, does not converge to zero, so Slutsky's theorem can be applied. Thus, H_{12} is consistently estimated. Moreover, Slutsky's theorem shows that, provided the denominator does not converge to zero, the asymptotic distribution will be characterized by the behaviour of the numerator. In particular, under regularity conditions and a martingale central limit theorem,

$$\sqrt{T}\left(\hat{H}_{12} - H_{12}\right) = \frac{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\eta_{1t}Z_t}{\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t} \xrightarrow{d} N\left(0, V_{strong}\right),$$

as long as $\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t \not\xrightarrow{p} 0$. V_{strong} is the usual White (1980) heteroskedasticity-robust TSLS asymptotic variance, $E\left[\eta_{2t}Z_t\right]^{-2}E\left[u_t^2Z_t^2\right]$. It can be consistently estimated for inference, but a bootstrap procedure is often used in practice.

What happens if the denominator is in fact close to zero? In the familiar IV setting, standard inference methods are not reliable (e.g. Staiger & Stock (1997)). As the first stage coefficient, Π, tends to zero, the instrument provides less information about the endogenous

regressor. As a result, virtually all work using instrumental variables now reports tests to detect instrument irrelevance (the first stage $F-$statistic). In identification via heteroskedasticity, Π goes to zero as $\sigma^2_{\varepsilon_2,2}$ approaches $\sigma^2_{\varepsilon_2,1}$, the case of no variance change.[5] However, up to now, no tests to detect this possibility have been proposed.

If $\sigma^2_{\varepsilon_2,2} = \sigma^2_{\varepsilon_2,1}$ $(\Pi = 0)$, so H_{12} is unidentified, then the denominator (and numerator) of (7) converges in probability to zero, so Slutsky's theorem cannot be applied as in (8). To obtain a limit distribution, multiplying (7) by $\frac{\sqrt{T}}{\sqrt{T}}$ leads both numerator and denominator to converge in distribution to mean-zero normal random variables. This means that $\hat{H}_{12}$ converges in distribution to the ratio of two correlated normal random variables, a Cauchy-like distribution, so the standard normal approximation is not a good one. Thus, the convergence of (7) is non-uniform with respect to $\left(\sigma^2_{\varepsilon_2,2}, \sigma^2_{\varepsilon_2,1}\right)$: if $\sigma^2_{\varepsilon_2,2} \neq \sigma^2_{\varepsilon_2,1}$ it is normal, but if $\sigma^2_{\varepsilon_2,2} = \sigma^2_{\varepsilon_2,1}$ it is not. To derive an asymptotic distribution that well-approximates the behaviour of $\hat{H}_{12}$ when $\sigma^2_{\varepsilon_2,2}$ is close to, but not equal to, $\sigma^2_{\varepsilon_2,1}$, I model the difference as "small". In particular,

$$\frac{\sigma^2_{\varepsilon_2,2}}{\sigma^2_{\varepsilon_2,1}} = 1 + \frac{d}{\sqrt{T}}. \tag{9}$$

Rearranging yields

$$\sigma^2_{\varepsilon_2,2} = \sigma^2_{\varepsilon_2,1}\left(1 + d/T^{1/2}\right) \equiv \sigma^2_{\varepsilon_2,1} + d_\varepsilon/T^{1/2},$$

so $\sigma^2_{\varepsilon_2,2}$ is "local to $\sigma^2_{\varepsilon_2,1}$".[6] Employing this device means that, even as $T \to \infty$, the probability of rejecting the hypothesis $\sigma^2_{\varepsilon_2,2} = \sigma^2_{\varepsilon_2,1}$ tends to neither zero nor one, capturing the intermediate case of weak identification.

With this model of $\sigma^2_{\varepsilon_2,2}$ and $\sigma^2_{\varepsilon_2,1}$ in hand, I derive an approximation to the asymptotic distribution of $\hat{H}_{12}$ under weak identification:

Proposition 1. *Under the device (9), if η_t is ergodic and stationary within regimes, then*

$$\hat{H}_{12} - H_{12} \xrightarrow{d} \frac{z_1}{d_\varepsilon + z_2}, \ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(0, V_{weak}\right), \tag{10}$$

where V_{weak} is determined by the parameters of the model and distribution of the data.

Proposition 1 follows from an argument in the spirit of Staiger & Stock (1997), presented in the Appendix. It shows that the estimator is no longer consistent; $\hat{H}_{12}$ itself, as opposed

[5] Hébert & Schreger (2017) note that weak identification can also result using the other two possible closed-form expressions for H_{12} if $H_{12} = 0$, as in those expressions, H_{12} scales the denominator as well as the numerator. This is a particular concern when the null hypothesis is $H_{12} = 0$. However, this is not a concern with (5).

[6] While in the present *simple case* it might be more direct to simply model $\Delta\left(\sigma^2_2\right) = \sigma^2_{\varepsilon_2,2} - \sigma^2_{\varepsilon_2,1}$ as local to zero, using relative changes will later allow me to unify the treatment of both the case where the variances barely change and that where they change but by similar factors, which both represent weak identification.

to $\sqrt{T}\left(\hat{H}_{12} - H_{12}\right)$, has a non-degenerate sampling distribution. The reason for this is that, asymptotically, the denominator $\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t \xrightarrow{p} 0$. As this identifying variation becomes small, the sampling variation in the otherwise consistently estimated means matters for the asymptotic distribution of $\hat{H}_{12}$.

The estimator's asymptotic distribution is thus better represented as the ratio of two correlated normals. Inference approaches based on the normal approximation break down. A bootstrap approach for $\hat{H}_{12}$ (for Wald-type inference) is also invalid, as shown in Moreira, Porter, & Suarez (2005). Similarly, a GMM application of the IV estimator will fare no better (Stock, Wright, & Yogo (2002)). Instead, robust methods developed for weak instruments should be used. While this discussion is specialized to the *simple case* to allow closed-form expressions for $\hat{H}_{12}$, the intuition applies to the fully general case developed in Section 3.

2.3 An illustration in simulation

To augment this theoretical interpretation of the weak identification problem, I now present simulation evidence that illustrates the threat posed to valid inference. I calibrate my analysis to the bivariate model of Nakamura & Steinsson's (2018) study of forward guidance effects. They use a bivariate model with one-day changes in instantaneous Treasury forward rates as a dependent variable and one-day changes in 2-year treasury yields as a policy measure (or alternatively 30-minute changes in a "policy news" series). They compare two regimes: monetary policy announcement dates and analogous days without such announcements. They seek to identify the effects of forward guidance by using heteroskedasticity to decompose changes in the forward rate into the response to the monetary policy shock and another unnamed shock. In particular, I adopt their specification where the policy series is the daily change in the 2-year nominal Treasury yield, and the "dependent variable" is the daily change in the 2-year instantaneous nominal Treasury forward rate. The authors assume that the non-policy shock's volatility is fixed; throughout, I instead allow both variances to change, estimating a slightly more general model. I do this for two reasons. First, I find evidence in Section 7 that both shock variances do change. Second, it allows my simulations to capture both potential sources of weak identification (negligible variance change and proportional variance change). For additional details, see Section 7.

The particular calibration is given by

$$H = \begin{bmatrix} 1 & -0.31 \\ 0.70 & 1 \end{bmatrix}, \Sigma_{\varepsilon,C} = \begin{bmatrix} 3.9 & 0 \\ 0 & 0.1 \end{bmatrix} \times 10^{-3}, \Sigma_{\varepsilon,P} = \begin{bmatrix} 7.1 & 0 \\ 0 & 0.5 \end{bmatrix} \times 10^{-3}, \qquad (11)$$

where C and P represent control and policy, with the forward rate ordered first and nominal Treasury yield second (under the unit effect normalization, the second shock is the policy shock). I calibrate strength of identification via $\delta = \frac{\sigma^2_{\varepsilon_{2,P}}/\sigma^2_{\varepsilon_{2,C}}}{\sigma^2_{\varepsilon_{1,P}}/\sigma^2_{\varepsilon_{1,C}}} - 1$. I center simulations at $T = 800$, with the fraction of control days $\frac{760}{834}$, in keeping with the data. I vary T by a factor of 2 in each direction and vary δ by a factor of 10 in each direction. Estimation proceeds via the Brunnermeier et al (2017) method, with inference based on efficient GMM using those estimates.[7]

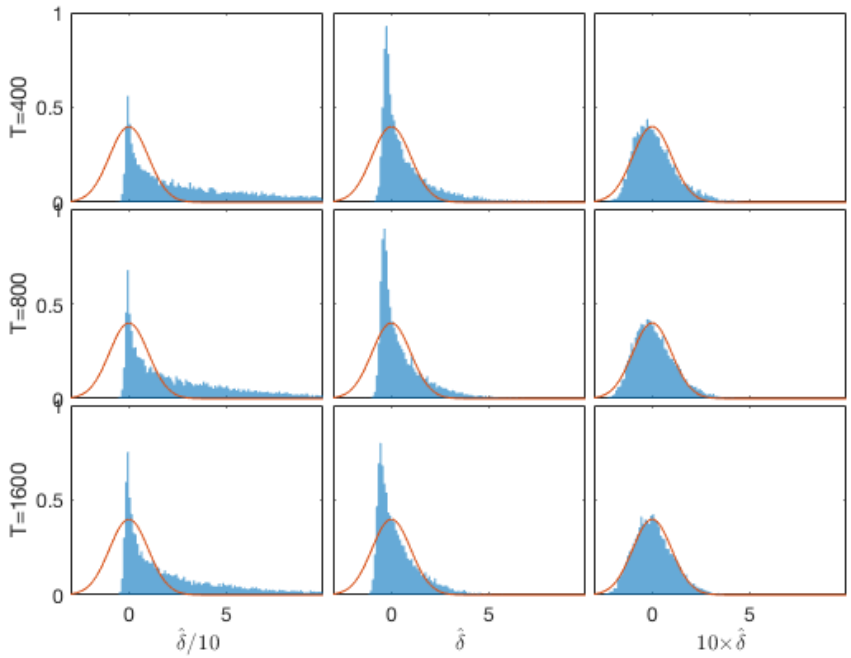
Figure 2 presents histograms of the $t-$ratio, $\frac{\hat{H}_{12}-H_{12}}{se\left(\hat{H_{12}}\right)}$, for 10,000 draws. It is clear that the estimates are not normally distributed for low degrees of identification, even as T grows large. However, for the "strong identification" specifications, the distribution is closer to a normal distribution. This is prima facie evidence of weak identification. It is clear that relying on standard inference, assuming asymptotic normality for estimates, may lead to unreliable tests under weak identification, as it is a poor approximation to the true distribution of the estimator. The distributions are strikingly similar to those for the IV problem presented by Stock (2008).

The simulations for the empirical degree of identification illustrate just how far from truly proportional the variances changes can be, while still causing problems. The policy shock variance changes by a factor of 5, while the other changes by a factor of 1.8. Though appreciably different factors, the results display the symptoms of weak identification. It is also worth quantifying what sort of variance changes in the policy shock correspond to the standard approximations being "reasonable" – the $10 \times \delta$ calibrations. For δ, $\sigma^2_{\varepsilon_{2,P}}/\sigma^2_{\varepsilon_{2,C}} = 5$ and for $10 \times \delta$, $\sigma^2_{\varepsilon_{2,P}}/\sigma^2_{\varepsilon_{2,C}} \approx 34$. Translated to the variance of η_{2t}, daily changes in the Treasury yield, this means an increase from 0.002 to 0.007 on policy days; empirically, the increase is only from 0.002 to 0.004.

To further illustrate the implications of this problem in a macroeconomic context, I consider how weak identification can propagate to IRFs. I impose a reduced form VAR(1) specification using the lag coefficients A_1 estimated in Section 7. I take A_1 as known since, in most settings, it will be estimated consistently, and contribute relatively little to the variation in distribution of the IRFs. Figure 3 plots IRFs using 200 draws of $\hat{H}_{12}$. These should be concerning to macroeconomists – under weak identification, the distribution of the IRFs is highly diffuse.

[7]Given that the problem is just-identified, the weighting matrix is irrelevant to point estimates. Therefore, estimation can be conducted without resorting to numerical optimization via the eigenvector method described by Brunnermeier et al (2017). Given these point estimates, inference can be conducted by using the efficient GMM moment function.

Figure 2: Distribution of $t-$ratio for $\hat{H}_{12}$



The figure presents $t-$ratios, $\frac{\hat{H}_{12}-H_{12}}{se\left(\hat{H}_{12}\right)}$, calculated from 10,000 Monte Carlo draws, using the sample length in the left margin and the degree of identification in the bottom margin. Extreme outliers are truncated to allow comparison on the same axes. Calibration details are given in equation (11). Point estimation proceeds via Brunnermeier et al's (2017) eigenvector method, with inference using this solution for efficient GMM.

3 Weak identification in the general case

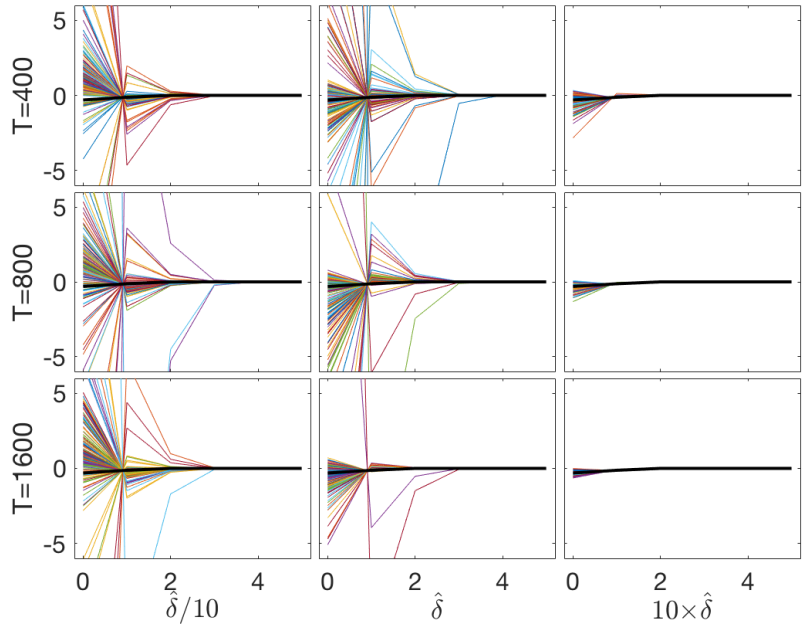
Having illustrated the problems posed by weak identification and presented asymptotic results for the *simple case*, I now consider the general model where all variances are permitted to change across regimes. This is the case in much empirical work (e.g. Rigobon (2003), Rigobon & Sack (2003), Craine & Martin (2008)). I first present the model more formally and discuss identification and estimation under strong identification. Then, I characterize how weak identification may arise.

3.1 Identification and estimation under strong identification

Having outlined the problem of weak identification in a simple bivariate setting, I now develop the general model, identification, and estimation more formally. As before,

$$\eta_t = H\varepsilon_t,$$

Figure 3: Distribution of IRFs



IRF paths for a unit shock for 200 Monte Carlo draws of $\hat{H}_{12}$, taking A_1 as given. Sample length is given in the left margin and the degree of identification in the bottom margin. Calibration details are given in equation (11). The horizontal axis is calibrated to days. Point estimation of H proceeds via Brunnermeier et al's (2017) eigenvector method.

where η_t and ε_t are now $n \times 1$ vectors, and H is $n \times n$. Often, η_t results from a reduced-form VAR of the form

$$A(L) Y_t = \eta_t. \tag{12}$$

Under standard assumptions, (correct specification, ergodicity, stationarity, and predetermined regressors) the lag coefficients of $A(L)$, and thus moments of the reduced form residuals η_t, can be estimated consistently by OLS, see e.g. Hayashi (2000), pg. 109. The approach of this paper applies equally in cross-sectional settings – the innovations need not come from a VAR. Indeed, the i.i.d. assumptions common with cross-sectional observations immediately imply many of the assumptions laid out for time series settings below. I generally treat η_t as observed, but account for the use of estimated VAR residuals in Section 6 when I consider inference on IRFs.

The second moments of η_t do not offer enough equations to recover H, furnishing only $\frac{n^2+n}{2}$ equations in n^2 unknowns ($n^2 - n$ in the normalized H, n in Σ_ε). Additional variance regimes may offer additional identifying variation. For simplicity of exposition, I henceforth focus on the two variance regime case.[8]

[8]For the most part, this is without loss of generality. It is mathematically straightforward to extend most results to settings with additional regimes. However, the results become much more cumbersome to state

To make the discussion precise, Assumption 1 outlines standard assumptions on the model.

Assumption 1. *For all* $t = 1, 2, \ldots, T$ *and regimes* $r \in \{P, C\}$,

1. H *is fixed over time, invertible, and has a unit-diagonal,*
2. $E\left[\varepsilon_t \mid t \in R_r, \mathcal{F}_{t-1}\right] = 0$, $E\left[\varepsilon_t \varepsilon_t' \mid t \in R_r, \mathcal{F}_{t-1}\right] = \Sigma_{\varepsilon,r}$,
3. $\Sigma_{\varepsilon,r}$ *is diagonal.*

where the filtration $\mathcal{F}_{t-1} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{t-1}\}$. The unit diagonal normalization takes the impact of a unit realization of structural shock i on the reduced form innovation to the i^{th} series to be one. Assumption 1.2 guarantees covariance stationarity within each regime. The addition of the filtration in the conditioning is less familiar, but satisfies the conditions for a martingale central limit theorem (used to derive asymptotic distributions below) without requiring full independence of the shocks over time.[9] The use of P (policy) and C (control) to denote the two regimes represents a typical framework that contrasts "event" observations and "Control" observations, arguing that on the event days, when, for example, a Policy announcement is made, the relevant structural shocks are likely to be more volatile than on a typical day.[10] This is the case in Nakamura & Steinsson (2018) (and many others - see e.g. Boyarchenko, Haddad, & Plosser (2017) and Wright (2012)).

Under Assumption 1, each regime has reduced form covariances given by

$$\Sigma_{\eta,r} = H\Sigma_{\varepsilon,r}H' \text{ for } r \in \{P, C\}. \tag{13}$$

Each variance regime offers $\left(n^2 + n\right)/2$ identifying equations. Thus, two variance regimes yield $2 \times \frac{n(n+1)}{2} = n^2 + n$ equations, with $n^2 - n + 2 \times n = n^2 + n$ unknowns (adding an additional n structural variances to the basic model). The system defined by (13) just meets the order condition (the number of equations equals the number of parameters). There is a unique solution (Rigobon (2003)), as long as the variances change non-proportionally across regimes.

clearly. Since the majority of empirical work uses two regimes, I restrict my focus to that case.

[9] Instead, weaker conditions like $m-$dependence can be substituted, and a central limit theorem for $m-$dependent sequences can be used. However, such deviations may also necessitate the use of long run variance estimators for each regime, adding further complexity.

[10] When regimes and breaks must instead be estimated, there can be substantial bias in identified parameters, as discussed in Lewis (2018).

Equation (13) can be rewritten as moment equations of the form

$$m\left(\eta_t, H, \Sigma_{\varepsilon,r}\right) = vech\left(\Sigma_{\eta,r} - H\Sigma_{\varepsilon,r}H'\right), \tag{14}$$

where date t is in regime r. Then, stacking the moment equations (14) from each regime, and defining the vector $\theta \in \Theta$ as the unique elements of H, $\Sigma_{\varepsilon,C}$, and $\Sigma_{\varepsilon,P}$, the resulting moment function is

$$\phi\left(\theta, \eta_t\right) = \begin{bmatrix} 1\left[t \in C\right] vech\left(\eta_t\eta_t' - H\Sigma_{\varepsilon,C}H'\right) \\ 1\left[t \in P\right] vech\left(\eta_t\eta_t' - H\Sigma_{\varepsilon,P}H'\right) \end{bmatrix}, \tag{15}$$

abusing notation to replace R_C with C and R_P with P in the indicator functions. Clearly, $E\left[\phi\left(\theta_0, \eta_t\right)\right] = 0$ at θ_0, the true parameter value. The GMM objective function is defined as

$$S_T\left(\theta; \tilde{\theta}\right) = \left[T^{-1/2}\sum_{t=1}^{T}\phi\left(\theta, \eta_t\right)\right]' W_T\left(\tilde{\theta}\right)\left[T^{-1/2}\sum_{t=1}^{T}\phi\left(\theta, \eta_t\right)\right]. \tag{16}$$

where $\tilde{\theta}$ is the parameter used to compute the weighting matrix, $W_T\left(\cdot\right)$. For the purposes of this paper, I focus on a continuous updating estimator (CUE) with the efficient weighting matrix (on which most weak identification results are based). This means $\tilde{\theta} = \theta$ and $W_T\left(\theta\right) = \Omega_T\left(\theta\right)^{-1}$, $\Omega_T\left(\theta\right) = \frac{1}{T}\sum \phi\left(\theta, \eta_t\right)\phi\left(\theta, \eta_t\right)'$. To characterize the asymptotic distribution of GMM estimates, regularity conditions such as those of Assumption 2 are required:

Assumption 2. *Assume*

1. *The process η_t is ergodic and stationary within regimes,*
2. *$E\left[vech\left(\eta_t\eta_t'\right) vech\left(\eta_t\eta_t'\right)' \mid t \in R_r\right] < \infty$ for $r \in \{P, C\}$,*
3. *$T_r/T = \tau_r > 0$, for $T_r = \left|R_r\right|, r \in \{P, C\}$.*

The first two points allow for the application of a martingale central limit theorem within each regime. The first point strengthens the covariance stationarity assumed within regimes in Assumption 1.2. The second is a standard moment existence condition. The third point guarantees that the sample size within each regime increases at the same rate as the overall sample size. Finally, I make a standard assumption on the parameter space:

Assumption 3. *Θ is compact.*

Under these assumptions, if there is a unique solution to (13), the usual arguments show that the GMM estimates of θ will be consistent and have the standard asymptotically normal GMM limiting distribution.

3.2 Weak identification

The intuition for identification was presented in 2.1 and sketched for the $n-$variable setting in 3.1. Now, I examine the argument in detail and show how weak identification can arise. Sentana & Fiorentini (2001) provide identification results for models characterized by equations of the form (13).[11] In particular, they define conditions for H to be globally identified (up to column order) in the presence of time-varying volatility from (13), which I simplify in Proposition 2:

Proposition 2. *Under Assumption 1, H is globally identified from $\Sigma_{\eta,C}$ and $\Sigma_{\eta,P}$ up to column order provided the rows of $\left[\begin{array}{cc} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right) \end{array} \right]$ are not proportional.*

Under an additional assumption, distinguishing the columns of H (or the shocks), point identification holds, instead of identification up to column order. I adopt Assumption 4, a common choice to this effect:

Assumption 4. *The shock of interest experiences the largest relative change in variance across regimes.*

Proposition 2 implies that identification can break down in two related ways. First, if the variances do not change, then $diag\left(\Sigma_{\varepsilon,C}\right) = diag\left(\Sigma_{\varepsilon,P}\right)$, and the rows are clearly proportional. Second, if two (or more) variances change by the same (potentially substantial) factor, those two rows are again proportional. This means that even if ample heteroskedasticity is present, identification is not guaranteed. As described in detail in the *simple case*, I model the relationship between the variances of two shocks, i and j, as local-to-unity:

$$\frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1 + \frac{d}{\sqrt{T}},$$

where d is finite. In economic terms, the Great Moderation or Financial Crisis were offered above as examples where variances might change together. This could also be the case for many modern treatments of monetary policy that posit the existence of simultaneous shocks to multiple dimensions of monetary policy; the volatility of all such shocks is likely to increase together on announcement days. If instead the variances barely differ across regimes, that too can be captured in this device, as both the numerator and denominator on the left-hand-side are close to unity.

[11] I choose to work from their results, in conjunction with the argument offered by Brunnermeier et al (2017) as the rank-like conditions for identification are immediately apparent.

Now, denote $\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}} = \gamma$, so

$$\sigma^2_{\varepsilon_{i,P}} = \gamma\sigma^2_{\varepsilon_{i,C}}\left(1 + \frac{d}{\sqrt{T}}\right) \equiv \gamma\sigma^2_{\varepsilon_{i,C}} + \frac{d_\varepsilon}{\sqrt{T}}. \tag{17}$$

The impact on identification is then characterized in Proposition 3:

Proposition 3. *Adopting the modeling device in (17) and Assumption 1, H is asymptotically unidentified.*

Intuitively, under the modified local-to-unity modeling device, the non-proportionality requirement of Proposition 2 fails asymptotically in population, as the variances converge to the knife-edge case $\sigma^2_{\varepsilon_{i,P}} = \gamma\sigma^2_{\varepsilon_{i,C}}$, resulting in an unidentified system. However, the limiting probability of rejecting the hypothesis $\frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1$ from (infeasible) observations of ε_t is neither zero nor one, capturing the spirit of the intermediate case of weak identification. As identification breaks down, H cannot be consistently estimated, as argued by Stock & Wright (2000). Similarly, standard asymptotic approximations used for inference also fail. Dufour (1997) (Section 4) demonstrates the dramatic impact that such deficiencies can have on testing problems. He shows that the size of Wald tests for affected parameters tends to unity as such a system tends towards non-identification. In this setting, it is straightforward to confirm, fixing d, that the size of a Wald test on the full parameter vector is unity asymptotically. In fact, this is borne out by the Monte Carlo simulations in Table 1, illustrating the severity of the problems posed by weak identification.

4 Weak identification robust inference

Having characterized how weak identification may arise, I now present inference approaches robust to weak identification. I first consider existing results that apply when the object of interest is the full parameter vector. I then extend these results to allow for non-conservative inference on subsets of parameter vector. Finally, I demonstrate the power improvements offered by these subset tests.

4.1 Asymptotic distribution of test statistics

The asymptotic behaviour of GMM estimators, robust to weak identification, was established in Stock & Wright (2000). Instead of providing an asymptotic distribution for the parameter estimates, as in strongly identified GMM problems, they show that $S_T(\theta_0)$ follows a Chi-square distribution. Many refinements have since been developed, including

the "$K-$statistic" of Kleibergen (2005), which is efficient under strong identification, additionally making use of an estimated Jacobian. Most of this literature is limited to joint tests on the full parameter vector or the subset of parameters that are weakly identified; inference results for generic subsets of the parameter vector have proven difficult, even on a case-by-case basis. However, the parameter(s) of interest in applied work is generally such a subset. In this section, I present standard results for tests on the full parameter vector, and then establish conditions under which test statistics for subsets of the parameter vector have a more precise limiting distribution. Robust inference on the full parameter vector (and the subset of all weakly identified parameters) in models identified via heteroskedasticity has already been considered as a motivating example in Magnusson & Mavroeidis (2014), who propose a variety of tests. However, the subset tests I develop constitute a generalization of these results important for applied work.

I present my results using Kleibergen's (2005) "$K-$statistic". In the leading two-regime case considered here, the $K-$statistic coincides with the $S-$statistic of Stock & Wright (2000) since the model is just-identified. If additional regimes are used, the $K-$statistic will be asymptotically efficient under strong identification, at the cost of possible power loss otherwise. Numerous further refinements exist, frequently using convex combinations of K and S (for example Conditional Linear Combination tests as in Andrews' (2016)), and can be considered by researchers using over-identified models. Readers interested in such test statistics should consult Magnusson & Mavroeidis (2014), who establish the validity of several test statistics for the full parameter vector; these tests can the be generalized for subset inference using results I prove below.

Full vector inference

I begin by demeaning the moment function and its Jacobian,

$$\begin{aligned}
\bar{\phi}(\theta, \eta_t) &= \phi(\theta, \eta_t) - E(\phi(\theta, \eta_t)), \\
q(\theta, \eta_t) &= vec\left(\frac{\partial \phi(\theta, \eta_t)}{\partial \theta'}\right), \\
\bar{q}(\theta, \eta_t) &= q(\theta, \eta_t) - E(q(\theta, \eta_t)),
\end{aligned}$$

as in Kleibergen (2005), with ϕ replacing his f. Lemma 1 provides asymptotic distributions for $\bar{\phi}(\theta_0, \eta_t)$ and $\bar{q}(\theta_0, \eta_t)$:

Lemma 1. *Under Assumptions 1 & 2,*

$$\psi_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \bar{\phi}(\theta_0, \eta_t) \\ \bar{q}(\theta_0, \eta_t) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_\phi \\ \psi_{\theta_0} \end{pmatrix}$$

where $\psi = \begin{pmatrix} \psi_\phi \\ \psi_{\theta_0} \end{pmatrix}$ *is a* $2(n^2+n)$*-dimensional normally distributed random variable with mean zero and positive semi-definite* $2(n^2+n) \times 2(n^2+n)$*-dimensional covariance matrix*

$$V(\theta) = \begin{pmatrix} V_{\phi\phi}(\theta) & V_{\phi\theta}(\theta) \\ V_{\theta\phi}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix} = \lim_{T\to\infty} var\left[\frac{1}{\sqrt{T}} \begin{pmatrix} \phi_T(\theta) \\ q_T(\theta) \end{pmatrix}\right]$$

where $\phi_T(\theta) = \sum_{t=1}^{T} \phi(\theta, \eta_t)$ *and* $q_T(\theta) = \sum_{t=1}^{T} q(\theta, \eta_t)$*.*

Lemma 2 provides additional properties needed for the use of the estimated Jacobian in the $K-$statistic:

Lemma 2. *Under Assumptions 1 & 2, the moment covariance matrix estimator* $\hat{V}(\theta_0)$ *satisfies*

$$\hat{V}(\theta_0) \xrightarrow{p} V(\theta_0)$$

and

$$\frac{\partial vec\left(\hat{V}_{\phi\phi}(\theta_0)\right)}{\partial \theta'} \xrightarrow{p} \frac{\partial vec\left(V_{\phi\phi}(\theta_0)\right)}{\partial \theta'}$$

These lemmata, proven in the Appendix, mirror those of Kleibergen (2005), and similarly establish Theorem 1 of that paper, which I replicate here:

Theorem 1. *If Lemmata 1 and 2 hold,*

$$K_T(\theta_0) \xrightarrow{d} \chi^2_{n^2+n},$$

where $K_T(\theta_0) = \frac{1}{T}\phi_T(\theta_0)' \hat{V}_{\phi\phi}(\theta_0)^{-1/2} P_{\hat{V}_{\phi\phi}(\theta_0)^{-1/2}\hat{D}_T(\theta_0)} \hat{V}_{\phi\phi}(\theta_0)^{-1/2} \phi_T(\theta_0)$*.*

P_A is the projection matrix $A(A'A)^{-1}A'$ and $\hat{D}_T(\theta_0)$ is Kleibergen's (2005) Jacobian

estimator. In particular,

$$\hat{D}_T(\theta_0) = \Big[q_{1,T}(\theta_0) - \hat{V}_{\theta\phi,1}(\theta_0)\hat{V}_{\phi\phi}(\theta_0)^{-1}\phi_T(\theta_0) \dots \\ q_{(n^2+n),T}(\theta_0) - \hat{V}_{\theta\phi,(n^2+n)}(\theta_0)\hat{V}_{\phi\phi}(\theta_0)^{-1}\phi_T(\theta_0) \Big],$$

where $\hat{V}_{\theta\phi}(\theta_0) = \left(\hat{V}_{\theta\phi,1}(\theta_0)', \dots, \hat{V}_{\theta\phi,(n^2+n)}(\theta_0)'\right)'$ ($\hat{V}_{\theta\phi,i}(\theta_0)$ are n^2+n square matrices) and $q_T(\theta) = \left(q_{1,T}(\theta), \dots, q_{(n^2+n),T}(\theta)\right)$ ($q_{i,T}(\theta)$ are n^2+n vectors). Theorem 1 provides an asymptotic distribution for the $K-$statistic under the enumerated assumptions. This is not novel; it is the *split-KLM* test considered by Magnusson & Mavroeidis (2014).

Monte Carlo: full vector

Monte Carlo evidence shows that tests based on these identification-robust asymptotics perform far better than Wald tests under weak identification. The distributions of $t-$ratios in the previous section suggests that strong identification asymptotics may break down. To investigate the performance of such tests, I again calibrate Monte Carlo simulations to equation (11). For each configuration, I take 10,000 draws and compute Wald statistics and $S/K-$statistics testing the null hypothesis of the true parameter vector (a six-restriction test). Rejection rates for nominally 5% tests are reported in Table 1. The Wald tests exhibit extremely large size distortions, aligned with Dufour's (1997) theoretical result that the size of such tests will tend to unity as the degree of identification tends to zero. The distortions improve with the strength of identification. The $S/K-$tests, however, are not systematically affected by the degree of identification, as expected of robust tests. Their size does decrease with sample size, which is indicative of small sample behaviour, not a lack of robustness. It appears that the performance of Wald-based inference approaches an acceptable level only for variance changes an order of magnitude larger than those observed empirically. As noted in Section 2.3, the stronger calibration imposes over six times the empirically observed change in the structural variance of the policy shock.

Subset inference

Inference on a subset of the parameter vector is challenging under weak identification. Previous work with identification via heteroskedasticity has skirted the problem of subset inference besides in the *simple case*, leaving interested researchers, should they worry about weak identification, to rely on projection methods based on a full-vector test statistic, like the $S-$statistic (see Dufour (1997), Dufour & Taoumouti (2005), Chaudhuri (2008), Chaudhuri

Table 1: Size of tests on the full parameter vector

	$\hat{\delta}/10$		$\hat{\delta}$		$\hat{\delta} \times 10$	
	Wald	S/K	Wald	S/K	Wald	S/K
$T = 400$	94.6	14.0	75.9	14.7	33.6	13.9
$T = 800$	94.3	10.4	70.4	10.2	24.1	10.2
$T = 1600$	93.5	7.7	63.3	8.3	16.2	7.7

Rejection rate of the true parameter vector for a nominally 5% test, based on 10,000 Monte Carlo draws. Calibration details are given in equation (11). Estimation via CUE GMM.

et al (2010), and Chaudhuri & Zivot (2011) for approaches to projection methods and possible refinements therein). Projection methods are notoriously conservative; the test statistic is minimized conditional on the parameter(s) of interest, but is compared to the same critical values as for the full-vector test.

However, Kleibergen (2005) provides a refinement over Theorem 1 for tests on a subset of parameter vector. Partition θ into the parameter(s) of interest, β, and the remainder, α. Assuming the asymptotic Jacobian conditional on the parameters in β,

$$J_\alpha(\alpha, \beta) = \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\partial \phi((\alpha,\beta),\eta_t)}{\partial \alpha'}\right)\Big|_{\alpha,\beta}\right]\right\}, \tag{18}$$

has rank equal to α, then the degrees of freedom of the limiting distribution of $K-$statistic is lowered to the dimension of β (Kleibergen (2005), Theorem 2). This is a generalization of the condition Stock & Wright (2000) use to derive their concentrated $S-$statistic for tests on the subset of all weakly identified parameters. I present a global analogue to this Jacobian assumption and then characterize conditions under which the class of models considered here satisfies this assumption.

Assumption 5. *Conditional on β, α is asymptotically strongly identified (globally).*

Define $K_T(\beta) = K_T(\beta, \alpha(\beta))$, where $\alpha(\beta) = \underset{\alpha}{\operatorname{argmin}}\, K_T(\beta, \alpha)$. Theorem 2 of Kleibergen (2005) implies Theorem 2:

Theorem 2. *If Lemmata 1 and 2 hold, then under Assumption 5,*

$$K_T(\beta_0) \xrightarrow{d} \chi^2_{p_{int}},$$

where p_{int} is the dimension of β.

This lowers the degrees of freedom of the limiting distribution from $n^2 + n$ for the full parameter vector (or projection tests) to p_{int}. It nests the full-vector results of Theorem 1, as $\beta = \theta$ clearly satisfies Assumption 5. I henceforth refer to the test comparing $K_T(\beta_0)$ to the $\chi^2_{p_{int}}$ critical values as the "reduced" test due to the degrees of freedom reduction.

When does the model satisfy Assumption 5? I begin by extending Proposition 2 to allow for partial identification of H. First, I introduce a partition of H:

Definition 1. Partition H as $H_I \vdots H_W$ such that $H^{(k)} \in H_I$ if and only if $\begin{pmatrix} \sigma^2_{k,C} & \sigma^2_{k,P} \end{pmatrix}$ is proportional to no other row in $\begin{bmatrix} \sigma^2_{1,C} & \sigma^2_{1,P} \\ \vdots & \vdots \\ \sigma^2_{n,C} & \sigma^2_{n,P} \end{bmatrix}$, and conversely for H_W.

Note that, by definition, if not empty, H_W contains at least two columns. Now, H_I is always identified without reference to H_W:

Proposition 4. *Under Assumption 1, H_I is identified from the covariance matrices across regimes.*

This result shows that Assumption 5 is satisfied if β includes H_W and associated variances. This is analogous to Theorem 3 of Stock & Wright (2000), the asymptotic distribution of the concentrated $S-$statistic, for tests on all weakly identified parameters. This is not necessarily helpful unless a researcher's subset of interest contains all weakly identified parameters (in the bivariate model, weak identification contaminates the full parameter vector, so this is still just a test on the full parameter vector).

In empirical work, the object of interest is generally either the immediate impact of one shock on one variable or its impact on all variables. The former consists of a single element of H; the latter pertains to a full column. Therefore, I extend the identification result of Proposition 4 from H_I to all of H in Theorem 3, by conditioning on either an element of H or a full column:

Theorem 3. *Under Assumption 1, if H_W contains two columns, H is conditionally identified from the covariance matrices provided*

1. *A single element H_{lk} is fixed and $H_{lk} \neq H_{lm}/H_{km}$ for $H^{(k)}, H^{(m)} \in H_W$, or*

2. *The full column $H^{(k)} \in H_W$ is fixed.*

By explicitly incorporating the information to be used in the null hypothesis of the subset test (fixing H_{lk} or $H^{(k)}$), I obtain conditional (strong) identification for the remainder of H_W by Theorem 3. This means that a system of equations satisfying the conditions

of Theorem 3 meets Assumption 5, so Theorem 2 applies. The ancillary condition on the relative magnitudes of elements of H_W can be thought of as strengthening the standard invertibility condition on H to an invertibility assumption on a sub-block of H; in the empirically common bivariate case, it coincides with the invertibility assumption.

Condition 1 interprets the result of Theorem 3 through the lens of the model, abstracting from the knife-edge $H_{lk} = H_{lm}/H_{km}$ case.

Condition 1. *If there are at most two variances, i, j, for which $\lim_{T\to\infty} \frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1$, and i or j is the shock of interest, then Assumption 5 is satisfied for tests where β contains a single element of the corresponding column of H or the full column of H (plus any additional parameters conditioned upon), and Theorem 2 holds.*

The result constitutes an improvement on the previous possibilities for subset inference since it permits the use of smaller critical values. Five remarks clarify its impact.

Remark 1. Condition 1 nests the case where $\beta = \theta$ and the case where β is the set of weakly identified parameters.

Remark 2. The shock of interest must be one of those affected by any variance pathology. Otherwise, fixing a parameter(s) in the column of interest of H conveys no information about the columns impacted by the pathology.

Remark 3. Given how few variances are permitted to be affected by these pathologies, a researcher should err towards minimizing the number of series in the system of equations subject to the constraint that the reduced form innovations span the structural shocks (invertibility).

Remark 4. In empirical practice, it is not uncommon to work with bivariate systems. In this case, the limit of two proportional series is non-binding, and the condition on the relative magnitudes of elements in H_k and H_m collapses to the usual invertibility assumption of H.

Remark 5. Condition 1 can be generalized further for less empirically common scenarios, where particular combinations of parameters from across the columns of H are to be jointly tested. For example, it could be extended to the case where there are two pairs of proportional variances if columns of H from each pair are jointly fixed.

It is also important to note the relation between these subset results and partial identification. Proposition 4 shows H may be partially identified; the columns of H_I can be estimated, and robust tests on the full parameter vector will be valid. However, the subset results do not immediately apply in the case of partial identification since, conditional on the subset of interest, they require the remainder of the model to be identified, in line with

Table 2: Size of $t-$test, projection $S-$test, and reduced test on $\hat{H}_{12}$

	$\hat{\delta}/10$			$\hat{\delta}$			$10 \times \hat{\delta}$		
	t	S_{proj}	$S_{reduced}/K_{reduced}$	t	S_{proj}	$S_{reduced}/K_{reduced}$	t	S_{proj}	$S_{reduced}/K_{reduced}$
$T = 400$	53.7	0.0	4.4	16.8	0.0	4.6	8.5	0.0	4.8
$T = 800$	48.3	0.0	4.7	12.8	0.0	5.1	6.6	0.0	4.6
$T = 1600$	40.4	0.0	4.9	9.7	0.1	4.9	5.4	0.0	4.8

Rejection rate of the true parameter value for H_{12} based on 10,000 Monte Carlo draws. The S_{proj} results are not identically zero, but round to 0.0. Calibration details are given in equation (11). Estimation via CUE GMM.

Remark 2. An exception arises if the reason for partial identification is that only the variance of interest changes, when the results for the *simple case* apply.

Nakamura & Steinsson (2018) compute robust confidence intervals for a single parameter of interest using what they refer to as a "Fieller's method" bootstrap, drawing on Staiger, Stock, & Watson (1997) (and Fieller (1954)). This approach only works in their *simple case*, since this setting means means that their test statistic depends only on H_{12}. With multiple variance changes, the test statistic depends on structural parameters other than H_{12}, and thus cannot be used to test values of H_{12} without specifying values for the other parameters, returning to the full parameter vector/projection problem. This is unsurprising, since their test asymptotically coincides with an $S-$test.

Monte Carlo: subset

Repeating the Monte Carlo assessment of size-distortion for subset tests demonstrates scope for improvement over standard procedures. I compare three testing approaches: the Wald test, the projected $S-$test based on Theorem 1, and the newly proposed reduced $S/K-$test, based on Theorem 2 and Condition 1. The results are displayed in Table 2. The projected $S-$test represents the best available option for inference in this general setting, absent the results of Condition 1. First, like Wald tests on the full parameter vector, the standard $t-$test is substantially oversized, though the distortion is not as large as for the full vector. As identification gets stronger, the distortions shrink. The $S-$test based on projection methods is substantially undersized, with a rejection rate of effectively zero in simulation. However, the reduced test is consistently well-sized, regardless of the degree of identification. These improvements in size-control over previously available tests establish the usefulness of Condition 1 for applied work.

4.2 Power improvements in subset testing

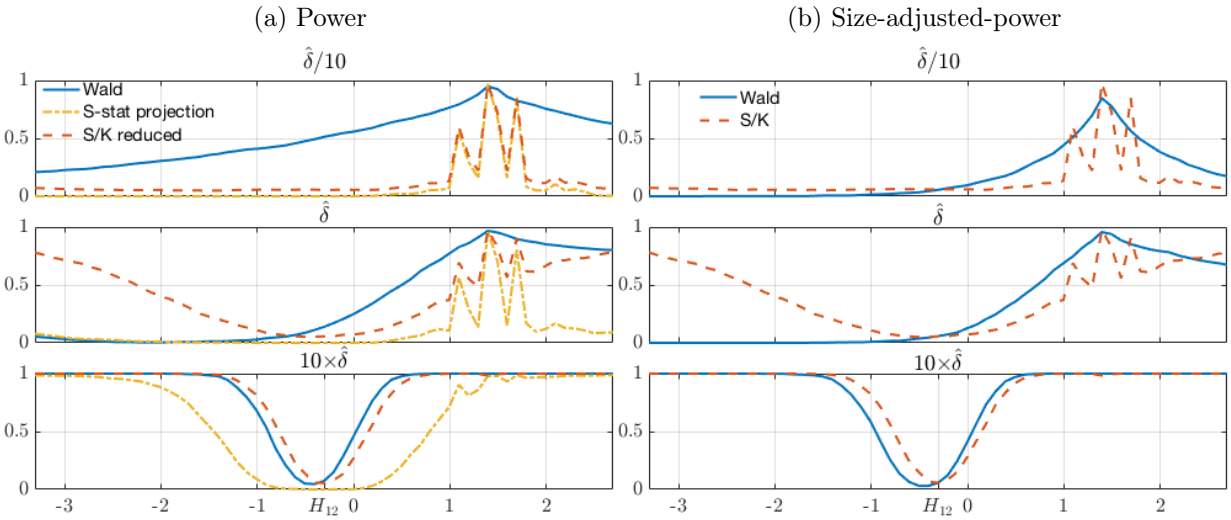
The use of smaller critical values for the subset tests justified by Condition 1 relative to projection tests imply automatic power improvements, which I explore in simulation. I fix $T = 800$ and consider a range of strengths of identification, testing the null hypothesis of $H_{12} = -0.31$ against a sequence of local alternatives. Panel (a) of Figure 4 computes power curves based on these simulations. For the weakest identification calibration, the power of the $t-$test unsurprisingly dominates that of the robust tests, but is still very low for alternatives far left of the null. The reduced test is more powerful than the projection test, as expected; they use the same test statistic, but the reduced test uses smaller critical values. For the main calibration, the results are similar, except that the reduced test surpasses the $t-$test with power tending to unity for alternatives left of the null, and approaches the $t-$test for alternatives to the far right. It also diverges dramatically from the projection test. For the strong identification calibration, the t and reduced tests are very similar, as expected in this context. The projection test is dominated, but does approach the others for more distant alternatives. The non-monotonic behaviour between $H_{12} = 1$ and $H_{12} = 2$ is discussed in the Appendix.

Panel (b) repeats the above exercise, but now plots size-adjusted power instead of power for the $t-$test and $S/K-$test (the projection vs. reduced distinction is mute since the size-adjusted power of both tests is identical mechanically). Across the board, this increases performance of the robust test compared to the $t-$test, even surpassing it for some distant alternatives in weak calibrations. There is some power loss due to using robust inference, particularly for alternatives close to the null, but for some alternatives, the reduced test actually dominates the $t-$test. Any power loss must be weighed against the substantial size distortion of standard inference. The reduced test certainly (mechanically) offers a substantial improvement over projection tests. Recall that in some contexts, there may also be scope for more powerful tests (for example, the Conditional Linear Combination test of Andrews (2016)).

5 Tests for weak Identification

Tests for weak identification pose a challenge in general GMM settings. It is straightforward to reparametrize the GMM framework to afford tests of a complete lack of identification (i.e. proportional variance changes, $\delta = 0$); however, a suitable null is rather the presence of *weak* identification (δ close enough to zero to cause some measure of unreliability in inference). Existing tests for heteroskedasticity (e.g. Lütkepohl & Milunovich (2016)) are

Figure 4: Power curves



Power curves formed from estimates of rejection rates of the null hypothesis ($H_{12} = -0.31$) against a sequence of local alternatives ($x-$axis) based on 1000 Monte Carlo draws. Panel (a) reports power and panel (b) size-adjusted-power. The size-adjusted critical values are based on quantiles from 10,000 Monte Carlo draws. The far right spikes in the robust methods is discussed in the Appendix and Figure 6. Calibration details are given in equation (11). Estimation via CUE GMM.

parametric and test for non-identification, not weak identification. The *simple case* with a single variance change coincides with just-identified linear IV with a single endogenous regressor, and a rule of thumb based on bias is provided based on Montiel Olea & Pflueger (2013), in the spirit of Staiger & Stock (1997) and Stock & Yogo (2005). For the general case, I recommend the approach of Andrews (2017).

5.1 Single-variance change rule of thumb

The analogy to the just-identified linear IV model with a single endogenous regressor is an appealing feature of the *simple case*, which has proven popular in the literature. However, the *simple case* comes at the cost of the assumption that only the variance of the shock of interest changes across regimes. The linear IV setting with a single endogenous regressor is that of Staiger & Stock (1997), who derive the now ubiquitous "$F > 10$" rule of thumb. Here, the first-stage $F-$statistic is based on the first-stage regression,

$$\begin{aligned}\eta_{2t} &= \Pi Z_t + \nu_t \\ &= \Pi \left[\mathbf{1}\left(t \in P\right) \times \frac{T}{T_P} - \mathbf{1}\left(t \in C\right) \times \frac{T}{T_C} \right] \eta_{2t} + \nu_t,\end{aligned}$$

and should be computed allowing for heteroskedasticity. Both Staiger & Stock (1997) and Stock & Yogo (2005) assume homoskedasticity in deriving bias- and size-based rules of thumb.

Table 3: Critical values for first-stage $F-$test based on TSLS bias

Bias	0.05	0.1	0.2	0.3
Critical Value	37.42	23.11	15.06	12.05

Critical values for the first-stage $F-$statistic from Montiel Olea & Pflueger (2013); estimates deliver bias less than or equal to that listed in 95% of samples.

However, Montiel Olea & Pflueger (2013) develop a test that is valid under the presence of arbitrary heteroskedasticity, autocorrelation, or clustering. They use a "scaled $F-$statistic", which, in this simplest setting, reduces to the standard $F-$statistic. They compare the Nagar bias of a TSLS estimator to a "worst case" benchmark. The Nagar bias approximates the distribution of the TSLS estimator under weak identification using a second-order Taylor approximation and computes the bias of that distribution. The exercise coincides with that of Staiger & Stock (1997) and Stock & Yogo (2005) under those papers' distributional assumptions. Table 3 presents the critical values for the first-stage weak instruments test based on TSLS bias (with a single instrument) for the 5% level from Montiel Olea & Pflueger (2013), analogous to Table 5.1 in Stock & Yogo (2005). Additional unreported simulations, based on an identification via heteroskedasticity-specific DGP, produced critical values within Monte Carlo error of those of Montiel Olea & Pflueger. The size exercise of Stock & Yogo (2005) is not pursued as their sufficient statistic approach does not work in this nonlinear model. The generally-adopted threshold in IV of $F > 10$ corresponds to the critical value for relative bias of 10%, which here translates to a new rule of thumb of $F > 23$, which can easily be adopted in future work in this setting.

5.2 Method for the general case

In the general case, more complex methods are required. Tests of the null of non-identification are straightforward, as noted in Wright (2001), but testing for weak identification in GMM has long been a problem. However, Andrews (2017) offers a two-step strategy. The necessary assumptions for its application are discussed and verified in the Appendix. Briefly, given a user-specified maximum allowable size distortion, ξ, a preliminary robust confidence set is constructed to have asymptotic coverage $1 - \nu - \xi$, regardless of identification. Then, $1 - \nu$ non-robust and robust confidence sets are constructed. If the preliminary confidence set is contained by the non-robust confidence set, the non-robust confidence set can be adopted (strong identification); otherwise, a robust set should be used (weak identification). Asymptotically, the probability of making the correct determination converges to unity. These confidence sets can be constructed using $K-$statistics and Wald tests as elsewhere in this paper; Andrews, however, considers a linear combination of the

$K-$ and $S-$statistics due to favourable power properties. Determination of the strength of identification can be conducted with respect to the full parameter vector or a subset of parameters of interest by employing confidence sets for the corresponding parameters. For the full parameter vector, this test can be applied using full-vector confidence sets (established by Magnusson & Mavroeidis (2014)). However, to test the strength of identification of a single parameter, the exercise most relevant in empirical work, subset confidence sets are required. Projection tests lead to highly conservative subset confidence sets (as demonstrated in Section 4), so are not well-suited to this exercise. Because they are too large, they are unlikely to be contained by the non-robust confidence sets, even under strong identification. This means that the test will detect weak identification even when the parameter of interest is strongly identified, which is highly unappealing in practice. However, the non-conservative subset tests valid under Condition 1 do not face these problems and thus make it possible to conduct tests of identification for the parameter(s) of interest without systematically under-rejecting weak identification. Both this test for the general model and the first stage $F-$test for the *simple case* are applied in the empirical application.

6 Robust inference on impulse responses

The ability to perform robust inference on subsets of H enables robust inference on another object of interest, the impulse response function (IRF). Much macroeconomic policy analysis relies on dynamic effects traced out by IRFs. Based on (12), these are formed via non-linear combinations of the SVAR lag coefficients that comprise $A(L)$ and H. Little work considers robust inference on IRFs. Montiel Olea, Stock & Watson (2016) develop a method that exploits the linearity of the external instruments problem to offer an elegant solution in that context. Chevillon, Mavroeidis, & Zhan (2016) consider the case of long-run restrictions and develop a projection method that also accounts for cointegration issues they face. Neither of those methods applies to identification via heteroskedasticity. I propose a robust inference method for IRFs that can be extended to other SVAR identification schemes for which conditional identification results analogous to Theorem 3 validate subset inference on a column of H.

A structural impulse response function, Λ^h, at horizon h is computed as

$$\Lambda^0 = H,$$

$$\Lambda^h = \left[\sum_{v=1}^{h} \Lambda^{h-v} A_v\right] \quad h = 1, 2, \ldots, \tag{19}$$

where A_v denotes the lag coefficient matrix corresponding to the v^{th} lag. The response of

the l^{th} variable to a unit shock to k can be read off as the lk element of this object. It is helpful to define the related object, B^h given by

$$B^0 = I,$$

$$B^h = \sum_{v=1}^{h} B^{h-v} A_v \; h = 1, 2, \ldots,$$

such that $\Lambda^h = B^h H$. Thus, $B^h = B^h\left(A\left(L\right)\right)$, entirely a function of the lag coefficients. An element of interest in Λ^h, Λ^h_{lk}, is the product of the l^{th} row of B^h and the k^{th} column of H.

With necessary objects defined, I now present conditions under which the remainder of Λ^h is strongly identified *conditional* on Λ^h_{lk}. First, I partition Λ^h as H was partitioned in Definition 1 and state an analogue to the partial identification result of Proposition 4:

Definition 2. Partition Λ^h as $\Lambda^h_I \vdots \Lambda^h_W$ such that $\Lambda^{h(k)} \in \Lambda^h_I$ if and only if $\begin{pmatrix} \sigma^2_{k,C} & \sigma^2_{k,P} \end{pmatrix}$ is proportional to no other row in $\begin{bmatrix} \sigma^2_{1,C} & \sigma^2_{1,P} \\ \vdots & \vdots \\ \sigma^2_{n,C} & \sigma^2_{n,P} \end{bmatrix}$, and conversely for Λ^h_W.

Proposition 5. *Under Assumption 1, if B^h is invertible, Λ^h_I is identified up to scale from the covariance matrices across regimes.*

This follows the same argument as Proposition 4, simply replacing H with $B^h H$; invertibility of the product is guaranteed by invertibility of B^h and H. IRFs for shocks with non-proportional variance processes are strongly identified. As in the case of H, the crucial step is to extend this result (conditionally) to weakly identified shocks:

Theorem 4. *Under Assumption 1, if Λ^h_W contains two columns, Λ^h is conditionally identified up to scale from the covariance matrices provided*

1. *B^h is invertible,*
2. *A single element Λ^h_{lk} is fixed and $\Lambda^h_{lk} \neq \Lambda^h_{lm}/\Lambda^h_{km}$ for $\Lambda^{h(k)}, \Lambda^{h(m)} \in \Lambda^h_W$.*

This is an analogue of Theorem 3, replacing H with $\Lambda^h = B^h H$. Identification up to scale may be unfamiliar for IRFs, but recall that IRFs for each shock are always implicitly scaled by the normalization of the corresponding column of H. Theorem 4 implies that, after fixing an element of interest in Λ^h, conditional identification in Assumption 5 is satisfied, subject to the same remarks made following Condition 1.

Following Theorem 4, robust inference for Λ^h_{lk} may proceed using the appropriate objective function, the corresponding $S-$statistic, and χ^2_1 critical values.[12] Let A denote the

[12] I focus on the $S-$statistic to facilitate the simplification of the test inversion problem discussed below.

vectorization of the parameters found in the lag coefficients $A(L)$. Denote the moment function for the reduced form VAR (12) as $f^A(A, Y_t)$. With η_t implicitly a function of Y_t and A, define

$$f^{\Lambda}\left(A, \Lambda^h, \Sigma_C, \Sigma_P, Y_t\right) = vech\begin{pmatrix} 1[t \in C]\left(B^h \eta_t \eta_t' B^{h'} - \Lambda^h \Sigma_C \Lambda^{h'}\right) \\ 1[t \in P]\left(B^h \eta_t \eta_t' B^{h'} - \Lambda^h \Sigma_P \Lambda^{h'}\right) \end{pmatrix}$$

The objective function, $S_T^{stack}\left(A, \Lambda^h, \Sigma_C, \Sigma_P\right)$, is computed for a horizon h using the moments

$$f^{stack}\left(A, \Lambda^h, \Sigma_C, \Sigma_P, Y_t\right) = \begin{pmatrix} f^A \\ f^{\Lambda} \end{pmatrix},$$

which stacks the moments yielding the reduced form VAR coefficients, $f^A(A)$, and those pertaining to Λ^h, $f^{\Lambda}(\cdot)$. Based on $S_T^{stack}(\cdot)$, a standard test-inversion procedure yields a confidence set for Λ_{lk}^h. The resulting set need not be connected, a common feature in weak identification settings. It is trivial to extend this method to cumulative IRFs. The uncertainty of the test will generally be dominated by uncertainty over $H^{(k)}$. The critical values are from the χ_1^2 distribution, as in the method of Montiel Olea, Stock, & Watson (2016) for the external instruments setting.

7 Empirical application

I demonstrate the use of my robust inference methods by studying the identification of monetary policy shocks in the setting of Nakamura & Steinsson (2018).[13] The authors analyze the impact of policy shocks on nominal and real Treasury instantaneous forward rates of varying maturities. They argue that the response of these forward rates captures forward guidance effects. They use identification via heteroskedasticity as a robustness check on their main results. They adopt a bivariate model with daily changes in a forward rate as the "dependent" variable and a second series that serves as a policy instrument. They consider two such instruments: the daily change in nominal 2-year Treasury yields and the 30-minute or daily change in a "policy news" series, which they construct as the first principal component of several interest rate series. They assume that the only shock exhibiting a variance change on announcement days is the monetary policy shock. They use announcement days as the "high-variance" regime, and a sample of analogous dates as the control period, or "low variance" sample. I examine specifications using both the daily Treasury yields and the authors' 30-minute window "policy news" series as the policy instrument, with either nominal

[13] I am very grateful to Emi Nakamura and Jòn Steinsson for making their "policy news" series available to me.

or real 2-year Treasury instantaneous forward rates as the "dependent" variable. Thus,

$$\eta_t = \begin{pmatrix} \Delta s_t \\ \Delta i_t \end{pmatrix}$$

where s_t is a forward rate and i_t is the policy instrument.

7.1 Specifications & tests of identification

Nakamura & Steinsson (2018) assume only the variance of policy shocks changes on announcement days. This places their analysis in the *simple case*, with analogy to just-identified linear IV with a single endogenous regressor. However, this paper focuses on estimators using all moments, allowing for the possibility that the variances of both structural shocks might change. Economically, it might be the case that only the variance of the policy shock should change, but if that is the case, the restriction need not be imposed mechanically, as estimation will bear it out. I thus focus on the unrestricted model.

Table 5 reports estimates for this model. For the 30-minute "policy news" shock, the results are extremely close to Nakamura & Steinsson's restricted model; compare 1.07 to 1.10 and 0.97 to 0.96. Thus, at high frequencies, the single variance change assumption has little impact on estimates of H_{12}; this is because H_{21} is near-zero, eliminating the possible bias in (6). Using daily changes in the nominal yield as the policy series gives varied results. The point estimates for the real forward rate are in keeping with the intraday results and those of Nakamura & Steinsson. In contrast, the negative pass-through to nominal forward rates is starkly at odds with the other estimates, and theory. The strongly positive value for H_{21} – the pass through of the second shock, interpretable as noise in the forward rates, to the policy series – also differs from the near-zero estimates in the original paper (and here when using the "policy news" shocks). These discrepancies suggest weak identification, and motivate a closer examination.

I test formally for weak identification using the methods proposed in Section 5. Nakamura & Steinsson reduce the model to the *simple case*; the first-stage $F-$statistic tests for weak identification under this assumption. These results are reported in the first panel of Table 4. For the daily nominal Treasury yield series, weak identification cannot be rejected at any level considered. In contrast, for the 30-minute "policy news" series, the first stage $F-$statistic is high and weak identification is rejected for all levels of bias. The general 2-step test for the unrestricted model is reported in the second panel. The daily nominal Treasury yield displays weak identification for all distortions. The 30-minute "policy news" series shows only mild evidence of weak identification at the 5 and 10% distortion thresholds (owing to the far right tail of the asymmetric robust confidence sets). These test results corroborate

Table 4: Tests of Identification

	First-stage F (bias)				Andrews 2-step (size)			
	F	0.2	0.1	0.05	0.2	0.15	0.1	0.05
Nominal, daily shock	8.15	×	×	×	×	×	×	×
Real, daily shock					×	×	×	×
Nominal, 30-min shock	6891.94	✓	✓	✓	✓	✓	×	×
Real, 30-min shock					✓	✓	×	×

The first panel tests each shock series using the first-stage $F-$statistic bias-based critical values in Table 3. The second panel conducts the Andrews 2-step size test for each specification. The acceptable distortions are those greater than or equal to the maximum threshold, Andrews' γ_{min}, the value at which the $1-\alpha-\gamma_{min}$ robust set is just contained by the strong identification set.

the less formal observations of Nakamura & Steinsson, who suspect moving from intraday to daily data weakens identification.

7.2 Performance of tests

I now compare confidence sets robust to weak identification to those computed assuming identification is strong. For the daily yield shocks (exhibiting weak identification), the robust confidence intervals are much wider than standard confidence intervals. However, they are substantially asymmetric, so do not always contain the standard confidence interval. Notably, the surprising estimate of $\hat{H}_{21} = 0.70$ (the impact of the unnamed second shock, potentially noise in the forward rates) for the nominal forward rate specification is highly statistically significant using standard methods, but not at all using the robust interval. For the 30-minute window "policy news" shocks (exhibiting strong identification), the robust confidence intervals are comparable with the standard ones, and the estimates of H_{12} remain statistically significant at the 5% or 1% level.[14] For H_{21}, I obtain (reasonably) precisely estimated zeros. These two conclusions replicate Nakamura & Steinsson's findings for the restricted model.

For the specifications using 30-minute shocks (treated as strongly identified), I can also test the null hypothesis that the non-policy shock variance is fixed across regimes using standard methods. This is the additional over-identifying assumption used in the original paper to reduce the model to the *simple case*. For the model with nominal forwards, $p = 0.12$ for a simple Wald test. While equality may not be soundly rejected, it is neither strong enough evidence of equality to maintain equality as an identifying assumption. For real forwards, $p = 0.65$, which is more compelling evidence of equality. This ambiguity supports the use of the unrestricted model in simulations in this paper, particularly since they are calibrated to the nominal forward data.

[14] Under strong identification, they should be asymptotically equivalent, but even if the model is strongly identified, this need not be true in finite samples.

Table 5: Estimates

dep. var., policy inst.	Nominal fwd., one-day yield	Real fwd., one-day yield	Nominal fwd., 30-min news	Real fwd., 30-min news
H_{21}	0.70	-0.01	0.01	0.00
stnd. CI	$[0.52, 0.89]$	$[-0.14, 0.12]$	$[-0.00, 0.02]$	$[-0.00, 0.01]$
robust CI	$[-42.25, 0.79]$	$[-0.51, 0.49]$	$[-0.01, 0.02]$	$[-0.00, 0.91]$
H_{12}	-0.31	0.95	1.07^{**}	0.97^{***}
stnd. CI	$[-5.12, 4.50]$	$[-0.60, 1.3]$	$[0.14, 2.01]$	$[0.41, 1.53]$
robust CI	$[-48.95, 1.50]$	$[-59.05, 2.36]$	$[0.25, 3.49]$	$[0.50, 2.46]$
$10^3 \times \sigma^2_{s,C}$	3.9	5.9	3.9	5.8
$10^3 \times \sigma^2_{i,C}$	0.1	2.1	0.02	0.02
$10^3 \times \sigma^2_{s,P}$	7.1	2.9	6.2	7.8
$10^3 \times \sigma^2_{i,P}$	0.5	4.0	0.8	0.8

GMM estimates allowing for changes in all variances. The "dependent variable" is the one-day change in either the nominal or real 2-year instantaneous forward rate on treasuries. The policy instrument is either one-day changes in the 2-year nominal Treasury yield or 30-minute changes in Nakamura & Steinsson's "policy news" series. For the variances, i denotes the monetary policy shock and s the second shock. The standard confidence interval is based on a $t-$statistic. The robust confidence interval is based on the reduced $K-$test proposed in the text. Stars indicate significance from zero at the 5 or 1% levels based on the more conservative of the two tests.

7.3 Application to IRFs

I also consider robust inference on the dynamic effects of monetary policy through IRFs. The policy-control event-study approach of Nakamura & Steinsson is not immediately suited to computing IRFs (which require a continuous sequence of observations for the VAR). Thus, I consider the full time series for daily changes in the yield on 2-year nominal treasuries and the 2-year nominal Treasury instantaneous forward rate for all trading days between January 2000 and March 2014, an extension of the Nakamura & Steinsson sample. I estimate a VAR(1) on this data:

$$y_t = \begin{pmatrix} \Delta s_t \\ \Delta i_t \end{pmatrix} = A_1 y_{t-1} + \eta_t.$$

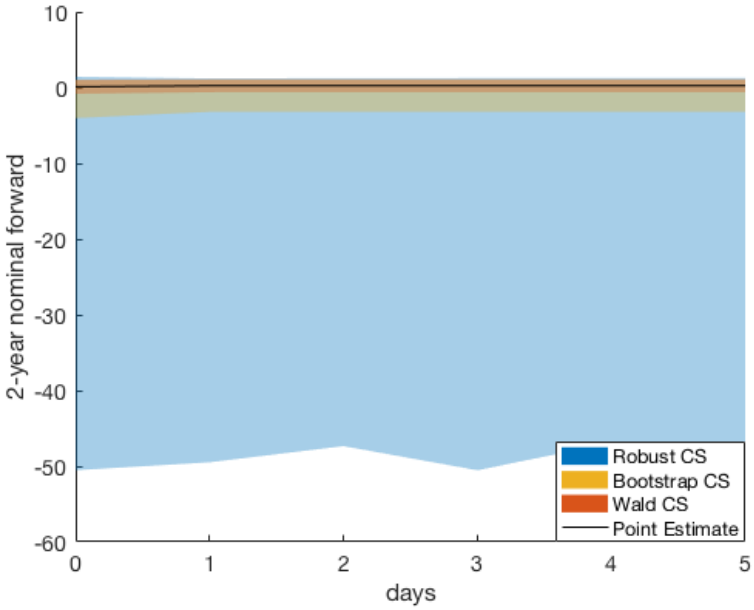
Estimation yields

$$\hat{A}_1 = \begin{pmatrix} -0.10 & 0.15 \\ -0.16 & 0.18 \end{pmatrix}.$$

Unsurprisingly, given weak identification, $\hat{H}$ differs from that for the main specifications; in particular, now $\hat{H}_{12} = 0.12$.

I apply the test-inversion procedure proposed in Section 6. For comparison, I compute

Figure 5: IRF Confidence Sets



Cumulative response (in b.p.) of 2-year nominal forward rates on Treasuries to a 1 b.p. monetary policy shock. The IRF path is computed based on the horizon-by-horizon IRF estimator (computing based on H_{12} and A_1 estimates from the 0^{th} horizons results in negligible differences). A Wald 95% confidence set is plotted, along with a bootstrap confidence interval following a block bootstrap. The collapse of the confidence sets towards zero with the point estimate is due to the precision with which the near-zero A_1^h values are estimated.

two standard confidence sets. The first is a Wald confidence set using GMM and the delta method. The second is computed using a conventional block bootstrap procedure with block length equal to sixty days. Figure 5 plots the cumulative IRF with these 95% confidence intervals. For the sake of clarity in the figure, the robust set is plotted as a connected set, even though there are some "gaps" within it that can be rejected. The bootstrap set captures the asymmetry of the robust set, not present in the Wald set. However, the Wald and bootstrap sets are both dramatically under-sized. These results show that inference based on non-robust confidence sets is likely to be misleading and exhibit substantial size distortions.

8 Conclusion

This paper provides a comprehensive framework allowing researchers to conduct inference robust to weak identification in models identified via heteroskedasticity. I describe and model the deficiencies that can lead to such weak identification, and show that these properties can significantly impact the reliability of standard inference in empirical data. I propose tests to detect weak identification, allowing researchers to determine whether they ought consider

these concerns.

I show that robust inference for a subset of the parameter vector can use smaller critical values than those required for projection methods. Tests based on this result are consistently well-sized, while standard tests are substantially over-sized and projection tests are highly conservative. Given the difficult problem posed by subset robust inference in nonlinear models, the approach taken here of deriving conditional identification results suggests an outline for those interested in other models. I extend my results to IRFs in a way that can also be followed for other cases of weakly-identified SVARs.

I apply these methods to the identification of monetary policy shocks, as in Nakamura & Steinsson (2018). Daily data exhibits several symptoms of weak identification, but intraday data strongly identifies monetary policy shocks. Daily data is frequently used in macro-financial contexts, so this finding has implications for the design of empirical studies. It remains to examine whether weak identification arises in lower frequency (e.g. monthly, quarterly) data.

Following Staiger & Stock (1997), papers using IV report first-stage $F-$statistics to justify instrument relevance. Up to now, this has not been possible for identification via heteroskedasticity. I hope that it can become best practice to do so using the tests that I provide.

9 Appendix

Notation

M_{ij}	denotes the ij^{th} element of matrix M
$M^{(j)}$	denotes the j^{th} column of matrix M
$M_{(i)}$	denotes the i^{th} row of matrix M
$vech\,(M)$	denotes the unique vectorization of matrix M
A_i	denotes the matrix of coefficients corresponding to lag i of the lag polynomial $A\,(L)$

Proof of Proposition 1

Proposition 1. *Under the device (9), if η_t is ergodic and stationary within regimes, then*

$$\hat{H}_{12} - H_{12} \xrightarrow{d} \frac{z_1}{d_\varepsilon + z_2}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\,(0, V_{weak})\,, \tag{20}$$

where V_{weak} is determined by the parameters of the model and distribution of the data.

Proof. The weak instruments literature models the first stage parameter, Π, as local-to-zero, $\Pi = \frac{C}{\sqrt{T}}$. However, the modified local-to-unity device I adopt implies a different structure, $\Pi = \frac{C_T}{\sqrt{T}}$, where the numerator depends on T. I show that the asymptotic distribution of $\hat{H}_{12}$ under weak identification still takes the same form as that found in the weak instruments literature. To derive this distribution, I examine the limit of $\hat{H}_{12} - H_{12}$ itself (as opposed to $\sqrt{T}\left(\hat{H}_{12} - H_{12}\right)$) as $T \to \infty$. Under the modified local-to-unity device, $\sigma^2_{2,2} = \sigma^2_{2,1} + \frac{d_\varepsilon}{\sqrt{T}}$ so

$\sigma^2_{\eta_2,2} = \sigma^2_{\eta_2,1} + \frac{d_\varepsilon}{\sqrt{T}}$ and $\sigma_{\eta_1\eta_2,2} = \sigma_{\eta_1\eta_2,1} + H_{12}\frac{d_\varepsilon}{\sqrt{T}}$. Writing the estimator in (7) yields

$$
\begin{aligned}
\hat{H}_{12} - H_{12} &= \frac{\frac{1}{T_2}\sum_{t\in R_2}\eta_{1t}\eta_{2t} - \frac{1}{T_1}\sum_{t\in R_1}\eta_{1t}\eta_{2t}}{\frac{1}{T_2}\sum_{t\in R_2}\eta^2_{2t} - \frac{1}{T_1}\sum_{t\in R_1}\eta^2_{2t}} - H_{12} \\
&= \frac{\frac{1}{T_2}\sum_{t\in R_2}\sigma_{\eta_1\eta_2,2} + (\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,2}) - \frac{1}{T_1}\sum_{t\in R_1}\sigma_{\eta_1\eta_2,1} + (\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,1})}{\frac{1}{T_2}\sum_{t\in R_2}\sigma^2_{2,2} + \left(\eta^2_{2t} - \sigma^2_{\eta_2,2}\right) - \frac{1}{T_1}\sum_{t\in R_1}\sigma^2_{2,1} + \left(\eta^2_{2t} - \sigma^2_{\eta_2,1}\right)} - H_{12} \\
&= \frac{H_{12}\frac{d_\varepsilon}{\sqrt{T}} + \frac{1}{T_2}\sum_{t\in R_2}\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,1} - H_{12}\frac{d_\varepsilon}{\sqrt{T}} - \frac{1}{T_1}\sum_{t\in R_1}\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,1}}{\frac{d_\varepsilon}{\sqrt{T}} + \frac{1}{T_2}\sum_{t\in R_2}\eta^2_{2t} - \sigma^2_{\eta_2,1} - \frac{d_\varepsilon}{\sqrt{T}} - \frac{1}{T_1}\sum_{t\in R_1}\eta^2_{2t} - \sigma^2_{\eta_2,1}} - H_{12} \\
&= \frac{\frac{1}{\sqrt{T}}\left(H_{12}d_\varepsilon + \sqrt{\frac{T}{T_2}}\frac{1}{\sqrt{T_2}}\sum_{t\in R_2}\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,1} - H_{12}\frac{d_\varepsilon}{\sqrt{T}} - \sqrt{\frac{T}{T_1}}\frac{1}{\sqrt{T_1}}\sum_{t\in R_1}\eta_{1t}\eta_{2t} - \sigma_{\eta_1\eta_2,1}\right)}{\frac{1}{\sqrt{T}}\left(d_\varepsilon + \sqrt{\frac{T}{T_2}}\frac{1}{\sqrt{T_2}}\sum_{t\in R_2}\eta^2_{2t} - \sigma^2_{\eta_2,1} - \frac{d_\varepsilon}{\sqrt{T}} - \sqrt{\frac{T}{T_1}}\frac{1}{\sqrt{T_1}}\sum_{t\in R_1}\eta^2_{2t} - \sigma^2_{\eta_2,1}\right)} - H_{12} \\
&\xrightarrow{d} \frac{H_{12}d_\varepsilon + z_{12,2} - z_{12,1}}{d_\varepsilon + z_{2,2} - z_{2,1}} - H_{12} = \frac{H_{12}d_\varepsilon + z_{12}}{d_\varepsilon + z_2} - H_{12} = \frac{z_1}{d_\varepsilon + z_2}
\end{aligned}
$$

where $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim \mathcal{N}(0, V_{weak})$. The convergence follows from a martingale central limit theorem for each of the summations, which holds since η_t is assumed to be ergodic and stationary within regimes. In the last line, $z_{12} = z_{12,2} - z_{12,1}$, $z_2 = z_{2,2} - z_{2,1}$, and $z_1 = z_{12} - H_{12}z_2$ to simplify the limiting distribution. □

Proof of Proposition 2

Proposition 2. *Under Assumption 1, H is globally identified from $\Sigma_{\eta,C}$ and $\Sigma_{\eta,P}$ up to column order provided the rows of $\left[\begin{array}{cc} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right)\end{array}\right]$ are not proportional.*

Proof. The result follows directly from an argument due to Brunnermeier et al (2017). The columns of H are the right eigenvectors of $\Sigma_{\eta,C}\Sigma^{-1}_{\eta,P}$, corresponding to eigenvalues given by the diagonal of $\Sigma_{\varepsilon,C}\Sigma^{-1}_{\varepsilon,P}$. Eigenvectors corresponding to unique eigenvalues are uniquely determined, up to normalization. If no rows of $\left[\begin{array}{cc} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right)\end{array}\right]$ are proportional, then there are no repeated elements in $\Sigma_{\varepsilon,C}\Sigma^{-1}_{\varepsilon,P}$, so no repeated eigenvalues, and H is uniquely determined from $\Sigma_{\eta,C}$ and $\Sigma_{\eta,P}$. This is a slight modification of Proposition 3 of Sentana and Fiorentini (2001), who require the stronger condition of linear independence of the rows of $\left[\begin{array}{cc} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right)\end{array}\right]$. □

Proof of Proposition 3

Proposition 3. *Adopting the modeling device in (17), H is asymptotically unidentified.*

Proof. I model the variance deficiency as

$$\frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1 + \frac{d}{\sqrt{T}}.$$

Under this device, the i^{th} row of $\left[\begin{array}{cc} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right)\end{array}\right]$ is equal to

$$\left[\begin{array}{ccc} \sigma^2_{\varepsilon_i,C} & \sigma^2_{\varepsilon_j,P}\frac{\sigma^2_{\varepsilon_i,C}}{\sigma^2_{\varepsilon_j,C}} & \left(1+d/T^{1/2}\right)\end{array}\right].$$

In the limit, for finite d, this equals $\left[\begin{array}{cc} \sigma^2_{\varepsilon_i,C} & \sigma^2_{\varepsilon_j,P}\frac{\sigma^2_{\varepsilon_i,C}}{\sigma^2_{\varepsilon_j,C}}\end{array}\right]$. However, this expression is $\frac{\sigma^2_{\varepsilon_i C}}{\sigma^2_{\varepsilon_j C}}$ times the j^{th} row, $\left[\begin{array}{cc} \sigma^2_{\varepsilon_j,C} & \sigma^2_{\varepsilon_j,P}\end{array}\right]$, so the condition of Proposition 2 is violated, and only the column space of $H^{(i)}$ and $H^{(j)}$ is identified. □

Proof of Lemma 1

Lemma 1. *Under Assumptions 1 & 2,*

$$\psi_T\left(\theta_0\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\begin{array}{c}\bar{\phi}\left(\theta_0,\eta_t\right)\\ \bar{q}\left(\theta_0,\eta_t\right)\end{array}\right) \xrightarrow{d} \left(\begin{array}{c}\psi_\phi\\ \psi_{\theta_0}\end{array}\right)$$

where $\psi = \left(\begin{array}{c}\psi_\phi\\ \psi_{\theta_0}\end{array}\right)$ *is a* $2\left(n^2+n\right)$*-dimensional normally distributed random variable with mean zero and positive semi-definite* $2\left(n^2+n\right)\times 2\left(n^2+n\right)$*-dimensional covariance matrix*

$$V\left(\theta\right) = \left(\begin{array}{cc} V_{\phi\phi}\left(\theta\right) & V_{\phi\theta}\left(\theta\right)\\ V_{\theta\phi}\left(\theta\right) & V_{\theta\theta}\left(\theta\right)\end{array}\right) = \lim_{T\to\infty} var\left[\frac{1}{\sqrt{T}}\left(\begin{array}{c}\phi_T\left(\theta\right)\\ q_T\left(\theta\right)\end{array}\right)\right]$$

where $\phi_T\left(\theta\right) = \sum_{t=1}^{T}\phi\left(\theta,\eta_t\right)$ and $q_T\left(\theta\right) = \sum_{t=1}^{T} q\left(\theta,\eta_t\right)$.

Proof. First, note that each block of $\bar{\phi}\left(\theta_0,\eta_t\right)$ forms a martingale difference sequence with respect to $\mathcal{F}_{t-1} = \{\eta_1,\eta_2,\ldots,\eta_{t-1}\}$. This follows from observing that the r^{th} block of $\phi\left(\theta_0,\eta_t\right)$, denoted $\phi^r\left(\theta_0,\eta_t\right)$, takes the form

$$1\left[t\in R_r\right]\left(vech\left(\eta_t\eta_t'\right) - vech\left(H\Sigma_{\varepsilon,j}H'\right)\right)$$

and Assumption 1.2. Then

$$\begin{aligned} E\left[\phi^r\left(\theta_0,\eta_t\right)\mid\mathcal{F}_{t-1}\right] &= E\left[1\left[t\in R_r\right]vech\left(\eta_t\eta_t'\right)\mid\mathcal{F}_{t-1}\right]-1\left[t\in R_r\right]vech\left(H\Sigma_{\varepsilon,r}H'\right) \\ &= \frac{T_r}{T}\left(vech\left(\Sigma_{\eta,r}\right)-vech\left(H\Sigma_{\varepsilon,r}H'\right)\right)=0 \end{aligned}$$

by Assumption 1. Finally, $E\left|\phi\left(\theta,\eta_t\right)\right|<\infty$ by Assumption 2.2, so $\phi^r\left(\theta_0\right)$ is a martingale difference sequence. This means that, stacking the blocks, $\phi\left(\theta,\eta_t\right)$ is a martingale difference sequence. By Billingsley's (1961) Ergodic Stationary Martingale Differences CLT, given Assumption 2.1,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\phi}\left(\theta_0,\eta_t\right)\xrightarrow{d}\mathcal{N}\left(0,E\left[\bar{\phi}\left(\theta_0,\eta_t\right)\bar{\phi}\left(\theta_0,\eta_t\right)'\right]\right).$$

Note that

$$\begin{aligned} E\left[\bar{\phi}\left(\theta,\eta_t\right)\bar{\phi}\left(\theta,\eta_t\right)'\right] &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}var\left(\phi\left(\theta,\eta_t\right)\right) \\ &= \lim_{T\to\infty}var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\phi\left(\theta,\eta_t\right)\right)\right] \\ &= V_{\phi\phi}\left(\theta\right) \end{aligned}$$

as required, where the second-last equality follows from the fact that $cov\left(\phi\left(\theta,\eta_t\right),\phi\left(\theta,\eta_s\right)\right)=0, t\neq s$ by Assumption 1.2 (using a similar argument to establishing the MDS property above).

By definition, $\bar{q}\left(\cdot\right)=0$ deterministically; note that $\frac{\partial\phi(\theta,\eta_t)}{\partial\theta'}=-\frac{\partial\left(vech\left(\Sigma_{\eta_C}\right)',vech\left(\Sigma_{\eta_P}\right)'\right)'}{\partial\theta'}=E\left[\frac{\partial\phi(\theta,\eta_t)}{\partial\theta'}\right]$ since it contains only parameters and no data (the moment equations are separable in data and parameters). This is true for any $\theta\in\Theta$; θ need not equal θ_0. Thus ψ_θ is a degenerate random variable. It remains to show that $V\left(\theta\right)$ is positive semi-definite. Since all but the top left block, $V_{\phi\phi}\left(\theta\right)$, will be zeros, it suffices to show that $V_{\phi\phi}\left(\theta\right)$ is positive semi-definite. This follows as $V_{\phi\phi}\left(\theta_0\right)$ has the form $E\left[\bar{\phi}\left(\theta,\eta_t\right)\bar{\phi}\left(\theta,\eta_t\right)'\right]$. □

Proof of Lemma 2

Lemma 2. *Under Assumptions 1 & 2, the covariance matrix estimator $\hat{V}\left(\theta_0\right)$ satisfies*

$$\hat{V}\left(\theta_0\right)\xrightarrow{p}V\left(\theta_0\right)$$

and

$$\frac{\partial vec\left(\hat{V}_{\phi\phi}\left(\theta_0\right)\right)}{\partial\theta'} \xrightarrow{p} \frac{\partial vec\left(V_{\phi\phi}\left(\theta_0\right)\right)}{\partial\theta'}.$$

Proof. By the Ergodic Theorem (e.g. Karlin & Taylor (1975), Theorem 9.5.5) and Assumption 2, the natural covariance estimator is consistent, $\frac{1}{T}\sum\phi\left(\theta_0,\eta_t\right)\phi\left(\theta_0,\eta_t\right)' \xrightarrow{p} E\left[\phi\left(\theta_0,\eta_t\right)\phi\left(\theta_0,\eta_t\right)'\right]$. Then

$$V\left(\theta_0\right) = \lim_{T\to\infty} var\left[\frac{1}{\sqrt{T}}\ \phi_T\left(\theta_0\right)\ \right] = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} var\left(\phi\left(\theta_0,\eta_t\right)\right),$$

by the same assumptions, which simplifies to $E\left[\phi\left(\theta_0,\eta_t\right)\phi\left(\theta_0,\eta_t\right)'\right]$ since $cov\left(\phi\left(\theta,\eta_t\right),\phi\left(\theta,\eta_s\right)\right) = 0, t\neq s$ by Assumption 1.2. Since $q\left(\theta,\eta_t\right)$ is deterministic, this establishes the first part of the Lemma.

For the second part, note that $\frac{\partial\hat{V}\phi\phi(\theta_0)}{\partial\theta'} = \frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\phi(\theta_0,\eta_t)\phi(\theta_0,\eta_t)'\right]}{\partial\theta'} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial\phi(\theta_0,\eta_t)}{\partial\theta'}\phi\left(\theta_0,\eta_t\right)'\right]$. $\frac{\partial\phi(\theta_0,\eta_t)}{\partial\theta'}$ is a matrix of zeros, ones, and continuous functions of elements of θ; it is entirely deterministic. Similarly, $\frac{\partial V\phi\phi(\theta_0)}{\partial\theta'} = E\left[\frac{\partial\phi(\theta_0,\eta_t)}{\partial\theta'}\phi\left(\theta_0,\eta_t\right)'\right] = \frac{\partial\phi(\theta_0,\eta_t)}{\partial\theta'}E\left[\phi\left(\theta_0,\eta_t\right)'\right]$, and since $E\left[\phi\left(\theta_0,\eta_t\right)'\right]$ is consistently estimated, so too is $\frac{\partial V_{\phi\phi}(\theta_0)}{\partial\theta'}$ by Slutsky's Theorem.

□

Proof of Theorem 1

Theorem 1. *If Lemmata 1 and 2 hold,*

$$K_T\left(\theta_0\right) \xrightarrow{d} \chi^2_{n^2+n},$$

where $K_T\left(\theta_0\right) = \frac{1}{T}\phi_T\left(\theta_0\right)'\hat{V}_{\phi\phi}\left(\theta_0\right)^{-1/2}P_{\hat{V}_{\phi\phi}(\theta_0)^{-1/2}\hat{D}_T(\theta_0)}\hat{V}_{\phi\phi}\left(\theta_0\right)^{-1/2}\phi_T\left(\theta_0\right)$.

Proof. The result follows directly from Kleibergen (2005). Lemmata 1 and 2 establish Assumptions 1 and 2 from that paper, which are used to prove Theorem 1 therein. Note that Lemma 1 and part of Lemma 2 also establish the required conditions of Stock & Wright (2001) Theorem 2 (their Assumption A and the consistency of the covariance matrix for the weighting matrix) so

$$S_T\left(\theta_0\right) \xrightarrow{d} \chi^2_{n^2+n}$$

as an immediate corollary.

□

Proof of Theorem 2

Theorem 2. *If Lemmata 1 and 2 hold, then under Assumption 5,*

$$K_T(\beta_0) \xrightarrow{d} \chi^2_{p_{int}},$$

where p_{int} is the dimension of β.

Proof. As above, Theorem 2 follows directly from Theorem 2 of Kleibergen 2005. Again, this also implies $S_T(\beta_0) \xrightarrow{d} \chi^2_{p_{int}}$ as an immediate corollary. □

Proof of Proposition 4

Proposition 3. *Under Assumption 1, H_I is identified from the covariance matrices across regimes.*

Proof. This follows directly from the proof of Proposition 2; columns of H corresponding to unique eigenvalues are uniquely identified, while those corresponding to repeated eigenvalues are not. This aligns with Sentana & Fiorentini (2001) Proposition 4. □

Proof of Theorem 3

Theorem 3. *Under Assumption 1, if H_W contains two columns, H is conditionally identified from the covariance matrices provided*

1. *A single element H_{lk} is fixed and $H_{lk} \neq H_{lm}/H_{km}$ for $H^{(k)}, H^{(m)} \in H_W$, or*
2. *The full column $H^{(k)} \in H_W$ is fixed.*

Proof. The proof follows from extending the argument of Proposition 4 in Sentana & Fiorentini (2001). They show that for a similarly partitioned H, the columns of H_I are identified. However, the columns of H_W are identified only up to an orthogonal rotation Q, $QQ' = Q'Q = I$. H_W represents the portion of H pertaining to proportional variance processes, and as such cannot contain just a single column. If H_W contains two columns, then Q is 2×2. Consider first a single fixed element of $H^{(k)}$, the subject of the null hypothesis for the subset test. Without loss of generality, let it be $H_{2k} = x$. This yields the system of

equations

$$\begin{bmatrix} 1 & H_{1m} \\ x & 1 \\ \vdots & \vdots \\ H_{nk} & H_{nm} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} 1 & \tilde{H}_{1m} \\ x & 1 \\ \vdots & \vdots \\ \tilde{H}_{nk} & \tilde{H}_{nm} \end{bmatrix}. \tag{21}$$

Placing $H^{(k)}$ and $H^{(m)}$ as the first and second columns, with the associated unit normalization, is without loss of generality, as identification is only up to scale of each column. Since Q is orthogonal, fixing column order, $Q_{11}^2 + Q_{21}^2 = 1$. Given this and the equation

$$xQ_{11} + Q_{21} = x,$$

Q_{11} and Q_{21} can be solved for where the sign is pinned down by the unit normalization. This yields two solutions for Q_{11} and Q_{21}: $\{Q_{11} = 1, Q_{21} = 0\}$ and $\left\{Q_{11} = \frac{x^2-1}{x^2+1}, Q_{21} = \frac{2x}{x^2+1}\right\}$. However, using an additional equation implied by (21), $Q_{11} + H_{1m}Q_{21} = 1$, rules out the second solution unless $H_{1m} = 1/x$. Generalizing away from the case where $H^{(k)}$ and $H^{(m)}$ are the first two columns yields the first condition of the theorem, $H_{km} \neq H_{lm}/H_{lk}$. With Q_{11} and Q_{21} thus pinned down, the other column of Q is unique, and thus the entirety of H is identified.

This argument extends to the case where the entirety of $H^{(k)}$ is fixed. Now, however, the solution is unique unless $H_{lm}/H_{mm} = H_{lk}/H_{mk}$ for all l, in which case column m is a scalar multiple of column k, making H non-invertible, which is false by Assumption 1.1. Thus, the solution when a full column of H is specified is unique.

The restriction that H_W contain at most two columns is necessary to yield conditional identification without any assumptions on the variances. If three columns pertaining to proportional variance processes were included, and column k's value conditioned upon, the remaining columns could not be distinguished. □

Proof of Proposition 5

Proposition 5. *Under Assumption 1, if B^h is invertible, Λ_I^h is identified up to scale from the covariance matrices across regimes.*

Proof. Proposition 5 follows similarly to Proposition 4 by replacing H with $\Lambda^h = B^h H$, where both are full-rank $n \times n$ matrices. □

Proof of Theorem 4

Theorem 4. *Under Assumption 1, if Λ_W^h contains two columns, Λ^h is conditionally identified up to scale from the covariance matrices provided*

1. *B^h is invertible,*
2. *A single element Λ_{lk}^h is fixed and $\Lambda_{lk}^h \neq \Lambda_{lm}^h/\Lambda_{km}^h$ for $\Lambda^{h(k)}, \Lambda^{h(m)} \in \Lambda_W^h$.*

Proof. This follows from the proof of Theorem 3. The condition that B^h is full rank guarantees that, like H, Λ^h is full-rank, which is needed to yield Proposition 5, embedded in the theorem. Then, H can simply be replaced in the proof of Theorem 3 with Λ^h. Note that this requires a unit normalization of Λ^h, which is unfamiliar. However, Λ^h is always normalized, at least implicitly, by any normalization applied to H. Thus, the unit normalization is without loss of generality. Any identification results must similarly hold under a different normalization, where Λ^h is rescaled back to the conventional $B^h H$, where H has a unit diagonal. □

Verification of Andrews (2017) Assumptions

Andrews' (2017) framework for weak identification relies on the necessary conditions for his Theorem 1 being satisfied. This requires conditions on both the robust confidence sets constructed and the non-robust confidence sets, as laid out in his Assumptions 2-6.

His Assumption 2 coincides with Lemma 1. His Assumption 3 largely coincides with Lemma 2, requiring the consistent estimation of $V(\theta)$. It also requires consistent estimation of the weighting matrix, which, given the focus on efficient CUE GMM in the present paper, also follows from Lemma 2. His andAssumption 4 requires, for any draw of data, the existence of normalizing matrices for both the Jacobian and orthogonalized Jacobian evaluated at θ_0 such that when normalized they converge to full-rank matrices. Since the Jacobian is deterministic, it is identical to the orthogonalized Jacobian. Under strong identification, the identity matrix serves as such a normalizing matrix; it is trivial to construct a matrix with judiciously placed $O\left(T^{1/2}\right)$ elements under the weak identification device considered here. Unsurprisingly, if the system is unidentified, no such matrices exist.

Assumption 5 states four conditions underlying the performance of Wald tests under strong identification. First, given the separability of $\phi(\theta, \eta_t)$ in data and parameters, $\phi_T(\theta)$ converges uniformly over Θ since the sample second moments converge uniformly. Uniform boundedness follows similarly. The third condition requires that the estimated weighting matrix converges uniformly over Θ, which follows since $W(\theta)$ is continuous in θ and measurable, Θ is compact, and $E\left[vech(\eta_t\eta_t)\, vech(\eta_t\eta_t)'\right] < \infty$ by Assumption 2.2. Since efficient

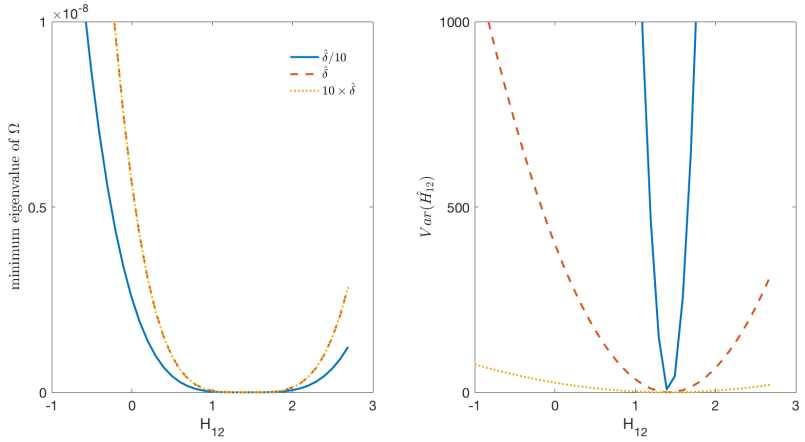
CUE GMM is considered here, positive definiteness of $W(\theta)$ follows from the positive definiteness of $V_{\phi\phi}(\theta)$, which is established in the final point of Assumption 6 below. Finally, its maximal eigenvalue is bounded by Assumption 2.2 and the minimal eigenvalue is bounded away from zero since $W(\theta)$ is positive definite. The second and fourth conditions are identification requirements, ensuring the population objective is small if and only if it is evaluated in neighbourhood of the true parameter value; these are satisfied if the system meets the requirements for strong identification.

Assumption 6 imposes five conditions guaranteeing asymptotic normality of $\hat{\theta}$ under strong identification. First, it requires that θ_0 be in the interior of Θ. Second, it requires $\phi_T(\theta)$ and $\hat{\Omega}(\theta)$ be continuously differentiable; the first part holds as the Jacobian is a deterministic matrix of ones, zeros, and products of elements of θ, and the second part holds since $\frac{\partial \hat{\Omega}(\theta)}{\partial \theta'} = 2\frac{\partial \phi(\theta,\eta_t)}{\partial \theta'}\frac{1}{T}\phi_T(\theta)'$ and $\frac{1}{T}\phi_T(\theta)$ is continuous in θ. The third set of conditions, on the Jacobian, are satisfied trivially since it is deterministic and of full rank under strong identification. $\sup_{\theta \in B(\theta_0)} \left\| \frac{\partial \hat{\Omega}(\theta)}{\partial \theta'} \right\| = O_p(1)$ (stochastic boundedness) holds since $\frac{1}{T}\phi_T(\theta)$ consists of linear combinations of estimated second moments of the data (which have finite second moments by Assumption 2.2) and multiplicative functions of θ (which are finite for θ in a ball around θ_0). For the fifth condition, since the moment function is continuous in θ and the asymptotic variance is a continuous function of the moment function, the asymptotic variance is continuous in θ. Thus, the covariance estimator is uniformly consistent on a ball around θ_0. Since, for all θ, $V_{\phi\phi}(\theta) = E\left[\phi(\theta,\eta_t)\phi(\theta,\eta_t)'\right]$, an outer product, and ϕ_t are linearly independent under strong identification, $V_{\phi\phi}(\theta)$ is positive definite on a ball around θ_0.

Non-monotonicity of power curves

The non-monotonic behaviour between $H_{12} = 1$ and 2 is potentially surprising. Figure 6 suggests a possible explanation. Under this particular calibration, the minimum eigenvalue of the moment covariance matrix approaches zero in this range. This matrix is assumed to be positive definite and must be inverted to calculate the robust test statistics. The variance of $\hat{H}_{12}$ also approaches zero around $H_{12} = 1.4$, where there is a kink in the Wald power curves. As these values get closer to zero, it is natural that the rejection rate of tests based on the inverse of the objects increase. The associated kinks in the power curves should be viewed as calibration-specific artifacts, and not indicative of general power properties of the tests in this context.

Figure 6: Calibration details



Numerically calculated values for the minimum eigenvalue of the moment covariance matrix and the variance (standard CUE GMM) of the estimator of $\hat{H}_{12}$ computed for various values of H_{12} and δ, $T = 100,000$, and 100 samples. The asymptote of the minimum eigenvalue of the moment covariance matrix, which is assumed to be positive definite and must be inverted, may explain the unexpected behaviour of the robust tests between values of H_{12} of 1 and 2. The asymptote of the variance to zero around a local alternative of 1.4 contributes to the kink in the Wald rejection rate visible in the power curves, Figure 4.

References

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84, 2155–2182.

——— (2017): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics (forthcoming)*.

Billingsley, P. (1961): "Statistical Methods in Markov Chains," *The Annals of Mathematical Statistics*, 32, 12–40.

Boyarchenko, N., V. Haddad, and M. Plosser (2017): "The Federal Reserve and Market Confidence," *Federal Reserve Bank of New York Staff Reports*.

Chaudhuri, S. (2008): "Projection-type Score Tests for Subsets of Parameters," Ph.D. thesis, University of Washington.

Chaudhuri, S., T. Richardson, J. Robins, and E. Zivot (2010): "A New Projection-type Split-Sample Score Test in Linear Instrumental Variables Regression," *Econometric Theory*, 26, 1820–1837.

Chaudhuri, S. and E. Zivot (2011): "A New Method of Projection-based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 164, 239–251.

Craine, R. and V. L. Martin (2008): "International Monetary Policy Surprise Spillovers," *Journal of International Economics*, 75, 180–196.

Dufour, J. M. (1997): "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

Dufour, J. M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

Ehrmann, M. and M. Fratzscher (2017): "Euro Area Government Bonds - Fragmentation and Contagion During the Sovereign Debt Crisis," *Journal of International Money and Finance*, 70, 26–44.

Eichengreen, B. and U. Panizza (2016): "A Surplus of Ambition: Can Europe Rely on Large Primary Surpluses to Solve its Debt Problem?" *Economic Policy*, 31, 5–49.

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups, and US Welfare," *Journal of Political Economy*, 125, 1040–1074.

Fernandez-Perez, A., B. Frijns, and A. Tourani-Rad (2016): "Contemporaneous Interactions among Fuel, Biofuel and Agricultural Commodities," *Energy Economics*, 58, 1–10.

Fieller, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society*, 16, 175–185.

Gong, X., S. Yang, and M. Zhang (2017): "Not Only Health: Environmental Pollution Disasters and Political Trust," *Sustainability*, 9, 1–28.

Hayashi, F. (2000): *Econometrics*, Princeton, NJ: Princeton University Press.

Hébert, B. and J. Schreger (2017): "The Costs of Sovereign Default: Evidence from Argentina," *American Economic Review (forthcoming)*, 107, 3119–3145.

Hogan, V. and R. Rigobon (2009): "Using Unobserved Supply Shocks to Estimate the Returns to Education," manuscript.

Islam, A., F. Islam, and C. Nguyen (2017): "Skilled Immigration, Innovation, and the Wages of Native-Born Americans," *Industrial Relations*, 56, 459–488.

Jahn, E. and E. Weber (2016): "Identifying the Substitution Effect of Temporary Agency Employment," *Macroeconomic Dynamics*, 20, 1264–1281.

Karlin, S. and H. M. Taylor (1975): *A First Course in Stochastic Processes*, San Diego: Elsevier.

Khalid, U. (2016): "The Effect of Trade and Political Institutions on Economic Institutions," *Journal of International Trade and Economic Development*, 26, 89–110.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that they Are Identified," *Econometrica*, 73, 1103–1123.

Klein, R. and F. Vella (2009): "Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments," *Journal of Human Resources*, 44, 1047–1065.

Lewis, D. J. (2018): "Identification via Time-Varying Volatility," manuscript.

Lin, F., E. O. Weldemicael, and X. Wang (2016): "Export Sophistication Increases Income in Sub-Saharan Africa: Evidence from 1981-2000," *Empirical Economics*, 52, 1627–1649.

Magnusson, L. M. and S. Mavroeidis (2014): "Identification using Stability Restrictions," *Econometrica*, 82, 1799–1851.

Millimet, D. L. and J. Roy (2016): "Empirical Tests of the Pollution Haven Hypothesis when Environmental Regulation is Endogenous," *Journal of Applied Econometrics*, 31, 652–677.

Mönkediek, B. and H. A. Bras (2016): "The Interplay of Family Systems, Social Networks and Fertility in Europe Cohorts Born Between 1920 and 1960," *Economic History of Developing Regions*, 31, 136–166.

Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2018): "Inference in Structural VARs with External Instruments," manuscript.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2005): "Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak," manuscript.

Nakamura, E. and J. Steinsson (2018): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133, 1283–1330.

Rigobon, R. (2003): "Identification through Heteroskedasticity," *The Review of Economics and Statistics*, 85, 777–792.

Rigobon, R. and D. Rodrik (2005): "Rule of law, Democracy, Openness, and Income," *Economics of Transition*, 13, 533–564.

Rigobon, R. and B. Sack (2003): "Measuring the Reaction of Monetary Policy to the Stock Market," *Quarterly Journal of Economics*, 118, 639–669.

——— (2004): "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics*, 51, 1553–1575.

Sentana, E. and G. Fiorentini (2001): "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143–164.

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

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Staiger, D., J. H. Stock, and M. W. Watson (1997): "How Precise Are Estimates of the Natural Rate of Unemployment?" in *Reducing Inflation: Motivation and Strategy*, ed. by C. Romer and D. Romer, University of Chicago Press, January, 195–246.

Stock, J. H. (2008): "What's New in Econometrics- Time series: Weak Instruments, Weak

Identification, and Many Instruments, Part I," in *NBER Summer Institute: Methods Lectures*.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalize Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–529.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, 80–108.

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

Wright, J. H. (2001): "Detecting Lack of Identification in GMM," manuscript.

——— (2012): "What Does Monetary Policy Do to Long-term Interest Rates at the Zero Lower Bound?" *Economic Journal*, 122, 447–466.

Zaefarian, G., V. Kadile, S. C. Henneberg, and A. Leischnig (2017): "Endogeneity Bias in Marketing Research: Problem, Causes and Remedies," *Industrial Marketing Management*, 65, 39–46.