GMM for Panel Count Data Models

Frank Windmeijer

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN

GMM for Panel Count Data Models[*]

Frank Windmeijer

cemmap

and

Department of Economics

University of Bristol

Abstract

This chapter gives an account of the recent literature on estimating models for panel count data. Specifically, the treatment of unobserved individual heterogeneity that is correlated with the explanatory variables and the presence of explanatory variables that are not strictly exogenous are central. Moment conditions are discussed for these type of problems that enable estimation of the parameters by GMM. As standard Wald tests based on efficient two-step GMM estimation results are known to have poor finite sample behaviour, alternative test procedures that have recently been proposed in the literature are evaluated by means of a Monte Carlo study.

JEL Classification: C12, C13, C23
Keywords: GMM, Exponential Models, Hypothesis Testing

---

1   Introduction

This chapter gives an account of the recent literature on estimating models for panel count data. Specifically, the treatment of unobserved individual heterogeneity that is correlated with the explanatory variables and the presence of explanatory variables that are not strictly exogenous are central. Moment conditions are discussed for these type of problems that enable estimation of the parameters by the Generalised Method of Moments (GMM). Interest in exponential regression models has increased substantially in recent years. The Poisson regression model for modelling an integer count dependent variable is an obvious example where the conditional mean function is routinely modelled to be exponential. But also models for continuous positive dependent variables that have a skewed distribution are increasingly being advocated to have an exponential conditional mean function. Although for these data the log transformation can be applied to multiplicative models, the "retransformation" problem often poses severe difficulties if the object of interest is the level of for example costs, see e.g. Manning, Basu and Mullahy (2005). Santos Silva and Tenreyro (2006) also strongly recommend to estimate the multiplicative models directly, as the log transformation can be unduly restrictive. Although the focus of this chapter is on models for count data, almost all procedures can directly be applied to models where the dependent variable is a positive continuous variable and the conditional mean function is exponential. The one exception is the linear feedback model as described in Section 4, which is a dynamic model specification specific to discrete count data.

Section 2 discusses instrumental variables estimation for count data models in cross sections. Section 3 derives moment conditions for the estimation of models for count panel data allowing for correlated fixed effects and weakly exogenous regressors. Section 4 discusses GMM estimation. Section 5 reviews some of the applied literature and software to estimate the models by nonlinear GMM. As standard Wald tests based on efficient two-step GMM estimation results are known to have poor finite sample behaviour, Section 6 considers alternative test procedures that have recently been proposed in the literature. It

also considers estimation by the continuous updating estimator (CUE) as this estimator has been shown to have a smaller finite sample bias than one- and two-step GMM. As asymptotic standard errors for the CUE are downward biased in finite samples we use results from alternative, many weak instrument asymptotics (Newey and Windmeijer (2005)) that lead to a larger asymptotic variance of the CUE. The various estimation and test procedures are evaluated by means of a small Monte Carlo study.

2 GMM in cross-sections

The Poisson distribution for an integer count variable y_i, $i = 1, ..., N$, with mean μ_i is given by

$$P(y_i) = \frac{e^{-\mu_i} \mu^{y_i}}{y_i!}$$

and the Poisson regression model specifies $\mu_i = \exp(x_i'\beta)$, where x_i is a vector of explanatory variables and β a parameter vector to be estimated. The log-likelihood function for the sample is then given by

$$\ln L = \sum_{i=1}^{N} y_i \ln(\mu_i) - \mu_i - \ln(y_i!)$$

with first-order condition

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} x_i (y_i - \mu_i) = 0. \tag{1}$$

It is therefore clear that the Poisson regression estimator is a method of moments estimator. If we write the model with an additive error term u_i as

$$y_i = \exp(x_i'\beta) + u_i = \mu_i + u_i$$

with

$$E(x_i u_i) = E(x_i (y_i - \mu_i)) = 0,$$

this is clearly the same as the first order condition in the Poisson regression model.

An alternative moment estimator is obtained by specifying the error term as multiplicative in the model

$$y_i = \exp(x_i'\beta) \nu_i = \mu_i w_i$$

with associated moment conditions

$$E\left((w_i - 1)|x_i\right) = E\left(\left(\frac{y_i - \mu_i}{\mu_i}\right)|x_i\right) = 0. \tag{2}$$

Mullahy (1997) was the first to introduce GMM instrumental variables estimation of count data models with endogenous explanatory variables. He used the multiplicative setup with x_i being correlated with the unobservables w_i such that $E\left((w_i - 1)|x_i\right) \neq 0$, and the moment estimator that solves (2) is no longer consistent. There are instruments z_i available that are correlated with the endogenous regressors, but not with w_i such that

$$E\left((w_i - 1)|z_i\right) = E\left(\left(\frac{y_i - \mu_i}{\mu_i}\right)|z_i\right) = 0. \tag{3}$$

Denote[1]

$$g_i = z_i\left(\frac{y_i - \mu_i}{\mu_i}\right),$$

then the GMM estimator for β that minimises

$$Q_N(\beta) = \left(\frac{1}{N}\sum_{i=1}^{N} g_i\right)' W_N^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} g_i\right)$$

is consistent, where W_N is a weight matrix. The efficient two-step weight matrix is given by

$$W_N\left(\widehat{\beta}_1\right) = \frac{1}{N}\sum_{i=1}^{N} g_i\left(\widehat{\beta}_1\right) g_i\left(\widehat{\beta}_1\right)'$$

where

$$g_i\left(\widehat{\beta}_1\right) = z_i\left(\frac{y_i - \exp\left(x_i'\widehat{\beta}_1\right)}{\exp\left(x_i'\widehat{\beta}_1\right)}\right)$$

with $\widehat{\beta}_1$ an initial consistent estimator. Angrist (2001) strengthens the arguments for using these moment conditions for causal inference as he shows that in a model with endogenous treatment and a binary instrument, the Mullahy procedure estimates a proportional local average treatment effect (LATE) parameter in models with no covariates.

Windmeijer and Santos Silva (1997) propose use of the additive moment conditions

$$E\left((y_i - \mu_i)|z_i\right) = 0, \tag{4}$$

---

[1]From the conditional moments (3) it follows that any function $h(z)$ are valid instruments, which raises the issue of optimal instruments. Here, we will only consider $h(z) = z$.

estimating the parameters β again by GMM, with in this case $g_i = z_i (y_i - \mu_i)$. They and Mullahy (1997) compare the two sets of moment conditions and show that both sets cannot in general be valid at the same time. One exception is when there is classical measurement error in an explanatory variable, as in that case both additive and multiplicative moment conditions are valid. Consider the simple model

$$y_i = \exp\left(\alpha + x_i^* \beta\right) + u_i$$

but x_i^* is not observed. Instead we observe x_i

$$x_i = x_i^* + \varepsilon_i$$

and estimate β in the model

$$y_i = \exp\left(\alpha + x_i \beta - \varepsilon_i \beta\right) + u_i.$$

When instruments z_i are available that are correlated with x_i and independent of the i.i.d measurement errors ε_i, then the multiplicative moment conditions

$$E\left(\left(\frac{y_i - \widetilde{\mu}_i}{\widetilde{\mu}_i}\right) | z_i\right) = 0$$

are valid, where

$$\begin{aligned} \widetilde{\mu}_i &= \exp\left(\widetilde{\alpha} + x_i \beta\right) \\ \widetilde{\alpha} &= \alpha + \ln\left(E\left[\exp\left(-\varepsilon\beta\right) | z\right]\right), \end{aligned}$$

and the latter expectation is assumed to be a constant. However, also the additive moment conditions are valid as

$$E\left(\left(y_i - \widetilde{\widetilde{\mu}}_i\right) | z_i\right) = 0$$

where

$$\begin{aligned} \widetilde{\widetilde{\mu}}_i &= \exp\left(\widetilde{\widetilde{\alpha}} + x_i \beta\right) \\ \widetilde{\widetilde{\alpha}} &= \alpha - \ln\left(E\left[\exp\left(-\varepsilon\beta\right) | z\right]\right), \end{aligned}$$

and so an estimate for α can be recovered as well as the average of the two intercept estimates.

3 Panel Data Models

Let y_{it} denote the discrete count variable to be explained for subject i, $i = 1, ..., N$, at time t, $t = 1, ..., T$; and let x_{it} denote a vector of explanatory variables. An important feature in panel data applications is unobserved heterogeneity or individual fixed effects. For count data models these effects are generally modelled multiplicatively as

$$
\begin{aligned}
y_{it} &= \exp\left(x_{it}'\beta + \eta_i\right) + u_{it} \\
&= \mu_{it}\nu_i + u_{it},
\end{aligned}
$$

where $\nu_i = \exp\left(\eta_i\right)$ is a permanent scaling factor for the individual specific mean. In general, it is likely that the unobserved heterogeneity components η_i are correlated with the explanatory variables, $E\left(x_{it}\eta_i\right) \neq 0$, and therefore standard random effects estimators for β will be inconsistent, see Hausman, Hall and Griliches (1984). This section will describe moment conditions that can be used to consistently estimate the parameters β when there is correlation between η_i and x_{it} and allowing for different exogeneity properties of the explanatory variables, i.e. the regressors being strictly exogenous, predetermined or endogenous. Throughout we assume that the u_{it} are not serially correlated and that $E\left(u_{it}|\nu_i\right) = 0$, $t = 1, ..., T$.

3.1 Strictly Exogenous Regressors

When the x_{it} are *strictly exogenous*, there is no correlation between any of the idiosyncratic shocks u_{is}, $s = 1, ..., T$ and any of the x_{it}, $t = 1, ...T$, and the conditional mean of y_{it} satisfies

$$
E\left(y_{it}|\nu_i, x_{it}\right) = E\left(y_{it}|\nu_i, x_{i1}, ..., x_{iT}\right).
$$

For this case, Hausman, Hall and Griliches (1984) use the Poisson conditional maximum likelihood estimator (CMLE), conditioning on $\sum_{t=1}^{T} y_{it}$, which is the sufficient statistic for η_i. This method mimics the fixed effect logit approach of Chamberlain (1984). However, the Poisson maximum likelihood estimator (MLE) for β in a model with separate individual specific constants does not suffer from the incidental parameters problem, and

is therefore consistent and is the same as the CMLE. To see this, note that the maximum likelihood first order conditions for the ν_i are given by

$$\frac{\partial \ln L}{\partial \nu_i} = \sum_{t=1}^{T} \frac{\partial \left(y_{it} \ln (\mu_{it}\nu_i) - \mu_{it}\nu_i\right)}{\partial \nu_i} = \sum_{t=1}^{T} \left(\frac{y_{it}}{\nu_i} - \mu_{it}\right) = 0$$

and therefore the MLE for ν_i is given by

$$\widehat{\nu}_{i(ML)} = \frac{\overline{y}_i}{\overline{\mu}_i},$$

where $\overline{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ and $\overline{\mu}_i = T^{-1} \sum_{t=1}^{T} \exp(x_{it}'\beta)$. The MLE of the fixed effect is independent of ν_i. Substituting the fixed effects estimates in the first order conditions for β results in he moment conditions

$$\frac{\partial \ln L}{\partial \beta}(\widehat{\nu}_i) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i}\right) x_{it} = 0.$$

When x_{it} is strictly exogenous,

$$\text{plim}_{N \to \infty} \frac{1}{N} \frac{\partial \ln L}{\partial \beta}(\widehat{\nu}_i) = \text{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(u_{it} - \frac{\mu_{it}}{\overline{\mu}_i} \overline{u}_i\right) x_{it} = 0,$$

with $\overline{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$, and therefore the MLE for β is consistent.[2] It is further identical to the CMLE. The latter can be seen as follows. The Poisson conditional log likelihood function is given by

$$\ln CL = \sum_{i=1}^{N} \sum_{t=1}^{T} \Gamma(y_{it}+1) - \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \ln \left[\sum_{s=1}^{T} \exp\left(-(x_{it} - x_{is})'\beta\right)\right],$$

where $\Gamma(.)$ is the gamma function, see Hausman, Hall and Griliches (1984, p. 919). The first-order condition for β is

$$\begin{aligned}
\frac{\partial \ln CL}{\partial \beta} &= \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{it}}{\sum_{s=1}^{T} \exp\left(-(x_{it}-x_{is})'\beta\right)} \sum_{s=1}^{T} \exp\left(-(x_{it}-x_{is})'\beta\right)(x_{it}-x_{is}) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} x_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \frac{\sum_{s=1}^{T} x_{is} \exp\left(x_{is}'\beta\right)}{\sum_{s=1}^{T} \exp\left(x_{is}'\beta\right)} \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \left(y_{it} - \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i}\right),
\end{aligned}$$

---

[2]Lancaster (2002) finds the same result for the Poisson model by means of a decomposition of the likelihood.

which is exactly the same as the MLE first order condition for β.

The first order conditions imply that the Poisson MLE for β is equivalent to the moment estimator in a model where the ratio of individual, or within group, means are used to approximate the individual specific effects. This *mean scaling* model is given by

$$y_{it} = \mu_{it}\frac{\overline{y}_i}{\overline{\mu}_i} + u_{it}^*, \tag{5}$$

where $u_{it}^* = u_{it} - \frac{\mu_{it}}{\overline{\mu}_i}\overline{u}_i$. Blundell, Griffith and Windmeijer (2002) call the resulting estimator the *within group mean scaling* estimator.[3]

3.2 Predetermined Regressors

A regressor is predetermined when it is not correlated with current and future shocks, but it is correlated with past shocks:

$$\begin{aligned}
E\left(x_{it}u_{it+j}\right) &= 0, \quad j \geq 0 \\
E\left(x_{it}u_{it-s}\right) &\neq 0, \quad s \geq 1.
\end{aligned}$$

With predetermined regressors, the within group mean scaling estimator is no longer consistent. Chamberlain (1992) has proposed transformations that eliminate the fixed effect from the multiplicative model and generate orthogonality conditions that can be used for consistent estimation in count data models with predetermined regressors. The quasi-differencing transformation is

$$s_{it} = y_{it}\frac{\mu_{it-1}}{\mu_{it}} - y_{it-1} = u_{it}\frac{\mu_{it-1}}{\mu_{it}} - u_{it-1}.$$

Let $x_i^{t-1} = (x_{i1}, ..., x_{it-1})$. When x_{it} is predetermined, the following moment conditions hold:

$$E\left(s_{it}|x_i^{t-1}\right) = E\left(\left(E\left(u_{it}|x_i^t\right)\frac{\mu_{it-1}}{\mu_{it}} - u_{it-1}\right)|x_i^{t-1}\right) = 0. \tag{6}$$

Wooldridge (1991) proposed the following quasi-differencing transformation

$$q_{it} = \frac{y_{it}}{\mu_{it}} - \frac{y_{it-1}}{\mu_{it-1}} = \frac{u_{it}}{\mu_{it}} - \frac{u_{it-1}}{\mu_{it-1}},$$

---

[3]Clearly, the Poisson pseudo-likelihood results are preserved, see also Wooldridge (1999).

with moment conditions

$$E\left(q_{it}|x_i^{t-1}\right) = E\left(\left(\frac{E\left(u_{it}|x_i^t\right)}{\mu_{it}} - \frac{u_{it-1}}{\mu_{it-1}}\right)|x_i^{t-1}\right) = 0.$$

It is clear that a variable in x_{it} can not have only non-positive or non-negative values, as then the corresponding estimate for β is infinity. A way around this problem is to transform x_{it} in deviations from its overall mean, $\widetilde{x}_{it} = x_{it} - \overline{x}$, with $\overline{x} = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T x_{it}$, see Windmeijer (2000).

Both moment conditions can also be derived from a multiplicative model specification

$$y_{it} = \exp\left(x_{it}'\beta + \eta_i\right)w_{it} = \mu_{it}\nu_i w_{it},$$

where x_{it} is now predetermined w.r.t. w_{it}. Again, we assume that the w_{it} are not serially correlated and not correlated with ν_i, and $E\left(w_{it}\right) = 1$. The Chamberlain quasi-differencing transformation in this case is equivalent to

$$s_{it} = y_{it}\frac{\mu_{it-1}}{\mu_{it}} - y_{it-1} = \nu_i\mu_{it-1}\left(w_{it} - w_{it-1}\right),$$

with moment conditions

$$E\left(s_{it}|x_i^{t-1}\right) = E\left(\nu_i\mu_{it-1}E\left(\left(w_{it} - w_{it-1}\right)|\nu_i, x_i^{t-1}\right)|x_i^{t-1}\right) = 0.$$

Equivalently, for the Wooldridge transformation,

$$q_{it} = \frac{y_{it}}{\mu_{it}} - \frac{y_{it-1}}{\mu_{it-1}} = \nu_i\left(w_{it} - w_{it-1}\right)$$

and

$$E\left(q_{it}|x_i^{t-1}\right) = E\left(\nu_i E\left(\left(w_{it} - w_{it-1}\right)|\nu_i, x_i^{t-1}\right)|x_i^{t-1}\right) = 0.$$

3.3 Endogenous Regressors

Regressors are endogenous when they are correlated with current (and possibly past) shocks $E\left(x_{it}u_{it-s}\right) \neq 0$, $s \geq 0$, for the specification with additive errors u_{it}, or when $E\left(x_{it}w_{it-s}\right) \neq 0$, $s \geq 0$, for the specification with multiplicative errors w_{it}. It is clear from the derivations in the previous section that we cannot find valid sequential conditional

moment conditions for the specification with additive errors due to the non-separability of the u_{it} and μ_{it}. For the multiplicative error specification, there is again non-separability of μ_{it-1} and $(w_{it} - w_{it-1})$ for the Chamberlain transformation and so

$$E\left(s_{it}|x_i^{t-2}\right) = E\left(\nu_i\mu_{it-1}E\left(\left(w_{it} - w_{it-1}\right)|\nu_i, x_i^{t-1}\right)|x_i^{t-2}\right) \neq 0.$$

In contrast, the Wooldridge transformation does not depend on μ_{it} or μ_{it-1} in this case. Valid moment conditions are then given by

$$E\left(q_{it}|x_i^{t-2}\right) = E\left(\nu_i E\left(\left(w_{it} - w_{it-1}\right)|\nu_i, x_i^{t-2}\right)|x_i^{t-2}\right) = 0.$$

Therefore, in the case of endogenous explanatory variables, only the Wooldridge transformation can be used for the consistent estimation of the parameters β. This includes the case of classical measurement error in x_{it}, where the measurement error is not correlated over time.

3.4 Dynamic Models

Specifying dynamic models for count data by including lags of the dependent count variables in the explanatory part of the model is not as straightforward as with linear models for a continuous dependent variable. Inclusion of the lagged dependent variable in the exponential mean function may lead to rapidly exploding series. A better starting point is to specify the model as in Crépon and Duguet (1997)

$$y_{it} = h\left(y_{it-1}, \gamma\right)\exp\left(x_{it}'\beta + \eta_i\right) + u_{it},$$

where $h\left(.,.\right) > 0$ is any given function describing the way past values of the dependent variable are affecting the current value.

Let

$$d_{it} = 1_{\{y_{it}=0\}},$$

then a possible choice for $h\left(.,.\right)$ is

$$h\left(y_{it}, \gamma\right) = \exp\left(\gamma_1 \ln\left(y_{it-1} + cd_{it-1}\right) + \gamma_2 d_{it-1}\right),$$

where c is a pre-specified constant. In this case, $\ln(y_{it-1})$ is included as a regressor for positive y_{it-1}, and zero values of y_{it-1} have a separate effect on current values of y_{it}. Crépon and Duguet (1997) considered

$$h(y_{it}, \gamma) = \exp(\gamma(1 - d_{it-1})),$$

and extensions thereof to several regime indicators.

Blundell, Griffith and Windmeijer (2002) propose use of a linear feedback model for modelling dynamic count panel data process. The linear feedback model of order 1 (LFM(1)) is defined as

$$\begin{aligned} y_{it} &= \gamma y_{it-1} + \exp(x'_{it}\beta + \eta_i) + u_{it} \\ &= \gamma y_{it-1} + \mu_{it}\nu_i + u_{it}, \end{aligned}$$

where the lag of the dependent variable enters the model linearly. Extensions to include further lags are straightforward. The LFM has its origins in the Integer-Valued Autoregressive (INAR) process and can be motivated as an entry-exit process with the probability of exit equal to $(1 - \gamma)$. The correlation over time for the INAR(1) process without additional regressors is similar to that of the AR(1) model, $corr(y_{it}, y_{it-j}) = \gamma^j$.

For the patents-R&D model, Blundell, Griffith and Windmeijer (2002) consider the economic model

$$P_{it} = k\left(R_{it}^{\beta} + (1 - \delta)R_{it-1}^{\beta} + (1 - \delta)^2 R_{it-2}^{\beta}...\right)\nu_i + \varepsilon_{it} \tag{7}$$

where P_{it} and R_{it} are the number of patents and R&D expenditures for firm i at time t respectively, k is a positive constant and R&D expenditures depreciate geometrically at rate δ. The long run steady state for firm i, ignoring feedback from patents to R&D, can be written as

$$P_i = \frac{k}{\delta}R_i^{\beta}\nu_i,$$

and β can therefore be interpreted as the long run elasticity. Inverting (7) leads to

$$P_{it} = kR_i^{\beta}\nu_i + (1 - \delta)P_{it-1} + u_{it}$$

11

and so in the LFM model

$$P_{it} = \gamma P_{it-1} + \exp\left(k^* + \beta \ln\left(R_{it}\right)\right)\nu_i + u_{it},$$

the estimate for γ is an estimate of the depreciation factor $(1 - \delta)$ and the estimate for β is an estimate of the *long run* elasticity of the R&D returns to patents.

Even when the x_{it} are strictly exogenous, the within group mean scaling estimator will be inconsistent for small T, as the lagged dependent variable is a predetermined variable. For estimation by GMM, the Chamberlain quasi-differencing transformation for the LFM(1) model is given by

$$s_{it} = (y_{it} - \gamma y_{it-1})\frac{\mu_{it-1}}{\mu_{it}} - (y_{it-1} - \gamma y_{it-2}) \tag{8}$$

and the Wooldridge quasi-differencing transformation is given by

$$q_{it} = \frac{y_{it} - \gamma y_{it-1}}{\mu_{it}} - \frac{y_{it-1} - \gamma y_{it-2}}{\mu_{it-1}}.$$

For predetermined x_{it} the following moment conditions hold

$$E(s_{it}|y_i^{t-2}, x_i^{t-1}) = 0;$$
$$E(q_{it}|y_i^{t-2}, x_i^{t-1}) = 0,$$

while for endogenous x_{it}, only the Wooldridge moment conditions are valid

$$E(q_{it}|y_i^{t-2}, x_i^{t-2}) = 0.$$

4   GMM

The orthogonality conditions as described in the sections above can be used to consistently estimate the model parameters by the GMM estimation technique (see Hansen, 1982). Let θ be the k-vector of parameters to be estimated, e.g. for the LFM(1) model $\theta = (\gamma, \beta')'$. The model has a true parameter θ_0 satisfying the q moment conditions

$$E\left[g_i\left(\theta_0\right)\right] = 0.$$

The GMM estimator $\widehat{\theta}$ for θ_0 is defined as

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \left[\frac{1}{N} \sum_{i=1}^{N} g_i(\theta) \right]' W_N^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} g_i(\theta) \right],$$

where Θ is a compact set of parameter values; W_N satisfies $\text{plim}_{N\to\infty} W_N = W$, with W a positive definite matrix. Regularity conditions are assumed such that $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} g_i(\theta) = E[g_i(\theta)]$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g_i(\theta_0) \to N(0, \Psi)$ where $\Psi = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left[\left(g_i(\theta_0) g_i(\theta_0)' \right) \right]$. Let $\Gamma(\theta) = E[\partial g_i(\theta)/\partial\theta']$ and $\Gamma_{\theta_0} \equiv \Gamma(\theta_0)$, then $\sqrt{N}\left(\widehat{\theta} - \theta_0\right)$ has a limiting normal distribution,

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \to N(0, V_W),$$

where

$$V_W = \left(\Gamma_{\theta_0}' W^{-1} \Gamma_{\theta_0} \right)^{-1} \Gamma_{\theta_0}' W^{-1} \Psi W^{-1} \Gamma_{\theta_0} \left(\Gamma_{\theta_0}' W^{-1} \Gamma_{\theta_0} \right)^{-1}.$$

The efficient two-step GMM estimator, denoted $\widehat{\theta}_2$, is based on a weight matrix that satisfies $\text{plim}_{N\to\infty} W_N = \Psi$, with $V_W = \left(\Gamma_{\theta_0}' \Psi^{-1} \Gamma_{\theta_0} \right)^{-1}$ in that case. A weight matrix that satisfies this property is given by

$$W_N\left(\widehat{\theta}_1\right) = \frac{1}{N} \sum_{i=1}^{N} g_i\left(\widehat{\theta}_1\right) g_i\left(\widehat{\theta}_1\right)',$$

where $\widehat{\theta}_1$ is an initial consistent estimator for θ_0.

Denote $\overline{g}(\theta) = \frac{1}{N} \sum_{i=1}^{N} g_i(\theta)$. The standard test for overidentifying restrictions is N times the minimised GMM criterion

$$N\overline{g}\left(\widehat{\theta}_2\right)' W_N^{-1}\left(\widehat{\theta}_1\right) \overline{g}\left(\widehat{\theta}_2\right),$$

which has an asymptotic chi-squared distribution with $q - k$ degrees of freedom when the moment conditions are valid.

For the Chamberlain quasi-differencing transformation the GMM estimator $\widehat{\theta}$ minimises

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \left(\frac{1}{N} \sum_{i=1}^{N} s_i(\theta)' Z_i \right) W_N^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' s_i(\theta) \right),$$

where, for the LFM(1) model, $s_i(\theta)$ is the $T-2$ vector $(s_{i3}, s_{i4}, ..., s_{iT})'$, with s_{it} as defined in (8), Z_i is the matrix of instruments and W_N is a weight matrix. When the full

sequential set of instruments is used and x_{it} is predetermined, the instrument matrix for the LFM(1) model is given by

$$
Z_i = \begin{bmatrix} y_{i1} & x_{i1} & x_{i2} & & & & & \\ & & & \ddots & & & & \\ & & & & y_{i1} & \cdots & y_{iT-2} & x_{i1} & \cdots & x_{iT-1} \end{bmatrix}.
$$

The efficient weight matrix is

$$
W_N\left(\widehat{\theta}_1\right) = \frac{1}{N}\sum_{i=1}^{N} Z_i' s_i(\widehat{\theta}_1) s_i(\widehat{\theta}_1)' Z_i,
$$

where $\widehat{\theta}_1$ can be a GMM estimator using for example $W_N = \frac{1}{N}\sum_{i=1}^{N} Z_i' Z_i$ as the initial weight matrix. As stated above, under the assumed regularity conditions both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are asymptotically normally distributed. The asymptotic variance of $\widehat{\theta}_1$ is estimated by

$$
\begin{aligned}
\widehat{\mathrm{var}}\left(\widehat{\theta}_1\right) &= \frac{1}{N}\left(C\left(\widehat{\theta}_1\right)' W_N^{-1} C\left(\widehat{\theta}_1\right)\right)^{-1} C\left(\widehat{\theta}_1\right)' W_N^{-1} W_N\left(\widehat{\theta}_1\right) W_N^{-1} C\left(\widehat{\theta}_1\right) \\
&\quad \times \left(C\left(\widehat{\theta}_1\right)' W_N^{-1} C\left(\widehat{\theta}_1\right)\right)^{-1}
\end{aligned}
$$

where

$$
C\left(\widehat{\theta}_1\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial Z_i' s_i\left(\theta\right)}{\partial\theta}|_{\widehat{\theta}_1}.
$$

The asymptotic variance of the efficient two-step GMM, estimator is estimated by

$$
\widehat{\mathrm{var}}\left(\widehat{\theta}_2\right) = \frac{1}{N}\left(C\left(\widehat{\theta}_2\right)' W_N^{-1}\left(\widehat{\theta}_1\right) C\left(\widehat{\theta}_2\right)\right)^{-1}.
$$

5 Applications and software

The instrumental variables methods for count data models with endogenous regressors using cross section data, as described in Section 2, have often been applied in the health econometrics literature. For example, Mullahy (1997) uses the multiplicative moment conditions to estimate cigarette demand functions with a habit stock measure as endogenous regressor. Windmeijer and Santos Silva (1997) estimate health care demand functions with a self reported health measure as possible endogenous variable, while Vera-Hernandez (1999) and Schellhorn (2001) estimate health care demand functions with

14

endogenous insurance choice. An example outside the health econometrics literature is Kelly (2000) who models the number of crimes with police activity as an endogenous regressor.

The count panel data literature has largely focused on estimating models for patenting and the returns to R&D investments, which started with the seminal paper of Hausman, Hall and Griliches (1984). Following the development of the quasi-differencing approach of Chamberlain (1992) and Wooldridge (1991,1997), Montalvo (1997), Cincera (1997), Crépon and Duguet (1997), Blundell, Griffith and Van Reenen (1999) and Blundell, Griffith and Windmeijer (2002) developed and/or estimated patent (or innovation) production functions using the GMM framework allowing for correlated firm specific effects and weakly exogenous inputs. More recently, Kim and Marschke (2005) use the GMM framework to find a relationship between a firms' patenting behaviour and scientist turnover, whereas Salomon and Shaver (2005) estimate a linear feedback model and find that exporting has a positive effect on innovating behaviour of the firm.

The latter two publications estimated the models using *ExpEnd*, Windmeijer (2002). This is a user friendly open source GAUSS (Aptech Systems, 2005) code for the nonlinear GMM estimation of the models described in the previous sections.[4] For cross-section data, *ExpEnd* estimates simple robust Poisson regression models using moment conditions (1); and instrumental variables regressions using Mullahy's (1997) multiplicative moment conditions (3) or the additive moment conditions (4). For panel data, *ExpEnd* estimates pooled robust Poisson regression models; fixed effects models, using the mean scaling model (5); and the quasi-differencing models using the Chamberlain (1992) or the Wooldridge (1991,1997) transformation, both for static and linear feedback models. For the quasi-differencing models, sequential and so-called stacked IV type instruments can be specified, in both cases allowing for a flexible lag length of the instruments. For overidentified models one- and two-step GMM parameter estimates are reported, together with asymptotic standard errors. The test for overidentifying restrictions is reported and for the panel data models the output further includes a test of first and second order

---

[4]For a review, see Romeu (2004).

serial correlation of the quasi-differencing "residuals" $s_{it}\left(\widehat{\theta}\right)$ or $q_{it}\left(\widehat{\theta}\right)$. If the model is correctly specified one expects to find an MA(1) serial correlation structure.

Another package that enables researchers to estimate these model types is TSP, Hall and Cummins (2005). Kitazawa (2000) provides various TSP procedures for the estimation of count panel data models. Also LIMDEP, Greene (2005), provides an environment where these models can be estimated.

6   Finite sample inference

Standard Wald tests based on two-step efficient GMM estimators are known to have poor finite sample properties (see e.g. Blundell and Bond (1998)). Bond and Windmeijer (2005) therefore analysed the finite sample performance of various alternative test procedures for testing linear restrictions in linear panel data models. The statistics they found to perform well in Monte Carlo exercises were an alternative two-step Wald test that uses a finite sample correction for the asymptotic variance matrix, the LM test, and a simple criterion-based test. In this section we briefly describe these procedures and adapt them to the case of nonlinear GMM estimation where necessary.

Newey and Smith (2004) have shown that the GMM estimator can further also suffer from quite large finite sample biases and advocate use of Generalized Empirical Likelihood (GEL) estimators that they show to have smaller finite sample biases. We will consider here the performance of the Continuous Updating Estimator (CUE) as proposed by Hansen, Heaton and Yaron (1996), which is a GEL estimator. The Wald test based on the CUE has also been shown to perform poorly in finite samples by e.g. Hansen, Heaton and Yaron (1996). Newey and Windmeijer (2005) derive the asymptotic distribution of the CUE when there are many weak moment conditions. The asymptotic variance in this case is larger than the usual asymptotic one and we will analyse the performance of an alternative Wald test that uses an estimate for this larger asymptotic variance, together with a criterion based test for the CUE as proposed by Hansen, Heaton and Yaron (1996).

The estimators and test procedures will be evaluated in a Monte Carlo study of testing linear restrictions in a static count panel data model with an explanatory variable that is correlated with the fixed unobserved heterogeneity and which is predetermined. The Chamberlain quasi-differencing transformation will be used with sequential moment conditions.

6.1 Wald test and finite sample variance correction

The standard Wald test for testing r linear restrictions of the form $r(\theta_0) = 0$ is calculated as

$$\text{Wald} = r\left(\widehat{\theta}\right)' \left(R'\widehat{\text{var}}\left(\widehat{\theta}\right) R\right)^{-1} r\left(\widehat{\theta}\right),$$

where $R = \partial r(\theta) / \partial \theta'$, and has an asymptotic χ_r^2 distribution under the null. Based on the two-step GMM estimator and using its conventional asymptotic variance estimate, the Wald test has often been found to overreject correct null hypotheses severely compared to its nominal size. This can occur even when the estimator has negligible finite sample bias, due to the fact that the estimated asymptotic standard errors can be severely downward biased in small samples. Windmeijer (2005) proposed a finite sample variance correction that takes account of the extra variation due to the presence of the estimated parameters $\widehat{\theta}_1$ in the weight matrix. He showed by means of a Monte Carlo study that this correction works well in linear models, but it is not clear how well it will work in nonlinear GMM.

To derive the finite sample corrected variance, let

$$\overline{g}(\theta) = \frac{1}{N} \sum_{i=1}^{N} g_i(\theta); \quad C(\theta) = \frac{\partial \overline{g}(\theta)}{\partial \theta'}; \quad G(\theta) = \frac{\partial C(\theta)}{\partial \theta},$$

and

$$
\begin{aligned}
b_{\theta_0, W_N} &= \frac{1}{2} \frac{\partial Q_{W_N}}{\partial \theta}|_{\theta_0} = C(\theta_0)' W_N^{-1} \overline{g}(\theta_0); \\
A_{\theta_0, W_N} &= \frac{1}{2} \frac{\partial^2 Q_{W_N}}{\partial \theta \partial \theta'}|_{\theta_0} = C(\theta_0)' W_N^{-1} C(\theta_0) + G(\theta_0)' \left(I_k \otimes W_N^{-1} \overline{g}(\theta_0)\right).
\end{aligned}
$$

A standard first order Taylor series approximation of $\widehat{\theta}_2$ around θ_0, conditional on $W_N\left(\widehat{\theta}_1\right)$, results in

$$\widehat{\theta}_2 - \theta_0 = -A_{\theta_0, W_N\left(\widehat{\theta}_1\right)}^{-1} b_{\theta_0, W_N\left(\widehat{\theta}_1\right)} + O_p\left(N^{-1}\right).$$

A further expansion of $\widehat{\theta}_1$ around θ_0 results in

$$\widehat{\theta}_2 - \theta_0 = -A^{-1}_{\theta_0,W_N(\theta_0)}b_{\theta_0,W_N(\theta_0)} + D_{\theta_0,W_N(\theta_0)}\left(\widehat{\theta}_1 - \theta_0\right) + O_p\left(N^{-1}\right), \qquad (9)$$

where

$$W_N\left(\theta_0\right) = \frac{1}{N}\sum_{i=1}^{N} g_i\left(\theta_0\right) g_i\left(\theta_0\right)'$$

and

$$D_{\theta_0,W_N(\theta_0)} = \frac{\partial}{\partial\theta'}\left(-A^{-1}_{\theta_0,W_N(\theta)}b_{\theta_0,W_N(\theta)}\right)|_{\theta_0}$$

is a $k \times k$ matrix.

Let $\widehat{\theta}_1$ be a one-step GMM estimator that uses a weight matrix W_N that does not depend on estimated parameters. An estimate of the variance of $\widehat{\theta}_2$ that incorporates the term involving the one-step parameter estimates used in the weight matrix can then be obtained as

$$
\begin{aligned}
\widehat{\mathrm{var}}_c\left(\widehat{\theta}_2\right) \ = \ & \frac{1}{N}A^{-1}_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}C\left(\widehat{\theta}_2\right)'W_N^{-1}\left(\widehat{\theta}_1\right)C\left(\widehat{\theta}_2\right)A^{-1}_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)} \\
& + \frac{1}{N}D_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}A^{-1}_{\widehat{\theta}_1,W_N}C\left(\widehat{\theta}_1\right)'W_N^{-1}C\left(\widehat{\theta}_2\right)A^{-1}_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)} \\
& + \frac{1}{N}A^{-1}_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}C\left(\widehat{\theta}_2\right)'W_N^{-1}C\left(\widehat{\theta}_1\right)A^{-1}_{\widehat{\theta}_1,W_N}D'_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)} \\
& + D_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}\widehat{\mathrm{var}}\left(\widehat{\theta}_1\right)D'_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)},
\end{aligned}
$$

where the jth column of $D_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}$ is given by

$$D_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}[.,j] = A^{-1}_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}C\left(\widehat{\theta}_2\right)'W_N^{-1}\left(\widehat{\theta}_1\right)\frac{\partial W_N\left(\theta\right)}{\partial\theta_j}|_{\widehat{\theta}_2}W_N^{-1}\left(\widehat{\theta}_1\right)\overline{g}\left(\widehat{\theta}_2\right),$$

and

$$\frac{\partial W_N\left(\theta\right)}{\partial\theta_j} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial g_i\left(\theta\right)}{\partial\theta_j}g_i\left(\theta\right)' + g_i\left(\theta\right)\frac{\partial g_i\left(\theta\right)'}{\partial\theta_j}\right).$$

The alternative two-step Wald test that uses a finite sample correction for the asymptotic variance matrix is then defined as

$$\mathrm{Wald}_c = r\left(\widehat{\theta}_2\right)'\left(R'\widehat{\mathrm{var}}_c\left(\widehat{\theta}_2\right)R\right)^{-1}r\left(\widehat{\theta}_2\right).$$

The term $D_{\theta_0, W(\theta_0)} \left(\widehat{\theta}_1 - \theta_0 \right)$ in (9) is itself $O_p \left(N^{-1} \right)$ and in this general setting, incorporating non-linear models and/or non-linear moment conditions, whether taking account of it will improve the estimation of the small sample variance substantially depends on the other remainder terms which are of the same order.

6.2   Criterion-Based Tests

Using the notation as in Bond and Windmeijer (2005), the standard two-step Wald test can be computed as a criterion difference

$$\text{Wald} = N \left(\overline{g} \left(\widetilde{\theta}_{\widehat{2}} \right)' W_N^{-1} \left(\widehat{\theta}_1 \right) \overline{g} \left(\widetilde{\theta}_{\widehat{2}} \right) - \overline{g} \left(\widehat{\theta}_2 \right)' W_N^{-1} \left(\widehat{\theta}_1 \right) \overline{g} \left(\widehat{\theta}_2 \right) \right),$$

where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are the one-step and two-step GMM estimators in the unrestricted model, whereas $\widetilde{\theta}_{\widehat{2}}$ is a two-step GMM estimator in the restricted model, but using as a weight matrix the consistent estimate of Ψ based on the unrestricted one-step GMM estimator, $W_N \left(\widehat{\theta}_1 \right)$, see Newey and West (1987).

The LM test can also be computed as a criterion difference

$$\text{LM} = N \left(\overline{g} \left(\widetilde{\theta}_2 \right)' W_N^{-1} \left(\widetilde{\theta}_1 \right) \overline{g} \left(\widetilde{\theta}_2 \right) - \overline{g} \left(\widehat{\theta}_{\widetilde{2}} \right)' W_N^{-1} \left(\widetilde{\theta}_1 \right) \overline{g} \left(\widehat{\theta}_{\widetilde{2}} \right) \right),$$

where $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ are the one-step and two-step GMM estimators in the restricted model, whereas $\widehat{\theta}_{\widetilde{2}}$ is a two-step GMM estimator in the unrestricted model, but using as a weight matrix the consistent estimate of Ψ under the null, based on the restricted one-step GMM estimator, $W_N \left(\widetilde{\theta}_1 \right)$. The LM test has an asymptotic χ_r^2 distribution under the null.

The criterion-based test statistic considered by Bond, Bowsher and Windmeijer (2001) is given by

$$D_{RU} = N \left(\overline{g} \left(\widetilde{\theta}_2 \right)' W_N^{-1} \left(\widetilde{\theta}_1 \right) \overline{g} \left(\widetilde{\theta}_2 \right) - \overline{g} \left(\widehat{\theta}_2 \right)' W_N^{-1} \left(\widehat{\theta}_1 \right) \overline{g} \left(\widehat{\theta}_2 \right) \right).$$

D_{RU} is the "likelihood ratio" test equivalent for GMM, and is the difference between the test statistics for overidentifying restrictions in the restricted and unrestricted models. Under the null, D_{RU} has an asymptotic χ_r^2 distribution.

6.3 Continuous Updating Estimator

The Continuous Updating Estimator (CUE) is given by

$$\widehat{\theta}_{CU} = \arg\min_{\theta \in \Theta} Q(\theta);$$

$$Q(\theta) = \frac{1}{2}\overline{g}(\theta)' W_N^{-1}(\theta)\,\overline{g}(\theta),$$

where, as before,

$$W_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\theta)\,g_i(\theta)'$$

and so the CUE minimises the criterion function including the parameters in the weight matrix. The limiting distribution under standard regularity conditions is given by

$$\sqrt{N}\left(\widehat{\theta}_{CU} - \theta_0\right) \to N(0, V); \quad V = \left(\Gamma'_{\theta_0}\Psi^{-1}\Gamma_{\theta_0}\right)^{-1}$$

and is the same as the efficient two-step GMM estimator. The asymptotic variance of the CUE is computed as

$$\widehat{\mathrm{var}}\left(\widehat{\theta}_{CU}\right) = \frac{1}{N}\left(C\left(\widehat{\theta}_{CU}\right)' W_N^{-1}\left(\widehat{\theta}_{CU}\right) C\left(\widehat{\theta}_{CU}\right)\right)^{-1},$$

which is used in the calculation of the standard Wald test. Again, it has been shown by e.g. Hansen, Heaton and Yaron (1996) that the asymptotic standard errors are severely downward biased, leading to overrejection of a true null hypothesis using the Wald test.

Newey and Windmeijer (2005) derive the asymptotic distribution of the CUE under many weak instrument asymptotics. In these asymptotics, the number of instruments is allowed to grow with the sample size N, with the increase in number of instruments accompanied by an increase in the concentration parameter. The resulting limiting distribution of the CUE is again the normal distribution, but convergence is at a slower rate than $\sqrt{N}$. The asymptotic variance is in this case larger than the asymptotic variance using conventional asymptotics, and can be estimated consistently by

$$\widehat{\mathrm{var}}\left(\widehat{\theta}_{CU}\right)_c = \frac{1}{N}H^{-1}\left(\widehat{\theta}_{CU}\right) S\left(\widehat{\theta}_{CU}\right)' W_N^{-1}\left(\widehat{\theta}_{CU}\right) S\left(\widehat{\theta}_{CU}\right) H^{-1}\left(\widehat{\theta}_{CU}\right),$$

where

$$
\begin{aligned}
H\left(\theta\right) &= \frac{\partial^2 Q\left(\theta\right)}{\partial\theta\partial\theta'}; \quad S\left(\theta\right) = \left(S_1\left(\theta\right), S_2\left(\theta\right), ..., S_k\left(\theta\right)\right) \\
S_j\left(\theta\right) &= \left(\frac{\partial \overline{g}\left(\theta\right)}{\partial\theta_j} - \Lambda_j\left(\theta\right) W_N^{-1}\left(\theta\right) \overline{g}\left(\theta\right)\right) \\
\Lambda_j\left(\theta\right) &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g_i\left(\theta\right)}{\partial\theta_j} g_i\left(\theta\right)'.
\end{aligned}
$$

Here, unlike the usual asymptotics, the middle matrix $S\left(\widehat{\theta}_{CU}\right)' W_N^{-1}\left(\widehat{\theta}_{CU}\right) S\left(\widehat{\theta}_{CU}\right)$ estimates a different, larger object than the Hessian. Also, the use of the Hessian is important, as the more common formula $C\left(\widehat{\theta}_{CU}\right)' W_N^{-1}\left(\widehat{\theta}_{CU}\right) C\left(\widehat{\theta}_{CU}\right)$ has extra random terms that are eliminated in the Hessian under the alternative asymptotics.

Hansen, Heaton and Yaron (1996) proposed the use of a criterion-based test similar to D_{RU}, but based on the CUE. Their test statistic D_{RU}^{CU} is defined as

$$
D_{RU}^{CU} = 2N\left(Q\left(\widetilde{\theta}_{CU}\right) - Q\left(\widehat{\theta}_{CU}\right)\right),
$$

where $\widehat{\theta}_{CU}$ and $\widetilde{\theta}_{CU}$ are the CUEs for the unrestricted and restricted models respectively. Under the null, D_{RU}^{CU} has an asymptotic χ_r^2 distribution.

6.4 Monte Carlo results

In this section we will illustrate the finite sample performance of the GMM estimators and the test statistics as discussed in the previous sections by means of a small Monte Carlo study. The data generating process is given by

$$
\begin{aligned}
y_{it} &\sim Poisson\left(\exp\left(x_{it}\beta + \eta_i + \varepsilon_{it}\right)\right) \\
x_{it} &= \rho x_{it-1} + \delta\eta_i + \theta\varepsilon_{it-1} + \omega_{it} \\
\eta_i &\sim N\left(0, \sigma_\eta^2\right); \ \varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right); \ \omega_{it} \sim N\left(0, \sigma_\omega^2\right), \\
\beta &= 0.5; \ \delta = 0.1; \ \theta = 0.3; \ \sigma_\eta^2 = 0.3; \ \sigma_\varepsilon^2 = 0.3; \ \sigma_\omega^2 = 0.25 \\
\rho &= \{0.5, 0.8\}.
\end{aligned}
$$

The dependent variable is a count variable, generated from the Poisson distribution with unobserved fixed normally distributed heterogeneity η_i and further idiosyncratic normally

distributed heterogeneity ε_{it}. The x_{it} are correlated with the η_i and ε_{it-1} and are therefore predetermined.

Table 1 presents estimation results from $10,000$ Monte Carlo replications for the one- and two-step GMM estimators as well as the continuous updating estimator for $T = 6$, $N = 250$ and $\rho = 0.5$ or $\rho = 0.8$, using the moment conditions (6) as proposed by Chamberlain (1992). The instruments set is given by

$$
Z_i = \begin{bmatrix} x_{i1} & & & & \\ & x_{i1} & x_{i2} & & \\ & & & \ddots & \\ & & & x_{i1} & \cdots & x_{iT-1} \end{bmatrix}
$$

and hence there are a total of 15 moment conditions. The one-step GMM estimator uses $W_N = \frac{1}{N} \sum_{i=1}^{N} Z_i' Z_i$ as the weight matrix.

When $\rho = 0.5$, the instruments are quite strong. The one-step GMM estimator, denoted GMM1 in the table, has a moderate downward bias of -0.0408. Its standard deviation is 0.1053, which seems well approximated by the asymptotic standard error. The mean of the estimated standard errors is equal to 0.1031. The two-step GMM estimator, denoted GMM2, has a smaller bias of -0.0211 and a smaller standard deviation of 0.0803, representing a substantial efficiency gain with more than a 23% reduction in standard deviation. In contrast to the one-step estimator, the mean of the usual asymptotic standard errors is 19% smaller than the standard deviation. However, taking account of the extra variation due to the presence of the one-step estimates in the weight matrix results in finite sample corrected standard errors with a mean of 0.0799, which is virtually identical to the standard deviation. The CUE has a very small bias of 0.0043, with a standard deviation of 0.0904, which is larger than that of the two-step GMM estimator, but smaller than that of the one-step estimator. The mean of the usual asymptotic standard errors is exactly the same as that of the two-step GMM estimator and in this case it is almost 28% smaller than the standard deviation. The standard errors resulting from the many weak instruments asymptotics have a mean of 0.0918, which is virtually the same as the standard deviation.

Table 1. Estimation results

$\rho = 0.5$	mean bias	st dev	se	sec	median bias	IQR
GMM1	-0.0408	0.1053	0.1031		-0.0409	0.1381
GMM2	-0.0211	0.0803	0.0652	0.0799	-0.0209	0.1077
CUE	0.0043	0.0904	0.0652	0.0918	0.0024	0.1165
$\rho = 0.8$						
GMM1	-0.1136	0.2094	0.1773		-0.0974	0.2435
GMM2	-0.0537	0.1335	0.0908	0.1365	-0.0498	0.1558
CUE	0.0033	0.1885	0.0879	0.1459	0.0029	0.1742

Note: $T = 6$, $N = 250$, 10000 replications, sec denotes finite sample corrected standard errors for GMM2 and those resulting from many weak instrument asymptotics for CUE, IQR=Inter Quartile Range

Figure 1 shows p-value plots for the hypothesis $H_0 : \rho = 0.5$, comparing nominal size with rejection frequencies. The various Wald tests are denoted W_1, W_2, W_{2C}, W_{CU} and W_{CUC} based on one-step GMM, two-step GMM with usual standard errors, two-step GMM with finite sample corrected standard errors, CUE and CUE with standard errors resulting from the many weak instruments asymptotics, respectively. As expected, W_2 and W_{CU} overreject the null hypothesis substantially. W_1 and W_{2C} perform much better, but are still moderately oversized due to the bias of the estimators. W_{CUC} has a very good performance in terms of size of the test, the rejection frequencies being very close to the 45° line. The two-step GMM based LM and D_{RU} tests also perform very well, their p-value plots being quite similar to that of W_{CUC}. Finally, the CUE based D_{RU}^{CU} performs well, but tends to overreject moderately at higher values of the nominal size.

When $\rho = 0.8$, the instruments become weaker as the x_{it} series become more persistent. The one-step GMM estimator now has a quite large downward bias of -0.1136. Its standard deviation is 0.2094, which is now less well approximated by the asymptotic standard error, with the mean of the estimated standard errors being equal to 0.1773. The two-step GMM estimator has a smaller, but still quite substantial bias of -0.0537 and a smaller standard deviation of 0.1335. The mean of the usual asymptotic standard errors is 0.0908, again substantially smaller than the standard deviation. The mean of the finite sample corrected standard errors is 0.1365, which is again very close to the standard deviation. The CUE, also with these weaker instruments, has a very small bias
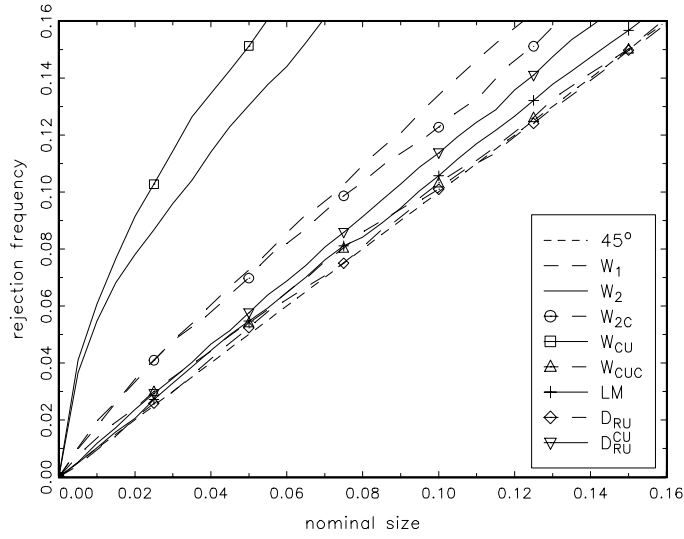
23

Figure 1. P-value plot, $H_0 : \rho = 0.5$

of 0.0033, with a standard deviation of 0.1885. In this case the so-called no moment-problem starts to become an issue for the CUE, though, with some outlying estimates inflating the standard deviation, see Guggenberger (2005). It is therefore better to look at the median bias and inter quartile range (IQR) in this case, which shows that the CUE is median unbiased with an IQR which is only slightly larger than that of the two-step GMM estimator, 0.1742 versus 0.1558 respectively.

Figure 2 shows p-value plots for the hypothesis $H_0 : \rho = 0.8$. W_2 and W_{CU} overreject the null hypothesis even more than when $\rho = 0.5$. W_1 performs better, but is still substantially oversized. W_{2C} and W_{CUC} perform quite well and quite similar, slightly overrejecting the null. The two-step GMM based LM and D_{RU} are again the best performers in terms of size, whereas the CUE based D_{RU}^{CU} performs worse than W_{2C} and W_{CUC}.
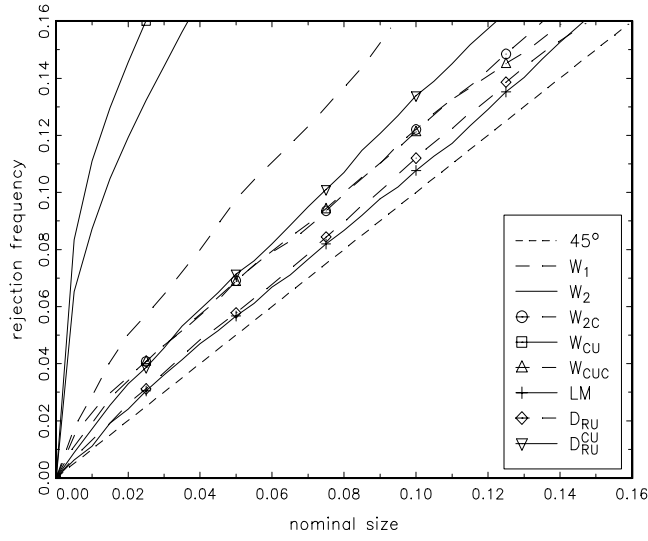
24

Figure 2. P-value plot, $H_0 : \rho = 0.8$

Summarising, it is clear that use of the finite sample corrected standard errors for the two-step GMM estimator and the standard errors from the many weak instrument asymptotics for the CUE improve the size performance of the Wald tests for these estimators considerably. The simple criterion based D_{RU} test performs very well in these examples, as was the case in Bond and Windmeijer (2005) for linear panel data models.

References

Angrist J.D., 2001, Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice, Journal of Business & Economic Statistics 19, 2-16.

Aptech Systems, Inc., 2005, GAUSS, Advanced Mathematical and Statistical System, Black Diamond, WA, USA.

Blundell R., R. Griffith and J. van Reenen, 1999, Market share, market value and innovation in a panel of British manufacturing firms, Review of Economic Studies 66, 529-554.

Blundell R., R. Griffith and F. Windmeijer, 2002, Individual effects and dynamics in count data models, Journal of Econometrics 108, 113-131.

Bond S.R., C. Bowsher and F. Windmeijer, 2001, Criterion-Based inference for GMM in autoregressive panel data models, Economics Letters 73, 379-388.

Bond S.R. and F. Windmeijer, 2005, Reliable inference for GMM estimators? Finite sample procedures in linear panel data models, Econometric Reviews 24, 1-37.

Chamberlain G., 1992, Comment: Sequential moment restrictions in panel data, Journal of Business & Economic Statistics 10, 20-26.

Cincera M., 1997, Patents, R&D, and technological spillovers at the firm level: some evidence from econometric count models for panel data, Journal of Applied Econometrics 12, 265-280.

Crépon B. and E. Duguet, 1997, Estimating the innovation function from patent numbers: GMM on count panel data, Journal of Applied Econometrics 12, 243-263.

Greene, W.H., 2005, LIMDEP 8.0, Econometric Software, Inc., Plainview, NY, USA.

Guggenberger P., 2005, Monte-carlo evidence suggesting a no moment problem of the continuous updating estimator, Economics Bulletin 3, 1-6.

Hall B. and C. Cummins, 2005, TSP 5.0, TSP International, Palo Alto, CA, USA.

Hansen L.P., 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

Hansen L.P., J. Heaton and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business & Economic Statistics 14, 262-280.

Hausman J., B. Hall and Z. Griliches, 1984, Econometric models for count data and an application to the patents-R&D relationship, Econometrica 52, 909-938.

Kelly M., 2000, Inequality and crime, The Review of Economics and Statistics 82, 530-539.

Kitazawa Y., 2000, TSP procedures for count panel data estimation, Kyushu Sangyo University.

Lancaster T., 2002, Orthogonal parameters and panel data, Review of Economic Studies 69, 647-666.

Kim J. and G. Marschke, 2005, Labor mobility of scientists, technological diffusion and the firm's patenting decision, The RAND Journal of Economics 36, 298-317.

Manning W.G., A. Basu and J. Mullahy, 2005, Generalized modeling approaches to risk adjustment of skewed outcomes data, Journal of Health Economics 24, 465-488.

Montalvo J.G., 1997, GMM estimation of count-panel-data models with fixed effects and predetermined instruments, Journal of Business and Economic Statistics 15, 82-89.

Mullahy J., 1997, Instrumental variable estimation of Poisson regression models: application to models of cigarette smoking behavior, Review of Economics and Statistics 79, 586-593.

Newey W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica 72, 219-255.

Newey W.K. and K.D. West, 1987, Hypothesis testing with efficient method of moments estimation, International Economic Review 28, 777-787.

Newey W.K. and F. Windmeijer, 2005, GMM with many weak moment conditions, cemmap Working Paper No CWP18/05.

Romeu A., 2004, *ExpEnd*: Gauss code for panel count data models, Journal of Applied Econometrics 19, 429-434.

Salomon R.M. and J.M. Shaver, 2005, Learning by exporting: new insights from examining firm innovation, Journal of Economics and Management Strategy 14, 431-460.

Santos Silva J.M.C. and S. Tenreyro, 2006, The log of gravity, The Review of Economics and Statistics 88.

Schellhorn M., 2001, The effect of variable health insurance deductibles on the demand for physician visits, Health Economics 10, 441-456.

Vera-Hernandez, A.M., 1999, Duplicate coverage and demand for health care. The case of Catalonia, Health Economics 8, 579-598.

Windmeijer F., 2000, Moment conditions for fixed effects count data models with endogenous regressors, Economics Letters 68, 21-24.

Windmeijer F., 2002, *ExpEnd*, a Gauss programme for non-linear GMM estimation of exponential models with endogenous regressors for cross section and panel data, cemmap Working Paper No. CWP14/02.

Windmeijer F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of Econometrics 126, 25-517.

Windmeijer F. and J.M.C. Santos Silva, 1997, Endogeneity in count data models: an application to demand for health care, Journal of Applied Econometrics 12, 281-294.

Wooldridge J.M., 1991, Multiplicative panel data models without the strict exogeneity assumption, MIT Working Paper No. 574.

Wooldridge J.M., 1997, Multiplicative panel data models without the strict exogeneity assumption, Econometric Theory 13, 667-678.

Wooldridge J.M., 1999, Distribution-free estimation of some nonlinear panel data models, Journal of Econometrics 90, 77-97.