ROBUST INFERENCE IN MODELS IDENTIFIED VIA HETEROSKEDASTICITY

Daniel J. Lewis*

Abstract—Identification via heteroskedasticity exploits variance changes between regimes to identify parameters in simultaneous equations. Weak identification occurs when shock variances change very little or multiple variances change close to proportionally, making standard inference unreliable. I propose an F-test for weak identification in a common simple version of the model. More generally, I establish conditions for validity of nonconservative robust inference on subsets of the parameters, which can be used to test for weak identification. I study monetary policy shocks identified using heteroskedasticity in high-frequency data. I detect weak identification, invalidating standard inference, in daily data, while intraday data provide strong identification.

I. Introduction

UNOBSERVED structural shocks, like those in the structural vector autoregressions (SVARs) of Sims (1980), are ubiquitous in economic models, 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, is increasingly popular 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. The intuition dates from at least Fisher (1965). This identification scheme is most popular in macrofinancial contexts but has also been adopted in many other fields.[1] However, little work has addressed the possibility of weak identification in these studies.

The identifying variation is the difference in innovation covariances across regimes. If the structural variances are in fact the same across regimes, then so too are these 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 also no new identifying information, as the covariance matrices are just scalar multiples. Both scenarios 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 instrumental variables (IV) context, where an IV 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 nonstandard. Standard inference methods will be unreliable, as will any empirical conclusions based on them. If not properly detected and accounted for, this phenomenon can undermine the credibility of empirical work.

I provide a framework for inference in models identified via heteroskedasticity when weak identification causes standard methods to provide a poor approximation to the asymptotic distribution. First, I propose a straightforward method to test for weak identification in an empirically common bivariate case, where it is assumed that only one variance changes across regimes. The model can be written as an IV regression using dummy variables for regimes. I propose a rule of thumb of $F > 23$ for the heteroskedasticity-robust first-stage F-statistic,

$$F = \frac{\hat{\Pi}^2 \left(\sum_{t=1}^{T} Z_t^2 \right)^2}{\sum_{t=1}^{T} Z_t^2 \hat{v}_t^2},$$

where Z_t is an instrument constructed based on the observed innovations and regime dummies, $\hat{\Pi}$ is the OLS estimator of the coefficient in the first-stage regression of an innovation on Z_t, and $\hat{v}_t$ are the corresponding OLS residuals (see section IIB for details). Papers using IV report first-stage F-statistics to support their results, based on Staiger and Stock (1997); now this pretest can be reported for models identified via heteroskedasticity.

In a fully general model, I establish conditions under which the asymptotic distributions of common identification-robust test statistics (K-statistic of Kleibergen, 2005; S-statistic of Stock & Wright, 2000) may be more tightly characterized when the object of interest is a subset of the parameter vector. I derive primitive conditions under which all weakly identified nuisance parameters are uniquely determined once values are

Received for publication December 9, 2019. Revision accepted for publication July 9, 2020. Editor: Olivier Coibion.
*Lewis: Federal Reserve Bank of New York.
The views expressed in this paper are my own and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. I thank Jim Stock, Isaiah Andrews, Gabriel Chodorow-Reich, Richard Crump, Domenico Giannone, Adam McCloskey, Jose Luis Montiel Olea, Emi Nakamura, Mikkel Plagborg-Möller, Neil Shephard, and Jòn Steinsson for their helpful feedback, as well as Olivier Coibion and two anonymous referees.
A supplemental appendix is available online at https://doi.org/10.1162/rest_a_00963.

[1]The macrofinance literature includes Rigobon (2003), Rigobon and Sack (2003, 2004), Craine and Martin (2008), Ehrmann and Fratzscher (2017), Eichengreen and Panizza (2016), Hébert and Schreger (2017), Nakamura and Steinsson (2018), Wright (2012), and Gürkaynak, Kisacikoglu, and Wright (2019). Examples in other fields include public finance (Jahn & Weber, 2016), growth (Islam, Islam, & Nguyen, 2017), trade (Feenstra & Weinstein, 2017; Lin, Weldemicael, & Wang, 2016), political economy (Khalid, 2016; Rigobon & Rodrik, 2005), environmental economics (Gong, Yang, & Zhang, 2017; Millimet & Roy, 2016), agriculture and energy (Fernandez-Perez, Frijns, & Tourani-Rad, 2016), education (Hogan & Rigobon, 2003; Klein & Vella, 2009), marketing Zaefarian et al. (2017), and even fertility studies (Mönkediek & Bras, 2016).

The Review of Economics and Statistics, May 2022, 104(3): 510–524

https://doi.org/10.1162/rest_a_00963

specified for the parameter of interest under the null hypothesis, meaning they may be concentrated out.[2] I provide an economic interpretation of the types of variance changes that are compatible with this procedure. Significantly, these conditions will often be satisfied in small models identified via heteroskedasticity. Inference can thus proceed using a familiar test statistic but potentially much smaller critical values. These more tightly characterized confidence sets can be used with the two-step procedure of Andrews (2018) to test for weak identification. This procedure determines how much lower the level of a robust confidence set must be before it is contained by a strong identification confidence set to quantify the strength of identification. Such tests would be conservative (likely prohibitively so) using robust sets for a subset of the parameters computed using the only previously available option, projection inference.

I demonstrate, in both data and empirically calibrated simulations, that weak identification does in fact cause standard inference approaches to perform poorly. I consider the application of Nakamura and Steinsson (2018, henceforth NS), who 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 provide strong identification. In daily data, robust confidence intervals for the effect of forward guidance are drastically wider than their standard counterparts; in intraday data, conclusions are the same whether standard or robust inference methods are used. In simulations based on the daily data, estimates of the effect of monetary policy shocks are not well approximated by a normal distribution. Additional simulations show that standard tests suffer accordingly from serious size distortions and projection methods are severely undersized, while the procedures I propose remain well sized.

With the tools I outline, research using heteroskedasticity for identification can address concerns of weak identification head-on. It is possible to verify the strength of identification using these methods, much like it is now common practice to do for IV following the work of Staiger and Stock (1997).

The paper is organized as follows. Section II presents the model, shows how weak identification arises, and demonstrates its effects on parameter estimates. Section III proposes a pretest for weak identification in a simple bivariate model commonly adopted in practice. Section IV outlines standard robust inference results, establishes conditions on proportional variance changes under which subset inference can proceed using reduced critical values, describes a test to detect weak identification, and reports simulation results. Section V applies the methods to the data of NS. Section VI concludes. Proofs are in the appendix.

I use the following standard matrix notation: M_{ij} denotes the ijth element of matrix M; $M_{\cdot j}$ denotes the jth column; $M_{i\cdot}$ denotes the ith row; $vech\,(M)$ denotes the unique vectorization; $P_M = M\left(M'M\right)^{-1}M'$ denotes the projection matrix.

[2]A parameter is "concentrated out" from a test statistic by making it a function of some other parameter, eliminating its influence on the limiting distribution.

II. Strong and Weak Identification via Heteroskedasticity

In this section, I outline the model and identification argument. I provide intuition for when identification might fail and illustrate the consequences analytically in an empirically popular simple case. I then characterize weak identification in the fully general model.

A. *Identification and When It Might Fail*

The observed data consist of an $n \times 1$ vector of serially uncorrelated mean-zero innovations η_t. These could be observed (asset price changes) or as-if observed (residuals from a consistently estimated VAR). While I focus on the time series setting, the results of this paper apply equally in cross-sectional settings. Innovations are related to an $n \times 1$ vector of structural shocks, ε_t, by a time-invariant invertible matrix H:[3]

$$\eta_t = \begin{pmatrix} \eta_{1t} \\ \vdots \\ \eta_{nt} \end{pmatrix} = \begin{bmatrix} 1 & H_{12} & \cdots & H_{1n} \\ H_{21} & 1 & \cdots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} & H_{n2} & \cdots & 1 \end{bmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix} = H\varepsilon_t. \quad (1)$$

The diagonal of H is unit-normalized without loss of generality. The object of interest is generally elements of H, which represent the contemporaneous responses of the innovations to structural shocks. In contrast to the standard SVAR identification problem, assume there are two-regimes for η_t. While I focus on the two-regime case, most of the following results can be directly extended to allow for additional regimes. For consistency with my empirical application to NS, I denote regimes C and P, which contrasts "*C*ontrol" observations and "event" observations, arguing that on the event days, when, for example, a *P*olicy announcement is made, the relevant structural shocks are likely to be more volatile than on a typical day. Assumption 1 details basic assumptions.

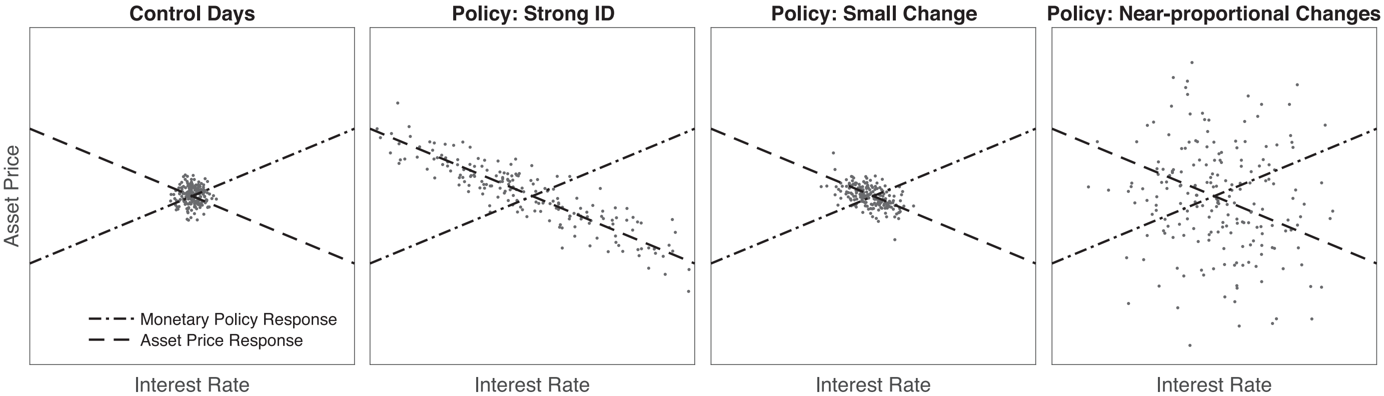
Assumption 1. For all $t = 1, 2, \ldots, T$ and regimes $r \in \{C, P\}$,

1. H is fixed over time, invertible, and has a unit-diagonal,
2. $E\left[\varepsilon_t \mid t \in r, \mathcal{F}_{t-1}\right] = 0,\ E\left[\varepsilon_t\varepsilon_t' \mid t \in r, \mathcal{F}_{t-1}\right] = \Sigma_{\varepsilon,r}$, $\mathcal{F}_{t-1} = \{\varepsilon_1, \ldots, \varepsilon_{t-1}\}$,
3. $\Sigma_{\varepsilon,r}$ is diagonal.

The first point imposes necessary assumptions on H. The second and third jointly state martingale difference sequence and orthogonality assumptions as well as covariance

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

FIGURE 1.—DISTINGUISHING SIMULTANEOUS RESPONSES USING HETEROSKEDASTICITY



Observations are simulated to correspond to the illustration in figures 1 and 2 of Rigobon and Sack (2004) and represent hypothetical innovations to asset prices and interest rates in different regimes.

stationarity of η_t within regimes. Consequently, the covariance of η_t for regime r is

$$\Sigma_{\eta,r} = E\left[\eta_t \eta_t' | t \in r\right] = H\Sigma_{\varepsilon,r}H', \, r \in \{C, P\}, \tag{2}$$

which is consistently estimated by $\frac{1}{T_r}\sum_{t\in r}\eta_t\eta_t'$, where T_r is the number of observations from regime r. I treat the regimes C, P as known; in practice, they are frequently chosen using external information about volatility, like monetary policy announcement dates. This leaves H, $\Sigma_{\varepsilon,C}$, and $\Sigma_{\varepsilon,P}$ to be identified.

A single covariance of η_t yields $\left(n^2+n\right)/2$ equations with n^2 unknowns between H and the shock variances. However, adding a second regime doubles the number of identifying equations to n^2+n, while only adding n new shock variances. Thus, with two regimes, the equations in (2) are potentially just identified. Rigobon (2003) establishes conditions under which these equations do indeed have a unique solution, laid out in proposition 1. Let $\sigma^2_{\varepsilon,r}$ denote the diagonal of $\Sigma_{\varepsilon,r}$.

Proposition 1. *Under assumption 1, H is globally identified from equation (2) up to column order provided the rows of* $\left[\sigma^2_{\varepsilon,C} \quad \sigma^2_{\varepsilon,P}\right]$ *are not proportional.*

Under additional assumptions, distinguishing the columns of H, point identification holds. To that end, I adopt assumption 2, an empirically common choice to this effect:

Assumption 2. For shock of interest i, $\sigma^2_{\varepsilon i,P}/\sigma^2_{\varepsilon i,C} > \sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j}\ \forall j \neq i$; additional columns of H are ordered based on some other statistical rule.

"The shock of interest" refers to the shock whose effects the econometrician seeks to recover. The second part of the assumption is required to point-identify H when $n > 2$ and simply requires some ordering. Many choices are possible (such as continuing to order according to $\sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j,C}$).

Figure 1 presents the intuition of the identification approach. The first two panels follow the example from Rigobon and Sack (2004), who identify the response of asset prices to monetary policy via variance changes on policy announcement days. The first panel plots hypothetical data for η_t, asset price changes against interest rate changes, on "control" days—those with no monetary policy announcement. The lines represent the monetary policy and asset price response curves—the coefficients that the econometrician wishes to identify. Due to the simultaneity of the problem, with two structural shocks affecting η_t contemporaneously, neither response can be identified. The second panel plots what might happen on days with monetary policy announcements if the variance of the policy shock increases dramatically. Now, due to the increase in volatility in the monetary policy shock, the data begin to trace out the asset price response. Since there is still nonnegligible 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.

What happens when the condition in proposition 1 is close to failing? First, the variances might not change much at all across regimes. For example, if most 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 Great Recession and associated financial crisis, many volatilities increased together. On announcement days, there may be increased volatility in more than one shock if there are multiple dimensions of monetary policy shocks. The closer the comovement of variance is, 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 data do not trace the curve. Again, the policy regime offers little additional identifying information.

Turning to estimation, the identification approach is easily implemented via GMM. Defining the vector $\theta \in \Theta$ as the elements of H, $\Sigma_{\varepsilon,C}$, and $\Sigma_{\varepsilon,P}$, equations (2) can be written as a combined set of moments (as in Rigobon, 2003):

$$\phi(\theta, \eta_t) = \begin{bmatrix} 1[t \in C]\, vech\left(\eta_t \eta_t' - H\Sigma_{\varepsilon,C}H'\right) \\ 1[t \in P]\, vech\left(\eta_t \eta_t' - H\Sigma_{\varepsilon,P}H'\right) \end{bmatrix}. \quad (3)$$

Under assumption 1, $E[\phi(\theta_0, \eta_t)] = 0$ at θ_0, the true parameter value. The GMM objective function is defined as

$$S_T\left(\theta; \tilde{\theta}\right) = \left[T^{-1/2}\sum_{t=1}^{T}\phi(\theta, \eta_t)\right]' W_T\left(\tilde{\theta}\right)\left[T^{-1/2}\sum_{t=1}^{T}\phi(\theta, \eta_t)\right], \quad (4)$$

where $\tilde{\theta}$ is the parameter used to compute the weighting matrix, $W_T(\cdot)$. For the purposes of this paper, I focus on a continuous updating estimator (CUE) with the efficient weighting matrix. This means $\tilde{\theta} = \theta$ and $W_T(\theta) = \Omega_T(\theta)^{-1}$, $\Omega_T(\theta) = \frac{1}{T}\sum \phi(\theta, \eta_t)\phi(\theta, \eta_t)'$, so for compactness, I write $S_T(\theta) \equiv S_T(\theta; \theta)$. Since in practice there may also be heteroskedasticity within regimes, the estimator can be thought of as computing the average variance in each regime, which is then used for identification.[4]

To characterize the asymptotic distribution of GMM estimates, I adopt the regularity conditions of assumption 3:

Assumption 3. Assume

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

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. Under these assumptions, if additionally θ_0 is the unique solution to equation (3), standard arguments show that the GMM estimates of θ will be consistent and have the standard asymptotically normal GMM limiting distribution.

However, in contexts characterized by weak identification, it is this final assumption—the uniqueness of the solution to equation (3)—that is in doubt. I now consider formally how that condition may fail and the consequences when it does.

[4]While the literature has exclusively assumed that ε_t are serially uncorrelated (unconditionally and within regimes), consistent with the notion of structural shocks, in principle the expectations of $\eta_t\eta_t'$ could be replaced with (consistent) heteroskedasticity and autocorrelation-robust (HAR) variance estimators, should a researcher wish to consider a model with persistent shocks.

B. The Asymptotic Distribution in a Simple Case

Many empirical papers make the additional assumption that only the variance of the shock of interest changes across regimes (e.g., NS; Hébert & Schreger, 2017; Rigobon & Sack, 2004; Wright, 2012). Under this assumption, the parameter of interest can be estimated in closed form via analogy to IV. This means that the effects of weak identification can be clearly illustrated. Throughout the paper, I refer to this restricted model as the *simple case* and restrict attention therein to models with $n = 2$.[5] I assume that the first shock is the one with constant variance, $\sigma^2_{\varepsilon 1,r} \equiv \sigma^2_{\varepsilon 1}$; H_{12} is the parameter of interest, measuring the impact of ε_{2t} on η_{1t}. This aligns with NS (and many other empirical papers) where the shock of interest is some policy shock whose variance changes. In NS, H_{12} represents the impact of monetary policy shocks on Treasury forward rates.

Following Rigobon and Sack (2004), H_{12} can be recovered in closed form:

$$\begin{aligned} &\frac{\sigma_{\eta 1\eta 2,P} - \sigma_{\eta 1\eta 2,C}}{\sigma^2_{\eta 2,P} - \sigma^2_{\eta 2,C}} \\ &= \frac{H_{12}\left(\sigma^2_{\varepsilon 2,P} - \sigma^2_{\varepsilon 2,C}\right) + H_{21}\left(\sigma^2_{\varepsilon 1,P} - \sigma^2_{\varepsilon 1,C}\right)}{\left(\sigma^2_{\varepsilon 2,P} - \sigma^2_{\varepsilon 2,C}\right) + H^2_{21}\left(\sigma^2_{\varepsilon 1,P} - \sigma^2_{\varepsilon 1,C}\right)} \\ &= \frac{H_{12}\Delta\left(\sigma^2_{\varepsilon 2}\right)}{\Delta\left(\sigma^2_{\varepsilon 2}\right)} = H_{12}, \end{aligned} \quad (5)$$

where

$$\Sigma_{\eta,r} = \begin{pmatrix} \sigma^2_{\eta 1,r} & \sigma_{\eta 1\eta 2,r} \\ \sigma_{\eta 1\eta 2,r} & \sigma^2_{\eta 1,r} \end{pmatrix},$$

and the $\Delta(\cdot)$ operator takes the difference in the argument between regimes. If in fact $\sigma^2_{\varepsilon 1,P} \neq \sigma^2_{\varepsilon 1,C}$, then H_{12} will be misidentified, since the $\Delta\left(\sigma^2_{\varepsilon 1}\right)$ terms will not vanish and equation (5) does not hold.

I now move from H_{12}, identified in population, to possible estimators, $\hat{H}_{12}$. The sample analogs from the left-hand side of equation (5) can be simply estimated:

$$\hat{H}_{12} = \frac{\Delta\left(\hat{\sigma}_{\eta 1\eta 1}\right)}{\Delta\left(\hat{\sigma}^2_{\eta 2}\right)} = \frac{\frac{1}{T_P}\sum_{t\in P}\eta_{1t}\eta_{2t} - \frac{1}{T_C}\sum_{t\in C}\eta_{1t}\eta_{2t}}{\frac{1}{T_P}\sum_{t\in P}\eta^2_{2t} - \frac{1}{T_C}\sum_{t\in C}\eta^2_{2t}}.$$

However, this estimator is equivalent to that of an instrumental variables problem (Rigobon & Sack, 2004):

$$\frac{\frac{1}{T_P}\sum_{t\in P}\eta_{1t}\eta_{2t} - \frac{1}{T_C}\sum_{t\in C}\eta_{1t}\eta_{2t}}{\frac{1}{T_P}\sum_{t\in P}\eta^2_{2t} - \frac{1}{T_C}\sum_{t\in C}\eta^2_{2t}} = \frac{\frac{1}{T}\sum_{t=1}^{T}\eta_{1t}Z_t}{\frac{1}{T}\sum_{t=1}^{T}\eta_{2t}Z_t}, \quad (6)$$

[5]In principle, the single-variance change assumption is compatible with larger models (with similar results) but is rarely combined with them in practice.

where

$$Z_t = \left[\mathbf{1}(t \in P) \times \frac{T}{T_P} - \mathbf{1}(t \in C) \times \frac{T}{T_C} \right] \eta_{2t}. \tag{7}$$

Thus, $\hat{H}_{12}$ can also be estimated via IV or TSLS, using

$$\begin{aligned} \text{first stage:} \quad & \underbrace{\eta_{2t}}_{X} = \underbrace{\Pi Z_t}_{\Pi Z} + \upsilon_t, \\ \text{second stage:} \quad & \underbrace{\eta_{1t}}_{Y} = \underbrace{H_{12}\eta_{2t}}_{\beta X} + u_t, \end{aligned} \tag{8}$$

where standard IV notation is indicated below the terms as a function of the structural parameters, $\Pi = \left(\sigma^2_{\varepsilon 2,P} - \sigma^2_{\varepsilon 2,C}\right) / \left(\tau_P^{-1}\left(H_{21}^2 \sigma^2_{\varepsilon 1} + \sigma^2_{\varepsilon 2,P}\right) + \tau_C^{-1}\left(H_{21}^2 \sigma^2_{\varepsilon 1} + \sigma^2_{\varepsilon 2,C}\right)\right)$. If Z_t is strongly correlated with the innovation η_{2t} (exogeneity follows from equation (1) and assumption 1), standard asymptotic results for IV/TSLS apply. First,

$$\hat{H}_{12} = \frac{\frac{1}{T}\sum_{t=1}^{T} \eta_{1t} Z_t}{\frac{1}{T}\sum_{t=1}^{T} \eta_{2t} Z_t} \xrightarrow{p} \frac{E\left[\eta_{1t} Z_t\right]}{E\left[\eta_{2t} Z_t\right]} = H_{12}, \tag{9}$$

as long as the denominator, $\frac{1}{T}\sum_{t=1}^{T} \eta_{2t} Z_t$, does not converge to 0, so Slutsky's theorem can be applied. Moreover, Slutsky's theorem shows that provided the denominator does not converge to 0, the asymptotic distribution will be fully characterized by the behavior of the numerator. In particular, under a martingale central limit theorem,

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

V_{strong} is the usual White (1980) heteroskedasticity-robust IV/TSLS asymptotic variance, $E\left[\eta_{2t} Z_t\right]^{-2} E\left[u_t^2 Z_t^2\right]$.

If the denominator is in fact close to 0, standard inference methods are not reliable in the familiar IV setting (e.g., Staiger & Stock, 1997). As the first-stage coefficient, Π, tends to 0, the instrument provides less information about the endogenous regressor. Here, Π goes to 0 as $\sigma^2_{\varepsilon 2,P}$ approaches $\sigma^2_{\varepsilon 2,C}$, the case of no variance change.

If $\sigma^2_{\varepsilon 2,P} = \sigma^2_{\varepsilon 2,C}$ $(\Pi = 0)$, so H_{12} is unidentified, then the denominator (and numerator) of equation (6) converges in probability to 0. To obtain a limiting distribution, multiplying equation (6) by $\frac{\sqrt{T}}{\sqrt{T}}$ illustrates that both numerator and denominator converge in distribution to mean-0 normal random variables. $\hat{H}_{12}$ 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 equation (6) is nonuniform with respect to $\left(\sigma^2_{\varepsilon 2,P}, \sigma^2_{\varepsilon 2,C}\right)$: if $\sigma^2_{\varepsilon 2,P} \neq \sigma^2_{\varepsilon 2,C}$ the limiting distribution is normal, but if $\sigma^2_{\varepsilon 2,P} = \sigma^2_{\varepsilon 2,C}$, it is not. To derive an asymptotic distribution that well approximates the behavior of $\hat{H}_{12}$ when $\sigma^2_{\varepsilon 2,P}$ is close to, but not equal to, $\sigma^2_{\varepsilon 2,C}$, I follow convention and model the difference as "small." In particular,

$$\frac{\sigma^2_{\varepsilon 2,P}}{\sigma^2_{\varepsilon 2,C}} = 1 + \frac{d}{\sqrt{T}}, \tag{10}$$

where d is finite. Rearranging yields

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

so $\sigma^2_{\varepsilon 2,P}$ is "local to $\sigma^2_{\varepsilon 2,C}$." Employing this device means that even as $T \to \infty$, the probability of rejecting the hypothesis $\sigma^2_{\varepsilon 2,P} = \sigma^2_{\varepsilon 2,C}$ tends to neither 0 nor 1, capturing the intermediate case of weak identification.

With this model of $\sigma^2_{\varepsilon 2,P}$ and $\sigma^2_{\varepsilon 2,C}$ in hand, the asymptotic distribution of $\hat{H}_{12}$ under weak identification is similar to that for the standard IV model:

Proposition 2. *Under the device (10) and assumptions 1 and 3, $\hat{H}_{12}$ is not consistent for H_{12}; rather,*

$$\hat{H}_{12} - H_{12} \xrightarrow{d} \frac{z_1}{d_\sigma + z_2}, \; \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(0, V_{weak}\right), \tag{11}$$

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

Proposition 2 follows from an argument in the spirit of Staiger and Stock (1997), presented in the appendix. The estimator is no longer consistent. Likewise, V_{weak} cannot be consistently estimated. The reason is that, asymptotically, the denominator $\frac{1}{T}\sum_{t=1}^{T} \eta_{2t} Z_t \xrightarrow{p} 0$. As the identifying variation becomes small, sampling variation in the consistently estimated means matters for the asymptotic distribution of $\hat{H}_{12}$.

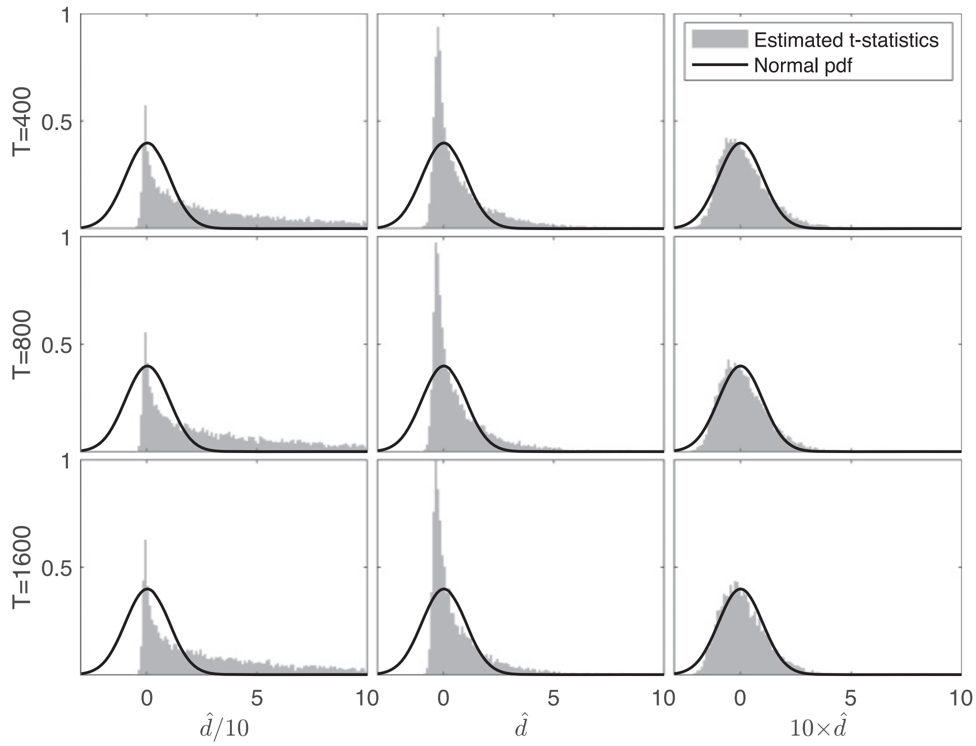
The estimator's asymptotic distribution is thus better represented as a ratio of two correlated normals. Inference approaches based on the normal approximation break down. A bootstrap approach for $\hat{H}_{12}$ (for Wald-type inference) is also invalid, as shown in Moreira, Porter, and Suarez (2004). Similarly, a GMM application of the IV estimator will fare no better (Stock, Wright, & Yogo, 2002). Instead, robust methods developed for weak instruments must be used.

In section IVA, I prove the validity of an inference procedure robust to weak identification for individual elements or subsets of H. The conditions required for the procedure will always hold in bivariate models like the one presented above, which are very common in practice. This procedure does not require the zero change assumption for $\sigma^2_{\varepsilon_1}$ discussed in this section but remains valid if it is adopted. This means that regardless of whether the model is weakly identified (which can be determined using the test in section III), valid inference can be conducted without resorting to potentially conservative projection methods.

C. *Weak Identification in the General Case*

I now discuss weak identification in the fully general model. Proposition 1 states the conditions for global identification, which can break down in the two related cases outlined previously. Similar to the simple case, I now model

FIGURE 2.—DISTRIBUTION OF t-RATIOS FOR $\hat{H}_{12}$



Distributions of estimated t-ratios, $(\hat{H}_{12}-H_{12})/SE(\hat{H}_{12})$, calculated from 10,000 Monte Carlo draws, using the sample length in the left margin and the degree of identification in the bottom margin, from the nominal forward and nominal yield specification estimated in section V. The black curve represents the standard normal pdf, the limiting distribution under standard asymptotic arguments. Extreme outliers are truncated to allow comparison on the same axes. Calibration details are given in equation (14) in the supplement. Estimation via CUE efficiently weighted GMM.

the relationship between the variances of two shocks, i and j, as local-to-unity:

$$\frac{\sigma^2_{\varepsilon i,P}/\sigma^2_{\varepsilon i,C}}{\sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j,C}} = 1 + \frac{d}{\sqrt{T}}, \tag{12}$$

where d is finite. In economic terms, the Great Moderation or Great Recession are offered above as examples where variances might change together. 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. The impact on identification is characterized in proposition 3:

Proposition 3. *Adopting the modeling device in equation (12) and assumption 1, equation (2) does not have a unique solution for H as $T \to \infty$.*

Intuitively, under the local-to-unity device, the nonproportionality requirement of proposition 1 fails asymptotically in population, as the variances converge to the knife-edge case $\sigma^2_{\varepsilon i,P} = \left(\sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j,C}\right)\sigma^2_{\varepsilon i,C}$, resulting in an unidentified system. However, the limiting probability of rejecting the hypothesis $\left(\sigma^2_{\varepsilon i,P}/\sigma^2_{\varepsilon i,C}\right)/\left(\sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j,C}\right) = 1$ from (infeasible) observations of ε_t is neither 0 nor 1, capturing the spirit of the intermediate case of weak identification. As identification breaks down, H cannot be consistently estimated, as Stock and Wright (2000) argued. Similarly, standard asymptotic approximations used for inference also fail.

To conclude this section, I demonstrate via simulation just how poor of an approximation standard asymptotic results may provide. I calibrate my simulations to NS's specification using daily changes in two-year nominal Treasury forward rates as a dependent variable and daily changes in two-year nominal Treasury yields as the policy series. Additional details can be found in the empirical application below and in section C of the supplement. I vary T from 400 to 1,600 (the empirical sample size is approximately 800) and vary the degree of identification, d, by a factor of 10 in each direction from the $\hat{d}$ implied by the data. I do this by adjusting the variance of monetary policy shocks on policy dates, setting

$$\sigma^2_{\varepsilon 2,P}(d) = \hat{\sigma}^2_{\varepsilon 2,C}\frac{\hat{\sigma}^2_{\varepsilon 1,P}}{\hat{\sigma}^2_{\varepsilon 1,C}}\left(1 + \frac{d}{\sqrt{T}}\right).$$

Figure 2 presents histograms of the t-ratio, $\left(\hat{H}_{12} - H_{12}\right)/SE\left(\hat{H}_{12}\right)$, for 10,000 draws. The estimates are not normally distributed for low degrees of identification, even as T grows large. For the "strong identification" specifications, the distribution is closer to a normal distribution; these specifications map to about seven times the empirical relative change in the policy shock variance and twice the change observed in the Treasury yield innovation variance. These distributions constitute prima facie evidence of weak identification. It is clear that relying on standard inference methods, assuming asymptotic normality for estimates, may lead to unreliable tests under weak identification, as such an assumption is a poor approximation to the true distribution of the estimator.

III. A Simple Pretest for Weak Identification

In the IV literature, following influential work by Staiger and Stock (1997) and Stock and Yogo (2005), it is now common practice to perform pretests to assess the strength of identification. The standard test compares the first-stage F-statistic to a threshold of 10, proposed by Staiger and Stock (1997). Given that the empirically common simple case of identification via heteroskedasticity outlined in section IIB can be recast as an IV problem, I propose a similar test to allow such a practice to be adopted in this setting too.

As demonstrated in equations (7) and (8), the simple case can be recast as a just-identified linear IV model with a single endogenous regressor. As a reminder, this is a bivariate model, where it is assumed that only the variance of the second structural shock changes across regimes. This means that a first-stage F-test approach can be adopted. In the setting of section IIB, the heteroskedasticity-robust first-stage F-statistic is

$$F = \frac{\hat{\Pi}^2 \left(\sum_{t=1}^T Z_t^2\right)^2}{\sum_{t=1}^T Z_t^2 \hat{\upsilon}_t^2},$$

where $\hat{\Pi}$ and $\hat{\upsilon}_t$ are OLS estimates from the first stage, and

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

Note that the heteroskedasticity-robust F-statistic must be used even if the shocks are homoskedastic within regimes, since υ_t^2 will generally be correlated with Z_t^2. However, the critical values of Stock and Yogo (2005) are valid only under homoskedasticity. Fortunately, Montiel Olea and Pflueger (2013) develop alternative critical values under weaker assumptions.[6] They allow for arbitrary heteroskedasticity and autocorrelation and calibrate critical values to the Nagar bias of TSLS (the first three terms of a Taylor expansion of the asymptotic distribution) relative to a "worst case" benchmark. Table 1 reports the relevant critical values. A threshold of $F > 23$ corresponds to 10% bias, the threshold motivating the $F > 10$ rule-of-thumb found in the IV literature. This test can easily be adopted in this specialized setting and I provide code to do so on my website.

Table 1.—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 are calculated by Montiel Olea and Pflueger (2013). For a given critical value, bias is greater than that indicated in 5% of repeated samples. Assumptions underlying these results are enumerated in note 6.

[6]In particular, in the TSLS framework, they assume $\frac{1}{\sqrt{T}}\left(\sum_{t=1}^T Z_t \upsilon_t \quad \sum_{t=1}^T Z_t u_t\right)'$ is asymptotically normal with consistently estimable positive definite covariance, the covariance of $(\upsilon_t \quad u_t)'$ is positive definite and consistently estimable, and a local-to-0 representation for Π. See Montiel Olea and Pflueger (2013) for additional details.

IV. Weak Identification Robust Inference

In this section, I present standard results for robust inference in the general model, followed by a new result justifying nonconservative inference on empirically relevant subsets of θ. I outline how these confidence sets can be used to test for weak identification and conclude with a simulation study.

A. *Parameter Inference*

The asymptotic behavior of GMM estimators, robust to weak identification, is established in Stock and Wright (2000). Instead of providing an asymptotic distribution for estimates $\hat{\theta}$, as in strongly identified GMM problems, they show that $S_T(\theta_0)$, the "S-statistic," follows a χ^2 distribution. Many refinements have since been developed, including the "K-statistic" of Kleibergen (2005). Much of this literature is limited to joint tests on the full parameter vector or subsets of the parameter vector including all parameters that are weakly identified, as only strongly identified nuisance parameters are "concentrated out." However, the parameter of interest in applied work is often some subset of the parameter vector; this is the case in Rigobon and Sack (2003) (response of three-month Treasury rate to S&P 500 shocks), Rigobon and Sack (2004) (response of equity indices and long-term rates to monetary policy shocks), Wright (2012) (response of long-term interest rates to monetary policy shocks), Hébert and Schreger (2017) (response of equities and exchange rates to sovereign default shocks), and NS (2018) (response of Treasury forward rates to monetary policy shocks), for example. 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 in this setting.

I state results using the K-statistic of Kleibergen (2005). In the leading two-regime case considered here, the K-statistic coincides with the S-statistic of Stock and Wright (2000) since the model is just identified. Further refinements may have better power properties in overidentified models (e.g., the conditional linear combination tests of Andrews, 2016).

Definition 1. *In the leading two-regime just-identified case,* $K_T(\theta) = S_T(\theta) = \frac{1}{T}\phi_T(\theta, \eta)'\Omega_T(\theta)^{-1}\phi_T(\theta, \eta)$, *where* $\phi_T(\theta, \eta) = \sum_{t=1}^T \phi(\theta, \eta_t)$ *and* $\Omega_T(\theta) = \frac{1}{T}\sum_{t=1}^T \phi(\theta, \eta_t)\phi(\theta, \eta_t)'$. *In the more general R-regime case,* $K_T(\theta) = \frac{1}{T}\phi_T(\theta, \eta)'$

$\Omega_T(\theta)^{-1/2} P_{\Omega_T(\theta)^{-1/2} J_T(\theta,\eta)} \Omega_T(\theta)^{-1/2} \phi_T(\theta,\eta)$, *where* $J_T(\theta,\eta) = \frac{\partial \phi_T(\theta,\eta)}{\partial \theta'} = -T \frac{\partial (\tau_1 vech(H\Sigma_{\varepsilon,1}H)', \ldots, \tau_R vech(H\Sigma_{\varepsilon,R}H)')'}{\partial \theta'}$.[7]

Full vector inference. Under the assumptions presented in section II, theorem 1 shows that inference on the full parameter vector can proceed using the K-statistic.

Theorem 1. *Under assumptions 1, 2, and 3, if the null hypothesis* $\theta = \theta_0$ *holds,*

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

As discussed below, Magnusson and Mavroeidis (2014) consider this test (their *split-KLM*) for identification via heteroskedasticity.

Inference for subsets of the parameter vector. Projection methods constitute the leading option for subset inference when some nuisance parameters are weakly identified. Such tests are notoriously conservative; the full-vector 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 (see Chaudhuri & Zivot, 2011, for a discussion and refinements).

However, Kleibergen (2005) provides a refinement over theorem 1 for tests on certain subsets of the parameter vector. Partition θ into the parameter(s) of interest, β, and the nuisance parameters, α. If the rank of the asymptotic Jacobian of the moment equations with respect to α is equal to the dimension of α, there exists a unique solution for α given a value for β. Thus, since β_0 is specified under the null hypothesis, inference may use degrees of freedom equal to the dimension of β (Kleibergen, 2005, theorem 2), as α can be concentrated out, even if it is not strongly identified ex ante. The elements of β, specified as the null hypothesis, may be either weakly or strongly identified. Assumption 4 and theorem 2 state this result formally.

Assumption 4. For a given value of β, a subset of the parameters in H, $\Sigma_{\varepsilon,C}$, and $\Sigma_{\varepsilon,P}$, equation (2) has a unique solution for α, the remaining parameters, as $T \to \infty$.

Assumption 4 is a global analog to the local assumption 3 in Kleibergen (2005). Although T does not appear explicitly in equation (2), uniqueness must hold asymptotically in light of the local-to-unity device (12). While it might be obvious that parameters α that can be written as continuous functions of β can be concentrated out when conducting inference, assumption 4 characterizes a more general class of models for which α can be concentrated out in theorem 2. Define $K_T(\alpha,\beta) = K_T(\theta)$ where $\theta = (\alpha' \quad \beta')'$ and further $K_T(\beta) = K_T(\alpha(\beta), \beta)$, where $\alpha(\beta) = \text{argmin}_\alpha K_T(\alpha,\beta)$; $K_T(\beta)$ is the subset K-statistic for β. Theorem 2 and the subsequent definition 2 of Kleibergen (2005) imply theorem 2:

[7]Note the simplification from D_T in Kleibergen (2005) to the Jacobian J_T results from the fact that the Jacobian of the moments is deterministic in this model; see lemma 1 for details.

Theorem 2. *Under assumptions 1, 2, and 3, if assumption 4 and the null hypothesis* $\beta = \beta_0$ *additionally hold, then*

$$K_T(\beta_0) \xrightarrow{d} \chi^2_p,$$

where p is the dimension of β.

The degrees of freedom of the limiting distribution for the full parameter vector (or projection tests for a subset) is lowered from $n^2 + n$ to p. I henceforth refer to the test comparing $K_T(\beta_0)$ to the χ^2_p critical values as the subset test, since it uses these critical values specific to the subset tested.

When does the present model satisfy assumption 4? Magnusson and Mavroeidis (2014) show that strongly identified parameters can be concentrated out in settings including the present model (theorems 6 and 7). However, a crucial aspect of Kleibergen's result, and that of Magnusson and Mavroeidis (2014), is that their definitions of strongly identified parameters include those that are only strongly identified after specifying a value β_0 (not just those that are strongly identified a priori). I exploit this distinction to derive subset robust inference results for identification via heteroskedasticity.

First, I introduce a partition of H:

Definition 2. *Partition H as* $H^I \vdots H^W$ *such that* $H_{\cdot k} \in H^I$ *if and only if* $\left(\sigma^2_{\varepsilon k,C} \quad \sigma^2_{\varepsilon k,P}\right)$ *is proportional to no other row in*

$$\begin{bmatrix} \sigma^2_{\varepsilon 1,C} & \sigma^2_{\varepsilon 1,P} \\ \vdots & \vdots \\ \sigma^2_{\varepsilon n,C} & \sigma^2_{\varepsilon n,P} \end{bmatrix},$$

and conversely for H^W.

H^I is uniquely determined from equation (2), while H^W is not.

In empirical work, the object of interest is generally either the immediate impact of one shock on one variable or the shock's impact on all variables. The former consists of a single element of H; the latter pertains to a full column (Rigobon & Sack, 2003, 2004; Wright, 2012; Hébert & Schreger, 2017; NS). Theorem 3 shows that if such parameters are in H^W, specifying them as β_0 ensures a unique solution for $\alpha(\beta_0)$.

Theorem 3. *Under assumptions 1 and 2, if* H^W *contains two columns, equation (2) has a unique solution for* α *after fixing as* β_0 *either*

1. *A single element* H_{lk}, *with* $H_{lk} \neq H_{lm}/H_{km}$ *for* $H_{\cdot k}, H_{\cdot m} \in H^W$, *or*
2. *The full column* $H_{\cdot k} \in H^W$.

By explicitly conditioning on the information to be used in the null hypothesis of the subset test (which fixes H_{lk} or $H_{\cdot k}$), theorem 3 provides primitive conditions under which there is a unique solution for α after fixing a value for the parameter of interest in β_0. This means that a system of equations satisfying the conditions of theorem 3 meets assumption 4, and theorem 2 may be applied. The ancillary condition on the relative magnitudes of elements of H^W can be seen as strengthening the standard invertibility condition on H to an invertibility assumption on a subblock of H. In the $n = 2$ case, this result may be obvious, since equation (4) of Rigobon (2003) shows that if one free parameter of H is known, the other is a simple function of that parameter and the reduced-form variances (and can thus be concentrated out). However, that is not the case for $n \geq 2$, where knowledge of one parameter of H^W is not sufficient to immediately solve for the remainder.

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 $\lim_{T\to\infty}\left(\sigma^2_{\varepsilon i,P}/\sigma^2_{\varepsilon i,C}\right)/\left(\sigma^2_{\varepsilon j,P}/\sigma^2_{\varepsilon j,C}\right) = 1$, and either i or j is the shock of interest, then assumption 4 is satisfied for tests where β contains the corresponding column of H or a single element of the column (and possibly additional parameters); theorem 2 holds.

Significantly, condition 1 may justify critical value refinements for subset inference even when no model parameters are strongly identified. Since most empirical papers focus on a single element of H or a column of H, this result means subset inference can frequently proceed using reduced degrees of freedom instead of projection methods, provided proportionality in variance changes is not too prevalent. Five remarks clarify the impact of condition 1:

Remark 1. Condition 1 nests the cases where $\beta = \theta$ or where β is the set of all 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 adds no new information.

Remark 3. Given at most two variances may evolve proportionally, a researcher should minimize the dimension of η_t subject to the constraint that η_t spans ε_t (invertibility).

Remark 4. In empirical practice, bivariate systems are common. In this case, both the limit on proportionality and the condition on relative magnitudes in H are nonbinding, so theorem 3 can always be applied.

Remark 5. It is straightforward to extend the results of theorem 3 to IRFs. For a detailed discussion, see section D of the online supplement.

To complete the discussion of robust inference, I relate my results to those in the existing literature. 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 and Mavroeidis (2014). They propose a variety of tests that accommodate the present setting. In section 3.5, they show that in this setting, the asymptotic distributions of the *split-S* and *split-KLM* test statistics are unaffected by estimating ζ, a strongly identified nuisance parameter, retaining degrees of freedom equal to the number of weakly identified parameters, their θ (their theorems 6 and 7). When weak identification arises, there need not be any a priori strongly identified parameters, but some parameters still may be uniquely determined once a null hypothesis β_0 is specified, and concentrated out on that basis. Theorem 3 and condition 1 offer economically interpretable primitive conditions under which such high-level results apply.

NS 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, and Watson (1997). This approach only works in their simple case, since their test statistic depends only on H_{12}, by virtue of the direct analogy to IV. With multiple variance changes, the test statistic they propose 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. Their test coincides asymptotically with an S-test.

A general test for weak identification. The two-step approach of Andrews (2018) for detecting weak identification can be adopted for the general model. This test can be applied to a subset of the parameter vector or the full vector. First, a researcher decides on a maximal acceptable size distortion that she believes is compatible with strong identification, say, $\gamma = 0.1$. Then a preliminary robust confidence set is constructed to have size $1 - \nu - \gamma$, where ν is the desired level of the test, say, $\nu = 0.05$. This robust set will be valid regardless of the true strength of identification. Next, a $1 - \nu$ confidence set is constructed under strong identification asymptotics (based on standard t inference, say). If this second set contains the preliminary robust set, then identification is not so weak that the size distortion surpasses the prespecified threshold. The parameter(s) can be said to be strongly identified and standard inference methods adopted. If the preliminary set is not contained, weak identification cannot be rejected, and a robust $1 - \nu$ set should be constructed for inference. More details can be found in Andrews (2018). Nonconservative subset tests based on theorem 3 are particularly valuable for this purpose. Given how conservative projection methods can be, it would be highly impractical to use them here, since the resulting confidence sets are so large; the sets are unlikely to be contained by a lower-size standard confidence set, even if strong identification truly holds.

In a recent paper, Lütkepohl et al. (2020) present a method to test whether variance changes are proportional in models identified via heteroskedasticity. Their approach assesses whether the changes differ in a statistically significant way; the mapping between p-values and the strength of identification requires further study. In addition, when $n > 2$, their theory does not yet offer guidance on how to conduct sequential tests to determine if the model is fully identified.

TABLE 2.—SIZE OF F-TEST AND K-TEST OF THE FULL PARAMETER VECTOR

	$\hat{d}/10$		$\hat{d}$		$\hat{d} \times 10$	
	Wald	K	Wald	K	Wald	K
$T = 400$	94.9	16.7	75.9	16.7	36.3	16.7
$T = 800$	95.5	11.2	76.5	11.2	30.7	11.2
$T = 1{,}600$	95.7	8.4	76.6	8.4	27.1	8.4

Rejection rates of the true parameter vector for a nominal 5% test, based on 10,000 Monte Carlo draws. Calibration details are given in equation (14) of the supplement. Estimation via efficiently weighted CUE GMM.

B. *Monte Carlo Performance of Tests*

I demonstrate the superior size-control properties of the proposed robust tests, as well as the power improvements provided by the subset inference results, in a Monte Carlo study. I calibrate the DGP to the two-variable empirical example from NS that I explore in section V. For additional details on the setting and calibration, see section C of the online supplement.

Rejection rates of tests on the full parameter vector. Rejection rates for both Wald and robust (K) tests with a nominal level of 5% are reported in Table 2. The value $\hat{d}$ implied by the data measures the empirical degree of identification, and $T = 800$ the approximate empirical sample size; I vary both across columns and rows by setting $\sigma^2_{\varepsilon 2,P}(d) = \hat{\sigma}^2_{\varepsilon 2,C}\left(\hat{\sigma}^2_{\varepsilon 1,P}/\hat{\sigma}^2_{\varepsilon 1,C}\right)\left(1 + d/\sqrt{T}\right)$.[8] The Wald tests exhibit extremely large size distortions, aligned with the theoretical result of Dufour (1997) that the size of such tests will tend to unity as the degree of identification tends to 0. The distortions improve with the strength of identification. The robust tests are unaffected by the degree of identification. Their size does decrease with sample size, which is due only to small-sample behavior. 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 II, the stronger calibration imposes nearly seven times the empirically observed change in the structural variance of the policy shock.

Rejection rates of subset tests on a single parameter of interest. I compare three subsample testing approaches: the t-test, the projected K-test based on theorem 1, and the newly proposed subset K-test with reduced critical values, based on theorem 2 and condition 1. Rejection rates for nominally 5% tests are displayed in Table 3. First, like Wald tests on the full parameter vector, the standard t-test is substantially oversized, although the distortion is not as large as for the full vector. As identification gets stronger, the distortions shrink. The K-test based on projection methods is substantially undersized, with a rejection rate of effectively 0 in simulation. However, K_{sub}, justified by condition 1, 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.

Power improvements in subset testing. The use of smaller critical values for the subset tests justified by condition 1 relative to projection tests implies automatic power improvements. I fix $T = 800$ and vary the strength of identification, testing a null of $H_{12} = -0.31$ (the empirical estimate; see section V and online supplement section C) against a sequence of alternative values of H_{12} used to simulate the data. Figure 3 computes power curves based on these simulations. For the weakest identification, the power of the t-test (solid line) unsurprisingly dominates that of the robust tests. The subset test (dashed line) is more powerful than the projection test (dot-dashed line), as expected; they use the same test statistic, but the subset test uses smaller critical values. For the baseline calibration, the results are similar, except that the subset test surpasses the t-test for alternatives left of the null (where its power is very low) and increases markedly in power for alternatives to the right, with smaller gains for the projection test. For the strong identification calibration, the t-and subset tests are comparable, and noticeably dominate the projection test. The online supplement contains results for a broader range of alternatives and size-adjusted power, along with a more detailed discussion.

V. Empirical Application

I demonstrate the use of the proposed robust inference methods by studying the identification of monetary policy shocks in the setting of NS.[9] The authors analyze the impact of policy shocks on nominal and real instantaneous Treasury 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 three such instruments: the daily change in nominal two-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 either the daily Treasury yields or the authors' 30-minute window policy news series as the policy instrument, with either nominal or real two-year instantaneous Treasury forward rates as the

[8] $\hat{d}$, with $\hat{\sigma}^2_{\varepsilon i,r}$ the estimated average variances in regime r, remains an accurate measure of the degree of identification even in the possible presence of within-regime heteroskedasticity, since identification is based on these average variances. Identification does not exploit any additional information contained within the regimes.

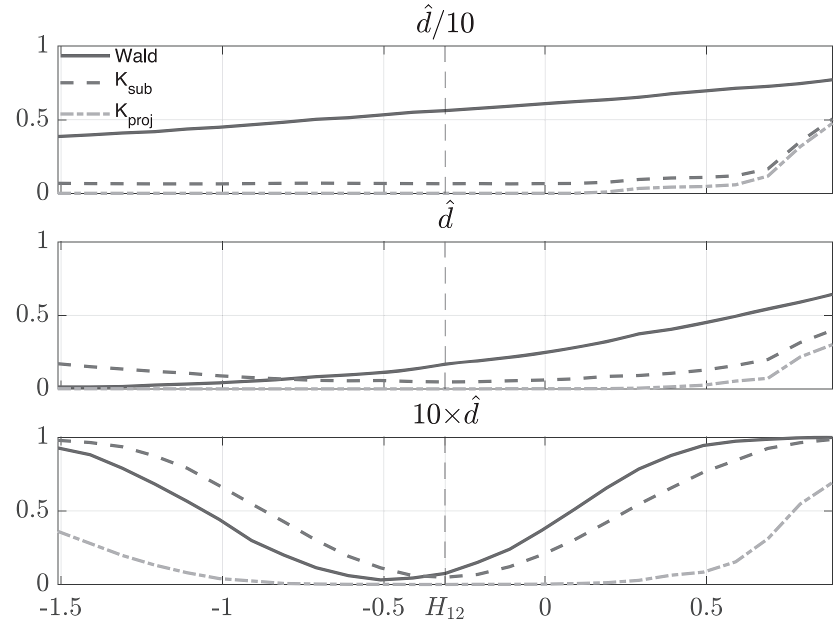
[9] I am grateful to Emi Nakamura and Jòn Steinsson for making their policy news series available to me.

TABLE 3.—SIZE OF t-TEST, PROJECTION K-TEST, AND SUBSET K-TEST OF $\hat{H}_{12}$

	$\hat{d}/10$			$\hat{d}$			$10 \times \hat{d}$		
	t	K_{proj}	K_{sub}	t	K_{proj}	K_{sub}	t	K_{proj}	K_{sub}
$T = 400$	54.1	0.0	6.7	16.6	0.0	4.5	7.9	0.0	4.4
$T = 800$	53.8	0.0	6.9	14.2	0.0	4.9	6.8	0.0	4.9
$T = 1,600$	52.9	0.0	6.8	13.3	0.0	5.2	6.2	0.0	5.2

Rejection rates of the true parameter value for H_{12} for a nominal 5% test based on 10,000 Monte Carlo draws. The K_{proj} results are not identically 0, but round to 0.0. Calibration details are given in equation (14) of the online supplement. Estimation via efficiently weighted CUE GMM.

FIGURE 3.—POWER CURVES FOR TESTS OF H_{12}



Power curves are formed from estimates of rejection rates of the null hypothesis ($H_{12} = -0.31$, dashed vertical line) against a sequence of alternatives (x-axis) based on 1,000 Monte Carlo draws. Estimation via efficiently weighted CUE GMM.

"dependent" variable. Thus, $\eta_t = (\Delta s_t \quad \Delta i_t)'$ where s_t is a forward rate and i_t is the policy instrument. For all specifications, I focus on the sample NS use for their real forward rate specifications, January 2004 to March 2014 (omitting July 2008 to June 2009), so I can maintain a consistent sample across specifications. The qualitative findings are robust to extending the sample back to 2000 where possible.

A. *Tests of Identification and Estimates*

NS 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 develops theory for the fully general model, allowing for the possibility that the variances of both structural shocks might change. Economically, it might make sense for only the variance of the policy shock to change, but if that is the case, the restriction need not be imposed mechanically, as estimation will bear it out. I thus primarily consider the unrestricted model. The monetary policy shock is labeled according to assumption 2.

I first test formally for weak identification using the methods proposed in section III. Under NS's restricted model, the first-stage F-statistic tests for weak identification. These results are reported in the left 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 large, and weak identification is rejected for all levels of bias. The Andrews (2018) test for the general model is reported in the right 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 intervals; see table 5). The differences between daily and intraday data arise since, while daily data do contain the high-volatility periods captured in the 30-minute windows on announcement days, the overall volatility of announcement days shrinks toward that of nonannouncement days because of nonelevated volatility during the rest of the day. These test results corroborate the less formal observations of NS, who suspect weaker identification in daily data.

Table 4.—Tests of Identification

	F	First-Stage F (bias) 0.2	0.1	0.05	Andrews Two-Step (Size) 0.2	0.15	0.1	0.05
Nominal, daily shock	8.04	×	×	×	×	×	×	×
Real, daily shock					×	×	×	×
Nominal, 30-min shock	7,343.78	✓	✓	✓	✓	✓	×	×
Real, 30-min shock					✓	✓	×	×

The first panel tests each shock series using the first-stage F-statistic bias-based critical values in table 1. The second panel conducts the Andrews (2018) two-step test for each specification for four candidate distortion thresholds, γ. The model is strongly identified in the sense of inference not being distorted more than γ if the $1-\nu-\gamma$ robust confidence set is contained by the $1-\nu$ strong identification set. Andrews's $\hat{\gamma}$, defined such that the $1-\nu-\hat{\gamma}$ robust set is just contained by the strong identification set, is 0.14 for both models using the 30-minute shock instrument.

Table 5.—Estimates of Structural Parameters

Policy Inst.	One-Day Yield		30 Min. News	
Dep. Var.	Nominal Fwd.	Real Fwd.	Nominal Fwd.	Real Fwd.
H_{21}	0.70***	−0.88	0.01*	0.01*
stnd. CI	[0.52, 0.89]	[−22.47, 20.71]	[0.00, 0.02]	[0.00, 0.01]
robust CI	[0.54, 0.78]	[−172.06, 1.14]	[−0.00, 0.01]	[−0.00, 0.01]
H_{12}	−0.31	1.30	1.07**	0.93***
stnd. CI	[−5.16, 4.54]	[−3.85, 6.46]	[0.17, 1.98]	[0.36, 1.51]
robust CI	[−78.46, 0.94]	[−388.68, 1.92]	[0.27, 3.25]	[0.43, 2.37]
$10^3 \times \sigma^2_{s,C}$	3.9	0.4	3.9	3.4
$10^3 \times \sigma^2_{i,C}$	0.1	1.8	0.02	0.02
$10^3 \times \sigma^2_{s,P}$	7.0	0.7	6.2	5.7
$10^3 \times \sigma^2_{i,P}$	0.4	3.4	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 two-year instantaneous Treasury forward rate. The policy instrument is either one-day changes in the two-year nominal Treasury yield or 30-minute changes in NS's policy news series. For the variances, i denotes the monetary policy shock and s the second shock. The standard 95% confidence interval is based on a t-statistic. The robust 95% confidence interval is based on the subset K-test. Asterisks indicate significance from 0 at the *10%, **5%, or ***1% levels based on the robust tests.

Table 5 reports estimates for the unrestricted model. Note that NS do not report estimates for H_{21}, preventing comparison on that dimension. For the 30-minute policy news instrument, the results for H_{12}, the pass-through of policy shocks to forwards (1.07 and 0.93) are extremely close to those for NS's restricted model (1.10 and 0.96), indicating a forward guidance/news effect that shifts expectations. Their assumption that σ_1^2 is fixed has little impact on estimates of H_{12} because H_{21} is near-zero, minimizing the possible bias in equation (5).

Using daily changes in the nominal yield as the policy instrument, the point estimates differ dramatically between the two specifications and from the strongly identified specifications. The estimated pass-through of monetary policy to the real forward rate is somewhat larger than the intraday results and that of NS (compare 1.30 to their 0.97 for H_{12}). On the other hand, the sizable negative value for H_{21} (−0.88) is at odds with the strongly identified zero estimates. The estimated negative pass-through of monetary policy to nominal forward rates is starkly at odds with the other estimates (−0.31; NS obtain 1.14); it indicates that a positive forward guidance shock lowers the two-year instantaneous forward rate, while raising the average rate over the next two years, altering the shape of the yield curve. For this specification, H_{21} is again estimated to be non-0, but this time is positive (0.70). The non-0 values for H_{21} in both weakly identified specifications are consistent with there being a second meaningful dimension of news, but the opposite signs make any such interpretation very different across specifications. These results are at odds with NS's background noise interpretation of the second shock and starkly opposed to the strongly identified specifications.

B. *Performance of Tests*

I now compare confidence sets robust to weak identification to those computed assuming identification is strong. For models using the daily yield series (exhibiting weak identification), the robust confidence intervals, which I compute using the subset K-test based on theorem 3, are much wider than standard confidence intervals for H_{12}; the same is true for H_{21} for the real forward rate.[10] However, the intervals are substantially asymmetric so do not contain the standard confidence intervals. Notably, the estimate $\hat{H}_{21} = 0.70$ for the nominal forward rate specification is statistically significant under both inference procedures, supporting the presence of a second dimension of news affecting nominal yields and forward rates at the daily frequency. For models using the 30-minute window policy news series (exhibiting strong identification), the robust confidence intervals are comparable to the standard ones, and the estimates of H_{12} remain

[10] If the robust confidence set is disjoint, I report the interval spanning all elements contained in the set.

statistically significant at the 5% or 1% level.[11] This conclusion replicates NS's findings using the restricted model. For H_{21}, I obtain precisely estimated 0s.

For the specifications using 30-minute shocks, I can also test the null hypothesis that the nonpolicy shock variance is fixed across regimes, adopting standard inference methods based on the evidence of strong identification. This is the overidentifying assumption used by NS to reduce the model to the simple case. Using a simple Wald test, $p = 0.13$ for the nominal forward specification and $p = 0.07$ for the real forward specification. While equality may not be overwhelmingly rejected, there is ample evidence against using it as an identifying assumption. This finding supports the use of the unrestricted model in the simulations conducted in the paper.

As an additional exercise, I compute confidence intervals for impulse responses based on the NS data. The results are broadly similar to those for the contemporaneous responses, with standard intervals far too narrow; details are in section D of the online supplement.

VI. Conclusion

This paper provides a framework for conducting 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 affect the reliability of standard inference methods in empirical data. I propose tests to detect weak identification, allowing researchers to determine whether they ought to confront these concerns.

I establish conditions under which robust inference for a subset of the parameter vector can use smaller critical values than those required for projection methods. Such tests constitute the first option for robust inference in this context that is not likely to be prohibitively conservative. Given the problem posed by robust subset inference in nonlinear models in general, the idea of concentrating out weakly identified parameters that are uniquely determined only after a null hypothesis is specified may be more broadly useful.

I apply these methods to the identification of monetary policy shocks, as in NS. Daily data exhibit several symptoms of weak identification, but intraday data strongly identify monetary policy shocks. Daily data are frequently used in macrofinancial contexts, so this finding has implications for the design of empirical studies. It remains to examine the prevalence weak identification in lower-frequency (e.g., monthly, quarterly) data.

Following Staiger and Stock (1997), papers using IV report first-stage F-statistics to justify instrument relevance. Up to now, reporting similar evidence has not been possible for the growing literature exploiting identification via heteroskedasticity, but the results presented in this paper enable researchers to do so.

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

REFERENCES

Andrews, Isaiah, "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica* 84:6 (2016), 2155–2182. 10.3982/ECTA12407

——— "Valid Two-Step Identification-Robust Confidence Sets for GMM," this REVIEW 100:2 (2018), 337–348. 29751022

Brunnermeier, Markus, Darius Palia, Karthik Sastry, and Christopher Sims, "Feedbacks: Financial Markets and Economic Activity," *American Economic Review* 111:6 (2021), 1845–1879.

Chaudhuri, Saraswata, and Eric Zivot, "A New Method of Projection-Based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics* 164:2 (2011), 239–251.

Craine, Roger, and Vance L. Martin, "International Monetary Policy Surprise Spillovers," *Journal of International Economics* 75:1 (2008), 180–196. 10.1016/j.jinteco.2007.06.005

Dufour, Jean Marie, "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica* 65:6 (1997), 1365–1387. 10.2307/2171740

Ehrmann, Michael, and Marcel Fratzscher, "Euro Area Government Bonds: Fragmentation and Contagion during the Sovereign Debt Crisis," *Journal of International Money and Finance* 70 (2017), 26–44. 10.1016/j.jimonfin.2016.08.005

Eichengreen, Barry, and Ugo Panizza, "A Surplus of Ambition: Can Europe Rely on Large Primary Surpluses to Solve its Debt Problem?," *Economic Policy* 31:85 (2016), 5–49. 10.1093/epolic/eiv016

Feenstra, Robert C., and David E. Weinstein, "Globalization, Markups, and US Welfare," *Journal of Political Economy* 125:4 (2017), 1040–1074. 10.1086/692695

Fernandez-Perez, Adrian, Bart Frijns, and Alireza Tourani-Rad, "Contemporaneous Interactions among Fuel, Biofuel and Agricultural Commodities," *Energy Economics* 58 (2016), 1–10. 10.1016/j.eneco.2016.05.014

Fisher, Franklin, "Near-Identifiability and the Variances of the Disturbance Terms," *Econometrica* 33:2 (1965), 409–419. 10.2307/1909798

Gong, Xun, Shenggang Yang, and Min Zhang, "Not Only Health: Environmental Pollution Disasters and Political Trust," *Sustainability* 9:4 (2017), 1–28.

Gürkaynak, Refet S., Burçin Kisacikoglu, and Jonathan H. Wright, "Missing Events in Event Studies: Identifying the Effects of Partially Measured News Surprises," *American Economic Review* 110:12 (2020), 3871–3912.

Hébert, Benjamin, and Jesse Schreger, "The Costs of Sovereign Default: Evidence from Argentina," *American Economic Review* 107:10 (2017), 3119–3145.

Hogan, Vincent, and Roberto Rigobon, "Using Heteroscedasity to Estimate the Returns to Education," Centre for Economic Research working paper series WP03/01 (2003).

Islam, Asadul, Faridul Islam, and Chau Nguyen, "Skilled Immigration, Innovation, and the Wages of Native-Born Americans," *Industrial Relations* 56:3 (2017), 459–488. 10.1111/irel.12182

Jahn, Elke, and Enzo Weber, "Identifying the Substitution Effect of Temporary Agency Employment," *Macroeconomic Dynamics* 20:5 (2016), 1264–1281. 10.1017/S1365100514000820

Khalid, Usman, "The Effect of Trade and Political Institutions on Economic Institutions," *Journal of International Trade and Economic Development* 26:1 (2016), 89–110. 10.1080/09638199.2016.1206142

Kleibergen, Frank, "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica* 73:4 (2005), 1103–1123. 10.1111/j.1468-0262.2005.00610.x

Klein, Roger, and Francis Vella, "Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments," *Journal of Human Resources* 44:4 (2009), 1047–1065. 10.1353/jhr.2009.0036

Lin, Faqin, Ermias O. Weldemicael, and Xiaosong Wang, "Export Sophistication Increases Income in Sub-Saharan Africa: Evidence from 1981–2000," *Empirical Economics* 52:4 (2016), 1627–1649. 10.1007/s00181-016-1103-7

Lütkepohl, Helmut, Mika Meitz, Aleksei Netšunajev, and Pentti Saikkonen, "Testing Identification via Heteroskedasticity in Structural Vector Autoregressive Models," *Econometrics Journal* 24:1 (2021), 1–22.

Magnusson, Leandro M., and Sophocles Mavroeidis, "Identification Using Stability Restrictions," *Econometrica* 82:5 (2014), 1799–1851. 10.3982/ECTA9612

Millimet, Daniel L., and Jayjit Roy, "Empirical Tests of the Pollution Haven Hypothesis When Environmental Regulation Is Endogenous," *Journal of Applied Econometrics* 31:4 (2016), 652–677. 10.1002/jae.2451

Mönkediek, Bastian, and Hilde A. J. Bras, "The Interplay of Family Systems, Social Networks and Fertility in Europe Cohorts Born between 1920 and 1960," *Economic History of Developing Regions*, 31:1 (2016), 136–166.

Montiel Olea, Jose Luis, and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics* 31:3 (2013), 358–369. 10.1080/00401706.2013.806694

Moreira, Marcelo J., Jack R. Porter, and Gustavo A. Suarez, "Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak," NBER working paper 302 (2004).

Nakamura, Emi, and Jón Steinsson, "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics* 133:3 (2018), 1283–1330. 10.1093/qje/qjy004

Rigobon, Roberto, "Identification through Heteroskedasticity," this REVIEW 85:4 (2003), 777–792.14585545

Rigobon, Roberto, and Brian Sack, "Measuring the Reaction of Monetary Policy to the Stock Market," *Quarterly Journal of Economics* 118:2 (2003), 639–669. 10.1162/003355303321675473

——— "The Impact of Monetary Policy on Asset Prices," *Journal of Monetary Economics* 51:8 (2004), 1553–1575. 10.1016/j.jmoneco.2004.02.004

Rigobon, Roberto, and Dani Rodrik, "Rule of Law, Democracy, Openness, and Income," *Economics of Transition* 13:3 (2005), 533–564. 10.1111/j.1468-0351.2005.00226.x

Sentana, Enrique, and Gabriele Fiorentini, "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics* 102:2 (2001), 143–164. 10.1016/S0304-4076(01)00051-3

Sims, Chris, "Macroeconomics and Reality," *Econometrica* 48:1 (1980), 1–48. 10.2307/1912017

Staiger, Doug, and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica* 65:3 (1997), 557–586. 10.2307/2171753

Staiger, Douglas O., James H. Stock, and Mark W. Watson, "How Precise Are Estimates of the Natural Rate of Unemployment?" (pp. 195–246), in Christina D. Romer and David H. Romer, eds., *Reducing Inflation: Motivation and Strategy* (Chicago: University of Chicago Press, 1997).

Stock, James H., and Jonathan H. Wright, "GMM with Weak Identification," *Econometrica* 68:5 (2000), 1055–1096. 10.1111/1468-0262.00151

Stock, James H., Jonathan H. Wright, and Motohiro Yogo, "A Survey of Weak Instruments and Weak Identification in Generalize Method of Moments," *Journal of Business and Economic Statistics* 29 (2002), 518–529. 10.1198/073500102288618658

Stock, James H., and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression" (pp. 80–108), in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg* (Cambridge: Cambridge University Press, 2005).

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48:4 (1980), 817–838. 10.2307/1912934

Wright, Jonathan H., "What Does Monetary Policy Do to Long-term Interest Rates at the Zero Lower Bound?," *Economic Journal* 122:564 (2012), 447–466. 10.1111/j.1468-0297.2012.02556.x

Zaefarian, Ghasem, Vita Kadile, Stephan C. Henneberg, and Alexander Leischnig, "Endogeneity Bias in Marketing Research: Problem, Causes and Remedies," *Industrial Marketing Management* 65 (2017), 39–46. 10.1016/j.indmarman.2017.05.006

Appendix

1. *Proof of Proposition 1*

Proof. The result owes to Rigobon (2003). Alternatively, Brunnermeier et al. (2021) show that the columns of H are the right eigenvectors of $\Sigma_{\eta,P}\Sigma_{\eta,C}^{-1}$, whose eigenvalues are the diagonal of $\Sigma_{\varepsilon,P}\Sigma_{\varepsilon,C}^{-1}$. Eigenvectors corresponding to nonrepeated eigenvalues (which implies nonproportionality) are uniquely determined.

2. *Proof of Proposition 2*

Proof. While the weak instruments literature models $\Pi = C/\sqrt{T}$, the device I adopt implies the more complicated expression:

$$\Pi = \frac{d_\sigma}{d_\sigma \tau_P^{-1} + \sqrt{T}\left(H_{21}^2\sigma_{\varepsilon 1}^2 + \sigma_{\varepsilon 2,C}^2\right)\left(\tau_P^{-1} + \tau_C^{-1}\right)}.$$

However, the asymptotic distribution of $\hat{H}_{12}$ is fundamentally unchanged. Under my local-to-unity device, $\sigma_{\varepsilon 2,P}^2 = \sigma_{\varepsilon 2,C}^2 + d_\sigma/\sqrt{T}$ so $\sigma_{\eta 2,P}^2 = \sigma_{\eta 2,C}^2 + d_\sigma/\sqrt{T}$ and $\sigma_{\eta 1\eta 2,P} = \sigma_{\eta 1\eta 2,C} + H_{12}d_\sigma/\sqrt{T}$. Asymptotically, the estimator in equation (6) yields

$$\hat{H}_{12} - H_{12} = \frac{\frac{1}{T_P}\sum_{t\in P}\eta_{1t}\eta_{2t} - \frac{1}{T_C}\sum_{t\in C}\eta_{1t}\eta_{2t}}{\frac{1}{T_P}\sum_{t\in P}\eta_{2t}^2 - \frac{1}{T_C}\sum_{t\in C}\eta_{2t}^2} - H_{12}$$

$$= \frac{\frac{1}{T_P}\sum_{t\in P}\left(\sigma_{\eta 1\eta 2,P} + \eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,P}\right) - \frac{1}{T_C}\sum_{t\in C}\left(\sigma_{\eta 1\eta 2,C} + \eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,C}\right)}{\frac{1}{T_P}\sum_{t\in P}\left(\sigma_{\eta 2,P}^2 + \eta_{2t}^2 - \sigma_{\eta 2,P}^2\right) - \frac{1}{T_C}\sum_{t\in C}\left(\sigma_{\eta 2,C}^2 + \eta_{2t}^2 - \sigma_{\eta 2,C}^2\right)} - H_{12}$$

$$= \frac{H_{12}\frac{d_\sigma}{\sqrt{T}} + \frac{1}{T_P}\sum_{t\in P}\left(\eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,P}\right) - \frac{1}{T_C}\sum_{t\in C}\left(\eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,C}\right)}{\frac{d_\sigma}{\sqrt{T}} + \frac{1}{T_P}\sum_{t\in P}\left(\eta_{2t}^2 - \sigma_{\eta 2,P}^2\right) - \frac{1}{T_C}\sum_{t\in C}\left(\eta_{2t}^2 - \sigma_{\eta 2,C}^2\right)} - H_{12}$$

$$= \frac{\frac{1}{\sqrt{T}}\left(H_{12}d_\sigma + \sqrt{\frac{T}{T_P}}\frac{1}{\sqrt{T_P}}\sum_{t\in P}\left(\eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,P}\right) - \sqrt{\frac{T}{T_C}}\frac{1}{\sqrt{T_C}}\sum_{t\in C}\left(\eta_{1t}\eta_{2t} - \sigma_{\eta 1\eta 2,C}\right)\right)}{\frac{1}{\sqrt{T}}\left(d_\sigma + \sqrt{\frac{T}{T_P}}\frac{1}{\sqrt{T_P}}\sum_{t\in P}\left(\eta_{2t}^2 - \sigma_{\eta 2,P}^2\right) - \sqrt{\frac{T}{T_c}}\frac{1}{\sqrt{T_c}}\sum_{t\in C}\left(\eta_{2t}^2 - \sigma_{\eta 2,C}^2\right)\right)} - H_{12}$$

$$\xrightarrow{d} \frac{H_{12}d_\sigma + z_{12,2} - z_{12,1}}{d_\sigma + z_{2,2} - z_{2,1}} - H_{12} = \frac{H_{12}d_\sigma + z_{12}}{d_\sigma + z_2} - H_{12}$$

$$= \frac{z_1}{d_\sigma + z_2},$$

where $(\, z_{12,2} \quad z_{12,1} \quad z_{2,2} \quad z_{2,1}\,)' \sim \mathcal{N}(0, V_z)$. Under assumptions 1 and 3, the convergence follows from a martingale central limit theorem for each of the summations, since η_t is assumed to be ergodic and stationary conditional on regime. In the last line, $z_{12} \equiv z_{12,2} - z_{12,1}$, $z_2 \equiv z_{2,2} - z_{2,1}$, and $z_1 \equiv z_{12} - H_{12}z_2$, with $(\, z_1 \quad z_2\,)' \sim \mathcal{N}(0, V_{weak})$.

3. *Proof of Proposition 3*

Proof. I model the variance deficiency as $\left(\sigma_{\varepsilon i,P}^2/\sigma_{\varepsilon i,C}^2\right)/\left(\sigma_{\varepsilon j,P}^2/\sigma_{\varepsilon j,C}^2\right) = 1 + d/\sqrt{T}$. The ith row of $\left[\, diag(\Sigma_{\varepsilon,C}) \quad diag(\Sigma_{\varepsilon,P})\,\right]$ is then equal to $\left[\, \sigma_{\varepsilon i,C}^2 \quad \sigma_{\varepsilon i,C}^2\left(\sigma_{\varepsilon j,P}^2/\sigma_{\varepsilon j,C}^2\right)\left(1 + d/T^{1/2}\right)\right]$. In the limit, this equals $\left[\, \sigma_{\varepsilon i,C}^2 \quad \sigma_{\varepsilon i,C}^2\left(\sigma_{\varepsilon j,P}^2/\sigma_{\varepsilon j,C}^2\right)\right]$. However, this expression is $\sigma_{\varepsilon i,C}^2/\sigma_{\varepsilon j,C}^2$ times the jth row, $\left[\, \sigma_{\varepsilon j,C}^2 \quad \sigma_{\varepsilon j,P}^2\,\right]$, so the condition of proposition 1 is violated.

4. Proof of Theorem 1

Proof. Define $\bar{\phi}(\theta, \eta_t) = \phi(\theta, \eta_t) - E(\phi(\theta, \eta_t))$, $q(\theta, \eta_t) = vec\left(\frac{\partial\phi(\theta,\eta_t)}{\partial\theta'}\right)$, $\bar{q}(\theta, \eta_t) = q(\theta, \eta_t) - E(q(\theta, \eta_t))$, as in Kleibergen (2005), with ϕ replacing his f. Lemma 1 in the supplement provides asymptotic distributions for $\bar{\phi}(\theta_0, \eta_t)$ and $\bar{q}(\theta_0, \eta_t)$, corresponding to assumption 1 of Kleibergen (2005). Lemma 2 in the online supplement establishes additional properties of the asymptotic variance, corresponding to assumption 2 of Kleibergen (2005). Theorem 1 of Kleibergen (2005) then immediately applies, and his definition 1 gives the limiting distribution of $K_T(\theta_0)$. Lemmas 1 and 2 also establish the required conditions of Stock and Wright's (2000) theorem 2, so $S_T(\theta_0) \xrightarrow{d} \chi^2_{n^2+n}$.

5. Proof of Theorem 2

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

6. Proof of Theorem 3

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 to column order; assumption 2 guarantees point identification. However, the columns of H^W are identified only up to an orthogonal rotation Q, $QQ' = Q'Q = I$. H_W contains at least two columns. If H^W contains exactly two columns, then Q is 2×2. Consider first a single fixed element of $H_{\cdot k}$, 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

$$\begin{bmatrix} 1 & H_{1m} \\ x & 1 \\ \vdots & \vdots \\ H_{nk} & H_{nm} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} 1 & \tilde{H}_{1m} \\ x & 1 \\ \vdots & \vdots \\ \tilde{H}_{nk} & \tilde{H}_{nm} \end{bmatrix}. \quad (13)$$

Placing $H_{\cdot k}$ and $H_{\cdot m}$ as the first and second columns, with the associated unit normalization, is without loss of generality, as identification is 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 $\{Q_{11} = (x^2 - 1)/(x^2 + 1), Q_{21} = 2x/(x^2 + 1)\}$. However, using an additional equation implied by equation (13), $Q_{11} + H_{1m}Q_{21} = 1$, rules out the second solution unless $H_{1m} = 1/x$. Generalizing this condition away from the case where $H_{\cdot k}$ and $H_{\cdot m}$ 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. If H is identified, then so too are $\sigma^2_{\varepsilon,C}$, $\sigma^2_{\varepsilon,P}$.

This argument extends to the case where the entirety of $H_{\cdot k}$ 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 noninvertible, which is false by assumption 1.1. Thus, the solution when a full column of H is specified is unique.