

Supplementary Appendix to the paper
Maximum Likelihood Inference in
Weakly Identified DSGE Models.

(proofs intended for web-posting)

By Isaiah Andrews¹ and Anna Mikusheva²

Abstract

This Supplementary Appendix contains details of the examples and proofs of some results given in the paper “Maximum Likelihood Inference in Weakly Identified Models,” by Isaiah Andrews and Anna Mikusheva. We also provide several additional examples illustrating ways in which weak identification can arise in a DSGE context.

S1 Stylized DSGE model from Section 2

S1.1 Solving the model

Here we solve the restricted linear rational expectations system:

$$\begin{cases} bE_t\pi_{t+1} + \kappa x_t - \pi_t = 0, \\ -[r_t - E_t\pi_{t+1} - \rho\Delta a_t] + E_t x_{t+1} - x_t = 0, \\ \frac{1}{b}\pi_t + u_t = r_t, \end{cases} \quad (\text{S1})$$

where x_t and π_t are observed endogenous variables. Exogenous shocks a_t and u_t evolve according to the system:

$$\begin{aligned} \Delta a_t &= \rho\Delta a_{t-1} + \varepsilon_{a,t}; & u_t &= \delta u_{t-1} + \varepsilon_{u,t}; \\ (\varepsilon_{a,t}, \varepsilon_{u,t})' &\sim iid N(0, \Sigma); & \Sigma &= \text{diag}(\sigma_a^2, \sigma_u^2). \end{aligned}$$

¹Department of Economics, M.I.T., Building E18, Cambridge, MA, 02142. Email: ian-drews@mit.edu. Support from the Ford Foundation and the NSF Graduate Research Fellowship under Grant No. 1122374 is gratefully acknowledged.

²Department of Economics, M.I.T., Building E18, Cambridge, MA, 02142. Email: amikushe@mit.edu. Financial support from the Castle-Krob Career Development Chair and the Sloan Research Fellowship is gratefully acknowledged.

To solve the system we substitute out r_t in the first two equations of (S1) and obtain the following system:

$$\begin{cases} bE_t\pi_{t+1} = -\kappa x_t + \pi_t, \\ E_t\pi_{t+1} + E_tx_{t+1} = x_t + \frac{1}{b}\pi_t + u_t - \rho\Delta a_t. \end{cases}$$

We solve for E_tx_{t+1} and get the expectation equation:

$$bE_tx_{t+1} = (b + \kappa)x_t + bu_t - b\rho\Delta a_t,$$

which we can rewrite as:

$$x_t = \frac{b}{b + \kappa}E_tx_{t+1} - \frac{b}{b + \kappa}u_t + \frac{b\rho}{b + \kappa}\Delta a_t.$$

Now we solve this expectation equation by iterating forward:

$$x_t = \sum_{j=0}^{\infty} \left(\frac{b}{b + \kappa}\right)^j E_t \left[-\frac{b}{b + \kappa}u_{t+j} + \frac{b\rho}{b + \kappa}\Delta a_{t+j} \right].$$

We notice that $E_tu_{t+j} = \delta^j u_t$ and $E_t\Delta a_{t+j} = \rho^j \Delta a_t$. As a result, we have:

$$\begin{aligned} x_t &= -\frac{b}{b + \kappa} \cdot \frac{1}{1 - \delta\frac{b}{b + \kappa}} u_t + \frac{b\rho}{b + \kappa} \cdot \frac{1}{1 - \rho\frac{b}{b + \kappa}} \Delta a_t = \\ &= -\frac{b}{b + \kappa - \delta b} u_t + \frac{b\rho}{b + \kappa - b\rho} \Delta a_t. \end{aligned}$$

We plug the last expression into the Euler equation and solve the resulting expectation equation for π_t :

$$\begin{aligned} \pi_t &= bE_t\pi_{t+1} + \kappa x_t = \\ &= bE_t\pi_{t+1} - \frac{b\kappa}{b + \kappa - \delta b} u_t + \frac{b\rho\kappa}{b + \kappa - b\rho} \Delta a_t = \\ &= \sum_{j=0}^{\infty} b^j E_t \left[-\frac{b\kappa}{b + \kappa - \delta b} u_{t+j} + \frac{b\rho\kappa}{b + \kappa - b\rho} \Delta a_{t+j} \right] = \\ &= -\frac{b\kappa}{(b + \kappa - \delta b)(1 - \delta b)} u_t + \frac{b\rho\kappa}{(b + \kappa - b\rho)(1 - b\rho)} \Delta a_t. \end{aligned}$$

Finally we obtain the following solution to the system (S1):

$$\begin{cases} x_t = -\frac{b}{b + \kappa - \delta b} u_t + \frac{b}{b + \kappa - b\rho} \rho \Delta a_t; \\ \pi_t = -\frac{b\kappa}{(b + \kappa - \delta b)(1 - \delta b)} u_t + \frac{b\kappa}{(b + \kappa - b\rho)(1 - b\rho)} \rho \Delta a_t. \end{cases} \quad (\text{S2})$$

S1.2 Identification of the model

In this subsection we check identification of the model (S1). We use the explicit solution written in equation (S2). Assume that $\sigma_a^2 > 0, \sigma_u^2 > 0, 0 < \delta, \rho, b < 1$ and $\kappa > 0$.

First we show that the model is point identified if $\delta < \rho$. Let $A_1(\theta) = -\frac{b}{b+\kappa-\delta b}$ and $A_2(\theta) = \frac{b}{b+\kappa-\rho b}$. We have

$$x_t = A_1(\theta)u_t + A_2(\theta)\rho\Delta a_t,$$

and

$$\pi_t = \frac{\kappa}{1-\delta b}A_1(\theta)u_t + \frac{\kappa}{1-\rho b}A_2(\theta)\rho\Delta a_t.$$

We can identify auto-covariances of all orders for the series x_t and π_t as well as all cross-covariances. In particular, we have

$$\begin{aligned} Var(x_t) &= A_1(\theta)^2 \frac{\sigma_u^2}{1-\delta^2} + A_2(\theta)^2 \rho^2 \frac{\sigma_a^2}{1-\rho^2}; \\ cov(x_t, x_{t-k}) &= A_1(\theta)^2 \frac{\sigma_u^2 \delta^k}{1-\delta^2} + A_2(\theta)^2 \rho^2 \frac{\sigma_a^2 \rho^k}{1-\rho^2}. \end{aligned}$$

It is easy to see that from the auto-covariance structure of process x_t one can identify $\delta < \rho, A_1(\theta)^2 \sigma_u^2$ and $A_2(\theta)^2 \sigma_a^2$. We also have the following expressions for the cross-covariances:

$$\begin{aligned} cov(x_t, \pi_t) &= A_1(\theta)^2 \frac{\sigma_u^2}{1-\delta^2} \frac{\kappa}{1-\delta b} + A_2(\theta)^2 \rho^2 \frac{\sigma_a^2}{1-\rho^2} \frac{\kappa}{1-\rho b}; \\ cov(x_t, \pi_{t-k}) &= A_1(\theta)^2 \frac{\sigma_u^2 \delta^k}{1-\delta^2} \frac{\kappa}{1-\delta b} + A_2(\theta)^2 \rho^2 \frac{\sigma_a^2 \rho^k}{1-\rho^2} \frac{\kappa}{1-\rho b} \end{aligned}$$

From cross-covariances we can additionally identify $A_1(\theta)^2 \sigma_u^2 \frac{\kappa}{1-\delta b}$ and $A_2(\theta)^2 \sigma_a^2 \frac{\kappa}{1-\rho b}$.

To sum up, the auto-covariance structure of the process x_t, π_t allows us to identify the following six quantities:

$$\delta, \rho, A_1(\theta)^2 \sigma_u^2, A_2(\theta)^2 \sigma_a^2, A_1(\theta)^2 \sigma_u^2 \frac{\kappa}{1-\delta b}, A_2(\theta)^2 \sigma_a^2 \frac{\kappa}{1-\rho b}$$

We can see from the last four quantities that $\frac{\kappa}{1-\delta b}$ and $\frac{\kappa}{1-\rho b}$ are identified, and thus $\frac{1-\rho b}{1-\delta b}$ is identified. Since ρ and δ are identified, we see that b is identified as well. This

implies that κ is also identified. Finally we notice that the $A_i(\theta)$ are functions of only b, κ, ρ and δ , and thus are identified. Looking at these six quantities, we can see that they imply identification of σ_u^2 and σ_a^2 .

Now we examine the identification in the case $\delta = \rho$. If $\delta = \rho$ we have that x_t and π_t satisfy the following system:

$$\begin{cases} x_t = \frac{b}{b+\kappa-\delta b}(\rho\Delta a_t - u_t); \\ \pi_t = \frac{b\kappa}{(b+\kappa-\delta b)(1-\delta b)}(\rho\Delta a_t - u_t) = \frac{\kappa}{1-\delta b}x_t \end{cases}$$

x_t and π_t are linearly dependent AR(1) processes with AR root $\delta = \rho$. The only functionally independent quantities that can be identified are the autoregressive parameter ($\delta = \rho$), the variance of x_t , and the ratio x_t/π_t . Hence we can only identify four quantities:

$$\delta = \rho, \quad \frac{b}{b+\kappa-\delta b}\sqrt{\rho^2\sigma_a^2 + \sigma_u^2}, \quad \frac{\kappa}{1-\delta b},$$

but we have six structural parameters. As a result, there are two degrees of under-identification.

S1.3 Assumption 1

We have that

$$Y_t = \begin{pmatrix} x_t \\ \pi_t \end{pmatrix} = C(\theta) \begin{pmatrix} u_t \\ \Delta a_t \end{pmatrix} = C(\theta)U_t,$$

and

$$U_t = \Lambda U_{t-1} + \varepsilon_t; \quad \Lambda = \begin{pmatrix} \delta & 0 \\ 0 & \rho \end{pmatrix} \text{ and } \varepsilon_t \sim N(0, \Sigma).$$

We can write the likelihood function:

$$\begin{aligned} \ell_T(\theta) = \text{const} - \frac{1}{2} \sum_{t=1}^T (C^{-1}(\theta)Y_t - \Lambda C^{-1}(\theta)Y_{t-1})' \Sigma^{-1} (C^{-1}(\theta)Y_t - \Lambda C^{-1}(\theta)Y_{t-1}) - \\ - \frac{T}{2} \log |\Sigma| - T \log |C(\theta)|. \end{aligned}$$

We derive the score for a similar likelihood in Section S3. Here we just note that the score at the true parameter value is a linear combination of terms $(\varepsilon_t \varepsilon_t' - \Sigma)$ and $\varepsilon_t Y_{t-1}'$. It thus trivially satisfies Assumption 1 from the paper for sequences of models with $\rho = \delta + \frac{C}{\sqrt{T}}$.

S2 Example 1: ARMA(1,1) with nearly canceling roots.

This section contains the details of Example 1 from the paper. Below we use the formulation of the weak ARMA(1,1) model from Andrews and Cheng (2012).

$$Y_t = (\pi + \beta)Y_{t-1} + e_t - \pi e_{t-1}, \quad e_t \sim i.i.d.N(0, 1).$$

The true value of parameter $\theta_0 = (\beta_0, \pi_0)'$ satisfies the restrictions $|\pi_0| < 1$, $\beta_0 \neq 0$ and $|\pi_0 + \beta_0| < 1$, which guarantee that the process is stationary and invertible. For simplicity we assume that $Y_0 = 0$ and $e_0 = 0$, though due to stationarity and invertibility the initial condition does not matter asymptotically. One can re-write the model as

$$(1 - (\pi + \beta)L)Y_t = (1 - \pi L)e_t, \quad \text{or} \quad Y_t = (1 - (\pi + \beta)L)^{-1}(1 - \pi L)e_t.$$

It is easy to see that if $\beta = 0$ the parameter π is not identified. Assume that the model is point identified, that is $\beta \neq 0$, but that identification is weak. This can be modeled as $\beta = \frac{C}{\sqrt{T}}$.

First, we write the log-likelihood function. Here we follow the derivation of Andrews and Cheng (2012) closely:

$$e_t = \sum_{j=0}^{t-1} \pi_0^j (Y_{t-j} - (\pi_0 + \beta_0)Y_{t-j-1}) = Y_t - \beta_0 \sum_{j=0}^{t-1} \pi_0^j Y_{t-j-1}.$$

$$\ell(\beta, \pi) = \text{const} - \frac{1}{2} \sum_{t=1}^T (Y_t - \beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1})^2.$$

Next, we introduce the following two time series:

$$u_t = \sum_{j=0}^t \pi_0^j Y_{t-j} = (1 - \pi_0 L)^{-1} Y_t = (1 - (\pi_0 + \beta_0)L)^{-1} e_t,$$

and

$$v_t = \sum_{j=0}^t j \pi_0^{j-1} Y_{t-j} = (1 - \pi_0 L)^{-2} Y_{t-1} = (1 - \pi_0 L)^{-2} (1 - (\pi_0 + \beta_0)L)^{-1} (1 - \pi_0 L) e_{t-1} =$$

$$= (1 - \pi_0 L)^{-1} (1 - (\pi_0 + \beta_0)L)^{-1} e_{t-1}.$$

Series u_t is an AR(1) process with coefficient $\pi_0 + \beta_0$; v_t is an AR(2) process with roots π_0 and $\pi_0 + \beta_0$.

One can see that the score is:

$$S_\beta(\theta) = \sum_{t=1}^T \left[(Y_t - \beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1}) \left(\sum_{j=0}^{t-1} \pi^j Y_{t-j-1} \right) \right];$$

$$S_\pi(\theta) = \beta \sum_{t=1}^T \left[(Y_t - \beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1}) \left(\sum_{j=0}^{t-1} j \pi^{j-1} Y_{t-j-1} \right) \right].$$

Notice that $Y_t - \beta_0 \sum_{j=0}^{t-1} \pi_0^j Y_{t-j-1} = e_t$. As a result,

$$S_T(\theta_0) = \begin{pmatrix} S_\beta(\beta_0, \pi_0) \\ S_\pi(\beta_0, \pi_0) \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T e_t u_{t-1} \\ \beta_0 \sum_{t=1}^T e_t v_{t-1} \end{pmatrix}.$$

We can now write the two measures of information:

$$J_T(\beta_0, \pi_0) = \begin{pmatrix} \sum_{t=1}^T e_t^2 u_{t-1}^2 & \beta_0 \sum_{t=1}^T e_t^2 u_{t-1} v_{t-1} \\ \beta_0 \sum_{t=1}^T e_t^2 u_{t-1} v_{t-1} & \beta_0^2 \sum_{t=1}^T e_t^2 v_{t-1}^2 \end{pmatrix},$$

$$I_T(\theta_0) = -\frac{\partial^2}{\partial \theta \partial \theta'} \ell =$$

$$= \begin{pmatrix} \sum_{t=1}^T u_{t-1}^2 & -\sum_{t=1}^T e_t v_{t-1} + \beta_0 \sum_{t=1}^T u_{t-1} v_{t-1} \\ -\sum_{t=1}^T e_t v_{t-1} + \beta_0 \sum_{t=1}^T u_{t-1} v_{t-1} & \beta_0^2 \sum_{t=1}^T v_{t-1}^2 - \beta_0 \sum_{t=1}^T e_t w_{t-1} \end{pmatrix},$$

here $w_{t-1} = \sum_{j=0}^{t-1} j(j-1) \pi_0^{j-2} Y_{t-j-1}$ is a weakly stationary series.

Assume weakly canceling roots, that is, $\beta = C/\sqrt{T}$. Then for a normalizing matrix $K_T = \text{diag}(1/\sqrt{T}, 1)$ we have

$$K_T J_T(\theta_0) K_T \rightarrow^p \begin{pmatrix} E[u_{t-1}^2] & C \cdot E[u_{t-1} v_{t-1}] \\ C \cdot E[u_{t-1} v_{t-1}] & C^2 \cdot E[v_{t-1}^2] \end{pmatrix}, \quad (\text{S3})$$

where we used the Law of Large Numbers.

We also can notice that

$$K_T (J_T(\theta_0) - I_T(\theta_0)) K_T = \begin{pmatrix} 0 & \frac{1}{\sqrt{T}} \sum e_t v_{t-1} \\ \frac{1}{\sqrt{T}} \sum e_t v_{t-1} & \frac{C}{\sqrt{T}} \sum e_t w_{t-1} \end{pmatrix} + o_p(1) \Rightarrow \begin{pmatrix} 0 & \xi \\ \xi & C\eta \end{pmatrix},$$

where $(\xi, \eta)'$ is a mean-zero normal vector with covariance matrix

$$\begin{pmatrix} E[v_{t-1}^2] & E[v_{t-1} w_{t-1}] \\ E[v_{t-1} w_{t-1}] & E[w_{t-1}^2] \end{pmatrix}.$$

Assumption 1 It is easy to see that Lindeberg's condition holds for sequences $\frac{e_t u_{t-1}}{\sqrt{T}}$ and $\frac{e_t v_{t-1}}{\sqrt{T}}$. We check Assumption 1(b) in equation (S3). As a result, Theorem 1 holds for the ARMA(1,1) model with near-canceling roots, and we have a robust test for a simple hypothesis $H_0 : \pi = \pi_0, \beta = \beta_0$.

Let us consider the problem of testing the weakly identified parameter π , treating β as a nuisance parameter. The hypothesis of interest is $H_0 : \pi = \pi_0$.

Assumptions 2

(a) We showed before:

$$\frac{1}{T} I_{\beta\beta}(\theta_0) = \frac{1}{T} \sum_{t=1}^T u_t^2 \xrightarrow{p} \lim \frac{1}{T} J_{\beta\beta}(\theta_0).$$

So, $J_{\beta\beta}^{-1}(\theta_0) I_{\beta\beta}(\theta_0) \xrightarrow{p} 1$.

(b) $I_{\beta\beta}(\pi_0, \beta)$ does not depend on β .

(c) Function $\ell(\pi_0, \beta)$ is quadratic in β , as a result $\hat{\beta}(\pi_0)$ is the OLS estimator in a regression of Y_t on u_t . The assumption trivially holds.

This means that Assumption 2 is satisfied, and thus the restricted ML estimate of β is asymptotically normal under the null.

Assumption 3 We have to check the conditions for the CLT for a pair $S_T(\theta_0)$ and

$$A_{\beta\pi}(\theta_0) = J_{\beta\pi}(\theta_0) - I_{\beta\pi}(\theta_0) = \beta_0 \sum_{t=1}^T (e_t^2 - 1) u_{t-1} v_{t-1} + \sum_{t=1}^T e_t v_{t-1}.$$

It is easy to see that for $\beta_0 = C/\sqrt{T}$ and $K_{\beta\pi} = \frac{1}{\sqrt{T}}$, Assumption 3 is satisfied, and $K_{\beta\pi} A_{\beta\pi} \Rightarrow N(0, E v_t^2)$.

Assumption 4

(a) We have $K_{\beta,T} = K_{\beta\pi,T} = \frac{1}{\sqrt{T}}$ and $K_{\pi,T} = 1$. Assumption 4(a) holds trivially.

- (b) Note that $\frac{\partial^3}{\partial^2\beta\partial\pi}\ell = -2\sum u_{t-1}v_{t-1}$. We may try to calculate $\Lambda_{\beta\beta\pi}$ from the third information equality, but it is enough to notice that $K_{\beta,T}^2 K_{\pi,T} \frac{\partial^3}{\partial^2\beta\partial\pi}\ell = -\frac{2}{T}\sum u_{t-1}v_{t-1}$ satisfies the Law of Large Numbers, and that all terms in the third information equality are normalized to converge to their expectations. This implies that $K_{\beta,T}^2 K_{\pi,T} \Lambda_{\beta\beta\pi}$ converges to its expectation (which is zero, since Λ is a martingale);
- (c) The argument here is exactly the same as in (b), with the additional observation that $\frac{\partial^4}{\partial^3\beta\partial\pi}\ell = 0$.

Since Assumptions 2, 3 and 4 are satisfied, according to Theorem 2 the two score test statistics $\widetilde{LM}_o(\pi_0)$ and $\widetilde{LM}_e(\pi_0)$ for testing hypothesis $H_0 : \pi = \pi_0$ have an asymptotic χ_1^2 distribution despite the weak identification of π .

S3 An additional example of weak identification: nearly reduced dynamics

This section contains an additional example showing how weak identification can arise in DSGE models. Specifically, we consider an example in which insufficiently rich dynamics for the observed variables gives rise to weak identification.

Assume that we observe a sample of 2×1 random vectors $Y_t, t = 1, \dots, T$ generated from the following model:

$$\left\{ \begin{array}{l} A(\tilde{\theta})Y_t = U_t, \\ U_t = \Lambda U_{t-1} + \varepsilon_t, \\ \varepsilon_t \sim i.i.d.N(0, \Sigma), \end{array} \right.$$

which is the form typically taken by log-linearized DSGE models. Here U_t and ε_t are 2×1 unobserved random vectors. Assume that the matrix of persistence parameters $\Lambda = \text{diag}(\rho, \delta)$ and the matrix of variances $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$ are both diagonal. The vector $\theta = (\tilde{\theta}, \sigma_1^2, \sigma_2^2, \rho, \delta)$ contains the unknown parameters. We will show that if the elements of Λ are equal, the parameter $\tilde{\theta}$ may become locally under-identified.

S3.1 Identification when $\delta \neq \rho$

According to Komunjer and Ng (2011), two parameter values θ_0 and θ_1 are observationally equivalent if and only if there exists matrix P such that

$$\begin{cases} P\Lambda_0P^{-1} = \Lambda_1; \\ PA(\tilde{\theta}_0) = A(\tilde{\theta}_1); \\ P\Sigma_0P' = \Sigma_1. \end{cases}$$

Assume that $\rho \neq \delta$. If there exists a matrix P such that for some diagonal matrices Λ_1 and Σ_1 we have $P\Lambda_0P^{-1} = \Lambda_1$ and $P\Sigma_0P' = \Sigma_1$, then matrix the P must be of the form $\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$ or $\begin{pmatrix} 0 & c_1 \\ c_2 & 0 \end{pmatrix}$ for some non-zero constants c_1 and c_2 . Thus the model is locally identified at θ_0 if and only if the transformation $f : (c_1, c_2, \tilde{\theta}) \rightarrow \text{vec} \left\{ \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} A(\tilde{\theta}) \right\}$ is locally injective at $(c_1, c_2, \tilde{\theta}) = (1, 1, \tilde{\theta}_0)$. The sufficient condition for this is that the derivative of f with respect to $(c_1, c_2, \tilde{\theta})$ have full rank at $(1, 1, \tilde{\theta}_0)$. The above mentioned matrix derivative is written below:

$$\begin{pmatrix} A_{11}(\tilde{\theta}_0) & 0 \\ 0 & A_{21}(\tilde{\theta}_0) \\ A_{12}(\tilde{\theta}_0) & 0 \\ 0 & A_{22}(\tilde{\theta}_0) \end{pmatrix} ; \frac{\partial}{\partial \tilde{\theta}} \text{vec}(A(\tilde{\theta}))$$

If this matrix has full rank, then parameter θ is locally identified at θ_0 . As we can see, for $\tilde{\theta}$ to be point-identified it must be of dimension at most two, which makes the dimension of $\theta = (\tilde{\theta}, \sigma_1^2, \sigma_2^2, \rho, \delta)$ equal to six. From now on we assume that $\tilde{\theta}$ is 2-dimensional and that the model is point identified for $\rho \neq \delta$.

S3.2 Identification at $\rho = \delta$

In order to show that identification fails at $\delta = \rho$ we write the likelihood for the model $\ell(\theta; Y_1, \dots, Y_T)$. Let $\Delta\ell_t(\theta) = \ell(\theta; Y_1, \dots, Y_t) - \ell(\theta; Y_1, \dots, Y_{t-1})$ be the increment of the likelihood in period t :

$$\Delta\ell_t = -\frac{1}{2}(A(\tilde{\theta})Y_t - \Lambda A(\tilde{\theta})Y_{t-1})'\Sigma^{-1}(A(\tilde{\theta})Y_t - \Lambda A(\tilde{\theta})Y_{t-1}) - \frac{1}{2}\log|\Sigma| + \log|A(\tilde{\theta})|.$$

Consider the score. First take the score with respect to the variances:

$$2 \frac{\partial \Delta \ell_t}{\partial \sigma_i^2}(\theta_0) = \frac{1}{\sigma_i^4}(\varepsilon_{i,t}^2 - \sigma_i^2).$$

Next, let s be a part of $\tilde{\theta}$. We have:

$$\begin{aligned} -\frac{\partial \Delta \ell_t}{\partial s}(\theta_0) &= (A(\tilde{\theta})Y_t - \Lambda A(\tilde{\theta})Y_{t-1})' \Sigma^{-1} \left(\frac{\partial A}{\partial s} Y_t - \Lambda \frac{\partial A}{\partial s} Y_{t-1} \right) - \text{trace} \left(\frac{\partial A}{\partial s} A^{-1} \right) = \\ &= \varepsilon_t' \Sigma^{-1} \left(\frac{\partial A}{\partial s} A^{-1} (\Lambda U_{t-1} + \varepsilon_t) - \Lambda \frac{\partial A}{\partial s} A^{-1} U_{t-1} \right) - \text{trace} \left(\frac{\partial A}{\partial s} A^{-1} \right). \end{aligned}$$

If $\rho = \delta$ then $\Lambda = \delta Id_2$ and $\frac{\partial A}{\partial s} A^{-1} \Lambda = \Lambda \frac{\partial A}{\partial s} A^{-1}$. As a result

$$-\frac{\partial \Delta \ell_t}{\partial s}(\theta_0) = \text{trace} \left((\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} \frac{\partial A}{\partial s} A^{-1} \right).$$

We can see that the score with respect to the four parameters $(\tilde{\theta}, \sigma_1^2, \sigma_2^2)$ is a linear function of the three-dimensional random variable $\sum_{t=1}^T (\varepsilon_t \varepsilon_t' - \Sigma)$. This implies that the Fisher information for parameters $\tilde{\theta}, \sigma_1^2, \sigma_2^2$, which is equal to covariance matrix of score, is degenerate and has rank at most three (which makes the rank for the full parameter vector θ at most five). Thus we lose one degree of identification compared with the case of $\rho \neq \delta$.

S3.3 Weak identification

We model weak identification as $\Lambda = \delta Id_2 + \frac{1}{\sqrt{T}} \mu$, where $\mu = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}$. Consider the score. First take the score with respect to the variances:

$$2 \frac{\partial \Delta \ell_t}{\partial \sigma_i^2}(\theta_0) = \frac{1}{\sigma_i^4}(\varepsilon_{i,t}^2 - \sigma_i^2).$$

Next let s be a part of $\tilde{\theta}$. We have:

$$\begin{aligned} -\frac{\partial \Delta \ell_t}{\partial s}(\theta_0) &= \text{trace} \left((\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} \frac{\partial A}{\partial s} A^{-1} \right) + \\ &\frac{1}{\sqrt{T}} \text{trace} \left(U_{t-1} \varepsilon_t \Sigma^{-1} \left(\frac{\partial A}{\partial s} A^{-1} \mu - \mu \frac{\partial A}{\partial s} A^{-1} \right) \right). \end{aligned} \quad (\text{S4})$$

Consider the following variables:

$$\begin{aligned} \xi_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (\varepsilon_{1,t}^2 - \sigma_1^2, \varepsilon_{2,t}^2 - \sigma_2^2, \varepsilon_{1,t} \varepsilon_{2,t})'; \\ \eta_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{vec}(U_{t-1} \varepsilon_t). \end{aligned}$$

Both ξ_T and η_T converge to mean-zero normal vectors (which are three and four dimensional respectively), all components of which are independent. We then see that

$$-\frac{1}{\sqrt{T}} \frac{\partial \ell_t}{\partial s}(\theta_0) = \gamma'_s \xi_T + \frac{1}{\sqrt{T}} \gamma_s'^* \eta_T. \quad (\text{S5})$$

Here γ_s and γ_s^* are fixed vectors.

Let $\theta^* = (\tilde{\theta}, \sigma_1^2, \sigma_2^2)$ be the subset of parameters excluding ρ and δ . What we have shown is that:

$$-\frac{1}{\sqrt{T}} \frac{\partial \ell_T}{\partial \theta^*}(\theta_0) = \Gamma \xi_T + \frac{1}{\sqrt{T}} \Gamma^* \eta_T,$$

where the score $-\frac{1}{\sqrt{T}} \frac{\partial \ell_t}{\partial \theta^*}(\theta_0)$ is 4×1 vector, Γ is 4×3 matrix, and Γ^* is 4×4 matrix. As a result, the 4×4 block of the normalized Fisher information matrix corresponding to the parameters θ^* has rank three asymptotically:

$$\frac{1}{T} \mathcal{I}_{\theta^*, T} = \Gamma \text{Var}(\xi_t) \Gamma' + \frac{1}{T} \Gamma^* \text{Var}(\eta_T) (\Gamma^*)' \rightarrow \Gamma \text{Var}(\xi_t) \Gamma'.$$

Now let us look at the components of the score corresponding to δ and ρ :

$$\begin{aligned} \frac{\partial \Delta \ell_t}{\partial \delta}(\theta_0) &= \varepsilon_t' \Sigma^{-1} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} U_{t-1} = \text{trace} \left(U_{t-1} \varepsilon_t' \Sigma^{-1} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right), \\ \frac{\partial \Delta \ell_t}{\partial \rho}(\theta_0) &= \text{trace} \left(U_{t-1} \varepsilon_t' \Sigma^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right). \end{aligned}$$

As a result

$$\frac{1}{\sqrt{T}} \frac{\partial \ell_T}{\partial (\rho, \delta)}(\theta_0) = \tilde{\Gamma} \eta_T,$$

where $\tilde{\Gamma}$ is 2×4 matrix of full rank. We see that the part of the normalized information matrix corresponding to the block of parameters ρ and δ has rank two asymptotically, and that the information matrix is asymptotically block-diagonal.

S3.3.1 Asymptotic behavior of Hessian

In the previous section we showed that the normalized (per observation) Fisher information for the 4-dimensional parameter θ^* is of rank three asymptotically, and as

a result there is a direction α along which this matrix is degenerate. We show that the normalized (per observation) Hessian of the log-likelihood is NOT asymptotically degenerate along this direction.

For simplicity of notation denote by \mathcal{I} the limit of the normalized (per observation) theoretical Fisher information for the block of parameters θ^* , that is,

$$\mathcal{I} = \lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{I}_{\theta^*, T} = \lim_{T \rightarrow \infty} \frac{1}{T} E \sum_{t=1}^T \left(\frac{\partial \Delta \ell_t}{\partial \theta^*} \right) \left(\frac{\partial \Delta \ell_t}{\partial \theta^*} \right)' = - \lim_{T \rightarrow \infty} \frac{1}{T} E \frac{\partial^2 \ell_T}{\partial \theta^* \partial \theta^{*t}}.$$

Let us also denote by $\mathcal{I}_{s, \tilde{s}}$ the entry of \mathcal{I} corresponding to parameters s and \tilde{s} .

First consider two parameters $s, \tilde{s} \in \tilde{\theta}$ and let $A_s = \frac{\partial A}{\partial s} A^{-1}$, $B_s = \frac{\partial A}{\partial s} A^{-1} \mu - \mu \frac{\partial A}{\partial s} A^{-1}$, $A_{s, \tilde{s}} = \frac{\partial^2 A}{\partial s \partial \tilde{s}} A^{-1}$, $B_{s, \tilde{s}} = \frac{\partial^2 A}{\partial s \partial \tilde{s}} A^{-1} \mu - \mu \frac{\partial^2 A}{\partial s \partial \tilde{s}} A^{-1}$. We have the following:

$$\begin{aligned} i_{T,t} &= - \frac{\partial^2 \Delta \ell_t}{\partial s \partial \tilde{s}}(\theta_0) = \left(A_s \varepsilon_t + \frac{1}{\sqrt{T}} B_s U_{t-1} \right)' \Sigma^{-1} \left(A_{\tilde{s}} \varepsilon_t + \frac{1}{\sqrt{T}} B_{\tilde{s}} U_{t-1} \right) + \\ &\quad + \varepsilon_t' \Sigma^{-1} \left(A_{s\tilde{s}} \varepsilon_t + \frac{1}{\sqrt{T}} B_{s\tilde{s}} U_{t-1} \right) + \text{trace}(A_s A_{\tilde{s}}) - \text{trace}(A_{s, \tilde{s}}) = \\ &= \{ \varepsilon_t' A_s' \Sigma^{-1} A_{\tilde{s}} \varepsilon_t + \text{trace}(A_s A_{\tilde{s}}) \} + \text{trace} [(\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} A_{s\tilde{s}}] + O_p(1/T). \end{aligned} \quad (\text{S6})$$

As a result we have

$$\mathcal{I}_{s\tilde{s}} = E \{ \varepsilon_t' A_s' \Sigma^{-1} A_{\tilde{s}} \varepsilon_t + \text{trace}(A_s A_{\tilde{s}}) \} = \text{trace}(\Sigma A_s' \Sigma^{-1} A_s) + \text{trace}(A_s A_{\tilde{s}}).$$

Let us define $C_s = \Sigma^{-1/2} A_s \Sigma^{1/2}$, then

$$\mathcal{I}_{s\tilde{s}} = \text{trace}(C_s' C_s) + \text{trace}(C_s C_s) = \text{trace}(D_s D_{\tilde{s}}),$$

where $D_s = \frac{1}{\sqrt{2}}(C_s + C_s')$ is a symmetric matrix.

In fact, all entries of the limit of the normalized Fisher information matrix \mathcal{I} have this form. Consider the entry corresponding to $s \in \tilde{\theta}$ and a variance σ_i^2 :

$$- \frac{\partial^2 \Delta \ell_t}{\partial s \partial \sigma_i^2}(\theta_0) = - \frac{\varepsilon_{it}}{\sigma_i^4} (A_s \varepsilon_t + \frac{1}{\sqrt{T}} B_s U_{t-1})_i$$

where the sub-index i stands for the i -th component. As a result,

$$\mathcal{I}_{s,i} = \text{trace}(\Sigma M_i \Sigma^{-1} A_s)$$

where M_i is matrix that has all entries equal to zero except entry ii which is $-\frac{1}{\sigma_i^2}$. Matrix $\Sigma^{-1/2} M_i \Sigma^{1/2}$ is symmetric. Define $D_i = \frac{1}{\sqrt{2}} \Sigma^{-1/2} M_i \Sigma^{1/2}$.

Thus, for any two parameters s and \tilde{s} out of $\theta^* = (\tilde{\theta}, \sigma_1^2, \sigma_2^2)$, the entry of the information matrix corresponding to this pair is

$$\mathcal{I}_{s,\tilde{s}} = \text{trace}(D_s D_{\tilde{s}}),$$

and all matrices D_s are symmetric 2×2 matrices. Because these matrices are symmetric

$$\text{trace}(D_s D_{\tilde{s}}) = \sum_{i,k} (D_s)_{ik} (D_{\tilde{s}})_{ik} = (\text{vec}(D_s))' \text{vec}(D_{\tilde{s}}).$$

Since D_s is symmetric there are two repeating entries. Let us define D_s^* to be 3×1 vector such that

$$\text{trace}(D_s D_{\tilde{s}}) = (D_s^*)' D_{\tilde{s}}^*.$$

If we put all the vectors D_s^* into one matrix D (of dimension 3×4), we get

$$\mathcal{I} = D' D$$

and so can see that \mathcal{I} is a 4×4 matrix of rank three, and the degenerate direction is the direction perpendicular to D_s^* for all $s \in \theta^*$. Call this direction α . Consider a linear combination of the parameters $\alpha' \theta^*$ and note that the limit of the normalized Fisher information along this direction is $\mathcal{I}_\alpha = \alpha' \mathcal{I} \alpha = \alpha' D' D \alpha = 0$.

The expression for \mathcal{I} is obtained as the expectation of the negative second derivative. Given the second information equality \mathcal{I} is also equal to the limit of the normalized covariance matrix of the score. From the formula for the score (S4) we have that for $S_s = \text{trace}((\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} A_s)$,

$$\text{cov}(S_s, S_{\tilde{s}}) = (D_s^*)' D_{\tilde{s}}^*,$$

where D_s^* is a 3×1 vector-function of A_s and Σ only (described above).

The Hessian is $I_T = \sum_{t=1}^T i_{T,t}$, where the explicit formula for $i_{T,t}$ is given in (S6).

We can see that :

$$\left(\frac{1}{T} I_T - \mathcal{I} \right)_{s,\tilde{s}} = \frac{1}{T} \sum_{t=1}^T \text{trace} [(\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} A_{s\tilde{s}}] + O_p(1/T).$$

The summands in the expression above have the same form as random variables S_s . As a result we have:

$$\lim_{T \rightarrow \infty} Tcov \left(\left(\frac{1}{T} I_T - \mathcal{I} \right)_{s, \tilde{s}}, \left(\frac{1}{T} I_T - \mathcal{I} \right)_{r, \tilde{r}} \right) = (D_{s, \tilde{s}}^*)' D_{r, \tilde{r}}^*,$$

where $D_{s, \tilde{s}}^*$ is 3×1 and constructed from $A_{s, \tilde{s}}$ in exactly the same manner as D_s^* is constructed from A_s .

Consider the direction $\alpha = (\alpha_s)_{s \in \theta}$ such that $\alpha' \mathcal{I} \alpha = 0$ and note that

$$\begin{aligned} \lim_{T \rightarrow \infty} Tvar \left(\alpha' \left(\frac{1}{T} I_T - \mathcal{I} \right) \alpha \right) &= \lim_{T \rightarrow \infty} var \left(\sum_{s, \tilde{s}} \left(\frac{1}{T} I_T - \mathcal{I} \right)_{s, \tilde{s}} \alpha_s \alpha_{\tilde{s}} \right) = \\ &= \sum_{s, \tilde{s}} \sum_{r, \tilde{r}} (D_{s, \tilde{s}}^*)' D_{r, \tilde{r}}^* \alpha_s \alpha_{\tilde{s}} \alpha_r \alpha_{\tilde{r}} = \left\| \sum_{s, \tilde{s}} D_{s, \tilde{s}}^* \alpha_s \alpha_{\tilde{s}} \right\|^2. \end{aligned}$$

In general the last expression is non-zero. For example, assume that Σ is identity matrix. Then the last expression is equal to zero if and only if the second derivative of matrix $A + A'$ along direction α is equal to zero. This is obviously true if for example A is a linear function of the parameter. In general, however, for non-linear functions the second derivative along the special degenerate direction does not have to be zero, and thus the stochasticity of I_T along this direction is non-trivial asymptotically.

S3.4 Assumptions 1-4

Assumption 1. Given the formula of score stated in equation (S5) it is easy to see that Assumption 1 holds.

Let us denote $\beta = \theta^* = (\tilde{\theta}, \sigma_1^2, \sigma_2^2)$, $\alpha = (\rho, \delta)$. Below we show that Assumptions 2-4 hold for testing $H_0 : \beta = \beta_0$ with the nuisance parameter α .

Assumption 2. Denote $e_1 = (1, 0)'$ and $e_2 = (0, 1)'$. Then $\Lambda = \rho e_1 e_1' + \delta e_2 e_2'$. It is easy to see that

$$\frac{\partial \ell_T}{\partial \rho} = \sum_{t=1}^T U'_{t-1} e_1 e_1' \Sigma^{-1} \varepsilon_t; \quad \frac{\partial \ell_T}{\partial \delta} = \sum_{t=1}^T U'_{t-1} e_2 e_2' \Sigma^{-1} \varepsilon_t.$$

We can also note that

$$-\frac{\partial^2 \ell_T}{\partial \rho^2} = \sum_{t=1}^T U'_{t-1} e_1 e_1' \Sigma^{-1} e_1 e_1' U_{t-1}; \quad -\frac{\partial^2 \ell_T}{\partial \delta^2} = \sum_{t=1}^T U'_{t-1} e_2 e_2' \Sigma^{-1} e_2 e_2' U_{t-1}$$

and

$$-\frac{\partial^2 \ell_T}{\partial \rho \partial \delta} = \sum_{t=1}^T U'_{t-1} e_1 e_1' \Sigma^{-1} e_2 e_2' U_{t-1}.$$

It is easy to see that the Law of Large Numbers implies that $\frac{1}{T} \frac{\partial^2 \ell_T}{\partial \alpha \partial \alpha'}$ and $\frac{1}{T} \sum_{t=1}^T \frac{\partial \Delta \ell_t}{\partial \alpha} \left(\frac{\partial \Delta \ell_t}{\partial \alpha} \right)'$ converge to the same matrix

$$\begin{pmatrix} \frac{EU_{t-1,1}^2}{\sigma_1^2} & 0 \\ 0 & \frac{EU_{t-1,2}^2}{\sigma_2^2} \end{pmatrix}.$$

Thus Assumption 2(a) holds. Assumption 2(b) holds trivially since the third derivative of ℓ_T with respect to α is zero. We also notice that estimator $\hat{\alpha}(\beta_0)$ is the usual OLS estimator, as such Assumption 2(c) holds trivially.

Assumption 3. We need only to check that some form of the CLT holds for the terms in the martingale $A_{\alpha\beta}$. Here we check one term, all others can be checked in the same manner. One can easily check that for $s \in \tilde{\theta}$

$$i_{\rho,s,t} = -\frac{\partial^2 \Delta \ell_t}{\partial \rho \partial s} = U'_{t-1} A_s e_1 e_1' \Sigma^{-1} \varepsilon_t + U'_{t-1} e_1 e_1' \Sigma^{-1} A_s \varepsilon_t + \frac{1}{\sqrt{T}} U'_{t-1} e_1 e_1' \Sigma^{-1} B_s U_{t-1}$$

while the score is

$$\begin{aligned} \frac{\partial \Delta \ell_t}{\partial \rho} &= U'_{t-1} e_1 e_1' \Sigma^{-1} \varepsilon_t; \\ \frac{\partial \Delta \ell_t}{\partial s} &= \varepsilon_t' \Sigma^{-1} A_s \varepsilon_t - \text{trace}(A_s) + \frac{1}{\sqrt{T}} \varepsilon_t' \Sigma^{-1} B_s U_{t-1} \end{aligned}$$

As a result,

$$\begin{aligned} \frac{1}{\sqrt{T}} A_{\rho,s,T} &= \text{trace} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t U'_{t-1} (A_s e_1 e_1' \Sigma^{-1} + e_1 e_1' \Sigma^{-1} A_s) \right) - \\ &\quad - \frac{1}{\sqrt{T}} \sum_{t=1}^T U'_{t-1} e_1 e_1' \Sigma^{-1} \varepsilon_t \text{trace}((\varepsilon_t \varepsilon_t' - \Sigma) \Sigma^{-1} A_s) + O_p(1/T). \end{aligned}$$

We can see that the CLT holds for the last expression, and $K_{\alpha_i, \beta_j, T} = \frac{1}{\sqrt{T}}$. For the terms that involve α and σ_i^2 we notice that

$$I_{\rho, \sigma_1^2, T} = - \sum_{t=1}^T \frac{U_{1,t-1} \varepsilon_{1,t}}{\sigma_1^4}$$

and $I_{\rho, \sigma_2^2, T} = 0$. So, $\frac{1}{\sqrt{T}} I_{\rho, \sigma_1^2, T}$ converges to a Gaussian random variable, and one can verify that the corresponding J_T entries converge in probability.

Assumption 4. Assumption (a) holds trivially since $K_{\alpha_i, T} = \frac{1}{\sqrt{T}}$, $K_{\alpha_i, \beta_j, T} = \frac{1}{\sqrt{T}}$, while $K_{\beta_j, T}$ is bounded (it is 1 for some directions while $\frac{1}{\sqrt{T}}$ for the others).

For part (b) we notice that $\Lambda_{\alpha_i, \alpha_j, \beta}$ is a linear combination of terms which are products of ε_t and U_{t-1} up to order 4. As a result all terms in $[\Lambda_{\alpha_i, \alpha_j, \beta}]$ satisfy the Law of Large Numbers and thus $\frac{1}{T} [\Lambda_{\alpha_i, \alpha_j, \beta}] \rightarrow^p \text{const}$. Thus, it is easy to see that the expression in Assumption 4(b) has too strong a normalization and converges to zero.

Assumption (c) holds trivially since $I_{\alpha, \alpha}(\alpha, \beta) = I_{\alpha, \alpha}(\alpha_0, \beta)$ for any α, α_0 and β .

S4 Additional example of weak identification: Weak VAR

The identification failure observed in our main example in Section 2 of the paper when $\rho = \delta$ results from the interplay of two problems, one of which is reduced dynamics, discussed in Section S3, while the other is that the structural VAR loses one degree of identification due to the fact that the 2×2 matrix $C(\theta)$ has rank 1. The example of this section deals with the second problem, in particular, we consider structural VAR models where part of parameter vector is weakly identified. Fernández-Villaverde et al. (2007) discuss the relationship between linearized DGSE models and VARs. To model weak identification in this context we follow the approach of Stock and Wright (2000) and consider a set of drifting functions that become asymptotically flat in some directions.

Consider an exponential family with joint density of the form

$$f_T(X_T|\theta) = h(X_T) \exp \left\{ \eta_T(\theta)' \sum_{t=1}^T H(x_t) - TA(\eta_T(\theta)) \right\}. \quad (\text{S7})$$

Here η is a p -dimensional reduced-form parameter, while $\sum_{t=1}^T H(x_t)$ is a p -dimensional sufficient statistic. Model (S7) covers structural VAR models for η a set of reduced-form VAR coefficients, structural variance terms, and functions thereof and $x_t = (Y_t', \dots, Y_{t-p}')'$, where Y_t is a vector of data observed at time t , and the sufficient statistics are the sample auto-covariances of the Y_t .

Suppose that we can partition the structural coefficient θ into sub-vectors α and β , $\theta = (\alpha', \beta')'$. For this example we consider an embedding similar to that of Stock and Wright (2000) for weak GMM, which we use to model β as weakly identified. In particular, we assume that

$$\eta_T(\theta) = m(\alpha) + \frac{1}{\sqrt{T}} \tilde{m}(\alpha, \beta),$$

where $\frac{\partial}{\partial \alpha'} m(\alpha_0)$ and $\frac{\partial}{\partial \theta'} \eta_T(\theta_0)$ are matrices of full rank k_α and $k = k_\alpha + k_\beta$ correspondingly. Assume that an infinitesimality condition holds for the sequence $\left\{ \frac{1}{\sqrt{T}} H(x_t) \right\}_{t=1}^T$ and a law of large numbers holds for $H(x_t)H(x_t)'$ (i.e. $\frac{1}{T} \sum_{t=1}^T H(x_t)H(x_t)' \rightarrow^p E[H(x_t)H(x_t)']$).

Let \dot{A} and \ddot{A} denote the first and the second derivatives of A with respect to η (they are a $p \times 1$ vector and $p \times p$ matrix respectively). From the normalization in the exponential family we have that $E[H(x_t)] = \dot{A}$ and $\text{Var}(H(x_t)) = \ddot{A}$. Assume that the parameter space for θ is compact, that θ_0 lies in the interior of the parameter space, and that the function $Q(\alpha) = m(\alpha)\dot{A}(m(\alpha_0)) - A(m(\alpha))$ is uniquely maximized at the point α_0 .

The score is

$$S_T = \sum_{t=1}^T \left(H(x_t) - \dot{A} \right)' \begin{pmatrix} \frac{\partial m(\alpha)}{\partial \alpha} + \frac{1}{\sqrt{T}} \frac{\partial \tilde{m}(\alpha, \beta)}{\partial \alpha} \\ \frac{1}{\sqrt{T}} \frac{\partial \tilde{m}(\alpha, \beta)}{\partial \beta} \end{pmatrix}.$$

Consider a set of normalizing matrices $K_T = \begin{pmatrix} \frac{1}{\sqrt{T}} Id_{k_\alpha} & 0 \\ 0 & Id_{k_\beta} \end{pmatrix}$. It is easy to see

that Assumption 1 is trivially satisfied. In particular, since

$$\frac{1}{T} \sum_{t=1}^T \left(H(x_t) - \dot{A} \right) \left(H(x_t) - \dot{A} \right)' \rightarrow^p \ddot{A},$$

we have that $K_T J_T K_T'$ converges in probability to a positive definite matrix.

Now consider the behavior of the Hessian. It is easy to see that

$$(I_T)_{ij} = - \sum_{t=1}^T \left(H(x_t) - \dot{A} \right)' \frac{\partial^2 \eta_T}{\partial \theta_i \partial \theta_j} + T \left(\frac{\partial \eta_T}{\partial \theta_i} \right)' \ddot{A} \frac{\partial \eta_T}{\partial \theta_j}. \quad (\text{S8})$$

Since $\text{Var}(H(x_t)) = \ddot{A}$, we have

$$\lim_{T \rightarrow \infty} K_T T \left(\frac{\partial \eta_T}{\partial \theta} \right)' \ddot{A} \frac{\partial \eta_T}{\partial \theta} K_T' = \lim_{T \rightarrow \infty} K_T J_T K_T' = \lim_{T \rightarrow \infty} K_T \mathcal{I}_T K_T'.$$

That is, the second term in (S8) reflects the Fisher information. The first term in (S8) also matters asymptotically, however. In particular,

$$(K_T (I_T - J_T) K_T')_{\beta_i, \beta_j} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(H(x_t) - \dot{A} \right)' \frac{\partial^2 \tilde{m}}{\partial \beta_i \partial \beta_j} \Rightarrow \varsigma' \frac{\partial^2 \tilde{m}}{\partial \beta_i \partial \beta_j},$$

where ς is a Gaussian vector. Thus $K_T I_T K_T'$ and $K_T J_T K_T'$ have different asymptotic limits and $K_T (I_T - J_T) K_T'$ converges in distribution to a matrix $\begin{pmatrix} 0_{k_\alpha \times k_\alpha} & 0_{k_\alpha \times k_\beta} \\ 0_{k_\beta \times k_\alpha} & \xi \end{pmatrix}$, where ξ is $k_\beta \times k_\beta$ symmetric matrix with Gaussian entries.

S4.1 Assumptions 2-4

Below we check Assumptions 2-4 for testing hypothesis $H_0 : \beta = \beta_0$ with strongly identified nuisance parameter α .

Assumption 2. Assumption 2(a) has been checked above. For the Assumption 2(b) we assume that non-stochastic functions $m(\alpha)$, $\tilde{m}(\alpha, \beta_0)$ and $A(\eta_T(\alpha, \beta_0))$ have third derivatives with respect to α that are bounded in absolute value over the whole parameter space for α . Indeed,

$$K_{\alpha, T} I_{\alpha_i, \alpha_j, T}(\alpha, \beta_0) K_{\alpha, T}' = - \left(\frac{1}{T} \sum_{t=1}^T H(x_t) \right) \frac{\partial^2 \eta_T}{\partial \alpha_i \partial \alpha_j} - \ddot{A} \frac{\partial^2 \eta_T}{\partial \alpha_i \partial \alpha_j} + \left(\frac{\partial \eta_T}{\partial \alpha_i} \right)' \ddot{A} \frac{\partial \eta_T}{\partial \alpha_j}$$

The last two terms are non-stochastic as well as term $\frac{\partial^2 \eta_T}{\partial \alpha_i \partial \alpha_j}$, the change in these terms when they evaluated at α_0 and α such that $K_{\alpha,T}^{-1}|\alpha - \alpha_0| \leq \delta$ is of order $O(K_{\alpha,T}) = O(\frac{1}{\sqrt{T}})$. The stochastic part of the first term $\frac{1}{T} \sum_{t=1}^T H(x_t)$ does not depend on α and converges to a constant by the Law of Large Numbers. Assumption 2(c) trivially follows from classical results, since $\hat{Q}(\alpha) = \frac{1}{T} \ell_T(\alpha, \beta_0)$ uniformly converges to $Q(\alpha) = m(\alpha) \dot{A}(m(\alpha_0)) - A(m(\alpha))$.

Assumption 3. It is easy to see that

$$A_{\alpha_i, \beta_j} = - \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (H(x_t) - \dot{A}) \right)' \frac{\partial^2}{\partial \alpha_i \partial \beta_j} - \left(\frac{\partial \eta_T}{\alpha} \right)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T ((H(x_t) - \dot{A})(H(x_t) - \dot{A}) - \ddot{A}) \right) \frac{\partial \tilde{m}}{\partial \beta_j}$$

Assume that the Law of Large Numbers holds for the fourth power of $H(x_t)$, then Assumption 3 holds with $K_{\alpha_i, \beta_j, T} = 1$.

Assumption 4. Assumption 4 (a) holds trivially. For Assumption 4(b) we assume that the Law of Large Numbers holds for any products of any up to 6 components of stochastic vectors $H(x_t)$, in such a case 4(b) holds due to the fact that $\frac{1}{T} [\Lambda_{\alpha_i, \alpha_j, \beta_n}]$ converges to a constant, while $K_{\alpha_i, T} K_{\alpha_j, T} K_{\beta_n, T} = \frac{1}{T}$. For Assumption 4(c) we assume that $\frac{\partial^3 \tilde{m}(\alpha, \beta_0)}{\partial \alpha^2 \partial \beta}$ is bounded everywhere.

S5 Additional Example: regime switching model

So far we have discussed only log-linearized DSGE models, which have been the primary focus of the DSGE literature to date. However, the robust tests we propose are applicable to non-linear models as well.

One class of non-linear DSGE models in the literature is that of models with regime switching, for example, Schorfheide (2005) whose model includes an exogenous state variable that determines the target inflation rate and the variance of Taylor-rule shocks. Such regime-switching mechanisms can produce additional weak identification

issues: for example, if the two regimes produce similar behavior for the observable variables, then the regime-switching probabilities will be weakly identified.

One difficulty of working with non-linear DSGE models is that it is often challenging to calculate the likelihood function and its derivatives, which we will need to evaluate our tests. For example, the frequently-used particle filter does not typically allow us to approximate derivatives to a sufficient level of accuracy. Nonetheless, there are some nonlinear models where the likelihood can be approximated using other methods which allow us to calculate derivatives. For examples, we refer the reader to Schorfheide (2005) as well as Amisano and Tristani (2011), who derive the exact likelihood of a second-order approximation for a class of models with regime-switching.

Below, we use a toy example to illustrate how regime switching models can generate weak identification, where to simplify the treatment we abstract from time-series behavior and consider an i.i.d. model.

We assume that we have a sample $X_t, t = 1, \dots, T$ drawn i.i.d. from the distribution

$$f(\cdot; \varphi_1, \varphi_2, \delta) = \delta f(\cdot; \varphi_1) + (1 - \delta) f(\cdot; \varphi_2),$$

where the one-dimensional parameters φ_1 and φ_2 belong to an open set Ω . To resolve the “label-switching” problem, assume that $0 < \delta < 1/2$. Consider a weak identification embedding in which the parameters φ_1 and δ are fixed while the parameter $\varphi_{2,T} = \varphi_1 + \frac{c}{\sqrt{T}}$ is drifting to the point of non-identification ($\varphi_1 = \varphi_2$).

Assume that for almost every realization of X_t the cdf $f(X_t; \varphi)$ is four times continuously differentiable in $\varphi \in \Omega$. Assume further that there exists a random variable η with the finite second moment such that almost surely

$$\max_{i=1, \dots, 4} \left\{ \left| \frac{f(X_t, \varphi)}{f(X_t, \varphi_1)} \right|, \left| \frac{f^{(i)}(X_t, \varphi)}{f(X_t, \varphi_1)} \right| \right\} \leq \eta$$

for all $\varphi \in \Omega$, where $f^{(i)}$ stands for i -th derivative with respect to φ . We also assume that $f^{(i)}(X_t, \varphi_1)$ for $i \in \{1, 2, 3\}$ are linearly independent random variables under $f(X_t, \varphi_1)$.

S5.1 Assumption 1.

The score is

$$S_T = \sum_{t=1}^T \frac{1}{\delta f(X_t; \varphi_1) + (1-\delta)f(X_t; \varphi_2)} \begin{pmatrix} \delta f^{(1)}(X_t; \varphi_1) \\ (1-\delta)f^{(1)}(X_t; \varphi_2) \\ f(X_t; \varphi_1) - f(X_t; \varphi_2) \end{pmatrix} =$$

$$= \sum_{t=1}^T \frac{1}{\omega_t} \begin{pmatrix} \delta f^{(1)}(X_t; \varphi_1) \\ (1-\delta) \left(f^{(1)}(X_t; \varphi_1) + f^{(2)}(X_t; \varphi_1) \frac{C}{\sqrt{T}} + \frac{1}{2} f^{(3)}(X_t; \varphi_1) \frac{C^2}{T} + O_p(T^{-3/2}) \right) \\ -f^{(1)}(X_t; \varphi_1) \frac{C}{\sqrt{T}} - f^{(2)}(X_t; \varphi_1) \frac{C^2}{2T} - \frac{1}{6} f^{(3)}(X_t; \varphi_1) \frac{C^3}{T^{3/2}} + O_p(T^{-2}) \end{pmatrix},$$

where $\omega_t = \delta f(X_t; \varphi_1) + (1-\delta)f(X_t; \varphi_2)$. We may notice that

$$\begin{pmatrix} \frac{1}{\delta\sqrt{T}} & 0 & 0 \\ \frac{2}{\delta} & \frac{1}{1-\delta} & \frac{3\sqrt{T}}{C} \\ \frac{C\sqrt{T}}{2\delta} & \frac{C\sqrt{T}}{2(1-\delta)} & T \end{pmatrix} S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{\omega_t} \begin{pmatrix} f^{(1)}(X_t; \varphi_1) \\ -\frac{C}{2} f^{(2)}(X_t; \varphi_1) + O_p(T^{-1/2}) \\ \frac{C^3}{12} f^{(3)}(X_t; \varphi_1) + O_p(T^{-1/2}) \end{pmatrix}.$$

Let us define

$$K_T = \begin{pmatrix} \frac{1}{\delta\sqrt{T}} & 0 & 0 \\ \frac{2}{\delta} & \frac{1}{1-\delta} & \frac{3\sqrt{T}}{C} \\ \frac{C\sqrt{T}}{2\delta} & \frac{C\sqrt{T}}{2(1-\delta)} & T \end{pmatrix}$$

then by the Law of Large Numbers

$$K_T J_T K_T' \rightarrow^p E \left\{ \frac{1}{\omega_t^2} \begin{pmatrix} f^{(1)}(X_t; \varphi_1) \\ -\frac{C}{2} f^{(2)}(X_t; \varphi_1) \\ \frac{C^3}{12} f^{(3)}(X_t; \varphi_1) \end{pmatrix} \begin{pmatrix} f^{(1)}(X_t; \varphi_1) \\ -\frac{C}{2} f^{(2)}(X_t; \varphi_1) \\ \frac{C^3}{12} f^{(3)}(X_t; \varphi_1) \end{pmatrix}' \right\}$$

where the limit is a finite positive definite matrix. We also may notice that the summands $K_T s_{T,t}$ satisfy Lindeberg's condition. As a result Assumption 1 of the paper is satisfied.

S5.2 Hessian

Now, let us look at the Hessian I_T . One can show that

$$J_T - I_T = \sum_{t=1}^T \frac{1}{\omega_t} \begin{pmatrix} \delta f^{(2)}(X_t, \varphi_1) & 0 & f^{(1)}(X_t, \varphi_1) \\ 0 & (1 - \delta)f^{(2)}(X_t, \varphi_2) & -f^{(1)}(X_t, \varphi_2) \\ f^{(1)}(X_t, \varphi_1) & -f^{(1)}(X_t, \varphi_2) & 0 \end{pmatrix}.$$

From the logic of the information equality it follows that

$$E \left(\frac{f^{(1)}(X_t, \varphi)}{\omega_t} \right) = E \left(\frac{f^{(2)}(X_t, \varphi)}{\omega_t} \right) = 0$$

for any φ . Thus we have the following Central Limit Theorem:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{\omega_t} (f^{(1)}(X_t, \varphi_1), f^{(2)}(X_t, \varphi_1)) \Rightarrow (\xi_1, \xi_2)$$

where (ξ_1, ξ_2) is a Gaussian vector with the covariance matrix $E \begin{pmatrix} \left(\frac{f^{(1)}}{f}\right)^2 & \frac{f^{(1)}f^{(2)}}{f^2} \\ \frac{f^{(1)}f^{(2)}}{f^2} & \left(\frac{f^{(2)}}{f}\right)^2 \end{pmatrix}$.

Further

$$\frac{1}{\sqrt{T}}(J_T - I_T) \Rightarrow \begin{pmatrix} \delta\xi_2 & 0 & \xi_1 \\ 0 & (1 - \delta)\xi_2 & -\xi_1 \\ \xi_1 & -\xi_1 & 0 \end{pmatrix}$$

from which it is easy to see that the matrix $K_T(J_T - I_T)K_T'$ is asymptotically explosive, and thus that I_T and J_T have asymptotically different behavior.

S6 A simplified non-linear model.

In this section we discuss an analytically solvable model with regime-switching that may suffer from identification issues.

Schorfheide (2005) discusses a model with learning and monetary policy shifts, whose log-linearized equilibrium conditions can be written:

$$\begin{cases} x_t = E_t x_{t+1} - \tau (r_t - E_t \pi_{t+1}) - E_t \Delta g_{t+1} + \tau E_t z_{t+1}, \\ \pi_t = \beta E_t \pi_{t+1} + \kappa (x_t - g_t), \\ r_t = (1 - \rho_r) \psi \pi_t + \rho_r r_{t-1} + (1 - \rho_r) (1 - \psi) \pi_t^*(s_t) + \varepsilon_{r,t} \end{cases}$$

and

$$\begin{pmatrix} \varepsilon_{g,t} \\ \varepsilon_{z,t} \\ \varepsilon_{r,t} \end{pmatrix} \sim N \left(0, \begin{bmatrix} \sigma_g^2 & 0 & 0 \\ 0 & \sigma_z^2 & 0 \\ 0 & 0 & \sigma_r^2(s_t) \end{bmatrix} \right),$$

where $s_t \in \{1, 2\}$ is an unobserved state that evolves exogenously according to a first order Markov chain with transition matrix

$$\mathcal{P} = \begin{bmatrix} \phi_1 & 1 - \phi_2 \\ 1 - \phi_1 & \phi_2 \end{bmatrix}.$$

Two parameters $\pi_t^*(s_t)$ and $\sigma_r^2(s_t)$ are functions of the state variable.

To solve the model analytically we make a few simplifying assumptions. In particular, we assume that $\pi_t^*(1) = \pi_t^*(2) = 0$, so there is no change in the target inflation across states. Let us further assume that $\tau = 1$, $\rho_r = 0$ and $\psi = \frac{1}{\beta}$. Under these assumptions the model becomes

$$\begin{cases} x_t = E_t x_{t+1} - r_t + E_t \pi_{t+1} + (1 - \rho_g) g_t + \rho_z z_t, \\ \pi_t = \beta E_t \pi_{t+1} + \kappa (x_t - g_t), \\ r_t = \frac{1}{\beta} \pi_t + \varepsilon_{r,t} \end{cases}$$

where the only state-dependence is regime-switching in the variance of $\varepsilon_{r,t}$. We have used the fact that $E_t z_{t+1} = \rho_z z_t$ and $E_t \Delta g_{t+1} = E_t [g_{t+1} - g_t] = (\rho_g - 1) g_t$.

We can solve this model forward in the same manner as the DSGE example in Section S1. We can write the solution in the following form:

$$Y_t = \begin{pmatrix} x_t \\ \pi_t \\ r_t \end{pmatrix} = \begin{bmatrix} 1 & \frac{\beta \rho_z}{\kappa + \beta - \beta \rho_z} & -\frac{\beta}{\kappa + \beta} \\ 0 & \frac{\beta^2 \kappa \rho_z}{(\kappa + \beta - \beta \rho_z)(1 - \beta \rho_z)} & -\frac{\beta \kappa}{\kappa + \beta} \\ 0 & \frac{\beta \kappa \rho_z}{(\kappa + \beta - \beta \rho_z)(1 - \beta \rho_z)} & \frac{\beta}{\kappa + \beta} \end{bmatrix} \begin{pmatrix} g_t \\ z_t \\ \varepsilon_{r,t} \end{pmatrix}.$$

S6.1 Identification failure

Let us impose that $0 < \beta, \rho_g, \rho_z < 1$, $\kappa > 0$, and assume all variances are strictly positive. Note that conditional on the state s_t

$$\text{Var}(Y_t | s_t) = C(\theta) \begin{bmatrix} \frac{\sigma_g^2}{1 - \rho_g} & 0 & 0 \\ 0 & \frac{\sigma_z^2}{1 - \rho_z} & 0 \\ 0 & 0 & \sigma_r^2(s_t) \end{bmatrix} C(\theta)'$$

while the auto-covariance of Y_t with Y_{t-j} for $j > 0$ is

$$\text{Cov}(Y_t, Y_{t-j} | s_t) = C(\theta) \begin{bmatrix} \rho_g^j \frac{\sigma_g^2}{1-\rho_g} & 0 & 0 \\ 0 & \rho_z^j \frac{\sigma_z^2}{1-\rho_z} & 0 \\ 0 & 0 & 0 \end{bmatrix} C(\theta)'$$

The state s_t has no effect on the auto-covariance of Y_t , but instead matters only through the variance. In the special case where the variance of $\varepsilon_{r,t}$ is the same across the two states, $\sigma_r^2(1) = \sigma_r^2(2)$, the state has no effect on the covariance structure of $\{Y_t\}_{t=1}^\infty$. Since $\{Y_t\}_{t=1}^\infty$ is jointly normal in this case, the covariance function is sufficient for all parameters, so this implies that for $\sigma_r^2(1) = \sigma_r^2(2)$ the state transition probabilities ϕ_1 and ϕ_2 are unidentified.

S7 Proof of Lemma 2 from the paper

Take any $\varepsilon > 0$,

$$\begin{aligned} \left| K_{i,T} K_{j,T} K_{l,T} \sum_{t=1}^T m_{i,t} m_{j,t} m_{l,t} \right| &\leq \max_t |K_{i,T} m_{i,t}| \left| K_{j,T} K_{l,T} \sum_{t=1}^T m_{j,t} m_{l,t} \right| = \\ &= \max_t |K_{i,T} m_{i,t}| |K_{j,T} K_{l,T} [M_j, M_l]_T|. \end{aligned}$$

Assumption 3(b) implies that $K_{j,T} K_{l,T} [M_j, M_l]_T \xrightarrow{p} \Sigma_{j,l}$ is bounded in probability.

$$\begin{aligned} E \left(\max_t |K_{i,T} m_{i,t}| \right) &\leq \varepsilon + E \left(K_{i,T} \max_t |m_{i,t}| \mathbb{I}\{|K_{i,T} m_{i,t}| > \varepsilon\} \right) \leq \\ &\leq \varepsilon + \sum_t E(K_{i,T} |m_{i,t}| \mathbb{I}\{|K_{i,T} m_{i,t}| > \varepsilon\}). \end{aligned}$$

The last term converges to 0 by Assumption 3(a). \square

S8 References

- Andrews, D.W.K., and X. Cheng (2012): “Estimation and Inference with Weak, Semi-strong and Strong Identification,” *Econometrica*, 80(5), 2153-2211.
- Amisano, G. and O. Tristani (2011): “Exact likelihood computation for nonlinear DSGE models with heteroskedastic innovations,” *Journal of Economic Dynamics and Control*, 35(12), 2167-2185.

- Fernández-Villaverde J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): “ABCs (and Ds) of Understanding VARs,” *American Economic Review*, 97(3), 1021-1026.
- Komunjer, I., and S. Ng (2011): “Dynamic Identification of Dynamic Stochastic General Equilibrium Models,” *Econometrica*, 79(6), 1995-2032.
- Schorfheide, F. (2005): “Learning and Monetary Policy Shifts,” *Review of Economic Dynamics*, 8, 392-419.
- Stock, J. H., and J.H. Wright (2000): “GMM With Weak Identification,” *Econometrica*, 68, 1055-1096.