Inference in the Presence of Redundant Moment Conditions and the Impact of Government Health Expenditure on Health Outcomes in England

Inference in the Presence of Redundant Moment Conditions and the Impact of Government Health Expenditure on Health Outcomes in England[1]

Martyn Andrews   Obbey Elamin   Alastair R. Hall[2]

University of Manchester

Kostas Kyriakoulis   Matthew Sutton

SAS Institute   University of Manchester

December 20, 2012

[1]This work was performed as part of a project entitled "Towards Improved Inferences in Health Economics Analyses Using Moment-Based Econometric Methods" supported by the National Institute of Health Research (NIHR) Research Methods Funding Scheme.

[2]Corresponding author: Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

Abstract

In his 1999 paper with Breusch, Qian and Wyhowski in the *Journal of Econometrics*, Peter Schmidt introduced the concept of "redundant" moment conditions. Such conditions arise when estimation is based on moment conditions that are valid and can be divided into two sub-sets: one that identifies the parameters and another that provides no further information. Their framework highlights an important concept in the moment-based estimation literature namely, that not all valid moment conditions need be informative about the parameters of interest. In this paper, we demonstrate the empirical relevance of the concept in the context of the impact of government health expenditure on health outcomes in England. Using a simulation study calibrated to this data, we perform a comparative study of the finite performance of inference procedures based on Generalized Method of Moment (GMM) and info-metric (IM) estimators. The results indicate that the properties of GMM procedures deteriorate as the number of redundant moment conditions increases; in contrast the IM methods provide reliable point estimators but the performance of associated inference techniques based on first order asymptotic theory, such as confidence intervals and overidentifying restriction tests, deteriorates as the number of redundant moment conditions increases. However, it is shown that bootstrap procedures can provide reliable inferences. We use IM and bootstrap methods to perform inference about the impact of government health expenditure on health outcomes in England.

Key words: Generalized Method of Moments, Info-metric estimation, Empirical Likelihood, Exponential Tilting

1 Introduction

The introduction by Lars Hansen (Hansen 1982) of Generalized Method of Moments (GMM) provided a method for obtaining estimators of the parameters of economic models based on the information in population moment conditions. Providing this information is both valid and (strongly) identifies the parameters, Hansen (1982) established the consistency and asymptotic normality of the estimator, and proposed a variant known as the "two-step" GMM estimator which is asymptotically efficient in class of semi-parametric estimators based on the population moment condition in question, see Chamberlain (1987).

In practice, the underlying economic/statistical model typically implies an array of possible moment conditions, and it has been recognized that the choice of which to use impacts on the comparative statistical properties of the resulting estimator. In essence, moment conditions contain differing amounts of information about the parameters of interest. To pursue this point further, we restrict attention to the class of population moment conditions associated with generalized instrumental variables (IV) estimation that is, in which the moment condition states the orthogonality of vector of instruments to a model residual. This is because this class of moment conditions is the most commonly encountered in econometrics and is the type involved in our analysis below.

A lot of attention has focussed on the two extreme cases, namely optimal instruments and weak instruments. Hansen (1985) characterized the asymptotic efficiency bound for IV, and, since then, various papers have examined how to construct so-called optimal instruments that achieve this bound in certain cases of interest.[1] However, a drawback to their use is that the construction of the optimal instrument can be complicated and may require additional assumptions about the data generation process beyond those implied by the economic model; this often proves a significant limitation and the use of optimal instruments is not common in empirical practice. At the other extreme is the weak instrument case. Following the insight in Nelson and Startz (1990), Staiger and Stock (1997) demonstrated that the standard first order statistical analysis of Hansen breaks down if the

[1] For references see the survey in Newey (1993) or Hall (2005)[Ch. 7.2].

instrument is weak; that is, the population moment condition provides insufficient information to (strongly) identify the parameters. Driven by a number of high profile empirical examples, the problem of inference in the presence of weak instruments has received a lot of attention in the literature.[2]

However, while both these extremes are of interest, they are not the only information scenarios of relevance in empirical applications. In his 1999 *Journal of Econometrics* paper with Breusch, Qian and Wyhowski, Peter Schmidt introduced the concept of redundant moments—or instruments—which represents an important information scenario that, in some sense, lies in between the two extremes described above. This covers the situation in which a subset of the instruments, z_1 say, lead to moment conditions that strongly identify the parameters and the remainder, z_2 say, provide no additional information. In such circumstances, Hansen's (1982) analysis still applies and implies that the first order asymptotic properties of the estimator are the same whether estimation is based on z_1 or z_1, z_2; in this case z_2 is said to be redundant given z_1. While this result implies no (first order) asymptotic cost to the inclusion of redundant instruments, there is evidence that the finite sample properties of IV are adversely affected by the inclusion of redundant moment conditions.

The concept of redundancy, as originally stated, is occasionally criticised for being unrealistically strict in the sense that z_2 provides *no* additional information beyond that in z_1. However, this seems pedantic to us: the key insight is to realize that there are situations where some instruments provide identification and most of the information, and the remainder of the instruments provide very little, for which redundancy, as defined above, is just the limit case.[3]

In this paper we illustrate these ideas using an important empirical example in which exactly this type of structure is present. For any economy, a key policy question is the extent to which the level of government expenditure on health influences the populations

[2]For a review of the weak instrument literature see Stock, Wright, and Yogo (2002) and Hall (2005)[Ch. 8.2].

[3]For example, the ideas can equivalently be expressed using the concept of *near*-redundancy as in Hall, Inoue, Jana, and Shin (2008).

health. Even though there are surprisingly few estimates of the elasticity of health outcomes with respect to government health expenditure in the literature, estimates of this parameter have been found by regressing (log) mortality on (log) health expenditures, typically exploiting cross-section variations in both variables by region/county/state etc. In the empirical example we use in this paper, the data are from England (in 2005–06) and the unit of observation is a so-called Primary Care Trust (PCT), of which there are 152. However, in this model, expenditures are correlated with the regression error because expenditure is determined by a funding rule that involves four key variables, one of which, a composite need index, is endogenous. As a result, OLS estimation is inappropriate as it leads to inconsistent estimators of the elasticity. However, instrumental variables estimation is feasible because the three other key variables in the funding rule are arguably exogenous and can be used as instruments. Given the construction of the funding rule, these three instruments are important determinants of expenditure. In addition to these three variables, it is possible to include other instruments, such as variables that are related to the needs index but not to mortality, which tend to be of lesser importance in the determination of expenditure. The funding rule instruments identify the parameters of interest and contain most of the information, whereas the remaining instruments add little extra information. Thus the funding rule instruments can be considered—and are referred to—as "stronger instruments" and the remaining instruments are "nearly redundant" given the stronger instruments.

The key question for policy maker is which instruments to use in the estimation: just the stronger instruments, or the stronger and nearly redundant instruments? Or put another way, does the presence of the near redundant instruments has an adverse effects on the GMM estimator when using small sample sizes as in our example? We explore this issue via a simulation study calibrated to our empirical example, and find evidence that the inclusion of the near redundant instruments does have an adverse effect.

This raises the question of how to proceed. One option is to use GMM with just the stronger instruments. However, a second option is to use a member of the class of info-metric

(IM) estimators that are argued to have better finite sample properties than GMM.[4] This class includes both Empirical Likelihood (EL) and Exponential Tilting (ET) estimators. We therefore also explore the performance of these IM estimators in our simulation study. While we find that IM estimators provide more reliable point estimates than GMM, we also find that first order asymptotic theory provides a poor approximation to the coverage probabilities of confidence intervals and rejection frequencies of model specification tests. However, we find these problems can be remedied by employing a bootstrap procedure proposed by Brown and Newey (2002) based on the probabilities obtained as part of the IM estimation.

We contrast the inferences based on GMM and IM, and find significant differences in the elasticity of interest. Because the elasticity can be interpreted as the "cost per life", if our estimates are taken at face value, this has important policy implications.

In Section 2 we report our own estimates of the elasticity of interest using English data, together with some background describing how funding is allocated in England. Section 3 provides the econometric analysis, and Section 4 reports the simulation study. In Section 5, we return to our empirical example. Section 6 concludes.

2 The impact of government health expenditure on health outcomes in England

From a policy perspective, a key question is whether, and, if so, to what extent, the allocation of funding to public sector health agencies can impact on population health. Such evidence is required to inform decisions about appropriate levels of overall funding and questions of distribution, such as whether differential allocation of funds can contribute to the reduction of inequalities in population health between areas. The extent to which the level of government expenditure on health influences the population's health is particularly important when assessing the wisdom of the UK Government's decision, in 2000, to increase

[4] IM estimators can be characterised as Generalized Empirical Likelihood (GEL) estimators (Smith 1997).

health expenditure to the EU average by 2006 (Appleby and Boyle 2000). It is also relevant for the measurement of public service productivity (ONS 2006).

As noted above, our empirical example comes from England. The National Health Service in England is financed almost entirely from national taxation. The Department of Health negotiates every year with HM Treasury over how much money the National Health Service can spend. The size of the budget in 2005-06 was £53.9 billion, which averaged at £1,097 per person.

The NHS is organised in geographical areas, with Primary Care Trusts taking responsibility for local administration and purchasing of services. These PCTs receive fixed annual budgets from central government and are required to meet their populations' expenditure needs on hospital and community-based services (including pharmaceuticals) and to improve their local population's health.

In England, a funding rule is used to allocate the overall budget to each PCT (DOH 2005). This funding rule creates shares of the overall budget for each PCT that reflect their population size, age and other measured need factors, and expected input prices. These target shares are used to calculate a "Distance From Target" (DFT) for each PCT, which measures the extent to which their actual share of the national budget last year differs from that indicated by their target share. All PCTs receive a minimum level of funding uplift and the residual funds are then distributed on the basis of the Distance From Target, with the most under-target PCTs receiving the largest increases in budget.

A PCT's budget can therefore be expressed as:

> Budget per head = (National budget per head) * (Age Index) * (Additional Needs Index) * (Input Price Index) * (DFT Index)

in which each of the four index adjustments takes a mean value of one.

The data are sourced from government websites. The health measure is a directly age-standardised mortality rate for the period 2005-2007, expressed as deaths per 100,000 European Standard population.[5] The funding variable is the 2005/6 allocations from the

[5] https://indicators.ic.nhs.uk/webview/

Unified exposition book: 2003/04, 2004/05 & 2005/06 PCT revenue resource limits.[6] The formula adjustments are those for the Hospital and Community Health Services element of the formula, taken from Table 5.12 of the same exposition book. The "Distances From Target" are the closing figures for 2005/6 taken from Table 4.2 of the same exposition book. The population counts used to calculate the allocations per head are based on the 2004 Attribution Data Set scaled to Office for National Statistics population projections.

In what follows, we estimate the following equation with PCT-level data:

$$ln(H) = \beta ln(E) + \text{controls} + u, \quad (1)$$

where $ln(H)$ denotes the log of the mortality rate, $ln(E)$ is the log of the allocation of health expenditure per head. The exact specification of our observed control variables is irrelevant at this stage. u denotes everything that is unobserved or not included in the model. The variable $ln(E)$ is potentially endogenous because it is easy to see why expenditure levels might be a function of historical mortality (reverse causality) and because expenditure levels may reflect unobserved area-specific effects (unobserved heterogeneity). It is easy to show that both sources of endogeneity mean that OLS is biased upwards.

With panel data, we might be able to deal with the latter, but the former can only be addressed using IV. But can suitable instruments be found? For this example, such variables naturally occur because of the funding rule duscussed above, namely: the Age Index (Z_1), the Additional Needs Index (N), the Input Price Index (Z_2), and the DFT Index (Z_3). Although N is endogenous, because it depends on historical mortality levels, the other three variables are arguably uncorrelated with u and can be used as instruments. In practice, the funding rule is not exact but Z_1, Z_2 and Z_3 are the main determinants of E; as a result, we refer to these variables as "stronger instruments" to reflect their relative importance in the determination of E. We also consider the inclusion of other other instruments, such as variables that are related to N (but not H), which tend to be of

[6] http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Allocations/DH_4000344

lesser importance in the determination of E; these variables we have already labelled "near redundant", and are similar to those used by Martin, Rice, and Smith (2008).

To give a flavour of the issues that this paper addresses, we estimate the model in Equation (1) using the following variables as controls: income deprivation among older people, education deprivation, and a constant term. The sample consists of the 152 PCT's in England in 2005-06. Summary statistics for the data variables are presented in Table 2.

When the model is estimated using OLS, β is estimated as 0.090 with a robust standard error of 0.064. If taken at face value, this elasticity would imply increases in health expenditure levels have a positive, but statistically, insignificant effect on mortality. From a policy perspective, this positive elasticity is counter intuitive. When re-estimated by 2SLS, using only the three "stronger instruments" for E, the estimate changes considerably, being -0.705 with a robust standard error of 0.245. In other words, increases in spending do have the expected negative effect on mortality, and the effect is significant, in spite of the increase in its standard error by a factor of 3.8.

This is an excellent example of where 2SLS works: the instruments can only plausibly work through the funding rule, and the upwards bias in OLS is ameliorated. The instruments are clearly not "weak" in the usual sense of the label, given the 3.8 factor. To illustrate the marginal contribution of some near redundant instruments, we re-estimate once more, adding the 7 further instruments (see Table 2 for full descriptions). Note that the R-squared from the first-stage regression with only the three stronger instruments is 0.793, and this rises to 0.833 when the 7 near redundant instruments are added. Now the 2SLS estimate is -0.587 with a robust standard error of 0.132. The question is, which estimate should the policy maker use? Or maybe use just a sub-set of the near redundant instruments? Given the small sample size, do ET and EL estimators give different answers? Do they perform better?

In the next two sections, we develop the appropriate econometric framework to address these issues and report what happens when we simulate the data generation process to assess the properties of the competing estimators. We then return to the above example in

Section 5.

Table 1: Summary statistics of variables in health expenditure example*

Variable	Mean	S.D.
Dependent variable		
Directly standardised mortality rate per 100,000: all causes ($ln(H)$)	614.5	76.33
Endogenous explanatory variable		
Allocation per head ($ln(E)$)	1,106	138.5
Controls		
Income deprivation among older people (proportion)	0.176	0.065
Education deprivation (proportion)	0.229	0.094
Stronger instruments		
Age index (Z_1)	0.994	0.051
Input price index (Z_2)	1.036	0.159
Distance from target (DFT) index (Z_3)	1.005	0.081
Near redundant instruments[a]		
A: Inflow of persons all ages (rate per 1,000 persons)	0.845	0.140
A: Outflow of persons all ages (rate per 1,000 persons)	0.864	0.183
B: Proportion of people aged 16+ who have never married	0.314	0.077
B: Proportion of people in households that own their home	0.693	0.118
B: Proportion of houses failing ODPM 'Decent Homes Standard'	0.346	0.053
C: Proportion of people aged 16-74 that have never worked	0.031	0.021
C: Proportion of people aged 16-74 that are long-term unemployed	0.011	0.005

* Notes: All variables are subsequently expressed in natural logarithms in the regressions.
[a] We collect the near redundant instruments into 3 groups later in the analysis, labelled A, B, and C.

3 Moment based inference and redundancy

The model in Section 2 fits the following generic linear specification:

$$y_t = x_t'\theta_0 + u_t, \qquad t = 1, 2, \dots T, \tag{2}$$

where y_t is the dependent variable, an observed scalar; x_t is a $(p \times 1)$ vector of observed explanatory (or regressor) variables; u_t is the unobserved error term. The t subscript indicates the observations pertain to the t^{th} member of the sample, and T denotes the sample size. The parameters of interest are denoted by the $p \times 1$ vector θ_0.

As noted above, IV involves the use of a set of variables as "instruments"; these are denoted by z_t, a $(q \times 1)$ vector of instruments. We assume these instruments are valid in the sense that they are orthogonal to the error so that the following population moment condition holds

$$E[z_t u_t(\theta_0)] \;=\; 0, \tag{3}$$

where $u_t(\theta) = y_t - x_t'\theta$. For IV to work, it must be the case (amongst other things) that there are at least as many instruments as parameters, and so we assume $q \geq p$. For ease of presentation we assume all variables, $v_t = (x_t', u_t, z_t')'$ are independently and identically distributed.

There are a number of ways in which the information in (3) can be exploited to produce estimators of θ_0. As discussed in the Introduction, we focus on Two Stage Least Squares, Generalized Method of Moments and the class of IM estimators. Below we describe both the methods and also their statistical properties, with particular emphasis on the impact of redundant moment conditions on the latter.

3.1 2SLS and GMM estimation

As the name suggests, 2SLS was originally presented as a method of IV estimation based on two least squares estimations; for example see Greene (2003)[p.398-400]. However, 2SLS can also be derived as a special case of GMM and so for economy of presentation we focus solely on the latter framework.

The GMM estimator based on (3) is defined to be:

$$\hat{\theta}_T \;=\; argmin_{\theta \in \Theta} Q_T(\theta)$$

where

$$Q_T(\theta) \;=\; g_T(\theta)'\, W_T\, g_T(\theta),$$

$g_T(\theta) = T^{-1}\sum_{t=1}^{T} z_t(y_t - x_t'\theta)$, and W_T is known as the "weighting matrix".

For the method to work, the weighting matrix needs to satisfy certain restrictions.[7] Nevertheless, there are many candidates for the weighting matrix, and, in general, the value of $\hat{\theta}_T$ depends on this choice. If we set

$$W_T = (T^{-1}\sum_{t=1}^{T} z_t z_t')^{-1}$$

then the GMM estimator based on (3) equals the 2SLS estimator. While the GMM estimator is consistent for θ_0 for all valid choices of weighting matrix, the specific choice of W_T affects inferences about θ_0 through the large sample variance of $\hat{\theta}_T$. It is therefore desirable to choose the weighting matrix that yields the smallest large sample variance for $\hat{\theta}_T$. Based on this criterion, and under the conditions assumed here, the optimal choice of W_T is a matrix converging in probability to $W = \{Var[z_t u_t]\}^{-1}$; the estimator obtained with this choice of weighting matrix is referred to as the "optimal GMM" estimator. As an important consequence, it follows that if u_t is conditionally homoscedastic given z_t, then the optimal choice of weighting matrix yields the 2SLS estimator, and so the latter is efficient in large samples. But if u_t is conditionally heteroscedastic given z_t, then 2SLS is inefficient in large samples relative to the optimal GMM, and is therefore sub-optimal. The latter provides the motivation for using GMM rather than just 2SLS to implement IV.

First order asymptotic analysis

To develop our analysis, we impose a number of conditions that are collected together in the following assumption.

Assumption 1 *(i)* $(u_t, x_t', z_t')'$ *are independently and identically distributed and* y_t *is generated via (2); (ii)* $E[z_t u_t] = 0$, *and* $rank\{E[x_t z_t']\} = p$*; (iii)* $Var[u_t|z_t] = \sigma^2$.

[7] W_T must be a positive semi-definite matrix that converges in probability to a positive definite matrix of constants; see, for example, Hall (2005)[Chap.1].

Assumption 1(ii) states that the instruments are both valid and relevant. Assumption 1(iii) states that the errors are conditionally homoscedastic (given the instruments) as a consequence of which 2SLS and optimal GMM are the same.

Under this assumption (and certain other regularity conditions), it can be shown that[8]

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, V_\theta). \quad (4)$$

This result can be interpreted as meaning the large sample distribution of $\hat{\theta}_T$ is approximately normal with a mean of θ_0 with variance V_θ.

Notice that the mean of this large sample distribution is the true value of the parameter provided that the instrument satisfies the generic conditions stated in Assumption 1. The variance of the large sample distribution, on the other hand, is affected by the choice of z_t. To demonstrate the nature of its dependence, we focus on the case that arises in our empirical examples, namely where there is one parameter of interest. Accordingly we partition $x_t = [w_t, z'_{1,t}]'$ where w_t is the scalar endogenous regressor and $z_{1,t}$ is the vector of controls, and then also partition θ conformably as $\theta = (\alpha, \phi')'$. Thus α is the scalar parameter of interest and ϕ is the $(p-1) \times 1$ vector of parameters on the control variables. Let $\hat{\alpha}_T$ be the GMM estimator of α_0 and V_α be the $(1,1)$ element of V_θ.[9]. We further partition $z_t = [z'_{1,t}, z'_{2,t}]$, where $z'_{2,t}$ are the identifying instruments, ie those not used as controls.

Proposition 1 *If Assumption 1 (and certain other regularity conditions) hold then:*

$$V_\alpha = \frac{\sigma^2}{\sigma_w^2}\left(\frac{1}{R^2_{w,z} - R^2_{w,z_1}}\right) \quad (5)$$

where R^2_{w,z_1} is the population multiple correlation coefficient from the regression of w_t on the controls $z_{1,t}$, $R^2_{w,z}$ is the population multiple correlation coefficient from the regression of w_t on z_t, and σ_w^2 denotes the variance of w.

[8] For example, see Hall (2005)[Chap. 2.3].

[9] It follows from (4) that $T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, V_\alpha)$.

Proposition 1 reveals that the large sample variance of the estimator depends inversely on additional explanatory power (over that of the controls) of the instruments $z_{2,t}$ for w_t. Thus the greater the additional explanatory power of $z_{2,t}$ then the more precise the estimator.

To consider the implications of this first order asymptotic result for the issue of instrument selection, we frame our discussion in terms of considering the consequences of augmenting an existing set of instruments z_t by the addition of one extra instrument denoted $z_{3,t}$. From Proposition 1, it follows that the addition of $z_{3,t}$ to the instrument vector can never increase V_α because $R^2_{w,[z,z_3]} \geq R^2_{w,z}$ by construction. If the inclusion of $z_{3,t}$ to the instrument vector has no effect on V_α (that is, $R^2_{w,[z,z_3]} = R^2_{w,z}$) then $z_{3,t}$ is said to be a *redundant* instrument for estimation of α_0 given z_t. Thus, a first order asymptotic analysis suggests that the inclusion of an extra instrument can never hurt and may help.

Second order asymptotic analysis

For this part of our analysis, we impose one additional condition.

Assumption 2 *(i)* $E[u_t^3|z_t] = 0$*; (ii)* $E[u_t v_t|z_t] = \sigma_{uv} \neq 0$.

Here, v_t is the implied reduced-form error term. Part (i) of this assumption states that the errors are symmetrically distributed conditional on z_t; part (ii) states implies that the covariance of u_t and v_t is non-zero. Using results in Newey and Smith (2004), we can show the following.

Proposition 2 *If Assumptions 1, 2 (and certain other regularity conditions) hold then:*

$$bias(\hat{\alpha}_T) = \frac{(q-p-1)\sigma_{uv}}{T\sigma_w^2}\left(\frac{1}{R^2_{w,z} - R^2_{w,z_1}}\right). \tag{6}$$

Proposition 2 reveals that the second order bias depends on the number of instruments (q), the explanatory power (over that of the controls) of the instruments for w_t ($R^2_{w,z} - R^2_{w,z_1}$), the covariance between u_t and v_t (σ_{uv}) and the sample size (T).

To consider the implications of this second order asymptotic result for the issue of instrument selection in our examples, we again frame our discussion in terms of considering the consequences of augmenting an existing set of instruments z_t by the addition of one extra instrument denoted $z_{3,t}$. Inspection of the formula for $bias(\hat{\alpha}_T)$ in Proposition 2 it can be seen that introduction of an additional instrument impacts both the numerator (by increasing q) and the denominator (by increasing $R^2_{w,z}$). The outcome is therefore ambiguous except in one special case: if $z_{3,t}$ is redundant given z_t then the denominator of the bias term is unaffected by the introduction of $z_{3,t}$ but the numerator increases, meaning the bias must also increase. However, note that this bias disappears as T increases: thus, *ceteris paribus*, the larger the sample, the less the bias.

3.2 Info-metric estimation

Concerns about the finite sample performance of GMM have led to interest in alternative methods of estimation based on the information in moment conditions. Leading examples of such estimators are Empirical Likelihood (EL) (Qin and Lawless, 1994 or Owen, 2001) and Exponential Tilting (ET) (Kitamura and Stutzer, 1997). While EL and ET can be derived from distinct estimation principles, it has been recognized that they have a common structure that has led to development of generic approaches of which both are special cases. The two such generic approaches are Generalized Empirical Likelihood (GEL) (Smith 1997) and Info-metric methods (Golan 2006). We focus on the second approach.

Within the Info-metric approach, the population moment condition (pmc) is viewed as a constraint on true probability distribution of data. If $\mathbf{M}$ is set of all probability measures then the subset that satisfies pmc for a given θ is

$$\mathbf{P}(\theta) \;=\; \left\{ P \in \mathbf{M} : \int f(v,\theta) dP \;=\; 0 \right\},$$

and the set that satisfies the pmc for all possible values of θ is

$$\mathbf{P} \;=\; \cup_{\theta\in\Theta}\mathbf{P}(\theta).$$

Estimation is based on the principle of finding the value of θ that makes $\mathbf{P}(\theta)$ as close as possible to true distribution of data.

To operationalize this idea, we work with discrete distributions. Let $p_t = P(v = v_t)$ and $P = [p_1, p_2, \ldots, p_T]$. Assuming no ties, the empirical distribution of the data is: $\hat{\mu}_t = T^{-1}$; let $\hat{\mu} = [\hat{\mu}_1, \ldots \hat{\mu}_T]$. The Info-metric (IM) estimator is then defined to be:

$$\hat{\theta}_{IM} \;=\; \arg\inf_{\theta}\; \rho_T(\theta, \hat{\mu})$$

where

$$\begin{aligned}
\rho_T(\theta, \hat{\mu}) &= inf_{\hat{P}}\; D(P \,\|\, \hat{\mu}), \\
\hat{\mathbf{P}}(\theta) &= \left\{ \hat{P} :\; p_t > 0, \sum_{t=1}^{T} p_t = 1, \sum_{t=1}^{T} p_t f(v_t, \theta) \right\},
\end{aligned}$$

and $D(\,\cdot \,\|\, \cdot)$ is a measure of distance. An interpretation of the estimator can be built up as follows. $\hat{\mathbf{P}}(\theta)$ is the set of all discrete distributions that satisfy the pmc for a given value of θ. $\rho_T(\theta, \hat{\mu})$ represents the shortest distance between any member of $\hat{\mathbf{P}}(\theta)$ and the empirical distribution for a particular value of θ. $\hat{\theta}_{IM}$ is the parameter value that makes this distance as small as possible over θ.

To implement the estimator, it is necessary to specify a distance measure. A popular choice in this literature is the Cressie and Read (1984) distance measure, defined as

$$D^{(\eta)}_{CR}(\mathbf{p}\|\mathbf{q}) \;=\; \frac{\eta}{1+\eta}\sum_{t=1}^{T} p_t \left\{ \left(\frac{p_t}{q_t}\right)^{\eta} - 1 \right\}$$

which is defined for $-\infty < \eta < \infty$. This distance measure nests EL and ET as special cases: $lim_{\eta\to 0} D^{(\eta)}_{CR}(\cdot\|\cdot)$ yields the ET optimand; $lim_{\eta\to -1} D^{(\eta)}_{CR}(\cdot\|\cdot)$ yields the EL optimand.

In terms of statistical properties, EL/ET are consistent and have same limiting distribution - and thus the same first order asymptotic properties - as optimal GMM. However, their second order asymptotic properties are different. Using the same set-up as before and the results in Newey and Smith (2004), we can show the second order bias properties of EL/ET are as follows.

Proposition 3 *If Assumptions 1, 2 (and certain other regularity conditions) hold then:*

$$bias(\hat{\alpha}_{IM}) = \frac{\sigma_{uv}}{T\sigma_w^2}\left(\frac{1}{R^2_{w,z} - R^2_{w,z_1}}\right) \tag{7}$$

A comparison of the results in Propositions 2 and 3 reveals that the denominator of the bias terms for GMM and IM estimators are the same but there is a crucial difference in the numerators: for GMM the numerator depends on the number of instruments, for IM it does not. So returning to the analysis of the consequence of including an extra instrument, the inclusion of $z_{3,t}$ never increases the absolute bias. So, for IM estimators, there are no potential negative consequences in terms of first or second order asymptotic properties from the inclusion of an additional instrument. This indicates that IM estimators can be expected to yield more reliable point estimators in moderate–sized samples.

4 Simulation study

4.1 Design

In the simulation, we mimic the key properties of our empirical health example introduced in Section 2. However, for ease of exposition, we dispense with all the controls in the regression model apart from the constant ($z_{1,t} \equiv 1$), and so the model being estimated is written

$$y_t = \alpha w_t + \phi_0 + u_t, \qquad t = 1, 2, \dots T. \tag{8}$$

In other words, there are $p = 2$ regressors. Because we need to distinguish our strong instruments from our near redundant instruments, we write the reduced-form explicitly as

$$w_t = z_{2,t}'\pi_2 + z_{3,t}'\pi_3 + \pi_0 + v_t, \qquad t = 1, 2, \ldots T. \tag{9}$$

$z_{2,t}$ is (3×1) vector of strong instruments, and $z_{3,t}$ is $(k \times 1)$ vector of near redundant instruments, with the reduced-form parameter vectors π_2 and π_3 having 3 and k elements respectively. In other words, there are $q = k + 4$ instruments in total, and the model is over-identified by $k + 2$. k is the first important parameter in the simulations, because the number of near redundant instruments has an ambiguous effect on the second order bias of $\hat{\alpha}$ identified by Proposition 2.

In the simulations, throughout we fix the following parameters as follows. First, $\alpha = -0.5$, so that the true elasticity of health outcomes with respect to health expenditure is negative; second, both constants ϕ_0 and π_0 are normalised to zero; third, $\pi_2 = 1_3$, so that the effect of the strong instruments on health expenditures is normalised to unity; and fourth, $\pi_3 = a1_k$, where 1_k is a $(k \times 1)$ vector of ones and a is a scalar. a is the second important choice parameter, because it captures the relative strength of the near redundant instruments compared with their stronger counterparts.

All $k + 3$ instruments are Normally distributed and are drawn independently of each other, each with a variance σ_z^2. σ_z^2 is the third choice parameter in the simulation design. The reduced-form error v_t and the regression error u_t are drawn independently of the $k+3$ instruments, but are jointly Normally distributed with the variance of v_t denoted σ_v^2, the variance of u_t normalised to unity, and the covariance between v_t and u_t denoted c. When c is non-zero, w_t is endogenous. c is the fourth choice parameter in the simulation design. As already explained, we restrict c to being positive because OLS is upwards biased. Also, because the correlation between u and v is c/σ_v, and is less than unity, c is ultimately restricted to $0 \leq c < \sigma_v$. Finally, all of $u_t, v_t, z_{2,t}, z_{3,t}$ have zero mean, which implies that y_t and w_t also have zero mean (given $\phi_0 = \pi_0 = 0$).

Of the 4 important parameters yet to be fixed, namely a, c, σ_z^2 and k, we note that the number of near redundant instruments k varies hereafter as 0, 4, 7, and 10. Although the other three parameters are allowed to vary, in what follows we report only what happens when $a = 1/\sqrt{10}$, $\sigma_z^2 = 1/4$, and $\sigma_v^2 = 3/2$.[10] Our choices are explained as follows. We choose $\sigma_v^2 = 3/2$ because this is what happens in the data. We then set the covariance to $c = 1$ to ensure a strong degree of endogeneity of the health expenditure variable w, with the correlation between u and v being $c/\sigma_v = 0.816$. Noting that

$$R^2_{w,z_2} = \frac{3\sigma_z^2}{3\sigma_z^2 + \sigma_v^2},$$

and setting $R^2_{w,z_2} = 1/3$ throughout, again because of the real data, this implies that $\sigma_z^2 = 1/4$ throughout. Next, note that

$$\sigma_w^2 = \sigma_z^2(3 + ka^2) \tag{10}$$

and

$$R^2_{w,[z_2,z_3]} = \frac{\sigma_z^2(3 + ka^2)}{\sigma_z^2(3 + ka^2) + \sigma_v^2}. \tag{11}$$

We now choose $ka^2 = 1$ so that the contribution of the near redundant instruments moves R^2_{w,z_2} from 1/3 to $R^2_{w,[z_2,z_3]} = 2/5$ when there are $k = 10$ near redundant instruments in the reduced form for w. Hence $a = 1/\sqrt{10}$ throughout. To check that these are sensible choices, when $k = 7$ and $\sigma_w^2 = 2.425$, the bias in the OLS estimator, c/σ_w^2, is 0.412. In other words, a true α of –0.5 is estimated, on average, as –0.088 using OLS, which is roughly consistent with the real data described in Section 2.

Table 2 summarises the population values of the key parameters for $k = 0, 4, 7, 10$, together with the second order population biases given in Propositions 2 and 3 above.

The table shows that magnitude of the second order bias for the IV and IM estimators is the same when there are no near redundant instruments, because $q - p - 1 = 1$. The bias

[10] A wider set of results is available on request.

Table 2: Summary of simulation design as number near redundant instruments varies

	Number near redundant instruments k			
	0	4	7	10
Variance of endogenous regressor σ_w^2 (Equation 10)	2.250	2.350	2.425	2.500
First stage $R^2_{w,[z_2,z_3]}$ (Equation 11)	0.333	0.362	0.381	0.400
Variance of IV estimator V_α (Equation 5)	1.333	1.176	1.081	1.000
Second order bias IV estimators (Equation 6)	0.00889	0.03921	0.05766	0.07233
Second order bias IM estimators (Equation 7)	0.00889	0.00784	0.00721	0.00667
First order bias OLS c/σ_w^2	0.444	0.425	0.412	0.400

* Data generation process given by Equations (8, 9). $T = 150$, $p = 2$, $q = k+4$, $\alpha = -0.5$, $\phi_0 = \pi_0 = 0$, $\pi_2 = 1_3$, $\pi_3 = a1_k$, and $\sigma_v^2 = 3/2$. In these simulations, $a = 1/\sqrt{10}$, $\sigma_z^2 = 1/4$, and $c = 1$.

is 0.00889. We now see what happens as more and more near redundant instruments are added. The variance of the endogenous regressor σ_w^2 increases, and so the first stage $R^2_{w,[z_2,z_3]}$ also increases. As the Propositions assert, the second order bias for the IM estimator falls, to 0.00667 for $k = 10$, whereas that for IV increases to 0.07233. For the latter, this is sizeable, as a true parameter of –0.5 will estimated as -0.4277 on average; in Equation (6), the effect of q in the numerator is outweighing the increase in the fit in the denominator.

We now examine the other properties of the IV estimators (2SLS and GMM) and IM estimators (ET and EL) assuming that the sample size is large whereas, in fact, it is a moderate $T = 150$. All estimations are performed using the Matlab® Optimization Toolbox. The EL and ET estimations utilize a GEL toolbox written by Kostas Kyriakoulis; this toolbox uses *fmincon* with the so-called *interior point* algorithm. The number of replications is $N = 1000$.

4.2 First order asymptotics

Table 3 summarises the properties of the estimators; that is, biases, coverage proportions, and rejection frequencies based on first order asymptotics (FOA).

The results for the 2SLS and GMM estimators are roughly the same throughout Table 3 because there is no heteroskedasticity in the simulation design. Given the analytical formulae already discussed, the biases reported in the first row of the table can be compared with Table 2: we see that the biases for the 2SLS/GMM estimators are close to their theoretical

Table 3: Coverage and rejection proportions assuming first order asymptotics

	2SLS				GMM			
	0	4	7	10	0	4	7	10
Bias[a]	0.006	0.043	0.052	0.066	0.006	0.042	0.052	0.067
Bias (median)[b]	0.014	0.051	0.059	0.069	0.015	0.050	0.059	0.071
Coverage prop, 95% nominal, t-stat[c]	0.930	0.882	0.837	0.773	0.925	0.845	0.807	0.732
Coverage prop, 99% nominal, t-stat	0.972	0.953	0.936	0.899	0.968	0.938	0.905	0.850
Rejection prop, 5% nominal, J[d]	0.046	0.075	0.066	0.079	0.042	0.053	0.045	0.047
Rejection prop, 1% nominal, J	0.009	0.017	0.017	0.019	0.005	0.008	0.007	0.009

	EL				ET			
	0	4	7	10	0	4	7	10
Bias	-0.012	-0.004	-0.010	-0.009	-0.012	-0.005	-0.010	-0.009
Bias (median)	-0.008	0.003	0.000	-0.002	-0.007	0.001	0.000	-0.003
Bias mean corrected est[e]	-0.003	0.003	-0.003	-0.004	-0.003	0.002	-0.003	-0.004
Coverage prop, 95% nominal, t-stat	0.944	0.914	0.906	0.889	0.943	0.909	0.899	0.882
Coverage prop, 99% nominal, t-stat	0.980	0.966	0.967	0.960	0.979	0.964	0.960	0.951
Rejection prop, 5% nominal, LR[f]	0.052	0.108	0.120	0.178	0.065	0.139	0.173	0.262
Rejection prop, 1% nominal, LR	0.010	0.028	0.041	0.064	0.015	0.044	0.072	0.115
Rejection prop, 5% nominal, LM	0.055	0.111	0.137	0.208	0.045	0.071	0.075	0.079
Rejection prop, 1% nominal, LM	0.010	0.032	0.045	0.087	0.007	0.015	0.019	0.018
Rejection prop, 5% nominal, W	0.055	0.111	0.137	0.208	0.070	0.165	0.232	0.350
Rejection prop, 1% nominal, W	0.010	0.032	0.045	0.087	0.019	0.077	0.121	0.215

[a] Bias is $N^{-1}\sum_r \hat{\alpha}_r - \alpha$, where $\hat{\alpha}_r$ is the estimate of α on the r-th replication.
[b] As [a], but using median rather than sample mean.
[c] Proportion of replications where $H_0 : \alpha = 0$ is not rejected using the t-statistic $\sqrt{T}\hat{\alpha}/\sqrt{\hat{V}_\alpha}$.
[d] Proportion of replications where usual GMM overidentifying restrictions test is rejected.
[e] Estimate of $\hat{\alpha}_r$ is bias-corrected using suggestion of Newey and Smith (2004).
[f] Same as [d], but for LR, LM, and Wald analogues of overidentifying restrictions test for ET/EL.

counterparts, but for ET/EL they have the wrong sign. The biases reported in the second row use the median rather then mean over $N = 1000$ replications; this is because the true sampling distributions might be non-symmetric. Now the biases for 2SLS/GMM get worse, whereas EL/ET get better.

For the EL/ET estimators, we implement a second-order bias-correction using a suggestion of Newey and Smith (2004). Now the biases for ET are very similar to EL, and are small, and so we conclude that our simulations confirm that EL/ET are not biased, even when there are $k = 10$ near redundant instruments. On the other hand, it is clear that the 2SLS/GMM estimators are biased, and the bias gets worse as more near redundant instruments are added.

We now examine whether we get correct inference when testing the null hypothesis that $\alpha = -0.5$. The table shows that the coverage proportions are all too small for $k = 0$ (for example, the ET coverage is 0.943 instead of 0.95 and is 0.979 instead of 0.99) and these get worse as k increases. This deterioration is much worse for the 2SLS/GMM estimators. When examining the standard overidentifying restrictions test, the so-called J-statistic, the table shows that the GMM estimator has the correct rejection proportions, and they are marginally worse for 2SLS. For the EL estimator, when there are no near redundant instruments, they are also correct, but slightly worse for ET. However, when k increases, for the ET/EL estimators they fall apart (for example, for ET and $k = 10$, the null hypothesis is rejected 35.0% of the time when it should be 5%).

Thus, while EL and ET provide accurate point estimators, inferences based on the first order asymptotic distribution of the estimator and overidentification restriction tests are unreliable in these samples when nearly redundant instruments are included. In the next sub-section, we explore whether a bootstrap can correct this problem.

4.3 Bootstrap

In this sub-section, we outline Brown and Newey's (2002) bootstrap procedure for constructing confidence intervals for the parameter estimators and performing the various versions of the overidentifying restrictions tests. For purposes of presentation, we let the random vector containing the data variables be $d = [y, w, z']'$, and d_t denote the t^{th} observation in the sample on d, that is $d_t = [y_t, w_t, z_t']'$. The statistics of interest are:

t-statistics:

$$\tau_{(\cdot)}(\hat{\theta}_j, \beta_{0,j}, \hat{V}_{jj}) \;=\; \left| \frac{\hat{\theta}_j^{(\cdot)} \;-\; \theta_{0,j}}{\sqrt{\hat{V}_{jj}}} \right|$$

for $(\cdot) = GMM$, EL, ET with $\hat{\theta}_j$ and $\hat{V}_{jj}$ denoting the appropriate estimator and its estimated first order asymptotic variance based on estimation method indicated by $(\cdot)$.

Overidentifying restrictions tests: $O_s^{(\cdot)}(d_1, d_2, \ldots d_T)$ where $(\cdot) = GMM, EL, ET$, and for $(\cdot) = GMM$ then $s = 1$ denotes the usual GMM overidentifying restrictions test, for $(\cdot) = EL$ or ET then $s = 1$ denotes the Wald statistic for the overidentifying restrictions, $s = 2$ denotes the LR statistic and $s = 3$ denotes the LM statistic.

Brown and Newey (2002) propose a version of the bootstrap based on GEL estimation of this model. To describe their procedure, we introduce the following definition.

- Let $\pi_t = P(d = d_t)$ and $\hat{\pi}_t$ denote the GEL estimator of π based on the sample.

For the purposes of the bootstrap, we treat d as discrete random vector with sample space $\mathcal{D}_T = \{d_t;\ t = 1, 2, \ldots, T\}$ and probability distribution function $P_T(d = d_t) = \hat{\pi}_t$.[11] The bootstrap samples are then created by sampling from replacement from this distribution for d. Let B be the total number of bootstrap samples generated, and index the bootstrap sample by b; so we have $b = 1, 2, \ldots, B$. Then, for each step of the bootstrap b, we proceed as follows.

1. On the b^{th} step of the bootstrap, draw a sample T observations $d_t^{(b)}$ with replacement from $P_T(d = d_t) = \hat{\pi}_t$.

[11] We ignore the possibility here that $d_t = d_s$ for some $t \neq s$ as this does not occur in our simulations or the health data.

2. Based on $\{d_t^{(b)};\ t = 1, 2, \ldots, T\}$, calculate:

   - the GEL and GMM estimators, denoted here by $\hat{\theta}^{(b)}_{GMM}$, $\hat{\theta}^{(b)}_{EL}$ and $\hat{\theta}^{(b)}_{ET}$.
   - the test statistics:
     - $\tau^{(b)}_{(\cdot)}(\hat{\theta}^{(b)}_j, \hat{\theta}_j, \hat{V}^{(b)}_{jj})$ where $\hat{\theta}^{(b)}_j$ and $\hat{V}^{(b)}_{jj}$ denote the appropriate estimator and its estimated first order asymptotic variance based on estimation method indicated by $(\cdot)$
     - $O_s^{(\cdot),b}(d_1^{(b)}, d_2^{(b)}, \ldots d_T^{(b)})$ where $(\cdot) = GMM, EL, ET$ and s is defined as above.

As a result of applying the bootstrap, this procedure generates a sampling distribution for each statistic of interest. It uses these distributions to provide bootstrap-based confidence intervals for the parameters and bootstrap-based p-values for the overidentifying restrictions tests as follows:

- *bootstrap-based confidence interval for* $\beta_{0,j}$: Let $\tau_b = \tau^{(b)}_{(\cdot)}(\hat{\theta}^{(b)}_j, \hat{\theta}_j, \hat{V}^{(b)}_{jj})$, that is the value of $\tau^{(b)}_{(\cdot)}(\hat{\theta}^{(b)}_j, \hat{\theta}_j, \hat{V}^{(b)}_{jj})$ based on the b^{th} bootstrap sample[12]. From the bootstrap, the procedure generates the following sampling distribution for $\tau_{(\cdot)}(\hat{\theta}_j, \beta_{0,j}, \hat{V}_{jj})$, $\{\tau_b\}_{b=1}^B$. Let q_α^B be the $100(1-\alpha)^{th}$ quantile of $\{\tau_b\}_{b=1}^B$: the $100(1-\alpha)\%$ bootstrap-based symmetric confidence interval for $\beta_{0,j}$ is:

$$\hat{\theta}_j \ \pm \ q_\alpha^B \sqrt{\hat{V}_{jj}}. \tag{12}$$

- *bootstrap-based p-values for the overidentifying restrictions tests*: To illustrate, consider the overidentifying restrictions test based on GMM, denoted $O_1^{GMM}(d_1, d_2, \ldots d_T)$ above. Put $O_b = O_1^{GMM}(d_1^{(b)}, d_2^{(b)}, \ldots d_T^{(b)})$, that is the value of the GMM overidentifying restrictions test based on the b^{th} bootstrap sample - again, for simplicity of notation, we suppress dependence (this time) on $(\cdot)$ and s. From the bootstrap, the procedure generates the following sampling distribution for $O_1^{GMM}(d_1, d_2, \ldots d_T)$,

[12]Note this depends on j but this suppressed to simplify the notation

$\{O^b\}_{b=1}^B$. Redefine q_α^B to be the $100(1-\alpha)^{th}$ quantile of $\{O^b\}_{b=1}^B$: then the bootstrap version of the decision rule for this test is as follows: reject $H_0 : E[z_t u_t(\theta_0)] = 0$ if $O_1^{GMM}(d_1, d_2, \ldots d_T) \geq q_\alpha^B$, that is test statistic from original data is compared to the appropriate quantile obtained from the bootstrapped sampling distribution.

Table 4 reports the outcomes, and shows that all the coverage and rejection probabilities are consistent with the nominal size.

Table 4: Coverage and rejection proportions using the bootstrap

	EL				ET			
	0	4	7	10	0	4	7	10
Coverage prop, 95% nominal, t-stat	0.958	0.941	0.944	0.942	0.955	0.941	0.938	0.938
Coverage prop, 99% nominal, t-stat	0.989	0.983	0.985	0.983	0.989	0.988	0.985	0.983
Rejection prop, 5% nominal, LR	0.038	0.051	0.045	0.039	0.043	0.057	0.047	0.045
Rejection prop, 1% nominal, LR	0.003	0.009	0.007	0.006	0.005	0.011	0.007	0.007
Rejection prop, 5% nominal, LM	0.043	0.057	0.045	0.043	0.039	0.041	0.029	0.019
Rejection prop, 1% nominal, LM	0.007	0.009	0.009	0.006	0.004	0.007	0.003	0.004
Rejection prop, 5% nominal, W	0.043	0.057	0.045	0.043	0.049	0.060	0.051	0.047
Rejection prop, 1% nominal, W	0.007	0.009	0.009	0.006	0.007	0.010	0.011	0.006

* See Tablenotes to Table 3. The number of bootstrap samples is $B = 999$.

5 The health expenditure example continued

We continue with the empirical example we introduced in Section 2 above. Table 5 reports results from 2SLS and GMM estimation of the model using various choices of instrument.[13] To recap Section 2, in contrast to the OLS estimate of 0.090 (0.064), the 2SLS estimate with no near redundant instruments is –0.705 (0.245) (see the column labelled "Base" and row labelled "2SLS"). We now report what happens when we add up to seven near redundant instruments, and re-estimate the models using GMM, EL and ET.

In the rest of the row labelled "2SLS", the near redundant instruments are added in

[13]We have applied standard tests available in the literature to confirm the relevance and validity of all choices of instruments considered here. Details are omitted for brevity.

Table 5: 2SLS, GMM, EL and ET estimates of β and standard errors*

Estimator	Base	A	B	C	AB	AC	BC	ABC[a]
2SLS	–0.705	–0.656	–0.705	–0.642	–0.608	–0.625	–0.641	–0.587
	(0.245)	(0.148)	(0.158)	(0.148)	(0.135)	(0.142)	(0.145)	(0.132)
GMM	–0.627	–0.593	–0.574	–0.518	–0.527	–0.536	–0.507	–0.511
	(0.200)	(0.192)	(0.175)	(0.172)	(0.162)	(0.171)	(0.160)	(0.157)
EL	–0.743	–0.787	–0.683	–0.658	–0.753	–0.694	–0.642	–0.734
	(0.218)	(0.233)	(0.192)	(0.191)	(0.204)	(0.200)	(0.179)	(0.196)
ET	–0.747	–0.786	–0.681	–0.665	–0.729	–0.709	–0.642	–0.710
	(0.218)	(0.231)	(0.190)	(0.192)	(0.197)	(0.201)	(0.178)	(0.189)
Bias corrected								
EL	–0.711	–0.756	–0.655	–0.633	–0.725	–0.670	–0.619	–0.708
ET[a]	–0.731	–0.770	–0.663	–0.653	–0.716	–0.698	–0.629	–0.698
Bootstrap-based p-values for the overidentifying restrictions tests[b]								
EL, LR	0.81	0.74	0.47	0.75	0.69	0.65	0.41	0.57
EL, LM	0.80	0.73	0.45	0.74	0.68	0.63	0.39	0.56
EL, Wald	0.80	0.73	0.46	0.74	0.68	0.64	0.39	0.56
ET, LR	0.80	0.71	0.47	0.74	0.65	0.63	0.41	0.53
ET, LM	0.83	0.77	0.55	0.77	0.72	0.68	0.49	0.61
ET, Wald	0.79	0.70	0.44	0.72	0.64	0.62	0.39	0.52
Bootstrapped-based confidence interval[c]								
EL (upper)	-0.204	-0.166	-0.171	-0.199	-0.216	-0.169	-0.175	-0.152
EL (lower)	-1.218	-1.347	-1.138	-1.068	-1.235	-1.172	-1.063	-1.264
ET (upper)	-0.160	-0.186	-0.088	-0.205	-0.193	-0.163	-0.119	-0.120
ET (lower)	-1.334	-1.387	-1.275	-1.125	-1.265	-1.254	-1.165	-1.299

* Notes: for definitions see Table 2. "Base" specification is 3 stronger instruments only. "A" adds 2 migration variables as near redundant instruments; "B" adds 3 further socioeconomic variables; "C" adds 2 further labour market variables; so that ... "ABC" adds all 7 variables as near redundant instruments.

[a] ET based on $B \approx 950$ because ET did not always converge.

[b] See Section 4.3 for full details. (The numbers of bootstrapped samples that were discarded are 41, 30, 51, 21, 38, 29, 43, and 54 resp.)

[c] See Equation (12).

groups, so that the final column has 7 near redundant instruments. Now the estimate is -0.587 (0.132). Recall that the first stage R-squared is 0.793 for the Base model and rises to 0.833 when there are 7 extra near redundant instruments. In the second row, all the models are re-estimated using GMM. All the estimates are smaller in absolute value, and have larger standard errors, except for the Base model.

The issue we have addressed in this paper is the fact that the 2SLS/GMM estimates are sensitive to the number of near redundant instruments. In particular, we note that the estimated elasticity tends to become smaller in absolute value as more instruments are included. By contrast, the EL/ET estimates exhibit far less sensitivity to the choice of instrument than their 2SLS/GMM counterparts. In particular, it is interesting to compare the estimates of the Base specification (stronger instruments only) with the "ABC" specification (all the instruments). For GMM, the estimates for Base are -0.627 and with "ABC" are -0.511; where as for EL they are -0.743 and -0.734 respectively, and for ET, -0.747 and -0.710 respectively. As is apparent, the EL and ET estimates are close and different from those obtained via 2SLS/GMM.

Given the insights from first and second order asymptotic theory described above, we believe that the EL/ET estimates are the more reliable. Our simulations also suggest that the Newey and Smith (2004) second order bias correction reduces the bias when $k = 0$; see the second panel of the table. However, our simulations also show that the "usual" inference techniques based on first order asymptotic theory are unreliable and this problem can be corrected using the bootstrap. Therefore, we apply Brown and Newey's (2002) procedure, described in Section 4.3, to our example. In the third panel, the p-values for the overidentifying restrictions tests all pass comfortably, and, in the fourth panel, we report the corresponding bootstrapped-based confidence intervals.

Of the second order biased corrected EL/ET estimates, which should we choose? Whilst the EL estimates range between –0.619 and –0.756 and the ET estimates range between –0.629 and –0.770, because this paper is concerned with the relevance of near redundant instruments, we focus on the ABC specifications, namely –0.708 for EL and –0.698 for ET.

These estimates are roughly in the middle of the 8 possibilities.

Thus we conclude that elasticity of mortality with respect to health expenditure is roughly -0.71. The equivalent GMM estimate is roughly -0.51, which we believe is biased because of the near redundant instruments. In terms of policy implications, the differences between the GMM and EL estimators can be demonstrated as follows: the GMM estimates imply that increasing NHS expenditure by 10% leads to a 5.1% reduction in deaths which translates to a cost per death averted of approximately £350,000, but the EL estimates imply that a 10% increase in NHS expenditure would lead to 7.3% fewer deaths which translates to a cost per death averted of £250,000.[14] However, the bootstrap confidence intervals suggest that there is considerable uncertainty about the estimate, with, for example, the EL estimate having a confidence interval of (–0.152, –1.264). The corresponding cost per death calculations turn out to be £1,120,000 and £140,000 respectively.

6 Concluding remarks

In his 1999 paper with Breusch, Qian and Wyhowski in the *Journal of Econometrics*, Peter Schmidt introduced the concept of "redundant" moment conditions. Such conditions arise when estimation is based on moment conditions that is valid and can be divided into two sub-sets: one that identifies the parameters and another that provides no further information. Their framework highlights an important concept in the moment-based estimation literature namely, that not all valid moment conditions need be informative about the parameters of interest.

In this paper, we demonstrate the empirical relevance of the concept in the context of the impact of government health expenditure on health outcomes in England because this is where exactly this type of structure is present. In estimating the elasticity of mortality

[14]These figures are obtained as follows. Using the summary statistics reported in Table 2 above, an increase in NHS expenditure by 10% would cost £107 per capita. The death probability is 0.006 and a 5.1% reduction in the death probability (from GMM) is a change of 0.000306. The cost per death averted is then £107/(0.000306) = £350,000 approximately. With the coefficient estimated from the EL (–0.71) the change in death probability is 0.000426, giving a cost per death averted of £250,000 approximately. In general, the cost per death averted is $£180,000/\beta$.

with respect to health expenditure using English data for 152 PCTs in 2005–06, the 2SLS estimate falls from –0.705 (0.245) with no near redundant instruments to −0.587 (0.132) when 7 are added. This raises the obvious question of which figure the policy maker should use.

Using a simulation study calibrated to these data, we perform a comparative study of the finite performance of inference procedures based on Generalized Method of Moment (GMM) and info-metric (IM) estimators. The results indicate that the properties of GMM procedures deteriorate as the number of redundant moment conditions increases; in contrast the IM methods provide reliable point estimators but the performance of associated inference techniques based on first order asymptotic theory, such as confidence intervals and overidentifying restriction tests, deteriorates as the number of redundant moment conditions increases. These results suggest that IM estimates combined with the Brown and Newey (2002) bootstrap provide reliable inferences.

When we return to the health example, we find that the IM point estimate implies a substantially lower cost per life saved than the GMM estimator. However, the bootstrapped confidence intervals suggest that there is considerable uncertainty about the estimate.

References

Appleby, J., and Boyle, S. (2000). 'Blair's billions: where will he find the money for the NHS?', *British Medical Journal*, March 25; 320(7238): 865867.

Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D. (1999). 'Redundancy of moment conditions', *Journal of Econometrics*, 91: 89–111.

Brown, B. W., and Newey, W. K. (2002). 'Generalized Method of Moments, efficient bootstrapping, and improved inference', *Journal of Business and Economic Statistics*, 20: 507–517.

Chamberlain, G. (1987). 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics*, 34: 305–334.

Cressie, N., and Read, T. R. C. (1984). 'Multinomial goodness-of-fit tests', *Journal of the Royal Statistical Society, series B*, 46: 440–464.

DOH (2005). 'Resource Allocation: Weighted Capitation Formula', Fifth Edition.

Golan, A. (2006). 'Information and entropy econometrics - a review and synthesis', *Foundations and Trends in Econometrics*, 2: 1–145.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, USA, fifth edn.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford, U.K.

Hall, A. R., Inoue, A., Jana, K., and Shin, C. (2008). 'Information in Generalized Method of Moments estimation and entropy based moment selection', *Journal of Econometrics*, 138: 488–512.

Hansen, L. P. (1982). 'Large sample properties of Generalized Method of Moments estimators', *Econometrica*, 50: 1029–1054.

——— (1985). 'A method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators', *Journal of Econometrics*, 30: 203–238.

Kitamura, Y., and Stutzer, M. (1997). 'An information-theoretic alternative to generalized method of moments estimation', *Econometrica*, 65: 861–874.

Martin, S., Rice, N., and Smith, P. C. (2008). 'Does health care spending improve health outcomes? Evidence from English programme budgeting data', *Journal of Health Economics*, 27: 826–842.

Nelson, C. R., and Startz, R. (1990). 'The distribution of the instrumental varaibles estimator and its t ratio when the instrument is a poor one', *Journal of Business*, 63: S125–S140.

Newey, W. K. (1993). 'Efficient estimation of models with conditional moment restrictions', in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, pp. 419–454. Elsevier Science Publishers, Amsterdam, The Netherlands.

Newey, W. K., and Smith, R. J. (2004). 'Higher order properties of GMM and generalized empirical likelihood estimators', *Econometrica*, 72: 219–255.

ONS (2006). 'Public Service Productivity: Health', Office of National Statistics, Economic Trends 628 March 2006.

Owen, A. (2001). *Empirical Likelihood*. Chapman Hall/CRC, New York, NY, USA.

Qin, J., and Lawless, J. (1994). 'Empirical likelihood and general estimating equations', *Annals of Statistics*, 22: 300–325.

Smith, R. J. (1997). 'Alternative semi-parametric likelihood approaches to generalized method of moments estimation', *Economics Journal*, 107: 503–519.

Staiger, D., and Stock, J. H. (1997). 'Instrumental variables regression with weak instruments', *Econometrica*, 65: 557–586.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics*, 20: 518–529.