Federal Reserve Bank of New York
Staff Reports

# Robust Inference in Models Identified via Heteroskedasticity

Daniel J. Lewis

Staff Report No. 876
December 2018
Revised August 2019

**Robust Inference in Models Identified via Heteroskedasticity**
Daniel J. Lewis
*Federal Reserve Bank of New York Staff Reports*, no. 876
December 2018; revised August 2019
JEL classification: C12, C32, C36, 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, outline two tests to detect weak identification, and establish conditions for the validity of nonconservative methods for robust inference on an empirically relevant 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, Jòn Steinsson, two anonymous referees, and an anonymous associate editor for helpful feedback. 

# 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. The intuition dates from at least Fisher (1965). This identification scheme is most popular in macro-financial contexts, but has also been adopted in many other fields. 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. Both 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 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 using dummy variables for regimes. I propose a rule of thumb of $F > 23$ for the first-stage $F-$statistic. In the general model, Andrews' (2017) two-step procedure can be used. This test would be (perhaps prohibitively) conservative using robust sets for a parameter of interest computed using the only previously available option, projection inference. However, I establish conditions under which the asymptotic distributions of common subset test statistics ($S-$statistic of Stock & Wright (2000), $K-$statistic of Kleibergen (2005)) may be more tightly characterized by concentrating out nuisance parameters strongly identified *conditional* on the null hypothesis. This means inference can proceed using the familiar test statistic but potentially much smaller critical values. The resulting confidence sets can be used with Andrews' (2017) approach to detect weak identification and then to conduct inference if necessary.

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) (henceforth NS), 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 renders all coefficients found to be statistically significant using standard methods insignificant; 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 are not well approximated by a normal distribution. Additional simulations show that standard tests suffer from serious size distortions and projection methods are severely undersized, while 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 identification using these methods, much like is now best practice for IV following the work of Staiger & Stock (1997).

The paper is organized as follows. Section 2 presents the model, shows how weak identification arises, and demonstrates its effects on parameter estimates. Section 3 presents standard robust inference results, establishes conditions under which subset inference can proceed using reduced critical values, and outlines tests to detect weak identification. Section 4 applies the methods to the data of NS. Section 5 concludes. Proofs are in the Appendix.

# 2 Strong and weak identification via heteroskedasticity

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

## 2.1 Identification and when it might fail

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

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

[1]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 follows exclusively from the heteroskedasticity.

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 $\eta_t$. While I focus on the two regime case, most results can be directly extended to allow for additional regimes. For consistency with my empirical application to NS, I denote $C$ (control) and $P$ (policy), 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 provides 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. $E[\varepsilon_t \mid t \in r, \mathcal{F}_{t-1}] = 0$, $E[\varepsilon_t \varepsilon_t' \mid t \in r, \mathcal{F}_{t-1}] = \Sigma_{\varepsilon,r}$, $\mathcal{F}_{t-1} = \{\varepsilon_1, \ldots, \varepsilon_{t-1}\}$,
3. $\Sigma_{\varepsilon,r}$ is diagonal.

The first point imposes necessary assumptions on $H$. The second and third jointly state serial uncorrelatedness and orthogonality assumptions as well as stationarity of $\eta_t$ within regimes. Consequently, the covariance of $\eta_t$ for regime $r$ is

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

which is easily estimated. 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 $\eta_t$ yields $(n^2 + n)/2$ 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 $\Sigma_{\eta,C}$ and $\Sigma_{\eta,P}$ up to column order provided the rows of* $\begin{bmatrix} \sigma^2_{\varepsilon,C} & \sigma^2_{\varepsilon,P} \end{bmatrix}$ *are not proportional.*

Under additional assumptions, distinguishing the columns of $H$ , point identification holds. I adopt Assumption 2, a common choice to this effect:

**Assumption 2.** *The shock of interest experiences the largest relative change in variance; additional columns are labeled using external information or a statistical rule.*

The second part of the assumption is required to point-identify $H$ when $n > 2$.

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. The first panel plots hypothetical data for $\eta_t$, 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 impacting $\eta_t$ 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 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.

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

Figure 1: Distinguishing simultaneous responses using heteroskedasticity


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

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 if there are multiple dimensions of monetary policy shocks. The closer the comovement of variances, 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 $\Sigma_{\varepsilon,P}$, the equations (2) can be written as a combined set of moments (as in Rigobon (2003)):

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

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

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

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 (on which most weak identification results are based). 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 $S_T(\theta) \equiv S_T(\theta;\theta)$. To characterize the asymptotic distribution of GMM estimates, regularity conditions such as those of Assumption 3 are required:

**Assumption 3.** *Assume*

1. *The process $\eta_t$ is ergodic and stationary within regimes,*
2. *$E\left[vech\left(\eta_t\eta_t'\right)vech\left(\eta_t\eta_t'\right)' \mid t \in r\right] < \infty$ for $r \in \{P, C\}$,*
3. *$T_r/T = \tau_r > 0$, for $T_r = |\{t : t \in r\}|, r \in \{P, C\}$.*
4. *$\Theta$ 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 $\theta_0$ is the unique solution to (3), standard arguments show that the GMM estimates of $\theta$ 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 (3) – that is in doubt. I now consider formally how that condition may fail and the consequences when it does.

## 2.2 The asymptotic distribution under weak identification 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, Rigobon & Sack (2004), Hébert & Schreger (2017), 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*". Without loss of generality, I assume that $\sigma^2_{\varepsilon_1,r} \equiv \sigma^2_{\varepsilon_1}$ is constant; $H_{12}$ is the parameter of interest. It measures the impact of a unit structural shock (say a policy shock) on $\eta_{1t}$. In NS, $H_{12}$ represents the impact of monetary policy shocks on Treasury forward rates.

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

$$\frac{\sigma_{\eta_1\eta_2,P} - \sigma_{\eta_1\eta_2,C}}{\sigma^2_{\eta_2,P} - \sigma^2_{\eta_2,C}} = \frac{H_{12}\left(\sigma^2_{\varepsilon_2,P} - \sigma^2_{\varepsilon_2,C}\right) + H_{21}\left(\sigma^2_{\varepsilon_1,P} - \sigma^2_{\varepsilon_1,C}\right)}{\left(\sigma^2_{\varepsilon_2,P} - \sigma^2_{\varepsilon_2,C}\right) + H^2_{21}\left(\sigma^2_{\varepsilon_1,P} - \sigma^2_{\varepsilon_1,C}\right)} = \frac{H_{12}\Delta\left(\sigma^2_{\varepsilon_2}\right)}{\Delta\left(\sigma^2_{\varepsilon_2}\right)} = H_{12}, \tag{5}$$

where $\sigma^2_{\eta_2,r}$, $\sigma_{\eta_1\eta_2,r}$ denote elements of $\Sigma_{\eta,r}$ and the $\Delta(\cdot)$ operator takes the difference in the argument between regimes. If in fact $\sigma^2_{\varepsilon_1,P} \neq \sigma^2_{\varepsilon_1,C}$ fails, then $H_{12}$ will be misidentified, since the $\Delta\left(\sigma^2_{\varepsilon_1}\right)$ terms will not vanish.

I now move from $H_{12}$, identified in population, to estimators, $\hat{H}_{12}$. The sample analogues from the left-hand-side of (5) can be simply estimated. This is equivalent to an instrumental variables problem (Rigobon & Sack (2004)):

$$\begin{aligned}\hat{H}_{12} = \frac{\Delta\left(\hat{\sigma}_{\eta_1\eta_2}\right)}{\Delta\left(\hat{\sigma}^2_{\eta_2}\right)} &= \frac{\frac{T}{T_P}\sum_{t\in P}\eta_{1t}\eta_{2t} - \frac{T}{T_C}\sum_{t\in C}\eta_{1t}\eta_{2t}}{\frac{T}{T_2}\sum_{t\in P}\eta^2_{2t} - \frac{T}{T_C}\sum_{t\in C}\eta^2_{2t}} \\ &= \frac{\sum_{t=1}^T \eta_{1t}Z_t}{\sum_{t=1}^T \eta_{2t}Z_t},\end{aligned} \tag{6}$$

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 TSLS, as suggested in Rigobon & Sack (2004), using

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

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

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

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. Moreover, Slutsky's theorem shows that, provided the denominator does not converge to zero, 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 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 zero, standard inference methods are not reliable in the familiar IV setting (e.g., Staiger & Stock (1997)). As the first stage coefficient, $\Pi$, tends to zero, the instrument provides less information about the endogenous regressor. Here, $\Pi$ goes to zero as $\sigma^2_{\varepsilon_2,P}$ approaches $\sigma^2_{\varepsilon_2,C}$, the case of no variance change.

If $\sigma^2_{\varepsilon_2,P} = \sigma^2_{\varepsilon_2,C}\left(\Pi = 0\right)$, so $H_{12}$ is unidentified, then the denominator (and numerator) of (6) converges in probability to zero. To obtain a limit distribution, multiplying (6) by $\frac{\sqrt{T}}{\sqrt{T}}$ leads both numerator and denominator to converge in distribution to mean-zero 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 (6) is non-uniform with respect to $\left(\sigma^2_{\varepsilon_2,P}, \sigma^2_{\varepsilon_2,C}\right)$: if $\sigma^2_{\varepsilon_2,P} \neq \sigma^2_{\varepsilon_2,C}$

it 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,

$$\frac{\sigma^2_{\varepsilon_2,P}}{\sigma^2_{\varepsilon_2,C}} = 1 + \frac{d}{\sqrt{T}}. \tag{10}$$

Rearranging yields

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

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 zero nor one, 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 & 3, $\hat{H}_{12}$ is not consistent for $H_{12}$; rather,*

$$\hat{H}_{12} - H_{12} \xrightarrow{d} \frac{z_1}{d_{\sigma^2} + z_2}, \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(0, V_{weak}\right), \tag{11}$$

*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 & 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, & 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 must be used.

### 2.3 Weak identification in the general case

Proposition 1 states the conditions for global identification, which can break down in the two related cases outlined above. As in detail in the *simple case*, I model the relationship between the variances of two shocks, $i$ and $j$, as local-to-unity:

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

where $d$ is finite. In economic terms, the Great Moderation or Financial Crisis were 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 (12) and Assumption 1, $H$ is asymptotically unidentified.*

Intuitively, under the local-to-unity device, the non-proportionality 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 $\frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1$ from (infeasible) observations of $\varepsilon_t$ 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.

To conclude this section, I demonstrate via simulations just how poor of an approximation standard asymptotic results provide. I calibrate my simulations to NS's specification using daily changes in 2-year Treasury forward rates as a dependent variable and daily changes in 2-year nominal Treasury yields as the policy series. Additional details can be found in the empirical application below and in Section B of the Supplement. I vary $T$ from 400 to 1600 (empirical size is approximately 800) and vary the empirical degree of identification ($\delta = \left(\sigma^2_{\varepsilon_{2,P}}/\sigma^2_{\varepsilon_{2,C}}\right) / \left(\sigma^2_{\varepsilon_{1,P}}/\sigma^2_{\varepsilon_{1,C}}\right) - 1$) by a factor of 10 in each direction.

Figure 2 presents histograms of the $t-$ratio, $\frac{\hat{H}_{12}-H_{12}}{se\left(\hat{H}_{12}\right)}$, 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, 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.

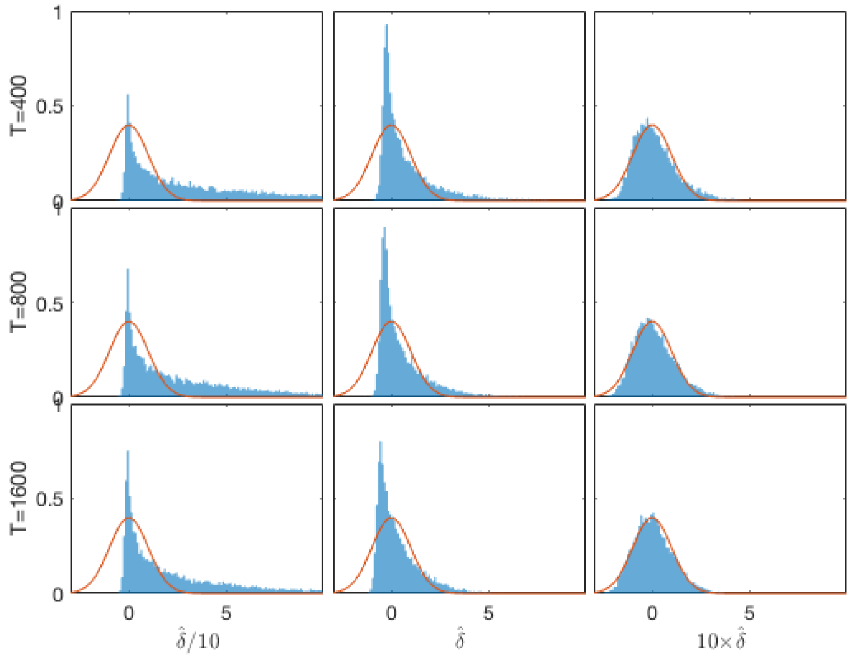
# 3 Weak identification robust inference

In this section, I present results for robust inference in GMM settings, before presenting a new result allowing non-conservative inference on empirically interesting subsets of $\theta$. I then outline how existing tests for weak identification may be used.

## 3.1 Parameter inference

The asymptotic behaviour of GMM estimators, robust to weak identification, is 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 $\chi^2$ distribution. Many refinements have since been developed, including the "$K-$statistic"

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 (14) in the Supplement. Point estimation proceeds via Brunnermeier et al's (2017) eigenvector method, with inference using this solution for efficient GMM.

of Kleibergen (2005), which is efficient under strong identification. Most 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 can be “concentrated out”. However, the parameter(s) of interest in applied work is often some generic subset of the parameter vector; this is the case in Rigobon & Sack (2003) (response of 3-month Treasury rate to S&P 500 shocks), Rigobon & 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 & Schreger (2017) (response of equities and exchange rates to sovereign default shocks), and NS (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.

I state 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. With additional regimes, the $K-$statistic will be asymptotically efficient under strong identification. Further refinements may have better power properties in over-identified models (e.g., the Conditional Linear Combination tests of Andrews (2016)).

**Full vector inference**

Under the assumptions presented in Section 2, Theorem 1 shows that inference on the full parameter vector can proceed using the $K-$statistic.

**Theorem 1.** *Under Assumptions 1 & 3,*

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

*where $K_T(\theta)$ is Kleibergen's $K-$statistic.*

As discussed below, Magnusson & Mavroeidis (2014) already consider this test (their *split-KLM*) for identification via heteroskedasticity. Simulations in the supplement calibrated to NS's data show the test performs well, while the size of standard methods exceeds 70%.

**Inference for subsets of the parameter vector**

Projection methods constitute the only previous 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 e.g., Dufour & Taamouti (2005), Chaudhuri & Zivot (2011)).

However, Kleibergen (2005) provides a refinement over Theorem 1 for tests on certain subsets of the parameter vector. Partition $\theta$ into the parameter(s) of interest, $\beta$, and the nuisance parameters, $\alpha$. If the rank of the asymptotic Jacobian conditional on $\beta$ is equal to the dimension of $\alpha$, $\alpha$ is strongly identified *conditional on* $\beta$ (but may be unconditionally weakly identified), so inference may use degrees of freedom equal to the dimension of $\beta$ (Kleibergen (2005), Theorem 2). The elements of $\beta$ may be either weakly or strongly identified. Assumption 4 and Theorem 2 state this result formally.

**Assumption 4.** *Conditional on $\beta$, $\alpha$ is asymptotically strongly identified.*

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.** *Under Assumptions 1 & 3, if Assumption 4 additionally holds, then*

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

*where $p_{int}$ is the dimension of $\beta$.*

The degrees of freedom of the limiting distribution for the full parameter vector (or projection tests) is lowered from $n^2 + n$ to $p_{int}$. 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 4? While existing work concentrates out parameters that are unconditionally strongly identified, I crucially exploit the fact that Kleibergen's result allows parameters that are only *conditionally* identified to be concentrated out. 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 $\begin{pmatrix} \sigma^2_{\varepsilon_k,C} & \sigma^2_{\varepsilon_k,P} \end{pmatrix}$ 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 unconditionally identified from (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 (e.g., Rigobon & Sack (2003, 2004), Wright (2012), Hébert & Schreger (2017), and NS). Theorem 3 shows that if such parameters are in $H_W$, conditioning on them may render the remainder of the model strongly identified.

**Theorem 3.** *Under Assumptions 1 & 2, 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 (which fixes $H_{lk}$ or $H^{(k)}$), I obtain conditional (strong) identification for the remainder of $H_W$. This means that a system of equations satisfying the conditions of Theorem 3 meets Assumption 4, so Theorem 2 applies. 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 sub-block of $H$.

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 4 is satisfied for tests where $\beta$ contains the corresponding column of $H$ or a single element of the column (and possibly additional parameters), and Theorem 2 holds.*

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 is not too prevalent, since the nuisance subset, $\alpha$, is conditionally strongly identified. Five remarks clarify the impact of Condition 1.

*Remark* 1. Condition 1 nests the cases where $\beta = \theta$ or where $\beta$ 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 $H$ adds no new information.

*Remark* 3. Given at most two variances may evolve proportionally, a researcher should minimize the dimension of $\eta_t$ subject to the constraint that $\eta_t$ spans $\varepsilon_t$ (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 non-binding.

*Remark* 5. It is straightforward to extend the results of Theorem 3 to IRFs. For a detailed discussion, see Section C of the Supplement.

Additional simulations in the supplement show that the reduced test is very well-sized, while a standard $t-$test delivers size distortions in the 10% range and projection tests have size of approximately zero.

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 & Mavroeidis (2014). They propose a variety of tests that accommodate the present setting. They only discuss subset inference in the context of concentrating out a strongly identified nuisance parameter, $\zeta$, besides $\theta$; they show that in this setting the asymptotic distributions of the *split-S* and *split-KLM* test statistics are unaffected by estimating $\zeta$, retaining degrees of freedom equal to the number of weakly identified parameters, $\theta$ (Theorems 6 and 7). This treatment requires all nuisance parameters, $\zeta$, to be strongly identified (unconditionally). The subset result I offer extends these results, by exploiting the fact that Kleibergen's results require nuisance parameters only be conditionally strongly identified (while perhaps unconditionally weakly identified) and establishing conditions under which this is the case for relevant $\beta$. Heuristically, in Magnusson & Mavroeidis' notation, this allows $\zeta$ to be a *subset* of the weakly identified $\theta$ (not just additional strongly identified parameters), such that their test statistics are now distributed $\chi^2_{p-p_\varsigma}$ instead of $\chi^2_p$.

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, & 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 asymptotically coincides with an $S-$test.

## 3.2 Tests for weak identification

It is desirable to assess the strength of identification ex ante in order to determine whether standard inference techniques will be reliable, or whether robust methods must be adopted. In the IV literature, the first-stage $F-$test of Staiger & Stock (1997) and Stock & Yogo (2005) is now ubiquitous. I address this problem for two cases: first, the *simple case* with a single variance change, and second, the fully general model.

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. This means that a first-stage $F-$test approach can be adopted. However, the critical values of Stock & Yogo (2005) are only valid under homoskedasticity. Fortunately, Montiel Olea & Pflueger (2013) develop alternative critical values under weaker assumptions.[2] They allow for arbitrary heteroskedasticity and autocorrelation and calibrate critical values to the Nagar bias of TSLS 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

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 for the first-stage $F-$statistic from Montiel Olea & Pflueger (2013). For a given critical value for the $F-$statistic, bias is greater than that indicated in 5% of repeated samples. Assumptions underlying these results are enumerated in the text.

IV literature. This test can be easily adopted in this restricted framework.

In the fully general model, the only option is the Andrews (2017) two-step approach. This test can be applied to a subset of the parameter vector. First, a researcher decides on a maximal size distortion that she believes is compatible with strong identification, say $\xi = 0.1$. Then, a preliminary robust confidence set is constructed to have size $1 - \nu - \xi$, where $\nu$ 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 passes the pre-specified 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 (2017). Non-conservative reduced tests based on Theorem 3 are particularly valuable here. Given how conservative projection methods are, 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.

# 4 Empirical application

I demonstrate the use of the robust inference methods by studying the identification

[2] In particular, in the TSLS framework, they assume $\frac{1}{\sqrt{T}}\begin{pmatrix} \sum_{t=1}^{T} Z_t v_t & \sum_{t=1}^{T} Z_t u_t \end{pmatrix}'$ is asymptotically normal with consistently estimable positive definite covariance, the covariance of $\begin{pmatrix} v_t & u_t \end{pmatrix}'$ is positive definite and consistently estimable, and a local-to-zero representation for $\Pi$. See Montiel Olea & Pflueger (2013) for additional details.

of monetary policy shocks in the setting of NS.[3] 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 the daily Treasury yields and the authors' 30-minute window "policy news" series as the policy instrument, with either nominal 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.

## 4.1 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 focuses on 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.

I first test formally for weak identification using the methods proposed in Section 3.2. Under NS's restricted model, the first-stage $F-$statistic tests for weak identification. These

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

results are reported in the first panel of Table 2. 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. Andrews' (2017) test for the general 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, see Table 3). These test results corroborate the less formal observations of NS, who suspect weaker identification in daily data.

Table 3 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" shock, the results for $H_{12}$, the pass-through of policy shocks to forwards (1.07 and 0.97) are extremely close to NS's restricted model (1.10 and 0.96), indicating a forward guidance/news effect that shifts expectations. Their assumption that $\sigma_1^2$ is fixed has little impact on estimates of $H_{12}$ because $H_{21}$ is near-zero, minimizing the possible bias in (5). Using daily changes in the nominal yield as the policy series, the point estimates for the real forward rate are in keeping with the intraday results and those of NS (who estimate 0.97 for $H_{12}$). In contrast, the *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. The positive value for $H_{21}$ is compatible with there being a second meaningful dimension of monetary policy news, as opposed to NS's "background noise" interpretation of the second shock. However, these results remain at odds with the strongly identified specifications.

### 4.2 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 ro-

Table 2: 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 1. 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' $\gamma_{min}$, the value at which the $1-\alpha-\gamma_{min}$ robust set is just contained by the strong identification set.

bust 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 anomalous estimate of $\hat{H}_{21} = 0.70$ for the nominal forward rate specification is highly statistically significant using standard methods, but this effect vanishes using the robust interval. Additionally, for the real forward rate specification, $\hat{H}_{12} = 0.95$ is highly statistically significant using standard methods, but not using robust intervals. 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.[4] For $H_{21}$, I obtain (largely) precisely estimated zeros. These conclusions replicate NS's findings for the restricted model.

For the specifications using 30-minute shocks, I can also test the null hypothesis that the non-policy shock variance is fixed across regimes, adopting standard inference methods based on the evidence of strong identification. This is the over-identifying assumption used by NS 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, there is not strong enough evidence of equality to justify 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 the simulations conducted in this paper and the Supplement.

[4]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 3: Estimates

| policy inst. | one-day yield | | 30 min. news | |
|---|---|---|---|---|
| dep. var. | Nominal fwd. | Real fwd. | Nominal fwd. | Real fwd. |
| $H_{21}$ | $0.70$ | $-0.01$ | $0.01$ | $0.00$ |
| 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 NS'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. Stars indicate significance from zero at the 5 or 1% levels based on the more conservative of the two tests.

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 can be found in Section C of the Supplement.

# 5 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 impact the reliability of standard inference methods in empirical data. I propose tests to detect weak identification, allowing researchers to determine whether they ought to 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. Such tests provide the first option for robust inference in this context that is not prohibitively conservative. Given the problem posed by robust subset inference in nonlinear models, the approach of focusing on deriving conditional identification results suggests an outline for those interested in other models.

I apply these methods to the identification of monetary policy shocks, as in NS. 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 the growing literature exploiting identification via heteroskedasticity, but the results presented in this paper enable researchers to do so.

# 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 $M$ |
| $vech\left(M\right)$ | denotes the unique vectorization of matrix $M$ |

## Proofs

### Proof of Proposition 1.

*Proof.* The result owes to Rigobon (2003). Alternatively, Brunnermeier et al (2017) show 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 of non-repeated eigenvalues (which implies non-proportionality) are uniquely determined. $\square$

**Proof of Proposition 2.**

*Proof.* While the weak instruments literature models $\Pi = \frac{C}{\sqrt{T}}$, the device I adopt implies $\Pi = \frac{d_{\sigma^2}}{d_{\sigma^2}+\sqrt{T}\sigma^2_{\varepsilon_2,C}}$. However, the asymptotic distribution of $\hat{H}_{12}$ is fundamentally unchanged. Under my local-to-unity device, $\sigma^2_{\varepsilon_2,P} = \sigma^2_{\varepsilon_2,C} + \frac{d_{\sigma^2}}{\sqrt{T}}$ so $\sigma^2_{\eta_2,P} = \sigma^2_{\eta_2,C} + \frac{d_{\sigma^2}}{\sqrt{T}}$ and $\sigma_{\eta_1\eta_2,P} = \sigma_{\eta_1\eta_2,C} + H_{12}\frac{d_{\sigma^2}}{\sqrt{T}}$. Asymptotically, the estimator in (6) yields

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

where $\begin{pmatrix} z_1 & z_2 \end{pmatrix}' \sim \mathcal{N}\left(0, V_{weak}\right)$. The convergence follows from a martingale central limit theorem for each of the summations, since $\eta_t$ 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 $z_1 \equiv z_{12} - H_{12}z_2$. □

**Proof of 3.**

*Proof.* I model the variance deficiency as $\frac{\sigma^2_{\varepsilon_{i,P}}/\sigma^2_{\varepsilon_{i,C}}}{\sigma^2_{\varepsilon_{j,P}}/\sigma^2_{\varepsilon_{j,C}}} = 1 + \frac{d}{\sqrt{T}}$. The $i^{th}$ row of $\begin{bmatrix} diag\left(\Sigma_{\varepsilon,C}\right) & diag\left(\Sigma_{\varepsilon,P}\right) \end{bmatrix}$ is equal to $\begin{bmatrix} \sigma^2_{\varepsilon_i,C} & \sigma^2_{\varepsilon_i,C}\frac{\sigma^2_{\varepsilon_j,P}}{\sigma^2_{\varepsilon_j,C}} & \left(1 + d/T^{1/2}\right)\end{bmatrix}$. In the limit, this equals $\begin{bmatrix} \sigma^2_{\varepsilon_i,C} & \sigma^2_{\varepsilon_i,C}\frac{\sigma^2_{\varepsilon_j,P}}{\sigma^2_{\varepsilon_j,C}} \end{bmatrix}$. However, this expression is $\frac{\sigma^2_{\varepsilon_i C}}{\sigma^2_{\varepsilon_j C}}$ times the $j^{th}$ row, $\begin{bmatrix} \sigma^2_{\varepsilon_j,C} & \sigma^2_{\varepsilon_j,P} \end{bmatrix}$, so the condition of Proposition 1 is violated. □

**Proof of 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 $\phi$ 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 & 3,* $\psi_T(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\bar{\phi}(\theta_0, \eta_t)\\ \bar{q}(\theta_0, \eta_t)\end{pmatrix} \xrightarrow{d} \begin{pmatrix}\psi_\phi\\ \psi_{\theta_0}\end{pmatrix}$ *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* $q_T(\theta) = \sum_{t=1}^{T} q(\theta, \eta_t)$ .

*Proof.* First, note that each block of $\bar{\phi}(\theta_0, \eta_t)$ 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 $\phi(\theta_0, \eta_t)$, denoted $\phi^r(\theta_0, \eta_t)$, takes the form

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

Then

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

by Assumption 1.2. Finally, $E|\phi(\theta, \eta_t)| < \infty$ by Assumption 3.2, so $\phi^r(\theta_0)$ is a martingale difference sequence. This means that, stacking the blocks, $\phi(\theta, \eta_t)$ is a martingale difference sequence. By Billingsley's (1961) Ergodic Stationary Martingale Differences CLT, given

Assumption 3.1,

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

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.

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 $q\left(\cdot\right)$ contains only parameters and no data (the moment equations are separable in data and parameters). This is true for any $\theta\in\Theta$; $\theta$ need not equal $\theta_0$. Thus $\psi_\theta$ 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 $E\left[\bar{\phi}\left(\theta,\eta_t\right)\bar{\phi}\left(\theta,\eta_t\right)'\right]$. □

Lemma 2 establishes additional properties of the asymptotic variance.

**Lemma 2.** *Under Assumptions 1 & 3, 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* $\frac{\partial vec\left(\hat{V}_{\phi\phi}(\theta_0)\right)}{\partial\theta'}\xrightarrow{p}\frac{\partial vec\left(V_{\phi\phi}(\theta_0)\right)}{\partial\theta'}$.

*Proof.* By the Ergodic Theorem (e.g., Karlin & Taylor (1975), Theorem 9.5.5) and Assumption 3, 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.

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 $\theta$; it is 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 $\frac{1}{T}\sum_{t=1}^{T}\phi\left(\theta_0,\eta_t\right)$ is consistent for $E\left[\phi\left(\theta_0,\eta_t\right)'\right]$, $\frac{\partial V_{\phi\phi}(\theta_0)}{\partial\theta'}$ is consistently estimated by Slutsky's Theorem.

□

Theorem 1 then follows directly from Kleibergen (2005). Lemmata 1 and 2 establish Assumptions 1 and 2 from that paper, under which his Theorem 1 holds. They also establish the required conditions of Stock & Wright (2001) Theorem 2 so $S_T\left(\theta_0\right)\xrightarrow{d}\chi^2_{n^2+n}$. □

**Proof of Theorem 2.**

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

**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 two columns, then $Q$ is $2\times 2$. 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

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

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 (13), $Q_{11} + H_{1m}Q_{21} = 1$, 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.

□

# References


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

——— (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 100, 337–348.

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

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2018): "Feedbacks, Financial Markets and Economic Activity," mansucript.

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.

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.

Fisher, F. (1965): "Near-Identifiability and the Variances of the Disturbance Terms," *Econometrica*, 33, 409–419.

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

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

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

Lewis, D. J. (2019): "Identification via Time-Varying Volatility," *Federal Reserve Bank of New York Staff Reports*.

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

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

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 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.

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