Info-metric Methods for the Estimation of Models with Group-Specific Moment Conditions

Martyn Andrews* Alastair R. Hall*[1]

Rabeya Khatoon† James Lincoln*

The University of Manchester (*) & The University of Dhaka (†)

September 30, 2016

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

Abstract

Motivated by empirical analyses in economics using repeated cross-section data, we propose info-metric methods (IM) for estimation of the parameters of statistical models based on the information in population moment conditions that hold at group level. The info-metric estimation can be viewed as the primary approach to a constrained optimization. The estimators can also be obtained via the dual approach to this optimization, known as Generalized Empirical Likelihood (GEL). In a companion paper, we provide a comprehensive framework for inference based GEL with the grouped-specific moment conditions. In this chapter, we compare the computational requirements of the primary and dual approaches. We also describe the IM/GEL inference framework in the context of a linear regression model that is estimated using the information that the mean of the error is zero for each group. For the latter setting, we use analytical arguments and a small simulation study to showthat the IM/GEL approach to estimation yields more reliable inference in finite samples than certain extant methods.

Key words: Microeconometrics, repeated cross-section data, pseudo-panel methods, Generalized Empirical Likelihood, Generalized Method of Moments

1 Introduction

Microeconometrics involves the use of statistical methods to analyze microeconomic issues. In this context the prefix "micro" implies that these economic issues relate to the behaviour of individuals, households or firms. Examples include: how do households choose the amount of their income to spend on consumer goods and the amount to save? How do firms choose the level of output to produce and the number of workers to employ?

The answers to these questions start with the development of an economic theory that postulates an explanation for the phenomenon of interest. This theory is most often expressed via an economic model, which is a set of mathematical equations involving economic variables and certain constants, known as parameters, that reflect aspects of the economic environment such as taste preferences of consumers or available technology for firms. While the interpretation of these parameters is known, their specific value is not. Therefore, in order to assess whether the postulated model provides useful insights, it is necessary to estimate appropriate values for these parameters based on observed economic data.

For microeconometric analyses, there are three main kinds of data available: cross-section, panel (or longitudinal) and repeated cross-section. Cross-section data consists of a sample of information on individuals, say, taken at a moment in time. Panel data consists of a sample of individuals who are then observed at regular intervals over time. Repeated cross-section data consists of samples from a population of individuals taken at regular intervals over time. Unlike in the case of panel data, repeated cross-section data involves a fresh sample of individuals is taken each time period, and so the same individuals are not followed over time.

A number of statistical methods are available for estimation of econometric models. In choosing between them an important consideration is that the implementation of the estimation method should not require the imposition of restrictions on the statistical behaviour of the economics variables beyond those implied by the economic model. For if these additional statistical restrictions turn out to be inappropriate then this may undermine subsequent inferences about the economic question of interest. For example, the estimation method known as Maximum Likelihood (ML) requires requires the specification of the complete probability distribution of the data but typically this information is not part of an economic model. As a result, ML is not an attractive choice in this context. While economic models usually do not imply the complete probability distribution, they do imply restrictions on functions of both the economic variables and unknown parameters. These restrictions, known as population moment conditions, can provide the basis for estimation of the parameters.

Lars Hansen was the first person to provide a general framework for population moment based estimation in econometrics. In his seminal article in *Econometrica* in 1982, Hansen introduced the Generalized Method of Moments (GMM) estimation method.[1] GMM has been widely applied in economics but with this familiarity has come an understanding that GMM-based inferences may be unreliable in certain situations of interest. This has stimulated alternative methods for estimation based on population moment conditions. Leading examples are the continuous updating GMM estimator (CUE; Hansen, Heaton, and Yaron, 1996), empirical likelihood (EL; Owen, 1988; Qin and Lawless, 1994) and exponential tilting (ET; Kitamura and Stutzer, 1997). While all three can be justified in their own right, it has been realized that they can also be regarded as special cases of more general estimation principles: info-metric (IM; Kitamura, 2007; Golan, 2008) or generalized empirical likelihood (GEL; Smith, 1997).

Both GMM and IM/GEL can be straightforwardly applied in the case where the data are a random sample from a homogeneous population, as is typically assumed for cross-section and panel data. In this case, the comparative properties of GMM and IM are well understood: under certain key assumptions about the information content of the population moment conditions, both estimators have the same large sample (first order asymptotic) properties. However, IM estimators exhibit fewer sources of finite sample bias, and IM-based inference procedures are more robust to circumstances in which the information content of

[1] Hansen was co-winner of the 2013 Nobel prize for Economics for his work on empirical analysis of asset pricing models, and especially the development of GMM which has been widely applied in empirical finance.

the population moment conditions is low.

However, the case of repeated cross-section data has received far less attention in the literature on moment-based estimation, even though such data is prevalent in the social sciences.[2] While certain GMM approaches have been proposed, to our knowledge IM methods have not been developed.[3] Part of the reason may be due to the fact that the original IM/GEL framework applies to samples from homogeneous populations but this does not match the assumptions typically applied in econometric analysis of repeated-cross section data. For example, one popular method - Deaton's (1985) "pseudo-panel" approach - requires the population to consist of a number of different homogeneous sub-populations. In a recent paper, Andrews, Hall, Khatoon, and Lincoln (2015) (AHKL, hereafter) propose an extension of the GEL framework to allow for estimation and inference based on population moment conditions that hold within the sub-populations. Since the sub-populations are often associated with groups of individuals or firms, the estimator is referred to as GEL with group specific moment conditions, GEL-GMC for short. AHKL establish the consistency, first order asymptotic normality and second order bias properties of the GEL-GMC estimator, and also the large sample properties of a number of model diagnostic tests.

Our first contribution in this chapter is to provide an IM counterpart to the GEL-GMC estimator and to compare the computational requirements of the two approaches. Our second contribution is to describe the GEL-GMC based inference framework in the leading case in which a linear regression model with potentially group-specific parameters is estimated using the information that the expectation of the regression error is zero in each group. This allows us to make a direct comparison with a pseudo-panel estimator which is a popular approach to linear model estimation based on repeated cross-section data. Using both theoretical analysis and evidence from a simulation study, it is shown that the IM/GEL-GMC estimator yields more reliable inference than the pseudo-panel estimator

[2]For example, the UK Government's Data Service identifies key data sets for analysis of various issues relevant to public policy: of the 22 data sets identified for their relevance to environmental and energy issues 6 consist of repeated cross-sections; of the 34 data sets identified for the relevance to health and health-related behaviour 13 consist of repeated cross-sections; see http://ukdataservice.ac.uk/get-data/themes.aspx.

[3]See *inter alia* Bekker and van der Ploeg (2005), and Collado (1997), and Inoue (2008).

considered here.

An outline of the chapter is as follows. Section 2 provides a brief review of GMM and IM/GEL where the data are a random sample from a homogeneous population. Section 3 describes Deaton's (1985) pseudo-panel approach, demonstrating how it depends crucially on the population being non-homogeneous. Section 4 provides the IM version of AHKL's GEL-GMC framework. Section 5 describes the GEL-GMC framework in the context of the linear regression model estimated, and compares GEL-GMC to the pseudo-panel approach, and Section 6 concludes.

2 GMM and IM/GEL

In this section we briefly review the GMM and IM/GEL estimation principles for data obtained as a random sample from a homogeneous population.

Our econometric model is indexed by a vector of parameters that can take values in Θ, a compact subset of $\mathbb{R}^p$. We wish to estimate the true value of these parameters, denoted θ_0. The economic variables are contained in the random vector $\mathbf{v}$ with sample space $\mathcal{V}$ and probability measure μ. It is assumed that we have access to a random sample from this population, denoted $\{v_i;\, i = 1, 2, \ldots, n\}$.

We consider the case where estimation of θ_0 is based on the population moment condition (pmc),

$$E[f(\mathbf{v}, \theta_0)] = 0, \tag{1}$$

where $f : \mathcal{V} \times \Theta \to \mathbb{R}^q$.

The population moment condition states that $E[f(\mathbf{v}, \theta)]$ equals zero when evaluated at θ_0. For the GMM or IM/GEL estimation to have the statistical properties described below, this must be a unique property of θ_0, that is $E[f(\mathbf{v}, \theta)]$ is not equal to zero when evaluated at any other value of θ. If that holds then θ_0 is said to be *identified* by $E[f(\mathbf{v}, \theta_0)] = 0$. A first order condition for identification is that $rank\{G(\theta_0)\} = p$, where $G(\theta_0) = E[\partial f(\mathbf{v}, \theta)/\partial \theta'|_{\theta=\theta_0}]$, and this condition plays a crucial role in standard asymptotic distribution theory for these estimators; $G(\theta_0)$ is commonly known as the "Jacobian". By definition the moment condition involves q pieces of information about p unknowns, therefore identification can only hold if $q \geq p$. For reasons that emerge below it is convenient to split this scenario into two parts: $q = p$, in which case θ_0 is said to be *just-identified*, and $q > p$, in which case θ_0 is said to be *over-identified.*

We illustrate this condition with two popular examples in microeconometrics.

Example 1: Instrumental Variable estimation based on cross-section data

Suppose it is desired to estimate θ_0 in the model

$$\mathbf{y} \;=\; \mathbf{x}'\theta_0 \,+\, \mathbf{u}$$

where $\mathbf{y}$ is the dependent variable, $\mathbf{x}$ a vector of explanatory variables and $\mathbf{u}$ represents an unobserved error term. If $E[\mathbf{u}|\mathbf{x}] = 0$ then θ_0 can be estimated consistently via Ordinary Least Squares (OLS). However, in many cases in econometrics, this moment condition will not hold, with common reasons for its violation being simultaneity, measurement error or an omitted variable.[4] These problems are commonly circumvented by seeking a vector of variables $\mathbf{z}$ - known as an instrument - that satisfies the population moment condition $E[\mathbf{zu}] = 0$ and the identification condition $rank\{E[\mathbf{zx}']\} = p$.[5] In this case $\mathbf{v} = (\mathbf{y}, \mathbf{x}', \mathbf{z}')'$ and $f(\mathbf{v}, \theta) = \mathbf{z}(\mathbf{y} - \mathbf{x}'\theta)$. $\diamond$

Our second example involves panel data in which individuals are observed for a number of time periods. In view of this structure, it is most convenient to index the random variables by both i, indicating the individual, and t, the time period. As is most often the case in microeconometric applications, the number of time periods, T, is treated as fixed, and the sample becomes large through the number of individuals, n, going to infinity.

[4]These occur respectively if: $\mathbf{y}$ and $\mathbf{x}$ are simultaneously determined; the true model is $y = x_*'\theta_0 + e$ and $x = x_* + w$; explanatory variables have been omitted from the right hand side of the regression equation.

[5]As the pmc is linear in θ_0 identification and first-order identification are identical.

Example 2: Linear panel data models

Suppose it is desired to estimate the scalar parameter θ_0 in the panel data model

$$\mathbf{y}_{i,t} \;=\; \mathbf{x}_{i,t}'\theta_0 \;+\; \mathbf{u}_{i,t},\; i=1,2,\ldots,n;\; t=2,\ldots T \tag{2}$$

where $\mathbf{y}_{i,t}$ is the scalar dependent variable, $\mathbf{x}_{i,t}$ is a vector of explanatory variables, $\mathbf{u}_{i,t} = \mathbf{a}_i + \mathbf{w}_{i,t}$. In this case, the unobserved error is of composite form consisting of an individual effect, $\mathbf{a}_i$, and an idiosyncratic component, $\mathbf{w}_{i,t}$. For each individual, the idiosyncratic error is mean zero, serially uncorrelated, and uncorrelated with the individual effect and $\mathbf{x}_{i,t}$. The individual effect has mean zero, but is correlated with $\mathbf{x}_{i,t}$, and so is known as a "fixed effect". As a result, $\mathbf{x}_{i,t}$ is correlated with the error $\mathbf{u}_{i,t}$, and so OLS estimation of θ_0 based on (2) would yield inconsistent estimators.[6] However, it can be shown that the following moment conditions hold: $E[\Delta\mathbf{x}_{i,t}\Delta\mathbf{u}_{i,t}(\theta_0)] = 0$ where $\Delta\mathbf{u}_{i,t}(\theta) = \Delta\mathbf{y}_{i,t} - \Delta\mathbf{x}_{i,t}'\theta$ and Δ is the first (time) difference operator. The intuition behind the form of the moment conditions can be obtained by noting that $\Delta\mathbf{u}_{i,t}(\theta_0) = \Delta\mathbf{w}_{i,t}$ and so first differencing eliminates the fixed effect which is the source of the correlation between the error and regressors. Identification holds provided $E[\Delta\mathbf{x}_{i,t}(\Delta\mathbf{x}_{i,t})']$ is full rank. In this case, $\mathbf{v}_{i,t} = (\Delta\mathbf{y}_{i,t}, \Delta\mathbf{x}_{i,t}')'$ and $f(\mathbf{v}_{i,t},\theta) = \Delta\mathbf{x}_{i,t}\Delta\mathbf{u}_{i,t}(\theta)$. $\diamond$

The Generalized Method of Moments estimator based on (1) is defined as

$$\hat{\theta}_{GMM} \;=\; arg\min_{\theta\in\Theta} g_n(\theta)'W_n g_n(\theta),$$

where $g_n(\theta) \;=\; n^{-1}\sum_{i=1}^{n} f(v_i,\theta)$ is the sample moment, W_n is known as the *weighting matrix* and is restricted to be a positive semi–definite matrix that converges in probability to W, some positive definite matrix of constants. The GMM estimator is thus the value of θ that is closest to setting the sample moment to zero. The measure of distance for $g_n(\theta)$ from zero depends on the choice of W_n, and we return to this feature below.

[6] If $\mathbf{a}_i$ is uncorrelated with $\mathbf{x}_{i,t}$ then it is said to be a "random effect".

"Info-metric" stands for a combination of Information and Econometric theory, and captures the idea that this approach synthesizes work from these two fields. However, we note its implementation in this context is also referred to as "minimum discrepancy", see Corcoran (1998). Whichever way we refer to it, the key to this approach is that the pmc is viewed as a constraint on true probability distribution of data. If $\mathbf{M}$ is a 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 the true distribution of data. To operationalize this idea, we work with discrete distributions. Let $\pi_i = P(\mathbf{v} = v_i)$ and $P = [\pi_1, \pi_2, \ldots, \pi_n]$. Assuming no two sample outcomes for $\mathbf{v}$ are the same, the empirical distribution of the data attaches the probability of $1/n$ to each outcome. It is convenient to collect these empirical probabilities into a $1 \times n$ vector $\hat{\mu}$ whose elements are all $1/n$ and whose i^{th} element can thus be interpreted as the empirical probability of the i^{th} outcome. The IM estimator is then defined to be:

$$\hat{\theta}_{IM} \;=\; \arg \inf_{\theta \in \Theta} \mathcal{D}_n(\theta, \hat{\mu})$$

where

$$\begin{aligned} \mathcal{D}_n(\theta, \hat{\mu}) \;&=\; \inf_{\hat{P}} \; D(\hat{P} \,\|\, \hat{\mu}), \\ \hat{\mathbf{P}}(\theta) \;&=\; \left\{ \hat{P} : \; \pi_i > 0, \sum_{i=1}^{n} \pi_i = 1, \sum_{i=1}^{n} \pi_i f(v_i, \theta) \;=\; 0 \right\}, \end{aligned}$$

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. $\mathcal{D}_n(\theta, \hat{\mu})$ 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. Following Kitamura (2007), this distance is defined as $n^{-1} \sum_{i=1}^{n} \phi(n\hat{\pi}_i)$ where $\phi(\,\cdot\,)$ is a convex function.[7] As noted in the introduction, the IM framework contains a number of other estimators as a special case; for EL $\phi(\cdot) = -log(\cdot)$, for ET $\phi(\cdot) = (\cdot)log(\cdot)$, for CUE $\phi(\cdot) = 0.5[(\cdot) - 1]^2$. This IM approach emphasizes the idea of economic modles placing restrictions on the probability distribution of the data.

While the IM perspective is intuitively appealing, it is often more convenient for the purposes of developing the statistical theory to take the GEL perspective, which is essentially the dual of the IM approach, even though it was derived by Smith (1997) via a different route.[8] Smith (1997) defines the GEL estimator of θ_0 to be

$$\hat{\theta}_{GEL} \equiv \arg \min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n} C_n(\theta, \lambda),$$

where

$$C_n(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' f_i(\theta)) - \rho_0],$$

$\rho(a)$ is a continuous, thrice differentiable and concave function on its domain $\mathcal{A}$, an open interval containing 0, $\rho_0 = \rho(0)$, and λ is an auxiliary parameter vector restricted so that $\lambda' f_i(\theta) \in \mathcal{A}$ (with probability approaching one), for all $(\theta', \lambda')' \in \Theta \times \Lambda_n$ and $i = 1, ..., n$.[9] The auxiliary parameter vector λ is the Lagrange Multiplier on the constraint that the moment condition holds in the IM formulation. Once again particular choices of $\rho(\cdot)$ yield the CU, EL and ET estimators: $\rho(a) = log(1-a)$ for EL; $\rho(a) = -e^a$ for ET; $\rho(a)$ quadratic for CU. Within the GEL, the probabilities are defined implicitly. Smith (1997) and Newey

[7]This is known as the f-divergence between the two discrete distributions in this case $\{\hat{p}_i\}$ and $\{\hat{\mu}_i\}$.

[8]See Newey and Smith (2004), Kitamura (2007) and Parente and Smith (2014) for further discussion.

[9]Specifically, Λ_n imposes bounds on λ that "shrink" with n, but at a slower rate than $n^{-1/2}$ which is the convergence rate of the GEL estimator for λ.

and Smith (2004) show that the GEL estimator of π_i is given by

$$\hat{\pi}_{i,GEL} \;=\; \frac{\rho_1\left(\hat{\lambda}' f(v_i, \hat{\theta}_{GEL})\right)}{\sum_{i=1}^{n} \rho_1\left(\hat{\lambda}' f(v_i, \hat{\theta}_{GEL})\right)},$$

where $\rho_1(\bar{\kappa}) = \partial \rho(\kappa)/\partial \kappa|_{\kappa=\bar{\kappa}}$. Newey and Smith (2004) show that $\hat{\pi}_{i,GEL}$ is guaranteed to be positive provided $\hat{\lambda}' f(v_i, \hat{\theta}_{GEL})$ is uniformly (in i) small.

A crucial difference between GMM and IM/GEL optimizations is that the latter not only estimates θ_0 but also provides estimated probabilities for the outcomes in the data $\{\hat{\pi}_i\}$ that are constructed to ensure the sample analog to (1) is satisfied at $\hat{\theta}$. As a result, the estimated sample moment is set equal to zero in IM/GEL but is not in GMM. This difference turns out to be key for understanding key differences in the statistical properties of the estimators, as we now discuss.

If θ_0 is just-identified by the pmc then GMM and IM/GEL estimators are identical, being equal to the Method of Moments estimator based on (1). If θ_0 is over-identified then subject to certain regularity conditions - including identification and first-order identification - it can be shown that: (i) the GMM and IM/GEL are consistent for θ_0; (ii) the so-called "two-step" GMM estimator and IM/GEL have the large sample distributions given by: $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\,0,\, V_\theta\,)$ where $V_\theta = \{G(\theta_0)' S(\theta_0)^{-1} G(\theta_0)\}^{-1}$ and $S(\theta_0) = Var[f(\mathbf{v}, \theta_0)]$;[10] (iii) the two-step GMM and IM/GEL estimators achieve the semi-parametric asymptotic efficiency bound for estimation of θ_0 based on (1), see Chamberlain (1987).

While the first-order asymptotic properties of the estimators are the same, Newey and Smith (2004) show that their second-order properties are different. Specifically, they show that IM/GEL estimators have fewer sources of second-order bias than GMM, and that within the IM/GEL class, EL has the fewest sources of bias. This suggests that EL should

[10] The "two-step" GMM estimator is calculated using W_n that converges in probability to $S(\theta_0)$. Hansen (1982) shows this choice leads to GMM estimator based on (1) with the smallest variance in large samples. Its name comes from the fact that to implement this GMM estimator requires $W_n = \hat{S}^{-1}$ where $\hat{S} \xrightarrow{p} S$, and so a consistent estimator of θ_0 to form $\hat{S}$. As noted by Hansen, this can be achieved by the following two-step estimation procedure: estimate θ_0 by GMM with a sub-optimal choice of W_n and use this to construct $\hat{S}$, then re-estimate θ_0 by GMM with $W_n = \hat{S}^{-1}$.

exhibit the smallest second-order bias.[11] These differences can be traced to the form of the first order conditions associated with GMM and IM/GEL. Newey and Smith (2004) show that in each case the first order conditions take the form

$$(\text{Jacobian})' \times (\text{variance of sample moment})^{-1} \times \text{sample moment} \;=\; 0,$$

with the differences in the estimators arising from how the Jacobian and variance terms are estimated. EL uses the probabilities $\{\hat{\pi}_i\}$ to construct efficient estimators for both; the other members of the GEL class use their associated probabilities to construct an efficient estimator for the Jacobian but use an inefficient estimator for the the variance term; and GMM uses an inefficient estimator for both.

3 The pseudo-panel data approach to estimation based on repeated cross-section data

In this section, we expand on the pseudo-panel approach to estimation of linear regression models based on repeated cross-section data. For ease of exposition, we present this discussion in the context of a specific example involving the relationship between an individual's level of education and subsequent earnings. More specifically, the estimation of the "returns to education" that is, the impact of an additional year of education on wages.

To this end, suppose we have a repeated cross-section data set containing the values of the log of hourly wages, $\mathbf{y}$, and the number of years of education, $\mathbf{ed}$, for cross-sections of individuals sampled from a population in each of T consecutive years. We thus index observations by the pair $(j(t), t)$ where t denotes the year and $j(t)$ denotes the j^{th} individual sampled in year t.

[11] See Andrews, Elamin, Hall, Kyriakoulis, and Sutton (2014) for further discussion of this issue in the context of the model in Example 1 and an empirical illustration of where these differences are important for estimation of a policy parameter in health economics.

Suppose further that wages and education are related by the following model

$$\mathbf{y}_{j(t),t} \;=\; \alpha \,+\, \beta \mathbf{ed}_{j(t),t} \,+\, \mathbf{u}_{j(t),t}, \tag{3}$$

where α, β are unknown parameters, and $\mathbf{u}_{j(t),t}$ is an unobservable error. Within this example, the key parameter of interest is β: 100β equals the implied percentage response in wages to one more year of education. As in our panel data example above, the error is assumed to have a composite form,

$$\mathbf{u}_{j(t),t} \;=\; \mathbf{a}_{j(t)} \,+\, \mathbf{w}_{j(t),t}. \tag{4}$$

The component $\mathbf{a}_{j(t)}$ is an individual-specific effect which captures unobserved characteristics about individual $j(t)$ that may affect the wage earned, and $\mathbf{w}_{j(t),t}$ is the idiosyncratic error. The unobserved characteristic, $\mathbf{a}_{j(t)}$, captures such factors as innate ability of the individual and government education policy at the time the individual was at school. Both are correlated with education as well, and so $\mathbf{a}_{j(t)}$ is a fixed effect. We assume that for any individual j in the population the fixed effect is generated via

$$\mathbf{a}_j \;=\; \alpha_{c(j)} \,+\, \mathbf{a}_j^* \tag{5}$$

where $\alpha_{c(j)}$ is an unknown constant that depends on $c(j)$, the birth cohort of individual j, and $\mathbf{a}_j^*$ is a mean-zero random variable that accounts for variation in the fixed effect across individuals from the same birth cohort.

This specification can be justified in our example as follows. There is no reason to suppose that the distribution of innate ability in the population has changed over time and so the effect of this component on wages is captured by the constant α in (3), and the remaining variation contributes to $\mathbf{a}_j^*$. However, government education policy has changed over time and the systematic component of this change is captured by $\alpha_{c(j)}$ with variation about this level also contributing to $\mathbf{a}_j^*$. Note this effect is indexed by the cohort of birth because this indicates the calendar years when the individual attended primary and secondary education.

Notice the model cannot be estimated using the moment condition in Example 2 because we cannot construct the first time differences of the variables for an individual as we do not observe the same individuals in each year. Instead, Deaton (1985) proposes creating a "pseudo-panel" data set from the original repeated cross-section data by constructing (birth) cohort-time averages leading to the estimation of the regression model

$$\bar{y}_{c,t} \;=\; \text{"cohort specific intercept"} \;+\; \beta\, \bar{ed}_{c,t} \;+\; \text{"error"}, \tag{6}$$

where $\bar{(\cdot)}_{c,t}$ represents the sample mean value of $(\cdot)$ over all $j(t)$ for which $c(j(t)) = c$. This is an example of a grouped-data estimation in which the number of groups is $G = CT$, where C is the total number of birth cohorts and T is the total number of time periods.[12] As noted by Angrist (1991), Durbin (1954) shows that OLS regression with group means is equivalent to Two Stage Least Squares (2SLS) estimation using individual level data with regressors on the first stage being a complete set of dummy variables for group membership. Since 2SLS is a GMM estimator,[13] OLS estimation based on group-mean data can be viewed as an estimation method based on population moment conditions.[14]

To present the moment conditions in question, we introduce a group index

$$g \;=\; (T-1)c \,+\, t, \qquad c = 1, 2, \ldots C;\; t = 1, 2, \ldots T.$$

[12]For ease of exposition, we assume each time period contains observations from each cohort.

[13]For example, see Hall (2005) Chapter 2.

[14]We note that Deaton (1985) does not propose inference based on the OLS estimator but instead a modified version referred to as the "Errors in Variables"(EVE) estimator. This is because Deaton (1985) considers asymptotics in which the number of groups gets large but the number of observations in each group is finite. Within that framework, Deaton (1985) shows the OLS estimator is inconsistent. However, if the number of groups is fixed and the number of observations in each group becomes large - the framework we adopt below - then OLS is consistent, see Angrist (1991). See Devereux (2007) for analysis of the second order properties of the EVE.

Now define the group level model

$$\mathbf{y}_g \;=\; \mathbf{x}_g'\theta_0 \,+\, \mathbf{u}_g \tag{7}$$

where $\mathbf{y}_g$ is a random variable modeling the log wage for members of group g, $\mathbf{x}_{g,i} = \left[\mathcal{I}_g^{(1)}, \mathcal{I}_g^{(2)}, \ldots, \mathcal{I}_g^{(C)}, \mathbf{ed}_g\right]'$, $\mathbf{ed}_g$ denotes the number of years for education for a member of group g, $\mathcal{I}_{g,i}^{(c)}$ is an indicator variable that takes the value 1 if group g involves individuals born in cohort c, θ_0 is the parameter vector whose first $p-1$ elements are the cohort specific intercepts and whose last element is β (here $p = C+1$). Following the reasoning in the previous paragraph, OLS estimation based on group-mean data is equivalent to estimation of the group-level model in (7) based on the information that

$$E[\mathbf{u}_g(\theta_0)] \;=\; 0, \qquad g = 1, 2, \ldots G, \tag{8}$$

where $\mathbf{u}_g(\theta) = \mathbf{y}_g - \mathbf{x}_g'\theta$. In this case, identification requires not only that $p = dim(\theta) < G$ but also that the data follow a different distribution in each group. To demonstrate why this is the case, it is sufficient to consider the case of $C = T = 2$ so that $G = 4$. Expressing (8) as a single 4×1 population moment condition with g^{th} element $E[\mathbf{u}_g(\theta_0)] = 0$, the Jacobian is

$$\begin{bmatrix} 1 & 0 & E[\mathbf{ed}_1] \\ 1 & 0 & E[\mathbf{ed}_2] \\ 0 & 1 & E[\mathbf{ed}_3] \\ 0 & 1 & E[\mathbf{ed}_4] \end{bmatrix}$$

and this matrix must have rank equal to three for θ_0 to be identified.[15] For this rank condition to hold, it must be that $E[\mathbf{ed}_1] \neq E[\mathbf{ed}_2]$ or $E[\mathbf{ed}_3] \neq E[\mathbf{ed}_4]$.

As can be seen from the preceding discussion, the pseudo-panel approach to estimation with repeated cross-section data involves dividing the population into groups, and basing estimation on group-specific moment conditions. Thus to develop IM estimators for this

[15] As the model is linear in θ_0, identification and first-order identification are equivalent.

kind of estimation scenario, we need to extend the IM framework of Section 2 to cover the case in which the the population is heterogeneous and consists of homogeneous sub-populations. This is the topic of the next section.

4 IM estimation with group-specific moment conditions

In this section we extend the IM framework to the case in which the population is heterogeneous consisting of a finite number of homogeneous sub-populations. In view of the discussion in the previous section, we refer to these sub-populations as "groups".

We first describe the group-data structure. It is assumed that there are G groups and for each group g, the model involves random variables $\mathbf{v}_g$ with probability measure μ_g and sample space $\mathcal{V}_g$. We impose the following condition on the groups.

Assumption 1 *(i)* $\mathbf{v}_g$ *is independent of* $\mathbf{v}_h$ *for all* $g, h = 1, 2, \ldots, G$ *and* $g \neq h$*; (ii) there is a random sample of size* n_g *on* $\mathbf{v}_g$ *for each* g*.*

Note that Assumption 1 states that observations are independent both across and within groups. This rules out many forms of clustering, as for example in worker-firm data. It also rules out serial correlation if g is defined on time.

We further assume that the model implies that each group satisfies a set of population moment conditions involving $f_g : \mathcal{V}_g \times \mathbb{R}^{q_g}$.

Assumption 2 $E[f_g(\mathbf{v}_g, \theta_0)] = 0$ *where* $\theta_0 \in \Theta \subset \mathbb{R}^p$*, for* $g = 1, 2, \ldots, G$*.*

Note that the moment conditions are allowed to vary by g. This may happen because the functional form of the moment conditions is the same across g but they are evaluated at group specific parameters that is,

$$f_g(\mathbf{v}_g, \theta_0) \;=\; f(\mathbf{v}_g, \gamma_g(\psi), \beta)$$

where $\theta = (\psi', \beta')'$. Our example in Section 3 fits this structure with $\psi = (\psi_1, \psi_2, \ldots, \psi_C)$ and ψ_c denoting the intercept for cohort c. However, the key element is that certain parameters appear in the population moment conditions associated with more than one group.

To present the IM estimator, we need the following additional notation. Let $\mathbf{v} = vec(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_g)$ and $\mu = \mu_1 \times \mu_2 \times \ldots \mu_G$; note that Assumption 1 implies μ is the probability measure of $\mathbf{v}$. Now define the following sets of measures: $\mathbf{M}$, the set of all possible probability measures for $\mathbf{v}$;

$$\mathbf{P}(\theta) \;=\; \left\{ P = P_1 \times P_2 \times \ldots \times P_G \in \mathbf{M} : \int f_g(v_g, \theta) dP_g \;=\; 0,\; g = 1, 2, \ldots G \right\},$$

the set of all measures for $\mathbf{v}$ which satisfy the population moment conditions in each group for a given value of θ;

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

the set of all measures for $\mathbf{v}$ which satisfy the population moment condition in each group for some $\theta \in \Theta$. As for the homogenous population case in Section 2, 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 and that this approach is operationalized using discrete distributions. To this end, suppose we have a random sample on $\mathbf{v}_g$ consisting of observations $\{v_{g,i};\, i = 1, 2, \ldots n_g\}$. The total sample size is then $N = \sum_{g=1}^{G} n_g$. Define $\pi_{g,i} = P(\mathbf{v}_g = v_{g,i})$, and $\hat{P}_g = [\pi_{g,1}, \pi_{g,2}, \ldots, \pi_{g,n_g}]$, $\hat{P} = [\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_G]$. Assuming no two sample outcomes are the same, the empirical distribution of the data is: $\hat{\mu}_{g,i} = n_g^{-1}$; let $\hat{\mu}_g = n_g^{-1} \iota'_{n_g}$ where ι_{n_g} is a $n_g \times 1$ vector of ones, $\hat{\mu} = [\hat{\mu}_1, \ldots \hat{\mu}_G]$.

The Info-metric estimator based on group-specific moment conditions (IM-GMC) is then defined to be:

$$\hat{\theta}_{IM} \;=\; \arg \inf_{\theta \in \Theta} \; \mathcal{D}_N(\theta, \hat{\mu}) \tag{9}$$

where

$$\begin{aligned}
\mathcal{D}_N(\theta, \hat{\mu}) &= \inf_{\hat{P}(\theta) \in \hat{\mathbf{P}}(\theta)} D(\hat{P}(\theta) \,\|\, \hat{\mu}), \\
\hat{\mathbf{P}}(\theta) &= [\hat{\mathbf{P}}_{\mathbf{1}}(\theta), \hat{\mathbf{P}}_{\mathbf{2}}(\theta), \ldots, \hat{\mathbf{P}}_{\mathbf{G}}(\theta)], \\
\hat{\mathbf{P}}_{\mathbf{g}}(\theta) &= \left\{ \hat{P}_g : \pi_{g,i} > 0, \sum_{i=1}^{n_g} \pi_{g,i} = 1, \sum_{i=1}^{n_g} \pi_{g,i} f(v_{g,i}, \theta) \;=\; 0 \right\},
\end{aligned}$$

and the distance measure is:

$$D(\hat{P}(\theta) \,\|\, \hat{\mu}) \;=\; N^{-1} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \phi(n_g \hat{\pi}_{g,i}).$$

The natural choices for $\phi(\,\cdot\,)$ are the same as for the IM estimator and the specific choices listed in Section 2 would give grouped-data versions of EL, ET and CUE.

AHKL define the GEL-GMC estimator of θ_0 as,

$$\hat{\theta}_{GEL} = \underset{\boldsymbol{\theta} \in \Theta}{argmin} \; \underset{\lambda \in \Lambda_N}{sup} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \left[\rho \left(\lambda_g' f_{g,i}(\theta) \right) \;-\; \rho(0) \right] / N, \tag{10}$$

where $f_{g,i}(\theta) = f_g(v_{g,i}, \theta)$, $\rho(.)$ is a concave function on its domain $\mathcal{A}$, an open interval containing 0, $\rho_0 = \rho(0)$ and $\lambda = (\lambda_1, \lambda_2, \ldots \lambda_G)'$ is vector of auxiliary parameters vector restricted so that with probability approaching 1, $\lambda_g' f_g(v_{g,i}, \theta) \in \mathcal{A}$ for all $(\theta', \lambda')' \in \Theta \times \Lambda_N$ and $g = 1, 2 \ldots G$.[16] The auxiliary parameter λ_g is the Lagrange Multiplier on the constraint that the group g moment condition holds in the IM formulation (9).

We find the info-metric (primal) approach more intuitively appealing because it is formulated explicitly in terms of the population moment conditions placing restrictions on the distribution of the data. However, GEL approach is often more appealing for the purposes of developing the asymptotic analysis of the estimator.

While either method can be used to calculate the estimator, the primal (IM-GMC) and dual (GEL-GMC) approaches have differing computational requirements. In the primal ap-

[16] Setting $\rho(\,\cdot\,)$ equal to $log[1 - (\,\cdot\,)]$, $-exp(\,\cdot\,)$ and $\rho_0 - (\,\cdot\,) - 0.5(\,\cdot\,)^2$ yields grouped-data versions of EL, ET and CUE respectively.

proach, the optimization is over both the probabilities $\{\hat{\pi}_{g,i};\ i = 1, 2, \ldots n_g;\ g = 1, 2, \ldots G\}$ and θ. In the dual approach, the optimization is over λ and θ.[17] While the latter involves fewer parameters, the associated optimizations can be problematic as we now discuss. The computation is performed by iterating between the so-called *inner* and *outer* loops. The inner loop involves optimization over λ for given θ that is,

$$\hat{\lambda}(\theta) = \arg \sup_{\lambda \in \Lambda_N} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \left[\rho \left(\lambda_g' f_{g,i}(\theta) \right) \; - \; \rho(0) \right] / N,$$

and the outer loop involves optimization over θ given λ that is,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \left[\rho \left(\hat{\lambda}_g(\theta)' f_{g,i}(\theta) \right) \; - \; \rho(0) \right] / N.$$

While the inner loop is well suited to gradient methods because $\rho(\cdot)$ is strictly concave, the outer loop can be more problematic.[18] In terms of calculating the estimators in our context using numerical routines in MATLAB, we have found optimization associated with the primal approach far more reliable because, due to the convexity of the primal approach, estimation is robust to the initial parameter starting values provided to the optimizer.[19] In contrast, the solution to the min-max problem in the GEL approach is extremely sensitive to starting values; in cases where the information content of the population moment conditions is low, the optimizing routine would often fail to move away from the initial values.

5 Statistical properties and inference

In this section, we describe the asymptotic properties of the IM/GEL-GMC estimator and its associated inference framework. To simplify the exposition, we consider the case in

[17] The probabilities can be estimated from estimators of λ_g, θ and the data.

[18] For example, see Guggenberger (2008).

[19] Specifically, in the simulations reported in Section 5, the procedure *fmincon* was used to enforce the constraints in equation (9). Convergence time is greatly improved by providing analytical form for the Jacobian and Hessian of the IM objective function.

Section 3 where the group level model is:

$$\mathbf{y}_g \;=\; \mathbf{x}_g'\theta_0 \;+\; \mathbf{u}_g, \tag{11}$$

where $\mathbf{x}_g = [1, \mathbf{r}_g']'$, $\mathbf{r}_g$ is a $k \times 1$ vector of observable random variables, and $\mathbf{y}_g$ and $\mathbf{u}_g$ are scalar random variables. Estimation is based on the information that

$$E[\mathbf{u}_g(\theta_0)] \;=\; 0, \qquad g = 1, 2, \ldots G. \tag{12}$$

This model fits into the framework of Section 4 with $\mathbf{v}_g = (\mathbf{y}_g, \mathbf{r}_g')'$ and $f_g(\mathbf{v}_g, \theta) = \mathbf{y}_g - \mathbf{x}_g'\theta$. It is assumed that $p = dim(\theta) < G$.

In addition to Assumption 1, the data must satisfy certain regularity conditions but for brevity these are suppressed here.[20] Define $\sigma_g^2 = Var[\mathbf{u}_g]$, and $B_g = E[\mathbf{x}_g]$. Samples are assumed to satisfy:

Assumption 3 *n_g is deterministic sequence such $n_g/N \to \nu_g \in (0,1)$ as $N \to \infty$.*

Notice this assumption implies the sample size for each group increases with N and so become asymptotically large. One consequence of this assumption is that the Weak Law of Large Numbers (WLLN) and Central Limit Theorem (CLT) can be used to deduce the behaviour of the group averages. Let $\{\mathbf{x}_{g,i}\}_{i=1}^{n_g}$ and and $\{\mathbf{u}_{g,i}\}_{i=1}^{n_g}$ be random draws from the distributions of $\mathbf{x}_g$ and $\mathbf{u}_g$ respectively, $\bar{(\cdot)}_g = n_g^{-1}\sum_{i=1}^{n_g}(\cdot)_{g,i}$, $\bar{(\cdot)} = vec[\bar{(\cdot)}_1, \bar{(\cdot)}_2, \ldots, \bar{(\cdot)}_G]$ and $\hat{\nu}_N = diag(n_1/N, n_2/N, \ldots, n_G/N)$. Then under Assumptions 1 and 3, and assuming (12) holds, we can invoke the WLLN and the CLT respectively to deduce:

$$(n_g/N)\bar{x}_g \;\xrightarrow{p}\; \nu_g B_g,\; g = 1, 2, \ldots G \tag{13}$$

$$N^{1/2}\hat{\nu}_N \bar{u} \;\xrightarrow{d}\; N(0_G,\, \Psi_u) \tag{14}$$

where 0_G denotes the $G \times 1$ null vector, $\Psi_u = diag(\nu_1\sigma_1^2, \nu_2\sigma_2^2, \ldots, \nu_G\sigma_G^2)$.

[20] See Khatoon (2014) and AHKL.

It can be shown using similar arguments to Newey and Smith (2004) that $\hat{\theta}$ is consistent for θ_0.[21]. Using this consistency property, it is possible to take a Mean Value expansion of the first order conditions of the optimization from which can be deduced the following result.

Proposition 1 *Under Assumptions 1 and 3 and certain other regularity conditions, we have:*

$$\begin{bmatrix} \sqrt{N}(\hat{\theta} - \theta_0) \\ \sqrt{N}\hat{\lambda} \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0_k \\ 0_G \end{bmatrix}, \begin{bmatrix} V_\theta & 0 \\ 0 & V_\lambda \end{bmatrix} \right),$$

where $V_\theta = (B'\Psi_u^{-1}B)^{-1}$, $V_\lambda = \Psi_u^{-1} - \Psi_u^{-1}BV_\theta B'\Psi_u^{-1}$, *and* $B = [\nu_1 B_1, \nu_2 B_2, \ldots, \nu_G B_G]'$.

Proposition 1 implies that $\sqrt{N}(\hat{\theta} - \theta_0)$ and $\sqrt{N}\hat{\lambda}$ converge to normal distributions and are asymptotically independent.

Within this framework, two types of inference are naturally of interest: inference about θ_0 and tests of the validity of the population moment condition upon which the estimations rests. We now discuss these in turn.

An approximate $100a\%$ confidence interval for $\theta_{0,k}$ is given by

$$\left(\hat{\theta}_k \pm z_{1-a/2} se(\hat{\theta}_k) \right) \tag{15}$$

where $\hat{\theta}_k$ is the k^{th} element of $\hat{\theta}$, $z_{1-a/2}$ is the $100(1 - a/2)^{th}$ percentile of the standard normal distribution, $se(\hat{\theta}_k)$ is the k^{th} main diagonal element of $\hat{V}_\theta = (\hat{B}\hat{\Psi}_u^{-1}\hat{B})^{-1}$, $\hat{B} = \hat{\nu}_N [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_G]'$, $\hat{\Psi}_u = \hat{\nu}_N diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots \hat{\sigma}_G^2)$, $\hat{\sigma}_g^2 = \sum_{i=1}^{n_g} \hat{\pi}_{g,i}(y_{g,i} - x'_{g,i}\hat{\theta})^2$, and $\{y_{g,i}, x_{g,i}\}_{i=1}^{n_g}$ are the sample realizations in group g.

As is apparent from the discussion in Section 3, the estimation exploits the information that the moment condition holds in the population. Typically, this moment condition is derived from some underlying economic and/or statistical model and so it is desirable to test whether the data is consistent with this moment condition. Within the info-metric

[21]Formal proofs of all results are omitted for brevity but are presented in Khatoon (2014) for the model in this section and in AHKL for the general model described in Section 4.

framework this can be done in three ways. First, using Proposition 1, we can base inference on the Lagrange Multipliers $\hat{\lambda}$. Inspection reveals V_λ is singular, being of rank $G-p$. One option is to use a generalized inverse of V_λ to construct the test, but following Imbens, Spady, and Johnson (1998) we propose using the asymptotically equivalent and computationally more convenient version of the LM statistic[22]

$$LM \;=\; N\hat{\lambda}'\hat{\Psi}_u\hat{\lambda}. \tag{16}$$

Second, inference can be based directly on the estimated sample moments $\bar{u}(\hat{\theta})$. While the optimization forces the weighted sample moments, $\sum_{i=1}^{n_g}\hat{\pi}_{g,i}u_{g,i}(\hat{\theta})$, to zero, the estimated sample moments $n_g^{-1}\sum_{i=1}^{n_g}u_{g,i}(\hat{\theta})$ are not so constrained but should be approximately zero if the moment condition is valid. This leads to the Wald statistic

$$Wald \;=\; N\{\hat{\nu}\bar{u}(\hat{\theta})\}'\hat{\Psi}_u^{-1}\hat{\nu}\bar{u}(\hat{\theta}). \tag{17}$$

Finally, inference can be based on the optimand. Within the GEL approach, this approach leads to the statistic:

$$LR_{GEL} \;=\; -2\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left[\rho\left(\hat{\lambda}_g u_{g,i}(\hat{\theta})\right) \;-\; \rho(0)\right], \tag{18}$$

in which the unconstrained version does not impose the population moment condition and so amounts to $\lambda_g = 0$ - and hence $\lambda_g u_{g,i} = 0$ - in the GEL-GMC framework.

The Wald and LM are all easily calculated with both the primal and dual approach to estimation, but while the LR_{GEL} is a natural side product of the dual approach, it is not so with the primal approach. For the latter, more convenient test statistics based on the primal optimand are as follows. If the estimation is performed using EL then, following Qin and Lawless (1994), a suitable statistic is based on the difference between the log likelihood evaluated at the constrained probabilities and the unconstrained probabilities. Adapting

[22] Imbens, Spady, and Johnson (1998) report that this version has better finite sample properties in their simulation study than the version based on the generalized inverse of V_λ. See also Imbens (2002).

their approach to our setting yields the test statistic:

$$LR_{EL} = -2\sum_{i=1}^{G}\sum_{i=1}^{n_g} ln(n_g\hat{\pi}_{g,i}). \tag{19}$$

If the estimation is based on ET then, following Imbens, Spady, and Johnson (1998), a suitable statistic can be based on the Kullback-Liebler distance between constrained and unconstrained probabilities. In our setting, this approach yields the test statistic

$$KLIC - R_{ET} = 2\sum_{i=1}^{G}\sum_{i=1}^{n_g} n_g\hat{\pi}_{g,i}\{ln(n_g\hat{\pi}_{g,i})\}. \tag{20}$$

The following proposition gives the limiting distributions of the above tests under the null hypothesis that the moments are valid.[23]

Proposition 2 *If Assumptions 1 and 3 and certain other regularity conditions hold then: (i) LM, $Wald$, and LR_{GEL} are asymptotically equivalent and they all converge in distribution to χ^2_{G-p} as $N \to \infty$ (ii) LR_{EL} and $KLIC - R$ converge in distribution to χ^2_{G-p} as $N \to \infty$.*

The first order asymptotic properties above are the same for all members of the IM/GEL class described in Section 3. However, the second order properties are different. Newey and Smith (2004) derive the second order bias of GEL estimators for the case described in Section 2. Their approach is based on taking a third order expansion of the first order conditions of GEL estimation and can be adapted to derive analogous results for the Im/GEL-GMC estimator to yield the following result.[24]

Proposition 3 *Under Assumptions (1) and (3) and certain other regularity conditions,*

$$Bias(\hat{\theta}_{GEL}) = -\Xi\left[\mathcal{B}_1 + \left(1 + \frac{\rho_3(0)}{2}\right)\mathcal{B}_2\right]/N, \tag{21}$$

[23]Part (i) is proved in Khatoon (2014) and part (ii) can be proved by adapting the arguments in Qin and Lawless (1994) (for EL) and Imbens, Spady, and Johnson (1998) (for ET) to our groped-data context.

[24]See Khatoon (2014).

where $\mathcal{B}_1 = diag\left[\Xi' M_{xu}\right]$*,* $\mathcal{B}_2 = M_u^{(3)} diag(V_\lambda)$*,* $\Xi = V_\theta B' \Psi_u^{-1}$*, (here)* $V_\theta = (B'\Psi_u^{-1}B)^{-1}$*,* M_{xu} *is a* $p \times G$ *matrix* g^{th} *column* $\nu_g E[\mathbf{u}_g \mathbf{x}_g]$*,* $M_u^{(3)} = diag(\nu_1\mu_1^{(3)}, \nu_2\mu_2^{(3)}, \ldots, \nu_G\mu_G^{(3)})$*,* $\mu_g^{(3)} = E[\mathbf{u}_g^3]$*, and* $\rho_3(\star) = \partial^3\rho(a)/\partial a^3|_{a=\star}$*.*

As can be seen, the second bias depends on $\rho(\cdot)$ and so is potentially different for different members of the GEL class of estimators. Specializing the bias formula to the three leading cases, the biases of EL, ET and CUE are given respectively by $-\Xi\mathcal{B}_1/N$, $-\Xi\left(\mathcal{B}_1 + \frac{1}{2}\mathcal{B}_2\right)/N$ and $-\Xi\left(\mathcal{B}_1 + \mathcal{B}_2\right)/N$.[25] So it can be seen that in general, as in Newey and Smith's (2004) analysis, EL has fewer sources of bias than the other two. The formula reveals that in our model the sources of the bias are correlation between u_g and x_g and asymmetry of the distributions of $\{\mathbf{u}_g\}_{i=1}^{n_g}$. Note that within our repeated cross-section model of Section 3, $\mathbf{u}_g$ and $\mathbf{x}_g$ are correlated through the stochastic part of the fixed effect. If $E[\mathbf{u}_g^3] = 0$ for all g then the bias is the same for all three estimators.

It is natural to consider whether the IM/GEL approach has similar advantages over GMM in the case of grouped-specific moment conditions as those described for the homogeneous population case in Section 2. For the model in this section, the GMM estimator is

$$\tilde{\theta} \;=\; argmin_{\theta\in\Theta}\bar{u}(\theta)' W_N \bar{u}(\theta), \tag{22}$$

Under our assumptions, the optimal choice of weighting matrix, W_N, is one that converges in probability to Ψ_u^{-1} and with this choice it is straightforward to show that $N^{1/2}\left(\tilde{\theta} - \theta_0\right)$ has the same limiting distribution as $\sqrt{N}(\hat{\theta} - \theta_0)$ (given in Proposition 1). However, once again the second order properties of the GMM and IM/GEL-GMC estimators differ.

Proposition 4 *If Assumptions 1 and 3 and certain other regularity conditions hold then*

$$Bias(\hat{\boldsymbol{\beta}}_{GMM}) = [V_\theta\mathcal{A} - \Xi(\mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3)]/N,$$

where $\mathcal{A} = M_{xu} diag(V_\lambda)$*,* $\mathcal{B}_3$ *is a term that depends on* M_{xu} *and the difference between the*

[25] $\rho_3(0)$ equals -2, -1 and 0, respectively for EL, ET and CUE.

first and second step weighting matrices.[26]

A comparison of Propositions 3 and 4 indicates that the grouped-data GMM estimator has more sources of bias than the corresponding GEL-GMC estimator. One of these additional sources is attributable to the two-step nature of GMM estimation and arises if a sub-optimal weighting matrix is used on the first step. A similar finding is reported in Newey and Smith (2004), and they argue that these extra sources of bias are likely to translate to an estimator that exhibits more bias in finite samples.

These results can be used to compare GEL-GMC to the pseudo-panel data approach of regression based on group averages.[27] As noted in Section 3, this pseudo-panel approach amounts to estimation of the individual level model via 2SLS using group dummies and is therefore a GMM estimator based on the moments in (12). 2SLS is only equivalent to two-step GMM if the variance of $\mathbf{u}_g$ is the same for all g, and so is only as efficient asymptotically as GEL-GMC under that condition. Even then, the arguments above indicate that finite sample inferences based on the the pseudo-panel approach are likely less reliable than those based on GEL-GMC. In the remainder of this section, we explore whether is the case in our setting via a small simulation study.

Artificial data generated for groups $g = 1, 2, \ldots, G$ via:

$$\mathbf{y}_g = \beta \mathbf{r}_g + \delta + \mathbf{u}_g \tag{23}$$

$$\mathbf{r}_g = 12 + \sum_{j=2}^{G} \mathcal{I}_g^{(j)} \delta_j + \mathbf{a}_g \tag{24}$$

where $\beta = 0.05$, $E[\mathbf{u}_g, \mathbf{a}_g] = 0$, $Var[\mathbf{u}_g] = \sigma_u^2$, $Var[\mathbf{a}_g] = \sigma_a^2$, $Cov[\mathbf{u}_g, \mathbf{a}_g] = \rho\sigma_u\sigma_a$, and $\mathcal{I}_g^{(j)}$ is an indicator variable that takes the value one if $j = g$. For the results reported below, we set the parameter values as follows: $\beta = 0.05$, $\sigma_u^2 = 0.2$, $\sigma_a^2 = 3.38$, $\sigma_{u,a} = \rho\sigma_u\sigma_a$ with $\rho = 0.2, 0.5, 0.9$, and $\delta_2 = 0.9$, $\delta_j = \delta_{j-1} + 0.9$ for $j = 3, \ldots G$. Notice that within this design, $\mathbf{r}_g$ is correlated with $\mathbf{u}_g$. We report results for the cases where $(\mathbf{u}_g, \mathbf{a}_g)'$ has a

[26] See Khatoon (2014) for details of this term, the details of which are omitted as they are not relevant to the exposition.

[27] The discussion in the paragraph does not cover the EVE estimator, see footnote 14.

bivariate normal and a bivariate Student-t distribution with 7 degrees of freedom.

We report results for different numbers of groups and sample sizes. Specifically, we consider the scenarios $G = 3, 4, 6, 8$, $N = 96, 144, 312$ with sample size within each group determined via $n_g = N/G$ for all g. Note that within this scheme, there is an inverse relationship between the number of groups and the number of observations within each group. This enables us to examine whether the accuracy of the asymptotic theory depends on just N *per se* or on the number of groups these observations are spread across. Ten thousand replications are performed for each parameter configuration.

Estimation is based on the moment conditions,

$$E[\mathbf{y}_g - \mathbf{x}_g'\theta_0] = 0, \text{ for } g = 1, 2, \ldots G.$$

where $\mathbf{x}_g = [1, \mathbf{r}_g]'$, $\theta = (\delta, \beta)'$. We consider two versions of IM/GEL-GMC: the first, denoted hereafter as EL, involves $\phi(\cdot) = -log(\cdot)$ and the second, denoted hereafter ET, involves $(\phi(\cdot) = (\cdot)log(\cdot)$. We also consider estimation based on 2SLS and two-step GMM. Specifically we report the following statistics: the mean of the simulated distributions of $\hat{\beta}$; the rejection frequency of the 5% approximate significance level test of $H_0 : \beta_0 = 0.05$ (its true value) based on the t-statistic, $\hat{\beta}/s.e.(\hat{\beta})$;[28]the rejection frequency for LM and $Wald$ (both for EL and ET), LR_{EL}, $KLIC - R_{ET}$ and the overidentifying restrictions test for GMM.[29]

Before discussing the results, we note that there are reasons to expect that the finite sample behaviour of GMM may be more sensitive to G than EL/ET. Specializing the result in Proposition 4 to the model in our simulations then the second order bias of the GMM estimator is:

$$Bias(\hat{\beta}_{GMM}) = \frac{(G-3)\sigma_{u,a}}{NR_{x,z}^2\sigma_x^2}. \tag{25}$$

[28]From Proposition 1 it follows that the t-statistic is distributed approximately as a standard normal random variable in large samples.

[29]The GMM overidentifying restrictions is calculated as N times the minimand on the right-hand side of (22). Like the other model specification tests, the overidentifying restrictions test converges to a χ^2_{G-p} distribution under the null hypothesis that the moments are valid.

where $R^2_{x,z}$ is the population multiple correlation coefficient from the pooled (over g) regression of $\mathbf{v}_g$ on $z = [\mathcal{I}_g^{(1)}, \mathcal{I}_g^{(2)}, \ldots, \mathcal{I}_g^{(G)}]$ and σ_x^2 is the population variance of x. In contrast, second order bias of the IM/GEL-GMC estimator is:

$$Bias(\hat{\beta}_{GEL}) \;=\; -\frac{\sigma_{u,a}}{NR^2_{x,z}\sigma_x^2}. \tag{26}$$

Clearly the second order bias of GMM increases with G (for $G > 3$) but that of IM/GEL-GMC is invariant to G.

Tables 1-3 report the results for the normal distribution for N equal to 96, 144 and 312 respectively, whereas Tables 4-6 report the results for the Student-t distribution. First consider the bias. From (25), it can be seen that the GMM estimator is second order unbiased in this case, and this is approximately true in the simulations. The bias formulae also indicate that the second order bias of IM/GEL-GMC is unaffected by the number of groups but that of the grouped GMM estimator increases with the number of groups. This is again what we find: interestingly, the GMM estimator exhibits more bias when the data come from the Student-t distribution whereas the bias of the IM/GEL-GMC estimators is comparable for both distributions. All three t-statistics show size distortion as G increases and/or ρ increases. The GMM t-tests have empirical size closer to the nominal size for low degrees of endogeneity ($\rho = 0.2$), but the IM/GEL-GMC t-tests have empirical size closer to the nominal size for high degrees of endogeneity ($\rho = 0.9$). For the largest sample size, all the empirical sizes are close, albeit systematically above, the nominal level. Now consider the overidentifying restrictions tests. The G-IM Wald tests exhibit empirical size very close to the nominal level in all settings, the GMM test has rejects slightly more than it should, but the other tests reject far too often in the smaller samples ($N = 96, 144$). As would be expected, the degree of over-rejection is reduced as the sample size increases; however, even in the larger sample size the LR_{EL}, $KLIC - R_{ET}$ tests over reject by 1.5-2.5% in the normal distribution case and by 2.5 and 3.1% in the Student-t distribution case, and the LM tests over reject by between 2.3 and 8.3%.

While our simulation study is limited, the results provide some interesting insights into the comparative properties of the estimators. If the degree of overidentification is small and the degree of endogeneity low then there is little to choose between the estimators. However, if the degree of overidentification is relatively large and/or the degree of endogeneity is high then the EL version of IM/GEL-GMC yields the most reliable inferences.

6 Concluding remarks

In this paper, we have introduced an IM estimator for the parameters of statistical models using information in population moment conditions that hold at group-level. The IM estimation can be viewed as the primary approach to a constrained optimization. The estimators can also be obtained via the dual approach to this optimization, known as Generalized Empirical Likelihood (GEL). In a companion paper (AHKL), we provide a comprehensive framework for inference based GEL with the grouped specific moment conditions. In this chapter, we compare the computational requirements of the primary and dual approaches. We also describe an inference framework based IM/GEL estimators. Using analytical arguments and a small simulation study, it is shown that the IM/GEL approach to estimation based on grouped specific moment conditions yields more reliable inference in finite samples than certain extant methods.

Table 1: Normal distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=96	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0446	0.0475	0.0488	0.0493	0.0378	0.0445	0.0482	0.0489	0.0321	0.0418	0.0465	0.0481
ET beta	0.0451	0.0475	0.0488	0.0493	0.0383	0.0446	0.0483	0.0489	0.0321	0.0419	0.0465	0.0481
2SLS beta	0.0500	0.0513	0.0518	0.0517	0.0495	0.0539	0.0555	0.0546	0.0505	0.0581	0.0591	0.0584
GMM beta	0.0499	0.0513	0.0518	0.0517	0.0495	0.0538	0.0555	0.0546	0.0507	0.0581	0.0590	0.0583
EL rej	0.0451	0.0639	0.0854	0.1002	0.0592	0.0663	0.0910	0.1001	0.0806	0.0679	0.0782	0.0942
ET rej	0.0449	0.0633	0.0839	0.0961	0.0588	0.0656	0.0882	0.0967	0.0795	0.0680	0.0765	0.0920
GMM rej	0.0357	0.0491	0.0656	0.0788	0.0611	0.0662	0.0881	0.0957	0.1095	0.1080	0.1221	0.1302
GMM J-test	0.0432	0.0471	0.0527	0.0493	0.0564	0.0572	0.0568	0.0554	0.0743	0.0803	0.0662	0.0610
EL WALD	0.0380	0.0405	0.0479	0.0498	0.0439	0.0466	0.0485	0.0529	0.0478	0.0521	0.0465	0.0457
ET WALD	0.0378	0.0400	0.0479	0.0495	0.0436	0.0463	0.0481	0.0521	0.0477	0.0519	0.0463	0.0440
EL LM	0.0410	0.0593	0.1213	0.2212	0.0485	0.0628	0.1217	0.2207	0.0533	0.0713	0.1209	0.2196
ET LM	0.0498	0.0774	0.1482	0.2555	0.0581	0.0781	0.1516	0.2550	0.0620	0.0893	0.1516	0.2525
EL LR	0.0421	0.0522	0.0823	0.1235	0.0485	0.0571	0.0821	0.1266	0.0524	0.0649	0.0822	0.1215
ET LR	0.0439	0.0574	0.0914	0.1352	0.0509	0.0603	0.0924	0.1379	0.0547	0.0694	0.0926	0.1330

[a] EL beta & ET beta denote the simulated means of the G-IM estimator using respectively $\phi(\cdot) = -log(\cdot)$ & $\phi(\cdot) = (\cdot)log(\cdot)$; GMM (2SLS) beta are the corresponding figures for the two-step GMM (2SLS) estimator; the true value value is $\beta_0 = 0.05$.

[b] EL rej & ET rej are the empirical rejection rates of the tests of $H_0 : \beta_0 = 0.05$ based on the G-IM estimators using $\phi(\cdot) = -log(\cdot)$ & $\phi(\cdot) = (\cdot)log(\cdot)$ respectively; GMM rej is the corresponding figure based on the two-step GMM estimator; nominal 5% rejections rates

[c] GMM J-test denotes the empirical rejection rate for the GMM overidentifying restrictions test, see footnote 29; EL (ET) Wald & EL (ET) LM denote the corresponding figure for the Wald test in (17) & (16) respectively with $\phi(\cdot) = -log(\cdot)$ ($\phi(\cdot) = (\cdot)log(\cdot)$); EL LR denotes the corresponding figure based on LR_{EL} in (19); ET LR denotes the corresponding figure based on $KLIC - R_{ET}$ in (20); nominal rejection rate is 5%.

Table 2: Normal distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=144	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0479	0.0483	0.0498	0.0500	0.0431	0.0474	0.0489	0.0494	0.0398	0.0447	0.0480	0.0485
ET beta	0.0479	0.0483	0.0498	0.0500	0.0431	0.0474	0.0489	0.0494	0.0398	0.0447	0.0480	0.0485
2SLS beta	0.0505	0.0506	0.0517	0.0514	0.0493	0.0531	0.0538	0.0533	0.0505	0.0549	0.0565	0.0555
GMM beta	0.0505	0.0507	0.0517	0.0515	0.0494	0.0531	0.0538	0.0533	0.0504	0.0550	0.0566	0.0555
EL rej	0.0487	0.0562	0.0736	0.0786	0.0510	0.0611	0.0716	0.0806	0.0645	0.0643	0.0675	0.0800
ET rej	0.0487	0.0554	0.0728	0.0777	0.0506	0.0615	0.0711	0.0793	0.0646	0.0641	0.0669	0.0793
GMM rej	0.0404	0.0462	0.0617	0.0668	0.0523	0.0641	0.0761	0.0795	0.0852	0.0904	0.0951	0.1059
GMM J-test	0.0537	0.0547	0.0492	0.0488	0.0546	0.0517	0.0563	0.0492	0.0667	0.0683	0.0631	0.0630
EL WALD	0.0488	0.0516	0.0452	0.0464	0.0481	0.0439	0.0508	0.0459	0.0505	0.0489	0.0466	0.0497
ET WALD	0.0485	0.0514	0.0449	0.0463	0.0480	0.0435	0.0505	0.0457	0.0505	0.0488	0.0463	0.0495
EL LM	0.0500	0.0574	0.0835	0.1362	0.0492	0.0530	0.0884	0.1288	0.0515	0.0574	0.0846	0.1401
ET LM	0.0564	0.0732	0.1117	0.1745	0.0555	0.0677	0.1188	0.1673	0.0596	0.0733	0.1145	0.1786
EL LR	0.0508	0.0568	0.0672	0.0884	0.0499	0.0505	0.0722	0.0861	0.0528	0.0559	0.0695	0.0944
ET LR	0.0529	0.0610	0.0750	0.0997	0.0519	0.0537	0.0810	0.0952	0.0542	0.0599	0.0771	0.1068

[a] See notes to Table 1

Table 3: Normal distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=312	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0490	0.0496	0.0498	0.0499	0.0469	0.0490	0.0493	0.0499	0.0452	0.0472	0.0490	0.0493
ET beta	0.0490	0.0496	0.0498	0.0499	0.0469	0.0490	0.0493	0.0499	0.0452	0.0472	0.0490	0.0493
2SLS beta	0.0501	0.0507	0.0508	0.0507	0.0495	0.0518	0.0516	0.0518	0.0501	0.0521	0.0531	0.0527
GMM beta	0.0501	0.0507	0.0508	0.0507	0.0495	0.0518	0.0517	0.0519	0.0501	0.0521	0.0531	0.0527
EL rej	0.0525	0.0559	0.0614	0.0624	0.0499	0.0581	0.0597	0.0622	0.0539	0.0552	0.0595	0.0616
ET rej	0.0524	0.0564	0.0614	0.0619	0.0497	0.0581	0.0595	0.0608	0.0540	0.0555	0.0586	0.0598
GMM rej	0.0484	0.0514	0.0572	0.0578	0.0510	0.0589	0.0611	0.0644	0.0658	0.0703	0.0727	0.0732
GMM J-test	0.0502	0.0491	0.0500	0.0511	0.0509	0.0523	0.0488	0.0483	0.0599	0.0555	0.0559	0.0550
EL WALD	0.0483	0.0461	0.0480	0.0482	0.0477	0.0478	0.0451	0.0443	0.0511	0.0458	0.0475	0.0476
ET WALD	0.0483	0.0461	0.0480	0.0478	0.0477	0.0477	0.0449	0.0444	0.0511	0.0459	0.0473	0.0474
EL LM	0.0484	0.0478	0.0570	0.0723	0.0479	0.0508	0.0557	0.0698	0.0520	0.0475	0.0581	0.0730
ET LM	0.0529	0.0565	0.0753	0.1040	0.0525	0.0596	0.0759	0.1019	0.0560	0.0561	0.0767	0.1048
EL LR	0.0494	0.0488	0.0562	0.0638	0.0489	0.0519	0.0546	0.0628	0.0523	0.0491	0.0551	0.0651
ET LR	0.0510	0.0516	0.0612	0.0738	0.0504	0.0541	0.0590	0.0708	0.0533	0.0511	0.0617	0.0734

[a] See notes to Table 1

Table 4: t_7-distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=96	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0421	0.0481	0.0487	0.0492	0.0361	0.0424	0.0475	0.0484	0.0246	0.0367	0.0440	0.0465
ET beta	0.0421	0.0481	0.0487	0.0492	0.0363	0.0427	0.0476	0.0485	0.0249	0.0369	0.0442	0.0466
2SLS beta	0.0481	0.0537	0.0534	0.0528	0.0512	0.0575	0.0598	0.0584	0.0522	0.0628	0.0651	0.0637
GMM beta	0.0481	0.0538	0.0533	0.0528	0.0511	0.0573	0.0592	0.0579	0.0519	0.0625	0.0651	0.0634
EL rej	0.0536	0.0669	0.0829	0.0963	0.0568	0.0660	0.0842	0.1013	0.0619	0.0640	0.0783	0.0936
ET rej	0.0522	0.0642	0.0787	0.0914	0.0554	0.0630	0.0781	0.0946	0.0612	0.0625	0.0745	0.0880
GMM rej	0.0462	0.0576	0.0703	0.0770	0.0567	0.0611	0.0779	0.0850	0.0743	0.0774	0.0882	0.1003
GMM J-test	0.0553	0.0551	0.0529	0.0474	0.0583	0.0610	0.0562	0.0527	0.0712	0.0650	0.0605	0.0532
EL WALD	0.0481	0.0464	0.0448	0.0441	0.0493	0.0504	0.0496	0.0469	0.0553	0.0483	0.0482	0.0467
ET WALD	0.0477	0.0456	0.0440	0.0432	0.0486	0.0497	0.0482	0.0454	0.0545	0.0479	0.0477	0.0447
EL LM	0.0604	0.0840	0.1580	0.2817	0.0632	0.0876	0.1689	0.2876	0.0702	0.0886	0.1707	0.2910
ET LM	0.0692	0.0981	0.1777	0.2918	0.0706	0.1028	0.1863	0.2999	0.0783	0.1007	0.1912	0.3005
EL LR	0.0563	0.0658	0.0948	0.1445	0.0560	0.0695	0.0993	0.1460	0.0624	0.0670	0.1033	0.1466
ET LR	0.0579	0.0697	0.1006	0.1459	0.0577	0.0728	0.1040	0.1486	0.0638	0.0692	0.1082	0.1486

[a] See notes to Table 1

Table 5: t_7-distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=144	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0478	0.0477	0.0499	0.0501	0.0390	0.0466	0.0484	0.0479	0.0342	0.0416	0.0471	0.0474
ET beta	0.0478	0.0478	0.0499	0.0501	0.0391	0.0466	0.0485	0.0479	0.0343	0.0417	0.0472	0.0475
2SLS beta	0.0513	0.0517	0.0529	0.0529	0.0483	0.0556	0.0564	0.0544	0.0515	0.0584	0.0613	0.0593
GMM beta	0.0514	0.0516	0.0530	0.0527	0.0484	0.0558	0.0564	0.0542	0.0514	0.0585	0.0612	0.0589
EL rej	0.0531	0.0606	0.0714	0.0856	0.0531	0.0575	0.0721	0.0873	0.0596	0.0577	0.0664	0.0832
ET rej	0.0520	0.0588	0.0672	0.0839	0.0525	0.0562	0.0680	0.0831	0.0595	0.0561	0.0641	0.0793
GMM rej	0.0477	0.0534	0.0595	0.0729	0.0537	0.0570	0.0658	0.0786	0.0696	0.0663	0.0744	0.0829
GMM J-test	0.0495	0.0574	0.0498	0.0467	0.0529	0.0538	0.0528	0.0511	0.0633	0.0606	0.0602	0.0548
EL WALD	0.0462	0.0504	0.0463	0.0433	0.0480	0.0478	0.0464	0.0457	0.0525	0.0505	0.0481	0.0457
ET WALD	0.0460	0.0495	0.0453	0.0426	0.0473	0.0472	0.0464	0.0458	0.0525	0.0501	0.0475	0.0443
EL LM	0.0512	0.0714	0.1120	0.1859	0.0560	0.0657	0.1169	0.1932	0.0594	0.0715	0.1198	0.1884
ET LM	0.0582	0.0846	0.1340	0.2113	0.0609	0.0781	0.1395	0.2174	0.0650	0.0826	0.1402	0.2142
EL LR	0.0495	0.0631	0.0752	0.1040	0.0517	0.0570	0.0773	0.1094	0.0569	0.0618	0.0830	0.1007
ET LR	0.0505	0.0670	0.0794	0.1105	0.0526	0.0598	0.0836	0.1167	0.0581	0.0639	0.0878	0.1077

[a] See notes to Table 1

Table 6: t_7-distribution: results for beta, t test based on conventional standard error, model specification test rejection rates

N=312	ρ_{ua} =0.2				ρ_{ua} =0.5				ρ_{ua} =0.9			
	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8	G=3	G=4	G=6	G=8
EL beta	0.0477	0.0506	0.0489	0.0504	0.0454	0.0495	0.0488	0.0495	0.0424	0.0452	0.0476	0.0487
ET beta	0.0477	0.0506	0.0489	0.0504	0.0454	0.0495	0.0487	0.0495	0.0425	0.0452	0.0477	0.0488
2SLS beta	0.0494	0.0524	0.0505	0.0516	0.0500	0.0540	0.0526	0.0528	0.0502	0.0535	0.0549	0.0547
GMM beta	0.0495	0.0525	0.0505	0.0516	0.0499	0.0541	0.0527	0.0527	0.0502	0.0536	0.0548	0.0546
EL rej	0.0503	0.0538	0.0591	0.0668	0.0503	0.0580	0.0630	0.0666	0.0543	0.0529	0.0580	0.0659
ET rej	0.0497	0.0530	0.0569	0.0630	0.0496	0.0573	0.0606	0.0635	0.0542	0.0521	0.0555	0.0643
GMM rej	0.0486	0.0500	0.0532	0.0596	0.0496	0.0569	0.0596	0.0622	0.0582	0.0582	0.0641	0.0682
GMM J-test	0.0500	0.0538	0.0537	0.0495	0.0500	0.0533	0.0551	0.0494	0.0495	0.0557	0.0527	0.0555
EL WALD	0.0474	0.0513	0.0494	0.0468	0.0476	0.0495	0.0515	0.0452	0.0459	0.0495	0.0465	0.0499
ET WALD	0.0472	0.0509	0.0490	0.0467	0.0473	0.0493	0.0512	0.0448	0.0459	0.0493	0.0458	0.0493
EL LM	0.0496	0.0571	0.0774	0.1006	0.0517	0.0557	0.0767	0.1027	0.0472	0.0571	0.0759	0.1063
ET LM	0.0556	0.0672	0.0981	0.1274	0.0543	0.0657	0.0950	0.1284	0.0518	0.0659	0.0931	0.1337
EL LR	0.0499	0.0554	0.0652	0.0709	0.0505	0.0543	0.0643	0.0735	0.0477	0.0551	0.0622	0.0749
ET LR	0.0509	0.0580	0.0688	0.0767	0.0506	0.0560	0.0675	0.0782	0.0478	0.0571	0.0649	0.0818

[a] See notes to Table 1

References

Andrews, M., Hall, A. R., Khatoon, R., and Lincoln, J. (2015). 'Generalized Empirical Estimation based on group-specific moment conditions', Discussion paper, University of Manchester, work in progress.

Andrews, M. J., Elamin, O., Hall, A. R., Kyriakoulis, K., and Sutton, M. (2014). 'Inference in the presence of redundant moment conditions and the impact of government health expenditure on health outcomes in England', *accepted for publication in Econometric Reviews*.

Angrist, J. (1991). 'Grouped-data estimation and testing in simple labor-suppy models', *Journal of Econometrics*, 47: 243–266.

Bekker, P. A., and van der Ploeg, J. (2005). 'Instrumental variable estimation based on grouped data', *Statistica Neerlandica*, 59: 239–267.

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

Collado, M. D. (1997). 'Estimating dynamic models from time series of independent cross-sections', *Journal of Econometrics*, 82: 37–62.

Corcoran, S. (1998). 'Bartlett adjustment of empirical discrepancy statistics', *Biometrika*, 85: 965–972.

Deaton, A. (1985). 'Panel data from a time series of cross-sections', *Journal of Econometrics*, 30: 109–126.

Devereux, P. J. (2007). 'Improved errors-in-variables estimators fro grouped data', *Journal of Business and Economic Statistics*, 25: 278–287.

Durbin, J. (1954). 'Errors in variables', *Review of the International Statistical Institute*, 22: 23–32.

Golan, A. (2008). 'Information and entropy econometrics - a review and synthesis', *Foundations and Trends in Econometrics*, 2: 1–145.

Guggenberger, P. (2008). 'Finite sample evidence suggesting a heavy tail problem of the Generalized Empirical Likelihood estimator', *Econometric Reviews*, 26: 526–541.

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

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

Hansen, L. P., Heaton, J., and Yaron, A. (1996). 'Finite sample properties of some alternative GMM estimators obtained from financial market data', *Journal of Business and Economic Statistics*, 14: 262–280.

Imbens, G. (2002). 'Generalized method of moments and empirical likelihood', *Journal of Business and Economic Statistics*, 20: 493–506.

Imbens, G., Spady, R. H., and Johnson, P. (1998). 'Information theoretic approaches to inference in moment condition models', *Econometrica*, 66: 333–357.

Inoue, A. (2008). 'Efficient estimation and inference in linear pseudo-panel data models', *Journal of Econometrics*, 142: 449–466.

Khatoon, R. (2014). 'Estimation and Inference of Microeconometric Models based on Moment Condition Models', Ph.D. thesis, University of Manchester, Manchester, UK.

Kitamura, Y. (2007). 'Empirical likelihood methods in econometrics: theory and practice', in R. Blundell, W. K. Newey, and T. Personn (eds.), *Advances in Economics and Econometrics: Ninth World Congress of the Econometric Society*, pp. 174–237. Cambridge University Press, Cambridge, UK.

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

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

Owen, A. B. (1988). 'Empirical likelihood ratio confidence intervals for a single functional', *Biometrika*, 75: 237–249.

Parente, P. M. D. C., and Smith, R. J. (2014). 'Recent developments in empirical likelihood and related methods', *Annual Review of Economics*, 6: 77–102.

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.