

14.384 Time Series Analysis, Fall 2008  
 Recitation by Paul Schrimpf  
 Supplementary to lectures given by Anna Mikusheva  
 November 7, 2008

## Recitation 10

### GMM Estimation of the NKPC

One popular use of GMM in applied macro has been estimating the Neo-Keynesian Phillips Curve. An important example is Galí and Gertler (1999). This is an interesting paper because it involves a good amount of macroeconomics, validated a model that many macroeconomists like, and (best of all for econometricians) has become a leading example of weak identification in GMM. These notes describe Galí and Gertler's paper, then give a quick overview of identification robust inference in GMM, and finally describe the results of identification robust procedures for Galí and Gertler's models.

#### Deriving the NKPC

You should have seen this in macro, so I'm going to go through it quickly. Suppose there is a continuum of identical firms that sell differentiated products to a representative consumer with Dixit-Stiglitz preferences over the goods. Prices are sticky in the sense of Calvo (1983). More specifically, each period each firm has a probability of  $1 - \theta$  of being able to adjust its price each period. If  $p_t^*$  is the log price chosen by firms that adjust at time  $t$ , then the evolution of the log price level will be

$$p_t = \theta p_{t-1} + (1 - \theta) p_t^* \quad (1)$$

The first order condition (or maybe a first order approx to the first order condition) for firms that get to adjust their price at time  $t$  is

$$p_t^* = (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t[mc_{t+k}^n + \mu] \quad (2)$$

where  $mc_t^n$  is log nominal marginal cost at time  $t$ , and  $\mu$  is a markup parameter that depends on consumer preferences. This first order condition can be rewritten as:

$$\begin{aligned} p_t^* &= (1 - \beta\theta) mc_t^n + (1 - \beta\theta) \sum_{k=1}^{\infty} (\beta\theta)^k E_t[mc_{t+k}^n + \mu] \\ &= (1 - \beta\theta)(\mu + mc_t^n) + (1 - \beta\theta)\beta\theta E_t p_{t+1}^* \end{aligned}$$

Substituting in  $p_t^* = \frac{p_t - \theta p_{t-1}}{1 - \theta}$  gives:

$$\begin{aligned} \frac{p_t - \theta p_{t-1}}{1 - \theta} &= \frac{1 - \beta\theta}{1 - \theta} E_t[p_{t+1} - \theta p_t] + (1 - \beta\theta)(\mu + mc_t^n) \\ p_t - p_{t-1} &= \beta E_t[p_{t+1} - p_t] + \frac{(1 - \theta)(1 - \theta\beta)}{\theta} (\mu + mc_t^n - p_t) \\ \pi_t &= \beta E_t[\pi_{t+1}] + \lambda(\mu + mc_t^n - p_t) \end{aligned} \quad (3)$$

(4)

This is the NKPC. Inflation depends on expected inflation and real marginal costs (or the deviation of log marginal costs from the steady state. In the steady state  $\mu - p = mc$ ).

## Estimation

**Using the Output Gap** Since real marginal costs are difficult to observe, people have noted that in a model without capital,  $mc_t^r - p_t \approx \kappa x_t$  where  $x_t$  is the output gap (the difference between current output and output in a model without price frictions). This suggests estimating:

$$\beta\pi_t = \pi_{t-1} - \lambda\kappa x_t - \lambda\mu + \epsilon_t$$

When estimating this equation, people general find that  $-\widehat{\lambda\kappa}$  is positive, contradicting the model.

**GG** Galí and Gertler (1999) argued that there at least two problems with this model: (i) the output gap is hard to measure and (ii) the output gap may not be proportional to real marginal costs. Galí and Gertler argue that the labor income share is a better proxy for real marginal costs. With a Cobb-Douglas production function,

$$Y_t = A_t K_t^{\alpha_k} L_t^{\alpha_l}$$

marginal cost is the ratio of the wage to the marginal product of labor,

$$\begin{aligned} MC_t &= \frac{W_t}{P_t(\partial Y_t / \partial Y)} = \frac{W_t L_t}{P_t \alpha_l Y_t} \\ &= \frac{1}{\alpha_l} S_{Lt} \end{aligned}$$

Thus the deviation of log marginal cost from its steady state should equal the deviation of log labor share from its steady state,  $mc_t = s_t$ . This leads to moment conditions:

$$E_t[(\pi_t - \lambda s_t - \beta\pi_{t+1})z_t] = 0 \quad (5)$$

$$E_t[(\theta\pi_t - (1-\theta)(1-\beta\theta)s_t - \theta\beta\pi_{t+1})z_t] = 0 \quad (6)$$

where  $z_t$  are any variables in firms' information sets at time  $t$ . As instruments, Galí and Gertler use four lags of inflation, the labor income share, the output gap, the long-short interest rate spread, wage inflation, and commodity price inflation. Galí and Gertler estimate this model and find values of  $\beta$  around 0.95,  $\theta$  around 0.85, and  $\lambda$  around 0.05. In particular,  $\lambda > 0$  in accordance with the theory unlike when using the output gap. The estimates of  $\theta$  are a bit high. They imply an average price duration of five to six quarters, which is much higher than observed in the micro-data of Bils and Klenow (200?).

## Hybrid Philips Curve

The NKPC implies that price setting behavior is purely forward looking. All inflation inertia comes from price stickiness in this model. One might be concerned whether this is enough to capture the observed dynamics of inflation. To answer this question, Galí and Gertler consider a more general model that allows for backward looking behavior. In particular, they assume that a fraction,  $\omega$  of firms set prices equal to the optimal price last period plus an inflation adjustment:  $p_t^b = p_{t-1}^* + \pi_{t-1}$ . The rest of the firms behave optimally. This leads to the following inflation equation:

$$\begin{aligned} \pi_t &= \frac{(1-\omega)(1-\theta)(1-\beta\theta)mc_t + \beta\theta E_t \pi_{t+1} + \omega\pi_{t-1}}{\theta + \omega(1-\theta(1-\beta))} \\ &= \lambda mc_t + \gamma^f E_t \pi_{t+1} + \gamma^b \pi_{t-1} \end{aligned} \quad (7)$$

As above, Galí and Gertler estimate this equation using GMM. They find  $\hat{\omega} \approx 0.25$  with a standard error of 0.03, so a purely forward looking model is rejected. Their estimates of  $\theta$  and  $\beta$  are roughly the same as above.

## Identification Issues

Galí and Gertler note that they can write their moment condition in many ways, for example the HNKPC could be estimated from either of the following moment conditions:

$$E_t [((\theta + \omega(1 - \theta(1 - \beta)))\pi_t - (1 - \omega)(1 - \theta)(1 - \beta)\theta s_t - \beta\theta\pi_{t+1} - \omega\pi_{t-1}) z_t] = 0 \quad (8)$$

$$E_t \left[ \left( \pi_t - \frac{(1 - \omega)(1 - \theta)(1 - \beta)\theta}{\theta + \omega(1 - \theta(1 - \beta))} s_t - \frac{\beta\theta}{\theta + \omega(1 - \theta(1 - \beta))} \pi_{t+1} - \frac{\omega}{\theta + \omega(1 - \theta(1 - \beta))} \pi_{t-1} \right) z_t \right] = 0 \quad (9)$$

Estimation based on these two moment conditions gives surprisingly different results. In particular, (8) leads to an estimate of  $\omega$  of 0.265 with a standard error of 0.031, but (9) leads to an estimate of 0.486 with a standard error of 0.040. If the model is correctly specified and well-identified, the two equations should, asymptotically, give the same estimates. The fact that the estimates differ suggests that either the model is misspecified or not well identified.

## Analyzing Identification

There's an old literature about analyzing identification conditions in rational expectations models. Pesaran (1987) is the classic paper that everyone seems to cite, but I have not read it. Anyway, the idea is to solve the rational expectations model (7) to write it as an autoregression, write down a model for  $s_t$  to complete the system, and then analyze identification using familiar SVAR or simultaneous equation tools. I will follow Mavroeidis (2005). Another paper that does this is Nason and Smith (2002). Solving (7) and writing an equation for  $s_t$  gives a system like:

$$\pi_t = D(L)\pi_{t-1} + A(L)s_t + \epsilon_t \quad (10)$$

$$s_t = \rho(L)s_{t-1} + \phi(L)\pi_{t-1} + v_t \quad (11)$$

$D(L)$  and  $A(L)$  are of order the maximum of 1 and the order of  $\rho(L)$  and  $\phi(L)$  respectively. An order conditions for identification is that the order of  $\rho(L)$  plus  $\phi(L)$  is at least two, so that you have at least two valid instruments to instrument for  $s_t$  and  $\pi_{t+1}$  in (7). This condition can be tested by estimating (11) and testing whether the coefficients are 0. Mavroeidis does this and finds a p-value greater than 30%, so non-identification is not rejected. Mavroeidis then picks a wide range of plausible values for the parameters in the model and calculates the concentration parameter for these parameters. He finds that concentration parameter is often very close to zero. Recall from 382 that in IV, a low concentration parameter indicates weak instrument problems.

## Weak Identification in GMM

As with IV, when a GMM model is weakly identified, the usual asymptotic approximations work poorly. Fortunately, there are alternative inference procedures that perform better.

**GMM Bias** <sup>1</sup> The primary approaches are based on the CUE (continuously updating estimator) version of GMM. To understand why, it is useful to write down the approximate finite sample bias of GMM. If our moment conditions are  $g(\beta) = \sum g_i(\beta)/T$  and  $\Omega(\beta) = E[g_i(\beta)g_i(\beta)']$  (in the iid case, for time series replace with an appropriate auto-correlation consistent type estimator) CUE minimizes:

$$\hat{\beta} = \arg \min g(\beta)' \Omega(\beta)^{-1} g(\beta)$$

That is, rather than plugging in a preliminary estimate of  $\beta$  to find the weighting matrix, CUE continuously updates the weighting matrix as a function of  $\beta$ . Suppose we used a fixed weighting matrix,  $A$  and do GMM.

<sup>1</sup>This section is based on Whitney's notes from 386.

What is the expectation of the objective function? Well, for iid data (if observations are correlated, we will get an even worse bias) we have:

$$\begin{aligned} E[g(\beta)'Ag(\beta)] &= E\left[\sum_{i,j} g_i(\beta)'Ag_j(\beta)/T^2\right] \\ &= \sum_{i \neq j} E[g(\beta)]AE[g(\beta)]/T^2 + \sum_i E[g_i(\beta)'Ag_i(\beta)]/n^2 \\ &= (1 - T^{-1})E[g(\beta)]AE[g(\beta)] + tr(A\Omega(\beta))T^{-1} \end{aligned}$$

The first term is the population objective function, so it is minimized at  $\beta_0$ . The second term, however, is not generally minimized at  $\beta_0$ , causing  $E[\hat{\beta}_T] \neq \beta$ . However, if we use  $A = \Omega(\beta)^{-1}$ , then the second term vanishes and we have an unbiased estimator. This is sort of what CUE does. It is not exactly since we use  $\hat{\Omega}(\beta)$  instead of  $\Omega(\beta)$ . Nonetheless, it can be shown to be less biased than two-step GMM. See Newey and Smith (2004).

Another view of the bias can be obtained by comparing the first order conditions of CUE and two-step GMM. The first order condition for GMM is

$$0 = G(\beta)\hat{\Omega}(\tilde{\beta})^{-1}g(\beta) \tag{12}$$

where  $G(\beta) = \frac{\partial g}{\partial \beta} = \sum \frac{\partial g_i}{\partial \beta}$  and  $\tilde{\beta}$  is the first step estimate of  $\beta$ . This term will have bias because the  $i$ th observation in the sum used for  $G$ ,  $\hat{\Omega}$ , and  $g$  will be correlated. Compare this to the first order condition for CUE:

$$\begin{aligned} 0 &= G(\beta)\hat{\Omega}(\beta)^{-1}g(\beta) - g(\beta)\hat{\Omega}(\beta)^{-1}\left(\sum\left(\frac{\partial g_i}{\partial \beta}g_i(\beta)' + g_i(\beta)\frac{\partial g_i'}{\partial \beta}\right)/T\right)\hat{\Omega}(\beta)^{-1}g(\beta) \\ &= \left[G(\beta) - \left(\sum\frac{\partial g_i}{\partial \beta}g_i(\beta)'\right)\hat{\Omega}(\beta)^{-1}\right]\hat{\Omega}(\beta)^{-1}g(\beta) \end{aligned}$$

The term in brackets is the projection of  $G(\beta)$  onto the space orthogonal to  $g(\beta)$ . Hence, the term in brackets is uncorrelated with  $g(\beta)$ . This reduces bias.<sup>2</sup>

**Identification Robust Inference** The lower bias of CUE suggests that inference based on CUE might be more robust to small sample issues than traditional GMM inference. This is indeed the case. Stock and Wright (2000) showed that under  $H_0 : \beta = \beta_0$  the CUE objective function converges to a  $\chi_m^2$  where  $m$  is the number of moment conditions. Moreover, this convergence occurs whether the model is strongly, weakly<sup>3</sup>, or non-identified. Some authors call the CUE objective function the  $S$ -statistic. Others call it the  $AR$ -statistic because in linear models, the  $AR$  statistic is the same as the CUE objective function. The  $S$ -stat has the same properties as the  $AR$ -stat discussed in 382. Importantly, its degrees of freedom grows with the number of moments, so it may have lower power in very over identified models. Also, an  $S$ -stat test may reject either because  $\beta \neq \beta_0$  or because the model is misspecified. This can lead to empty confidence sets.

The Kleibergen (2005) developed an analog of the Lagrange Multiplier that, like the  $S$ -stat, has the same limiting distribution regardless of identification. The LM stat is based on the fact that under  $H_0 : \beta = \beta_0$ , the derivative of the objective function at  $\beta$  should be approximately zero. Kleibergen applies this principal to the CUE objective function. Let  $\hat{D}(\beta) = G(\beta) - \widehat{acov}(G(\beta), g(\beta))\hat{\Omega}(\beta)^{-1}$  (as above for iid data,  $\widehat{acov}(G(\beta), g(\beta)) = \sum \frac{\partial g_i}{\partial \beta}g_i(\beta)'$ ). Kleibergen's statistic is

$$KLM = g(\beta)\hat{\Omega}(\hat{\beta})^{-1}\hat{D}(\hat{\beta})(\hat{D}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{D}(\hat{\beta}))^{-1}\hat{D}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}g(\hat{\beta}) \xrightarrow{d} \chi_p^2 \tag{13}$$

<sup>2</sup>There is still some bias due to parts of  $\hat{\Omega}$  being correlated with  $g$  and  $G$ .

<sup>3</sup>Defining weak GMM asymptotics involves introducing a bunch of notation, so I'm not going to go through it. The idea is essentially the same as in linear models. See Stock and Wright (2000) for details.

Figure 1: Kleibergen and Mavroeidis Power Curve

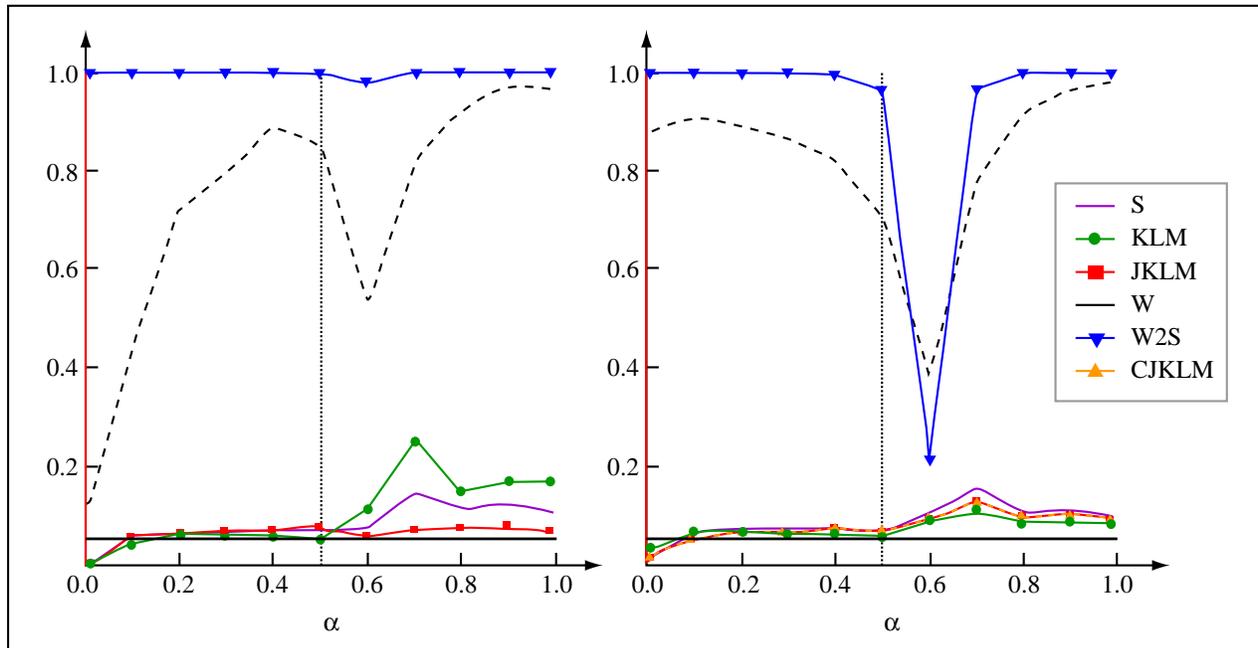


Figure by MIT OpenCourseWare.

It is asymptotically  $\chi^2$  with  $p = (\text{number of parameters})$  degrees of freedom. The degrees of freedom of KLM does not depend on the degree of overidentification. This can give it better power properties than the AR/S stat. However, since it only depends on the first order condition, in addition to being minimized at the minimum of the CUE objective function, it will also be minimized at local minima and maxima and inflection points. This property leads Kleibergen to consider an identification robust version of the Hansen’s J-statistic for testing overidentifying restriction. Kleibergen’s J is

$$J(\beta) = S(\beta) - KLM(\beta) \xrightarrow{d} \chi_{m-p}^2 \tag{14}$$

Moreover,  $J$  is asymptotically independent of  $KLM$ , so you can test using both of them, yielding a joint test with size  $\alpha = \alpha_J + \alpha_K - \alpha_J\alpha_K$ .

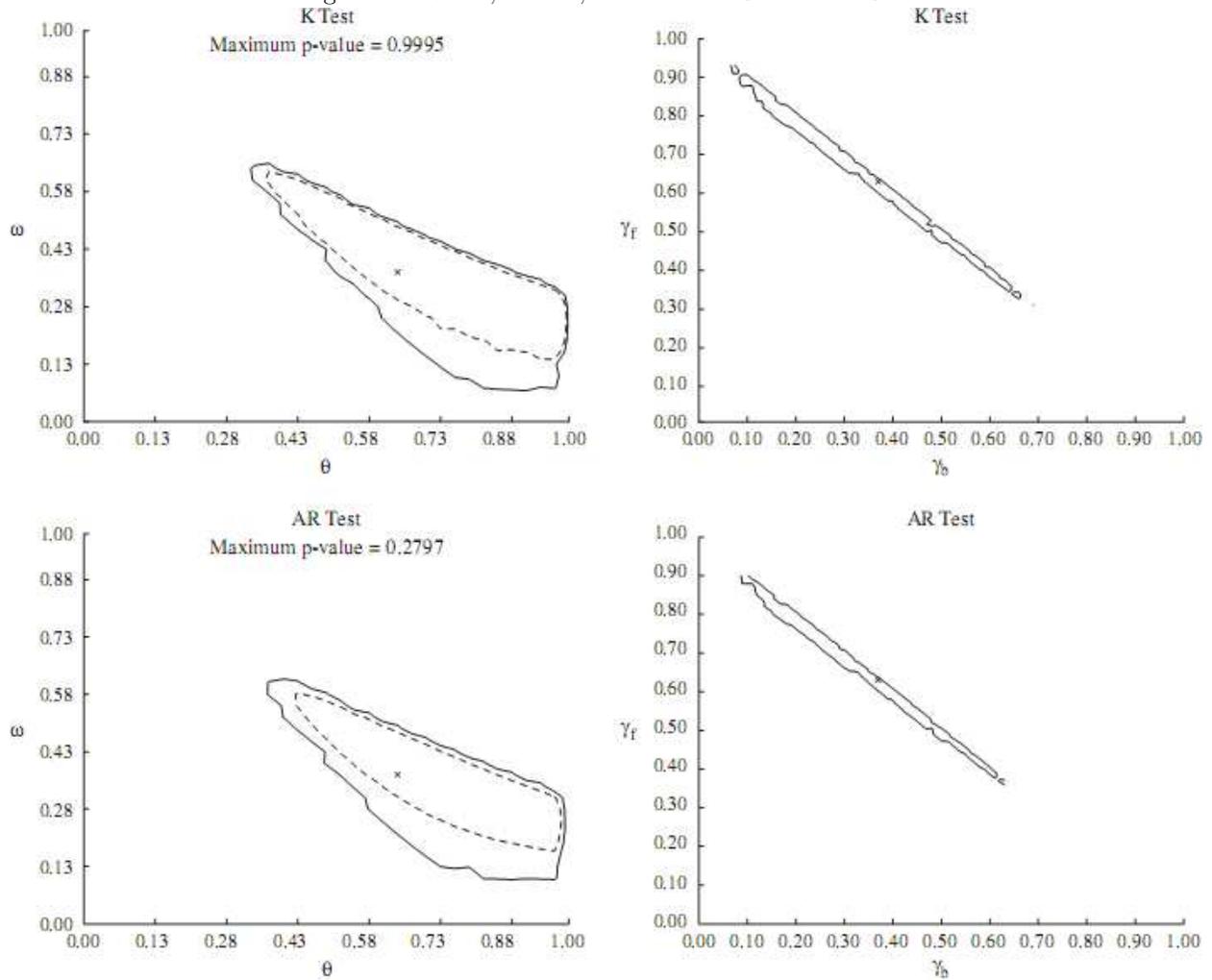
If you have a great memory, you might also remember Moreira’s conditional likelihood ratio test from covering weak instruments in 382. There’s also a GMM version of this test discussed in Kleibergen (2005).

**Results of Weak Identification Robust Inference for HNKPC**

**Kleibergen and Mavroeidis** Kleibergen and Mavroeidis (2008) extend Kleibergen’s tests described above, which only work for testing a the full set of parameters, to tests for subsets of parameters. As an application, Kleibergen and Mavroeidis (2008) simulate a HNKPC model and consider testing whether the fraction of backward looking firms (which they call  $\alpha$ , but GG and I call  $\omega$ ) equals one half. Figure 1 shows the frequency of rejection for various true values of  $\alpha$ . The Wald test badly overrejects when the true  $\alpha$  is 1/2. The KLM and JKLM have the correct size under  $H_0$ , but they also have no power against any of the alternatives. It looks like identification is a serious issue.

**Dufour, Khalaf, and Kichian (2006)** Use the  $AR$  and  $K$  statistics to construct confidence sets for Galí and Gertler’s model. Figure 2 shows the results. The confidence sets are reasonably informative. The point estimates imply an average price duration of 2.75 quarters, which is much closer to the micro-data evidence

Figure 2: Dufour, Khalaf, and Kichian Confidence Sets



Courtesy Elsevier, Inc., <http://www.sciencedirect.com>. Used with permission.

(Bils and Klenow's average is 1.8 quarters) than Galí and Gertler's estimator. Also, although not clear from this figure, Dufour, Khalaf, and Kichian find that Galí and Gertler's point estimates lie outside their 95% confidence sets.

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.384 Time Series Analysis  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.