

NCER Working Paper Series

Weak Instruments: A Guide to the Literature

Adrian Pagan

Working Paper #13

April 2007

Weak Instruments: A Guide to the Literature*

Adrian Pagan
School of Economics and Finance
Queensland University of Technology

March 2007

Contents

1	Introduction	2
2	Distributional Problems of the IV Estimator with Weak Instruments	2
3	Detection of Weak Instruments	6
4	Inferences with the IV Estimator	8
5	Constructing Useful Test Statistics	9
6	Examples of Weak Instruments	13
6.1	Estimating the Phillips Curve in a New Keynesian Model . . .	13
6.2	Euler Equations for Inventory Problems	14
7	References	17

Abstract

Weak instruments have become an issue in many contexts in which econometric methods have been used. Some progress has been made

*Research Supported by ESRC Grant 000 23-0244. This is an English version of a paper published in Russian in *Quantile*, Issue #2, March 2007.

into how one diagnoses the problem and how one makes an allowance for it. The present paper gives a partial survey of this literature, focussing upon some of the major contributions and trying to provide a relatively simple exposition of the proposed solutions.

1 Introduction

These notes are meant to be a selective guide to the literature on weak instruments. They use simple models to illustrate the problems raised by weak instruments and the suggestions that have been made to solve them. Because the literature is one that is still developing we only briefly touch on some recent developments. The origin of the present form of these lectures were the courses Economics 4202 at the University of New South Wales in 2002-4 and Economics 607 at Johns Hopkins University in 2004.

2 Distributional Problems of the IV Estimator with Weak Instruments

Consider the simple model

$$y_t = x_t\theta + u_t$$

where u_t is *i.i.d.* $(0, \sigma_u^2)$, $E(x_t u_t) \neq 0$, and there exist variables z_t , instruments, such that $E(z_t u_t) = 0$. We assume that

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{zz} & \sigma_{zx} \\ \sigma_{xz} & \sigma_{xx} \end{bmatrix} \right)$$

Thus there is a single regressor and a single instrument. Then we have the simple IV estimator.

$$\begin{aligned}
\hat{\theta} - \theta_0 &= \left(\sum z_t x_t \right)^{-1} \left(\sum z_t u_t \right) \\
&= \left(\frac{\frac{1}{T} \sum z_t x_t}{\hat{\sigma}_z \hat{\sigma}_x} \right)^{-1} \frac{1}{\hat{\sigma}_z \hat{\sigma}_x} \frac{1}{T} \sum z_t u_t \\
&= \left(\frac{\frac{1}{T} \sum z_t x_t}{\hat{\sigma}_z \hat{\sigma}_x} \right)^{-1} \left(\frac{\frac{1}{T} \sum z_t u_t}{\hat{\sigma}_z \sigma_u} \right) \frac{\sigma_u}{\hat{\sigma}_x} \\
&= \hat{\rho}_{zx}^{-1} \hat{\rho}_{zu} \cdot \frac{\sigma_u}{\hat{\sigma}_x}. \tag{1}
\end{aligned}$$

where $\hat{\sigma}_z$, $\hat{\sigma}_x$ are the estimated standard deviations of z_t and x_t and the $\hat{\rho}'s$ are the estimated correlation coefficients ($\hat{\rho}_{zu}$ involves the true standard deviation of u_t , σ_u). From this expression the distribution of $\hat{\theta} - \theta_0$ depends on the product of three random variables. Of these, we would expect $\sigma_u/\hat{\sigma}_x$ to converge quickly to a constant, leaving $\hat{\rho}_{zu}/\hat{\rho}_{zx}$ as the random variable to be studied. Any problems with the distribution of $\hat{\theta}$ clearly come from $\hat{\rho}_{zx}$; specifically the issue is how random $\hat{\rho}_{zx}$ is and whether small realizations of this random variable are possible. If we get such realizations then, because $\hat{\rho}_{zx}$ is on the denominator, we get a big value for $\hat{\theta} - \theta_0$, and this will tend to produce a skewed density for $\hat{\theta} - \theta_0$.

Now asymptotic theory proceeds by assuming that the sample size is large enough to treat $\hat{\rho}_{zx}$ as a constant i.e. we would normally look at

$$T^{1/2}(\hat{\theta} - \theta_0) = \hat{\rho}_{zx}^{-1} (T^{1/2} \hat{\rho}_{zu}) \cdot \frac{\sigma_u}{\hat{\sigma}_x}$$

and, because $T^{1/2}(\hat{\rho}_{zu} - \rho_{zu}) = T^{1/2} \hat{\rho}_{zu}$, we expect that $T^{1/2} \hat{\rho}_{zu}$ would be asymptotically normally distributed as $N(0, 1)$ (when $\rho_{zu} = 0$). We also expect that the other quantities all converge to their true values. So, provided $\rho_{zx} \neq 0$, we find that $T^{1/2}(\hat{\theta} - \theta_0)$ will be $N(0, \frac{\sigma_u^2}{\sigma_{zx} \rho_{zx}^2})$.

Now things get messy when $\rho_{zx} = 0$ since, if we followed the approach above, then we would be dividing by zero in large samples. In this case, when $\rho_{zx} = 0$, the instruments are said to be *irrelevant*. They are *valid* instruments, since $\rho_{zu} = 0$, but they are of little use to us. But we can say a bit more than that. In particular let us assume that asymptotically the estimator of ρ_{zx} is also normally distributed around its true value (zero in this case) i.e. $T^{1/2} \hat{\rho}_{zx}$ is $N(0, v)$ (when $\rho_{zx} = 0$ then $v = 1$). Then we have

$$(\hat{\theta} - \theta_0) = (T^{1/2} \hat{\rho}_{zx})^{-1} (T^{1/2} \hat{\rho}_{zu}) \cdot \frac{\sigma_u}{\hat{\sigma}_x}$$

and so in large samples this becomes

$$\frac{N(0, 1)}{N(0, 1)} \cdot \frac{\sigma_u}{\hat{\sigma}_x}$$

From this it is clear that $\hat{\theta}$ is not a consistent estimator i.e. there is always a random gap between $\hat{\theta}$ and θ_0 . Another way of saying this is that, unlike the standard case where the $var(\hat{\theta} - \theta_0)$ goes to zero as $T \rightarrow \infty$, here the variance of $(\hat{\theta} - \theta_0)$ does not decline with the sample size. Thus this benchmark case shows that, if instruments are irrelevant, all the nice properties of IV estimators fail to hold. Moreover, if the two random variables $N(0, 1)$ and $N(0, 1)$ were independent, $(\hat{\theta} - \theta_0)$ would be distributed as a Cauchy random variable i.e. have no moments.

Now it seems unlikely that ρ_{zx} would be exactly zero. It is more likely that it will have a small value. Thus we want to look more closely at the analysis in this case. To do so we will assume that

$$x_t = z_t\pi + \xi_t$$

where $E(z_t\xi_t) = 0$. This is essentially a reduced form equation where the z_t are regarded as being exogenous. Assuming that $E(z_t) = E(x_t) = 0$ (for convenience) we have

$$E(z_tx_t) = E(z_t^2)\pi$$

so that

$$\rho_{zx} = \frac{E(z_tx_t)}{\sigma_x\sigma_z} = \frac{E(z_t^2)\pi}{\sigma_x\sigma_z} = \frac{\pi\sigma_z}{\sigma_x}$$

Now we will describe a *weak instrument* as one in which ρ_{zx} is small. This will be taken to mean that π is small.

At first glance it seems as if the asymptotic theory still applies when there are weak rather than irrelevant instruments since $\hat{\rho}_{zx}$ converges to a non-zero quantity. Indeed this is correct but one has to be nervous about such a conclusion. After all if $\rho_{zx} = 0$ the theory fails, so, if ρ_{zx} departed from zero by an extremely small amount, one has the intuition that the effects on the estimator would be more like the $\rho_{zx} = 0$ case than the $\rho_{zx} \neq 0$ case. Of course, this is really a statement about the finite sample properties of the

estimator i.e. how big does T have to be to get the asymptotic theory to cut in.

Perhaps it's useful to think about this in a heuristic way by writing $\hat{\rho}_{zx}$ as $\rho_{zx} + \eta$ where η is $N(0, v/T)$ so that we get

$$T^{1/2}(\hat{\theta} - \theta_0) = \left\{ \frac{T^{1/2}\hat{\rho}_{zx}}{\rho_{zx} + N(0, v/T)} \right\} \frac{\sigma_u}{\hat{\sigma}_x}$$

It is clear that, as T becomes large, the term $N(0, v/T)$ disappears, making realizations of the term in the denominator, $\rho_{zx} + N(0, v/T)$, become closer and closer to the non-zero value ρ_{zx} . But, for a small sample, and with v big enough, it would be possible that a realization for $\hat{\rho}_{zx}$ is produced that is close to zero. In that case a large value for $T^{1/2}(\hat{\theta} - \theta_0)$ will eventuate. In this case the asymptotic theory would not provide a good guide to the behaviour of the estimator of θ in small samples. Thus it is clear that a very small value of ρ_{zx} may mean that the sampling properties look more like the $\rho_{zx} = 0$ case than the $\rho_{zx} \neq 0$ case.

So we want to get some idea of what happens as we shrink ρ_{zx} towards zero. The situation is like that encountered with test statistics that always reject a false null in large samples i.e. are consistent. In order to compare such tests we use the idea of a local alternative i.e. the distributions of the tests are found as the alternative becomes closer to the null as the sample size increased. We will therefore do the same thing here. Specifically we will make $\pi = \frac{\phi}{\sqrt{T}}$. Then

$$\begin{aligned} T^{1/2}\hat{\rho}_{zx} &= T^{1/2}\rho_{zx} + T^{1/2}\eta \\ &= \frac{\phi\sigma_z}{\sigma_x} + T^{1/2}\eta \\ &= \frac{\phi\sigma_z}{\sigma_x} + \varepsilon \end{aligned}$$

where ε is $N(0, v)$. So now let's analyze what happens with *weak instruments*. We have

$$\begin{aligned} (\hat{\theta} - \theta_0) &= (T^{1/2}\hat{\rho}_{zx})^{-1}(T^{1/2}\hat{\rho}_{zx}) \cdot \frac{\sigma_u}{\hat{\sigma}_x} \\ &\rightarrow \frac{N(0, 1)}{N(\frac{\phi\sigma_z}{\sigma_x}, v)} \frac{\sigma_u}{\sigma_x} \end{aligned}$$

Now this is bad news since it implies that asymptotically $(\hat{\theta} - \theta_0)$ is the ratio of two random variables and so does not converge to zero i.e. $\hat{\theta}$ is not a consistent estimator of θ_0 . Moreover $(\hat{\theta} - \theta_0)$ can't be normal. Therefore, when ρ_{zx} is small, we expect that this analysis is a better predictor of the properties of $\hat{\theta} - \theta_0$ than the standard asymptotics, and that has been shown to be true in simulation experiments. Essentially it is a relatively simple way of doing small sample theory. Consequently, it is likely that the small sample distribution of $(\hat{\theta} - \theta_0)$ will be highly non-normal when there are weak instruments and there will be a substantial bias in the coefficient estimators.

3 Detection of Weak Instruments

How would we detect weak instruments?. Since the problem arises when ρ_{xz} is close to zero this suggests that we test the hypothesis that $\rho_{zx} = 0$ using $\hat{\rho}_{xz}$. But, because $\hat{\rho}_{xz}$ is just a non-zero multiple of $\hat{\pi}$, where $\hat{\pi}$ is the OLS estimate of π from the regression of x_t on z_t , we can test $\pi = 0$ instead. One could either use the t ratio for this or the F test that the regressors z_t contribute nothing to x_t .

The latter interpretation becomes important once we depart from the simple model. One important modification would be if extra regressors, w_t , appear in the equation being estimated and they also end up in the reduced form.

$$\begin{aligned} y_t &= x_t\theta + w_t\alpha + u_t \\ x_t &= z_t\pi + w_t\gamma + \xi_t \end{aligned}$$

To analyze this we note that, in matrix form, the equation becomes

$$\begin{aligned} y &= X\theta + W\alpha + u \\ X &= Z\pi + W\gamma + \xi \end{aligned}$$

and, if we pre-multiply both equations by $M_W = I - W(W'W)^{-1}W'$, we will get

$$\begin{aligned} M_W y &= M_W X\theta + u^* \\ M_W X &= M_W Z\pi + \xi^* \end{aligned}$$

i.e.

$$\begin{aligned}y^* &= X^*\theta + u^* \\X^* &= Z^*\pi + \xi^*\end{aligned}$$

so the analysis is done with y^*, X^*, Z^* instead of y, X and Z . The quantities like y^* are the residuals from the regression of y on W , and $\hat{\pi}$ will now be the estimate of π from the regression of X^* on Z^* which is the same as the estimate from the regression of X on Z and W . The F test that π is zero therefore has to allow for the presence of W in the regression i.e. it is the correlation between x_t and z_t *after w_t has been partialled out* that is the right measure of whether there are any weak instrument problems. Since the F test that $\pi = 0$ in the model with no regressors is just $\frac{1-R^2}{R^2}$, and the test when w_t is included in the equation has the same form, but with R^2 replaced by a partial R^2 , it is clear that, even when R^2 is high - x_t is well explained by z_t and w_t - the partial R^2 may be very low i.e. most of the explanation of x_t comes from w_t and not z_t . In this case the z_t are effectively weak instruments. The problem of course is that w_t is not available as an instrument for x_t as it has been “used up ” in estimating α i.e. it is needed as an instrument for itself. In fact, most of the explanation of x_t comes from w_t and not from z_t , meaning that the R^2 is a poor guide to how useful z_t are as instruments.

Staiger and Stock (1997) recommended that one should classify instruments as weak if the F test that $\pi = 0$ in the regression of x_t on z_t and w_t is less than 10.¹ In R^2 terms this means that the partial R^2 must be greater than .1. This seems a reasonable rule of thumb and has been widely used. Exactly what one should test if $\dim(X) > 1$ is less clear as then one is regressing a vector of variables x_t against a set of regressors z_t - one suggestion by Hall et al. (1996) is to use canonical correlations. Stock and Yogo (2005) have suggested that one use a multivariate analogue of the concentration coefficient and take the smallest eigenvalue of that. They tabulate conservative critical values for such a test as it is difficult to find the exact distribution.

¹As discussed in that paper this is an estimator of the concentration parameter, the index known to be a major determinant of the distribution of the 2SLS estimator in finite samples. Shea (1995) proposes an extension of this test to the case where X is multidimensional which has proven useful e.g. in the analysis in Pagan and Robertson (1996).

4 Inferences with the IV Estimator

Now let's consider making inferences with the IV estimator. We are really interested in two issues. One is the derivation of a test statistic for testing the hypothesis $H_0 : \theta = \theta_0$ and the other is how we might construct confidence intervals. The latter problem has been extensively discussed in Zivot et al (1998) and it is too complex to deal with here. Therefore we look at the issue of deriving a test statistic that can be used to deliver a p -value for the test of the null hypothesis. To do this we conventionally compute a t ratio.

Since the moments used for estimation are

$$E(m_t) = E(z_t u_t) = E(z_t(y_t - x_t \theta)) = 0,$$

method of moments theory gives the variance of $T^{1/2}(\hat{\theta} - \theta_0)$ as

$$V_{\hat{\theta}} = [E(\frac{\partial m_t}{\partial \theta})]^{-2} \text{var}(m_t) = \sigma_{zx}^{-1} (\sigma_u^2 \sigma_z^2) \sigma_{xz}^{-1},$$

since

$$\frac{\partial m_t}{\partial \theta} = -x_t z_t, \text{var}(m_t) = \sigma_u^2 E(z_t^2).$$

In standard derivations of a t ratio $V_{\hat{\theta}}$ is estimated by $\hat{\sigma}_{zx}^{-2} (\hat{\sigma}_u^2 \hat{\sigma}_z^2)$ giving

$$\begin{aligned} t_{\hat{\theta}} &= \frac{T^{1/2}(\hat{\theta} - \theta_0)}{\sqrt{V_{\hat{\theta}}}} = \frac{T^{1/2} \hat{\sigma}_{zx}^{-1} \sigma_{zu}}{\hat{\sigma}_{zx}^{-1} (\hat{\sigma}_u \hat{\sigma}_z)} \\ &= \frac{T^{1/2} \sigma_{zu}}{\hat{\sigma}_u \hat{\sigma}_z} \\ &= (T^{1/2} \rho_{zu}) \frac{\sigma_z \sigma_u}{\hat{\sigma}_z \hat{\sigma}_u} \end{aligned}$$

Now it is generally true that $T^{1/2} \rho_{zu}$ is asymptotically $N(0, 1)$ so that any problems with the distribution of the t ratio has to come from the term $\frac{\sigma_z \sigma_u}{\hat{\sigma}_z \hat{\sigma}_u}$. In normal circumstances this term converges to 1, giving the well known result that asymptotically $t_{\hat{\theta}}$ is distributed as $N(0, 1)$. So how does this change when there are weak instruments? The answer lies in the fact that

$$\begin{aligned} \hat{\sigma}_u^2 &= \frac{1}{T} \sum (y_t - x_t \hat{\theta})^2 = \frac{1}{T} \sum (y_t - x_t \theta_0 - x_t (\hat{\theta} - \theta_0))^2 \\ &= T^{-1} \sum (y_t - x_t \theta_0)^2 + (\hat{\theta} - \theta_0)^2 (T^{-1} \sum x_t^2) - 2(\hat{\theta} - \theta_0) \left(\frac{1}{T} \sum_{t=1}^T u_t x_t \right) \end{aligned}$$

Asymptotically the first term is σ_u^2 and normally the other two terms disappear as $\hat{\theta}$ is a consistent estimator of θ_0 . But with weak instruments (under the local to zero device) $\hat{\theta}$ is not consistent, so that $\hat{\sigma}_u$ is asymptotically a random variable, making the distribution of the t ratio non-standard as it is the product of an $N(0, 1)$ random variable and the inverse of whatever distribution $\hat{\sigma}_u$ has. Of course, unless one knows what this distribution is one can't easily form a confidence interval.

5 Constructing Useful Test Statistics

Various solutions have emerged for the general case when $\dim(X) = 1$ and $\dim(Z) \geq \dim(X)$. The oldest one is to engage in some lateral thought and perform an hypothesis test that $H_0 : \theta = \theta^*$ by solving a different problem.² Specifically, in estimating θ we had to assume that $E(z_t u_t) = 0$. Now this condition is $E(z_t(y_t - x_t \theta_0)) = 0$, where θ_0 is the true value of θ , so that

$$E(z_t(y_t - x_t' \theta^*)) = E(z_t x_t' (\theta_0 - \theta^*)). \quad (2)$$

Provided the instruments are not irrelevant we can test the null hypothesis that $\theta_0 = \theta^*$ by testing if $E(z_t(y_t - x_t \theta^*)) = 0$.

(2) can be formulated as a conditional moment test of the form

$$E[z_t(y_t - x_t' \theta^*) - \gamma] = 0$$

and $\gamma = 0$ can be tested using $\hat{\gamma} = \frac{1}{T} \sum_{t=1}^T z_t(y_t - x_t' \theta_0)$. Writing this in matrix form (and leaving out the T) we would be looking at $Z'(y - X\theta_0)$. If the null hypothesis $H_0 : \theta = \theta_0$ is correct and we have that u_t is *i.i.d.*($0, \sigma_0^2$) then $\text{var}(Z'(y - X\theta_0)) = \sigma_0^2 Z'Z$, pointing to the test statistic

$$AR = \frac{(y - X\theta_0)' Z (Z'Z)^{-1} Z (y - X\theta_0)}{\tilde{\sigma}_0^2},$$

where $\tilde{\sigma}_0^2 = \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \theta_0)^2$. This is the *Anderson-Rubin test statistic (AR)*. It is asymptotically a $\chi^2(\dim(Z))$ and is quite well behaved in finite

²It's important to note that what we mean by θ are the endogenous variable coefficients. If there are exogenous variables in the original relation we eliminate them as described in section 2.1 so that y_t, z_t and x_t would be residuals after they are regressed out from the original variables.

samples. The fact that it is $\chi^2(\dim(Z))$ and not $\chi^2(\dim(\theta))$ is an unsatisfactory aspect since $\dim(Z)$ may be much greater than $\dim(\theta)$. Provided *all* the parameters θ are being tested it readily extends to the general case where $\dim(\theta) > 1$, but that is probably a rarity.

Now, there are other ways of performing a test that is more direct than the *AR* test. What we have described in the previous section is the Wald test. How about doing the LM test? The Wald and LM tests are identical when $\dim(Z) = 1$ except that $\hat{\sigma}^2$ is replaced by $\tilde{\sigma}_0^2$ in the LM test. It immediately follows that the distribution of the t ratio version of the LM test will in fact be asymptotically $N(0, 1)$. This is a neat result (first observed by Wang and Zivot (1996)). Unfortunately it doesn't survive extension to more relevant models. In particular, the case when there are more instruments than regressors that need to be instrumented for, i.e. $\dim(z_t) > \dim(x_t)$. To see why, we note that the Two Stage Least Squares Estimator of θ would be

$$\begin{aligned}\hat{\theta} - \theta_0 &= (\hat{\pi}' Z' Z \hat{\pi})^{-1} (\hat{\pi}' Z' u) \\ sd(\hat{\theta}) &= \hat{\sigma}_u (\hat{\pi}' Z' Z \hat{\pi})^{-1/2}\end{aligned}$$

so that

$$t_{\hat{\theta}} = \frac{(\hat{\pi}' Z' Z \hat{\pi})^{-1} (\hat{\pi}' Z' u)}{\hat{\sigma}_u (\hat{\pi}' Z' Z \hat{\pi})^{-1/2}}$$

In the case that $\dim(Z) = \dim(X) = 1$, $\hat{\pi}$ is a scalar and cancels from numerator and denominator leaving

$$t_{\hat{\theta}} = \frac{Z' u}{\hat{\sigma}_u (Z' Z)^{1/2}}$$

and so it is enough to replace $\hat{\sigma}_u$ by $\tilde{\sigma}_0$ in order to ensure that $t_{\hat{\theta}}$ is asymptotically $N(0, 1)$. But when $\dim(Z) > \dim(X)$ we are left with $\hat{\pi}$ on the bottom line and we need to ask what its distribution now looks like in the local to zero (weak instrument case)

$$\hat{\pi} = \frac{\phi}{\sqrt{T}} + (Z' Z)^{-1} Z' \xi.$$

So if we want to stop $\hat{\pi}$ from becoming zero we need to recognize that $T^{1/2} \hat{\pi}$ will be asymptotically

$$N(\phi, \sigma_{\xi}^2 \sigma_z^{-2}).$$

To see what the effect of this is note that

$$t_{\hat{\theta}} = \frac{[(T^{1/2}\hat{\pi}')(\frac{1}{T}Z'Z)(T^{1/2}\hat{\pi})]^{-1}(T^{1/2}\hat{\pi}')(T^{-1/2}Z'u)}{\hat{\sigma}_u[(T^{1/2}\hat{\pi}')(T^{-1}Z'Z)(T^{1/2}\hat{\pi})]^{-1/2}},$$

and, given that $\hat{\pi}$ and $T^{-1/2}Z'u$ have limiting distributions, the t ratio is the product (and ratio) of many (asymptotically) normally distributed random variables. Hence, even if $\hat{\theta}$ is replaced by θ_0 when estimating σ_u , we do not get around the problems raised by weak instruments.

There are other cases in which $\hat{\pi}$ disappears from the test statistic e.g. if $\dim(\theta) = \dim(Z)$, $\hat{\pi}$ will be square, and it disappears from the quadratic form producing the conventional χ^2 statistic. Therefore, in that instance, provided we utilize σ_0^2 as the estimate of σ_u^2 the test statistic is indeed a $\chi^2(\dim(\theta))$ random variable. This case is of interest because it occurs in structural VARs with long-run restrictions, Pagan and Robertson(1998), although there the weak instruments come from near unit root behaviour of the x_t , and the local to zero assumption made in the weak instruments literature is not really appropriate, since π would now need to be $\frac{\delta}{T}$ rather than $\frac{\delta}{\sqrt{T}}$.

Returning to the $\dim(\theta) = 1, \dim(Z) > 1$ case, suppose that we consider a different estimator of π . To derive one look at the system

$$\begin{aligned} y_t &= x_t\theta + u_t \\ x_t &= z_t'\pi + \xi_t \end{aligned}$$

If $\begin{bmatrix} u_t \\ \xi_t \end{bmatrix}$ is $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{uu} & \sigma_{u\xi} \\ \sigma_{\xi u} & \sigma_{\xi\xi} \end{bmatrix}\right)$ then we can write $\xi_t = \frac{\sigma_{u\xi}}{\sigma_{uu}}u_t + \eta_t$, where η_t is independent of u_t . Setting $\theta = \theta_0$ we can estimate u_t and also $\frac{\sigma_{u\xi}}{\sigma_{uu}}$ so that regressing $x_t - \frac{\hat{\sigma}_{u\xi}}{\hat{\sigma}_{uu}}\hat{u}_t$ against z_t will produce an alternative estimator for π of $\tilde{\pi}$. In fact this would be the LIML estimate given that $\theta = \theta_0$. It's clear that $\tilde{\pi}$ is a function of η_t , and so independent of u_t , implying that (conditional upon Z) $Z'u$ is independent of $\tilde{\pi}$. This independence means we can now condition upon $\tilde{\pi}$ when evaluating the distribution of a t ratio which uses $\tilde{\pi}$ in place of $\hat{\pi}$. Thus the t ratio will now be asymptotically normal. Replacing $\hat{\sigma}^2$ by σ_0^2 and $\hat{\pi}$ by $\tilde{\pi}$ means that one is performing an *LM* test that $\theta = \theta_0$, and this was Kleibergen's (2002) suggestion. This statistic clearly extends to handle $\dim(\theta) > 1$. As with much of this literature a crucial requirement is that Z is exogenous (or can be conditioned upon), which is not the case if the weak instruments stem from unit root type behaviour.

The problems caused by the behaviour of $\hat{\pi}$ can be solved in other ways. One suggestion is that we should condition upon some function of $\hat{\pi}$, in particular the the concentration parameter $\phi = \sigma_{\xi\xi}^{-1}\pi'Z'Z\pi$. An estimate of this, $\hat{\phi}$, is the F statistic that tests if π is zero. Hillier and Forchini (2004) discuss the logic of such conditioning, arguing that it is natural to condition the distribution of $\hat{\theta}$ upon the outcome of a test that $\pi = 0$, since that summarizes the information in a sample about θ . They derive the moments of this conditional distribution and compare them to OLS. If the concentration parameter is very small the results suggest that OLS might be the preferred estimator. Hillier and Forchini also provide distributions for the $\dim(\theta) > 1$ case, conditioning upon the minimum eigenvalue of the multivariate analogue of the concentration parameter. They observe however that these expressions are so complex and hard to evaluate that some other function of π might be a better conditioning choice.

Moreira (2003) considered the likelihood ratio (LR) test that $\theta = \theta_0$. If we let the F statistic that the coefficients of z_t are zero in the regression of $x_t - \frac{\hat{\sigma}_{u\xi}}{\hat{\sigma}_{uu}}\hat{u}_t$ against z_t be \tilde{r} , which estimates a modified "concentration parameter", $r = \sigma_{\eta\eta}^{-1}\pi'Z'Z\pi$, then Kleibergen (2005) gives

$$LR = \frac{1}{2}[AR - \tilde{r} + \sqrt{(AR + \tilde{r})^2 - 4\tilde{r}(AR - LM)}].$$

As mentioned before AR and \tilde{r} must be independent and this suggests we condition upon \tilde{r} in this case. The distribution of this conditional LR test (CLR) can be found numerically, a recent algorithm being given in Andrews, Moreira and Stock (2005b).

Most of the attention in the literature above has been paid to getting a test statistic that is in some ways independent of π . But one also wants a test that is powerful, and this has recently been studied by Poskitt and Skeels (2005) (using small concentration parameter asymptotics) and Andrews, Moreira and Stock (2005a). The latter consider the construction of point optimal invariant tests. They also note that the power function of the LM test is not monotonic and that the CLR test seems best of the trio discussed above.

A related issue that is only now being dealt with is the testing of sub-sets of θ i.e. if $\theta' = [\theta'_1 \ \theta'_2]$ we wish to test $H_0 : \theta_1 = \theta_{10}$. If the parameters θ_2 have no weak instrument issues relating to them, then the results established above continue to hold, since the IV estimator of θ_2 is well behaved. If however they are also subject to weak instrument bias some adjustments

will need to be made. Recent papers by Dufour and Taamouti (2005) and (Kleibergen (2007)) consider the question of how to do this; the former use projection methods to get conservative tests while the latter establishes some bounds on the distributions for $\hat{\theta}_1$.

There are a lot of other issues raised by this literature One involves the implications of it for GMM estimators. Basically the GMM estimator can be thought of as

$$\hat{\theta} - \theta_0 = \left(-\sum \frac{\partial m_t}{\partial \theta}\right)^{-1} \left(\sum m_t(\theta_0)\right)$$

and we normally assume that $\frac{1}{T} \sum \frac{\partial m_t}{\partial \theta}$ converges to a constant. But it might also be that $E\left(\frac{\partial m_t}{\partial \theta}\right)$ is either zero or very small in such a case and the same issues arise as were raised above with weak instruments (of course in the simple IV case $-\frac{\partial m_t}{\partial \theta} = z_t x_t$). So it may be that the asymptotic theory for the GMM estimator can fail. Using the Generalized Information Equality we know that $E\left(-\frac{\partial m_t}{\partial \theta}\right) = E(m_t L_{\theta,t})$, where $L_{\theta,t}$ are the scores for θ , and so there are very real problems for the GMM estimator if the moments chosen, m_t , are not correlated with the scores, so one should experiment a little to determine what would be good moments (it helps to have a theoretical model here that one can estimate since one can then simulate it for chosen parameter values and then study the issue of moment selection numerically). The failure of GMM estimators to have a distribution as predicted by asymptotic theory has been well documented e.g. Kocherlakota (1990).

6 Examples of Weak Instruments

6.1 Estimating the Phillips Curve in a New Keynesian Model

The New Keynesian model that has become increasingly popular in macro-economic research has the form.

$$\begin{aligned} \pi_t &= \delta_1 E_t(\pi_{t+1}) + \delta_2 \pi_{t-1} + \lambda x_t + \varepsilon_{AS,t} \\ x_t &= \mu E_t(x_{t+1}) + (1 - \mu)x_{t-1} - \phi(r_t - E_t \pi_{t+1}) + \varepsilon_{IS,t} \\ r_t &= \rho r_{t-1} + (1 - \rho)[\beta E_t \pi_{t+1} + \gamma x_t] + \varepsilon_{MP,t}, \end{aligned}$$

where π_t is the inflation rate, y_t is the output gap and r_t is an interest rate determined by policy. Three shocks appear in the model - an aggregate

supply shock ($\varepsilon_{AS,t}$), an IS (demand) shock ($\varepsilon_{IS,t}$), and a monetary policy shock ($\varepsilon_{MP,t}$). Suppose we try to estimate the Phillips curve in the standard way i.e. to assume rational expectations and thereby replace $E_t\pi_{t+1}$ with π_{t+1} . Then we need an instrument for π_{t+1} . But we will also need one for x_t as that is an endogenous variable. What instruments are available? We know that the solution to this system (if there is no serial correlation in the shocks) expresses π_t, x_t and r_t as functions of their first lagged values only. Therefore the only instruments available are x_{t-1}, r_{t-1} and π_{t-1} . But π_{t-1} already appears in the Phillips curve so we are effectively left with x_{t-1} and r_{t-1} as instruments for π_{t+1} . As you would discover from any regression these are very poor instruments and so it is virtually impossible to estimate δ_1 and δ_2 . Of course in practice it is often the case that the restriction that $\delta_2 = 1 - \delta_1$ is applied (producing a "hybrid model") and this changes the Phillips curve to

$$\Delta\pi_t = \delta_1 E_t(\pi_{t+1} - \pi_{t-1}) + \lambda x_t + \varepsilon_{AS,t},$$

and often π_{t-1} will be a good instrument for $\pi_{t+1} - \pi_{t-1}$. Hence sometimes the weak instrument problem can be overcome via adopting a model design that is compatible with some theoretical reasoning.

6.2 Euler Equations for Inventory Problems

Weak instruments can arise in many ways. Sometimes not enough care has been put into their selection e.g. as seen in the debate over the returns to schooling where it has been shown that the answers are very sensitive to the choice of instrument. In other cases difficulties may arise either from the nature of the model being used or the interaction of model features and the data. A possible example of the latter would be the “common factors” test of Vahid and Engle (1993) where one needs instruments for the growth rates of variables in output and consumption, and it is rare to find much correlation in these growth rates.³ In some cases one can locate the source of the weak instrument and we turn to two of these now.

³Their test is essentially one of serial correlation after instrumental variables and therefore the distribution of the statistic depends upon another estimated quantity. Since the latter is the IV estimator of the parameters of a linear relation it is affected by weak instruments, and may therefore may impact upon the distribution of the statistic of interest. This issue has not been examined much in the literature but we have found that serial correlation tests can be affected in strange ways when one is basing them on weak instruments.

The parameters appearing in the set of first order conditions associated with Euler equations defining the optimal choices for decision variables are frequently estimated by GMM and potentially involve weak instruments. Many situations arise when the potential becomes reality. Gregory et al (1993) showed that, if one used the Euler equations from a linear quadratic optimization model, instruments are irrelevant when used to identify the discount parameter whenever the forcing variables are I(1) variables. A related example is the influential paper of Fuhrer et al (1995). They reported extremely poor performance of the GMM estimator in certain circumstances, even with thousands of observations.

Fuhrer et al's paper derived the first order conditions defining optimal inventory choice in a linear quadratic context with agents minimizing

$$\sum_{j=0}^{\infty} \beta^j E_t [C_Y(Y_{t+j}) + C_N(t_{t+j}, S_{t+j})],$$

subject to the constraint $N_{t+j} = N_{t+j-1} + Y_{t+j} - S_{t+j}$, where Y_t is output, S_t is sales, N_t is the level of inventories and

$$C_Y(Y_{t+j}) = (\delta/2)Y_{t+j}^2 + (\alpha/2)(\Delta Y_{t+j})^2$$

$$C_N(N_{t+j}, S_{t+j}) = (\phi/2)(N_{t+j} - \omega S_{t+j})^2.$$

The performance of the GMM estimator is so poor as to lead them to recommend that it might be preferable to use a maximum likelihood estimator, even if an incorrect likelihood is used (the latter qualification arising from the need to specify a form for the driving process for sales in order to determine a likelihood). It is worth looking at the origins of the poor performance of the GMM estimator in the model investigated by Fuhrer et al., as an understanding of the causes is important when assessing recommendations such as that just mentioned.

The Euler equation coming from the optimization above is

$$E_t [\delta (Y_t - \beta Y_{t+1}) + \alpha (\Delta Y_t - 2\beta \Delta Y_{t+1} + \beta \Delta Y_{t+2}) + \phi (N_t - \omega S_t)] = 0, \quad (3)$$

and this provides a set of moment conditions

$$E [z_t \{\delta (Y_t - \beta Y_{t+1}) + \alpha (\Delta Y_t - 2\beta \Delta Y_{t+1} + \beta \Delta Y_{t+2}) + \phi (N_t - \omega S_t)\}] = 0, \quad (4)$$

where z_t are instruments drawn from the information set used in the conditional expectation. Now it is clear from (4) that one cannot identify all of the parameters entering into it and researchers have chosen particular normalizations. Fuhrer et al work with five normalizations. Two of these, designated A and B , are $A : \delta = 1, B : \phi = 1$. To aid the analysis of these different normalizations we employ two simplifying assumptions. First we assume that $S_t = S_{t-1} + e_t$, where e_t is *i.i.d.*($0, \sigma^2$), i.e. sales are a pure unit root process. Second we put $\beta = 1$.⁴ With S_t being $I(1)$ it can be shown that N_t is $I(1)$ and ω is the cointegrating parameter. With these assumptions, normalization A involves a linear model in which ΔY_{t+1} is the dependent variable while $N_t - \omega S_t$ and $\Delta^2 Y_{t+2}$ are regressors, and instruments are needed for the latter. When the normalization is B , $N_t - \omega S_t$ is the dependent variable and ΔY_{t+1} and $\Delta^2 Y_{t+2}$ are regressors, and instruments are required for both of these. The instruments used are effectively the changes in Y_t and S_t as the levels will not be useful as an instrument for ΔY_{t+1} as one would be using an $I(1)$ variable as an instrument for $I(0)$ variables. Now $\Delta Y_{t+1} = \Delta S_{t+1} + \Delta^2 N_{t+1}$ and therefore, for lags of ΔY_t to be effective as instruments, it would be necessary for $\Delta^2 N_t$ to have serial correlation. If one solves for the N_t process under the specification used for S_t above, it can be established that there is very little serial correlation in $\Delta^2 N_t$. This means that normalizations such as B will result in very weak instruments. Thus one can see the origin of their conclusion that (p. 143)

“in the case of the smoothing model, normalization I requires 30,000 observations to converge to the true value”.⁵

This example shows why it is hard to come up with general propositions regarding the possibility of weak instruments. If S_t was not close to a pure unit root process then the instruments could be very effective so that the outcome is very dependent on the nature of the underlying forcing processes.⁶

⁴In their data analyses and simulations $\beta = .995$ and the univariate process used for sales effectively features a unit root with some weak autocorrelation. Fuhrer et al comment that when the sales process is mis-specified by fitting an AR(1) rather than an AR(3) that “the estimated lag coefficient in the AR(1) model tends to approximate the sum of the AR(3) coefficients, the dominant root” (p. 143). This would give a root of .956.

⁵What I have called normalization B they term I .

⁶The same conclusion holds for the example in Gregory et al (1993)

7 References

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2005a), "Optimal Invariant Similar Tests for Instrumental Variables Regression", *Econometrica*, (forthcoming)

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2005b), "Performance of the Conditional Wald Tests in IV Regression with Weak Instruments", *Journal of Econometrics* (forthcoming)

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

Forchini, G. and G.Hillier (2003), "Conditional Inference for Possibly Unidentified Structural Equations", *Econometric Theory*, 19, 707-743.

Fuhrer, J., G. Moore, and S. Schuh (1995), "Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood Versus Generalized Method of Moments", *Journal of Monetary Economics*, 35, 115-158

Hall, A.R., G.D. Rudebusch, and D.W. Wilcox (1996), "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283-298.

Gregory, A.W., A.R. Pagan and G.W. Smith (1993), "Estimating Linear Quadratic Models with Integrated Processes" in P.C.B. Phillips (ed.) *Models, Methods and Applications of Econometrics* (Blackwell).

Kleibergen, F. (2002), "Pivotal Statistics for Testing Parameters in Instrumental Variables Regression", *Econometrica*, 70, 1781-1803.

Kleibergen, F. (2006), "Testing" in S. Durlauf and L. Blume, *The New Palgrave Dictionary of Economics*, 2nd edition. The New Palgrave (forthcoming)

Kleibergen, F. (2007), "Subset Statistics in the Linear IV Regression Model", (mimeo, Brown University)

Kocherlakota, N.R. (1990), "On Tests of Representative Consumer Asset Pricing Models", *Journal of Monetary Economics*, 26, 285-304.

Moreira, M.J. (2003), "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica*, 71, 1027-1-48.

Nelson, C. R., and R. Startz (1990), "The Distribution of the Instrumental Variables Estimator and its t-Ratio When the Instrument is a Poor One", *Journal of Business*, 63, S125-140.

Pagan, A.R., and J.C. Robertson (1998), "Structural Models of the Liquidity Effect", *Review of Economics and Statistics*, 80, 202-217.

Poskitt and Skeels (2005), "Small Concentration Asymptotics and Instrumental Variables Inference", *Research Paper no 948*, Department of Economics, University of Melbourne

Shea, J. (1997), "Instrument Relevance in Multivariate Linear Models: A Simple Measure", *Review of Economics and Statistics*, 79,348-352.

Staiger, D., and J.H. Stock (1995), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557-586..

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

Stock, J.H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression", in J.H. Stock and D.W.K. Andrews (ed) *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*, Cambridge University Press

Stock, J.H., J.H. Wright and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in GMM", *Journal of Business and Economic Statistics*

Vahid, F. and R. Engle (1993), "Common Trends and Common Cycles", *Journal of Applied Econometrics*, 8, 341-360.

Wang, J., and E. Zivot (1996), "Inference on a Structural Parameter in Instrumental Variables Regression with Weakly Correlated Instruments", *Econometrica*, 66, 1389-1404.

West, K.D., and D.W. Wilcox (1996), "A Comparison of Alternative Instrumental Variable Estimators of a Dynamic Linear Model", *Journal of Business and Economic Statistics*, 14, 281-293.