# Department of Economics

## Cross-sectional Averaging and Instrumental Variable Estimation with Many Weak Instruments

George Kapetanios and Massimiliano Marcellino

Working Paper No. 627 March 2008 ISSN 1473-0278

Queen Mary
University of London

# Cross-sectional Averaging and Instrumental Variable Estimation with Many Weak Instruments

G. Kapetanios and M. Marcellino
Queen Mary, University of London    Bocconi University and EUI

March 14, 2008

**Abstract**

Instrumental variable estimation is central to econometric analysis and has justifiably been receiving considerable and consistent attention in the literature in the past. Recent developments have focused on cases where instruments are either weak, in terms of correlations with the endogenous variables, or many or both. The present paper suggests a new way to deal with many, possibly weak, instruments. Our suggestion is to cross-sectionally average the instruments and use these averages as instruments. Intuition and interesting recent work by Hahn (2002) suggest that parsimonious devices used in the construction of the final instruments, may provide effective estimation strategies. Our use of cross-sectional averaging promotes parsimony and therefore falls within the context of such arguments. We provide a theoretical analysis of this approach in terms of its consistency properties and also show, via a Monte Carlo study, that the approach can provide improved estimation compared to standard instrumental variables estimation.

*Keywords*: Instrumental Variable Estimation. 2SLS, cross-sectional average.

*JEL classification*: C13, C23, C51.

## 1 Introduction

Recent work in instrumental variable estimation has considered two distinct routes. The first is one where instrumental variables are only weakly correlated with the endogenous explanatory variables of an instrumental variables (IV) regression. Work by, e.g., Phillips (1983), Rothenberg (1984), Stock and Yogo (2003b) and Chao and Swanson (2005) consider a natural measure of instrument weakness (or strength) in a linear IV framework to be the so-called concentration parameter. In standard analysis the concentration parameter is taken to grow at the rate of the sample size whereas in the case of weak instruments this parameter grows more slowly or in the extreme case introduced and considered by Staiger and Stock (1997) it remains finite asymptotically. In the case of weak instruments, the properties of IV estimators such two stage least squares (2SLS) and limited information maximum likelihood (LIML) are affected relative to the case of strong instruments and the estimators may, in fact, be inconsistent.

Another direction in IV research involves the case where the number of available instruments is large. This approach was first taken by Morimune (1983) and later generalized by Bekker (1994). Other relevant papers include Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), Hahn (2002), and Chao and Swanson (2004). More recently, the two different stands have been combined to provide a comprehensive framework for the analysis of the properties of IV estimators in the case of many weak instruments. Work on this includes Hansen, Hausman, and Newey (2006), Stock and Yogo (2003a), Newey (2004) and Chao and Swanson (2005). A clear conclusion from this work suggests that inconsistency of IV estimators is a probable outcome when many weak instruments are used.

With this in mind, a further recent development focuses on considering parsimonious modeling assumptions for the set of instruments to avoid IV estimator inconsistency. In particular, Kapetanios and Marcellino (2006) suggest that imposing a factor structure on the set of instruments, extracting estimates of these factors and using them as instruments can be very useful. Of course, an issue with this approach is the need to assume a factor structure, albeit a possibly weak one, as discussed in detail in Kapetanios and Marcellino (2006). Simulation evidence suggests that if no factor structure exists then assuming one is problematic for IV estimation as one would expect.

The present paper aims to provide a new method in a similar spirit to Kapetanios and Marcellino (2006) but designed to parsimoniously summarise large sets of instruments in the complete absence of a factor structure. There is a reasonably strong case for parsimony to be made for IV estimation. In a very interesting and stimulating paper, Hahn (2002) provides grounds for parsimony in terms of optimal inference when many instruments are available. The basic idea of our paper is that a finite number of cross-sectional weighted averages of the available instruments can, under certain conditions, be valid instruments themselves. We explore in some detail the necessary condition for validity of cross-sectional averages as instruments. A Monte Carlo study provides support for the new method.

The paper is structured as follows: Section 2 presents the theoretical results. Section 3 reports results of the Monte Carlo study. Finally, Section 5 concludes.

## 2 Theoretical Considerations

The model is given by

$$y_{1n} = Y_{2n}\beta + u_n$$

$$Y_{2n} = Z_n\Pi_n + V_n$$

where $y_{1n}$ and $Y_{2n}$ are respectively an $n \times 1$ vector and an $n \times G$ matrix of observations on the $G+1$ endogenous variables of the system, $Z_n$ is an $n \times K_n$ matrix of observations on the $K_n$ instrumental variables, and $u_n = (u_1, ..., u_i, ..., u_n)'$ and $V_n = (v_1, ..., v_i, ..., v_n)'$ are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of random disturbances.

Define a weight matrix $W_n = [w_{ij}]$ to be an $K_n \times G$ matrix of weights. Then, the cross-sectional average (CSA) instrumental variables are defined to be $\bar{Y}_{2n} = Z_n W_n$. We make the following assumption

**Assumption 1** *(i) $Z_n$ and $\eta_i = (u_i, v_i')'$ are independent for all $i, n$, (ii) $\eta_i \sim i.i.d.(0, \Sigma)$, where* $\Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_{VV} \end{pmatrix}$

Then, we have the following theoretical result:

**Theorem 1** *Let Assumption 1 hold. Let $\hat{\beta}_{2SLS} = \left(\bar{Y}_{2n}'Y_{2n}\right)^{-1}\bar{Y}_{2n}'y_{1n}$ be the 2SLS estimator of $\beta$ using the CSA instrumental variables. Assume that there exists $r_n \to \infty$ such that $r_n/n \to \kappa$ and $0 \leq \kappa < \infty$, and an invertible matrix $\Psi$, for which $W_n'Z_n'Z_n\Pi_n/r_n \overset{a.s.}{\to} \Psi$. Further, assume that $W_n'Z_n'Z_nW_n/r_n^2 \overset{a.s.}{\to} 0$. Then, $\hat{\beta}_{2SLS} - \beta = o_p(1)$.*

*Proof of Theorem 1:*

We have that

$$\hat{\beta}_{2SLS} - \beta = \left(\frac{\bar{Y}_{2n}'Y_{2n}}{r_n}\right)^{-1}\frac{\bar{Y}_{2n}'u_n}{r_n}$$

We first examine $\frac{\bar{Y}_{2n}'Y_{2n}}{r_n}$. We have

$$\bar{Y}_{2n}'Y_{2n} = W_n'Z_n'Z_n\Pi_n + W_n'Z_n'V_n$$

By the assumption of the Theorem $W_n'Z_n'Z_n\Pi_n/r_n \overset{a.s.}{\to} \Psi$. We next examine $W_n'Z_n'V_n/r_n$. We have

$$E\left[\left\|\frac{W_n'Z_n'V_n}{r_n}\right\|^2\right] = E\left(Tr\left[\frac{W_n'Z_n'V_nV_n'Z_nW_n}{r_n^2}\right]\right) =$$

$$Tr\left[\Sigma_V\right]E\left(Tr\left[\frac{W_n'Z_n'Z_nW_n}{r_n^2}\right]\right) \leq$$

$$C_1 E\left(Tr\left[\frac{W_n' Z_n' Z_n W_n}{r_n^2}\right]\right) = o(1)$$

for some constant $C_1$. Thus, overall,

$$\frac{\bar{Y}_{2n}' Y_{2n}}{r_n} \xrightarrow{a.s.} \Psi$$

Next, we examine

$$\frac{\bar{Y}_{2n}' u_n}{r_n} = \frac{W_n' Z_n' u_n}{r_n}$$

Then, similarly to above we have

$$E\left[\left\|\frac{W_n' Z_n' u_n}{r_n}\right\|^2\right] =$$

$$\sigma_u^2 E\left(Tr\left[\frac{W_n' Z_n' Z_n W_n}{r_n^2}\right]\right) \leq$$

$$C_2 E\left(Tr\left[\frac{W_n' Z_n' Z_n W_n}{r_n^2}\right]\right) = o(1)$$

for some constant $C_2$. Overall, it then follows that $\hat{\beta}_{2SLS} - \beta = o_p(1)$ proving the Theorem. Q.E.D.

The main condition of the Theorem is given by $W_n' Z_n' Z_n \Pi_n / r_n \xrightarrow{a.s.} \Psi$, where $\Psi$ is an invertible matrix, and $r_n/n \to \kappa$ and $0 \leq \kappa < \infty$. This condition needs to be further explored and for this we provide a number of specific examples below.

**Example 1** *Let $G = 1$. Let $\Pi_n = (\varpi_1, ..., \varpi_{K_n})'$ and $\varpi_i = \varpi/\sqrt{K_n}$, $\varpi \neq 0$. For simplicity, assume all instruments have mean zero. $E(z_{in} z_{in}') = diag(\sigma_1^2, ..., \sigma_{K_n}^2)$ where $Z_n = (z_{1n}, ..., z_{nn})'$ and $z_{in}$ is an $K_n \times 1$ vector and $\sigma_i^2$, $i = 1, ..., K_n$, are finite positive constants. We set $w_{ij} = 1/\sqrt{K_n}$ and consider $W_n' Z_n' Z_n \Pi_n / r_n$, which for this case becomes $1/n \sum_{i=1}^n \left(\varpi/\sqrt{K_n}\iota' z_{in}\right)\left(1/\sqrt{K_n}\iota' z_{in}\right)$. Under either sequential asymptotics whereby $n \to \infty$ followed by $K_n \to \infty$, or joint asymptotics where $n, K_n \to \infty$ jointly,*

$$1/n \sum_{i=1}^n \left(\varpi/\sqrt{K_n}\iota' z_{in}\right)\left(1/\sqrt{K_n}\iota' z_{in}\right) \xrightarrow{a.s.} \varpi\sigma^2$$

*where $\sigma^2 = \lim_{K_n \to \infty} 1/K_n \sum_i^{K_n} \sigma_i^2$.*

**Example 2** *Consider the setup of Example 1, but in this case set $\varpi = 0$. Then, obviously the main condition of Theorem 1 is not satisfied as the instruments are not related to the endogenous variables. Slightly more subtly, we can extend Example 1 to have $\varpi_i \sim i.i.d.(\varpi, \sigma_\varpi^2)$ where $\varpi_i$ are independent of all other stochastic quantities in the model and $\sigma_\varpi^2$ is a finite* *constant. Then, it is straightforward to see that the result of Example 1 holds as long as* $\varpi \neq 0$. *If* $\varpi = 0$, *then although the original instruments are relevant, IV estimation using CSA instrumental variables fails. Note that, in this case it is also straightforward to see that* $\Pi_n' Z_n' Z_n \Pi_n / n \overset{a.s.}{\to} \sigma_{\varpi}^2 \sigma^2$ *thus showing that the last part of Assumption 1 of Chao and Swanson (2005) holds thereby making standard IV estimation valid. But if only a finite number of* $\varpi_i$ *are non zero then* $\Pi_n' Z_n' Z_n \Pi_n / r_n \overset{a.s.}{\to} 0$ *for all* $r_n \to \infty$. *Then, standard IV estimation fails as well. It is clear that the last part of Assumption 1 of Chao and Swanson (2005), although less strict than the main Condition of Theorem 1, is not innocuous either. It is also easy to see that the above results extend straightforwardly to the case where* $E(z_{in} z_{in}')$ *is not diagonal but has bounded column sum norm.*

**Example 3** *For this example we extend the setup of Example 1 to* $G > 1$. *Thus, let* $\Pi_n = (\tilde{\varpi}_1, ..., \tilde{\varpi}_G)$ *where* $\tilde{\varpi}_j = (\varpi_{j1}, .., \varpi_{jK_n})'$, $j = 1, ..., G$ *and* $\varpi_{ij} = \varpi_j$. *Further, we set* $w_{ij} = w_j / \sqrt{K_n}$ *The rest of the setup of Example 1 is kept intact. Then,*

$$\Pi_n' Z_n' Z_n W_n / n \overset{a.s.}{\to} \begin{pmatrix} \varpi_1 w_1 \sigma^2 & \varpi_1 w_2 \sigma^2 & ... & \varpi_1 w_G \sigma^2 \\ \varpi_2 w_1 \sigma^2 & ... & ... & \varpi_2 w_G \sigma^2 \\ ... & ... & ... & ... \\ \varpi_G w_1 \sigma^2 & ... & ... & \varpi_G w_G \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \varpi_1 \\ ... \\ ... \\ \varpi_G \end{pmatrix} (w_1, ... w_G)$$

*and*

$$\Pi_n' Z_n' Z_n \Pi_n / n \overset{a.s.}{\to} \begin{pmatrix} \varpi_1^2 \sigma^2 & \varpi_1 \varpi_2 \sigma^2 & ... & \varpi_1 \varpi_G \sigma^2 \\ \varpi_2 \varpi_1 \sigma^2 & ... & ... & \varpi_2 \varpi_G \sigma^2 \\ ... & ... & ... & ... \\ \varpi_G \varpi_1 \sigma^2 & ... & ... & \varpi_G \varpi_G \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \varpi_1 \\ ... \\ ... \\ \varpi_G \end{pmatrix} (\varpi_1, ... \varpi_G)$$

*In both cases the rank of the limit is 1 implying that neither the main condition of Theorem 1 not the last part of assumption 1 of Chao and Swanson (2005) holds. If we instead assume that* $\varpi_{ij} \sim i.i.d.(\varpi_j, \sigma_{\varpi_j}^2)$ *then the limit of* $\Pi_n' Z_n' Z_n W_n / n$ *will still be of rank equal to 1. But*

$$\Pi_n' Z_n' Z_n \Pi_n / n \overset{a.s.}{\to} \begin{pmatrix} \sigma_{\varpi_1}^2 \sigma^2 & \varpi_1 \varpi_2 \sigma^2 & ... & \varpi_1 \varpi_G \sigma^2 \\ \varpi_2 \varpi_1 \sigma^2 & ... & ... & \varpi_2 \varpi_G \sigma^2 \\ ... & ... & ... & ... \\ \varpi_G \varpi_1 \sigma^2 & ... & ... & \sigma_{\varpi_G}^2 \sigma^2 \end{pmatrix}$$

*The limit now is a full rank matrix. Again, though, if all but a finite number of the elements of the* $j$*-th row of* $\Pi_n'$ *are equal to* $\varpi_j$ *then standard IV estimation fails.*

**Example 4** *The problem that became apparent in Example 3 relates to the fact that a necessary (but not sufficient) condition for validity of standard IV is that* $\Pi_n$ *is full column rank, and a necessary (but, again, not sufficient) condition for CSA IV is that both* $\Pi_n$ *and* $W_n$ *are full column rank. Exploring further, the condition for validity of CSA IV estimation we can* *see that $W_n$ can be made full rank by associating different groups of instruments to different columns of $W_n$. So for example, in the context of Example 3, the g-th column of $W_n$ can be given by*

$$(\overbrace{0,...,0}^{\sum_{s=1}^{g-1}K_n^s},\overbrace{1/\sqrt{n},...,1/\sqrt{n}}^{K_n^g},\ \overbrace{0,...,0}^{\sum_{s=g+1}^{G}K_n^s}\ ) \tag{1}$$

*where $\sum_{s=1}^{G} K_n^s = K_n$. Of course, different columns of $W_n$ can overlap as long as they are not identical. As we said, full column rank for $W_n$ is not a sufficient condition for the main condition of Theorem 1 to hold. A number of further assumptions can be made though to get such a result. One set of such assumptions imposes mild structures on $\Pi_n$. For example, partition $z_{in}$ comformably to $W_n$ in the case where the columns of $W_n$ are constructed as in (1), to get $z_{in} = (z_{in}^{1\prime}, ..., z_{in}^{G\prime})'$ where $z_{in}^g$, $g = 1, ..., G$ is a $K_n^g \times 1$ vector. Next, partition $\Pi_n$ comformably to $z_{in}$ to get $\Pi_n = (\Pi_n^{1\prime}, ..., \Pi_n^{G\prime})'$ where $\Pi_n^g = [\varpi_{ij}^g]$ are $K_n^g \times G$ matrices. Then, setting $\varpi_{ij}^g \sim i.i.d.(\varpi_j^g, \sigma^2_{\varpi_j^g})$ and restricting $\varpi_j^g$ to take different values across $j$ and $g$ ensures that the limit of $\Pi_n' Z_n' Z_n \tilde{W}_n/n$ is full rank.*

**Example 5** *Examples 1-4 have provided some detailed analysis of particular cases where the main condition of Theorem 1 holds or does not hold. We saw that allowing the elements of $\Pi_n$ to be stochastic may pose problems for the validity of the condition. However, it is also worth noting that previous work in the literature has mainly focused on non-stochastic elements for $\Pi_n$. In this case constructing particular designs that allow an exploration of the validity of the main condition of Theorem 1 are much more difficult. In this case it is worth noting that a sufficient condition for the condition of Theorem 1 is that there exist $r_{1n} \to \infty$ and $r_{2n} \to \infty$ such that $\lim_{n\to\infty} \Pi_n' W_n/r_{1n}$ and $plim_{n\to\infty} Z_n' Z_n/r_{2n}$ have nonsingular and positive definite limits respectively. A necessary condition for the first of the above conditions is, of course, as we noted earlier, that for all n both $\Pi_n$ and $W_n$ have full column rank.*

**Remark 1** *The importance of parsimony for IV estimation has been pointed out by Hahn (2002) who conjectured that a 2SLS estimator using small subset of available instruments, when the number of available instruments is large, may be optimal. We view our cross-sectional averaging estimator in the same spirit as the estimator suggested by Hahn (2002).*

As the above discussion makes clear, the cross-sectional average instrumental variable estimator has the potential to provide consistent estimation when standard instrumental variable estimation cannot. On the other hand the main condition of Theorem 1 is not necessarily true and, therefore, it would be useful to have some means for its verification. This condition is essentially needed for making the cross-sectional averages relevant instruments. In the case where the condition is not satisfied the instruments are completely irrelevant.

However, the study of the relevance of instruments has received some attention in the literature. We therefore suggest that standard existing tools may be used on the cross-sectional average instrumental variables to ascertain their relevance. As we are dealing with a finite number of instruments standard theory applies. Examples of work that provides tools for investigating instrument relevance include Hall, Rudebusch, W., and Wilcox (1996), Bound, Jaeger, and Baker (1995), Shea (1997) and Poskitt and Skeels (2002). The last paper is especially relevant given its main focus on completely irrelevant instruments, which is the case for the cross-sectional averages relevant instruments if the main condition of Theorem 1 fails, rather than weak instruments.

# 3 Monte Carlo Evidence

In this section we provide a Monte Carlo study of the cross-sectional average instrumental variables estimator and its relative performance compared to the standard instrumental variable estimator and to the Factor IV estimator introduced by Kapetanios and Marcellino (2006). We focus on 2SLS estimation. The basic setup of the Monte Carlo experiments is:

$$y_i = x_i + \epsilon_i,\ i = 1, ..., n \tag{2}$$

$$z_{ij} = e_{ij}, \quad j = 1, ..., K_n,\ i = 1, ..., n \tag{3}$$

$$x_i = \sum_{j=1}^{K_n} K_n^{-1/2}(1 + \alpha_j) z_{ij} + u_i, \tag{4}$$

where $e_{ij} \sim i.i.d.N(0,1)$, $\alpha_j \sim N(0, c^2)$ and $cov(e_{il}, e_{sj}) = 0$ for $i \neq s$ or $l \neq j$ . $c = 0.1, 0.2, 0.5, 2$. Let $\kappa_i = (\epsilon_i, u_i)'$. Then, $\kappa_i = P\eta_i$, where $\eta_i = (\eta_{1,i}, \eta_{2,i})'$, $\eta_{j,i} \sim i.i.d.N(0,1)$ and $P = [p_{ij}]$, $p_{ij} \sim i.i.d.N(0,1)$. The errors $e_{ij}$ and $u_s$ are independent for each $i$ and $s$. The CSA is computed using equal weigths of $1/K_n$, and the factor 2SLS is based on one factor, estimated as the first principal component of $z_1, ..., z_{K_n}$ where $z_j = (z_{1j}, ..., z_{nj})'$.

In all cases the 2SLS and CSA 2SLS estimators have negligible biases, while the bias of the Factor 2SLS estimator is slightly larger, and we therefore concentrate on their variances, which are reported in Table 1. Results make interesting reading. Focusing first on the comparison CSA 2SLS - standard 2SLS estimator, the former dominates the latter in most cases in terms of variance. More specifically, results in general improve as $n$ increases for both estimators and low values of $K_n$, but only for the CSA 2SLS when $K_n$ is large. This is in line with the existing literature, since 2SLS is not consistent for large values of $K_n$. Therefore, CSA 2SLS is clearly superior in this case.

Another feature that deserves a comment is that the variation of the coefficients that explain $x_i$ in terms of $z_i$ make a difference (i.e., the value of the parameter $c$). This effect seems to work in opposite directions for the CSA 2SLS and 2SLS. For CSA 2SLS, small variation seems to improve performance, whereas large variation seems to do so for 2SLS. However, this is only a small sample effect, with both estimators performing very similarly for large values of $n$ and small values of $K_n$, and CSA 2SLS outperforming the standard 2SLS for large values of $n$ and $K_n$, in accordance with our asymptotic results. Note further that some variation in the coefficients is needed for the CSA 2SLS to be consistent according to Theorem 1 (see also Examples 1 and 2).

As fas as the performance of the Factor 2SLS is concerned, it is very poor, even worse than standard 2SLS, in line with the findings of Kapetanios and Marcellino (2006) for the case of a very weak factor structure. Two additional comments are worth making. First, for fixed $K_n$ the performance improves with the sample size $n$. Second, for increasing $K_n$ the variance of the Factor 2SLS estimator increases, in line with its non consistency in this case.

Overall, the conclusion is clear: CSA 2SLS systematically outperforms 2SLS when the main condition of Theorem 1 is satisfied, as is the case in our Monte Carlo study, and it is better than the Factor 2SLS in the absence of a clear factor structure for the large set of instruments.

# 4 Empirical Examples

In this Section we discuss two empirical applications of the CSA IV estimation. The former concerns estimation of a forward looking Taylor rule, along the lines of Clarida, Galí, and Gertler (1998) (CGG), Clarida, Galí, and Gertler (2000) (CGG2)) and Favero, Marcellino, and Neglia (2005). The latter focuses on estimation of a New-Keynesian Phillips curve, along the lines of Galí and Gertler (1999) (GG 1999) and Beyer, Farmer, Henry, and Marcellino (2005). Kapetanios and Marcellino (2006) (KM) have considered Factor IV estimation of the parameters of these two equations, and shown that it produces efficiency gains with respect to standard IV. Here we are particularly interested in the comparison among standard, Factor and CSA IV. More precisely, since the underlying economic models are fairly complicated, we will use GMM estimation with standard variables, cross sectional averages or factors as instruments. The extension of the theoretical results from IV to GMM is straightforward, see e.g. Kapetanios and Marcellino (2006) for details on the Factor IV case.

## 4.1 Taylor rule

For the Taylor rule, we adopt the following specification[1] :

$$r_t = \alpha + (1-\rho)\beta(\pi_{t+12} - \pi_t^*) + (1-\rho)\gamma(y_t - y_t^*) + \rho r_{t-1} + \epsilon_t, \tag{5}$$

where $\epsilon_t = (1-\rho)\beta(\pi_{t+12}^e - \pi_{t+12}) + v_t$, and $v_t$ is an i.i.d. error. We use the federal funds rate for $r_t$, annual cpi inflation for $\pi_t$, 2% as a measure of the inflation target $\pi_t^*$, and the potential output $y_t^*$ is the Hodrick Prescott filtered version of the IP series. Since $\pi_{t+12}$ is correlated with the error term $\epsilon_t$, and the error term has an MA structure, we adopt GMM estimation with a correction for the MA component in the error $\epsilon_t$ and a proper choice of instruments.

In particular, as in KM, we use a HAC estimator for the weighting matrix, based on a Bartlett kernel with Newey and West (1994) automatic bandwith selection. For the set of instruments, in the base case the choice is similar to that in CGG and CGG2. We use one lag of the output gap, inflation, commodity price index, unemployment and interest rate. We focus on the period 1985-2003, since Beyer, Farmer, Henry, and Marcellino (2005) have detected instability in Phillips curves and Taylor rules estimated on a longer sample with an earlier start date.

For the Factor GMM estimator, as in KM, we add to the set of instruments the (one period lagged) factors extracted from a large dataset of 132 monthly macroeconomic and financial variables for the US, extracted from the dataset in Stock and Watson (2005). The number of factors is eight, as indicated by the Bai and Ng (2002) criteria, which suggests that the factor structure is rather weak. We also consider a subset of 12 of the 132 variables, those with an absolute correlation with inflation higher than 0.40, since this can strenghten the factor structure and improve the information content of the factors for future inflation. In fact, in this case one factor explains over 60% of the variance of all variables, and we use one to twelve lags of this factor as instruments, in addition to those in the base case.

For the CSA GMM, we add to basic set of instruments the simple average of either all the (standardized) 132 macroeconomic variables, or of only the subset of 12 variables mostly correlated with inflation. In both cases, we included one to twelve lags of the averages as instruments.

[1] Note that for this subsection which deals with time series data we change notation so that the observation index is $t$ rather than $i$.

Finally, we also considered one lag of the 12 selected macroeconomic variables as instruments, to compare the performance of standard and CSA IV with a relatively small set of instruments.

The results from the six estimation methods (Base, Factor-GMM All data, CSA-GMM All data, Factor GMM Select data, CSA GMM Select data, and Select data as instruments) are reported in Table 2. For the base case, which is the same as in KM, the estimated values for $\beta$ and $\gamma$ are, respectively, about 2.3 and 1, and the fact that the output gap matters less than inflation is not surprising. The persistence parameter, $\rho$, is about 0.88, in line with other studies. An LM test for the null hypothesis of no correlation in the residuals of an MA(11) model for $\widehat{\epsilon}_t$ does not reject the null hypothesis, which provides evidence in favor of the correct dynamic specification of the Taylor rule in (5). The p-value of the J-statistic for instrument validity is 0.11, so that the null hypothesis is not rejected at the conventional level of 10%.

Adding the "All data" factors to the instrument set does not improve the precision of the estimators of $\rho$, $\gamma$ and $\beta$. Instead, the CSA GMM using "All data" produces a major reduction in the variance of the estimators, about 100% for $\rho$, 20% for $\gamma$, and 15% for $\beta$. This suggests that CSA GMM can be useful in cases where the large set of instruments presents a weak factor structure, in line with the results of the Monte Carlo experiments.

Using the "Select data" factors, the precision of the Factor GMM improves, and becomes comparable to that of the CSA GMM based on the "Select data". The ranking between "All data" and "Select data" for CSA GMM is not clear cut.

Using directly the lagged "Select data" as additional instruments produces bad results in terms of variances of the estimators, even worse than in the base case for $\gamma$ and $\beta$. The point parameter estimates are also fairly different from the other five cases. These findings indicate that GMM estimation based on 18-20 macroeconomic instruments can already be problematic.

Finally, a regression of future (12 months ahead) inflation on the alternative sets of instruments indicates that each set of factors is significant at the 10% level when added to the macro variables, while the CSA from the "All data" are not, and those from "Select data" only marginally so. However, a few of the lagged CSA variables are strongly significant in both cases. Moreover, the values of the adjusted $R^2$ in these equations are all of comparable size.

## 4.2 Phillips curve

For the second empirical example, as in KM, the New-Keynesian Phillips curve is specified as,

$$\pi_t = c + \gamma\pi_{t+1} + \alpha x_t + \rho\pi_{t-1} + \epsilon_t, \tag{6}$$

where $\epsilon_t = \gamma(\pi^e_{t+1} - \pi_{t+1}) + v_t$, and $v_t$ is an i.i.d. error. Moreover, $\pi_t$ is annual CPI inflation, $\pi^e_{t+1}$ is the forecast of $\pi_{t+1}$ made in period $t$, and $x_t$ is unemployment, with reference to Okun's law, as in e.g. Beyer and Farmer (2003).

As for the Taylor rule, $\pi_{t+1}$ is correlated with the error term $\epsilon_t$, which in turn is correlated over time. Hence, we estimate the parameters of (6) by GMM, with a correction for the MA component in the error $\epsilon_t$, and the same six sets of instruments as for the Taylor rule.

The results are reported in Table 3. For the base case, the coefficient of the forcing variable is not statistically significant (though it has the correct sign), while the coefficients of the backward and forward looking components of inflation, $\rho$ and $\gamma$, are similar and close to 0.5.

Adding the "All data" factors to the instrument set improves the precision of the estimators of all parameters, but the gains are much larger with the "Select" data factors. For the latter, the gains are about 10% for $\alpha$ and 120% for $\gamma$ and $\rho$. Moreover, a regression of future (1 month ahead) inflation on the instruments indicates that only the Select data factors are strongly significant when added to the set of macroeconomic regressors.

As for the Taylor rule, the CSA GMM based on "All data" performs much better than the corresponding Factor GMM. However, CSA and Factor GMM based on "Select data" produce very similar results in terms of both point estimates of the parameters, and the variances of the estimates. The CSA from "Select data" are also strongly jointly significant in a regression of future (1 month ahead) inflation on the instruments.

Finally, in this case using directly the "Select data" as instruments is slightly better in terms of efficiency than the base case, but much worse than either CSA or Factor GMM.

In summary, the two empirical examples in this Section confirm that CSA GMM is often better than standard GMM. It can be even better than Factor GMM, in particular when the factor structure is weak.

# 5 Conclusions

Instrumental variable estimation is central to econometric analysis and has justifiably been receiving considerable and consistent attention in the literature in the past. Recent developments have focused on the cases where instruments are either weak, in terms of correlations with the endogenous variables, or many or both.

A clear conclusion of past work is that the number of instruments can be too large in the sense that too many instruments can make estimators inconsistent. The exact conditions on the number of instruments is closely related to the extent to which instruments are weak, making the two issues closely interlinked. The case for parsimony in this context has been made convincingly, in an interesting paper by Hahn (2002), which advocates parsimony as a prerequisite for optimal inference when a large number of instruments is available.

In a similar spirit as Hahn (2002), the present paper suggests a new way to deal with many, possibly weak, instruments. Our suggestion is to cross-sectionally average the instruments and use these averages as instruments. We have provided a theoretical analysis of this approach in terms of its consistency properties and also showed, via a Monte Carlo study and two detailed empirical applications, that the approach can provide improved estimation and inference compared to standard instrumental variables estimation.

# References

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

Beyer, A., and R. Farmer (2003): "On the indeterminacy of New-Keynesian economics," *ECB Working Paper No. 323.*

Beyer, A., R. Farmer, J. Henry, and M. Marcellino (2005): "Factor analysis in a new-Keynesian model," *ECB Working Paper No. 510.*

Boivin, J., and S. Ng (2006): "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 127, 169–194.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450.

Chao, J. C., and N. R. Swanson (2004): "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments," in *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, ed. by D. Corbae, S. Durlauf, and B. Hansen. Cambridge, U.K.: Cambridge University Press.

——— (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692.

Clarida, R., J. Galí, and M. Gertler (1998): "Monetary policy rules in practice: Some international evidence," *European Economic Review*, 42, 1033–1067.

——— (2000): "Monetary policy rules and macroeconomic stability: evidence and some theory," *Quarterly Journal of Economics*, 115, 147–180.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

Favero, C., M. Marcellino, and F. Neglia (2005): "Principal components at work: the empirical analysis of monetary policy with large datasets," *Journal of Applied Econometrics*, Forthcoming.

Galí, J., and M. Gertler (1999): "Inflation Dynamics: A structural econometric approach," *Journal of Monetary Economics*, 44, 195–222.

Hahn, J. (2002): "Optimal Inference with Many Instruments," *Econometric Theory*, 18, 140–168.

Hahn, J., J. Hausman, and G. Kuersteiner (2001): "Accuracy of Higher Order Bias and MSE Approximations," *Econometrics Journal*, 7, 272–306.

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

Hansen, C., J. Hausman, and W. K. Newey (2006): "Many Instruments, Weak Instruments and Microeconometric Practice," *Working Paper, MIT.*

Kapetanios, G., and M. Marcellino (2006): "Factor-GMM Estimation with Large Sets of Possibly Weak Instruments," *Queen Mary Working Paper No. 577.*

Morimune, K. (1983): "Approximate Distributions of $k$-class Estimators when the Degree of Overidentification is Large Compared with the Sample Size," *Econometrica*, 51, 821–842.

Newey, W., and K. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Newey, W. K. (2004): "Many Weak Moment Asymptotics for the Continuously Updated GMM Estimator," *Working Paper, MIT.*

Phillips, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics, Vol. I*, ed. by Z. Griliches, and M. D. Intriligator, p. 449516. Amsterdam: North-Holland.

Poskitt, D. S., and C. L. Skeels (2002): "Assessing Instrumental Variable Relevance: An Alternative Measure and Some Exact Finite Sample Theory," *University of Melbourn Working Paper No. 862.*

Rothenberg, T. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches, and M. D. Intriligator, pp. 881–935. Amsterdam: North-Holland.

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

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H., and M. W. Watson (2005): "Implications of dynamic factor models for VAR analysis," *mimeo.*

Stock, J. H., and M. Yogo (2003a): "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge, U.K.: Cambridge University Press.

——— (2003b): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock, pp. 80–108. Cambridge, U.K.: Cambridge University Press.

| Table 1: Monte Carlo results on the variance of alternative 2SLS estimators | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | CSA 2SLS | | | | | 2SLS | | | | | Factor 2SLS | | | | |
| $c$ | n/K$_n$ | 30 | 50 | 100 | 200 | 400 | 30 | 50 | 100 | 200 | 400 | 30 | 50 | 100 | 200 | 400 |
| | 30 | 0.210 | 0.205 | 0.217 | 0.238 | 0.214 | 0.373 | 0.369 | 0.362 | 0.379 | 0.366 | 1.751 | 1.835 | 1.785 | 1.804 | 1.867 |
| | 50 | 0.157 | 0.167 | 0.160 | 0.152 | 0.156 | 0.276 | 0.358 | 0.364 | 0.357 | 0.366 | 1.550 | 1.699 | 1.782 | 1.878 | 1.900 |
| 0.1 | 100 | 0.104 | 0.103 | 0.103 | 0.104 | 0.105 | 0.178 | 0.245 | 0.353 | 0.363 | 0.363 | 1.496 | 1.570 | 1.646 | 1.665 | 1.906 |
| | 200 | 0.072 | 0.074 | 0.073 | 0.070 | 0.075 | 0.107 | 0.146 | 0.239 | 0.353 | 0.348 | 1.239 | 1.396 | 1.624 | 1.628 | 1.688 |
| | 400 | 0.051 | 0.052 | 0.050 | 0.053 | 0.050 | 0.066 | 0.090 | 0.144 | 0.234 | 0.354 | 1.148 | 1.188 | 1.345 | 1.404 | 1.533 |
| | 30 | 0.219 | 0.224 | 0.223 | 0.200 | 0.212 | 0.373 | 0.368 | 0.367 | 0.367 | 0.357 | 1.721 | 1.958 | 1.691 | 1.801 | 1.997 |
| | 50 | 0.146 | 0.154 | 0.158 | 0.155 | 0.153 | 0.269 | 0.361 | 0.359 | 0.363 | 0.352 | 1.441 | 1.506 | 1.789 | 1.651 | 1.952 |
| 0.2 | 100 | 0.100 | 0.109 | 0.103 | 0.101 | 0.105 | 0.169 | 0.242 | 0.354 | 0.356 | 0.356 | 1.353 | 1.742 | 1.757 | 1.750 | 1.630 |
| | 200 | 0.072 | 0.071 | 0.074 | 0.074 | 0.071 | 0.104 | 0.144 | 0.233 | 0.352 | 0.350 | 1.431 | 1.346 | 1.481 | 1.459 | 1.597 |
| | 400 | 0.049 | 0.051 | 0.050 | 0.050 | 0.050 | 0.063 | 0.086 | 0.143 | 0.235 | 0.352 | 1.206 | 1.312 | 1.480 | 1.411 | 1.648 |
| | 30 | 0.228 | 0.228 | 0.221 | 0.202 | 0.257 | 0.336 | 0.339 | 0.329 | 0.340 | 0.331 | 1.642 | 1.604 | 1.705 | 1.658 | 1.872 |
| | 50 | 0.161 | 0.159 | 0.163 | 0.157 | 0.152 | 0.251 | 0.333 | 0.324 | 0.327 | 0.324 | 1.372 | 1.558 | 1.588 | 1.761 | 1.777 |
| 0.5 | 100 | 0.103 | 0.107 | 0.107 | 0.105 | 0.105 | 0.157 | 0.214 | 0.322 | 0.329 | 0.324 | 1.312 | 1.513 | 1.647 | 1.570 | 1.754 |
| | 200 | 0.074 | 0.075 | 0.073 | 0.072 | 0.069 | 0.093 | 0.125 | 0.203 | 0.319 | 0.309 | 1.348 | 1.474 | 1.276 | 1.653 | 1.652 |
| | 400 | 0.050 | 0.051 | 0.050 | 0.054 | 0.050 | 0.056 | 0.075 | 0.127 | 0.204 | 0.320 | 1.233 | 1.186 | 1.331 | 1.481 | 1.503 |
| | 30 | 0.645 | 0.582 | 0.753 | 0.596 | 0.455 | 0.146 | 0.138 | 0.141 | 0.139 | 0.137 | 1.306 | 1.375 | 1.447 | 1.455 | 1.456 |
| | 50 | 0.400 | 0.346 | 0.272 | 0.278 | 0.471 | 0.097 | 0.135 | 0.131 | 0.130 | 0.131 | 1.125 | 1.292 | 1.172 | 1.148 | 1.448 |
| 2 | 100 | 0.227 | 0.323 | 0.157 | 0.132 | 0.140 | 0.060 | 0.080 | 0.124 | 0.126 | 0.120 | 1.190 | 1.086 | 1.201 | 1.151 | 1.396 |
| | 200 | 0.143 | 0.096 | 0.115 | 0.080 | 0.074 | 0.038 | 0.045 | 0.072 | 0.119 | 0.120 | 0.924 | 0.991 | 1.123 | 0.996 | 1.182 |
| | 400 | 0.096 | 0.062 | 0.058 | 0.056 | 0.052 | 0.025 | 0.028 | 0.040 | 0.067 | 0.122 | 0.889 | 0.853 | 0.948 | 0.974 | 1.222 |

Notes: The Monte Carlo design is $y_i = x_i + \epsilon_i$, $z_{ij} = e_{ij}, j = 1, ..., K_n$, $x_i = \sum_{j=1}^{K_n} K_n^{-1/2}(1 + \alpha_j)z_{ij} + u_i$, , where $e_{ij} \sim i.i.d.N(0,1)$, $\alpha_j \sim N(0, c^2)$ and $cov(e_{il}, e_{sj}) = 0$ for $i \neq s$ or $l \neq j$ . Let $\kappa_i = (\epsilon_i, u_i)'$. Then, $\kappa_i = P\eta_i$, where $\eta_i = (\eta_{1,i}, \eta_{2,i})'$, $\eta_{j,i} \sim i.i.d.N(0,1)$ and $P = [p_{ij}]$, $p_{ij} \sim i.i.d.N(0,1)$. The errors $e_{ij}$ and $u_s$ are independent for each $i$ and $s$. The standard 2SLS estimator uses the $z$ variables as instruments, the CSA 2SLS estimator uses their cross sectional average with weights 1/N, and the Factor 2SLS the first principal component of $z_1, ..., z_{K_n}$.

Table 2. Results for alternative GMM estimators of the parameters of a Taylor rule

| | | | | | | | | **First stage regression (infl+12)** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\rho$ | $\gamma$ | $\beta$ | **R2-adj** | **S.E. regr** | **Pval J-stat** | **R2-adj** | **S.E. regr** | **Pval F-stat** |
| **Base** | | 0.883 | 0.993 | 2.310 | 0.98 | 0.27 | 0.11 | 0.12 | 0.002 | |
| | st. err | 0.037 | 0.241 | 0.278 | | | | | | |
| **Factors** | | 0.908 | 1.261 | 2.905 | 0.98 | 0.27 | 0.13 | 0.15 | 0.002 | 0.05 |
| **All data** | st. err | 0.024 | 0.291 | 0.394 | | | | | | |
| **Average** | | 0.901 | 1.102 | 2.206 | 0.98 | 0.24 | 0.47 | 0.20 | 0.002 | 0.47 |
| **All data** | st. err | 0.018 | 0.189 | 0.228 | | | | | | |
| **Factors** | | 0.884 | 1.122 | 2.251 | 0.98 | 0.27 | 0.52 | 0.15 | 0.002 | 0.08 |
| **Select data** | st. err | 0.028 | 0.233 | 0.204 | | | | | | |
| **Average** | | 0.877 | 1.086 | 2.353 | 0.98 | 0.28 | 0.50 | 0.15 | 0.002 | 0.10 |
| **Select data** | st. err | 0.030 | 0.227 | 0.216 | | | | | | |
| **All** | | 0.940 | 1.723 | 2.953 | 0.99 | 0.23 | 0.18 | 0.14 | 0.002 | 0.14 |
| **Select data** | st. err | 0.019 | 0.412 | 0.479 | | | | | | |

Notes: The estimated equation is $r_t = \alpha + (1-\rho)\beta(\pi_{t+12} - \pi_t^*) + (1-\rho)\gamma(y_t - y_t^*) + \rho r_{t-1} + \epsilon_t$ (see text for details). The parameters are estimated by GMM over 1986.01-2003.12. In the base case (no factors) the set of instruments used includes lags of the output gap, unemployment, inflation, interest rate and commodity price index. In the Factors cases, the SW factors are added to the instruments. In particular, in "All data" the (8) factors are extracted from the whole dataset; in "Select data" the (1) factor extracted from.a subset of the variables selected with the Boivin and Ng (2006) criterion. The number of factors is based on the Bai and Ng (2002) criteria for "All data", while it is set to one for "Select data". We use one lag of each factor, but 12 lags for the "Select data" factor. In the Average cases, the instruments are one to 12 lags of the simple average of the standardized variables in "All data" or in "Select data". In the "All select data" case, the instruments are one lag of all the variables selected with the Boivin and Ng (2006) criterion. The last three columns contain statistics related to the first-stage regression of the one-year ahead expected inflation on the set of instruments used. In particular, we report the adjusted $R^2$, the standard error of the regression and the F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.

| Table 3. Results on alternative GMM estimators of the parameters of a New Keynesian Phillips curve | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | **First stage regression (infl+1)** | | |
| | | $\alpha$ | $\gamma$ | $\rho$ | **R2-adj** | **S.E. regr** | **Pval J-stat** | **R2-adj** | **S.E. regr** | **Pval F-stat** |
| | | | | | | | | | | |
| **Base** | | -0.002 | 0.538 | 0.462 | 0.98 | 0.16 | 0.62 | 0.12 | 0.002 | |
| | st. err | 0.007 | 0.048 | 0.047 | | | | | | |
| | | | | | | | | | | |
| **Factors** | | -0.000 | 0.513 | 0.492 | 0.98 | 0.16 | 0.30 | 0.11 | 0.002 | 0.48 |
| **All data** | st. err | 0.006 | 0.038 | 0.038 | | | | | | |
| | | | | | | | | | | |
| **Average** | | -0.002 | 0.473 | 0.532 | 0.98 | 0.16 | 0.29 | 0.12 | 0.002 | 0.37 |
| **All data** | st. err | 0.006 | 0.030 | 0.029 | | | | | | |
| | | | | | | | | | | |
| **Factors** | | -0.002 | 0.500 | 0.509 | 0.98 | 0.15 | 0.12 | 0.23 | 0.002 | 0.00 |
| **Select data** | st. err | 0.006 | 0.021 | 0.020 | | | | | | |
| | | | | | | | | | | |
| **Average** | | -0.002 | 0.501 | 0.509 | 0.98 | 0.16 | 0.10 | 0.16 | 0.002 | 0.03 |
| **Select data** | st. err | 0.006 | 0.021 | 0.020 | | | | | | |
| | | | | | | | | | | |
| **All** | | -0.000 | 0.551 | 0.459 | 0.98 | 0.16 | 0.27 | 0.11 | 0.002 | 0.64 |
| **Select data** | st. err | 0.006 | 0.043 | 0.042 | | | | | | |
| | | | | | | | | | | |

Notes: The estimated equation is $\pi_t = c + \alpha(ur_t) + \gamma(\pi_{t+1}) + \rho\pi_{t-1} + \epsilon_t$ (see text for details). The parameters are estimated by GMM over 1986.01-2003.12. In the base case (no factors) the set of instruments used includes lags of the output gap, unemployment, inflation, interest rate and commodity price index. In the Factors cases, the SW factors are added to the instruments. In particular, in "All data" the (8) factors are extracted from the whole dataset; in "Select data" the (1) factor extracted from.a subset of the variables selected with the Boivin and Ng (2006) criterion. The number of factors is based on the Bai and Ng (2002) criteria for "All data", while it is set to one for "Select data". We use one lag of each factor, but 12 lags for the "Select data" factor. In the Average cases, the instruments are one to 12 lags of the simple average of the standardized variables in "All data" or in "Select data". In the "All select data" case, the instruments are one lag of all the variables selected with the Boivin and Ng (2006) criterion. The last three columns contain statistics related to the first-stage regression of the one-year ahead expected inflation on the set of instruments used. In particular, we report the adjusted $R^2$, the standard error of the regression and the F-test for the joint significance of the coefficients on factors, when factors are added to the baseline model.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London



Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm